Volume 283, number 2, 293–294 FEBS 09674 May 1991

ADONIS 001457939100383Q

The transcriptional control proteins c-Myb and v-Myb contain a basic region DNA binding motif

Mark D. Carr[1] and Richard F. Mott[2]

[1]*Laboratory of Molecular Structure and* [2]*Laboratory of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London, UK*

Received 13 February 1991

A class of transcriptional control proteins, typified by c-Myb, are characterised by a highly conserved N-terminal DNA binding domain, which is composed of either 2 or 3 imperfect repeats of about 52 amino acids. A sequence homology search of the SWISSPROT protein sequence data bank with this region from mouse c-Myb has allowed us to identify an area near the C-terminus of repeat 2 that contains a short sequence motif, known as the basic region, which forms the DNA binding site in both leucine zipper and helix–loop–helix transcription factors. We therefore propose that this region of repeat 2 and the homologous part of repeat 3 will form the Myb DNA binding site.

Myb; Basic region DNA binding motif

1. INTRODUCTION

The cellular proto-oncogene product c-Myb and its transforming viral derivative v-Myb are the original members of an expanding family of transcriptional control proteins, which are characterised by a highly conserved N-terminal region that has been shown to be responsible for the sequence-specific DNA binding activity of both c-Myb and v-Myb [1–5]. In the case of c-Myb this conserved region consists of 3 imperfect repeats of approximately 52 amino acids [4], however, only the second and third of these are required for c-Myb to bind to the consensus site PyAACG/TG [2,6]. Indeed, the v-Myb proteins carried by the transforming avian retroviruses AMV and E26 are active despite the fact that they are truncated at the N-terminus, resulting in the loss of the majority of the first repeat [1,3].

In this communication we would like to point out a previously unrecognised sequence similarity between the DNA binding domain of Myb proteins and that of the leucine zipper and helix–loop–helix transcription factor families.

2. MATERIALS AND METHODS

The sequence homology search through the SWISSPROT protein sequence data-bank was carried out on an NCUBE parallel processor using the Smith–Waterman local similarity algorithm [7]. The search string used was a 155-residue portion of the mouse c-Myb protein corresponding to the 3 highly conserved N-terminal repeats, which contain the DNA binding site.

Correspondence address: M.D. Carr, Laboratory of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London, UK

3. RESULTS AND DISCUSSION

The sequence homology search identified a number of short stretches of similarity between portions of the DNA binding domain from c-Myb and none-Myb family proteins, but in most cases their significance is unclear. However, a region near the C-terminus of repeat 2 from c-Myb was found to show substantial sequence homology (up to about 55% taking into account conservative substitutions) with a well characterised, 16-residue DNA binding motif, known as the basic region (Fig. 1), which is found in two other groups of transcription factors, the leucine zipper proteins such as c-Jun and helix–loop–helix proteins like MyoD [8]. In addition, it is clear from the sequence alignments shown in Fig. 1 that the homology between c-Myb and the Jun proteins extends beyond the basic motif to include the first 6 amino acids of the 29-residue leucine zipper dimerisation domain, even though Myb proteins do not contain a leucine zipper domain and bind to DNA as a monomer [2].

In the case of the Myb transcription factors the precise residues involved in sequence specific DNA binding have not been identified. However, on the strength of the sequence similarities described above we would like to propose that residues from the basic motif in repeat 2 of c-Myb and the homologous region of repeat 3 will mediate DNA binding. This suggestion is consistent with the known properties of both leucine zipper and helix–loop–helix transcription factors, which bind to DNA as dimers so that two basic motifs form the DNA binding site [4,8,9]. The results of site directed mutagenesis and chemical modification studies

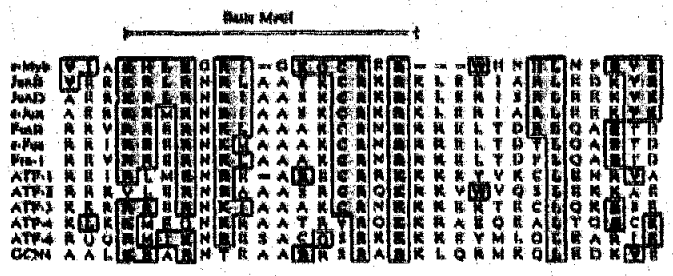

Fig. 1. A group of aligned partial protein sequences, which illustrate the strong sequence homology between the DNA binding basic motif from leucine zipper proteins and the c-terminal region of repeat 2 from the DNA binding domain of mouse c-Myb. It should be noted that over the stretch of c-Myb shown v-Myb differs in sequence by just the substitution of an aspartate for the first valine. In the figure conserved residues are contained within the shaded areas and single letter amino acid codes shown in bold denote identities.

on c-Fos/c-Jun leucine zipper heterodimers strongly suggest that the conserved cysteine residue in the basic motif is directly involved in sequence specific DNA binding [10]. Hence, the equivalent cysteine in repeat 2 of c-Myb might play a similar role.

A recent NMR study of the DNA binding domain from the yeast leucine zipper protein GCN4 showed that both the leucine zipper and basic motif adopt helical conformations in solution [11]. Thus, it seems likely that the C-termini of repeats 2 and 3 from c-Myb will also be predominantly helical. The Myb transcription factor family is therefore probably yet another example of where protein DNA contacts are mediated through residues located in protein helices.

REFERENCES

[1] Klempnauer, K.-H. and Sippel, A.E. (1987) EMBO J. 6, 2719–2725.
[2] Howe, K.M., Reakes, C.F.L. and Watson, R.J. (1990) EMBO J. 9, 161–169.
[3] Oehler, T., Arnold, H., Biedenkapp, H. and Klempnauer, K.-H. (1990) Nucleic Acids Res. 18, 1703–1710.
[4] Cole, M.D. (1990) Curr. Opin. Cell Biol. 2, 502–508.
[5] Luscher, B. and Eisenman, R.N. (1991) Genes Dev. 4, 2235–2241.
[6] Biedenkapp, H., Borgmeyer, U., Sippel, A.E. and Klempnauer, K.-H. (1988) Nature 335, 835–837.
[7] Smith, T.F. and Waterman, M.S. (1981) J. Mol. Biol. 147, 195–197.
[8] Prendergast, G.C. and Ziff, E.B. (1989) Nature, 341, 392.
[9] Sauer, R.T. (1990) Nature 347, 514–515.
[10] Abate, C., Patel, L., Rauscher III, F.J. and Curran, T. (1990) Science 249, 1157–1161.
[11] Saudek, V., Pastore, A., Castiglione Morelli, M.A., Frank, R., Gausepohl, H., Gibson, T., Weih, F. and Roesch, P. (1990) Protein Eng. 4, 3–10.