

# DNA-binding domain ancestry

SIR—We would like to point out a similarity between the DNA-binding domain of the Myb family of proteins and homoeodomain-containing factors. This relationship indicates a common ancestry for the DNA-binding domains of these two classes of eukaryotic nuclear regulatory proteins, and predicts specific roles for parts of the Myb DNA-binding domain.

Genes similar to the *c-myc* proto-oncogene have been found in birds, mice, humans, *Drosophila melanogaster* and plants (*Zea mays*) (refs 1–6). The region common to the products of all these genes consists of imperfect repeats of 51–53 amino acids; in general there are three such repeats, except in *Z. mays*, where there are two, and in the two versions of v-Myb (encoded by the AMV and E26 viruses), which are truncated so that most of the first repeat is missing<sup>7</sup>. These repeats are responsible for DNA binding<sup>8</sup>,

and it is assumed that the rest of the molecule is involved in transactivation of the target gene(s) of Myb (ref. 9).

In the Fig., *a* shows an alignment of the twenty known Myb repeat sequences, grouped as first, second and third repeats. (*Z. mays* sequences are best aligned if assigned as repeats 2 and 3.) Eighteen positions are highly conserved for all the sequences, an additional 11 positions being included if the occurrence stringency is lowered to greater than 60 per cent. Sixteen of these positions match with a homoeodomain (HD) consensus if two gaps are introduced (*b* in Fig.).

NMR spectroscopy of the homoeodomain from the *Antennapedia*-gene (*Anp*) product demonstrates three  $\alpha$ -helical segments (*b* in Fig.) and confirms an earlier prediction that this domain is similar to bacterial repressors, with helices 2 and 3 forming the helix-turn-helix structure<sup>10</sup>.

The similarity between Myb and the homoeodomain is reinforced by a prediction of the secondary structure of the Myb repeat<sup>3</sup> which suggests three  $\alpha$ -helical

segments — designated here as Myb $\alpha$ 1–3. On the basis of conserved hydrophobicity and frequently occurring proline and glycine residues, the extent of the Myb $\alpha$ -helices can be estimated, revealing a close correspondence with the three helices of the homoeodomain<sup>10</sup> (see *b* in Fig.).

As expected, the two alignment gaps fall within the loops interconnecting the helices. If these extrapolations are correct we would predict that the Myb $\alpha$ 2 and Myb $\alpha$ 3 helices would form a helix-turn-helix motif, with helix Myb $\alpha$ 3 making direct contact with bases in the DNA. Furthermore, the most conserved positions<sup>11</sup> in the bacterial repressor helix-turn-helix motif (which is related to the homoeodomain) align well with positions in the Myb consensus, especially those used in helix prediction (see *b* in Fig.). Applying a statistical algorithm derived from the bacterial sequences used to generate the helix-turn-helix consensus (ref. 11, with the modification suggested in ref. 12) to the homoeodomain consensus helix2-turn-helix3 region and the predicted Myb $\alpha$ 2-turn-Myb $\alpha$ 3 domain of the third repeat in chicken Myb (discounting the single insertion), gave scores of 1,363 and 1,358, respectively, compared with 1,675 for the  $\lambda$ Cro repressor and 1,336 for the  $\lambda$ CI repressor.

Myb sequences within a particular repeat grouping are most highly related at the C terminus, in the region of helix Myb $\alpha$ 3 (as expected if this is the intimate DNA contact point). A 'patch' of positive amino acids, as seen at the end of helix 3 in the homoeodomain, can also be seen at the end of the helix Myb $\alpha$ 3 in the repeat 3 sequences but not at the equivalent position in repeats 1 and 2.

A specific role for tryptophans in the Myb DNA-binding structure has been recently suggested<sup>13</sup>; it now seems likely that this may involve their hydrophobic character as, by analogy to the bacterial repressors, the tryptophans would be within a hydrophobic core.

JONATHAN FRAMPTON

ACHIM LEUTZ

TOBY J. GIBSON

THOMAS GRAF

European Molecular Biology Laboratory,  
Postfach 10.2209,  
Meyerhofstrasse 1,  
6900 Heidelberg, FRG

<i>a</i>	1			52	
Drosophila	gfgkrwsksedvllkqlvethg	enweiigphfk	drleqqvqqrwakvlnpe		
Chicken	lgktrwtreedeklklveqngtedkvwiasflp	nrtadvqqrhwqkvlndpe			First Myb repeat
Mouse	lgktrwtreedeklklveqngtddwkviandy	nrtadvqqrhwqkvlndpe			
Human	rwteeteklklveqngtddwkviandy	nrtadvqqrhwqkvlndpe			
Human A	wrvkwtredtdklklveqngtdwtliashly	nrsdfqqrhwqkvlndpe			
Human B	kckvkwtheetqqlralvrqfgqddkwiashfp	nrtadvqqrhwqkvlndpe			
<i>Zea mays</i>	vkrgawtskeddalaayvkahegkqrevpqqaglr	rrcgksclrlwlnylrpn			
Drosophila	likgpwtrdeddmviklvrnfgpkwtliaryln	grigkqcrerwhhnlndpe			Second Myb repeat
Mouse	likgpwtkeedqrviqlvqkygppkrwsviakhlk	grigkqcrerwhhnlndpe			
Human/chicken	likgpwtkeedqrviqlvqkygppkrwsviakhlk	grigkqcrerwhhnlndpe			
Human A	likgpwtkeedqrviqlvqkygppkrwsviakhlk	grigkqcrerwhhnlndpe			
Human B	lvkqpwtkeedqrviqlvqkygppkrwsviakhlk	grigkqcrerwhhnlndpe			
<i>Zea mays</i>	irrgnisydeedliirhlrlly	nrwsliaqrlp	grtdneiknywnstlgr		
Drosophila	ikktawtekedeiyyqahlelg	nqwakiakrlp	grtdnaiknhwnstmrk		Third Myb repeat
Chick/mouse/human	vkktswteeedriiyqahkrig	nrwaeiaklpl	grtdnaiknhwnstmrk		
Human A	vkktswteeedriiyqahkrig	nrwaeiaklpl	grtdnsiknhwnstmrk		
Human B	vkktswteeedriiceahkvig	nrwaeiakmip	grtdnavknhwnstmrk		

Myb consensus #.k..WTSeEd...#...#...G..\$W..IA..L. gRtd.q#..rW...lnp\$

Positions used in helix prediction #..gp\*.....#...#...#Gp...#...#p g.....#...#...#p.

<i>b</i>				
Predicted Myb helices		Myb $\alpha$ 1 hhhhhhhhhhhh	Myb $\alpha$ 2 hhhhhhhhhhhh	Myb $\alpha$ 3 hhhhhhhhhhhh
Myb consensus	#.k..WTSeEd...#...#...G..\$W..IA..L.			gRtd.q#..rW...lnp\$
Myb-HD consensus	..+..@T\$*...#...#...\$...IA..L.			..t-.q#..@.....
HD consensus	rkRgRqtYTRYQtLELEKEFHFRRRRIEIAhaL		CLTERQIKIWFQNRMRKwKKen	60
NMR-derived <i>Anp</i> HD helices		hhhhhhhhhhhh	hhhhhhhhhhhh	hhhhhhhhhhhh
Bacterial repressor helix-turn-helix		..S#A..#	G#...#...#..	

*a*, Aligned repeats from Myb-related sequences<sup>1–6</sup> (single letter amino-acid code), grouped as first, second and third repeats. The derived Myb repeat consensus is shown below (capitals and symbols, >85 % occurrence in the 20 Myb sequences, lower-case letters, residues which occur in >60 % of the sequences). Also indicated are the critical positions for helix prediction; conserved hydrophobicity (#), repeated every 3–4 residues, indicates an amphipathic helix; frequently occurring proline (P) and glycine (G) residues indicate interruptions. *b*, Derived information and comparison of the Myb consensus to the homoeodomain consensus and the bacterial repressor helix-turn-helix motif. The central line of the Fig. indicates the sixteen residues common to the Myb consensus and the homoeodomain (HD) consensus<sup>14</sup>. The extent of the experimentally determined and predicted helical segments are indicated by H (certainly helical) and h (possibly helical). The lower part of the Fig. shows a consensus for bacterial helix-turn-helix motifs derived from a compilation of 37  $\lambda$ Cro-like proteins<sup>11</sup>. Residues at conserved positions occur in >60 % of the sequences. #, hydrophobic; @, large hydrophobic; \$, charged; +, positive; -, negative; \*, E or Q.

- Rosson, D. & Reddy, E.P. *Nature* **319**, 604–606 (1986).
- Gerondakis, S. & Bishop, J.M. *Molec. cell Biol.* **6**, 3677–3684 (1986).
- Gonda, T.J. *et al.* *EMBO J.* **4**, 2003–2008 (1985).
- Nomura, N. *et al.* *Nucleic Acids Res.* **16**, 11075 (1988).
- Katzen, A.L., Kornberg, T.B. & Bishop, J.M. *Cell* **41**, 449–456 (1985).
- Paz-Ares, J. *et al.* *EMBO J.* **6**, 3553–3558 (1987).
- Nunn, M.F. *et al.* *Nature* **306**, 391–395 (1983).
- Klemppner, K.H. & Sippel, A.E. *EMBO J.* **6**, 2719–2725 (1987).
- Nishina, Y. *et al.* *Nucleic Acids Res.* **17**, 107 (1989).
- Otting, G. *et al.* *EMBO J.* **7**, 4305–4309 (1988).
- Dodd, I.B. & Egan, J.B. *J. molec. Biol.* **194**, 557 (1987).
- Yudkin, M.D. *Protein Eng.* **1**, 371–372 (1987).
- Anton, I.A. & Frampton, J. *Nature* **336**, 719 (1988).
- Muller, M.M. *et al.* *Nature* **336**, 544–551 (1988).