Bacteriophage Mu C protein is a new member of unusual leucine zipper-HTH class of proteins

Atanu Chakraborty[1], Bindu Diana Paul[2] and Valakunja Nagaraja[1,3,4]

[1]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India, [2]Laboratory of Gene Regulation and Development, National Institute of Child Health and Human Development, National Institute of Health, Bethesda, MD 20892, USA and [3]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064, India

[4]To whom correspondence should be addressed.
E-mail: vraj@mcbl.iisc.ernet.in

Transcription activator protein C of bacteriophage Mu activates transcription of the late genes, including *mom*, during the lytic cycle of the phage. C binding to its site leads to the alteration in DNA topology of the promoter elements resulting in RNA polymerase (RNAP) recruitment. At the next step, the transactivator enhances promoter clearance of RNAP from P_{mom}. The C protein binds DNA with a very high affinity using a carboxyl-terminal helix turn helix (HTH) motif which has similarity with the HTH from paired domain of *Drosophila* prd protein. Previous studies established that the protein is dimeric in free and DNA bound forms. We describe now the unique dimerization interface of the protein. Two heptad repeats of hydrophobic amino acids found in the protein were considered to be the candidates for dimerization region. Site-directed mutational analysis revealed that the amino-terminal coiled coil region is not the dimerization determinant. In contrast, similar mutagenesis studies indicated a role for the leucine zipper motif, located in the middle region of the protein, in dimerization. Mixed oligomerization assays confirmed the importance of leucine zipper in C dimer formation establishing the presence of an uncommon zipper-HTH domain in the transactivator.
Keywords: C protein/helix turn helix/leucine zipper/phage Mu/transcription activation

Introduction

The *mom* gene of bacteriophage Mu is unique in terms of the biochemical activity of its product and the regulation of expression (Kahmann and Hattman, 1987). Mom modifies the adenine residues in the DNA to *N*-acetamido adenine and makes the DNA resistant to an array of restriction enzymes (Hattman, 1999). The modification function is expressed only at the late lytic cycle of the phage life cycle. Premature expression of the gene is detrimental for both the phage and the host. Phage Mu has developed extraordinary regulatory mechanisms for repressing the *mom* gene expression both at the transcription and translation steps. The *mom* promoter comprises of poor −35 and −10 elements and suboptimal spacing of 19 bp between them. C protein, a middle gene product of the phage, is the transcription activator of the *mom* gene. In absence of C, RNA polymerase (RNAP) does not bind to the P_{mom} and instead transcribes from a divergent promoter *P2* (Balke *et al.*, 1992; Sun and Hattman, 1998). Transactivation of P_{mom} by C is unusual and involves at least two steps. The activator binds to its cognate site overlapping the −35 element of P_{mom} and asymmetrically untwists the DNA (Ramesh and Nagaraja, 1996, Basak and Nagaraja, 1998). C-mediated unwinding results in reorientation of promoter elements, allowing RNAP binding. Following RNAP recruitment, in the second step, C reduces abortive transcription and enhances promoter clearance (Chakraborty and Nagaraja, 2006).

C binds to its cognate site next to −35 element at P_{mom} as a dimer with very high affinity. Mutagenesis and homology modeling demonstrated that an extended helix turn helix (HTH) motif, consisting of a three-helix bundle preceded by a β turn, is responsible for DNA binding (Paul *et al.*, 2003). The analysis of the primary sequence of the protein revealed the presence of two putative hydrophobic repeat sequences as candidates for dimerization motif. Mutagenesis and chemical crosslinking studies ruled out the N-terminal heptad repeat to be involved in dimerization and suggested the second leucine zipper like motif as dimerization domain. We have characterized the dimerization interface of C further and show that C is a new member of the emerging family of bacterial dimeric DNA-binding proteins containing both HTH and leucine zipper.

Materials and methods

Bacterial strains and plasmids

Escherichia coli DH10B was used for cloning experiments. The expression strain *E. coli* BL26 (DE3) was obtained from Novagen (Madison, WI, USA). The plasmid pVR7 (Ramesh *et al.*, 1994), carrying *C* gene under the control of the T7 gene ϕ_{10} promoter was used both for expression of C protein and site-directed mutagenesis.

Mutagenesis and preparation of mutant proteins

Site-directed mutagenesis was carried out using the GeneEditor[TM] *in vitro* mutagenesis kit (Promega, Madison, WI, USA). The minimal heptad repeat (M44-Y92) was amplified with basic leucine zipper (bZip) forward and reverse primers using pVR7 as template. The PCR product was cloned into *Nco*I and blunted *Bam*HI sites of pET11d (Novagen) vector to obtain the minimal zipper polypeptide. Mutants were confirmed by sequencing the obtained clones. The *E. coli* BL26 (DE3) cells, harboring plasmids expressing wild-type (wt) or mutant C protein, were grown in LB containing 100 μg/ml ampicillin up to an OD of 0.6 at A_{600} and induced with 0.3 mM IPTG for 3 h at 37°C. After harvesting, cells were suspended in buffer containing 20 mM Tris-HCl, pH 8.0, 5 mM MgCl$_2$, 1 mM EDTA, 7 mM 2-mercaptoethanol, 5% glycerol, 100 mM NaCl and 1 mM PMSF and disrupted by sonication. The extract was centrifuged at 20 000g for 30 min at 4°C (S20 fraction). All the

mutant proteins were present in the insoluble S20 pellet fractions. The pellet fractions were resolved in 15% SDS-PAGE and visualized by negative staining using CuCl$_2$. The bands corresponding to the overexpressed proteins were excised and eluted from SDS polyacrylamide gel using a BioRad electroelutor. The eluted proteins were then subjected to acetone precipitation to remove SDS, denatured in refolding buffer (10 mM HEPES pH 7.4, 5 mM MgCl$_2$, 1 mM EDTA, 100 mM NaCl and 5% glycerol) containing 6 M urea and then renatured by stepwise dialysis against refolding buffer containing 4, 2 and 1 M urea, respectively. The proteins were then dialysed against the same buffer lacking urea.

Electrophoretic mobility shift assay

Five picomoles of the top strand of a 25 mer C-binding site with the sequence 5′-AGATCGATTATGCCCCAATAAC CAC-3′ was labeled using 40 μCi of [γ^{32}P]ATP (6000 Ci/ mmol, PerkinElmer) and mixed with 2.5-fold molar excess of the complementary strand and incubated for 5 min at 85°C. The mixture was allowed to cool slowly to room temperature following which the DNA was purified by passing through Sephadex G50.

The assays were carried out as described previously (Ramesh and Nagaraja, 1996), in 20 μl reactions in binding buffer (20 mM Tris-HCl pH 8.0, 5 mM MgCl$_2$, 1 mM EDTA, 7 mM 2-mercaptoethanol, 5% glycerol and 100 mM NaCl). DNA (10 000 cpm) was incubated with 100 ng of wt C protein or 1 μg of mutant proteins on ice for 15 min. The reactions were applied on a 10% native polyacrylamide gel and electrophoresed in 0.5 × TBE (45 mM Tris-borate, 1 mM EDTA pH 8.0) buffer at 150 V for 1 h at 4°C and subjected to autoradiography.

Dimerization assay

The ability of the C protein and the mutant proteins to form dimers was assessed by chemical crosslinking using glutaraldehyde (Merck) as described previously (Ramesh and Nagaraja, 1996). All the proteins under study were denatured in 6 M urea and sequentially renatured as described in the preceding section. Purified C or its mutants were dialyzed against crosslinking buffer (10 mM HEPES pH 7.4, 5 mM MgCl$_2$ pH 7.5, 1 mM EDTA, 100 mM NaCl, 7 mM 2-mercaptoethanol, 5% glycerol). C or mutant proteins (1 μg) were incubated with different concentrations of glutaraldehyde (0.0025 and 0.005%) at room temperature for 30 min. The reactions were stopped by the addition of SDS loading dye, subjected to 15% SDS-PAGE and visualized by silver staining. For mixed oligomerization assay, 1 μg C was mixed with 2 μg of minimal zipper polypeptide and incubated for 10 min in ice. Glutaraldehyde was added to a final concentration of 0.005% and the crosslinking was allowed for 15 min at room temperature. Reactions were stopped by the addition of SDS loading dye, analysed on 15% tricine SDS PAGE and visualized by silver staining.

Results and discussion

Primary sequence analysis of C protein

Transactivator C is a 16.5 kDa protein containing 140 amino acids. The protein is a dimer in solution and binds DNA in the dimeric form (De *et al.*, 1997). Within the short protein,

the HTH motif located towards the carboxy terminal is involved in DNA binding (Paul *et al.*, 2003). The protein contains two putative dimerization motifs as shown in Fig. 1. Towards the N-terminal region, a putative coiled coil motif was identified by analysing the primary sequence with COILS program (Lupas *et al.*, 1991). Coiled coil is characterized by the presence of hydrophobic residues in every first and fourth position in each heptad repeat. A 14 amino acid stretch (V35–R48) showed the probability of forming coiled coil where first and fourth amino acids are hydrophobic that could form hydrophobic interface for homodimerization. In the middle part of the protein, a putative leucine zipper motif (V63–L90) is also found. In a typical leucine zipper, every seventh amino acid is an L or an I. As represented by a helical wheel, four continuous heptad repeats in C protein contain L or I at the seventh position (Fig. 2). Both of the candidate regions were subjected to further analysis to determine the dimerization interface of C protein.

Characterization of the motifs in C protein

The amino-terminal coiled coil region has been analysed by site-directed mutagenesis. To disrupt the region, hydrophobic residues V or L were mutated to charged amino acid R. In spite of the non-conservative changes involving bulky replacements in the region, the mutants were efficient in dimerization as observed by glutaraldehyde crosslinking (Table I), indicating that the putative coiled coil region is not important for dimerization of the protein. In order to disrupt the putative leucine zipper, the repeated L residues were mutated individually by site-directed mutagenesis. They were substituted with M, R or S. All the single mutants of leucine zipper were efficient in dimerization as observed by glutaraldehyde crosslinking (Table I). They were also proficient in DNA binding indicating that dimerization is not significantly affected. However, a double mutant, L83RL90S, showed highly compromised dimerization. It has been observed in case of many leucine-zipper-containing proteins, *viz.* GCN4 homodimer (van Heeckeren *et al.*, 1992), Fos-Jun
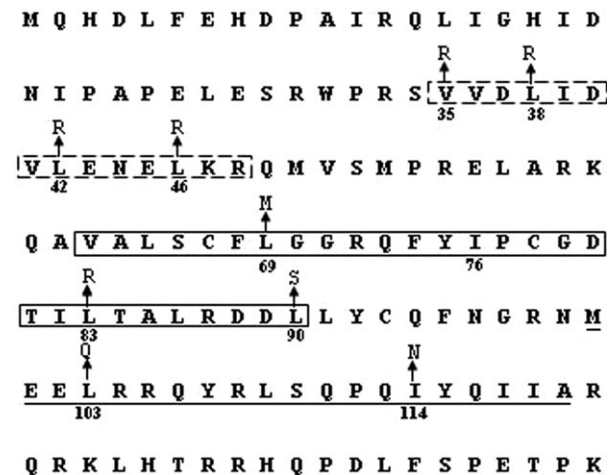

Fig. 1. The primary sequence of C protein. The coiled coil region is boxed with broken line and the leucine zipper is boxed with solid line. The residues in the HTH motif are underlined. The hydrophobic residues in the heptad repeats of coiled coil and leucine zipper are numbered. The mutations generated in the proteins are indicated.
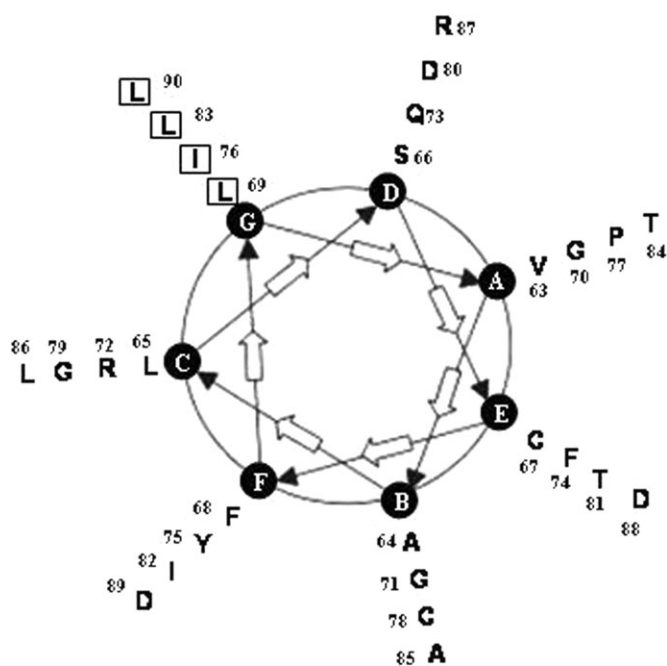
Fig. 2. Helical wheel representation of the leucine zipper. The heptad repeats are numbered from A to G. The positions of the amino acids are indicated. The hydrophobic residues, aligning along the G position of the wheel, are boxed.

heterodimer (Kouzarides and Ziff, 1988), MetR homodimer (Maxon *et al.*, 1990), that dimerization could not be effectively disrupted by a single mutation. Mutations in two conserved leucines were required in order to disrupt the leucine-zipper-mediated protein–protein interaction. The minimal leucine zipper polypeptide (minimal LZ, comprising residues 44 to 92 of the C protein, as shown in Fig. 3A), when tested, showed efficient crosslinking, indicating that the leucine zipper is responsible for dimerization (Table I).

In order to confirm the role for the leucine zipper in dimerization, we carried out a mixed oligomerization assay with full-length C protein and the minimal LZ (Fig. 3A) according to the strategy depicted in Fig. 3B. In this assay, full-length C would form homodimer upon crosslinking. If the minimal LZ possesses the dimerization domain, it would form a homodimer. When both the proteins are present, the minimal LZ would compete with the full-length C and form

a heterodimeric species. The size difference between the two proteins facilitates the separation of the three crosslinked species *viz.* C homodimer, minimal LZ homodimer and C-minimal LZ heterodimer, on a denaturing PAGE. The result of such mixed oligomerization assay is presented in Fig 3C. C protein showed dimerization upon crosslinking (Fig. 3C, lane 2). The minimal LZ also formed dimer under these assay conditions (Fig. 3C, lane 4). C protein (1 µg) was mixed with 2 µg of minimal LZ and subjected to glutaraldehyde crosslinking. In addition to C dimer and minimal LZ dimer, there was an additional crosslinked product with lower mobility than C monomer (Fig. 3C, lane 5). The product is the crosslinked species of C and minimal LZ. These results demonstrate that the leucine zipper region not only forms homodimer, it can also compete with full-length C to form a heterodimer, assigning a role for the leucine zipper in C dimer formation.
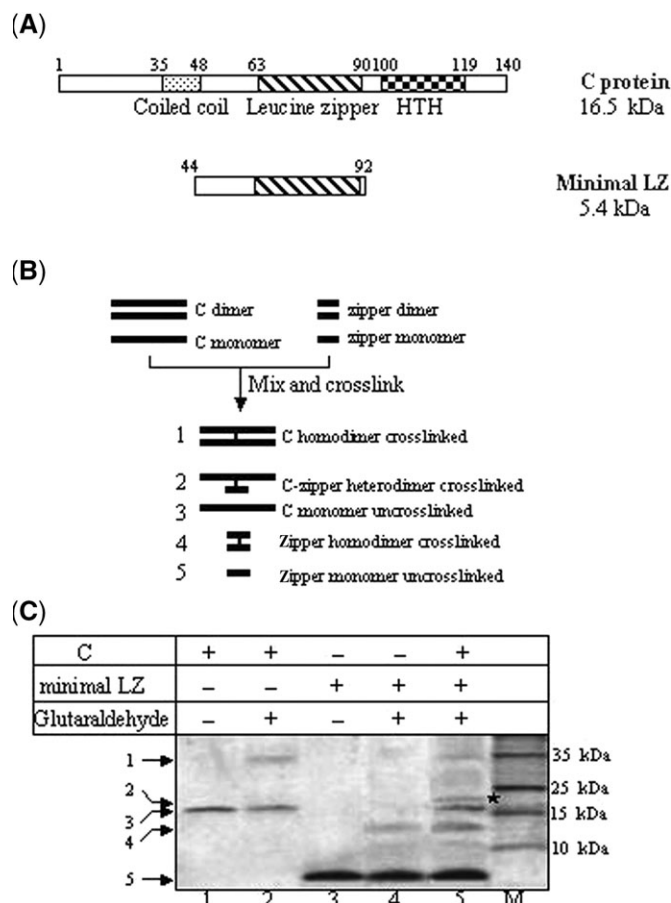

Fig. 3. (**A**) Proteins used in mixed oligomerization assay. The full-length C protein contains HTH motif, putative coiled coil and leucine zipper motifs shown in different shades. The minimal LZ contains the leucine zipper from residues 44 to 92 as shown. (**B**) Strategy for mixed oligomerization assay. Full-length C and minimal zipper polypeptides are shown with black bars of different sizes. The crosslinked products are numbered and shown according to their expected molecular weights. (**C**) Mixed oligomerization assay with C and minimal LZ. C (1 µg) and minimal LZ (2 µg) were crosslinked with 0.005% glutaraldehyde. Lanes 1 and 2, full-length C uncrosslinked and crosslinked, respectively. Lanes 3 and 4, minimal LZ uncrosslinked and crosslinked, respectively. Lane 5, 1 µg of C was mixed with 2 µg of minimal LZ and crosslinked. The reactions were analysed in 15% tricine SDS PAGE and silver stained. The numbers represent the protein species as shown in the above schematic representation in B. The heterodimer of C-minimal LZ is marked by asterix. Molecular weight markers are indicated.

Table I. Summary of mutagenesis studies

Mutant	Motif	Dimerization	DNA binding
V35R	Coiled coil	√	×
L38R	Coiled coil	√	×
L42R	Coiled coil	√	×
L46R	Coiled coil	√	×
L42RL46R	Coiled coil	√	×
L83R	Leucine zipper	√	√
L90S	Leucine zipper	√	√
L83RL90S	Leucine zipper	×	×
Minimal LZ	Leucine zipper	√	×
L103Q	HTH	√	×
I114 N	HTH	√	×

3

The leucine zipper does not participate in DNA binding by C protein

Presence of a stretch of basic amino acids preceding the leucine zipper in C protein prompted us to analyse the possible connection between leucine zipper and DNA binding. Eukaryotic bZIP motifs, which are responsible for dimerization and DNA binding, contain two conserved residues in the basic stretch, C and N, at −11 and −18 positions from the first L in the zipper (Lau *et al.*, 1997). The basic amino acid stretch in C protein does not contain the conserved C or N residues, suggesting that in this case the leucine zipper possibly does not influence DNA binding of the protein. Nevertheless, the DNA-binding efficiency of the dimerization-deficient double mutant of C, L83RL90S, and the minimal LZ was analysed using electrophoretic mobility shift assay (EMSA). The double mutant did not bind to the DNA, as the mutant is unable to dimerize (Fig. 4). The dimerization-efficient minimal LZ, containing the basic stretch, also did not bind to the DNA in EMSA, indicating that the basic stretch preceding the zipper is not contributing to DNA binding and the motif is not a typical bZIP motif (Fig. 4). Also, the DNA-binding mutants of C *viz.* L103Q and I114N that are part of the HTH motif, where zipper region is unaltered, did not show compromised dimerization. These results demonstrate that the HTH and the leucine zipper motifs are modular and dimerization activity of the leucine zipper is not affected by the HTH motif. The leucine zipper motif does not participate in DNA binding directly but influences indirectly as only dimeric C binds to the DNA.

A majority of the transcription activators bind to the DNA as dimers and use a variety of dimerization domains e.g. helix loop helix (HLH), leucine zipper, coiled coil etc. The present work demonstrates that C protein of phage Mu dimerizes through a leucine zipper motif, an uncommon oligomerization motif for a bacterial transcription activator. Leucine zipper motif was originally identified in eukaryotic
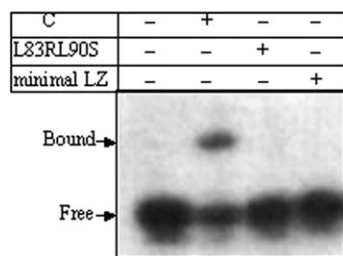

Fig. 4. Comparison of DNA binding of wt and mutant C. End-labeled DNA was incubated with 100 ng of wt C, 1 μg of L83RL90S or 2 μg of minimal LZ. The reactions were analysed on native PAGE and autoradiographed, as described in Materials and methods.

transcription activators GCN4, cMyc, Fos, Jun and C/EBP (Landschulz *et al.*, 1988). Subsequently, two motifs, basic HLH (bHLH) and bZIP have been shown to be involved in both dimerization and DNA binding in eukaryotic regulatory proteins where a stretch of basic amino acids is responsible for DNA recognition (Johnson and McKnight, 1989; Jones, 1990; Murre and Baltimore, 1992; Kadesch, 1993). A variant class of eukaryotic transactivators contains both HLH and zipper (Blackwood and Eisenman, 1991; Ferre-D'Amare *et al.*, 1993). To date, very few bacterial proteins have been shown to contain leucine zipper (Table II). MetR of *E. coli* is the first leucine zipper protein identified in prokaryotes (Maxon *et al.*, 1990). A bacterial histidine kinase, TodS, has been shown to dimerize and bind DNA through a bZIP motif (Lau *et al.*, 1997). Now, it appears that a small group of prokaryotic transcription activators could be classified as HTH-zipper family. Different kinds of HTH motifs are widespread in the bacterial transactivators carrying out DNA-binding function. However, the combination of HTH and zipper is rare, among a large number of DNA-binding proteins characterized so far. VirB, a regulator of the virulence cascade of *Shigella flexneri* has been shown to possess a HTH domain. The protein forms oligomers and mutagenesis studies have revealed that the dimerization and oligomerization are two separate functions. A leucine zipper motif is responsible for the dimerization of the protein while the oligomerization takes place through a carboxyl-terminal coiled coil (Beloin *et al.*, 2002). In contrast, in the case of well known HTH-containing protein, lac repressor (Weber *et al.*, 1982; Kaptein *et al.*, 1985), a leucine zipper motif in the carboxyl-terminal helical region, forms the tetramerization interface and does not have any role in dimerization. Mutation of the individual L residues of the leucine zipper of the lac repressor precluded tetramer formation and indicated a role for the motif in forming the dimer–tetramer interface (Chakerian *et al.*, 1991). In this context, it is worth noting that plasmid pPS10 of *Pseudomonas syringae patovar savastanoi* encodes a replication initiator protein containing an atypical leucine zipper. As in C protein, mutagenesis and biochemical studies showed that RepA contains an internal atypical leucine zipper and a typical HTH domain towards the C-terminal end of the protein involved in DNA binding (Giraldo *et al.*, 1989; Garcia de Viedma *et al.*, 1995). Recently, the crystal structure of RepA showed that the biochemically identified leucine repeats do not form a zipper motif for dimerization and instead form a 'leucine latch' that confers stability to a critical alpha helix of the dimer (Giraldo and Fernandez-Tresguerres, 2004). The elucidation of structure of C protein, eluded so far in spite of efforts from various groups, would conclusively resolve this further.

Table II. Prokaryotic leucine zipper like proteins

Protein	Origin	Size (kDa)	Motifs	Function	References
TodS	*Pseudomonus putida*	108	bZIP	Histidine kinase	Lau *et al.*, 1997
RepA	*Pseudomonas sp.*	26.7	HTH, leucine zipper	Replication initiator	Giraldo *et al.*, 1989
MetR	*Escherichia coli*	35	Leucine zipper	Transcription regulator	Maxon *et al.*, 1990
lac repressor	*Escherichia coli*	37.5	HTH, leucine zipper	Transcription regulator	Chakerian *et al.*, 1991
VirB	*Shigella flexneri*	35.4	HTH, leucine zipper	Transcription regulator	Beloin *et al.*, 2002
C protein	Bacteriophage Mu	16.5	HTH, leucine zipper	Transcription regulator	This work

C protein is thus a new member of emerging group of regulatory proteins that contain both HTH and zipper motifs. Although C appears to share the similar motif with the other members of zipper-HTH family of proteins, it does not exhibit any sequence similarities with them indicating its distant origin. The DNA-binding domain of C is unusual in that it has a striking similarity with the paired domain of *Drosophila* prd protein (Paul *et al.*, 2003). Thus, like the gene activated by C, *mom*, the activator also appears to be unique among the transactivators, having a unique combination of DNA-binding and dimerization domain. The activator functions by an unusual DNA unwinding activity to recruit RNAP to promoter and facilitation of promoter clearance at the second step of transactivation. Structure–function analysis of other transcription activators would probably identify more activators with this kind of unusual combination of motifs. It would be of interest to see whether there is any connection between the particular combination of dimerization and DNA-binding motifs and the mechanism of transcription activation.

Acknowledgements

References

Balke,V., Nagaraja,V., Gindlesperger,T. and Hattman,S. (1992) *Nucleic Acids Res.*, **20**, 2777–2784.

Basak,S. and Nagaraja,V. (1998) *J. Mol. Biol.*, **284**, 893–902.

Beloin,C., McKenna,S. and Dorman,C.J. (2002) *J. Biol. Chem.*, **277**, 15333–15344.

Blackwood,E.M. and Eisenman,R.N. (1991) *Science*, **251**, 1211–1217.

Chakerian,A.E., Tesmer,V.M., Manly,S.P., Brackett,J.K., Lynch,M.J., Hoh,J.T. and Matthews,K.S. (1991) *J. Biol. Chem.*, **266**, 1371–1374.

Chakraborty,A. and Nagaraja,V. (2006) *J. Biol. Chem.*, **281**, 8511–8517.

De,A., Paul,B.D., Ramesh,V. and Nagaraja,V. (1997) *Protein Eng.*, **10**, 935–941.

Ferre-D'Amare,A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K. (1993) *Nature, (London)*, **363**, 38–45.

Garcia de Viedma,D., Serrano-Lopez,A. and Díaz-Orejas,R. (1995) *Nucleic Acids Res.*, **23**, 5048–5054.

Giraldo,R., Nieto,C., Fernandez-Tresguerres,M.E. and Diaz,R. (1989) *Nature (London)*, **342**, 866.

Giraldo,R. and Fernandez-Tresguerres,M.E. (2004) *Plasmid*, **52**, 69–83.

Hattman,S. (1999) *Pharmacol. Ther.*, **84**, 367–388.

Johnson,P.F. and McKnight,L.S. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.

Jones,N. (1990) *Cell*, **61**, 9–11.

Kadesch,T. (1993) *Cell Growth Differ.*, **4**, 49–55.

Kahmann,R. and Hattman,S. (1987) In Symonds,N., Toussiant,A., van de Putte,P. and Howe,M.M. (ed.), *Phage Mu.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp.93–109

Kaptein,R., Zuiderweg,E.R.P., Scheek,R.M., Boelens,R. and van Gunsteren,W.F. (1985) *J. Mol. Biol.*, **182**, 179–182.

Kouzarides,T. and Ziff,E. (1988) *Nature*, **336**, 646–651.

Lau,C.K.P., Wang,Y., Patel,A., Labbe,D., Bergeron,H., Brousseau,R., Konishi,Y. and Rawlings,M. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 1453–1458.

Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.

Lupas,A., Van Dyke,M. and Stock,J. (1991) *Science*, **252**, 1162–1164.

Maxon,M.E., Wigboldus,J., Brot,N. and Weissbach,H. (1990) *Proc. Natl Acad. Sci., USA*, **87**, 7076–7079.

Murre,C. and Baltimore,D. (1992) In McKnight,S.L. and Yamamoto,Y.R. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p.861.

Paul,B.D., Kanhere,A., Chakraborty,A., Bansal,M. and Nagaraja,V. (2003) *Proteins*, **52**, 272–282.

Ramesh,V. and Nagaraja,V. (1996) *J. Mol. Biol.*, **260**, 22–33.

Ramesh,V., De,A. and Nagaraja,V. (1994) *Protein Eng.*, **7**, 1053–1057.

Sun,W. and Hattman,S. (1998) *J. Mol. Biol.*, **284**, 885–892.

van Heeckeren,W.J., Sellers,J.W. and Struhl,K. (1992) *Nucleic Acids Res.*, **20**, 3721–3724.

Weber,I.T., McKay,D.B. and Steitz,T.A. (1982) *Nucleic Acids Res.*, **10**, 5085–5102.