

The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins

Isobel A.Wadman, Hirotaka Osada¹,
Gerald G.Grütz, Alan D.Agulnick²,
Heiner Westphal², Alan Forster and
Terence H.Rabbitts³

MRC Laboratory of Molecular Biology, Hills Road,
Cambridge CB2 2QH, UK and ²Laboratory of Mammalian Genes &
Development, National Institute of Child Health & Human
Development, National Institutes of Health, Bethesda, MD 20892,
USA

¹Present address: Laboratory of Ultrastructure Research, Aichi Cancer
Center Research Institute, Chikusa-ku, Nagoya 464, Japan

³Corresponding author

I.A.Wadman and H.Osada contributed equally to this work

The LIM-only protein Lmo2, activated by chromosomal translocations in T-cell leukaemias, is normally expressed in haematopoiesis. It interacts with TAL1 and GATA-1 proteins, but the function of the interaction is unexplained. We now show that in erythroid cells Lmo2 forms a novel DNA-binding complex, with GATA-1, TAL1 and E2A, and the recently identified LIM-binding protein Ldb1/NLI. This oligomeric complex binds to a unique, bipartite DNA motif comprising an E-box, CAGGTG, followed ~9 bp downstream by a GATA site. *In vivo* assembly of the DNA-binding complex requires interaction of all five proteins and establishes a transcriptional transactivating complex. These data demonstrate one function for the LIM-binding protein Ldb1 and establish a function for the LIM-only protein Lmo2 as an obligatory component of an oligomeric, DNA-binding complex which may play a role in haematopoiesis.

Keywords: haematopoiesis/Ldb1/LIM/Rbtn2/transcription

Introduction

The *LMO2* gene (previously known as *RBTN2* or *TTG2*) and the *TAL1* gene (previously known as *SCL* or *TCL5*), are both involved with the development of T-cell acute lymphoblastic leukaemia (T-ALL) (reviewed in Rabbitts, 1994). *LMO2* and *TAL1* were identified via chromosomal translocations involving chromosome 11p13 in T-cell acute leukaemia (Boehm *et al.*, 1991; Royer-Pokora *et al.*, 1991) and 1p32 (Begley *et al.*, 1989; Finger *et al.*, 1989; Carroll *et al.*, 1990; Chen *et al.*, 1990) respectively. A normal function of the *LMO2* and *TAL1* genes is the regulation of haematopoiesis. Gene targeting experiments have shown that both the mouse *Lmo2* (Warren *et al.*, 1994) and *Tall* genes (Robb *et al.*, 1995; Shivdasani *et al.*, 1995) are essential during embryogenesis, as homozygous mutant mice die due to failure of yolk sac erythropoiesis. These results are very similar to the effect of a null mutation in

the *GATA-1* gene in embryonal stem (ES) cells, which are also unable to differentiate into mature erythrocytes (Pevny *et al.*, 1991; Weiss *et al.*, 1994). The phenotypes of the *Lmo2*, *Tall* and *GATA-1* null mutations suggest that these three genes have closely related, perhaps synergistic roles in erythroid differentiation. In support of this, we discovered that Lmo2 binds to both Tall and GATA-1 in erythroid cell lines, and that these two interactions may occur simultaneously with Lmo2 bridging between Tall and GATA-1 (Osada *et al.*, 1995). A more general role for Tall in haematopoiesis has been suggested recently from studies of *Tall* null mutant ES cells which fail to develop into any haematopoietic lineages in chimeric mice (Porcher *et al.*, 1996; Robb *et al.*, 1996).

The *LMO2* and *GATA-1* genes both encode proteins belonging to the zinc finger family (reviewed in Sanchez-Garcia and Rabbitts, 1993; Weiss and Orkin, 1995). Lmo2 is a member of the LIM-only class of LIM zinc finger proteins, which also includes the related proteins Lmo1 and Lmo3 (previously called RBTN1 or TTG1, and RBTN3 respectively) (Feroni *et al.*, 1992). In addition to its two LIM domains, Lmo2 also has a short amino-terminal domain which has transcriptional transactivation activity (Sanchez-Garcia *et al.*, 1995). GATA-1 was the first member of the GATA family of zinc finger proteins to be cloned (Tsai *et al.*, 1989), and these proteins all have two zinc fingers and bind the DNA sequence GATA (Weiss and Orkin, 1995). The structure of the LIM and GATA-1 zinc fingers are similar (Omichinski *et al.*, 1993; Perez-Alvarado *et al.*, 1994). Although the global fold of the LIM domain is unique, one part of the LIM zinc finger adopts an α -helical structure similar to the DNA recognition helix of GATA-1 (Perez-Alvarado *et al.*, 1994). Therefore, the LIM domain may function in a similar way to GATA-1 zinc fingers, which mediate both DNA binding and protein dimerization (Perez-Alvarado *et al.*, 1994; Crossley *et al.*, 1995; Merika and Orkin, 1995; Osada *et al.*, 1995; Yang and Evans, 1995). Although there is no evidence so far that LIM domains bind DNA, they do mediate protein dimerization. LIM-LIM interactions facilitate the formation of Lmo2 homodimers (Sanchez-Garcia *et al.*, 1995), as well as heterodimerization between zyxin and CRP (Feuerstein *et al.*, 1994; Schmeichel and Beckerle, 1994). Furthermore, interactions between LIM zinc fingers and other protein dimerization domains are known to occur, e.g. Lmo2 binds to the basic helix-loop-helix (bHLH) domain of TAL1 (Valge-Archer *et al.*, 1994; Wadman *et al.*, 1994). In addition to the interaction found between Lmo2, TAL1 and GATA-1, a new protein has been discovered, called Ldb1 (Agulnick *et al.*, 1996) or NLI (Jurata *et al.*, 1996), which binds to the LIM domains of LIM homeodomain and LIM-only proteins, including Lmo2 and the related protein Lmo1.

In contrast to the Lmo2 and GATA-1 zinc finger

proteins, TAL1 belongs to the bHLH class of transcription factors (Baer, 1993). The bHLH domain facilitates both the formation of protein dimers and sequence-specific DNA recognition, and bHLH dimers bind to E-box motifs which have the general sequence CANNTG (Murre *et al.*, 1989). bHLH proteins have been divided into different categories. The class A factors are broadly expressed, they form homo- and heterodimers, and include the E12 and E47 proteins encoded by the *E2A* gene (Murre *et al.*, 1989; Sun and Baltimore, 1991; Roberts *et al.*, 1993). TAL1 is a member of the B class of bHLH proteins which display tissue-specific expression (Baer, 1993) and interact with class A proteins to form DNA-binding heterodimers (Hsu *et al.*, 1991, 1994b).

Although both TAL1-E2A heterodimers and GATA-1 proteins bind to DNA, it was not known whether the protein complexes containing Lmo2 also associate with DNA. These questions were addressed by initially performing CASTing experiments with crude nuclear extracts from erythroid cells which express Lmo2, TAL1, E2A and GATA-1 (Valge-Archer *et al.*, 1994; Osada *et al.*, 1995) and Ldb1 (our unpublished data). We identified a consensus binding site containing an E-box-GATA motif, consisting of the E-box, CAGGTG, 9 bp upstream of a GATA site. Electrophoretic mobility shift assays with this sequence and the crude nuclear extracts demonstrate that a specific complex of the LIM-only protein Lmo2, with TAL1, E2A and GATA-1, binds to this motif, and the newly identified Ldb1 protein is part of this complex.

Results

DNA sequences obtained from CASTing with anti-Lmo2 antiserum

The possible involvement of Lmo2 in a DNA-binding complex was assessed with CASTing experiments (Wright *et al.*, 1991) using crude nuclear extracts from MEL cells, which express Lmo2 (Valge-Archer *et al.*, 1994; Wadman *et al.*, 1994). A pool of double-stranded oligonucleotides, containing a central core of 26 random nucleotides with conserved flanking regions (Pollock and Treisman, 1990), was incubated with the nuclear extract. Oligonucleotides bound by Lmo2 protein in the extract were immunoprecipitated with anti-Lmo2 antibody, purified and amplified by PCR. After a further four rounds of CASTing, the final PCR products were subcloned and sequenced.

Figure 1A shows the sequence of 31 clones obtained after CASTing with anti-Lmo2 antibody. All of these contained an E-box upstream of a GATA site (herein called the E-box-GATA sequence); furthermore, 18 out of the 31 clones (58%) encoded one particular E-box: CAGGTG. There was some variation in the number of base pairs of DNA separating the E-box and GATA sites, as eight clones (26%) had a spacing of 8 bp between the two sites, 20 clones (65%) had a 9 bp spacing, two clones (6%) had a 10 bp spacing and one clone (3%) had an 11 bp spacing. A number of additional clones contained sequences closely related to the E-box-GATA motif (see Figure 1B). Seven clones had GATT instead of GATA, and six encoded a non-canonical E-box motif (CANNNTG) in combination with either GATA or GATT. Three extra clones only contained GATA sites. The consensus binding site (Figure 1A) shows that neither the residues flanking

Anti-Lmo2 CASTing sequences

A E-box-GATA clones

1	CGGG	CATATG	GCTGGC	--GC	GATA	AGGCG
2	TC	CAGGTG	CACAAC	--AC	GATA	AGGGCGGCCG
3	CCGC	CAGCTG	CACTCC	--GC	GATA	CGGGCCCGA
4	TCGG	CAGCTG	GTGGCG	--CC	GATA	GCCTGTGGAGG
5	GGTA	CAGGTG	CAGACT	--GC	GATA	GTGGTGA
6	TCGG	CAGCTG	CAGAGC	--TT	GATA	AGGGCCGA
7	CGCCT	CAGTGT	CCGTCC	--CC	GATA	GGGGCCGA
8	CCCGA	CAGCTG	TTGAGA	--GC	GATA	CGAC
9	TCGA	CAGGTG	TTGTTT	--CA	GATA	AAAGCCG
10	CGA	CAGGTG	TNCTCTC	--TA	GATA	ACGCCCC
11	CGGCCG	CATCTG	GTGGCC	--TA	GATA	GTGAG
12	CGGC	CAGGTG	CTCCGTA	--CT	GATA	CCGGGAC
13	CCT	CAGGTG	CAGGTG	--NA	GATA	GGCAGTGC
14	CGG	CAGGTG	GTGTGA	--TA	GATA	CGGGCCCGA
15	CACATG	CAGGTG	GGGAGC	--CT	GATA	TGAGG
16	TGCA	CAGGTG	GGGAGC	--CA	GATA	GTGGCCCGA
17	AGCAC	CATCTG	CTTCGA	--CC	GATA	GGGACA
18	CTC	CAGCTG	CAAGTCC	--AC	GATA	CGGGCCCGA
19	GTG	CAGGTG	GCATAGC	--CC	GATA	TGGAGCG
20	GGGCC	CAGGTG	GCTCCGA	--CA	GATA	ACGA
21	CGGGAC	CAGGTG	CNTCCGT	--AC	GATA	GGGACA
22	CTCAAC	CAGGTG	CTGTATC	--CT	GATA	CCGACGACA
23	CGCCTC	CAGGTG	TACCGG	--AA	GATA	CCGGCCGACA
24	CCTCC	CAGGTG	TGCACTC	--AA	GATA	ACAGGT
25	TCGA	CAGCTG	TTAGTCG	--TT	GATA	ATTGC
26	GCCAAC	CAGGTG	TTCCAC	--CT	GATA	TCGAC
27	TCG	CATGTG	CAGTGA	--CA	GATA	ACC
28	TCGG	CATGTG	CCGGAG	--CA	GATA	GGGC
29	GGGCC	CAGGTG	TAGAGCG	--GA	GATA	GGGA
30	TCG	CAGGTG	CATTGCAC	--TA	GATA	AGGCTGGAG
31	CCT	CAATG	TGCTCTGCCA		GATA	GGGCCCGA

Base Frequency																					
C	16	10	12	31	0	0	8	0	0	14	5	8	14	11	0	0	0	0	7	11	8
A	3	4	6	0	31	2	2	0	0	10	3	5	14	0	31	0	31	10	1	3	
G	8	11	10	0	0	24	21	0	31	8	4	14	6	0	31	0	0	11	15	19	
T	3	6	3	0	0	5	0	31	0	9	10	6	5	6	0	31	0	3	4	1	
Consensus sequence																					
	C	A	G	G	T	G							G	A	T	A					G

B Motifs related to the E-box-GATA consensus

E-box-GATT clones

1	AGTCCGC	CAGGTG	TCAAGG	--GC	GATT	AGG
2	CGC	CAGGTG	GGTCCC	--TA	GATT	AGGCCGTG
3	AATTA	CAGCTG	TCAGCA	--CT	GATT	ACG
4	CCAC	CATGTG	GCATCA	--CT	GATT	AGCA
5	TCAC	CATCTG	GTACCGA	--AT	GATT	AGC
6	TCG	CAGCTG	TTGGATG	--CC	GATT	ATGGCTGGC
7	TCGC	CACCTG	ATAAGCACTA		GATT	ATGCATGA

E-box-like (CANNNTG) GATA and GATT clones

1	GCTCG	CAATCTG	CCAGAC	--GC	GATA	ACGG
2	CCT	CAAGCTG	GTTCGG	--AT	GATA	AGCTGGCC
3	TCG	CAAGGTG	CAGGGG	--AG	GATT	AGCGAA
4	CGGGC	CAATGTG	CAGCCATC		GATA	CGACA
5	CCT	CAAGGTG	GTCCCCACA		GATA	TGACCGC
6	TGGCC	CAAGGTG	GTACATACT		GATT	GCGACA

GATA clones

1	CGCCTCCTGGAGCGTCCA	GATA	GCATAGCCAGC
2	GGACCAAGGTGCAACCGGCTT	GATA	CGACA
3	TGGCCAGATT	GATA	AGCCCCAGGGTGACA

Fig. 1. Sequences of clones obtained after CASTing with MEL nuclear extract and anti-Lmo2 antiserum. DNA-binding site selection was performed by CASTing with nuclear extracts from MEL cells and the R76 oligonucleotide. After five rounds of CASTing the final selected oligonucleotide pools were cloned into pBSpt vector and sequenced using the Universal primer. The E-boxes and GATA sites were aligned and one or two padding characters inserted three residues 5' of the GATA site to accommodate the variation in spacing between the two sites in the different clones. The base frequency is expressed as the number of sequences with a particular nucleotide at each position. Where 50% or more of the sequences encoded a given nucleotide, that nucleotide is included in the consensus, and the number is shown in bold. The consensus extends to three residues upstream of the E-box and three residues downstream of the GATA site (the whole of the sequence between the E-box and GATA sites is not shown, as it was not conserved). (A) Nucleotide sequence of 31 E-box-GATA clones selected by CASTing with anti-Lmo2 antiserum. The consensus was calculated by analysing the 31 sequences. (B) Nucleotide sequence of 16 additional clones which encoded motifs which were related to the E-box-GATA consensus.

the E-box and GATA sites, nor the sequence of the DNA between the two sites, was conserved. It is noteworthy that the E-box and GATA sites have been identified previously as binding sites for non-LIM domain proteins, since bHLH proteins recognize E-box sequences (Murre

et al., 1989), and members of the GATA family recognize the GATA sequence (Weiss and Orkin, 1995).

DNA sequences bound by Tal1 and E2A proteins in MEL cells

Lmo2 complexes with the bHLH protein Tal1 in MEL cells (Valge-Archer *et al.*, 1994; Wadman *et al.*, 1994), and therefore it seemed likely that Tal1 itself might be contributing to the recognition of the E-box-GATA sequence obtained by anti-Lmo2 CASTing. The consensus DNA-binding site of TAL1 was compared with that of Lmo2 by conducting CASTing experiments with anti-TAL1 antiserum and nuclear extracts. Twenty-nine clones were obtained with an E-box upstream of a GATA site (15 of these sequences are shown in Figure 2A). As observed in the Lmo2 CASTing data, a majority of the E-box-GATA clones contained the E-box CAGGTG (17 clones, i.e. 59%). Furthermore, the number of base pairs of DNA separating the E-box and GATA sites also varied, with four clones (14%) having a spacing of 8 bp between the two sites, 22 clones (76%) having a 9 bp spacing, two clones (7%) having a 10 bp spacing and one clone (3%) having an 11 bp spacing. Another 15 CASTing clones were obtained (in addition to those shown in Figure 2A). Four of these sequences were similar to the E-box-GATA consensus and encoded E-box-like or GATT motifs, another four sequences contained isolated E-boxes, and seven contained single GATA sites. The consensus DNA-binding site (Figure 2A) was very similar to that derived from the anti-Lmo2 CASTing (Figure 1A), suggesting that a complex consisting of at least TAL1 and Lmo2 binds to the E-box-GATA motif.

DNA binding by TAL1 has been observed previously, but only in the presence of class A bHLH proteins (Hsu *et al.*, 1991, 1994a). The E2A gene is widely expressed, and it encodes two class A bHLH proteins, E12 and E47, which dimerize with TAL1 in leukaemic T cells (Hsu *et al.*, 1994a,b). Furthermore, mammalian two-hybrid experiments showed that Lmo2 can interact with TAL1-E2A complexes (Wadman *et al.*, 1994; Osada *et al.*, 1995). It is therefore possible that the Lmo2 in MEL cells binds to a complex of TAL1 and E2A, and the three proteins together contribute to recognition of the E-box-GATA motif. Accordingly, we repeated the CASTing experiments using anti-E2A antiserum (Figure 2B) and an E-box-GATA motif was observed, similar to that found after the anti-Lmo2 and anti-TAL1 CASTing. Three clones (22%) had an 8 bp spacing, nine clones (64%) had a 9 bp spacing and two clones (14%) had a 10 bp spacing. The consensus sequence (Figure 2B) was different from that previously observed, as only four clones had the CAGGTG E-box; the other features of the consensus were very similar to the results of the anti-Lmo2 and anti-TAL1 CASTing (Figures 1A and 2A). A large number of sequences closely related to the E-box-GATA consensus were also obtained (17 clones), which encoded E-box-like GATA and E-box-GATT motifs. Additional sequences contained either isolated E-boxes (11 clones) or GATA sites (nine clones). The high degree of similarity in the consensus DNA-binding sites of Lmo2, TAL1 and E2A suggests that a complex of these three proteins binds specifically to an E-box-GATA motif in MEL nuclei.

A Anti-TAL1 CASTing sequences

1	TCTA	CAGCTG	CTCCGC--TC	GATA	GTGGCCCG
2	CCT	CAGGTG	CGAGAC--AC	GATA	TTGGCCCCGA
3	CGGG	CAGGTG	GGCGGC-CA	GATA	CGCGGGGA
4	TCGC	CAGGTG	GGTACC-CC	GATA	CCGCCCG
5	CGGGC	CAGGTG	GGACTG-CT	GATA	GGAGG
6	CGA	CAGGTG	GCCAGCC-GT	GATA	CGAGGTGA
7	TC	CAGGTG	CCATGAC-GT	GATA	TCGGTGCCG
8	TC	CACCTG	TAGGCC-CT	GATA	CGGGGGCCG
9	CGC	CAGGTG	CTGAGT-GT	GATA	CCCGGGGA
10	CGGGCCAC	CAGGTG	CACNGT-CC	GATA	GAGG
11	TC	CAGGTG	CTGTGCT-AC	GATA	GTGGTGGCGA
12	CGG	CAGGTG	CAGCACC-AC	GATA	CGCGGTGAG
13	CGGGGCCA	CAGGTG	TTGCCGA-GT	GATA	CGAG
14	CGGGGC	CAGCTG	CCTGCAC-CT	GATA	GTGGAG
15	CGGGGCAG	CATGTG	TGCAGGCCG	GATA	CGA

Base Frequency	
C	19 4 15 29 0 3 3 0 0 16 7 9 - 13 13 0 0 0 0 11 7 3
A	0 7 8 0 29 0 1 0 0 0 0 9 3 - 5 6 0 29 0 29 7 1 7
G	6 11 5 0 0 21 24 0 29 8 3 14 - 8 0 29 0 0 0 8 15 19
T	0 7 1 0 0 5 1 29 0 5 10 3 - 3 10 0 0 29 0 3 6 0
Consensus sequence	
C	C C A G G T G C G A T A G G

B Anti-E2A CASTing sequences

1	CCTC	CAGGTG	TCCACT--AG	GATA	ACATGCAA
2	TCTG	CACCTG	GTCGTT--AC	GATA	AGCCA
3	CCTC	CACCTG	CGAGCT--AC	GATA	ACGTCGTC
4	ATGCCAC	CACCTG	CAGCTA-GC	GATA	CGA
5	CGA	CACCTG	CTGACTG-CT	GATA	CGGCAT
6	ACCAG	CACCTG	TTGTGCA-AC	GATA	AGCA
7	TGGGA	CATCTG	TGGTGT-TT	GATA	AGAGG
8	TCCAG	CATCTG	TGATCT-CC	GATA	CGA
9	AGCAC	CATCTG	GTCACGC-CT	GATA	ACGA
10	TCCGG	CACCTG	CACAAGT-TA	GATA	GTAG
11	CCTC	CAGGTG	CACTTG-GC	GATA	CCCCCT
12	TCCACA	CAGGTG	TCCGCC-CT	GATA	ACGA
13	GAGAC	CAGGTG	GCATGCTAGC	GATA	ACGA
14	AACAC	CACCTG	TACCACCCC	GATA	ACGA

Base Frequency	
C	10 1 7 14 0 6 8 0 0 5 3 7 - 5 8 0 0 0 0 4 8 2
A	1 6 3 0 14 0 0 0 0 0 4 3 - 4 1 0 14 0 14 9 0 5
G	3 3 4 0 0 5 6 0 14 3 4 - 3 1 14 0 0 0 1 5 7
T	0 4 0 0 0 3 0 14 0 6 4 0 - 2 4 0 0 14 0 0 1 0
Consensus sequence	
C	C C A C T G C C G A T A A C G

C Anti-GATA-1 CASTing sequences

1	CTGTCGGGGCCG	GATA	ACGATCAGTCTGCCCCCGA
2	GTCCGGCTGCAC	GATA	AGATCGCTCTGGCCGA
3	TCGGGT	GATA	AGGGCAGCCTCCGACGGGA
4	GTCCGTGC	GATA	AGCATACATCTGCCCGCGGA
5	TGTCGCCGC	GATA	AGGAGGGGGCGGTAGGGAG
6	TCGGCC	GATA	AGGTGTGTAGCTGGCCCGGGAG
7	TCGAGCGGAC	GATA	AGGATCGAGGGGGCTGAG
8	TCGCGGGGGAGATGGCC	GATA	CGGCTGGAG
9	TCGTGGGAGCC	GATA	ACGATCTGCTCCGGAG
10	TCGCC	GATA	AGCCATGACCGGGGGCGGAG
11	TCGGGGCACAC	GATA	AGCATCCACGGCACGAC
12	TCGCCA	GATA	GGGCATCGAGCTGGGGGAG
13	CTCCCCACGCC	GATA	ACGATCTGTCCCTCGA
14	TCGGGA	GATA	CGGTGAGCTATCGACCCCGAG
15	CTCGGCC	GATA	ACATCATGGCCCCCGGA

Base Frequency	
C	19 24 36 0 0 0 0 5 24 8
A	5 5 6 0 50 0 50 33 0 10
G	19 16 2 50 0 0 0 12 24 26
T	4 5 6 0 0 50 0 0 2 6
Consensus sequence	
C	C G A T A A G

Fig. 2. Sequences of clones obtained after CASTing with MEL nuclear extract and anti-TAL1, anti-E2A and anti-GATA-1 antisera. CASTing was carried out as in Figure 1 with nuclear extracts from MEL cells and antisera recognizing TAL1 (five rounds of CASTing with anti-TAL1 antiserum), E2A (seven rounds of CASTing with anti-E2A antiserum) and GATA-1 (five rounds of CASTing with an antibody against the NH₂-terminal GATA-1 peptide). Selected oligonucleotides were cloned in pBSpt prior to sequence analysis. Consensus binding sites were calculated by aligning the E-box and GATA motifs. The numbers of sequences with given nucleotides at each position are shown, and where the frequency of any nucleotide was 50% or above, that number is written in bold, and the nucleotide is included in the consensus. (A) Nucleotide sequence of 15 out of a total number of 29 E-box-GATA clones obtained by CASTing with anti-TAL1 antiserum. The consensus was calculated by analysis of all 29 sequences. (B) Sequences of 14 E-box-GATA clones selected by CASTing with anti-E2A antiserum. These sequences were aligned to derive the consensus binding site. (C) Nucleotide sequences of 15 out of a total of 46 GATA site-containing clones, obtained by CASTing with MEL nuclear extract and anti-GATA-1 antiserum. The consensus was calculated by aligning a total of 50 GATA sites.

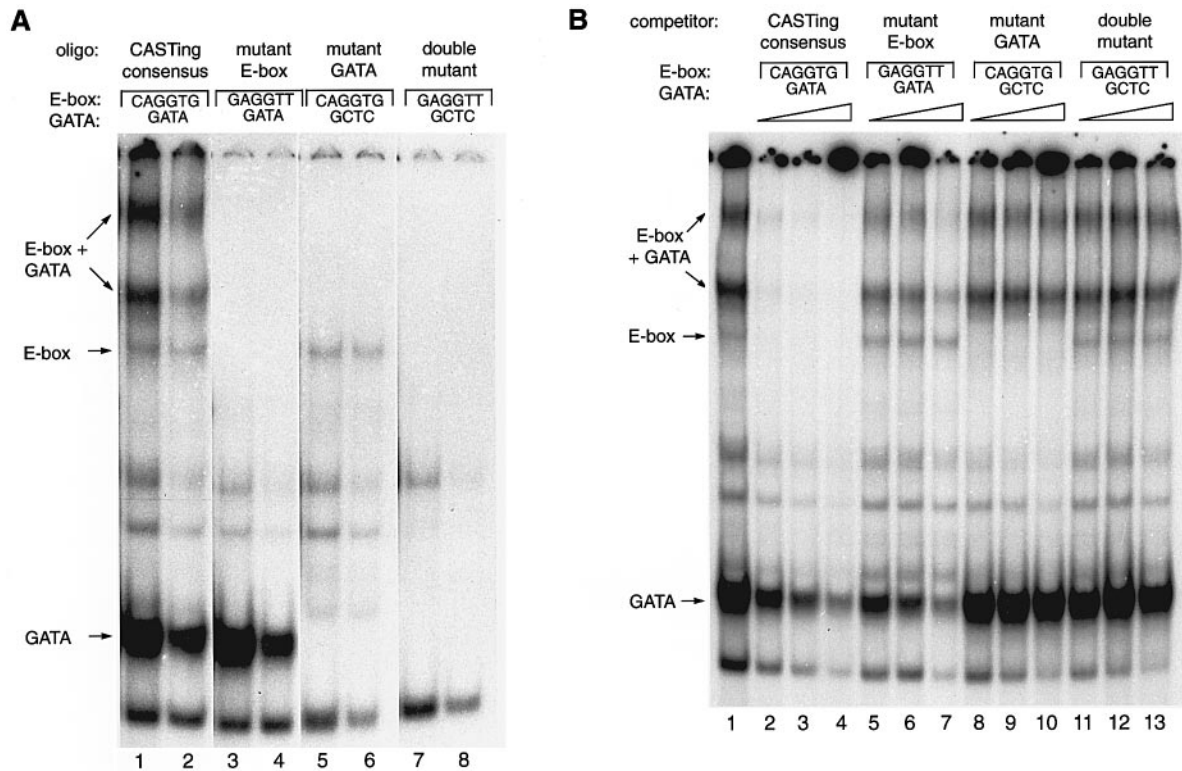


Fig. 3. Electrophoretic mobility shift of specific E-box–GATA-binding proteins in MEL nuclear extract. **(A)** A series of double-stranded, ^{32}P -labelled oligonucleotide probes encoding the E-box–GATA motif (CASTing consensus, lanes 1 and 2), a mutant E-box (lanes 3 and 4), a mutant GATA site (lanes 5 and 6) or both sites mutated (lanes 7 and 8), were incubated with MEL nuclear extract (5 μg of protein) in the presence of 40 $\mu\text{g}/\text{ml}$ (lanes 1, 3, 5 and 7) or 60 $\mu\text{g}/\text{ml}$ (lanes 2, 4, 6 and 8) *E. coli* DNA, for 20 min at room temperature. The samples were separated on a 4% non-denaturing polyacrylamide gel, which subsequently was dried and subjected to autoradiography. Protein–DNA complexes recognizing the E-box, GATA site or E-box–GATA motif are indicated with arrows. **(B)** MEL nuclear extract was incubated with ^{32}P -labelled oligonucleotides, containing the consensus E-box–GATA motif obtained in the CASTing experiments, in the presence of a 10- (lanes 2, 5, 8 and 11), 20- (lanes 3, 6, 9 and 12) or 50-fold (lanes 4, 7, 10 and 13), excess of the indicated unlabelled oligonucleotides. These encoded the consensus sequence (lanes 2, 3 and 4) or mutant versions of the motif (lanes 5–13). Samples were fractionated as above. Protein–DNA complexes recognizing the E-box, GATA site or E-box–GATA motif are indicated with arrows. EMSA experiments were carried out with excess probe and, in this and subsequent EMSAs, only the upper portions of each gel are shown in the figures.

Sequence specificity of DNA binding by GATA-1 protein in MEL cells

The GATA-1 protein is highly expressed in MEL cells, and a proportion of it is found complexed with Lmo2, suggesting that an oligomeric complex occurs which comprises Lmo2, Tal1, E47 and GATA-1 (Osada *et al.*, 1995). Because GATA-1 protein binds to the DNA sequence GATA, we considered whether GATA-1 might also be part of the Lmo2–Tal1–E2A complex which recognizes the E-box–GATA motif. We therefore repeated the CASTing procedure with MEL nuclear extract using anti-GATA-1 antiserum. The sequences of 15 clones derived from five rounds of CASTing are shown in Figure 2C. In contrast to the sequences shown in Figures 1A and 2A and B, only six clones out of a total of 54 contained an E-box as well as a GATA site, but the spacing between the two, as well as their relative orientation, was variable. The majority of clones (85%) encoded only one GATA site (Figure 2C).

Specific DNA recognition of the E-box–GATA motif by proteins in MEL nuclear extract

The sequences derived from the CASTing experiments probably corresponded to high affinity binding sites for the proteins analysed in MEL cells. This possibility was tested by performing electrophoretic mobility shift assays

(EMSA) with oligonucleotides corresponding to the consensus E-box–GATA motif, or mutant sequences where either the E-box or GATA site, or both sites, were mutated (see Materials and methods). The EMSA illustrated in Figure 3A shows that incubation of MEL nuclear extract with each of the four ^{32}P -labelled oligonucleotides resulted in the formation of a number of distinct protein–DNA complexes. The DNA sequence specificity of four of these complexes was defined by comparing the mobility shift characteristics of the nuclear extract with wild-type and mutant oligonucleotides. These specificities are indicated on the left hand side of Figure 3A. Two of the complexes migrated a short distance into the gel and were formed only when both the E-box and the GATA sites were present on the same oligonucleotide (Figure 3A, lanes 1 and 2, bands designated E-box + GATA). These particular low mobility complexes were not observed when the GATA site (Figure 3A, lanes 3 and 4), the E-box (lanes 5 and 6) or both sites (lanes 7 and 8) were mutated in the oligonucleotides. Other DNA-binding complexes were present in the MEL nuclear extracts which bound to the E-box sequence alone or the GATA sequence alone (designated E-box and GATA respectively in Figure 3A). The band indicated as corresponding to the E-box-binding complex was only observed with the E-box–GATA consensus oligonucleotide (Figure 3A, lanes 1 and 2), or the

mutant GATA oligonucleotide (lanes 5 and 6), but not with mutant E-box or double mutant oligonucleotides (lanes 3, 4, 7 and 8). Conversely, the band designated GATA only appeared in the presence of the consensus oligonucleotide (lanes 1 and 2) or the mutant E-box oligonucleotide in which the GATA site was retained (lanes 3 and 4).

In the two low mobility DNA-binding complexes observed in the band shift assays (Figure 3A, labelled E-box + GATA), there appeared to be simultaneous recognition of both E-box and GATA motifs. It was possible, however, that two independent protein complexes were bound simultaneously to the E-box and GATA sites. EMSAs were therefore used to examine the sequence requirements of DNA recognition by proteins in the MEL nuclear extract. These were incubated with the radiolabelled E-box GATA consensus oligonucleotide (CAGGTG-9 bp-GATA), in the presence of increasing concentrations of unlabelled consensus or mutant oligonucleotides, to determine the competition characteristics of the complexes (Figure 3B). Non-radiolabelled, E-box-GATA consensus oligonucleotide competed very effectively with the radiolabelled, consensus oligonucleotide, and inhibited the formation of all the protein-DNA complexes including the E-box-GATA and E-box-binding proteins, and to a lesser extent the GATA-binding activity (Figure 3B, compare lane 1 with lanes 2, 3 and 4). The apparent resistance of the GATA-binding protein to the competitor oligonucleotide was most likely due to the high concentration of this protein (GATA-1, see below) in the MEL nuclear extract.

By contrast, none of the oligonucleotides which had mutant E-boxes (Figure 3B, lanes 5, 6 and 7), mutant GATA sites (lanes 8, 9 and 10) or both sites mutated (lanes 11, 12 and 13) inhibited the binding of the two low mobility protein complexes to the E-box-GATA motif. The sole exception to this was a 50-fold excess of the oligonucleotide with a mutated E-box and an intact GATA site. At this level, the mutant oligonucleotide was able to inhibit, to some extent, the recognition of the E-box-GATA motif (lane 7). The data therefore show that MEL nuclei contain DNA-binding proteins which specifically recognize an E-box-GATA motif consisting of the E-box, CAGGTG, 9 bp upstream of a GATA site. These data also indicate that recognition by the low mobility complex of the E box GATA motif requires complementary binding to both sites, as oligonucleotides carrying the individual sites failed to compete for binding.

A complex of Lmo2, Ldb1/NLI, TAL1, GATA-1 and E2A proteins specifically recognizes the E-box-GATA motif

The possible presence of Lmo2, TAL1, E2A and GATA-1 in each of the protein-DNA complexes was assessed using antibody-mediated supershifts, in which an antibody binds to a pre-formed protein-DNA complex thereby increasing its molecular size, and decreasing its electrophoretic mobility in EMSAs. A number of the protein-DNA complexes in the MEL F4N nuclear extract were supershifted by the various antibodies. Most important is that the two E-box-GATA-dependent low mobility complexes were supershifted by anti-GATA-1, anti-Lmo2, anti-TAL1 and anti-E2A polyclonal antisera (Figure 4A, lanes 5, 9, 13

and 17), whereas pre-immune serum (lane 1), or an antiserum against GATA-3, had no effect (lane 2). In addition, antiserum binding to GATA-1 protein supershifted the band previously designated as the 'GATA' band (Figure 4A, lanes 5 and 6) and anti-E2A antiserum supershifted the E2A band (Figure 4, lanes 17 and 18).

The addition of antibody resulted in only a small retardation of electrophoretic mobility of the two bands corresponding to the E-box-GATA-binding complexes, reflecting the high molecular weight of the multi-protein complexes. Improved resolution of antibody-supershifted bands was achieved by inclusion of the immunoglobulin-binding protein, protein A, to increase the size of the complex further, thereby reducing the mobility of bands within the gel. In some cases, the addition of protein A actually prevented the entry of the supershifted complexes into the gel, but its presence did not alter the mobility of the protein-DNA complexes from samples incubated without added antibody (data not shown), nor with an antiserum against GATA-3 (Figure 4A, lane 2). However the addition of protein A did reduce the mobility of the complexes supershifted by antisera against Lmo2, TAL1, E2A and GATA-1 (lanes 6, 10, 14 and 18), while having no effect with the corresponding pre-immune sera (lanes 3, 4, 7, 8, 11, 12, 15 and 16). It is also noteworthy that antibodies recognizing TAL1 only affected the mobility of the two low mobility complexes. In summary, the two high molecular weight DNA-binding complexes in MEL nuclei, which specifically recognize the E-box 9 bp upstream of a GATA site, contain Lmo2, TAL1, E2A and GATA-1 proteins, previously demonstrated to form an interactive complex in MEL cells (Osada *et al.*, 1995). In addition, analogous Lmo2-containing complexes were observed in two other erythroid lines, 707 and HEL, but not in two non-erythroid cell lines, C3H10T1/2, and the mouse myeloma line NS0 (data not shown).

A novel protein has recently been identified which binds specifically to LIM domains of LIM-only and LIM homeobox proteins (Agulnick *et al.*, 1996; Jurata *et al.*, 1996). This novel LIM domain-binding protein, Ldb1 or NLI, can bind to both Lmo1 and Lmo2, and is expressed in a wide range of tissues. Thus it seemed a potential component of the E-box-GATA-binding complex, particularly in view of the large size of this complex. Therefore, we added anti-Ldb1 antiserum to MEL F4N nuclear extract in the presence and absence of protein A, and analysed the samples by EMSA with radiolabelled E-box-GATA oligonucleotide (Figure 4B). The anti-Ldb1 antiserum specifically supershifted the two high molecular weight Lmo2-GATA-1-TAL1-E47 complex bands. The presence of protein A reduced the mobility of the two Lmo2-GATA-1-TAL1-E47 complexes further (Figure 4, lanes 4 and 5) whilst pre-immune serum had no effect on the mobility of any of the DNA-binding complexes (Figure 4, lanes 2 and 3). Therefore, in MEL nuclei, the Ldb1 protein is part of a complex with Lmo2, TAL1, E2A and GATA-1, and together these five proteins constitute an oligomeric complex which binds to the E-box-GATA motif.

The E-box-GATA clones obtained after CASTing with the anti-Lmo2, anti-TAL1 and anti-E2A antisera displayed a restricted variation in the spacing between the E-box and GATA sites (Figures 1A and 2A and B). The effect

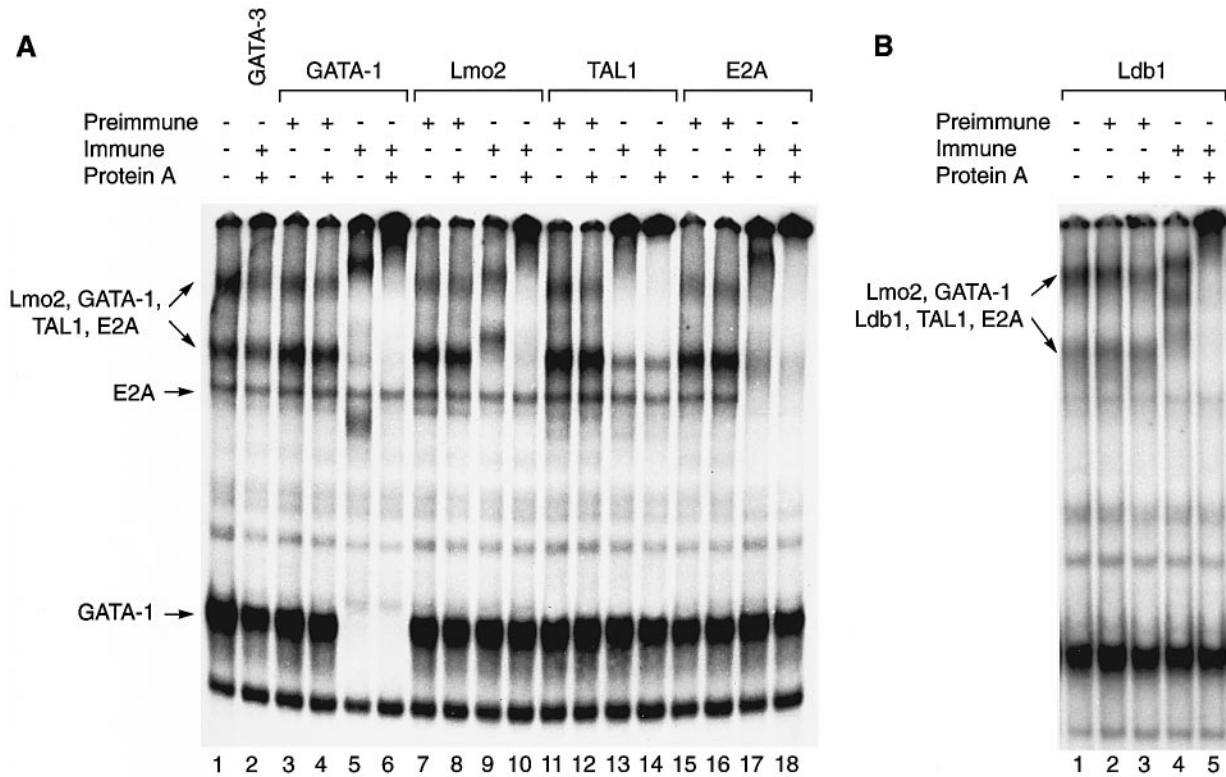


Fig. 4. Identification of proteins which recognize the E-box–GATA motif in MEL cells. **(A)** Aliquots of MEL subclone F4N nuclear extract were incubated with a ³²P-labelled oligonucleotide encoding the consensus E-box–GATA motif, for 5 min prior to the addition of specific pre-immune or immune antisera raised against GATA-3, GATA-1, Lmo2, TAL1 and E2A. After an additional 10 min, 1 µg of the immunoglobulin-binding protein, protein A, was added to some of the samples (even-numbered lanes), and the incubations were continued for a further 10 min. The extracts were separated on a 4% polyacrylamide gel, prior to autoradiography. The location of the free probe is not shown, as the figure depicts only the upper half of the gel. Bands corresponding to Lmo2, Ldb1, GATA-1, TAL1, E2A complexes, and E2A and GATA-1 proteins, are indicated with arrows. **(B)** Aliquots of MEL subclone F4N nuclear extract were incubated with a ³²P-labelled oligonucleotide encoding the consensus E-box–GATA motif, for 5 min prior to the addition of specific pre-immune (lane 2) or immune antisera raised against Ldb1 (lanes 4 and 5). Then 2 µg of protein A was added to one tube and incubation continued for 10 min (lane 5). The extracts were separated on a 4% polyacrylamide gel, prior to autoradiography. The bands corresponding to the oligomeric DNA-binding complex are indicated.

of varying the distance between the E-box and GATA sites on DNA binding to the motif by the E-box–GATA-binding complexes was investigated by performing EMSAs with oligonucleotides, in which the spacing between the E-box and GATA sites varied between 7 and 12 bp of DNA. Although the binding of E2A and GATA-1 to their individual binding sites was unaffected by the spacing between the E-box and GATA sequences (Figure 5, lanes 1–6), DNA binding by the low mobility complexes of Lmo2, Ldb1, Tal1, E2A and GATA-1 was dependent on the distance between these sites. The low mobility complexes were detected only when the spacing between E-box and GATA sites was 8, 9 or 10 bp (lanes 2–4), and not when reduced to 7 bp, or increased to 11 or 12 bp. It is interesting that this separation corresponds to approximately one turn of the DNA helix, and would place the E-box and GATA sites on the same face of the DNA. However, it should be noted that the design of the CASTing oligonucleotides restricts bound products to the 26 internal random oligonucleotides.

Specificity of TAL1–E2A complexed with Lmo2 and GATA-1

The E-box most frequently found associated with the GATA site in the CASTing data sequences was CAGGTG (39 out of 74 clones), which differs from the previously

published TAL1–E2A consensus E-box (Hsu *et al.*, 1994a). The sequence specificity of DNA recognition by Tal1–E2A-containing complexes in MEL cells was tested directly using EMSAs with MEL nuclear extract and three different oligonucleotides in the presence of either anti-Tal1 or anti-E2A antisera. The data in Figure 6 (lane 1) show the pattern of bands obtained with the consensus E-box–GATA oligonucleotide, and the bands corresponding to Lmo2, Ldb1, Tal1, E2A and GATA-1 are indicated. Using the consensus E-box–GATA oligonucleotide, specific Tal1–E2A complexes were only found in combination with Lmo2 and GATA-1, since the two low mobility bands were supershifted by both anti-TAL1 (lanes 3 and 4) and anti-E2A antisera (lanes 5 and 6). An E2A homodimer band which recognized the CAGGTG E-box was supershifted by anti-E2A antiserum (Figure 6, lanes 5, 6, 11 and 12).

The low mobility Lmo2, Ldb1, Tal1, E2A and GATA-1 oligomeric complexes were not observed by EMSA when the GATA site was absent (Figure 6, lanes 7–12), while Tal1–E2A heterodimers did bind to the CAGGTG E-box in the absence of Lmo2 or GATA-1, since a new band appeared in the EMSA with this oligonucleotide (Figure 6, lanes 13 and 14). This complex was supershifted by anti-TAL1 (lanes 15 and 16) and by anti-E2A antisera (lanes 17 and 18), but not by anti-GATA-1 or anti-

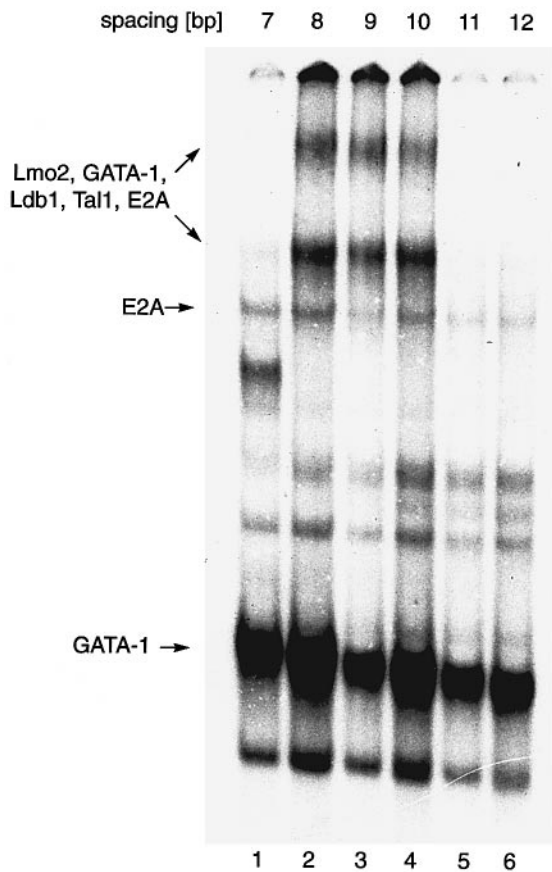


Fig. 5. The separation between the E-box and GATA sites affects recognition of the E-box–GATA motif. A series of ^{32}P -labelled oligonucleotides encoding the E-box CAGGTG, separated from a GATA site by either 7, 8, 9, 10, 11 or 12 bp of DNA, were incubated with MEL (F4N) nuclear extract as described in Materials and methods and the protein–DNA complexes separated on a 4% acrylamide gel before autoradiography. The locations of E-box–GATA-binding complexes containing Lmo2, Ldb1, Tal1, E2A and GATA-1 in lanes 2, 3 and 4 are indicated by arrows, as are the E2A and GATA-1 protein–DNA complexes.

Lmo2 antisera (data not shown). In addition, anti-GATA-2 antiserum had no effect on the EMSA pattern of the nuclear extract (lanes 2, 8 and 14). Therefore, two distinct complexes involving Tal1 exist in MEL nuclei, an oligomeric complex binding to the CAGGTG E-box (only when adjacent to a GATA site) and one binding to the CAGATG E-box (in the absence of the GATA site).

Assembly of the oligomeric complex *in vivo* establishes a function in transcriptional transactivation

The transcriptional activation potential of the constituent proteins, except Ldb1, has been investigated and, while GATA-1 is a potent activator, TAL1 may act as a transcriptional repressor (Hsu *et al.*, 1994c). No direct DNA binding has so far been demonstrated for Lmo2 or Ldb1. The possibility of the oligomeric complex activating transcription was tested by reconstituting the complex *in vivo* together with an E-box–GATA reporter in COS cells. The transcriptional activation of an E-box–GATA–luciferase reporter was related to the co-transfection of expression constructs (Figure 7A), the protein expression levels were assessed by Western blotting (Figure 7B) and

the DNA-binding activity of the extracts was ascertained by EMSA using the consensus E-box–GATA oligonucleotide (Figure 7C).

COS cells were transfected with combinations of expression plasmids (Figure 7A) together with the reporter. The low mobility bands in the EMSA, due to the oligomeric complex, were absent if any one of the five expression plasmids was omitted (unpublished results). Reporter construct activation was observed in four transfection combinations (3, 4, 7 and 8). GATA-1 alone is a potent transactivator (transfection 3), and when GATA-1 was produced in cells together with Lmo2 and Ldb1, activation was also observed, albeit at a consistently reduced level (Figure 7A, transfection 4). No transactivation was observed when TAL1 and E47 were co-transfected with or without Lmo2 and Ldb1 (transfections 6 and 5 respectively), presumably due to the weak transactivation by TAL1–E47 dimers (Hsu *et al.*, 1994a). Accordingly, when GATA-1 was co-transfected with TAL1 and E47, similar levels of activation were observed as with GATA-1 alone (transfections 7 and 3 respectively).

When the five protein components of the complex were co-expressed, assembly of a complex occurred which gave the most efficient transactivation (transfection 8), yielding ~2.5 times greater luciferase activity than GATA-1 expression alone. The sum of transactivation observed after transfection of combinations of GATA-1 with other components only marginally increases that found with GATA-1 alone, except in the case of expression of the five components together. These data indicate that full assembly of the oligomeric complex (i.e. Lmo2, Ldb1, Tal1, E47 and GATA-1) is necessary to facilitate efficient binding across the E-box and GATA motifs, and establish a function for this complex in transcriptional transactivation.

Since the *in vivo* assembly of the complex required the simultaneous transfection of many components, each of the cDNAs were cloned into the same expression vector, and the levels of protein were monitored by Western blotting of the cellular extracts from the eight transfections (Figure 7B). These data confirm the presence of each protein in the relevant transfection, and that the amounts of each synthesized are approximately equivalent between transfections.

The same transfected cells were also used as a source of nuclear extract for EMSAs with an E-box–GATA probe to correlate the transactivation with DNA binding to the E-box–GATA motif (Figure 7C). No protein complexes were observed when Lmo2 and Ldb1 were expressed together (transfection 2) or alone (data not shown), indicating that the LIM-only protein Lmo2 does not bind to the GATA site, at least in this context. Cells transfected with the GATA-1 clone alone (transfection 3) or together with Lmo2 and Ldb1 clones (transfection 4) yielded one band in the EMSA which could be supershifted with the anti-GATA-1 serum. Combined transfection with E47 and TAL1 expression vectors (transfection 5) yielded three main bands, one of which corresponds to an E47–TAL1 heterodimer, since it was supershifted by antisera recognizing either of the proteins, and a doublet of bands corresponding to E47 homodimers. Co-expression of Lmo2 and Ldb1 with E47 and TAL1 (transfection 6) had no effect on the binding of the E47 homodimer, but it is interesting that this combination ablated the binding of the

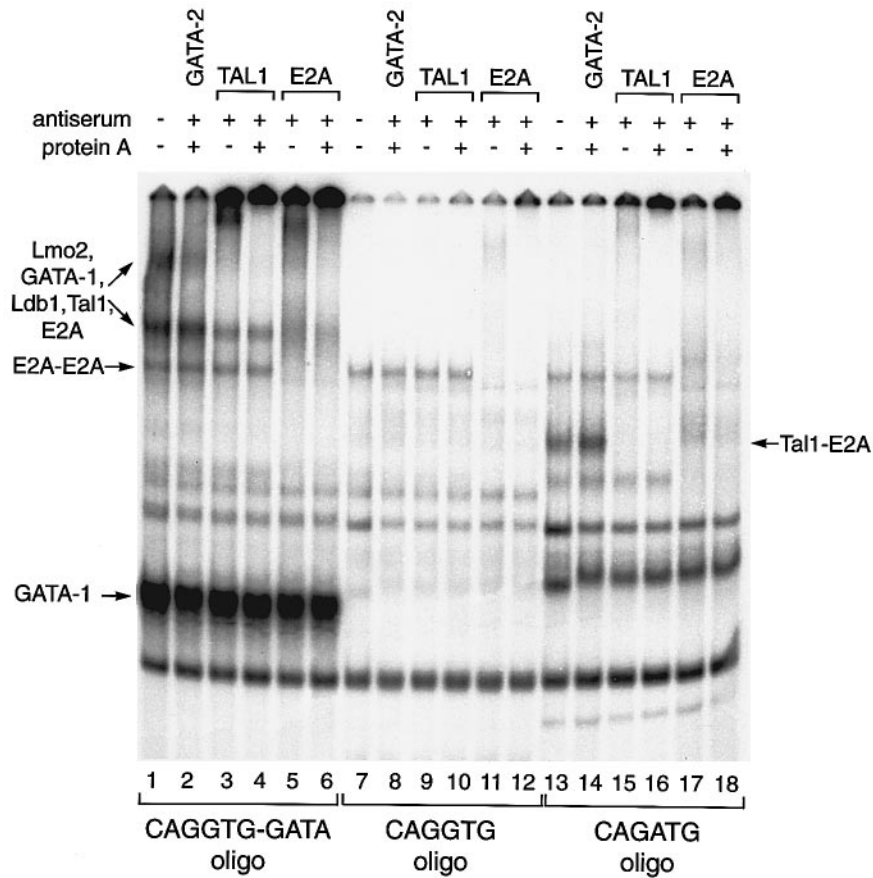


Fig. 6. The DNA-binding specificity of TAL1-E2A complexed with Lmo2 and GATA-1, differs from that of TAL1-E2A heterodimers. EMSA of MEL nuclear extract with three different oligonucleotide probes encoding (i) the E-box-GATA consensus which contains the CAGGTG E-box (lanes 1-6), (ii) the CAGGTG E-box without a downstream GATA site (lanes 7-12) and (iii) the TAL1-E2A consensus E-box CAGATG without an associated GATA site (lanes 13-18). Nuclear extracts were incubated with the particular oligonucleotide, in the presence or absence of specific antibody, followed by fractionation of 4% polyacrylamide gels and subsequent autoradiography. TAL1-E2A-containing protein-DNA complexes were identified by incubating the nuclear extract with anti-TAL1 (No. 370) and anti-E2A (No. 526) antisera, with (lanes 2, 4, 6, 10, 12, 16 and 18) and without protein A (lanes 1, 3, 5, 7, 8, 9, 11, 13, 14, 15 and 17). GATA-2 antiserum, together with protein A, was also added to some of the samples (lanes 2, 8 and 14). Bands corresponding to Lmo2, Ldb1, GATA-1, TAL1-E2A and E2A and GATA-1 protein-DNA complexes are indicated with arrows on the left-hand side of the figure, and the position of the TAL1-E2A heterodimer is indicated on the right-hand side.

E47-TAL1 heterodimer to the DNA. Conversely, expression of GATA-1 had no apparent effect on DNA binding by E47-E47 or E47-TAL1 dimers (transfection 7).

The transfected COS cells which showed the greatest transactivation of the E-box-GATA-luciferase reporter (Figure 7A, transfection 8) show a number of bands in the EMSA which can be attributed to various complexes by supershift assays. These bands include those attributable to E47 homodimer and GATA-1 bands (observed in transfections 3, 4, 5 and 7) and, in addition, two low mobility bands were found, due to oligomeric complexes comprising Lmo2, Ldb1, GATA-1, TAL1 and E2A as judged by supershifts. The oligomeric complex was only found in transfection 8, consistent with the need for the five components of the DNA-binding complex in activation of transcription from the E-box-GATA sequence.

Discussion

Lmo2 forms part of an oligomer binding to the E-box-GATA motif

Our previous data, using reporter and two-hybrid assays, indicated that a complex involving Lmo2, TAL1, E47 and

GATA-1 could form in haematopoietic cells (Osada *et al.*, 1995). We have now shown that such a complex exists in erythroid cells and that this complex binds to DNA. The oligomeric complex, which also involves Ldb1, specifically recognizes a unique bipartite E-box-GATA motif, consisting of an E-box followed 9 bp downstream by a GATA site. In addition, the oligomeric complex can function in transcriptional activation. The data we present here demonstrate for the first time that a LIM-only protein can be part of a sequence-specific DNA-binding complex. Therefore, two classes of LIM protein interact with DNA; the LIM homeodomain proteins bind directly to DNA via their homeodomains (Karlsson *et al.*, 1990) and the LIM-only protein Lmo2 is associated with DNA because it is a component of a DNA-binding complex. However, there is still no evidence that the LIM domains of either class of LIM protein specifically contact DNA.

The CASTing procedure, when performed with nuclear extracts, preferentially selects high affinity binding sites for complexes of proteins, as opposed to lower affinity binding sites for the individual protein components (Funk and Wright, 1992; Wright and Funk, 1993). Since the majority of clones from the anti-Lmo2 and anti-TAL1

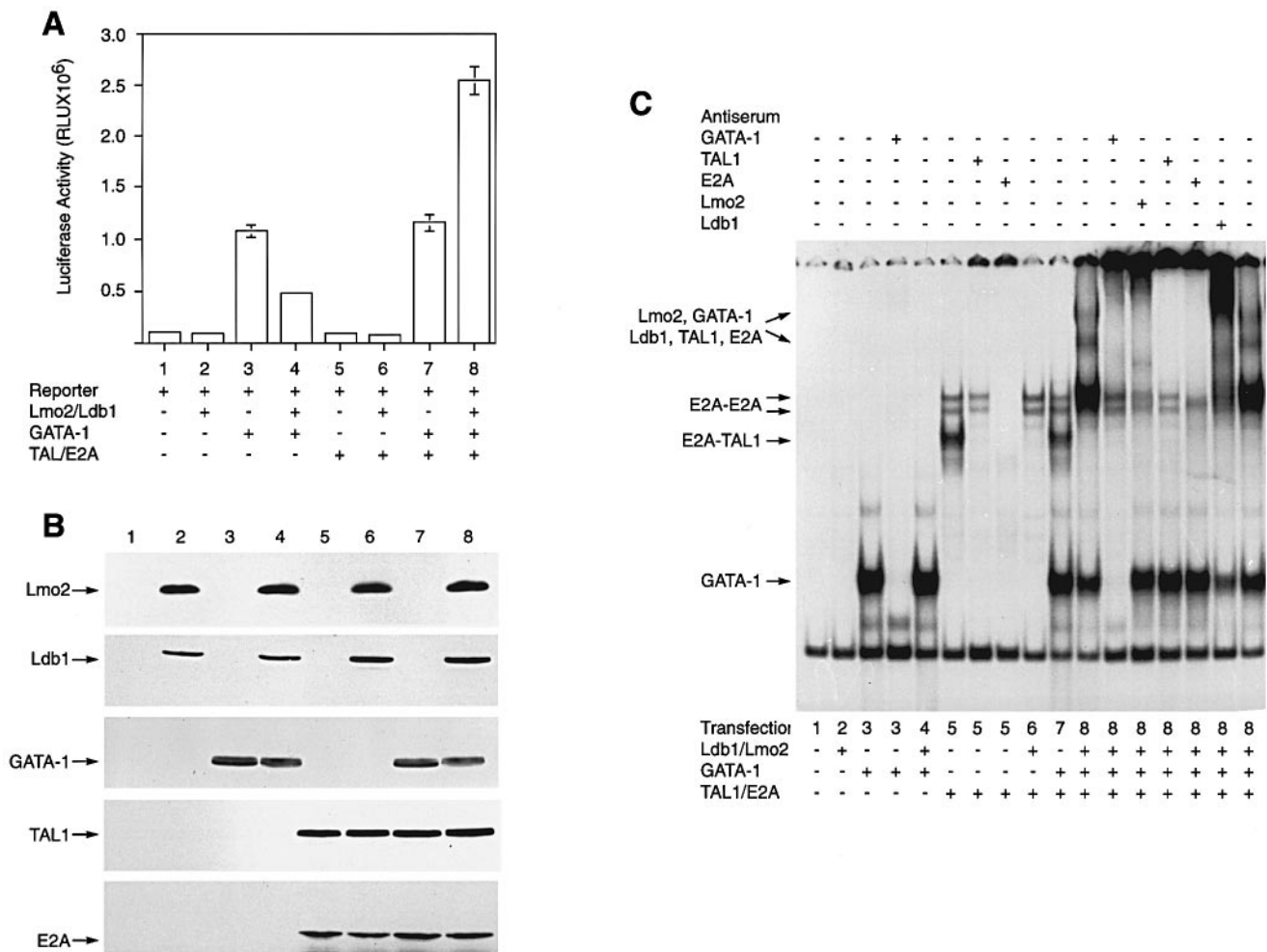


Fig. 7. Assembly and function of the E-box–GATA binding complex *in vivo*. COS cells were transiently transfected with a luciferase reporter plasmid with a minimal promoter linked to two E-box–GATA oligonucleotides, together with a β -galactosidase control plasmid, in the absence (transfection 1) or presence of mixtures of expression clones. After 48 h, cells were harvested, extracts made and aliquots used for luciferase and β -galactosidase assays (A), Western detection of proteins (B) or DNA-binding gel retardation assays (C). (A) Luciferase values were determined and normalized to β -galactosidase activity; each value is the mean from three transfections (\pm standard error). (B) Each of the transfections indicated in (A) (1–8) was assayed for protein expression by Western blotting with antisera raised against Lmo2, Ldb1, GATA-1, TAL1 and E47. The position of the relevant protein is arrowed. (C) EMSA of DNA–protein complexes. Extracts from transfections 1–8 were incubated with a 32 P-labelled consensus E-box–GATA oligonucleotide in the presence or absence of antisera, and protein–DNA complexes were analysed on 4% native polyacrylamide gels. The transfection number (and the proteins expressed therein) are indicated at the bottom of (C) (note some transfections, e.g. number 8, were used in several distinct band shift lanes) while the antisera, used to supershift complexes, are indicated at the top of (C). Positions of bands corresponding to relevant DNA–protein complexes are indicated.

CASTing experiments encoded an E-box in combination with a GATA site, this E-box–GATA motif probably corresponds to a high affinity binding site. Any lower affinity binding sites for Lmo2 or Tal1 would have been under-represented in the final PCR products. Although the DNA recognition motif is composed of two previously identified binding sites (an E-box and a GATA site) several features demonstrate that it is a novel recognition site for a distinct transcription factor complex. The relative orientation of the E-box and GATA sites was conserved in the CASTing sequences, as the E-box always occurred upstream of the GATA site. In addition, the spacing between the two sites appeared important because protein binding was only observed when the E-box and GATA sites were separated by 8, 9 or 10 bp of DNA, indicative of binding on one surface of the double helix. Taken together, these data suggest that the E-box–GATA-binding complex not only recognizes the spatial separation of the

E-box and GATA sites in the motif, but also binds to DNA in a sequence-specific manner.

A simple model for the molecular role of Lmo2 is that this protein forms a bridge, probably in conjunction with Ldb1/NLI, between DNA-binding elements, in this case TAL1–E47 heterodimers and GATA-1 (Figure 8). The complex of Lmo2, Ldb1, TAL1, E2A and GATA-1 bound to the E-box–GATA motif may therefore contain multimers of Lmo2 and Ldb1. An Lmo2 molecule may be in contact with TAL1 on one side and with a dimer of Ldb1 on the other, which in turn touches a second Lmo2 molecule which contacts GATA-1. The formation of the oligomeric complex may occur in the presence of DNA such that GATA-1 and/or TAL1–E47 may contact DNA recognition sites on the chromosome and form a ‘nucleation site’ for the assembly of the oligomeric complex. Thus the large quantity of free GATA-1, not bound to DNA, would provide a pool from which this process could be started.

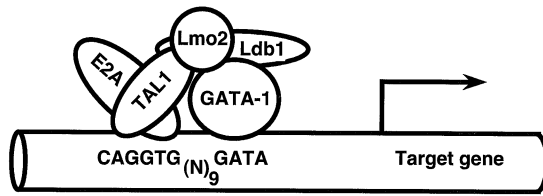


Fig. 8. Model of the Lmo2-containing oligomeric DNA-binding complex. The oligomeric complex binds to the E-box–GATA motif with a restricted spacing of 9 bp (at least as determined by CASTing with the R76 oligonucleotide). The stoichiometry of the complex is unknown, but DNA-binding modules within the oligomeric complex are provided by E47–TAL1 heterodimer (E-box binding) and GATA-1 (GATA binding). The Lmo2 LIM-only protein, together with Ldb1, links the two DNA-binding moieties. Target genes with E-box–GATA motif(s) provide recognition sites for the oligomeric complex. Transcriptional transactivation of these genes is a possible consequence of binding to target genes, but repressive functions may also occur, perhaps depending on the stage of haematopoiesis at which binding occurs.

The newly identified LIM-binding protein Ldb1/NLI plays a role in the E-box–GATA-binding complex presumably by binding to the LIM protein Lmo2. Unless Ldb1 has distinct DNA-binding or enhanced transcriptional transactivation functions, its function *in vivo* is probably to mediate protein–protein binding activity (Agulnick *et al.*, 1996; Jurata *et al.*, 1996). We consistently observed two low mobility protein–DNA complexes bound to the E-box–GATA consensus oligonucleotide which may be due to one or more of the components of the complex being present at a variable copy number. In addition, TAL1 and GATA-1 are both subject to post-translational modification (Cheng *et al.*, 1993a,b; Crossley and Orkin, 1994), which might also affect the electrophoretic mobility of the proteins. Alternatively, there may be other proteins in the complex which we have not yet identified.

The role of Lmo2 protein-associated complexes in haematopoiesis

Lmo2, TAL1 and GATA-1 are essential for stages of haematopoiesis as shown by gene targeting experiments (Pevny *et al.*, 1991; Warren *et al.*, 1994; Weiss *et al.*, 1994; Robb *et al.*, 1995, 1996; Shivdasani *et al.*, 1995; Porcher *et al.*, 1996). Our molecular data link these three gene products together into a DNA-binding complex (Figure 8), suggesting that the similar phenotypes result from the failure of recognition of common target genes, carrying the E-box–GATA motif in their regulatory regions. This might be achieved by modulating their transcription during haematopoiesis by activation or repression.

A search of DNA databases revealed a number of genes with consensus E-box–GATA motifs (with 8, 9 or 10 bp separation between E-box and GATA sequences) located adjacent to their promoters, including the genes for glycoporphin A and B and porphobilinogen deaminase. Other unidentified genes may also possess these regulatory elements. The sequences adjacent to the identified promoters, however, contain many other potential regulatory elements. While the oligomeric complex was found to bind to these promoters (unpublished data), reporter constructs built with these complex sequences did not yield data which allowed the distinction between the effects of

E-box–GATA and other control elements (unpublished results).

During the early events of haematopoiesis, GATA-2 predominates over GATA-1, and targeting experiments of the *GATA-2* gene have shown that null *GATA-2* mutations have an effect on early haematopoiesis (Tsai *et al.*, 1994). Therefore, complexes involving Lmo2 in earlier haematopoietic lineages might utilize GATA-2 rather than GATA-1 since both can bind Lmo2 (Osada *et al.*, 1995). In addition, Lmo2, GATA-1 and E2A can form homodimers (Murre *et al.*, 1989; Sun and Baltimore, 1991; Crossley *et al.*, 1995; Merika and Orkin, 1995; Sanchez-Garcia *et al.*, 1995) or heterodimers TAL1–E2A (Hsu *et al.*, 1991, 1994b) and Lmo2–GATA-1 (Osada *et al.*, 1995). Thus it is conceivable that various Lmo2-containing complexes, with or without GATA-1 or 2, bind to specific target genes, and presumably activate (or repress) their transcription during various stages of haematopoiesis (Figure 8). This would permit variations in both DNA-binding site recognition and in composition of the oligomers, thereby yielding precise control of target gene expression patterns during the differentiation process. In the situation of LMO2 and TAL1 protein co-expression in leukaemic T cells, analogous oligomeric complexes are conceivable. When LMO2 is expressed in the absence of TAL1, it may bind different molecules, thereby either activating, or repressing, different target gene populations during leukaemogenesis.

Materials and methods

Cell lines

MEL and subclone F4N, HEL and 707 cell lines were maintained in RPMI growth medium supplemented with 10% fetal bovine serum. C3H10T1/2, COS-7 and NS0 cell lines were cultured in Dulbecco's modified Eagle's medium plus 10% fetal bovine serum.

Antisera and Western blotting

Rabbit anti-peptide antisera recognizing Lmo2 (residues 2–17) and mouse GATA-1 (residues 5–20 or 376–391) have been described previously (Warren *et al.*, 1994; Osada *et al.*, 1995). Antisera against human GATA-2 (residues 5–20) and human GATA-3 (residues 413–428) were made as described (Warren *et al.*, 1994). The anti-TAL1 No. 1080 (residues 1–121), anti-TAL1 No. 370 (residues 238–331) and anti-E2A No. 526 (residues 217–371) rabbit antisera have been described elsewhere (Cheng *et al.*, 1993b; Hsu *et al.*, 1994a,b). Rabbit anti-peptide antiserum recognizing Ldb1 (A.D. Agulnick and H. Westphal, unpublished) was raised against a peptide corresponding to residues 256–270 of Ldb1 (Agulnick *et al.*, 1996). Western blotting was performed with 16 µg of nuclear extract as described in Larson *et al.* (1996) using the rabbit anti-Lmo2 and anti-Ldb1 antisera. TAL1 was detected with mouse monoclonal antiserum BTL73 (the kind gift of K. Pulford), E47 with mouse monoclonal antiserum YAE (Santa Cruz Biotechnology Inc.) and GATA-1 with rat monoclonal antiserum N6 (also from Santa Cruz Biotechnology, Inc.).

Preparation of nuclear extract for CASTing

Crude nuclear extracts were prepared from MEL, or MEL subclone F4N cells, essentially as described (Lee *et al.*, 1988). MEL cells (3×10^7) were harvested, washed twice with phosphate-buffered saline (PBS) and resuspended in 5 ml of hypotonic buffer [10 mM Tris–HCl, 1.5 mM MgCl₂, 10 mM KCl, 1 mM dithiothreitol (DTT), 50 µM ZnAc, pH 7.9] with protease inhibitors [1 mM phenylmethylsulfonyl fluoride (PMSF), 1 µg/ml pepstatin A, 1 µg/ml aprotinin, 1 µg/ml leupeptin]. After a 10 min incubation at 4°C, the cells were homogenized with a Dounce homogenizer, centrifuged at 1000 g for 5 min, and the pellet of nuclei resuspended in 500 µl of KCl-free buffer (20 mM Tris–HCl, 20% glycerol, 1.5 mM MgCl₂, 1 mM DTT, pH 7.9) with protease inhibitors. KCl was added to a final concentration of 0.3 M, the lysate was incubated for 30 min, and centrifuged at 13 000 g for 15 min. Dilution buffer

containing protease inhibitors was added to the supernatant (750 µl of 20 mM Tris-HCl, 20% glycerol, 1 mM DTT, 50 µM ZnAc, pH 7.9), and the sample was centrifuged again at 13 000 g for 10 min. The protein concentration of the nuclear extract was estimated by the Bradford assay (BioRad) and aliquots were stored at -70°C.

CASTing procedure

Enrichment for specific DNA-binding sites from a pool of random oligonucleotides was performed approximately as described (Blackwell *et al.*, 1990; Pollock and Treisman, 1990). MEL nuclear extract (200 µg) was mixed with 300 ng of double-stranded random oligonucleotide R76 (Pollock and Treisman, 1990) and 20 µg of poly(dI-dC), in 600 µl of binding buffer (20 mM HEPES, 100 mM NaCl, 10% glycerol, 0.1% NP-40, 50 µM ZnAc, 0.5% bovine serum albumin, 1 mM DTT, pH 7.4) at 4°C for 60 min. Five µl of the appropriate antiserum (anti-Lmo2, anti-TAL1 No. 1080, anti-E2A No. 526 and anti-NH₂-terminal GATA-1 peptide) and 0.6 µl of 1 M iodoacetamide were added, and the incubation was continued for a further 30 min. Fifty µl of a 50% protein A-Sepharose slurry was added, and after 30 min the beads were washed four times with binding buffer, incubated with 200 µg/ml proteinase K in digestion buffer (50 mM Tris-HCl, 20 mM EDTA, 1% SDS, 10 mM NaCl, pH 7.4) for 1 h at 50°C, and the sample was extracted with phenol. Oligonucleotides were purified by non-denaturing PAGE, precipitated with ethanol in the presence of 10 µg of glycogen, resuspended in 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.5, and amplified by PCR. PCR products from five (Figures 1A and 2A and C) or seven (Figure 2B) rounds of selection were subcloned and sequenced. Consensus binding sites for CASTing performed with anti-Lmo2 and anti-TAL1 antisera were calculated from data from two separate experiments.

Preparation of nuclear extract for EMSA

Cells (5×10^7) were washed twice with 10 ml of cold PBS, resuspended in 500 µl of buffer A (10 mM HEPES, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT pH 7.9) with protease inhibitors (100 µg/ml aprotinin, 5 µg/ml leupeptin, 1 µg/ml pepstatin A, 0.5 mM PMSF) (Lee *et al.*, 1988) and incubated on ice for 15 min. NP-40 was added to a final concentration of 0.5%, and the cells were vortexed for 10 s. The nuclei were pelleted by centrifugation at 6500 g for 20 s, and resuspended in 150 µl of buffer C (20 mM HEPES, 1.5 mM MgCl₂, 420 mM NaCl, 0.2 mM EDTA, 25% v/v glycerol pH 7.9) with protease inhibitors. The nuclear suspension was stirred vigorously on ice with a small, magnetic stirrer bar for 30 min. The sample was centrifuged at 13 000 g for 10 min, and aliquots of the nuclear extract were frozen immediately on dry ice. Samples were stored under liquid nitrogen. The protein concentration of the nuclear extract was determined by the Bradford assay (BioRad).

EMSA

EMSA were performed with ³²P-labelled double-stranded oligonucleotides, essentially as described previously (Hsu *et al.*, 1991). The sequence of the sense strands of individual oligonucleotides is shown below. MEL subclone F4N nuclear extract (5 µg) was incubated for 20 min at 25°C in binding buffer (20 mM HEPES, 50 mM KCl, 1 mM EDTA, 25% v/v glycerol, 1 mM DTT, pH 7.6) in the presence of 40 µg/ml of sonicated *Escherichia coli* DNA and 2 nM ³²P-labelled oligonucleotide. Where indicated, 1 µl of the appropriate rabbit antiserum [anti-Lmo2, anti-TAL1 Nos 370 (Figure 6) and 1080 (Figure 4), anti-E2A No. 526, anti-GATA-1 residues 376–391 and anti-Ldb1] and 1 µg of *Staphylococcus aureus* protein A (Sigma Cat. No. P-6650) were added. However, anti-Ldb1 supershifts were performed using 2 µg of protein A. Nuclear extracts from transfected COS-7 cells were treated identically, except that each sample contained 3 µg of protein and 3 nM oligonucleotide, and the anti-Lmo2 supershift was performed with 2 µl of antibody and 2 µg of protein A. Samples were loaded onto 4% native polyacrylamide gels, which were run for 4 h at 13 mA constant current in 0.5× TBE (45 mM Tris, 45 mM boric acid, 1.25 mM EDTA). After drying the gel, the DNA-protein complexes were visualized by autoradiography. For clarity, the band corresponding to the free probe has been cut from the autoradiograph; however, all EMSA experiments were carried out with an excess of free probe.

Plasmid clones

Complementary oligonucleotides encoding the E1b TATA box (GATCAC-TAGTAGGGTATATAATGGAATTCA and GATCTGAATCCATTAT-ATACCCTACTAGT) were phosphorylated, annealed and ligated into the *Bgl*III site of the luciferase reporter vector pGL3-Basic (Promega), to make the plasmid E1bLUC. Two copies of a pair of complementary

oligonucleotides encoding the E-box-GATA consensus (TCGACCGC-CAGGTGCTGCGTCCCAGATAGGGGCC and TCGAGGCCCTAT-CGGGACGACGACCTGGCGG) were then inserted in the same orientation into the *Xho*I site of E1bLUC to create the luciferase reporter (E-box GATA)₂-E1bLUC. The pEF-BOS-β-galactosidase control plasmid has been described before (Osada *et al.*, 1995). Expression plasmids encoding Lmo2, TAL1, murine GATA-1 and human E47 were constructed by subcloning the relevant cDNA sequences into the vector pEF-BOS, as described previously (Warren *et al.*, 1994; Osada *et al.*, 1995). A pEF-BOS-Ldb1 expression plasmid was constructed by excising a *Hind*III-*Xba*II fragment from the plasmid Ldb1-pcDNA 1/Amp, and blunt cloning it into pEF-BOS.

COS-7 transfections, luciferase and β-galactosidase assays

COS-7 cells (1.4×10^6), were seeded onto 100 mm plates in 10 ml of medium, 30 h prior to calcium phosphate-mediated transfection (Gibco-BRL). Each sample was transfected with 10 µg of (E-box GATA)₂-E1bLUC reporter plasmid, 0.25 µg of pEF-BOS-βGAL control plasmid and 17 µg of pEF-BOS expression plasmids, and the transfections were set up in triplicate. The amount of each expression plasmid used was 5 µg of pEF-BOS-Lmo2, 5 µg of pEF-BOS-TAL1, 3 µg of pEF-BOS-GATA-1, 3 µg pEF-BOS-E47 and 1 µg of pEF-BOS-Ldb1. In those samples where the total amount of pEF-BOS expression plasmids was <17 µg, pEF-BOS vector was added to provide a constant mass of DNA. The cells were harvested from each plate after 36 h and were resuspended in 150 µl of buffer A, lysed with 0.5% NP-40 and the nuclei were pelleted as described above. The luciferase activity of each post-nuclear supernatant was normalized with respect to the β-galactosidase activity of the sample, as described previously (Hsu *et al.*, 1994b). Nuclear extracts were prepared by pooling the nuclear pellets from the triplicate transfections in 200 µl of buffer C.

Sequence of oligonucleotides used in EMSAs

1. 9 bp spacing, consensus E-box-GATA oligonucleotide:
5'-TCGGCGCCAGGTGCTGCGTCCCAGATAGGGGCCG-3'
2. 9 bp spacing, mutant E-box oligonucleotide:
5'-TCGGCGCGAGGTCTGCGTCCCAGATAGGGGCCG-3'
3. 9 bp spacing, mutant GATA oligonucleotide:
5'-TCGGCGCCAGGTGCTGCGTCCCAGTCCGGGCCG-3'
4. 9 bp spacing, double mutant oligonucleotide:
5'-TCGGCGCGAGGTCTGCGTCCCAGTCCGGGCCG-3'
5. TAL1-E2A consensus E-box oligonucleotide:
5'-ACCTGAACAGATGGTTCGGCTACCGAGCGAGGGT-3'
6. 7 bp spacing, consensus E-box-GATA oligonucleotide:
5'-TCGGCGCCAGGTGCTGCCCGATAGGGGCCG-3'
7. 8 bp spacing, consensus E-box-GATA oligonucleotide:
5'-TCGGCGCCAGGTGCTGCGCCCCAGATAGGGGCCG-3'
8. 10 bp spacing, consensus E-box-GATA oligonucleotide:
5'-TCGGCGCCAGGTGCTGCGTACCCGATAGGGGCCG-3'
9. 11 bp spacing, consensus E-box-GATA oligonucleotide:
5'-TCGGCGCCAGGTGCTGCGTCCCGATAGGGGCCG-3'
10. 12 bp spacing, consensus E-box-GATA oligonucleotide:
5'-TCGGCGCCAGGTGCTGCGTCCCGATAGGGGCCG-3'.

Acknowledgements

We would like to thank Dr H.Axelsson for carrying out database searches and Dr G.Smith for helpful advice and discussions. The 707 and MEL cell lines were a gift from Dr M.Greaves, and the MEL subclone F4N was a gift from Dr A.Green. We are especially grateful to Dr R.Baer for providing the anti-TAL1 and anti-E2A antisera, and Dr C.Murre for anti-E2A antiserum. I.W. and H.O. were supported by the Leukaemia Research Fund, and G.G. was funded by an EMBO fellowship.

References

- Agulnick,A.D., Taira,M., Breen,J.J., Tanaka,T., Dawid,I.B. and Westphal,H. (1996) Functional and physical interaction of Ldb1, a novel LIM domain binding factor, with the LIM homeodomain protein Lhx1/Xlim-1. *Nature*, **384**, 270–272.
- Baer,R. (1993) TAL1, TAL2 and LYL1: a family of basic helix-loop-helix proteins implicated in T cell acute leukaemia. *Semin. Cancer Biol.*, **4**, 341–347.

- Begley, C.G., Aplan, P.D., Denning, S.M., Haynes, B.F., Waldmann, T.A. and Kirsch, I.R. (1989) The gene *SCL* is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. *Proc. Natl Acad. Sci. USA*, **86**, 10128–10132.
- Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) Sequence-specific DNA binding by the c-Myc protein. *Science*, **250**, 1149–1151.
- Boehm, T., Foroni, L., Kaneko, Y., Perutz, M.P. and Rabbitts, T.H. (1991) The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc. Natl Acad. Sci. USA*, **88**, 4367–4371.
- Carroll, A.J., Crist, W.M., Link, M.P., Amylon, M.D., Pullen, D.J., Ragab, A.H., Buchanan, G.R., Wimmer, R.S. and Vietti, T.J. (1990) The t(1;14)(p34;q11) is nonrandom and restricted to T-cell acute lymphoblastic leukemia: a Pediatric Oncology Group study. *Blood*, **76**, 1220–1224.
- Chen, Q. *et al.* (1990) The *tal* gene undergoes chromosome translocation in T cell leukaemia and potentially encodes a helix-loop-helix protein. *EMBO J.*, **9**, 415–424.
- Cheng, J.-T., Cobb, M.H. and Baer, R. (1993a) Phosphorylation of the TAL1 oncoprotein by the extracellular-signal-regulated protein kinase ERK1. *Mol. Cell. Biol.*, **13**, 801–808.
- Cheng, J.-T., Hsu, H.-L., Hwang, L.-Y. and Baer, R. (1993b) Products of the *TAL1* oncogene: basic helix-loop-helix proteins phosphorylated at serine residues. *Oncogene*, **8**, 667–683.
- Crossley, M. and Orkin, S.H. (1994) Phosphorylation of the erythroid transcription factor GATA-1. *J. Biol. Chem.*, **269**, 16589–16596.
- Crossley, M., Merika, M. and Orkin, S.H. (1995) Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Mol. Cell. Biol.*, **15**, 2448–2456.
- Feuerstein, R., Wang, X., Song, D., Cooke, N.E. and Liebhauer, S.A. (1994) The LIM/double zinc-finger motif functions as a protein dimerisation domain. *Proc. Natl Acad. Sci. USA*, **91**, 10655–10659.
- Finger, L.R., Kagan, J., Christopher, G., Kurtzberg, J., Hershfield, M.S., Nowell, P.C. and Croce, C.M. (1989) Involvement of the *TCL5* gene on human chromosome 1 in T-cell leukemia and melanoma. *Proc. Natl Acad. Sci. USA*, **86**, 5039–5043.
- Foroni, L. *et al.* (1992) The rhombotin gene family encode related LIM-domain proteins whose differing expression suggests multiple roles in mouse development. *J. Mol. Biol.*, **226**, 747–761.
- Funk, W.D. and Wright, W.E. (1992) Cyclic amplification and selection of targets for multicomponent complexes: myogenin interacts with factors recognizing binding sites for basic helix-loop-helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2, and COMP1 factor. *Proc. Natl Acad. Sci. USA*, **89**, 9484–9488.
- Hsu, H.-L., Cheng, J.-T., Chen, Q. and Baer, R. (1991) Enhancer-binding activity of the *tal-1* oncoprotein in association with the E47/E12 helix-loop-helix proteins. *Mol. Cell. Biol.*, **11**, 3037–3042.
- Hsu, H.-L., Huang, L., Tsan, J.T., Funk, W., Wright, W.E., Hu, J.-S., Kingston, R.E. and Baer, R. (1994a) Preferred sequences for DNA recognition by the TAL1 helix-loop-helix proteins. *Mol. Cell. Biol.*, **14**, 1256–1265.
- Hsu, H.-L., Wadman, I. and Baer, R. (1994b) Formation of *in vivo* complexes between the TAL1 and E2A polypeptides of leukemic T cells. *Proc. Natl Acad. Sci. USA*, **91**, 3181–3185.
- Hsu, H.-L., Wadman, I., Tsan, J.T. and Baer, R. (1994c) Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. *Proc. Natl Acad. Sci. USA*, **91**, 5947–5951.
- Jurata, L.W., Kenny, D.A. and Gill, G.N. (1996) Nuclear LIM interactor, a rhombotin and LIM homeodomain interacting protein, is expressed early in neuronal development. *Proc. Natl Acad. Sci. USA*, **93**, 11693–11698.
- Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. and Edlund, T. (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a cys-his domain. *Nature*, **344**, 879–882.
- Larson, R.C., Lavenir, I., Larson, T.A., Baer, R., Warren, A.J., Wadman, I., Nottage, K. and Rabbitts, T.H. (1996) Protein dimerisation between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. *EMBO J.*, **15**, 1021–1027.
- Lee, K.A.W., Bindereif, A. and Green, M.R. (1988) A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. *Gene Anal. Technol.*, **5**, 22–31.
- Merika, M. and Orkin, S.H. (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel family proteins Sp1 and EKLF. *Mol. Cell. Biol.*, **15**, 2437–2447.
- Murre, C., McCaw, P.S. and Baltimore, D. (1989) A new DNA binding and dimerisation motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell*, **56**, 777–783.
- Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438–446.
- Osada, H., Grutz, G., Axelson, H., Forster, A. and Rabbitts, T.H. (1995) Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc. Natl Acad. Sci. USA*, **92**, 9585–9589.
- Perez-Alvarado, G.C., Miles, C., Michelsen, J.W., Louis, H.A., Winge, D.R., Beckerle, M.C. and Summers, M.F. (1994) Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP. *Nature Struct. Biol.*, **1**, 388–398.
- Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.-F., D'Agati, V., Orkin, S.H. and Costantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**, 257–260.
- Pollock, R. and Treisman, R. (1990) A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acid Res.*, **18**, 6197–6204.
- Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W. and Orkin, S.H. (1996) The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell*, **86**, 47–57.
- Rabbitts, T.H. (1994) Chromosomal translocations in human cancer. *Nature*, **372**, 143–149.
- Robb, L., Lyons, L., Li, R., Hartley, L., Kontgen, F., Harvey, R.P., Metcalf, D. and Begley, C.G. (1995) Absence of yolk sac hematopoiesis from mice with a targeted disruption of the *scl* gene. *Proc. Natl Acad. Sci. USA*, **92**, 7075–7079.
- Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnet, L.D. and Begley, C.G. (1996) The *scl* gene product is required for the generation of all hematopoietic lineages in the adult mouse. *EMBO J.*, **15**, 4123–4129.
- Roberts, V.J., Steenbergen, R. and Murre, C. (1993) Localization of E2A mRNA expression in developing and adult rat tissues. *Proc. Natl Acad. Sci. USA*, **90**, 7583–7587.
- Royer-Pokora, B., Loos, U. and Ludwig, W.-D. (1991) TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t(1;14)(p13;q11). *Oncogene*, **6**, 1887–1893.
- Sanchez-Garcia, I. and Rabbitts, T.H. (1993) LIM domain proteins in leukaemia and development. *Semin. Cancer Biol.*, **4**, 349–358.
- Sanchez-Garcia, I., Axelson, H. and Rabbitts, T.H. (1995) Functional diversity of LIM proteins: amino-terminal activation domains in the oncogenic proteins RBTN1 and RBTN2. *Oncogene*, **10**, 1301–1306.
- Schmeichel, K.L. and Beckerle, M.C. (1994) The LIM domain is a modular protein-binding interface. *Cell*, **79**, 211–219.
- Shivdasani, R.A., Mayer, E. and Orkin, S.H. (1995) Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature*, **373**, 432–434.
- Sun, X.-H. and Baltimore, D. (1991) An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell*, **64**, 459–470.
- Tsai, S.-F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G. and Orkin, S.H. (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, **339**, 446–451.
- Tsai, F.-Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W. and Orkin, S.H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, **371**, 221–226.
- Valge-Archer, V.E., Osada, H., Warren, A.J., Forster, A., Li, J., Baer, R. and Rabbitts, T.H. (1994) The LIM protein RBTN2 and the bHLH protein TAL1 are present in a complex in erythroid cells. *Proc. Natl Acad. Sci. USA*, **91**, 8617–8621.
- Wadman, I., Li, J., Bash, R.O., Forster, A., Osada, H., Rabbitts, T.H. and Baer, R. (1994) Specific *in vivo* association between the bHLH and LIM proteins implicated in human T cell leukaemia. *EMBO J.*, **13**, 4831–4839.
- Warren, A.J., Colledge, W.H., Carlton, M.B.L., Evans, M.J., Smith, A.J.H. and Rabbitts, T.H. (1994) The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell*, **78**, 45–58.
- Weiss, M.J. and Orkin, S.H. (1995) GATA transcription factors: key regulators of hematopoiesis. *Exp. Hematol.*, **23**, 99–107.

- Weiss,M.J., Keller,G. and Orkin,S.H. (1994) Novel insights into erythroid development revealed through *in vitro* differentiation of GATA-1⁻ embryonic stem cells. *Genes Dev.*, **8**, 1184–1197.
- Wright,W.E. and Funk,W.D. (1993) CASTing for multicomponent DNA-binding complexes. *Trends Biochem. Sci.*, **18**, 77–80.
- Wright,W.E., Binder,M. and Funk,W. (1991) Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol. Cell. Biol.*, **11**, 4104–4110.
- Yang,H.-Y. and Evans,T. (1995) Homotypic interactions of chicken GATA-1 can mediate transcriptional activation. *Mol. Cell. Biol.*, **15**, 1353–1363.

Received on 12 December 1996; revised on 20 February 1997