The EMBO Journal Vol.2 No.12 pp.2375–2383, 1983

Sequence of the murine and human cellular *myc* oncogenes and two modes of *myc* transcription resulting from chromosome translocation in B lymphoid tumours

Ora Bernard, Suzanne Cory, Steven Gerondakis, Elizabeth Webb and Jerry M. Adams*

The Walter and Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital, Victoria 3050, Australia

Communicated by G. Klein
Received on 23 September 1983

The 15;12 chromosome translocation in murine plasmacytomas and the 8;14 in human Burkitt lymphomas often link the cellular *myc* oncogene to the locus for constant regions of immunoglobulin heavy chains (C_H locus). To clarify how and why c-*myc* translocation occurs, we have sequenced the mouse and human c-*myc* genes and correlated c-*myc* transcription with c-*myc* rearrangement. Both genes comprise three exons; the second and third encode the *myc* polypeptide, which is conserved between mammals and birds, particularly in its more basic C-terminal half. Southern blots showed that four of 12 Burkitt lines have c-*myc* linked near C_H switch regions and two near the joining region (J_H) locus. Hence, immunoglobulin recombination machinery may participate in translocation, although the common *myc* breakpoint region around exon 1 does not resemble a switch region. Tumours with breakpoints just 5′ to exon 1, or distant from c-*myc*, had normal c-*myc* mRNAs of 2.25 and 2.4 kb, which differ at their 5′ ends, while tumours with breakpoints within exon 1 or intron 1 had altered c-*myc* mRNAs (2.1–2.7 kb in Burkitt lines), initiated within intron 1. Both types of mRNAs probably yield the same polypeptide. Since the untranslocated c-*myc* allele was generally silent, translocation to the C_H locus must induce constitutive c-*myc* expression. The presence of c-*myc* mRNA in immortal but non-tumorigenic lymphoblastoid cell lines may implicate c-*myc* in an immortalization step.

Key words: B lymphocyte oncogenesis/human chromosome translocation t(8;14)/oncogene nucleotide sequence/altered *myc* transcription

Introduction

The cellular gene c-*myc*, which is homologous to the oncogene (v-*myc*) carried by avian retroviruses of the MC29 family, is strongly implicated in development of B lymphomas. In most chicken B lymphomas induced by avian leukosis virus, which lacks an oncogene, the retrovirus has integrated near c-*myc* and stimulated its expression (Hayward *et al.*, 1981). Moreover, two mammalian B lymphoid neoplasms, murine plasmacytomas and Burkitt lymphomas of man, exhibit specific chromosome translocations that involve c-*myc* (reviewed by Klein, 1981, 1983; Perry, 1983). The t(15;12) in most plasmacytomas (Ohno *et al.*, 1979) links c-*myc* on chromosome 15 to the immunoglobulin heavy (H) chain locus on chromosome 12, generally near switch recombination regions (Adams *et al.*, 1982, 1983; Shen-Ong *et al.*, 1982; Calame *et al.*, 1982; Harris *et al.*, 1982; Marcu *et al.*, 1983). The breakpoints on chromosome 15 usually occur within the 5′ portion of the c-*myc* transcriptional unit (Cory *et al.*, 1983b), accounting for the smaller c-*myc* mRNAs in plasmacytomas (Adams *et al.*, 1982, 1983; Shen-Ong *et al.*, 1982; Marcu *et al.*, 1983; Mushinski *et al.*, 1983). Similarly, the t(8;14) in Burkitt lymphomas often links the human c-*myc* gene on 8q24 to the H chain locus on chromosome 14 (Taub *et al.*, 1982; Adams *et al.*, 1983; Dalla-Favera *et al.*, 1983), but the effect on human c-*myc* transcription requires clarification. A minority of tumours display variant translocations, namely t(15;6) in mouse and t(8;2) and t(8;22) in man, which may link c-*myc* to light chain loci (Klein, 1983).

To clarify how c-*myc* translocation affects its expression, we have sequenced most of the mouse and human c-*myc* transcriptional units. The sequences considerably extend those reported recently for the mouse (Stanton *et al.*, 1983) and human (Colby *et al.*, 1983; Watson *et al.*, 1983b; Watt *et al.*, 1983) genes and demonstrate clearly that both genes comprise three exons within an ~5.2-kb transcriptional unit. Our hybridization studies show that the untranslocated allele in the tumours is usually silent, strongly implicating translocating in c-*myc* activation. We find that some translocations cleave the c-*myc* transcriptional unit while others do not; consequently tumours display alternative modes of transcription.

Results

Figure 1 shows that the murine and the human c-*myc* genes contain two exons homologous to v-*myc* (filled boxes) and an upstream exon (hatched), first identified in the mouse (Adams *et al.*, 1983). Using clones from murine plasmacytoma J558 and T lymphoma ST4 and two from a library of human embryonic DNA (Adams *et al.*, 1982, 1983), we determined sequences spanning all three exons; arrows delineate sequences obtained from different M13 phage clones. Human exons 2 and 3 (plus intron 2) have also been sequenced by Colby *et al.* (1983) and Watson *et al.* (1983b), a human cDNA clone including most of exon 1 by Watt *et al.* (1983), and murine sequences spanning exon 2 and most of exon 1 by Stanton *et al.* (1983).

Conservation of myc coding region

The *myc* coding region, which starts at nucleotide 16 of exon 2 (see below), has been substantially conserved in evolution. Figure 2 compares the murine (M) and human (H) nucleotide sequences and predicted amino acid sequences within exons 2 and 3. Figure 3 relates these amino acid sequences to those of chicken c-*myc* (Watson *et al.*, 1983a) and MC29 v-*myc*, which differ from each other by only five amino acids (Alitalo *et al.*, 1983) or seven amino acids (Reddy *et al.*, 1983). The two mammalian proteins are remarkably similar. In the 252 amino acid residues encoded by exon 2, the human gene has a one-codon deletion and one-codon insertion with respect to the mouse, but 92% of positions are identical, as are 94% in exon 3, which encodes 187 amino acids. The nucleotide sequences are 90% homologous in exon 2 and 88% in exon 3. With respect to chicken c-*myc* (Figure 3), the C-terminal portion (exon 3) is the more conserved: there are no insertions or deletions between the avian and mammalian polypeptides and 76% of amino acid residues are identical,

*To whom reprint requests should be sent.

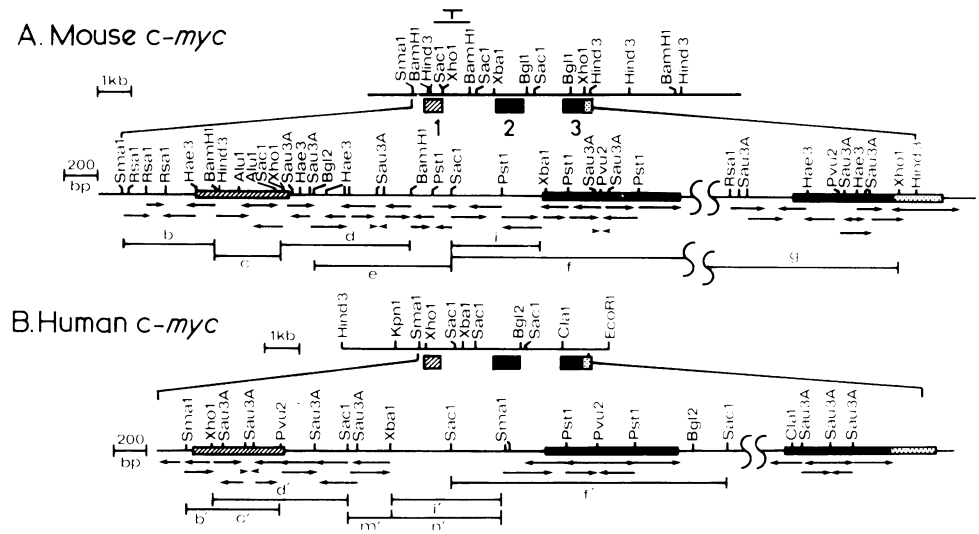

Fig. 1. Structure of the mouse and human c-*myc* genes and sequencing strategy. Boxes denote exons, filled boxes being homologous to v-*myc*. Exon 1 (hatched) is untranslated (see text) and the 3′-untranslated region of exon 3 is stippled. T denotes the murine region within which most translocations occur (see text). Restriction endonuclease sites on the expanded maps are those used in making M13 subclones for sequencing and/or as probes; a 700-bp gap has been left within mouse and human intron 2, which was not studied here. The murine sequence 5′ to the *Bam*HI site in exon I was determined from the ST4 clone, that in exon 1 from both the ST4 and J558 clones (no differences being found) and the rest of the sequence from the J558 clone.

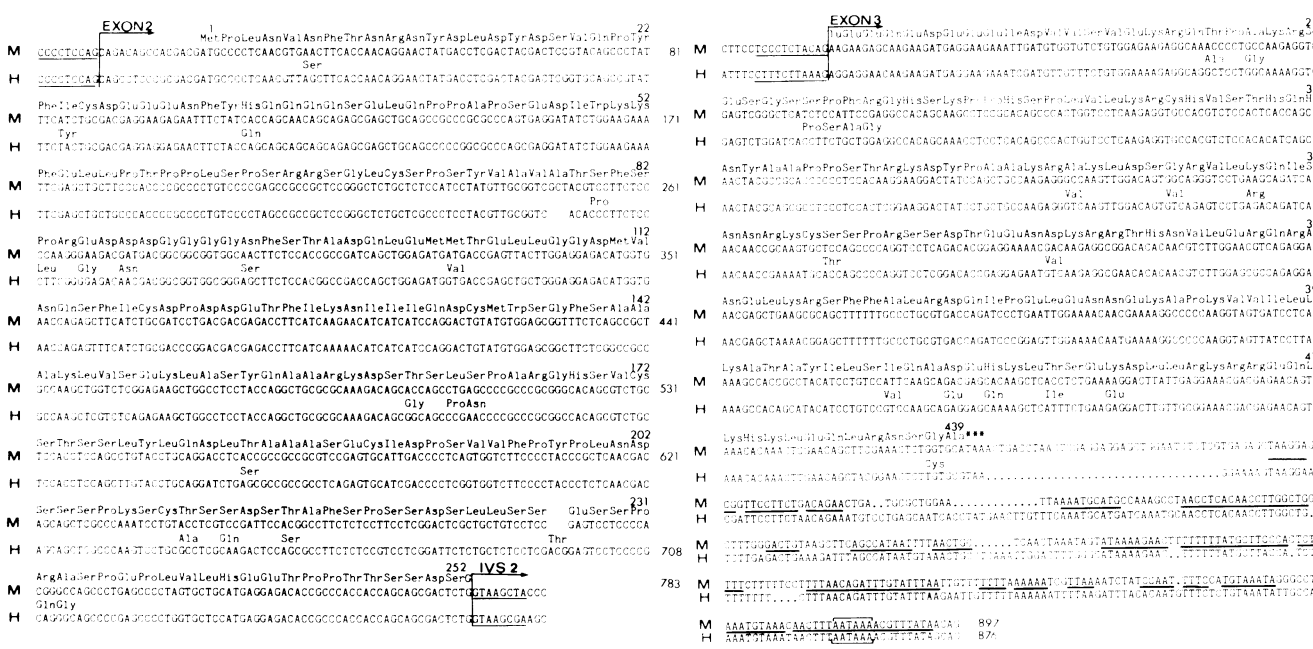

Fig. 2. Comparison of the murine (B) and human (H) *myc* coding and 3′-untranslated regions. Both nucleotide sequences are shown, but the human amino acid sequence is shown only where it differs from the mouse. Amino acids are numbered from the presumed initiating methionine in exon 2 (see text) through the entire gene, but the nucleotide sequences are numbered separately for the two exons, since we have not sequenced intron 2. Our sequence for human exons 2 and 3 agrees with that of Colby *et al.* (1983) and Watson *et al.* (1983b). The human 3′-untranslated sequence determined by Colby *et al.* (1983) is shown for comparison with that of the mouse. Dots denote 1-residue gaps introduced to maintain alignment and a line between the untranslated sequences indicates four or more identical residues. Poly(A) addition signals are bracketed.

whereas alignment of those encoded by exon 2 requires seven insertions or deletions (of 1 – 22 amino acids), with a resulting homology of ~70%. The blocks of conserved amino acids, marked by asterisks in Figure 3, presumably are particularly important to *myc* function(s).

In accordance with the hypothesis (Alitalo *et al.*, 1983) that ability of v-*myc* to bind DNA (Donner *et al.*, 1982) resides in its very basic C-terminal half, the c-*myc* region encoded by exon 3 has far more basic residues (e.g., 48 basic, 25 acidic in mouse) than that encoded by exon 2 (18 basic, 36 acidic). Both sections are fairly hydrophilic (49% of residues overall) and there is no long hydrophobic stretch like that for a transmembrane peptide. The mammalian polypeptides lack the central proline-rich region of v-*myc* (where 10 of 15 residues are proline), which might separate functional domains (Alitalo *et al.*, 1983).

Exon 1 of mouse and man comprises a long 5′-untranslated region

We previously detected a 5′ exon (now exon 1) in the

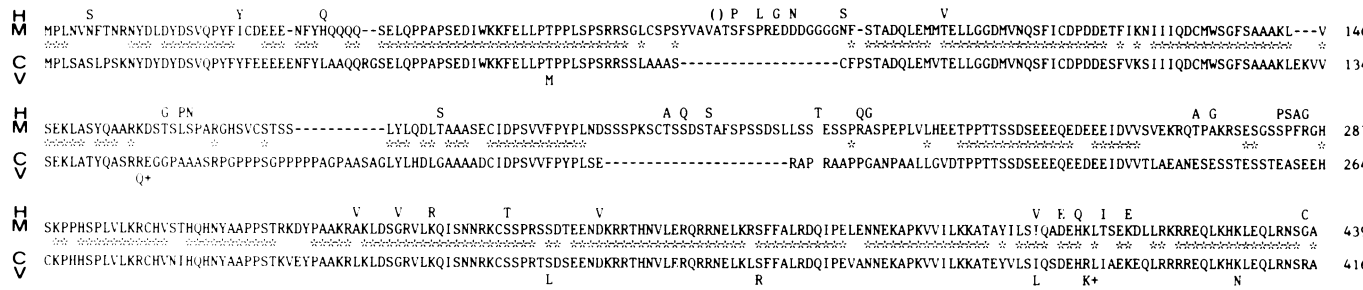

Fig. 3. Comparison of mouse (M) and human (H) c-*myc* amino acid sequences with that of avian c-*myc* (C) and v-*myc* (V). A vertical arrow marks the exon 2/ exon 3 boundary. Each human residue differing from the murine (see Figure 2) is indicated above it. Asterisks mark identity between mouse and chicken. The chicken c-*myc* sequence is from Watson *et al.* (1983a), and those residues which differ in v-*myc* are indicated below; the two marked + differ from c-*myc* in the sequence determined by Reddy *et al.* (1983) but not in that of Alitalo *et al.* (1983).

mouse (Adams *et al.*, 1983), subsequently confirmed by a cDNA clone (Stanton *et al.*, 1983), and one is also present in man (see below). Based upon the (conserved) splice sites defined by cDNA clones (Stanton *et al.*, 1983; Watt *et al.*, 1983), intron 1 is 1535 bp long in mouse and 1616 bp in man (Figure 4). Significantly, as concluded by Stanton *et al.* (1983), mouse exon 1 has no open reading frame in phase with exon 2 and that is also the case for human c-*myc*. Hence exon 1 of both species is an unusually long 5′-untranslated region. Lines between the sequences in Figure 4, which mark four or more conserved residues, indicate that much of exon 1 and intron 1 has been conserved. Homology in exon 1 is 69 or 76%, depending on whether each nucleotide insertion/deletion is included in the calculation, while that in intron 1 is 55 or 62%. Exon 1 is more conserved than might have been expected for an untranslated region, but the 3′-untranslated region (Figure 2) is as conserved (63% or 78%).

Eukaryotic promoters usually have a 'TATA box', consensus $TATA^{A}_{T}A^{A}_{T}$, located 20 – 30 bp 5′ of the mRNA cap (start) site, which is often an A residue within a pyrimidine-rich sequence (Breathnach and Chambon, 1981). Potential TATA boxes are bracketed or boxed in Figure 4. We present evidence below for two normal c-*myc* mRNA start sites and propose that these are near the horizontal arrows downstream from TATA boxes 1 and 2 in Figure 4, which are 160 bp apart. In line with this proposal, a 100-bp region spanning site 1 is conserved ~80%, as is that around site 2. Start site 1 gives an exon 1 of ~565 bp in mouse and ~555 in man. TATA boxes within intron 1, such as the conserved site 3, may generate some of the new transcripts in tumours (see below).

Breakpoint region of c-myc gene is not like a switch region

Since most mouse 15;12 and many human 8;14 translocations link sequences within or near c-*myc* exon 1 to a switch region, we searched the two c-*myc* sequences using the Diagon program (Staden, 1982) for homology with mouse S_α (Davis *et al.*, 1980) and S_μ (Takahashi *et al.*, 1980) sequences. None was detected. Hence, as hybridization results suggested (Adams *et al.*, 1982; Cory *et al.*, 1983b), homologous recombination is not the translocation mechanism. Moreover, the c-*myc* sequences (in either orientation) display only a few copies of the oligonucleotides repeated very frequently in all switch regions (for a review, see Marcu *et al.*, 1982). For instance, the 1.1-kb murine region T (Figure 1), where most breakpoints occur, contains (on its non-coding strand) only three GAGCT, two GGGGT and no YAGGTTG. The c-*myc* sequences also lack the 20 – 80 bp internal repeats characteristic of switch regions. Hence translocation does not result from similarity of the c-*myc* locus to a switch region.

Chromosome 8 breakpoints vary with respect to c-myc

Southern blots revealed that eight of 12 Burkitt lines had one c-*myc* allele rearranged (r in Table I). Germline (G) context was maintained for Daudi (8;14), JI (8;2) and Ly67 (8;22) in *Eco*RI, *Hind*III or *Bam*HI digests, and for Maku (8;22) in an *Eco*RI digest. From the c-*myc* locus map (Figure 5), we conclude for the first three lines that recombination occurred >17.8 kb 5′ to exon 1, or 7.6 kb 3′ to exon 3. Hence the 8q24 breakpoints in these and a number of other Burkitt lines (Taub *et al.*, 1982; Dalla-Favera *et al.*, 1983) are distant from c-*myc*. Of the eight lines with c-*myc* rearranged, all but W2BL displayed both products expected of a reciprocal exchange, since exon 2 and c-*myc* 5′-flanking regions were on different rearranged *Eco*RI fragments. Thus reciprocal exchange appears to be general, as in plasmacytomas (Cory *et al.*, 1983b).

We mapped seven breakpoints within a 4-kb region spanning exon 1 (Figure 5, top) by determining which probe from the c-*myc* locus labelled both reciprocal products, as shown in Figure 5A for selected lines. Two breakpoints mapped 5′ to exon 1 [r (5′) in Table I]. That in Ramos (lane 1) mapped to region *k′* since probe *k′* detected both the product bearing c-*myc* exons 2 and 3, which we term M (solid triangle), and the reciprocal (R) product (open triangle). Raji mapped further upstream because *k* labelled only the M product (not shown) while probe *o′* detected both M and R (lane 2). The five other breakpoints appear to split the c-*myc* transcriptional unit [r (tu)]. Those in Lou, Joy, WI and BM mapped to region *d′*, which spans exon 1 and part of intron 1, as shown for Joy (lane 3) and BM (lane 4). The order Lou-Joy-(W1, BM) was inferred from the relative labelling of the R and M product in each line with probes enriched for sequences near the 5′ or 3′ end of region *d′* (see Materials and methods). The LS breakpoint mapped within the same 400-bp *Xba*I-*Sac*I region of intron 1 as those in five other lines (Dalla-Favera *et al.*, 1983), since an exon 2 probe hybridized to a rearranged *Xba*I fragment but a normal *Sac*I fragment; hybridization to the reciprocal products confirmed this.

c-Myc translocates to different sites in the H chain locus

Linkage to the C_H locus was inferred for six of the seven lines (BM being uncertain) from hybridization results with probes from the human J_H-C_μ locus, kindly provided by Dr. T. Honjo. Recombination near C_μ was inferred for Joy, Ramos and W1, since c-*myc* exon 2 and C_μ probes hybridized

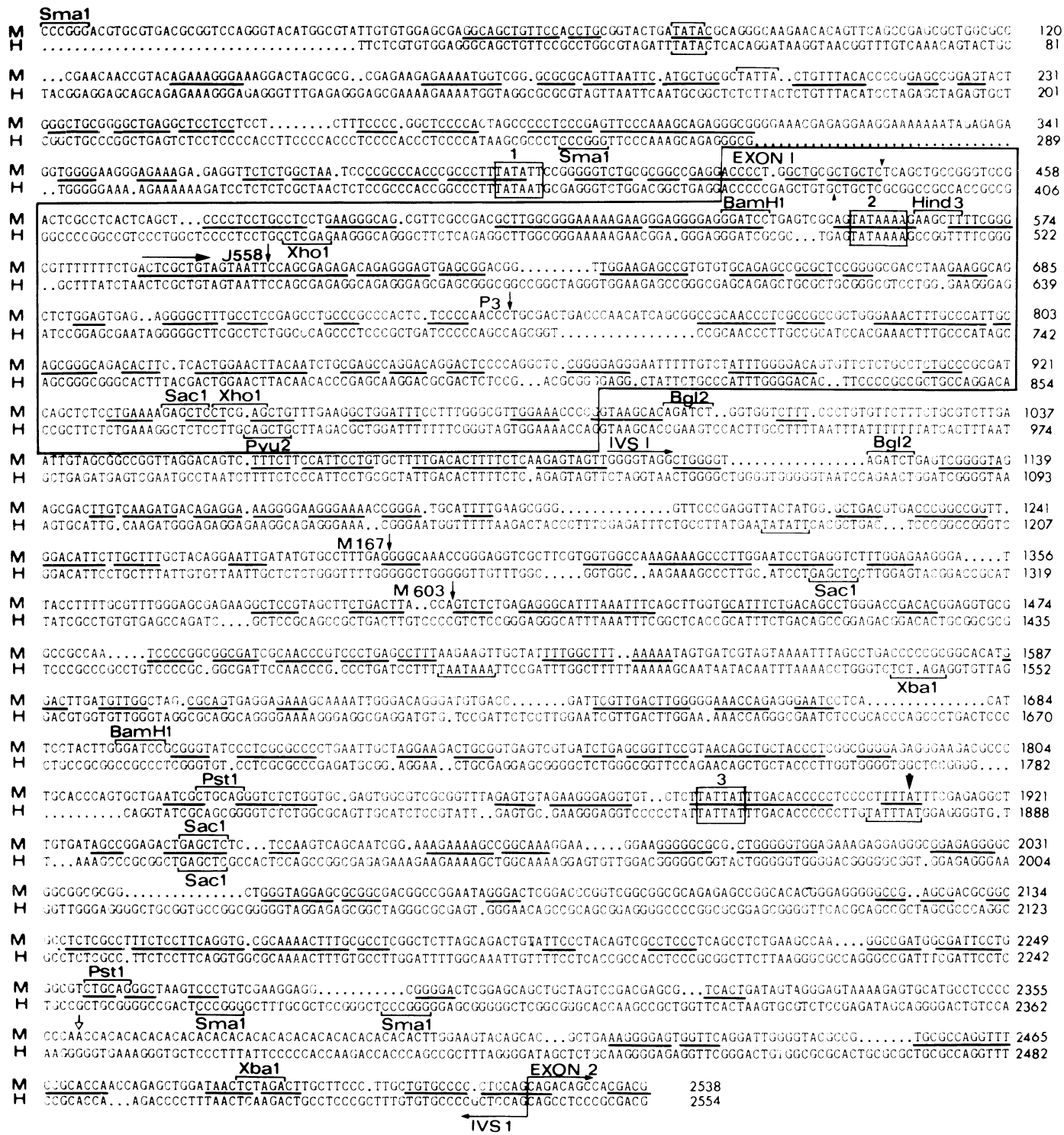

Fig. 4. Aligned sequences for c-*myc* exon 1 and intron 1 of mouse (M) and man (H). Dots denote one-residue gaps added to maintain alignment. Consensus TATA boxes are bracketed or boxed; the horizontal arrows downstream from boxes 1 and 2 are proposed start sites for the normal c-*myc* mRNAs (see text). Vertical arrowheads near the 5′ end of exon 1 mark the 5′-most residue in cDNA clones (Stanton *et al.*, 1983; Watt *et al.*, 1983); the splice points are based upon those clones. Restriction sites relevant to most probes in Figures 1, 5 and 7 are indicated. Plasmacytoma breakpoints for J558 (Adams *et al.*, 1982), P3 (S. Gerondakis, unpublished), M167 and M603 (Calame *et al.*, 1982) are indicated. An open arrowhead marks a start point proposed by Stanton *et al.*(1983) for plasmacytoma mRNA. The human sequence from the *Xba*I site (residue 1538) to *Sma*I (residue 2262) is from Colby *et al.* (1983) and is shown for comparison with our mouse sequence. The murine sequence, which includes some corrections within the 400-bp region of exon 1 that we published previously (Adams *et al.*, 1982), differs at a number of positions from that of Stanton *et al.* (1983) but some of these have been resolved (K. Marcu, personal communication).

to *Hind*III fragments of the same size (solid arrows, Figure 5B). Lou, Raji and LS instead involve linkage to a 'switched' C_H gene, one in which the $(V_HD_H)J_HS_\mu$ region has recombined 5′ to the C_H gene, as in the normal heavy chain switch. We inferred this because the M and/or R product in each hybridized to S_μ or J_H region probes (Figure 5C) but not to C_μ (not shown). Hamlyn and Rabbitts (1983) have recently identified the switched C_H gene linked to c-*myc* in Raji as

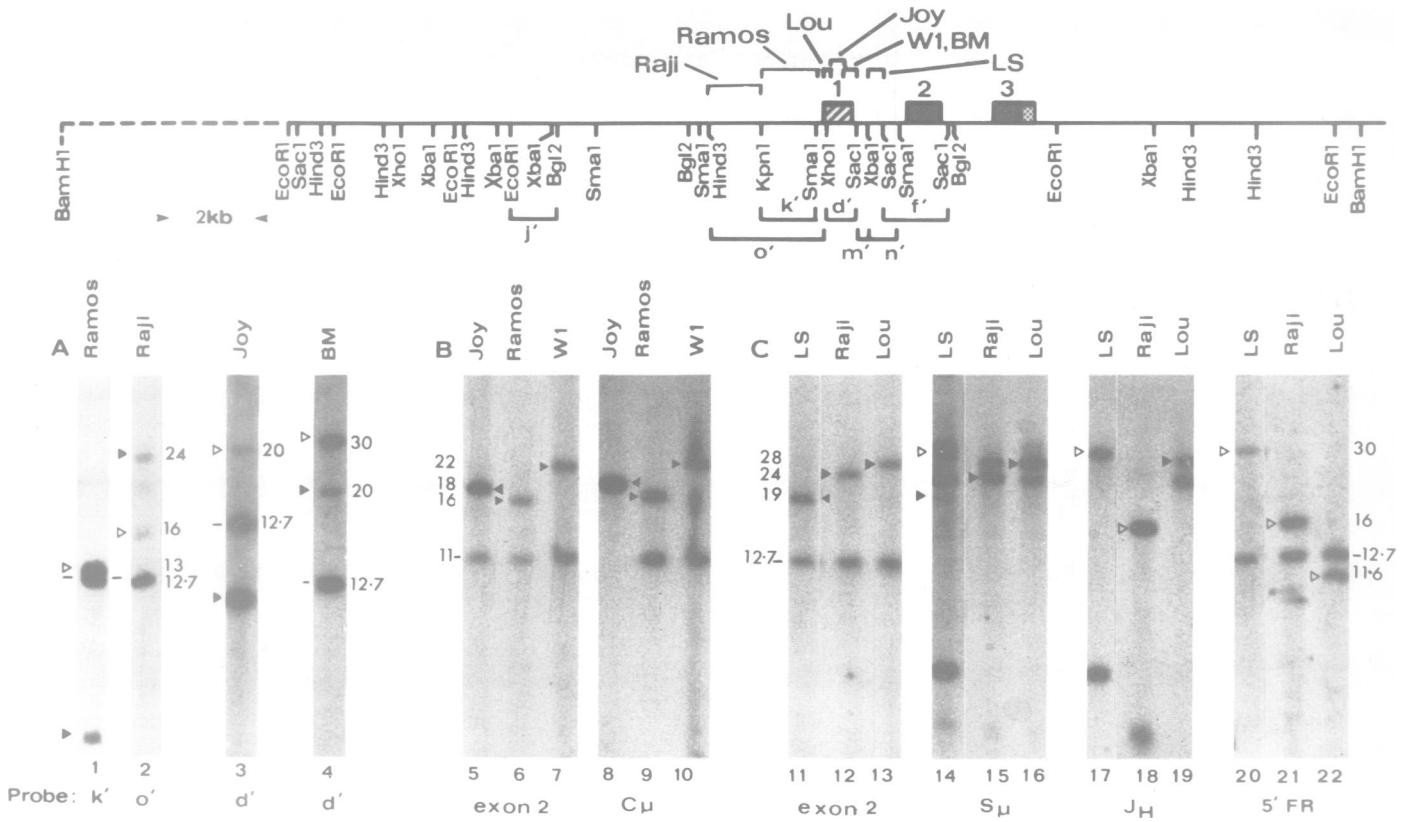

Fig. 5. Location of recombination points (upper brackets) near c-*myc* gene in Burkitt lymphomas. The broken portion of the restriction map was deduced from Southern blots, the remainder from cloned DNA (Adams *et al.*, 1983); the *Xba*I site within region *j'* is relatively resistant to digestion and was not detected previously. Probe regions are lettered here and above the J_H-C_μ map in Figure 6. In the Southern blots, **A** and **C** show *Eco*RI digests and **B**, *Hind*III digests. A filled triangle denotes the rearranged fragment bearing c-*myc* exons 2 and 3 (M product), an open triangle, the reciprocal product (R) bearing the c-*myc* 5'-flanking region (5' FR), and a bar, the germline c-*myc* fragment (see text and Figure 6). Sizes are in kb.

$C_{\gamma 1}$.

The structures of the reciprocal products in Figure 6 were deduced from blot analysis of *Eco*RI, *Hind*III, *Xba*I, and *Sac*I with c-*myc* probes (Figure 5, top) and J_H-C_μ probes (Figure 6, top). They are consistent with a simple cross-over and show that the c-*myc* and C_H genes have opposite orientation, as in other Burkitt lines (Dalla-Favera *et al.*, 1983; Hamlyn and Rabbitts, 1983) and all plasmacytomas so far described. The known orientation of the V_H-C_H locus on chromosome 14 suggests that the M product derives from the cytogenetically defined $14q^+$ and R from $8q^-$ (see Klein, 1983).

Significantly, the C_H locus breakpoint in four lines (Joy, Ramos, Raji and LS) lies within or near the S_μ region (stippled), implicating switch recombination enzyme(s) in the scission of chromosome 14. As in normal switch recombination, various sites within S_μ can recombine; for example,that in LS lies to the left of that in Joy. Switch regions however do not appear to be the only targets. Lou and W1 have breakpoints closer to the J_H locus (Figure 6), as did one line described by Dalla-Favera *et al.* (1983). V-D-J joining enzymes may be implicated in such lines.

An intact c-myc gene yields two mRNAs with different 5' ends

We found that tumours with breakpoints outside the normal transcriptional unit had c-*myc* mRNA like that in normal cells, comprised of two similar species, while those with breakpoints inside it yielded mRNA with new 5' ends. Our model (Figure 7) was derived by hybridization with the indicated c-*myc* probes, as illustrated by selected blots in Figure 8 of human RNA and in Figure 9 of murine RNA.

Burkitt lines with the c-*myc* transcriptional unit intact, such as Ly67 in Figure 8A, yielded ~2.25 and 2.4 kb c-*myc* mRNAs, as also observed by Hamlyn and Rabbitts (1983). These normal transcripts (N in Table I) were observed for lines with breakpoints just 5' to exon 1, which may have excised upstream regulatory elements (Ramos, Raji), and those with distant breakpoints (Ly67, Daudi, JI). The two mRNAs appear to differ at their 5' ends. Both hybridized to the 3' region of exon 1 (probe *c'*), as shown in Figure 8C, but only the 2.4-kb species hybridized to the probe *b'* from the 5' end of exon 1 (Figure 8B). Since this *Sma*I-*Xho*I fragment hybridized only weakly, the 2.4-kb mRNA probably starts 50–100 bp 5' of the *Xho*I site. Since box 1 in Figure 4 is the only TATA box within 400 bp 5' of that *Xho*I site, most likely the 2.4-kb mRNA starts 20–30 bp 3' to box 1, just upstream from the end of a human cDNA clone (Watt *et al.*, 1983). The 2.25-kb mRNA probably starts 160 bp downstream, after TATA box 2, the only consensus TATA box within region *c'*. Presumptive precursor RNAs are also evident in Daudi (Figure 8A).

Normal murine c-*myc* mRNA also partially resolved into 2.4- and 2.25-kb species and only the 2.4-kb was labelled by probe *b* from the 5' end of exon 1, as shown for T lymphoma W242 in Figure 9A. Hence these mRNAs also differ at their 5' ends, and most likely also initiate at the conserved sites 1 and 2 (Figure 4).

Table I. *Myc* rearrangement and transcription

Line[a]		*myc* gene[b]	*myc* mRNA[c]	RNA level[d]
Burkitt				
Daudi	(8;14)	G	N	50
JI	(8;2)	G	N	55
Ly67	(8;22)	G	N	15
Maku	(8;22)	G	nd	nd
Ramos	(8;14)	r (5′)	N	45
Raji	(8;14)	r (5′)	N[e]	50
LouBL36	(8;14)	r (tu)[f]	N[f]	50
JOY	(8;14)[g]	r (tu)	A	55
W1BL	(8;14)	r (tu)	A	40
BM		r (tu)	A	40
LSBL29	(8;14)	r (tu)	A	50
W2	(8;14)	r (5′)	nd	nd
Other				
HL60		G	2.25	50
Lou LCL		G	N	40
W1 LCL		G	N	20
Bou LCL		G	N	20
Lat LCL		nd	N	~25
W2 LCL		G	N	5–10

nd, not determined.

[a]See Materials and methods regarding the origin of the lines. The translocation, where reported, is indicated in parenthesis; those in JI and Ly67 were confirmed by R. Hutchinson and M. Garson of St. Vincent's Hospital (personal communication).

[b]Context of the *myc* gene: G denotes germline context for both alleles in *Eco*RI, *Hind*III and *Bam*HI digests except for Maku and Bou LCL, which were checked only with *Eco*RI; r (5′), rearrangement in 5′-flanking region; r (tu), rearrangement that alters the transcriptional unit for one allele. The rearrangement in W2 BL lies 5′ to exon 1, within the *Eco*RI fragment.

[c]N denotes the normal 2.25- and 2.4-kb mRNAs and A the altered mRNAs of 2.1–2.7 k.

[d]An order-of-magnitude estimate for molecules/cell, based upon the signal relative to that of c-*myc* DNA fragments run on the same gels and an average recovery of ~200 μg poly(A)$^+$ RNA from 10^9 cells. Relative levels should be accurate within a factor of 2 since several gels gave similar results.

[e]Raji mRNA involves the normal promoters, since it clearly bears sequences from exon 1 (both b′ and c′), but appeared to run mainly as a 2.3-kb mRNA; since Raji bears a deletion in the vicinity of exon 2 (see legend to Figure 6), conceivably part of exon 2 is deleted.

[f]See text regarding this apparent discrepancy.

[g]Cytogenetic data is not available but blot data indicates c-*myc*-C_H linkage (Figure 6).

The relative abundance of the two mRNAs varied in different murine and human cell lines. Daudi, Ramos and Raji had 2–3 times more of the larger species, while Ly67 had equivalent amounts (Figure 8A), and only the 2.25-kb mRNA was seen in murine B lymphoma W231 (Figure 9A) or, in longer runs, for the human promyelocytic line HL60. This variation suggests that the two normal c-*myc* promoters can be regulated independently.

Scission of the c-myc transcriptional unit activates new promoters

Lines in which translocation has clearly split the normal c-*myc* transcriptional unit display altered transcripts (A in Table I). Burkitt lines Joy, W1, BM and LS have at least

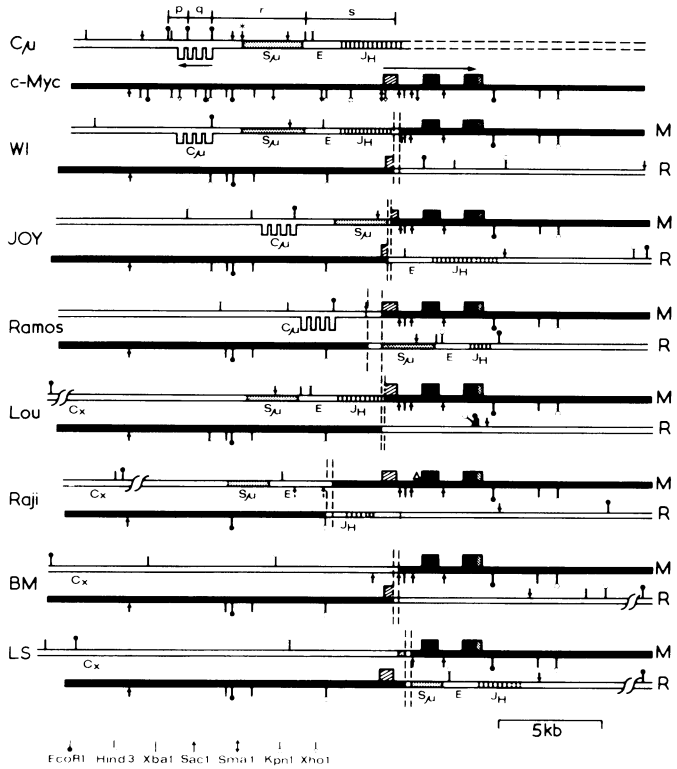

Fig. 6. Reciprocal translocation products in Burkitt lymphomas. The human C_μ (Ravetch *et al.*, 1981; Mignone *et al.*, 1983) and c-*myc* (Adams *et al.*, 1983) loci are orientated so that c-*myc* transcription would be from left to right, and C_μ from right to left. For the reciprocal translocation products M and R (see text), restriction sites derived from the c-*myc* locus (solid) are shown below the bar and sites from chromosome 14 (open) above it. Recombinant points were localized to the regions indicated by vertical broken lines by hybridization analysis (see text) and restriction mapping. The Lou, Raji and BM maps are foreshortened by 10 kb at the indicated break. E denotes an enhancer placed by analogy to that in the mouse (see text). An asterisk in the S_μ region (top) indicates a *Sac*I site absent from some individuals (Mignone *et al.*, 1983); the *Hind*III site to its right may also be polymorphic because it was absent from the W1 and Joy products. In LS, the *Sac*I site near exon 1 is absent from the rearranged allele, perhaps reflecting polymorphism. The Raji M product appears to contain a small deletion (△) within the *Sac*I fragment spanning exon 2, because, as well as the normal size fragments, probe *f′* detects *Sac*I and *Xba*I fragments ~0.3 kb smaller; Raji may have other secondary rearrangements because two rearranged *Sac*I fragments are labelled by probe *k′* (broken arrows mark the corresponding sites).

three or four c-*myc* species from 2.1 to 2.7 kb and some displayed larger species as well (Figure 8A). Unlike normal c-*myc* mRNAs, these mRNAs were not labelled by an exon 1 probe (Figure 8C), except for a low level in BM (see below). Significantly, probe *i′* from the 3′ half of intron 1 hybridized to these mRNAs (Figure 8D) but only very slightly to mRNA from lines like Ramos, Daudi and JI. Thus these mRNAs have new 5′ ends.

We surmise that the ~2.1–2.7 kb mRNAs reflect activation of cryptic promoters(s) within the 3′ half of intron 1 and/or altered splicing (Figure 7B). Since most of these mRNAs are common to all four lines, all probably initiate 3′ to the LS breakpoint (Figure 5). If so, since most are labelled by probe *i′* (*Xba*I-*Sma*I) and thus must include >100 bp 5′ of the *Sma*I site, the start point(s) probably fall between 375 and 1000 bp 5′ to exon 2. The 4-kb Joy RNA might instead initiate within the C_μ-S_μ region (Figure 6) and be spliced to exon 2; it cannot be merely an unspliced precursor because the intron 1 probe did not label it (Figure 8D).

The untranslocated c-myc allele is usually silent

Since the exon 1 probe labelled no c-*myc* mRNA in Joy, W1 or LS (Figure 8C), the untranslocated c-*myc* allele must be silent. The faint ~2.4 species in BM might reflect low level synthesis from that allele, or from the reciprocal translocated chromosome, but comparison with Figure 8A indicates that the rearranged BM c-*myc* gene is ~5-fold more active. We showed that six of seven plasmacytomas also lacked the normal transcripts (Adams *et al.*, 1983) and have found that to hold for six others (E109, T609, WEHI 267, HOPC 1, MOPC 41 and BAL1131). These results suggest that c-*myc* expression was repressed in the normal B cell from which the tumour derived and that translocation activated its expression.

Lou BL may be an exception. Its breakpoint (Figures 5, 6) appears to lie 3′ of the proposed promoter site 1, yet the mRNAs seem to be of normal type (Table I). They may derive from the untranslocated allele, or H locus promoters may have generated c-*myc* mRNAs of that size.

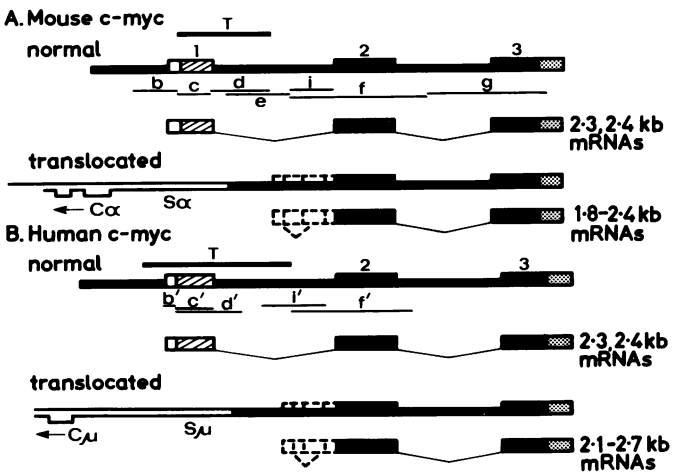

Fig. 7. Model for two modes of c-*myc* transcription following translocation of the mouse and human c-*myc* gene. In each species, the normal transcriptional unit, which gives the ~2.25- and 2.4-kb mRNAs, apparently involves two promoters, as indicated by the open box at the 5′ end of the primary transcript (see text). T denotes the most common translocation region. Translocations that disrupt the normal transcriptional unit uncover cryptic promoters within the former intron 1, generated new c-*myc* mRNAs (see text). The start points within intron 1 are not yet known. The lettered probe regions are defined by restriction sites in Figure 1.

Two modes of c-myc transcription in plasmacytomas

While the vast majority of plasmacytomas have an altered transcriptional unit, T1033 has a breakpoint just 5′ to exon 1 (Cory *et al.*, 1983b) and contains only the normal mRNAs, as assessed with a probe from exon 3 (Figure 9B) or exon 1 (not shown). Some plasmacytomas induced by Abelson murine leukaemia virus are also of this class (Mushinski *et al.*, 1983; our unpublished results). Hence altered c-*myc* mRNAs are not obligatory in plasmacytomas. Lines with a disrupted c-*myc* transcriptional unit, such as EPC 109, TEPC 609 and SAMM 368 in Figure 9, contain several c-*myc* mRNAs, 1.8–2.4 kb long, and some have ~1.2–1.5 kb species (Adams *et al.*, 1982, 1983). All the new mRNAs hybridized to probe *i* from the 3′ half of intron 1 (Figure 9C), whereas probe *e* (*Bgl*II-*Sac*I) from the 5′ section of intron 1 hybridized weakly to 2.4- and 2.6-kb species (Figure 9D). Hence the 2.4-kb mRNA appears to start 5′ to the intron *Sac*I site; a likely start point would be that arrowed 3′ to TATA box 3 in Figure 4, since mRNA bearing the rest of intron 1 (616 bp) + exons 2 and 3 would be ~2.4 kb long. The 1.8–2.2-kb species probably initiate closer to exon 2. Apparent start sites near exon 2 (Stanton *et al.*, 1983) might account for the 2.0-kb mRNA. Hybridization results suggesting that a 1.85-kb species might differ in its 3′-untranslated region (Adams *et al.*, 1983) did not hold for the mRNAs in Figure 9 and may have reflected cross-hybridization to residual 18S rRNA of that size.

Non-tumorigenic lymphoblastoid cells also express c-myc mRNA

Table I shows that most Burkitt lines had a very similar level of c-*myc* mRNA, of the order of 50 molecules/cell, although Ly67 (Figure 8E) had 3-fold less. Since the promyelocytic line HL60 (Figure 8E) had the typical Burkitt level (Table I), the translocations induce a transcription level comparable with the 20-fold gene amplification in HL60 (Collins and Groudine, 1982). No lymphoid line had an amplified c-*myc* gene. We also examined lymphoblastoid cell lines (LCL), B cells immortalized by Epstein-Barr (EB) virus; such lines have a normal karyotype and lack c-*myc* rearrangement

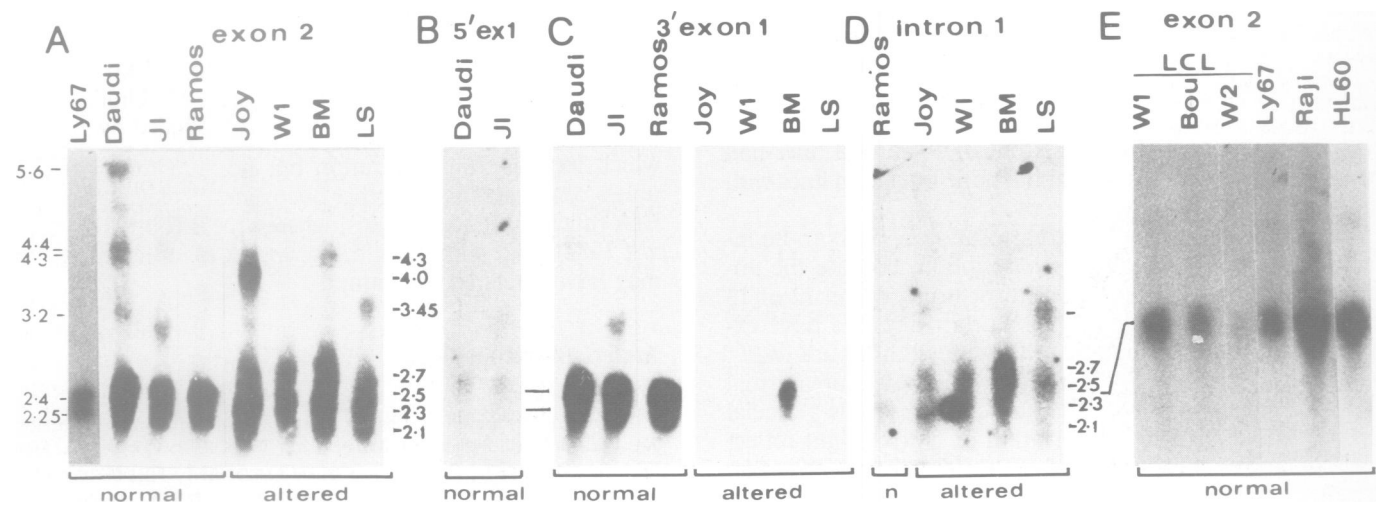

Fig. 8. c-*Myc* RNA in human cell lines. Blots of 5 μg of total cellular poly(A)$^+$ RNA, fractionated by electrophoresis on 1.5% agarose, were hybridized with probes indicated in Figure 7B and defined in Figure 1: **A** and **E**, exon 2 probe *f* (*Sac*I-*Sac*I); **B**, exon 1 probe *b′* (*Sma*I-*Xho*I); **C**, exon 1 probe *c′* (*Xho*I-*Pvu*II); and **D**, intron 1 probe *i′* (*Xba*I-*Sma*I). Probes *b′*, *c′* and *i′* were M13 probes that hybridize only to sequences from the c-*myc* mRNA strand.

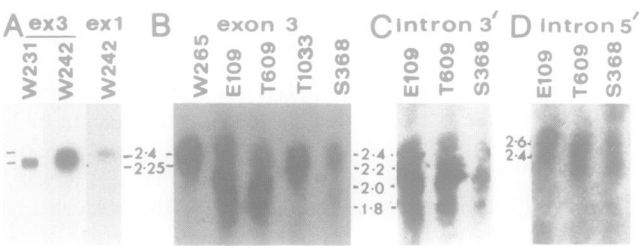

Fig. 9. c-*Myc* RNA in murine cell lines. Blots of poly(A)$^+$ RNA, fractionated as in Figure 8, were hybridized with the following probes (see Figure 7): **A**, exon 3 probe *g* (*Sac*I-*Xho*I) and exon 1 probe *b* (*Bam*HI-*Xho*I) compared on the same filter; **B**, exon 3 probe *g*; **C**, probe i (*Sac*I-*Xba*I) for the 3′ end of intron 1; **D**, probe *e* (*Bgl*II-*Sac*I) for the 5′ end of intron 1. Myeloid line WEHI-265 and T lymphoma 242 have no c-*myc* rearrangement.

(Table I). Of five such lines (three shown in Figure 8E), four had a c-*myc* mRNA level within the Burkitt range (Table I). These results, like those of Maguire *et al.* (1983) but unlike those of Nishikura *et al.* (1983), suggest that c-*myc* is usually expressed in lymphoblastoid cell lines. Hence c-*myc* may be implicated in immortalization (see Discussion).

Discussion

Our correlation of c-*myc* rearrangement and transcription in 10 Burkitt lymphomas (Table I) reveals marked similarities with murine plasmacytomas. In both species, translocation induces one of two modes of c-*myc* transcription: breakpoints within the 5′ end of the normal c-*myc* transcriptional unit trigger new mRNAs, while those just outside it or a large distance away yield the same two mRNAs (2.25 and 2.4 kb) as an untranslocated gene. While most plasmacytomas display the altered transcriptional mode, Burkitt lines of both classes abound, if our data is taken with that of Taub *et al.* (1982) and Dalla-Favera *et al.* (1983). Hence either mode of c-*myc* transcription is compatible with tumour formation.

The altered c-*myc* mRNAs in both species bear 5′-terminal sequences from the 3′ half of intron 1 (Figure 7). We surmise that the multiple new species reflect multiple promoters within that region, although splicing could be involved as well. Whether such promoters are ever used in a normal tissue, as with two tissue-specific promoters for α-amylase (Hagenbüchle *et al.*, 1981), is not known. The new mRNAs probably encode the same polypeptide as the normal transcripts, since intron 1 in both species lacks an ATG that would be in phase with exon 2. While the new 5′-untranslated region might allow altered translational control, that cannot be obligatory because five or six out of 10 Burkitt lines had only normal mRNAs, usually at the same level as in lines with altered mRNAs (Table I).

The major consequence of translocation appears to be induction of constitutive c-*myc* transcription, because the untranslocated c-*myc* allele was silent, or nearly so, in 12 of 13 plasmacytomas and four of five Burkitt lines we have examined, as in several somatic cell hybrids (Nishikura *et al.*, 1983). We infer that c-*myc* is turned off in the mature normal B lymphocyte (or plasma cell) and its expression is a crucial step towards malignancy. Activation of c-*myc* might reflect gross opening of chromatin on fusion with open regions of an immunoglobulin locus, and/or enhancers within that locus. The presence in certain murine and human tumours of transcripts from the opposite strand of c-*myc* intron 1 (J. Adams, unpublished) would fit with an enhancer, which can stimulate transcription in both directions. The enhancer between J_H and S_μ (Gillies *et al.*, 1983; Banerji *et al.*, 1983), indicated as E in Figure 6, would be linked to c-*myc* in Burkitt lines W1 and Lou, but not in the others, nor in any plasmacytoma so far described; however, other enhancers might exist closer to the different C_H genes.

Why does c-*myc* usually translocate to the H chain locus rather than another region of active chromatin? We imagine that breaks in various chromosomes provide ends that can be ligated to the H chain locus by enzymes transiently active in the B cell for immunoglobulin DNA rearrangement, in particular for switch recombination. Translocations that activated a relevant oncogene would be highly selected. Since the breakpoint region spanning c-*myc* exon 1 does not resemble a switch region, the high preference for breaks there probably mainly reflects selection for removal of normal c-*myc* regulatory elements but retention of the cryptic intron promoters.

The role of EB virus in Burkitt lymphoma remains obscure (see Klein, 1981). *Myc* activation in lines reported to lack EB markers (Ramos, Joy, BM) did not differ from the others (Table I). Any EB viral insertion near c-*myc* is unlikely because our mapping (Figure 6) would have revealed even a 0.5-kb insert within a 30-kb region spanning c-*myc* in 12 Burkitt lines or in several lymphoblastoid cell lines. EB virus might simply increase the number of B cells capable of undergoing translocation.

Myc activation is unlikely to suffice for full neoplasia. DNA from these tumours contains a fibroblast-transforming gene (Lane *et al.*, 1982) distinct from c-*myc* (Crews *et al.*, 1982; Adams *et al.*, 1983). That gene appears to be B-*lym* in several Burkitt lines (Diamond *et al.*, 1983) and N-*ras* in another (Murray *et al.*, 1983). *Myc* expression in immortal but non-tumourigenic lymphoblastoid cell lines (Table I) suggests that c-*myc* immortalizes but that a complementary oncogene is required for full B cell neoplasia. In line with that conclusion, the promyelocytic line HL60 contains activated *myc* and N-*ras* (Murray *et al.*, 1983) and combined action of *myc* and certain other oncogenes permits neoplastic conversion of primary fibroblast cells (Land *et al.*, 1983).

To provide clues to *myc* function, we searched the Dayhoff sequence library for proteins similar to each 10-residue segment of mouse c-*myc* but found only weak homology with diverse nucleic acid-binding proteins. Lack of resemblance to most oncogenes supports the notion that *myc* belongs to a separate oncogene class. However, R. Ralston and J.M. Bishop (personal communication) noted slight homology between *myc* and the E1A gene of adenovirus, an oncogene which induces immortalization but not full tumorigenicity, unless complemented by other oncogenes (Ruley, 1983). Since E1A functions in the nucleus, where *myc* is found (Donner *et al.*, 1982), and stimulates transcription of other genes, *myc* may have a related function.

Materials and methods

Tumour cell lines

The origin of the murine plasmacytomas and of most Burkitt lymphomas has been detailed elsewhere (Adams *et al.*, 1982, 1983; Cory *et al.*, 1983b). D. Moss (Queensland Institute of Medical Research) kindly provided Burkitt lines JOY and BM which he established from Australian patients and LS (IARC/BL29), established from a Reunion patient by G.M. Lenoir (Lyon).

Hybridization analysis

Total cellular poly(A)$^+$ RNA was isolated by the procedure of Gonda *et al.*

(1982); subsequent size fractionation on agarose gels, transfer to modified paper and hybridization were as described by Alwine *et al.*, (1979). DNA extraction and Southern blotting have been detailed (Cory *et al.*, 1983b). ^{32}P-labelled probes were prepared by nick-translation of restriction fragments, or by synthesis from M13 recombinant phage, using either the primer for sequence analysis or the 'hybridization' primer (Hu and Messing, 1982). Synthesis from the sequencing primer sometimes gave short products, enriching sequences near the primer; for example, the region *d'* insert in two orientations (i.e., synthesizing from the 5' or 3' restriction site) yielded probes enriched for its 5' or 3' end.

Nucleotide sequence analysis

Fragments of murine and human c-*myc* clones described previously (Adams *et al.*, 1983) were inserted into M13 phage vectors mp8, mp9, mp10 and mp11 (Messing, 1981), and sequences derived by the dideoxynucleotide method (Sanger *et al.*, 1980). Initially [^{32}P]dATP was used but much longer sequences were obtained using [^{35}S]dATP (Amersham). The sequencing strategy is outlined in Figure 1.

Acknowledgements

We thank Jan Mitchell, Scott Poynton, Filippa Brugliera and Linda Byrt for skilled assistance; A. Harris, M. Potter, D. Moss and G. Lenoir for cell lines; R. Hutchinson and M. Garson for cytogenetic analysis; and T. Honjo for J_H-C_μ probes. This work was supported by the National Health and Medical Research Council (Canberra), the U.S. National Cancer Institute (Grant CA12421), the American Heart Association and the Drakensberg Foundation.

References

Adams,J.M., Gerondakis,S., Webb,E., Mitchell,J., Bernard,O. and Cory,S. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 6966-6970.

Adams,J.M., Gerondakis,S., Webb,E., Corcoran,L.M. and Cory,S. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 1982-1986.

Alitalo,K., Bishop,J.M., Smith,D.H., Chen,E.Y., Colby,W.W. and Levinson,A.D. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 100-104.

Alwine,J.C., Kemp,D.J. and Stark,G.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5350-5354.

Banerji,J., Olson,L. and Schaffner,W. (1983) *Cell,* **33**, 729-740.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.,* **50**, 349-383.

Calame,K., Kim,S., Lalley,P., Hill,R., Davis,M. and Hood,L. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 6994-6998.

Colby,W.W., Chen,E.Y., Smith,D.H. and Levinson,A.D. (1983) *Nature,* **301**, 722-725.

Collins,S.J. and Groudine,M. (1982) *Nature,* **298**, 679-681.

Cory,S., Adams,J.M., Gerondakis,S.D., Miller,J.F.A.P., Gamble,J., Wiener,R., Spira,J. and Francke,U. (1983a) *EMBO J.,* **2**, 213-216.

Cory,S., Gerondakis,S. and Adams,J.M. (1983b) *EMBO J.,* **2**, 697-703.

Crews,S., Barth,R., Hood,L., Prehn,J. and Calame,K. (1982) *Science (Wash.),* **218**, 1319-1321.

Dalla-Favera,R., Martinotti,S., Gallo,R.C., Erikson,J. and Croce,C.M. (1983) *Science (Wash.),* **219**, 963-968.

Davis,M., Kim,S. and Hood,L. (1980) *Science (Wash.),* **209**, 1360-1365.

Diamond,A., Cooper,G.M., Ritz,J. and Lane,M.-A. (1983) *Nature,* in press.

Donner,P., Greiser-Wilke,I. and Moelling,K. (1982) *Nature,* **296**, 262-265.

Gillies,S., Morrison,S.L., Oi,V.T., and Tonegawa,S. (1983) *Cell,* **33**, 717-728.

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

Hagenbüchle,G., Tosi,M., Schibler,U., Borey,R., Wellauer,P.K. and Young, R.A. (1981) *Nature,* **289**, 643-646.

Hamlyn,P. and Rabbitts,T. (1983) *Nature,* **304**, 135-139.

Harris,L.J., Lang,R.B. and Marcu,K.B. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 4175-4179.

Hayward,W., Neel,B.G. and Astrin,S. (1981) *Nature,* **290**, 475-480.

Hu,N. and Messing,J. (1982) *Gene,* **17**, 271-277.

Klein,G. (1981) *Nature,* **294**, 313-318.

Klein,G. (1983) *Cell,* **32**, 311-313.

Land,H., Parada,L. and Weinberg,R. (1983) *Nature,* **304**, 596-602.

Lane,M.A., Sainten,A. and Cooper,G. (1982) *Cell,* **28**, 873-880.

Maguire,R.T., Robins,T.S., Thorgeirsson,S.S. and Heilman,C.A. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 1947-1950.

Marcu,K.B., Lang,R.B., Stanton,L.W. and Harris,L.J. (1982) *Nature,* **298**, 87-89.

Marcu,K.B., Harris,L.J., Stanton,L.W., Erikson,J., Watt,R. and Croce,C. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 519-524.

Messing,J. (1981) in Walton,A. (ed.), *Third Cleveland Symposium on Macromolecules: Recombinant DNA,* Elsevier, Amsterdam, pp. 143-153.

Mignone,N., Feder,J., Cann,H., van West,B., Hwang,J., Takahashi,N., Honjo,T., Piazza,A. and Cavalli-Storza,L. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 467-471.

Murray,M.J., Cunningham,J.M., Parada,L.F., Dantry,F., Lebowitz,P. and Weinberg,R.A. (1983) *Cell,* **33**, 749-756.

Mushinski,J.F., Bauer,S.R., Potter,M. and Reddy,E.P. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 1073-1077.

Nishikura,K., ar-Rushidi,A., Erikson,J., Watt,R., Rovera,G. and Croce, C.M. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 4822-4827.

Ohno,S., Babonits,M., Wiener,F., Spira,J., Klein,G. and Potter,M. (1979) *Cell,* **18**, 1001-1007.

Perry,R.P. (1983) *Cell,* **33**, 647-650.

Ravetch,J., Siebenlist,U., Korsmeyer,S., Waldmann,T. and Leder,P. (1981) *Cell,* **27**, 583-591.

Reddy,E.P., Reynolds,K., Watson,D.K., Shulz,D.R., Lautenberger,J. and Papas,T.S. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 2500-2504.

Ruley,H.E. (1983) *Nature,* **304**, 602-606.

Sanger,F., Coulson,A., Barrell,B., Smith,A. and Roe,B. (1980) *J. Mol. Biol.,* **143**, 161-173.

Shen-Ong,G., Keath,E., Piccoli,S.P. and Cole,M. (1982) *Cell,* **31**, 443-452.

Staden,R. (1982) *Nucleic Acids Res.,* **10**, 2951-2961.

Stanton,L.W., Watt,R. and Marcu,K.B. (1983) *Nature,* **303**, 401-405.

Takahashi,N., Kataoka,T. and Honjo,T. (1980) *Gene,* **11**, 117-127.

Taub,R., Kirsch,I., Morton,C., Lenoir,G., Swan,D., Tronick,S., Aaronson, S. and Leder,P. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 7837-7841.

Watson,D.K., Reddy,E.P., Duesberg,P.H. and Papas,T.S. (1983a) *Proc. Natl. Acad. Sci. USA,* **80**, 2146-2150.

Watson,D.K., Psallidopoulos,M.C., Samuel,K.P., Dalla-Favera,R. and Papas,T.S. (1983b) *Proc. Natl. Acad. Sci. USA,* **80**, 3642-3645.

Watt,R., Stanton,L.W., Marcu,K.B., Gallo,R.C., Croce,G.M. and Rovera, G. (1983) *Nature,* **304**, 725-727.