FEBS Letters 582 (2008) 1117–1122

Identification of operator sites within the upstream region of the putative *mce2R* gene from mycobacteria

Vaibhav Vindal, Enjapoori Ashwantha Kumar, Akash Ranjan*

Computational and Functional Genomics Group, Sun Centre of Excellence in Medical Bioinformatics, Centre for DNA Fingerprinting and Diagnostics, EMBnet India Node, Hyderabad 500 076, India

Received 29 January 2008; revised 26 February 2008; accepted 29 February 2008

Available online 7 March 2008

Edited by Ivan Sadowski

Abstract ***Mycobacterium tuberculosis*** **harbors four** ***mce*** **operons. Among them,** ***mce2*** **operon is preceded by a FadR-like regulator mce2R (Rv0586). Here, we report the operator sites of the mce2R and its orthologs in other sequenced mycobacteria and non-mycobacterial species** ***Nocardia farciana*****. All the identified DNA motifs illustrate the FadR subfamily specific nucleotide preference. Moreover, these motifs from the upstream region share sequence conservation, which is in agreement with the similarity of their DNA binding domain. Using electrophoretic mobility shift assay, we demonstrate that the predicted DNA motifs specifically interact with the recombinant Mce2R–Rv0586. Our present study has implications in the understanding of** ***cis*****-regulatory elements and the auto-regulatory nature of the FadR subfamily of regulators.**

Keywords: FadR; GntR; Operator site; Transcriptional regulator; Rv0586; *Mycobacterium tuberculosis*

1. Introduction

The biological activity of most of the genes in adaptive responses is regulated by a set of transcriptional regulators. These molecular players modulate cellular physiology at the level of transcription [1]. Apart from regulating other genes or operons, many transcriptional regulators interact with an upstream region to regulate their own expression. To decipher such *cis*-elements, comparative genomics has emerged as a major approach to explore the upstream sequences of Open Reading Frames (ORFs) from multiple genome sequences [2]. In the last 10 years, genomes of many mycobacterial species, including *Mycobacterium tuberculosis*, have been sequenced. The complete genome sequence of this bacillus is annotated with a large number of putative transcriptional regulators indicating that much of the gene regulation at the level of transcription is yet to be understood [3].

The *M. tuberculosis* genome is reported to contain four copies of *mce* operons (*mce1*, *mce2*, *mce3* and *mce4*) [3,4]. In vivo mutational studies of these operons have shown their importance in the virulence [5]. These operons share similarity in sequence and organization. Orthologs of these operons are reported in many mycobacterial species [6]. Out of five FadR-like transcriptional regulators from *M. tuberculosis*, the regulators Rv0165c and Rv0586 are associated with *mce1* and *mce2* operon, respectively [7,8]. Typically, members of this subfamily of transcriptional regulators are known to be auto-regulatory [9]. Hence in addition to regulating expression of a number of genes and operons these regulators also regulate their own expression. Identification of the operator sites is the primary step to understand the *cis–trans* relationship of transcriptional regulators. Conventionally, these DNA binding sites were determined by labor-intensive DNAase1 footprinting. However, many computational approaches have been developed to address these issues [10–12]. Understanding the regulatory elements from the same family of transcriptional regulators is one of the effective strategies [13,14]. Our earlier study classifies ORF Rv0586 to the FadR subfamily of transcriptional regulators [7]. Operator sites for many transcriptional regulators of this subfamily are characterized [15]. Based on known operator sites, members of this family of regulators were suggested to exhibit nucleotide preferences in their operator sites recognition. This study envisaged the nucleotide preferences in the upstream sequence of the putative *mce2R* gene along with the sequence conservation among the mycobacterial species [9,15]. Further, using recombinant Rv0586 protein, predicted DNA motifs were experimentally investigated. This computational analysis, coupled with the in vitro information of potential operator site in mycobacterial species, provides important leads to further analyze Rv0586 and its putative ortholog, Mce2R, in mycobacteria.

2. Materials and methods

2.1. Ortholog prediction and multiple sequence alignment

We used the reciprocal best BLAST hit method to predict orthologous proteins between each of the two proteomes by the BLASTP program with an *E*-value cutoff 10^{-6} for both directions [16,17]. Protein sequences of Rv0586 and its identified orthologs across the mycobacteria and *Nocardia farciana* were aligned using ClustalX [18]. Secondary structures of all protein sequences were studied using 3DPSSM, Jpred and SsPro [19–21]. All the published and annotated bacterial genome sequences for the study were downloaded from NCBI ftp site (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/).

2.2. Upstream sequence analysis

Upstream DNA sequences 400 bases upstream and 50 bases downstream of the translation start site of ORF Rv0586 and its putative orthologs were aligned to locate the palindromic conserved DNA block satisfying this preference. A further consensus sequence logo

*Corresponding author. Fax: +91 40 27155610.
E-mail addresses: vaibhav@cdfd.org.in (V. Vindal), ashwanth@cdfd.org.in (E. Ashwantha Kumar), akash@cdfd.org.in (A. Ranjan).

0014-5793/$34.00
doi:10.1016/j.febslet.2008.02.074

was drawn, using web logo, to show the relative frequency of each base at each position of the motif [22].

2.3. Cloning, expression and purification of *M. tuberculosis* Rv0586

ORF Rv0586 was amplified by PCR using a forward primer (CGCGGATCCATGGCGCTGCAGCCGGTGACTCG) with a BamH1 site and reverse primer (CCCAAGCTTTCATTGCCGACTCGCCTGGCTAAC) with a HindIII site. This PCR fragment (723 bp) cloned into pQE30 expression vector (Qiagen) with an N-terminal 6× His tag. Recombinant clones were identified and checked with restriction digestion and DNA sequencing. The recombinant vector was transformed into *Escherichia coli* M15 cells and a transformed single colony was inoculated in 5 ml of LB media containing 100 μg/ml of ampicillin and 25 μg/ml of kanamycin (starter culture). This starter culture was grown overnight at 37 °C with vigorous shaking and 2 ml of the above culture was inoculated into 200 ml LB medium containing appropriate antibiotics. The culture was grown at 37 °C until the OD reached 0.6 at $A_{600\,nm}$. A control culture was maintained in parallel. The cells were kept in an incubator shaker for another 8 h at 27 °C and 200 rpm, to allow protein expression at 0.5 mM IPTG concentration. Then, cells were harvested by centrifugation and resuspended in 10 ml of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole, pH 8.0) with 1 mM PMSF and disrupted using a sonicator. After a second round of centrifugation for 20 min at 12000 rpm, the supernatant was collected and applied to an Ni–NTA affinity column (Qiagen, USA). The recombinant protein was eluted with 200 mM imidazole and analyzed by SDS–PAGE, after washing the column with five bed-volumes of wash buffer containing 20 mM imidazole.

2.4. Electrophoretic mobility shift assay

Electrophoretic mobility shift assays were carried out using recombinant Rv0586 protein to show binding with the predicted operator site. An increasing amount of Rv0586 was incubated with 10 fmol of ^{32}P-labeled DNA motif at room temperature for 40 min (10 mM Tris–HCl [pH 8.0], 1 mM DTT, 10 mM NaCl, 1 mM EDTA, 10% glycerol, 10 μg of poly(dI–dC)/ml and 5 μg of bovine serum albumin per ml); and loaded onto 5% non-denaturing polyacrylamide gels containing 0.5× Tris–borate–EDTA buffer. Samples were separated by electrophoresis at 200 V for 2 h. Subsequently, the gel was dried and exposed to a storage phosphor image plate. The image plate was scanned in a storage phosphor imaging workstation.

3. Results and discussion

3.1. Conservation of Rv0586–Mce2R across the mycobacteria

Since ortholog prediction is an important application of comparative genomics, to help in functional annotation of a sequenced genome, we began our study by identifying putative orthologs of Rv0586 in the genome of other sequenced mycobacteria and closely related non-mycobacterial species like *N. farciana*. We observed that most of the mycobacterial genomes have an ortholog of this transcriptional regulator except *Mycobacterium leprae*, *Mycobacterium gilvum PYR-GCK* and *Mycobacterium ulcerans Agy99* (Table 1). The presence of this gene in genomes of multiple mycobacterial species highlights the importance of this regulator in mycobacterial physiology.

3.2. Rv0586-orthologs show a similar DNA binding domain

In order to assess conservation in the DNA binding domain of Mce2R, multiple sequence alignment of protein sequences were carried out. The N-terminal region, known as the DNA binding domain, was found conserved in comparison to the C-terminal ligand-binding domain across all the *mce2R* genes (Fig. 1). In general, conservation of the DNA binding domain in these closely related species constrains the evolution of the regulatory DNA motifs, relative to the neighboring DNA sequences [2]. This suggests that Rv0586 could recognize target DNA motifs specific to the putative Mce2R in closely related species [23,24]. Additionally, these orthologs showed conservation in the pattern of secondary structural elements known for FadR-like transcriptional regulators (Fig. 1) [9].

3.3. Upstream region of *mce2R* possess novel conserved operator site

Many FadR-like transcriptional regulators are reported to recognize DNA palindromes showing nucleotide preferences exhibited by members of FadR subfamily [9]. We analyzed DNA sequences for these preferences, in addition to conservation among the orthologous upstream region. DNA motifs satisfying the condition set for identification at maximum were listed (Table 1). Additionally, all upstream *mce2R* regions were aligned to find conservation in their operator sites in case they are recognized by a similar DNA binding domain. This operator site conservation was revealed from alignment of all the upstream DNA sequence from the translational start site (Fig. 2A and Table 1) [25]. Amongst all the identified operator sites DNA motifs for Rv0586 and Mb0601; Mkms_2771 and Mjls_2751 were found to be identical in sequence (Table 1). This sequence comparison was quite convincing as we expected to find conservation in the operator sites. A consensus sequence logo was produced, using web logo, to show the relative frequency of each base at each position of the motif (Fig. 2B). The figure shows that positions 4–9 in the consensus are nearly an inverted palindrome of positions 13–18 (Fig. 2B).

3.4. Mce2R binds to the identified operator site across the mycobacteria

To examine binding of Rv0586 to the predicted operator site, the protein was expressed as His-tagged recombinant protein in *E. coli* (Fig. 3A). This protein was subjected to electrophoretic mobility shift assays (EMSA), to validate the identified operator site in the upstream region of ORF Rv0586. Purified protein showed clear binding with increasing concentration to the synthesized double stranded DNA motif (Table 1). This binding was abolished with increasing concentration of unlabeled DNA as a specific competitor (10×, 25× and 50× molar excess), whereas similar excess of non-specific

Table 1
List of identified operator sites in the upstream region of putative orthologs of Rv0586

ORF	Organism	Potential operator site
Rv0586	*M. tuberculosis*	GGTGTCGGTCTGACCACTTGA
Mb0601	*M. bovis*	GGTGTCGGTCTGACCACTTGA
MAP4081	*M. avium*	GCCGGTGGTCTGACCACCTGA
Mkms_2771	*M. KMS*	GCTAACTGGTCAGACCACTTGAC
Mjls_2757	*M. JLS*	GCTAACTGGTCAGACCACTTGAC
MSMEG_3527	*M. smegmatis*	ACCACTGGTAAGACCACTTGA
Mvan_2942	*M. vanbaalenii PYR*	CACACTGGTCTGACCACTTGA
Nfa1630	*N. farciana*	ACGATTGGTCTTACCACTTGA

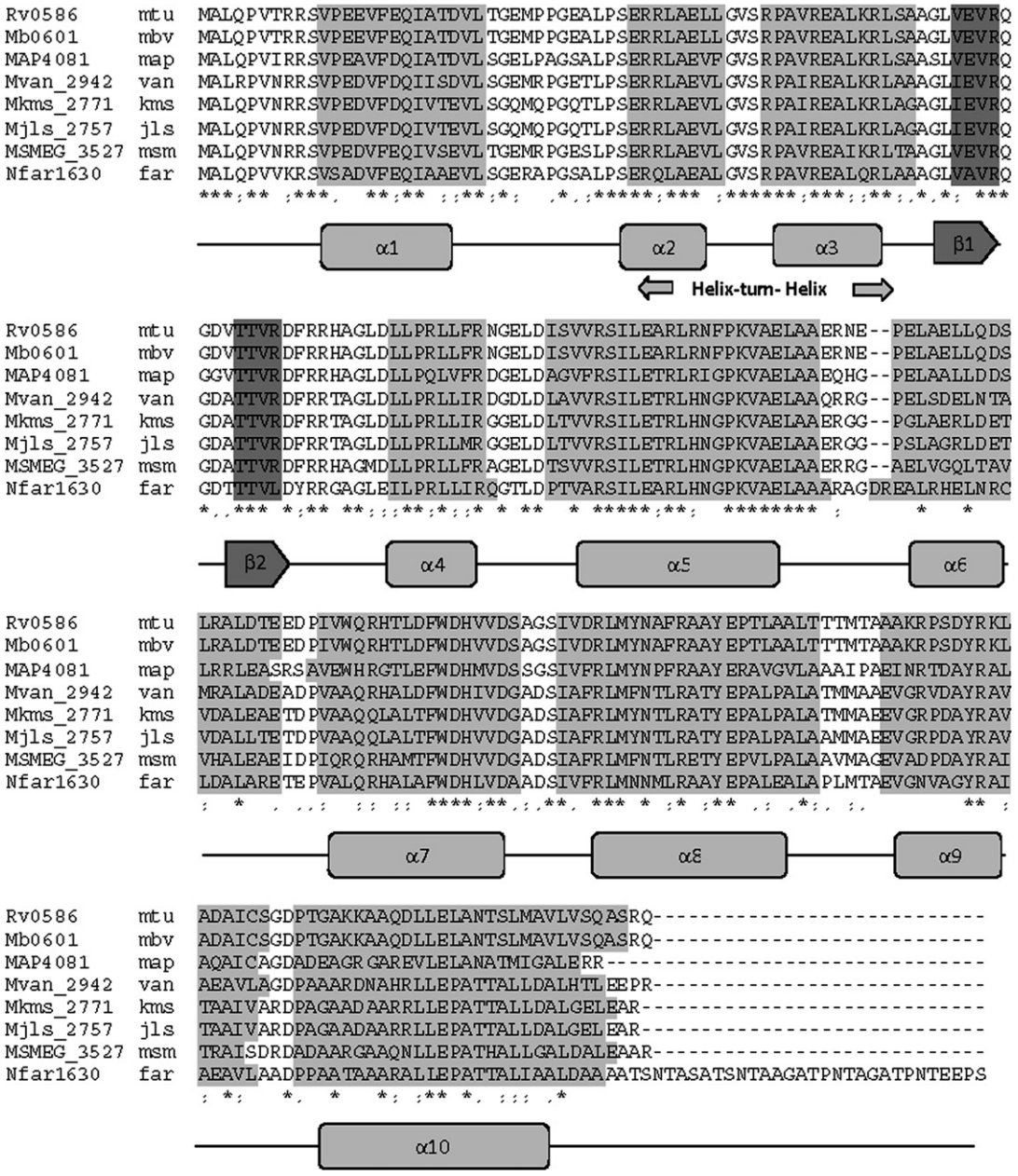

Fig. 1. Multiple sequence alignment of putative Mce2R from mycbacteria and *N. farciana*. In graphical representation α-helix region and β-sheet regions are highlighted with light and dark gray background, respectively. Helix-turn-helix region (α-2 and α-3) in the alignment is shown with an arrow mark (*abbreviations:* mtu – *M. tuberculosis*; mbv – *M. bovis*; map – *M. avium subsp. paratuberculosis*; van – *M. vanbaalenii PYR*; msm – *M. smegmatis*; kms – *M. sp KMS*; jls – *M. JLS*; far – *N. farciana*).

competitor DNA did not affect the DNA–protein complex (Fig. 3B). This non-specific DNA was also chosen from the upstream region of the ORF Rv0586 (CAACTTAGCCCGATAACTGCG). Additionally, Rv0586 protein was analyzed for DNA interaction with the predicted operator sites in the upstream region of the orthologs (Table 1). Using EMSA, we observe binding to these predicted DNA motifs in presence of increasing Rv0586 protein concentration. Binding was shown to be specific because the complex was not abolished by non-specific competitor DNA while specific unlabeled DNA successfully abolished the DNA–protein complex at 50-fold excess (Fig. 4A–E). These observations suggested that Rv0586 binds specifically to its upstream region to a conserved motif across the related species.

Earlier studies have demonstrated the critical role of the *mce2* operon, where animals infected with a strain bearing mutant *mce2* operon caused delay in granuloma formation [5]. This study identifies a set of DNA sequences that are likely to serve as operator sites for the Mce2R regulator in its upstream region, suggesting a typical auto-regulatory mechanism of FadRs. Identified DNA motifs display consistent nucleotide preferences known for many transcriptional regulators belonging to FadR subfamily. All the identified orthologs share similarity in the DNA binding domain of the proteins. This feature was also depicted in the sequence conservation of the predicted operator sites. Using recombinant Rv0586, specific DNA–protein interaction with the set of identified DNA motifs were experimentally verified. It is worth mentioning that

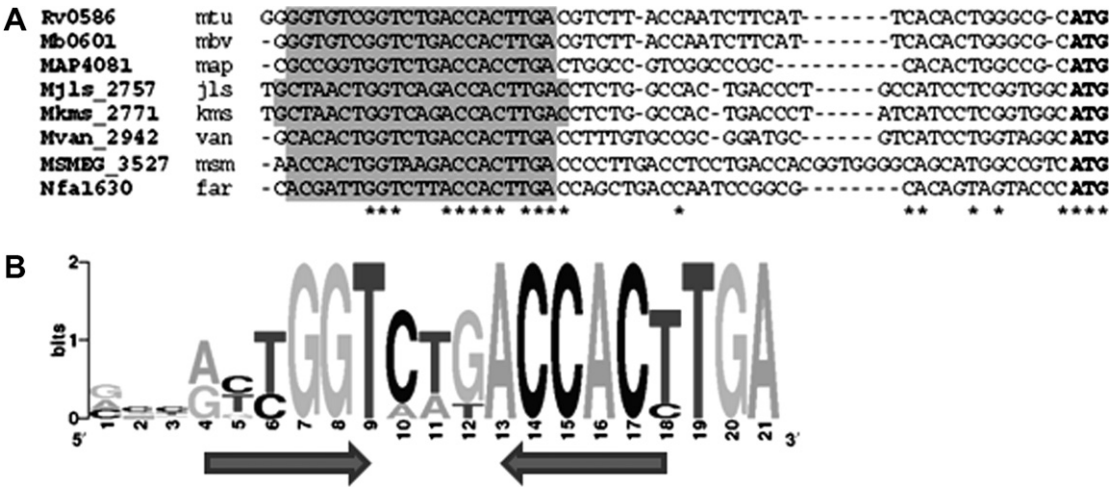

Fig. 2. Diagram showing identified operator sites in the upstream sequences of putative the *mce2R* gene in relation to the translations start sites. (A) Upstream sequence alignment of putative the *mce2R* ORF showing the position of identified DNA operator site in light gray background. Translational start sites in all the sequence are printed in bold. (B) A consensus logo was drawn using identified operator (21 bp) (abbreviations used are same as mentioned in Fig. 1).

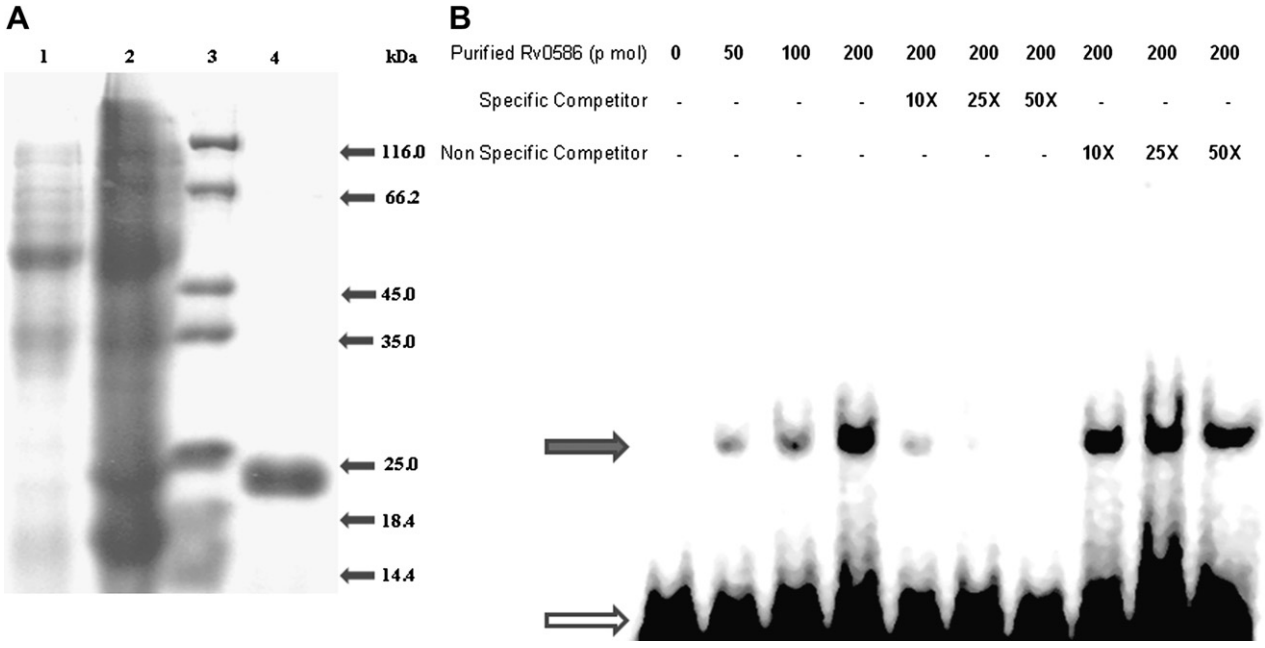

Fig. 3. Binding of the Rv0586 protein to the operator DNA from the upstream region of Rv0586 (or Mb0601). (A) Expression and purification of Rv0586. Lane 1, IPTG induced *E. coli* M15 cell lysate harboring pQE30 as a control; lane 2, IPTG induced *E. coli* M15 cell lysate harboring recombinant pQE30 vector cloned with ORF Rv0586; lane 3, protein marker; lane 4, Rv0586 purified protein. All samples were loaded on 12% SDS–PAGE followed by Coomassie blue staining. (B) Lane 1, labeled fragment; lanes 2–4, labeled fragment with 50, 100, 200 pmol of purified His_6–Rv0586 protein; lanes 5–7 contain an increasing amount of cold specific dsDNA oligonucleotide competitor (10-, 25- and 50-fold molar excess); lanes 8–10 contain an increasing amount of cold non-specific competitor (10-, 25- and 50-fold molar excess). The positions of DNA–protein complex and free probe are shown with solid and open arrows, respectively.

such interactions strengthen the idea of sharing nucleotide preferences among the transcriptional regulators belonging to the same family. It also reveals the influence of the conservation of DNA binding domain upon the specificity of the DNA targets. This study will have potential implications to understand the regulation of the *mce2* operon, as well its associated genes playing important role in host–pathogen interactions.

Acknowledgements: V.V. is supported by the UGC fellowship. E.A.K. is supported by the CSIR NMITLI Grant and Department of Biotechnology Grant to A.R. Research in A.R.'s laboratory is supported by the Grants from Department of Biotechnology, Department of Science and Technology, Indian Council of Medical Research (ICMR), Council of Scientific & Industrial Research (CSIR) NMITLI. Authors acknowledge the assistance of Ms. Rohini Kondam and Mr. Gopinathan Gokul, SRF for proof reading the manuscript. Authors thank the editor, Dr. Ivan Sadowski, for his valuable suggestions.

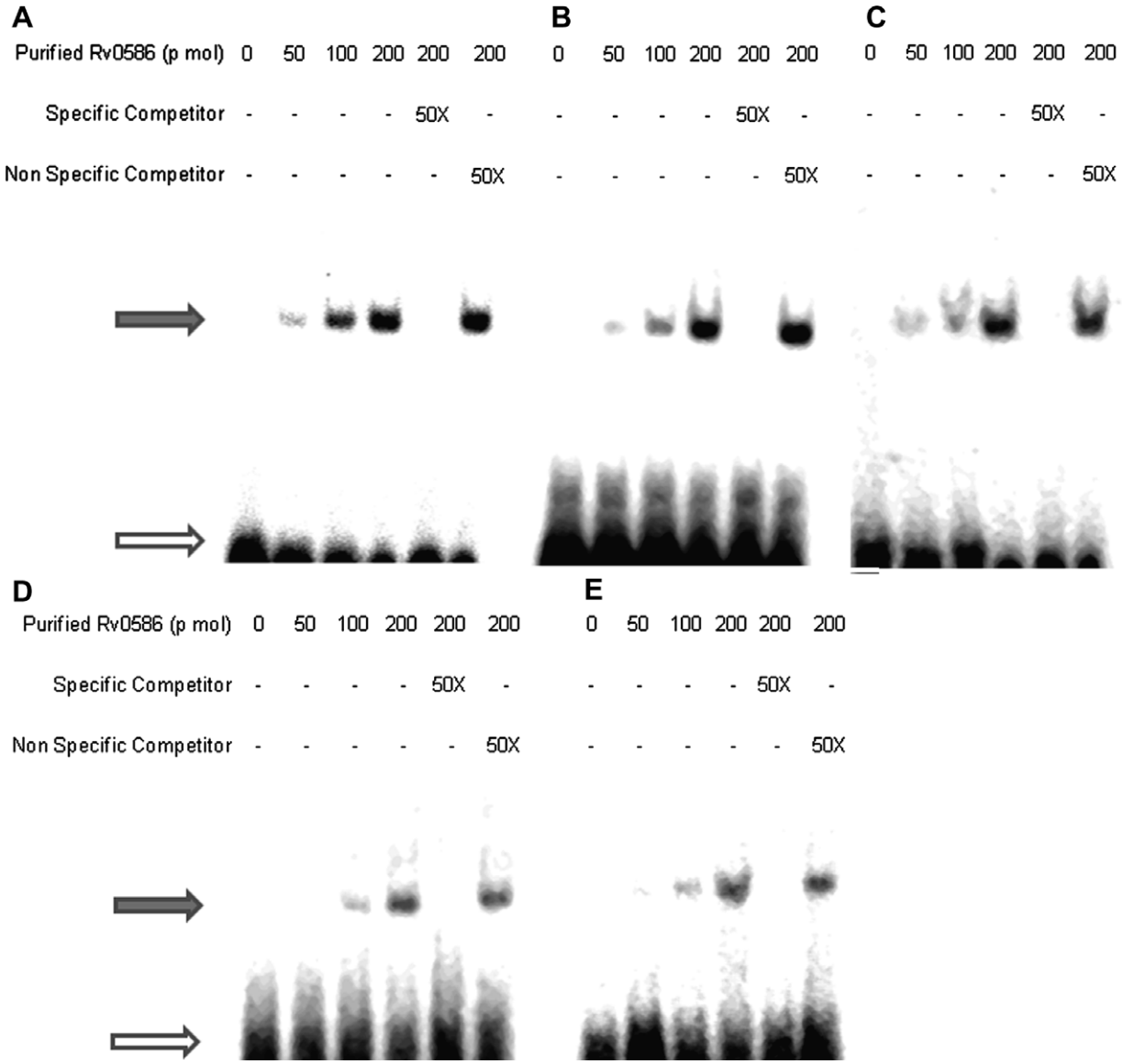

Fig. 4. EMSA for the predicted operator sites for other mycobacterial species. Lane 1, labeled fragment; lanes 2–4, labeled fragment with 50, 100, 200 pmol of purified His_6–Rv0586 protein; lane 5 contains 50-fold molar excess cold specific dsDNA competitor; lane 6 contains the same concentration of cold non-specific competitor. (A) MSMEG_3527 (*M. smegmatis MC2 155*); (B) MAP4081 (*M. avium paratuberculosis*); (C) Mvan_2942 (*M. vanbaalenii PYR*); (D) Mkms_2771/Mjls_2757 (*M. KMS/M. JLS*); (E) Nfa1630 (*N. farciana*).

References

[1] Seshasayee, A.S., Bertone, P., Fraser, G.M. and Luscombe, N.M. (2006) Transcriptional regulatory networks in bacteria: from input signals to output responses. Curr. Opin. Microbiol. 9, 511–519.

[2] Bailey, T.L., Williams, N., Misleh, C. and Li, W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 34, W369–W373.

[3] Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd., C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537–544.

[4] Camus, J.C., Pryor, M.J., Médigue, C. and Cole, S.T. (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. Microbiology 148, 2967–2973.

[5] Gioffré, A., Infante, E., Aguilar, D., Santangelo, M.P., Klepp, L., Amadio, A., Meikle, V., Etchechoury, I., Romano, M.I., Cataldi, A., Hernández, R.P. and Bigi, F. (2005) Mutation in *mce* operons attenuates *Mycobacterium tuberculosis* virulence. Microb. Infect. 7, 325–334.

[6] Casali, N. and Riley, L.W. (2007) A phylogenomic analysis of the Actinomycetales *mce* operons. BMC Genom. 8, 60.

[7] Vindal, V., Ranjan, S. and Ranjan, A. (2007) In silico analysis and characterization of GntR family of regulators from *Mycobacterium tuberculosis*. Tuberculosis 87, 242–247.

[8] Casali, N., White, A.M. and Riley, L.W. (2006) Regulation of the *Mycobacterium tuberculosis mce1* operon. J. Bacteriol. 188, 441–449.

[9] Vindal, V., Suma, K. and Ranjan, A. (2007) GntR family of regulators in *Mycobacterium smegmatis*: sequence and structure based characterization. BMC Genom. 8, 289.

[10] Ranjan, S., Seshadri, J., Vindal, V., Yellaboina, S. and Ranjan, A. (2006) iCR: a web tool to identify conserved targets of a regulatory protein across the multiple related prokaryotic species. Nucleic Acids Res. 34, W584–W587.
[11] Yellaboina, S., Seshadri, J., Kumar, M.S. and Ranjan, A. (2004) PredictRegulon: a web server for the prediction of the regulatory protein binding sites and operons in prokaryote genomes. Nucleic Acids Res. 32, W318–W320.
[12] Ranjan, S., Gundu, R.K. and Ranjan, A. (2006) MycoperonDB: a database of computationally identified operons and transcriptional units in mycobacteria. BMC Bioinform. 7, S9.
[13] Kaplan, T., Friedman, N. and Margalit, H. (2005) Ab initio prediction of transcription factor targets using structural knowledge. PLoS Comput. Biol. 1, e1.
[14] Wu, C.H., Huang, H., Yeh, L.S. and Barker, W.C. (2003) Protein family classification and functional annotation. Comput. Biol. Chem. 27, 37–47.
[15] Rigali, S., Derouaux, A., Giannotta, F. and Dusart, J. (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J. Biol. Chem. 277, 12507–12515.
[16] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
[17] Fulton, D.L., Li, Y.Y., Laird, M.R., Horsman, B.G., Roche, F.M. and Brinkman, F.S. (2006) Improving the specificity of high-throughput ortholog prediction. BMC Bioinform. 7, 270.
[18] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
[19] Kelley, L.A., MacCallum, R.M. and Sternberg, M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. 299, 499–520.
[20] Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) JPred: a consensus secondary structure prediction server. Bioinformatics 14, 892–893.
[21] Cheng, J., Randall, A.Z., Sweredoski, M.J. and Baldi, P. (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res. 33, W72–W76.
[22] Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.
[23] Roy, S., Sahu, A. and Adhya, S. (2002) Evolution of DNA binding motifs and operators. Gene 285, 169–173.
[24] Yellaboina, S., Ranjan, S., Vindal, V. and Ranjan, A. (2006) Comparative analysis of iron regulated genes in mycobacteria. FEBS Lett. 580, 2567–2576.
[25] McCue, L.A., Thompson, W., Carmack, C.S. and Lawrence, C.E. (2002) Factors influencing the identification of transcription factor binding sites by cross-species comparison. Genome Res. 12, 1523–1532.