Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

Biochimica et Biophysica Acta 1789 (2009) 460–468

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in *Pichia pastoris*

Balla Venkata Kranthi, Ritesh Kumar, Nallani Vijay Kumar, Desirazu N. Rao, Pundi N. Rangarajan *

Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India

ARTICLE INFO

Article history:
Received 27 November 2008
Received in revised form 1 May 2009
Accepted 12 May 2009
Available online 18 May 2009

Keywords:
Alcohol oxidase
Pichia pastoris
Zinc finger protein
Methanol
Transcription
DNase I foot printing

ABSTRACT

Mxr1p (methanol expression regulator 1) functions as a key regulator of methanol metabolism in the methylotrophic yeast *Pichia pastoris*. In this study, a recombinant Mxr1p protein containing the N-terminal zinc finger DNA binding domain was overexpressed and purified from *E. coli* cells and its ability to bind to promoter sequences of *AOXI* encoding alcohol oxidase was examined. In the *AOX1* promoter, Mxr1p binds at six different regions. Deletions encompassing these regions result in a significant decrease in *AOXI* promoter activity *in vivo*. Based on the analysis of *AOXI* promoter sequences, a consensus sequence for Mxr1p binding consisting of a core 5′ CYCC 3′ motif was identified. When the core CYCC sequence is mutated to CYCA, CYCT or CYCM (M = 5-methylcytosine), Mxr1p binding is abolished. Though Mxr1p is the homologue of *Saccharomyces cerevisiae* Adr1p transcription factor, it does not bind to Adr1p binding site of *S. cerevisiae* alcohol dehydrogenase promoter (*ADH2UAS1*). However, two point mutations convert *ADH2UAS1* into an Mxr1p binding site. The identification of key DNA elements involved in promoter recognition by Mxr1p is an important step in understanding its function as a master regulator of the methanol utilization pathway in *P. pastoris*.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The methylotrophic yeast *Pichia pastoris* is an excellent model organism for the study of eukaryotic gene expression and peroxisome biogenesis as well as production of recombinant proteins [1–4]. In *P. pastoris*, expression of a number of genes involved in methanol metabolism as well as peroxisome biogenesis is induced when methanol or fatty acids are used as the sole carbon source [2,3]. Among the methanol-inducible genes, the gene encoding alcohol oxidase I (*AOXI*), the first enzyme in methanol metabolism is expressed at very high levels and as a result *P. pastoris AOX1* promoter (P_{AOX1}) has become one of the most extensively used eukaryotic promoters for recombinant protein expression [1,2]. Despite its industrial importance, the mechanism by which methanol induces the transcription of genes involved in methanol utilization pathway is not well understood. Recently, a zinc finger transcription factor named Mxr1p (methanol expression regulator 1) was identified as the global regulator of genes involved in the methanol utilization pathway [5]. *P. pastoris* strains lacking Mxr1p are unable to grow on peroxisomal substrates such as methanol and oleic acid and expression of *AOXI* as well as several genes involved in the methanol utilization pathway is severely affected in the *mxr1*Δ strain [5]. Methanol-inducible expression of *AOXI* as well as several other genes of the methanol utilization pathway is severely affected in *mxr1*Δ strain [5]. A number of promoter sequences involved in methanol-inducible expression of P_{AOX1} have been recently identified and some of these harbour potential Mxr1p binding sites [6]. However, transcription factors binding to these sequences have not been biochemically characterized.

Mxr1p consists of a C_2H_2-type zinc finger in the amino terminus which shares 70% identity with the zinc finger DNA binding domain of *Saccharomyces cerevisiae* Adr1p (alcohol dehydrogenase II synthesis regulator), a transcription factor regulating the expression of genes involved in the metabolism of glycerol, ethanol and oleic acid [5]. In view of the high degree of homology between the DNA binding domains of Mxr1p and Adr1p, it was suggested that the Mxr1p and Adr1p may bind similar DNA sequences [5]. While the Adr1p binding sites are well characterized, the DNA sequence requirements for Mxr1p binding are not known. Electrophoretic mobility shift assays (EMSA) carried out with *P. pastoris* cell extracts overexpressing Mxr1p and a 243 bp P_{AOX1} DNA indicated the presence of an Mxr1p binding site between −415 and −172 bp. This region contains a sequence similar to that recognized by Adr1p suggesting that Mxr1p and Adr1p may recognize similar DNA sequences [5]. However, *in silico* analysis of P_{AOX1} for transcription factor binding sites revealed the presence of a putative Adr1p binding site between −577 and −571 bp [6]. Whether Mxr1p actually binds to any of these putative Adr1p binding

Abbreviations: Mxr1p, methanol expression regulator 1; AOXI, alcohol oxidase I; P_{AOX}, promoter of *AOXI*; DHAS, dihydroxyacetone synthase; PEX8, peroxin 8; MXRE, MXR response element; EMSA, electrophoretic mobility shift assay; Adr1p, alcohol dehydrogenase II synthesis regulator; HP0051, *Helicobacter pylori* C^5 cytosine methyltransferase

* Corresponding author. Tel./fax: +91 80 23601492.
E-mail addresses: pnr@biochem.iisc.ernet.in, pnrangarajan@yahoo.com (P.N. Rangarajan).

1874-9399/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagrm.2009.05.004

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

sites in P_{AOX1} is not known. In this study, the N-terminal region of Mxr1p containing the DNA binding domain was overexpressed in *E. coli* cells and the ability of the purified recombinant protein to bind to *AOXI* promoter sequences (−940 to −1 bp) was evaluated by EMSA as well as DNase I foot printing. Such studies have led to the identification of six Mxr1p response elements in P_{AOX1} (P_{AOX1}-MXREs), each of which is of 20 bp in length and consists of a core 5′ CYCC/GGRG 3′ sequence. Interestingly, Mxr1p does not recognize the Adr1p binding sites present in the *S. cerevisiae* of alcohol dehydrogenase promoter upstream activation sequence 1 (*ADH2UAS1*), unless two adenine residues are mutated to cytosine.

2. Materials and methods

2.1. Synthesis of oligonucleotides and preparation of radiolabeled DNA probes

Oligonucleotides were purchased from Sigma-Aldrich Chemicals (Bangalore, India). DNA containing Adr1p binding site present in *S. cerevisiae ADH2* promoter from −237 to −215 bp [7–9] was generated by annealing the oligonucleotide 5′ GGGTCTCCAACTTA-TAAGTTGGAGACCC 3′ with its complement. ^{32}P-labeled DNA probes were generated by radiolabeling the 5′ ends of double stranded oligonucleotides with [γ-^{32}P]ATP and T4 polynucleotide kinase using standard molecular biology protocols [10]. After radiolabeling, unincorporated radionucleotides were removed by sephadex G-50 chromatography.

2.2. Preparation of recombinant Mxr1p

The cDNA encoding the first 150 N-terminal amino acids of Mxr1p was obtained by PCR amplification of *P. pastoris* genomic DNA using the primer pair 5′ CGCGGATCCATGAGCAATCTACCCCCAAC 3′ and 5′ CCCAAGCTTTCAATT TGAGTCCCGGCGGCC 3′ and cloned into BamHI and HindIII sites of pRSETA vector (Invitrogen, USA) using standard molecular biology protocols [10]. The restriction sites in the primers are underlined. The recombinant plasmid was transformed into *E. coli* BL21(DE3)pLysS cells and grown at 37 °C until an A_{600} of 0.8 was reached. Expression of the recombinant protein was induced for 3 h by the addition of 1.0 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Cells were harvested and resuspended in buffer A (50 mM Tris–HCl, pH 8.0, 250 mM NaCl, 10% glycerol and 10 mM β-mercaptoethanol) containing 10 mM imidazole. Cells were lysed by sonication and the lysate was centrifuged at 10,000 g at 4 °C for 30 min. The clear supernatant was loaded onto a Ni^{2+}-NTA agarose column (Qiagen, USA) equilibrated with buffer A. The column was washed with buffer A containing 50 mM imidazole and the recombinant protein was eluted sequentially with buffer A containing 100 mM and 300 mM imidazole. The 300 mM eluate fractions containing histidine-tagged recombinant Mxr1p protein were pooled, dialysed against buffer A and stored in aliquots at −80 °C. Western blot analysis of the purified protein was carried out with anti-histidine tag antibodies (Bangalore Genei, Bangalore, India). The recombinant protein was subjected to trypsin digestion and the tryptic peptides were purified by reverse phase HPLC. One of the purified peptides was sequenced using Procise™ protein sequencing system (Perkin–Elmer, Applied Biosystems, Foster, CA).

2.3. Electrophoretic mobility shift assay (EMSA)

Purified recombinant Mxr1p (30 nM/10 ng) was incubated with ^{32}P-labeled DNA probes (20,000 cpm, ~0.2–0.5 nM) in a 30 μl reaction containing buffer B (50 mM Tris–HCl, pH 8.0, 50 mM NaCl, 1 mM DTT, 0.05% NP-40, 100 ng poly (dI-dC) and 6% glycerol) for 30 min at 4 °C. In some experiments, Mxr1p (~25–60 nM) was incubated first with excess amount of unlabeled oligonuceotides for 5 min prior to the addition of ^{32}P-labeled DNA probes. The reaction mixtures were subjected to electrophoresis at 4 °C on 5% non-denaturing polyacrylamide gel in a buffer containing 7 mM Tris–HCl (pH 7.4), 3 mM boric acid and 1 mM EDTA. The gels were dried and autoradiographed.

2.4. Measurement of the apparent dissociation constant

The apparent dissociation constant (K_d) of Mxr1p binding to *AOXI* promoter sequences was estimated as the protein concentration at which half of the radiolabeled DNA probe was shifted in an EMSA. EMSA was carried out with radiolabeled P_{AOX}*MXRE*s (~0.5 nM) and varying amounts of Mxr1p in the presence of 100 ng of poly(dI-dC) per 30 μl reaction as described above. The intensity of shifted bands and free probes was measured in a PhosphorImager (Molecular Dynamics). All quantitations were done using ImageQuant software, version 5.0 (Molecular Dynamics). Experiments were repeated twice and the mean ± standard deviation was calculated. The K_d values were also calculated by non-linear regression using GraphPad Prism 5

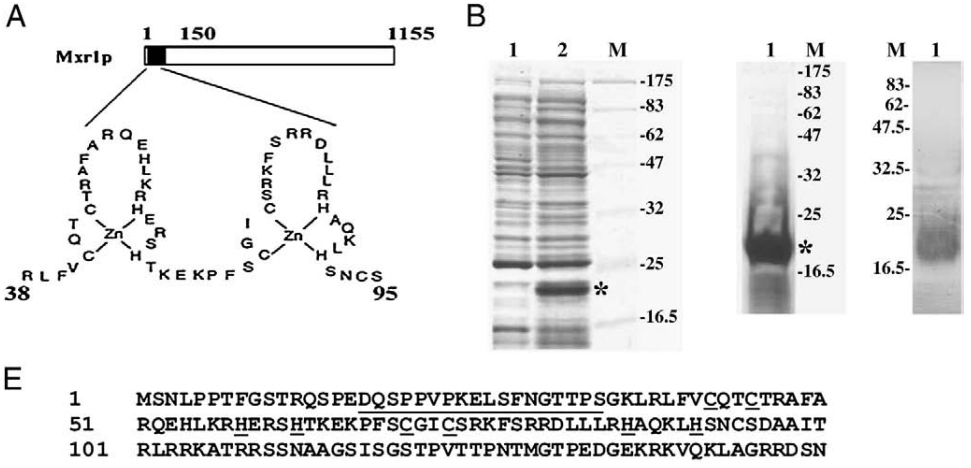


Fig. 1. Overexpression and purification of recombinant Mxr1p containing the first 150 N-terminal amino acids. (A) Schematic representation of Mxr1p overexpressed and purified from *E. coli* cells. The C_2H_2 zinc finger DNA binding domain located between amino acids 38 and 95 is indicated. (B) Coomassie blue-stained SDS gel pattern of lysates of recombinant *E. coli* cells cultured in the absence (lane 1) or presence (lane 2) of IPTG. (C) Mxr1p purified by Ni^{2+} affinity chromatography. Asterisk indicates Mxr1p. (D) Western blot analysis of recombinant Mxr1p with anti-Histidine tag antibodies. (E) Amino acid sequence of recombinant Mxr1p used in this study. The region homologous to Adr1p PAR [11–13] is underlined. The cysteine and histidine residues of Mxr1p zinc fingers which co-ordinate with the zinc atoms are also underlined.

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

Table 1
Oligonucleotides spanning the *P. pastoris AOX1* promoter.

Name	*AOX1* promoter region (bp)	Oligonucleotide sequence
940	−940 to −881	5′ AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCG**ACATCCACAG** 3′
940c		3′ TCTAGATTGTAGGTTTCTGCTTTCCAACTTACTTTGGAAAAACGGTAGGC**TGTAGGTGTC** 5′
890	−890 to −831	5′ **ACATCCACAG**GTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGGGG**ATACACTAGC** 3′
890c		3′ **TGTAGGTGTC**CAGGTAAGAGTGTGTATTCACGGTTTGCGTTGTCCTCCCC**TATGTGATCG**
840	−840 to −781	5′ **ATACACTAGC**AGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTTCTC**CTCAACACCC** 3′
840c		3′ **TATGTGATCG**TCGTCTGGCAACGTTTGCGTCCTGGAGGTGAGGAGAAGAG**GAGTTGTGGG**
790	−790 to −731	5′ **CTCAACACCC**ACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCTTGA**TTGGAGCTCG** 3′
790c		3′ **GAGTTGTGGG**TGAAAACGGTAGCTTTTTGGTCGGGTCAATAACCCGAACT**AACCTCGAGC**
740	−740 to −681	5′ **TTGGAGCTCG**CTCATTCCAATTCCTTCTATTAGGCTACTAACACCATGAC**TTTATTAGCC** 3′
740c		3′ **AACCTCGAGC**GAGTAAGGTTAAGGAAGATAATCCGATGATTGTGGTACTG**AAATAATCGG** 5′
690	−690 to −631	5′ **TTTATTAGCC**TGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGTTTA**TTTCCGAATG** 3′
690c		3′ **AAATAATCGG**ACAGATAGGACCGGGGGGACCGCTCCAAGTACAAACAAATA**AAAGGCTTAC**
640	−640 to −581	5′ **TTTCCGAATG**CAACAAGCTCCGCATTACACCCGAACATCACTCCAGATGA**GGGCTTTCTG** 3′
640c		3′ **AAAGGCTTAC**GTTGTTCGAGGCGTAATGTGGGCTTGTAGTGAGGTCTACT**CCCGAAAGAC** 5′
590	−590 to −531	5′ **GGGCTTTCTG**AGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGGCCCA**AAACTGACAG** 3′
590c		3′ **CCCGAAAGAC**TCACACCCCAGTTTATCAAAGTACAAGGGGTTTACCGGGT**TTTGACTGTC** 5′
540	−540 to −481	5′ **AAACTGACAG**TTTAAACGCTGTCTTGGAACCTAATATGACAAAAGCGTGA**TCTCATCCAA** 3′
540c		3′ **TTTGACTGTC**AAATTTGCGACAGAACCTTGGATTATACTGTTTTCGCACT**AGAGTAGGTT**
490	−490 to −431	5′ **TCTCATCCAA**GATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGCCAGT**TGGTCAAAAA** 3′
490c		3′ **AGAGTAGGTT**CTACTTGATTCAAACCAAGCAACTTTACGATTGCCGGTCA**ACCAGTTTTT** 5′
440	−440 to −381	5′ **TGGTCAAAAA**GAAACTTCCAAAAGTCGCCATACCGTTTGTCTTGTTTGGT**ATTGATTGAC** 3′
440c		3′ **ACCAGTTTTT**CTTTGAAGGTTTTCAGCGGTATGGCAAACAGAACAAACCA**TAACTAACTG** 5′
390	−390 to −331	5′ **ATTGATTGAC**GAATGCTCAAAAATAATCTCATTAATGCTTAGCGCAGTCT**CTCTATCGCT** 3′
390c		3′ **TAACTAACTG**CTTACGAGTTTTTATTAGAGTAATTACGAATCGCGTCAGA**GAGATAGCGA** 5′
340	−340 to −281	5′ **CTCTATCGCT**TCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAAAC**ACCCGCT** 3′
340c		3′ **GAGATAGCGA**AGACTTGGGGCCACGTGGACACGGCTTTGCGTTTACCCCTTTG**TGGGCGA** 5′
287	−287 to −228	5′ ACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGG 3′
287c		3′ TGGGCGAAAAACCTACTAATACGTAACAGAGGTGTAACATACGAAGGTTCTAAGACCACC 5′
227	−227 to −174	5′ GAATACTGCTGATAGCCTAACGTTCATGATCAAAATTTAACTGTTCTAACCCCT 3′
227c		3′ CTTATGACGACTATCGGATTGCAAGTACTAGTTTTAAATTGACAAGATTGGGGA 5′
173	−173 to −114	5′ ACTTGACAGCAATATATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCA 3′
173c		3′ TGAACTGTCGTTATATATTTGTCTTCCTTCGACGGGACAGAATTTGGAAAAAAAAATAGT 5′

Bgl II restriction site at −940 bp is underlined. Triangle indicates transcription start site (−114). Overlapping regions between two oligonucleotides are shown in bold face. Doublestranded oligonucleotides were radiolabeled at their 5′ ends by polynucleotide kinase and used in EMSA.

(Graphpad Software, Inc, San Diego, CA). The equation, $Y = (Yo - NS)(K^*) + NS$ was used, where Yo is the binding at time zero, NS is nonspecific binding at infinite time and K is the rate constant.

2.5. DNase I foot printing

Full length double stranded DNA was generated by annealing equimolar amounts of 5′ end labeled single stranded oligonucleotide with unlabeled complementary strand. These radiolabeled probes (20,000 cpm, ~0.5 nM) were incubated on ice with Mxr1p (60 nM/20 ng) and 100 ng of poly (dI-dC) in a 30 µl reaction mixture containing buffer B for 30 min. A 50 µl solution consisting of 10 mM $MgCl_2$ and 5 mM $CaCl_2$ was added followed by 0.02 U of DNase I (Sigma-Aldrich, USA) and the samples were incubated at room temperature for 1 min. DNase I digestion was arrested by the addition of EDTA to a final concentration of 50 mM. Following deproteinization

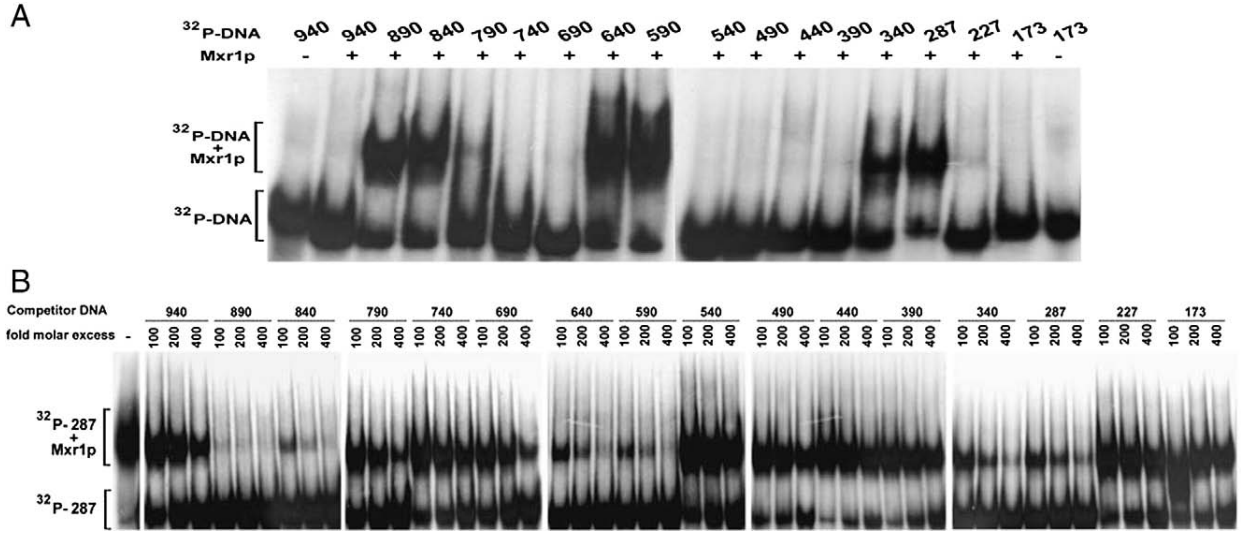


Fig. 2. Analysis of Mxr1p binding to P_{AOXI} DNA by EMSA. (A) Interaction of Mxr1p with ^{32}P-labeled P_{AOXI} oligonucleotides. Mxr1p forms a complex with six of the sixteen oligonucleotides. EMSA was carried out with radiolabeled probes (20,000 cpm, ~0.2–0.5 nM), 100 ng of (polydI-dC) and Mxr1p (30 nM). (B) Competition EMSA. Binding of Mxr1p (30 nM) to ^{32}P-labeled 287 oligonucleotide (20,000 cpm, ~0.5 nM) is outcompeted by 100–400 fold molar excess of unlabeled duplex oligonucleotides 890, 840, 640, 590, 340 and 287 but not 940, 790, 740, 540, 490, 440, 390, 227, and 173.

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

and ethanol precipitation, the DNA was dissolved in a buffer containing 80% formamide, 50 mM Tris–borate, 1 mM EDTA, 0.01% xylene cyanol and 0.1% bromophenol blue. The DNA is denatured and then subjected to electrophoresis on 12% polyacrylamide gels containing 8 M urea. Foot prints were visualized by autoradiography and the foot print regions were identified by comparing their positions on the A + G sequence ladder.

2.6. Purification of Helicobacter pylori C^5 cytosine methyltransferase (HP0051)

The gene encoding *Helicobacter pylori* C^5 cytosine DNA methyltransferase, HP0051 (GenBank accession # AAD07117) was cloned into pET28a vector and transformed into *E. coli* BL21(DE3)pLysS cells. Conditions for overexpression and purification of HP0051 enzyme from *E. coli* extracts were similar to those employed for Mxr1p with minor modifications. The purified enzyme was stored in aliquots at −80 °C in 10 mM Tris–HCl, pH 8.0 containing 100 mM NaCl, 11 mM β-mercaptoethanol and 30% glycerol.

2.7. In vitro methylation of P_{AOX}MXREs using HP0051 enzyme

Methylation of P_{AOX}MXREs by HP0051 enzyme was monitored by the incorporation of tritiated methyl groups into DNA. Briefly, methylation assays were carried out in a reaction mixture containing 50 mM Tris–HCl, pH 7.5, 5 mM β-mercaptoethanol, 0.05 μM P_{AOX}MXREs, [*methyl*-^{3}H]-S-adenosylmethionine (66 Ci/mmol, GE Healthcare, India) and 0.5 μM purified HP0051 enzyme. After incubation of enzyme and substrate at 37 °C for 3 h, reactions were stopped by snap-freezing in liquid nitrogen and aliquots were transferred onto Whatman DE81 filter paper discs. Filters were washed three times (15 min each) with 200 mM ammonium bicarbonate solution equilibrated at 4 °C and once with 95% (v/v) ethanol. Filters were air dried, and the tritium content was determined in 3 ml of scintillation fluid (0.5% 2, 5-diphenyloxazole and 0.05% 1, 4-bis [2-(5-phenyl-oxazoyl)] benzene in 1:1 (v/v) 2-methoxy ethanol and toluene) in a liquid scintillation counter (Beckman Coulter, Fullerton, CA).

For use in EMSA, ^{32}P-labeled double stranded DNA was methylated by HP0051 enzyme exactly as described above except

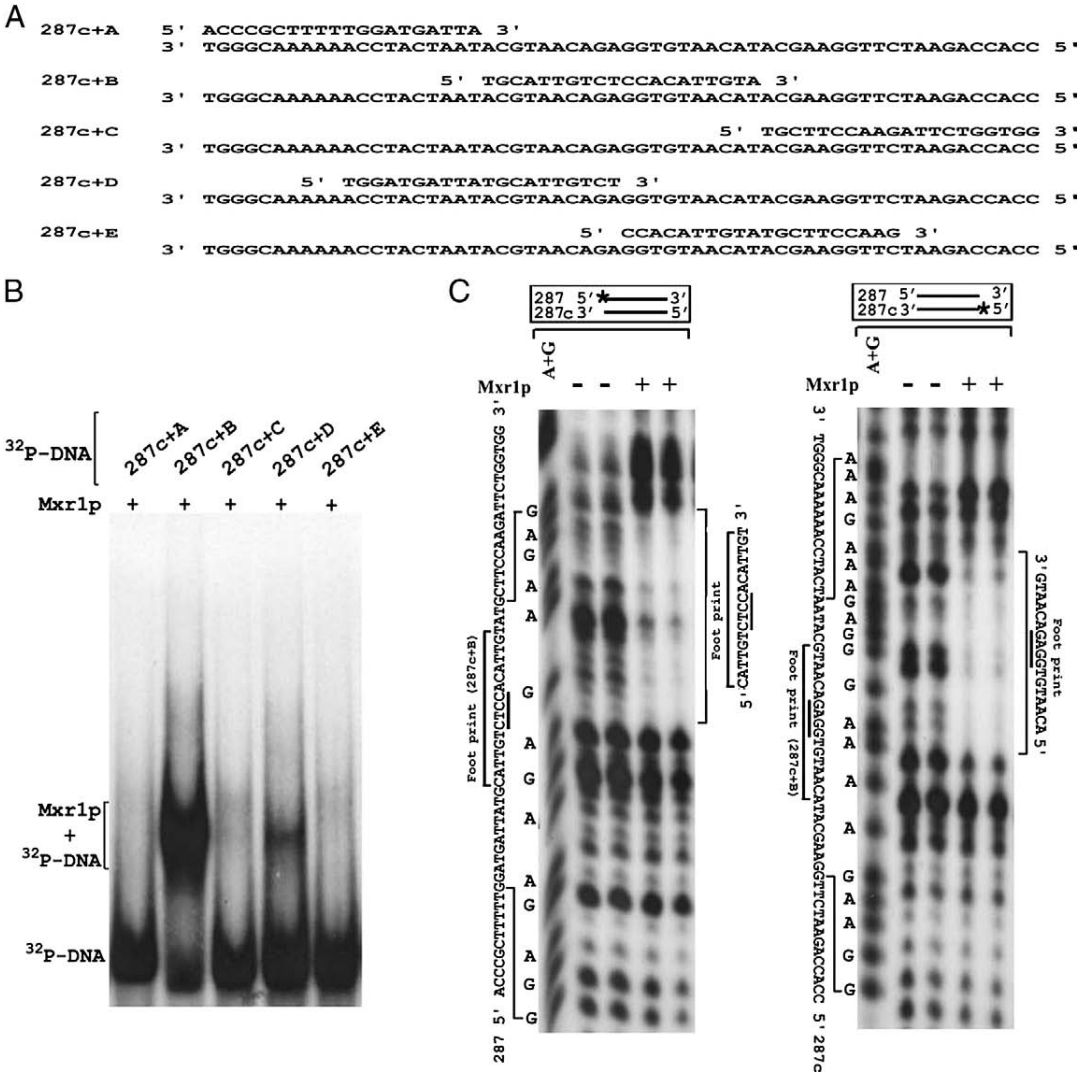


Fig. 3. Identification of P_{AOX}MXRE in the 287 oligonucleotide by EMSA and DNase I foot printing. (A) Nucleotide sequence of partially double stranded oligonucleotides generated for use in EMSA. (B) Binding of Mxr1p to partially double stranded, radiolabeled oligonucleotides (20,000 cpm, ~0.15–0.20 nM) and Mxr1p (30 nM) as evaluated by EMSA. (C) DNase I foot print generated by Mxr1p in the sense (287) and antisense (287c) strands of 287 double stranded DNA (60 bp) is shown. Reactions were carried out in duplicate, in the absence (−,−) or presence (+,+) of 60 nM Mxr1p as described in Materials and methods. The double stranded DNA carrying radiolabel at the 5′ end of one of the DNA strands (asterisk) is shown schematically. The core sequence 5′ CTCC 3′ (in 287) and its complement, 5′ GGAG 3′ (in 287c) within the foot print are underlined (see text for details).

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

that [*methyl*-^{3}H]-S-adenosylmethionine was replaced by 10 μM unlabeled S-adenosylmethionine. Following methylation, DNA was deproteinized, ethanol precipitated, redissolved in water and used in EMSA.

3. Results and discussion

3.1. Mxr1p binds at six different regions in P_{AOXI}

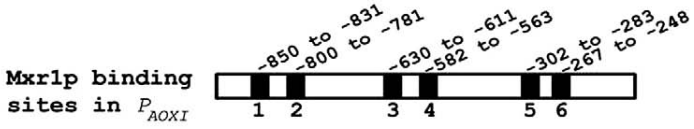
Mxr1p is a large protein of 1155 amino acids with an N-terminal zinc finger DNA binding domain located between amino acids 38 and 95 [5]. Our attempts to overexpress a recombinant Mxr1p protein containing only the zinc finger domain (amino acids 38–95) were not successful. However, Mxr1p containing 150 N-terminal amino acids could be expressed at high levels in *E. coli* cells and purified as a histidine-tagged protein (Figs. 1A–D). The recombinant protein was identified as Mxr1p by amino acid sequence analysis of one of the tryptic peptides (data not shown). This recombinant protein, referred to as Mxr1p in the rest of this study, contains a region homologous to the proximal accessory region (PAR) of Adr1p (Fig. 1E), in addition to the zinc finger domain [11–13].

To identify Mxr1p binding sites in P_{AOXI}, sixteen different double stranded oligonucleotides spanning the P_{AOX1} region between −940 and −1 were synthesized (Table 1) and their ability to bind to Mxr1p was examined by EMSA. Of the sixteen oligonucleotides, only six (890, 840, 640, 590, 340 and 287) generated a complex with Mxr1p (Fig. 2A). In competition experiments, only the duplex oligonucleotides 890, 840, 640, 590, 340 and 287 which bind to Mxr1p (Fig. 2A) efficiently outcompete Mxr1p binding to ^{32}P-labeled 287, when used as unlabeled competitors (Fig. 2B) indicating that Mxr1p binding to these DNA sequences is very specific.

3.2. Identification of minimal Mxr1p binding sites in P_{AOXI}

To identify the minimal Mxr1p binding sites within the six P_{AOX1} oligonucleotides, partially double stranded oligonucleotides

A



B

Deletion mutant[a]	Relative P_{AOXI} activity[a]	deleted region in P_{AOXI}[a]	Mxr1p binding site[b]
dHSF1	60%	Δ-805 to -798	2
d2	33%	Δ-643 to -597	3
dAdr1	30%	Δ-576 to -570	4
d5	40%	Δ-322 to -264	5,6
dMat1-MC	42%	Δ-264 to -260	6
d6	45%	Δ-253 to -224	6

[a] Hartner *et al.*,(6); [b] This study.

C

Fig. 4. Correlation between Mxr1p binding sites identified in this study and DNA sequences essential for P_{AOXI} function *in vivo*. (A) Schematic representation of the *AOXI* promoter and location of the six Mxr1p binding sites identified in this study. (B) Effect of deletion of different regions of *AOXI* promoter on methanol-inducible P_{AOXI} activity *in vivo* as reported by Hartner et al. [6]. The six Mxr1p binding sites (P_{AOX}MXREs) identified in this study (last column) are present within these deleted regions. (C) Nucleotide sequence of *AOXI* promoter and location of the P_{AOX}MXREs in *AOXI* promoter. The conserved 5′ CYCC/GGRG 3′ motifs in the MXREs are underlined. The promoter deletions which affect P_{AOXI} function *in vivo* as reported by Hartner et al. [6] are shown by broken lines and arrows.

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

were generated by annealing the antisense strands with short, sense strand oligonucleotides. For example, partially double stranded regions were generated in the P_{AOX1} region between −287 and −228 by annealing the 60 bp long antisense strand (287c) with five different sense strand oligonucleotides, each of which is of 20 bp in length (Fig. 3A). The five different partially double stranded oligonucleotides thus generated (287c+A–287c+E) were radiolabeled and used in EMSA. Interestingly, only one of them (287c+B) generated a prominent complex with Mxr1p (Fig. 3B) indicating the presence of an Mxr1p binding site within this 20 bp region. A faint Mxr1p–DNA complex was obtained with 287c+D and this may be due to the presence of a 10 bp region that is common between 287c+D and 287c+B. The Mxr1p binding site within double stranded 287+287c oligonucleotide was also identified by DNase I foot printing. A foot print was generated by Mxr1p in the 287c+B region in the sense as well as antisense strands (Fig. 2C). Thus, using a combination of EMSA and DNase I foot printing, Mxr1p binding sites were identified in all the remaining five oligonucleotides (Fig. 4A, Supplementary Figs. S1–S5).

3.3. Deletion of Mxr1p binding sites in P_{AOXI} reduces promoter activity in vivo

To examine whether the Mxr1p binding sites identified in P_{AOXI} have a role in P_{AOXI} function *in vivo*, we compared their location with that of cis-acting elements and putative transcription factor binding sites identified recently by Hartner et al. [6]. Such a comparison revealed a significant overlap between the Mxr1p binding sites identified in this study and putative binding sites for transcription factors such as HSF1, Mat1-MC and Adr1p reported by Hartner et al. [6]. Further, a number of promoter deletions which result in a significant decrease in P_{AOXI} activity *in vivo* involved deletions within Mxr1p binding sites (Fig. 4B). For example, the d2 deletion between −643 and −597, whose deletion results in a 67% decrease in P_{AOX1} activity, includes Mxr1p binding site 3 (Fig. 4B). Similarly, d5, dMat1-MC and d6 deletions include deletions within Mxr1p binding sites 5 and 6 while dHSF1 and dAdr1 deletions involve deletions within Mxr1p binding sites 2 and 4 respectively (Fig. 4B). In view of the strong correlation between deletions within Mxr1p binding sites and decrease in P_{AOXI} function, we have named the six P_{AOXI} Mxr1p binding sites identified in this study as P_{AOXI} Mxr1p response elements (P_{AOX}MXREs) (Figs. 4B, C).

3.4. Identification of a conserved motif in P_{AOX}MXREs

To further characterize the P_{AOX}MXREs, six different oligonucleotides of 20 bp length referred to as P_{AOX}MXRE1–P_{AOX}MXRE6 were synthesized (Fig. 4C). The 20 bp region was chosen based on the results of EMSA with partially double stranded oligonucleotides wherein the 287c+B partially double stranded DNA, which contains a 20 bp double stranded region was found to bind Mxr1p very efficiently (Fig. 3B). When used in EMSA, all of the six P_{AOX}MXREs generated a complex with Mxr1p (Fig. 5A). Analysis of the nucleotide sequence of the P_{AOX}MXREs revealed that the sequence 5′ CYCC 3′ (5′ GGRG 3′ in the complementary strand) is conserved in all the six P_{AOX1}-MXREs (Fig. 4C). A single point mutation within this conserved motif (5′ CYCC 3′ to 5′ CYCA 3′) abrogated Mxr1p binding, indicating a critical role for this sequence in Mxr1p–DNA interactions (Fig. 5B). The importance of this conserved motif in Mxr1p binding was further confirmed by DNA methylation studies employing a recombinant *Helicobacter pylori* C^5 cytosine methyltransferase (HP0051), which specifically recognizes the 5′ GGRG 3′ sequence in the double stranded DNA and methylates the terminal cytosine residue on the opposite strand within this motif (Supplementary Fig. S6, R.K. and D.N.R. unpublished data). Methylation of a specific cytosine residue within the core 5′ CYCC 3′ motif of P_{AOX}MXREs by HP0051 enzyme was confirmed by an *in vitro* methylation reaction using [^{3}H]-S-adenosylmethionine as methyl group donor. Except P_{AOX}MXRE5, all other P_{AOX}MXREs were methylated with approximately the same efficiency by HP0051 enzyme (Supplementary Fig. S6). However, P_{AOX}MXRE4-M1, P_{AOX}MXRE6-M1, which carry C to A point mutation within the 5′ CYCC 3′ motif, were not methylated by HP0051 enzyme confirming the specificity of methylation reaction (Supplementary Fig. S6). In EMSA, Mxr1p binding to methylated P_{AOX}MXREs was significantly reduced (Fig. 6A) confirming a critical role for the cytosine residue within the 5′ CYCC 3′ motif in Mxr1p binding. Since methylcytosine as well as thymine carry the bulky methyl group at position 5 of the pyrimidine ring, we synthesized P_{AOX}MXRE4-M2 and P_{AOX}MXRE6-M2 oligonucleotides which carry C to T point mutation within the 5′ CYCC 3′ motif and both these mutants do not bind to Mxr1p (Fig. 6B). These results are summarized in Fig. 6C. Based on the alignment of the P_{AOX}MXREs, we have derived a core consensus sequence for Mxr1p binding (Fig. 6D). Further, we calculated the apparent dissociation constant (K_d) for Mxr1p binding to three of the six *P_{AOX}MXREs* and the K_d values were found to be in the range of 50–150 nM (Fig. 7). In comparison, the K_d values reported for Adr1p-*ADH2UAS1* interactions are in the range of

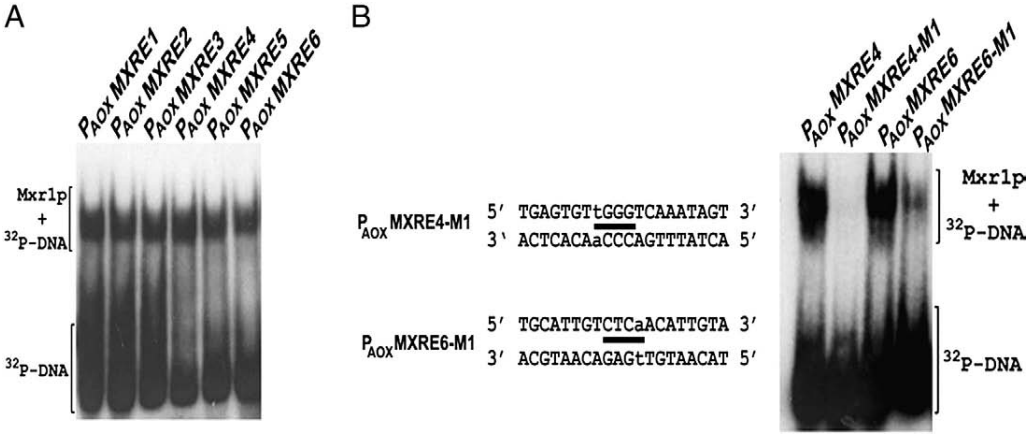


Fig. 5. Mxr1p binding to P_{AOX}MXREs as analyzed by EMSA. (A) Mxr1p (30 nM) binds to radiolabeled P_{AOX}MXREs (30,000 cpm, ~0.15–0.20 nM) in presence of 100 ng of poly (dI-dC). (B) Radiolabeled P_{AOX}MXREs (30,000 cpm, ~0.15–0.20 nM) carrying point mutations within the conserved 5′ CYCC 3′ motif (P_{AOX}MXRE4-M1, P_{AOX}MXRE6-M1) do not bind to Mxr1p (60 nM). The mutated nucleotides are indicated in lower case. The conserved core motif is underlined.

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

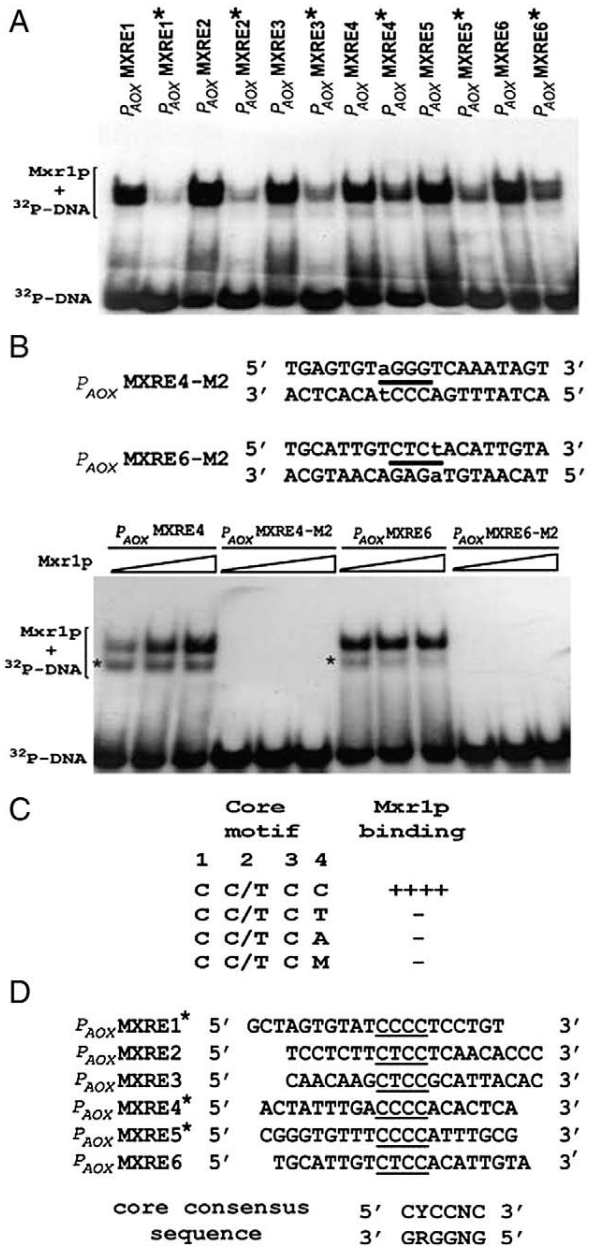


Fig. 6. Effect of methylation or mutation of a cytosine residue in the core motif of P_{AOX}MXREs on Mxr1p binding. (A) EMSA was carried out with radiolabeled, unmethylated as well as methylated (asterisk) P_{AOX}MXREs (30,000 cpm, ~0.15–0.25 nM) and Mxr1p (30 nM) as described in Materials and methods. (B) Effect of mutation of cytosine to thymine within the conserved 5′ CYCC 3′ motif of P_{AOX}MXREs on Mxr1p binding as evaluated by EMSA. The mutated nucleotides are indicated in lower case. The conserved core motif is underlined. Asterisk indicates complex formation by Mxr1p molecules which are not full length. These may arise due to incomplete translation or proteolytic degradation during isolation, purification and storage. (C) Summary of the mutagenesis and methylation studies. M is 5-methyl cytosine. (D) Alignment of the nucleotide sequences of P_{AOX}MXREs and derivation of a core consensus sequence for Mxr1p binding.

2–9 nM (7), suggesting that Mxr1p binds DNA with an affinity lower than that of Adr1p.

3.5. Analysis of Mxr1p binding to S. cerevisiae alcohol dehydrogenase promoter upstream activation sequence 1 (ADH2UAS1)

Mxr1p is considered to be a homologue of Adr1p [5,6] and therefore it was of interest to examine whether Mxr1p binds to Adr1p binding sites. An oligonucleotide containing the *S. cerevisiae ADH2*-UAS1 which harbours an Adr1p binding site was synthesized (Fig. 8A) and its ability to outcompete Mxr1p binding to radiolabeled P_{AOX}MXRE6 was evaluated by EMSA. The results presented in Fig. 8B indicate that *ADH2UAS1* failed to outcompete Mxr1p binding to ^{32}P-labeled P_{AOX}MXRE6. Increasing the length of *ADH2UAS1* to 38 bp such that the 5′ CTCC 3′ core motif is flanked by 9 bp, did not restore Mxr1p binding to *ADH2UAS1* (Fig. 8C). Alignment of the nucleotide sequences of Adr1p and Mxr1p binding sites indicated a key nucleotide difference between these two binding sites (Fig. 9A). We hypothesized that mutation of an adenine residue in the Adr1p monomer binding site of *ADH2UAS1* to cytosine should restore Mxr1p binding to *ADH2UAS1* (Fig. 9A). The results presented in Fig. 9B indicate that mutation of adenine to cytosine does restore Mxr1p binding to *ADH2UAS1*. Thus, despite extensive homology in the DNA binding domains and DNA recognition sequences of Mxr1p and Adr1p, the target gene specificity of these two zinc finger proteins is different.

In the case of Adr1p, the amino acid residues Arg115, His118 and Arg121 present in −1, +3, and +6 positions of the first zinc finger are known to contact the bases 3′ GA/GG 5′, of the core recognition motif (11). In the second zinc finger, only Arg143 present in the −1 position makes contact with the 5′ G residue of the core motif (11). Comparison of the zinc finger motif of Adr1p with that of Mxr1p indicates that all the amino acid residues mentioned above are also conserved in the Mxr1p zinc fingers (Supplementary Fig. S7A). Therefore, it is not surprising that both Adr1p and Mxr1p recognize the same core nucleotide quartet (5′ GGRG 3′). In addition to the two zinc fingers, a 20-amino acid domain referred to as the proximal accessory region (PAR) present N-terminal to the first zinc finger is also essential for high affinity DNA binding of Adr1p (12, 13). The Adr1p PAR exhibits 32% identity with a homologous region in Mxr1p and a number of amino acid residues within the PAR that are critical for DNA binding are conserved between these two proteins (Supplementary Fig. S7B). Further, in Adr1p PAR, Pro97 executes a turn so that Arg95, Gly94, Lys100 and Leu101 can contact the DNA (13). All these residues are also conserved in Mxr1p except Arg95 which is replaced by Thr31 in Mxr1p (Supplementary Fig. S7B). Since Mxr1p does not bind to *ADH2UAS1*, it will be interesting to examine whether T31R mutation in the putative PAR domain of Mxr1p will restore its binding to *ADH2UAS1*. Since Mxr1p binds to DNA with an affinity lower than that of Adr1p (Fig. 7), it is possible that amino acid differences in the PAR domains of Mxr1p and Adr1p as reported in this study, are responsible not only for altered DNA binding specificity but also weak affinity of Mxr1p for MXREs. Thus, it will be interesting to examine whether mutations in the putative PAR of Mxr1p affect the DNA binding specificity or DNA binding affinity or both.

Thus far, two putative Mxr1p binding sites have been reported in P_{AOXI}. Based on EMSA carried out with *P. pastoris* cell extracts overexpressing Mxr1p and a 243 bp DNA probe, an Mxr1p binding site was identified between −415 to −172 bp [5]. We now demonstrate that this region actually harbours two MXREs, P_{AOX}MXREs 5 and 6 (Fig. 4C). In another study, based on *in silico* analysis of P_{AOX1} for potential transcription factor binding sites, a putative Adr1p binding site was identified between −577 and −571 bp [6]. We now demonstrate that this region harbours P_{AOX}MXRE4. In addition, we have identified three more MXREs within P_{AOX1} (Fig. 4). The *AOXI* promoter deletion studies of Hartner et al. indicate that deletions encompassing five of the six P_{AOX}MXREs result in a significant decrease in P_{AOX1} function *in vivo* [6]. Based on the results presented in this study, it is now possible to carry out specific site-directed mutagenesis studies so that the effect of loss of Mxr1p binding to each of the six P_{AOX}MXREs on AOXI promoter function can be studied. It is likely that synergistic interactions amongst Mxr1p molecules binding to multiple P_{AOX}MXREs play a key role in the

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

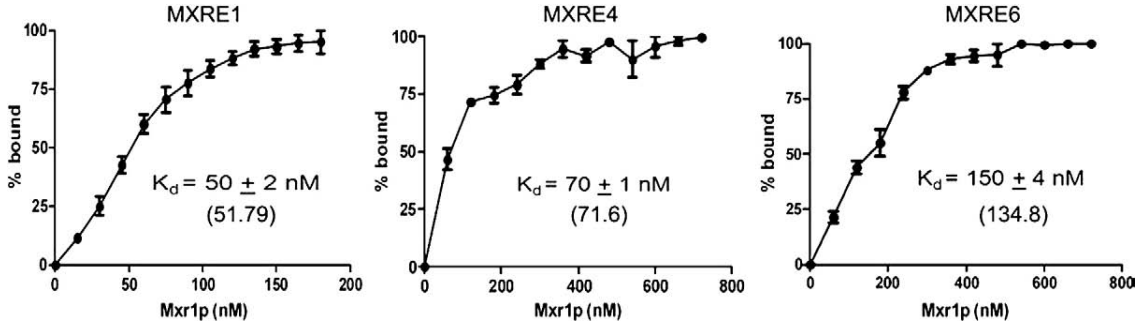


Fig. 7. Determination of K_d. EMSA reactions were carried out in the presence of 0.3–0.5 nM radiolabeled P_{AOX}MXREs and varying amounts of Mxr1p as indicated in the presence of 100 ng of poly dI-dC. In order to obtain K_d values of Mxr1p–P_{AOX}MXRE interactions, the percentage of radiolabeled DNA–Mxr1p complex was quantitated and plotted against the quantity of Mxr1p (nM) used. The data is the average of two independent saturation binding experiments and the error bars represent standard deviation of the mean. The numbers in parentheses represent K_d values calculated by non-linear regression using GraphPad Prism (for details, see Materials and methods).

high level of methanol-inducible expression of P_{AOX1}. Although Mxr1p appears to be the major regulator of P_{AOXI}, the fact that mutations and deletions in P_{AOX} outside the P_{AOX}MXREs also affect P_{AOX1} activity [6] indicate that cis-acting elements and transcription factors other than P_{AOX}MXREs and Mxr1p may also play a key role in the regulation of *AOXI* expression.

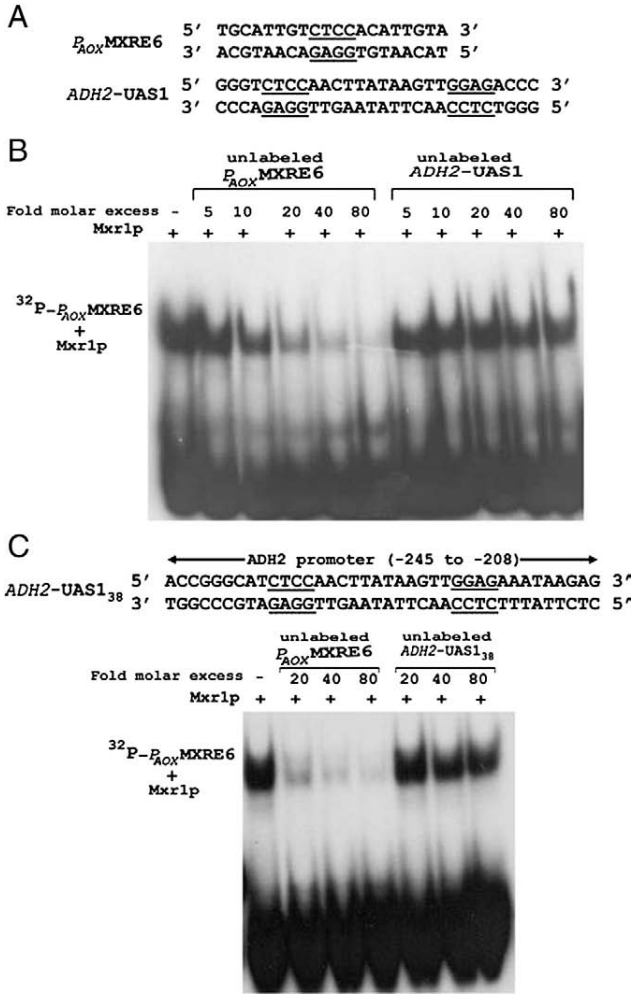


Fig. 8. Analysis of Mxr1p binding to *S. cerevisiae ADH2UAS1* by EMSA. (A) Nucleotide sequence of P_{AOX}MXRE6 and *ADH2UAS1*. The core CYCC motifs are underlined. (B) Competition EMSA was carried out with 0.5 nM radiolabeled probe and 30 nM Mxr1p. Unlabeled P_{AOXI}MXRE6 but not *ADH2UAS1* competes out binding of Mxr1p to ^{32}P-labeled P_{AOX}MXRE6. (C) Effect of increasing the length of *ADH2UAS1* on Mxr1p binding as analyzed by competition EMSA.

Acknowledgements

This work was supported by a research grant from the Department of Biotechnology, New Delhi, India to P.N.R.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbagrm.2009.05.004.

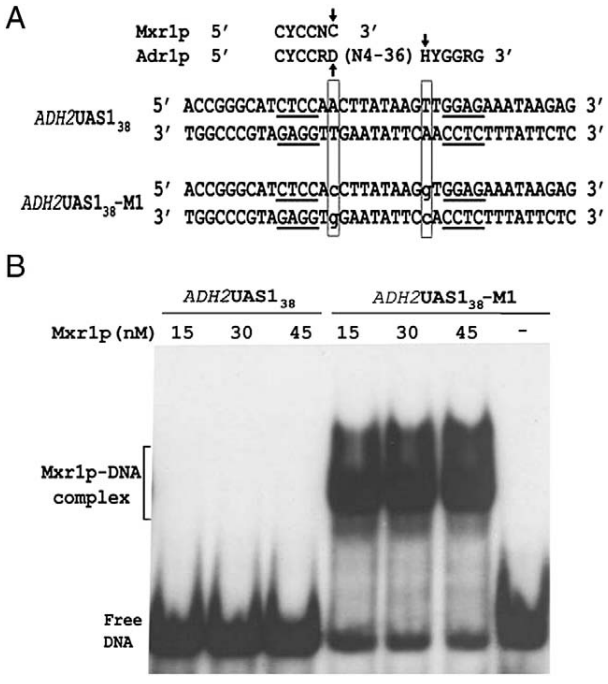


Fig. 9. Conversion of an Adr1p binding site into an Mxr1p binding site. (A) Comparison of consensus DNA sequences for the binding of Adr1p (9) and Mxr1p. Key nucleotide differences are indicated by arrows. Nucleotide sequence of the *ADH2UAS1*$_{38}$ corresponding to the *S. cerevisiae ADH2* promoter region between −245 and −208 bp is compared with that of the mutant oligonucleotide, *ADH2UAS1*$_{38}$-M1. The nucleotide differences are boxed and the mutated bases are indicated in lower case. The core 5′ CYCC 3′ motif is underlined in all the sequences. IUB nucleotide codes used are: H:A/C/T; D:A/G/T; Y:C/T; R:A/G; N:A/G/C/T. (B) Analysis of Mxr1p binding to ^{32}P-labeled *ADH2UAS1*$_{38}$ and *ADH2UAS1*$_{38}$-M1 by EMSA.

Case No.: IPR2023-00321
U.S. Patent No. 10,689,656

References

[1] J. Lin-Cereghino, J.M. Cregg, Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, FEMS Microbiol. Rev. 24 (2000) 45–66.
[2] J.M. Cregg, K.R. Madden, K.J. Barringer, G.P. Thill, C.A. Stillman, Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*, Mol. Cell Biol. 9 (1989) 1316–1323.
[3] F.S. Hartner, A. Glieder, Regulation of methanol utilisation pathway genes in yeasts, Microb. Cell Fact. 5 (2006) 39.
[4] S. Subramani, A. Koller, W.B. Snyder, Import of peroxisomal matrix and membrane proteins, Annu. Rev. Biochem. 69 (2000) 399–418.
[5] G.P. Lin-Cereghino, L. Godfrey, B.J. de la Cruz, S. Johnson, S. Khuongsathiene, I. Tolstorukov, M. Yan, J. Lin-Cereghino, M. Veenhuis, S. Subramani, J.M. Cregg, Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in *Pichia pastoris*, Mol. Cell Biol. 26 (2006) 883–897.
[6] F.S. Hartner, C. Ruth, D. Langenegger, S.N. Johnson, P. Hyka, G.P. Lin-Cereghino, J. Lin-Cereghino, K. Kovar, J.M. Cregg, A. Glieder, Promoter library designed for fine-tuned gene expression in *Pichia pastoris*, Nucleic Acids Res. 36 (2008) e76.
[7] S.K. Thukral, A. Eisen, E.T. Young, Two monomers of yeast transcription factor ADR1 bind a palindromic sequence symmetrically to activate ADH2 expression, Mol. Cell Biol. 11 (1991) 1566–1577.
[8] A. Eisen, W.E. Taylor, H.T. Blumberg, E.T. Young, The yeast regulatory protein ADR1 binds in a zinc-dependent manner to the upstream activating sequence of ADH2, Mol. Cell Biol. 8 (1988) 4552–4558.
[9] C. Cheng, N. Kacherovsky, K.M. Dombek, S. Camier, S.K. Thukral, E. Rhim, E.T. Young, Identification of potential target genes for Adr1p through characterization of essential nucleotides in UAS1, Mol. Cell Biol. 14 (1994) 3842–3852.
[10] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ednCold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[11] C. Cheng, E.T. Young, A single amino acid substitution in zinc Finger 2 of Adr1p changes its binding specificity at two positions in UAS1, J. Mol. Biol. 251 (1995) 1–8.
[12] S.K. Thukral, A. Eisen, E.T. Young, Two monomers of yeast transcription factor ADR1 bind a palindromic sequence symmetrically to activate ADH2 expression, Mol. Cell Biol. 11 (1991) 1566–1577.
[13] P.M. Bowers, L.E. Scvhaufler, R.E. Klevit, A folding transition and novel zinc finger accessory domain in the transcription factor ADR1, Nat. Struct. Biol. 6 (1999) 478–485.