

Eukaryotic Gene Expression: Basics & Benefits

P N RANGARAJAN

Lecture 5

**Eukaryotic Transcription factors:
the DNA binding domain**

So far.....

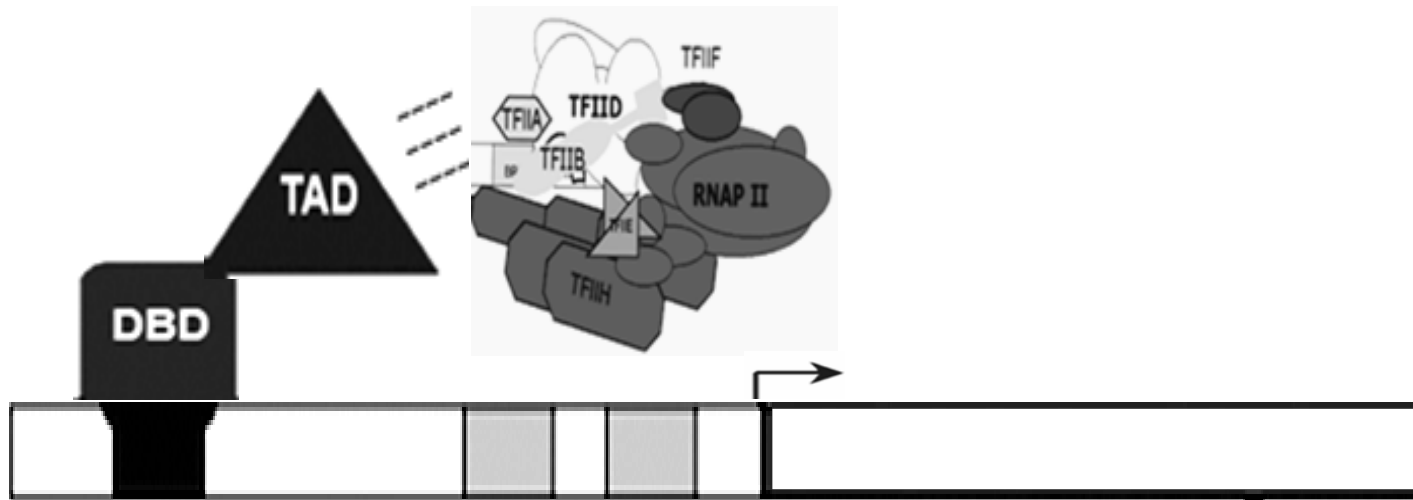
- **Eukaryotic RNA polymerases (lecture 1)**
- **Core promoter sequences and their variants (lecture 2)**
- **General transcription factors and their variants (lecture 3)**

Lecture 4:

- **Proximal & distal promoter elements & enhancers**
- **Identification and characterization of transcriptional activators**
- **Benefits arising out of study of promoters and transcription factors**

In this lecture.....

- **How eukaryotic transcription factors bind DNA in a sequence- specific manner?**



Transcription factors are modular proteins composed of distinct and separable functional domains:

1. DNA binding domain (DBD) interacts with specific DNA sequences
2. Transcription Activation domain (TAD) interacts with other proteins (general transcription machinery, coactivators etc.), to stimulate transcription from the core promoter

Many eukaryotic transcription factors bind DNA as dimers

Dimerization domain.....binding to each other

Effector binding domain

for binding to an effector molecule



Transcription factors that bind DNA do so in a sequence specific manner.

The three common features most DNA binding proteins are:

- they bind to the *major groove* (12Å wide and 8Å deep) of B- DNA through α –helices.
- the *minor groove* of B-DNA (5Å wide, 8Å deep) is generally too narrow to fit entire α –helices.
- they generally *do not disrupt the base pairs* of the DNA, but distort the conformation of the backbone by bending the double helix

The role of DNA binding domain is to bring the transcription activation domain into the vicinity of the preinitiation complex

DNA binding domains of eukaryotic transcription factors

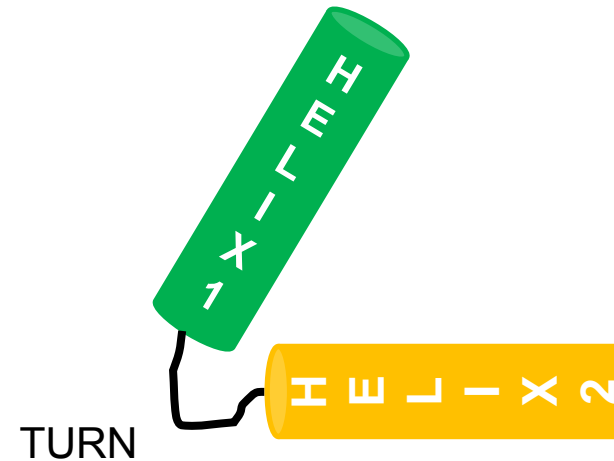
In general, transcription factors which bind DNA in a sequence-specific manner contain characteristic DNA binding domains or motifs.

In fact, eukaryotic transcription factors are categorized into distinct families based on the type of DNA binding motifs they contain

Eukaryotic transcription factor families

- 1 Helix-Turn-Helix proteins**
- 2 Zinc finger proteins**
- 3 Leucine zipper proteins**
- 4 Helix-Loop-Helix proteins**

The helix-turn-helix motif



The helix-turn-helix motif was initially identified as the DNA-binding domain of bacteriophage repressors

These repressors contain two alpha helices: one that lies in the major groove of DNA and the other lies at an angle across DNA.

Two adjacent α -helices separated by a "turn" of several amino acids enables the protein to bind to DNA. However, the HTH motif cannot fold or function alone, but is always part of a larger DNA-binding domain and amino acid residues outside the HTH motif are important in regulating DNA recognition and binding.

A motif similar to helix-turn-helix motif is present in certain eukaryotic transcription factors that play a key role in regulation of embryonic development.

The DNA binding domain in these proteins is called the homeodomain and the C-terminal region of homeodomain shows homology with the helix-turn-helix motif of procaryotic repressors. it contains 60 amino acid motif known as homeobox.



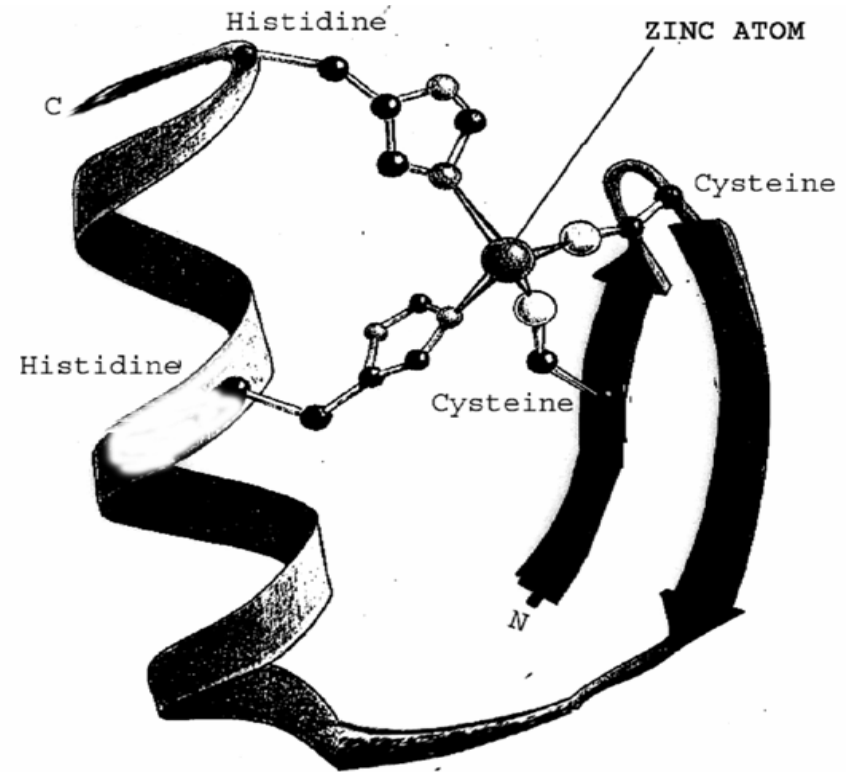
The zinc finger motif

Certain eukaryotic transcription factors contain a unique DNA binding motif called a zinc-finger where a zinc ion is coordinated by 2 Cysteine and 2 Histidine residues.

This motif was first discovered in a transcription factor known as TFIIIA in the frog, *Xenopus laevis*.

The number of zinc fingers vary in different transcription factors

The non-conserved aminoacids in the C-terminal side of each finger recognize specific target sites.



The C-terminal part of each finger forms α -helices that bind DNA while the N-terminal part form β -sheets.

A C2H2 zinc finger protein has a series of zinc fingers and the consensus sequence of a single finger is:

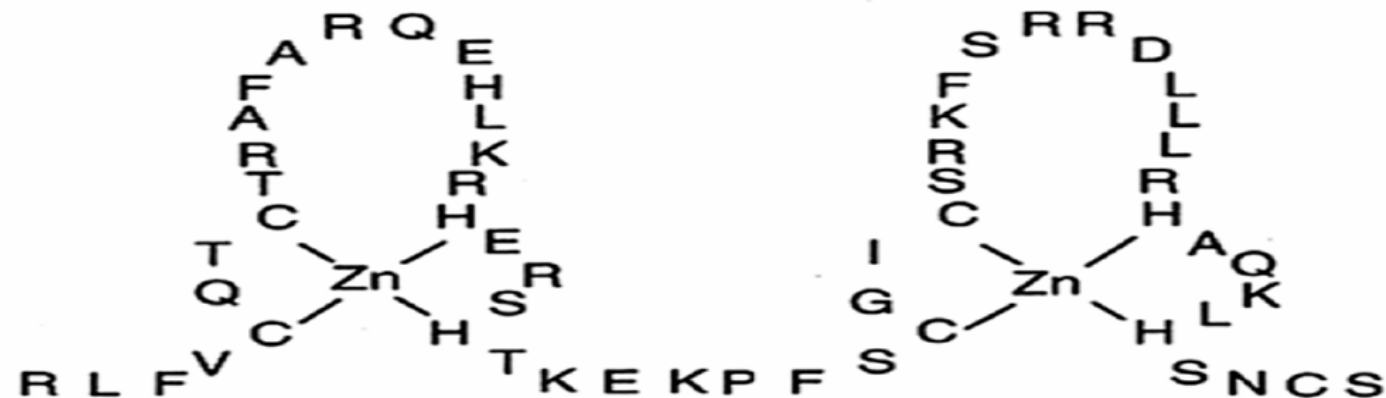
Cys-X2-4-Cys-X3-Phe-X3-Leu-X2-His-X3-His

The interspersed cysteine and histidine residues covalently bind zinc atoms, folding the amino acids into loops known as zinc fingers.

Each finger consists of approximately 23 amino acids (with a loop of 12 to 14 amino acids between the Cys and His residues) and a linker between loops consisting of 7 or 8 amino acids. The amino acids in the loop bind to specific DNA sequences.

MSNLPPTFGSTRQSPEDQSPPVPKELSFNGTTPSGKLRLFVCQTCTRAFARQEHLKRHER 60

SHTKEKPFSCGICSRKFSRRDLLLRHAQKLHSNCSDAAITRLRRKATRRESSNAAGSISGS 120

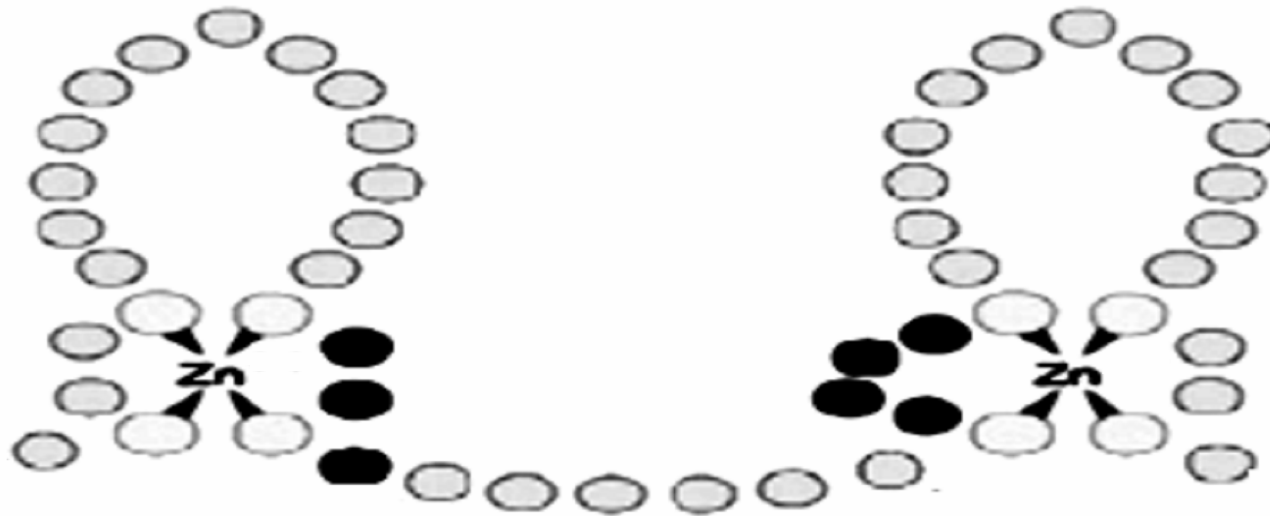




Zinc-finger proteins, represent one of the major families of eukaryotic transcription factors including those belonging to the steroid hormone receptor super family.

The steroid hormone receptors contain C2C2 zinc fingers instead of C2H2 fingers. They have the consensus sequence:

Cys-X2-Cys-X13-Cys-X2-Cys



In the case of steroid hormone receptors, the first zinc finger determines specificity of DNA binding and the second finger specifies dimerization specificity.

It has been shown that a single amino acid change in the base of the first finger can alter the DNA binding specificity of steroid hormone receptors.

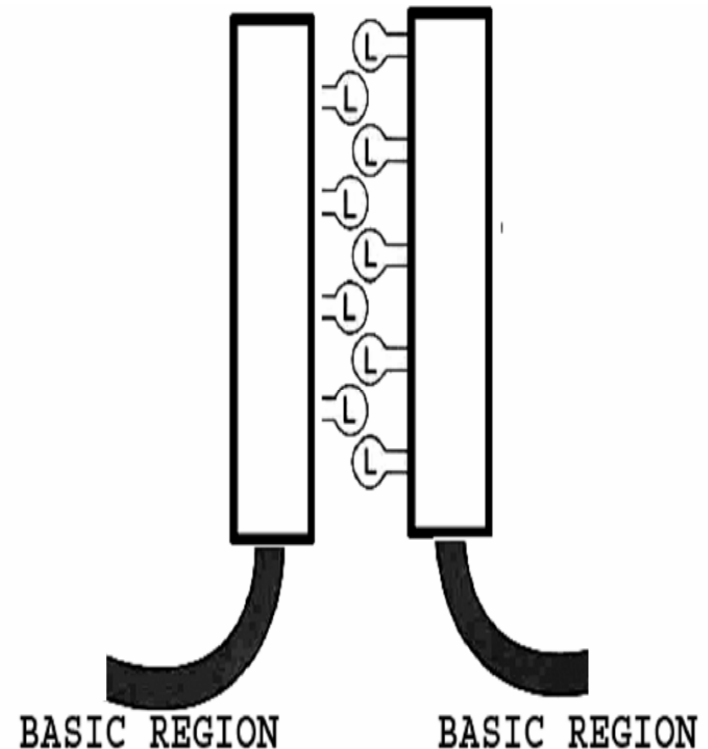
The basic leucine zipper (bzip)

In certain transcription factors the dimerization domain forms motif known as leucine zipper.

This motif consists of two amphipathic helices, one from each subunit, interacting with each other resulting in a left handed coiled-coil structure.

The leucine zipper is a interdigitation of regularly spaced leucine residues in one helix with leucines from the adjacent helix.

Leucine zipper motifs can mediate homodimer or heterodimer formation.

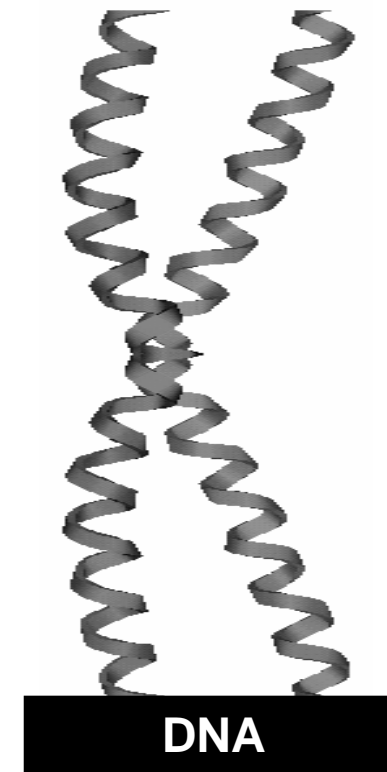


An important point to note is that the leucine zipper motif itself does not directly participate in DNA binding.

The DNA-binding domain is adjacent to the leucine zipper and the latter facilitates protein-protein dimerization. Dimerization allows the juxtaposition of the DNA-binding regions of each subunit.

A leucine zipper forms an amphipathic helix in which the leucines of the one protein protrude from the alpha helix and interdigitate with the leucines of the other protein in parallel to form a coiled coil domain.

The region adjacent to the leucine repeats is highly basic and forms the DNA-binding site.



The helix-loop-helix (HLH)

The amphipathic helix-loop-helix (HLH) motif was identified in certain transcription factors involved in developmental regulation. They contain a stretch of 40-50 aminoacids comprising of two amphipathic α -helices separated by a linker region (the loop) of varying length.

The HLH proteins form both homodimers and heterodimers by means of interactions between the hydrophobic residues on the corresponding faces of the 2 helices. The ability to form dimers resides with these amphipathic helices.

Most HLH proteins contain a basic region adjacent to the HLH motif that is involved in DNA binding and hence they are often called bHLH proteins.

Examples of HLH proteins include MyoD, Myf5, myogenin and MRF4 which are involved in myogenesis and are called myogenic regulatory factors (MRFs)

**Helix-turn-helix, zinc-finger, and
leucine-zipper motifs for
eukaryotic transcriptional
regulatory proteins**

Kevin Struhl

TIBS 14 – April 1989

1. *Cold-shock domain*
2. *Helix-turn-helix*
3. *Homeo-Box*
4. *POU domain*
5. *"bromo"domain*
6. *ZincFinger + types*
7. *CCAAT-binding dom*
8. *ETS-domain*
9. *Forkhead domain*
10. *AT-hooook domain*
11. *HSF-DNA domain*
12. *Histone H4 signature*
13. *MADS-box*
14. *Myb-DNA binding*
15. *Helix-loop-helix*
16. *Rel-homology domain*
17. *P-loop domain*
18. *Small acid-soluble proteins*
19. *T-box*
20. *TEA domain*
21. *Zinc ribbon*
22. *Tryptophan pentapeptide*
23. *Leucine Zipper*
24. *recF prot: signature*
25. *RUNT domain*
26. *Ethylene-responsive*
27. *'CxxCxxxHxxxC' retroviruses*
28. *HMG-box*
29. *Copper-fist*
30. *Dps protein signature*
31. *dnaA*
32. *p53 tumor*
33. *Histone-like*

Gene Regulation in yeast cells: the methanol switch

Methylotrophic yeasts

Four known genera: *Pichia*, *Hansenula*, *Candida* and *Torulopsis*

Can grow on methanol, as a sole carbon and energy source

Harbor highly efficient and regulated metabolic pathways for methanol, an otherwise toxic compound

The genes coding for methanol utilization pathway are under glucose repression and are induced to the maximum level by the cognate substrate i.e. methanol

**TRANSCRIPTIONAL REGULATION OF GENES INVOLVED IN
METHANOL UTILIZATION PATHWAY OF
*PICHA PASTORIS***

The promoter of the gene encoding alcohol oxidase (AOX), the first enzyme of methanol metabolism is widely used for the production of recombinant proteins

***AOX1* encodes alcohol oxidase,
the first enzyme of the methanol metabolism.**

		Glucose / Glycerol	Methanol
-937	<i>pAOX1</i>	-	+++++
	663 amino acids		

**HOW DOES METHANOL INDUCE AOX GENE
TRANSCRIPTION IN *P. pastoris*?**

A key transcription factor regulating *AOX1* as well as other genes of methanol utilization pathway was identified in the year 2006.

MOLECULAR AND CELLULAR BIOLOGY, Feb. 2006, p. 001

Vol. 26, No. 3

0270-7306/06/\$08.00+0 doi:10.1128/MCB.26.3.000-000.2006

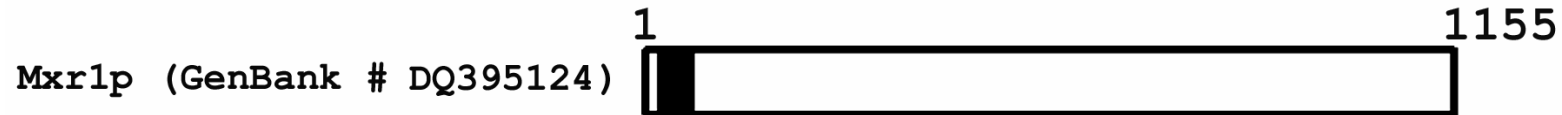
Copyright © 2006, American Society for Microbiology. All Rights Reserved.

Mxr1p, a Key Regulator of the Methanol Utilization Pathway and Peroxisomal Genes in *Pichia pastoris*

Geoffrey Paul Lin-Cereghino,^{1,2,†} Laurie Godfrey,¹ Bernard J. de la Cruz,² Sabrina Johnson,[†]
Samone Khuongsathiene,^{1,‡} Ilya Tolstorukov,² Mingda Yan,³ Joan Lin-Cereghino,^{1,2,†}
Marten Veenhuis,⁴ Suresh Subramani,³ and James M. Cregg^{1,2*}

MXR1

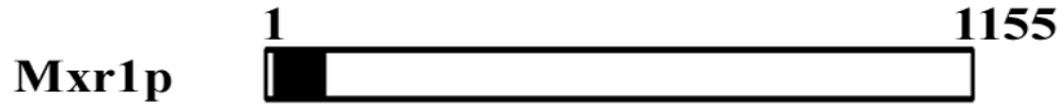
***MXR1* encodes a protein (Mxr1p) of 1155 amino acids with a zinc finger DNA-binding domain near its amino terminus that has homology to *Saccharomyces cerevisiae* Adr1p.**



Mxr1p is expressed constitutively at a low level.

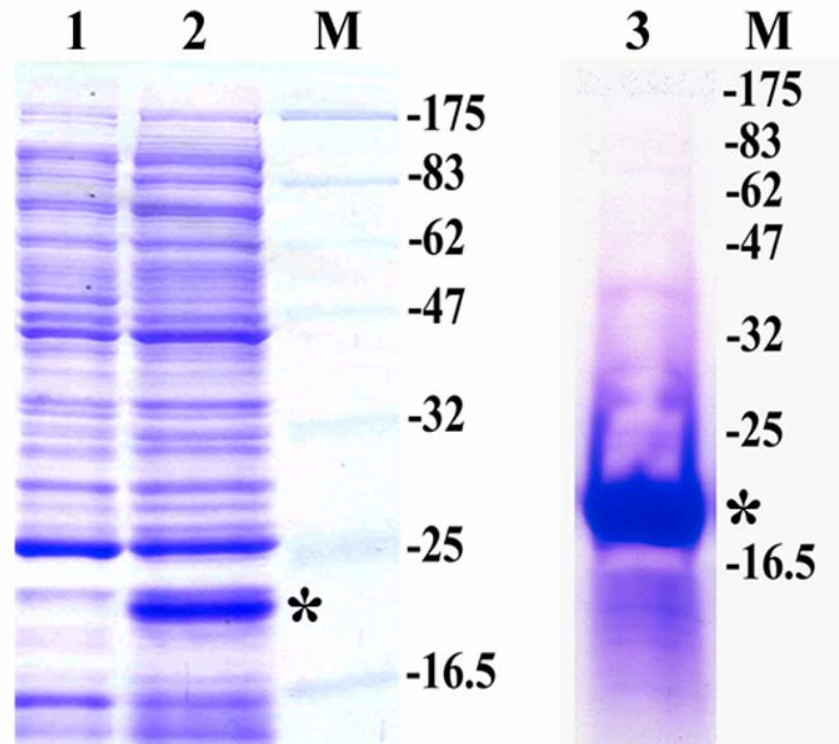
It is cytoplasmic in glucose grown cells but gets translocated to the nucleus in methanol and oleate grown cells.

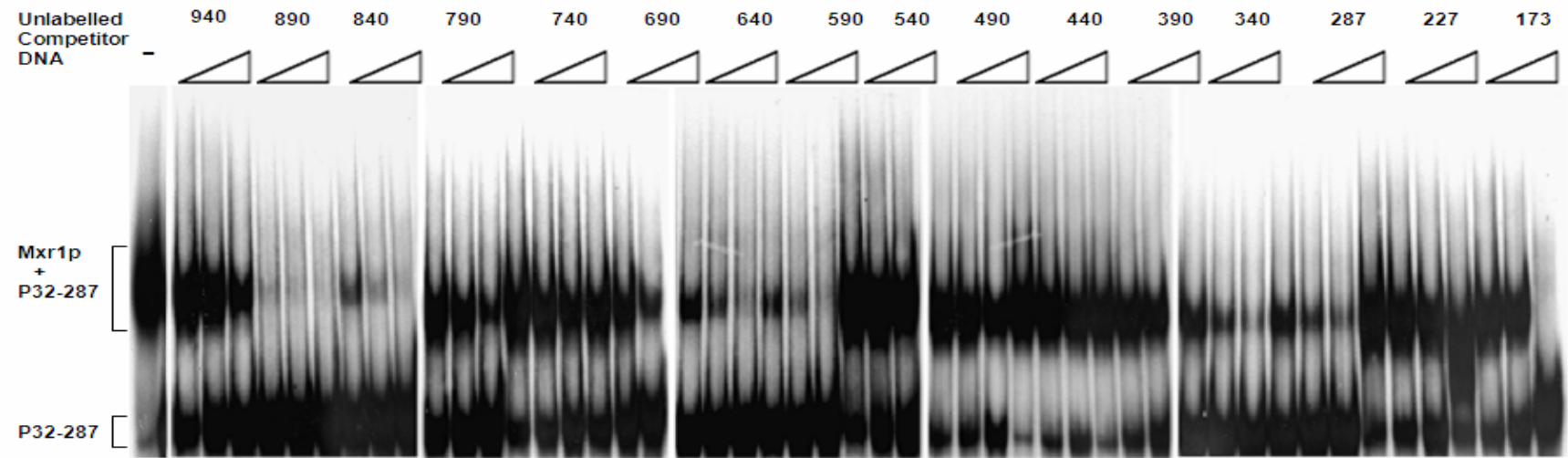
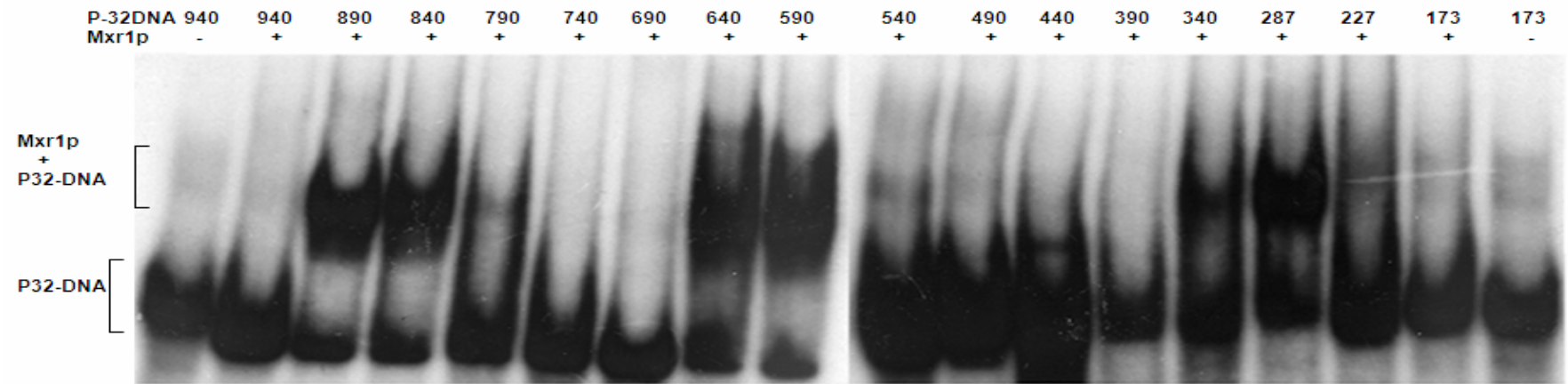
Mxr1p is a transcription factor that is necessary for the activation of many genes involved in methanol metabolism.



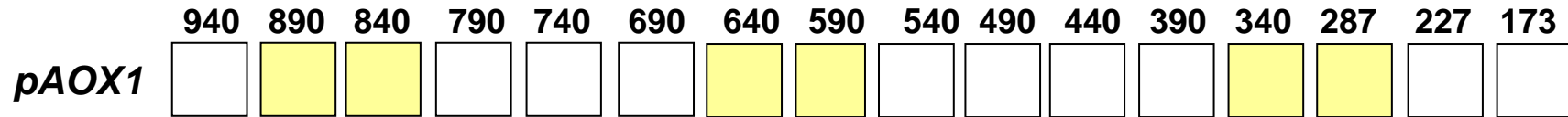
1	MSNLPPTFGS	TRQSPEDQSP	PVPKELSFNG	TTPSGKLRLF	VCQTCTRAFA	50
51	RQEHLKRHER	SHTKEKPFSC	GICSRKFSRR	DLLLRHAQKL	HSNCSDAAIT	100
101	RLRRKATRRS	SNAAGSISGS	TPVTTPNTMG	TPEDGEKRKV	QKLAGRRDSN	150

~17 kDa



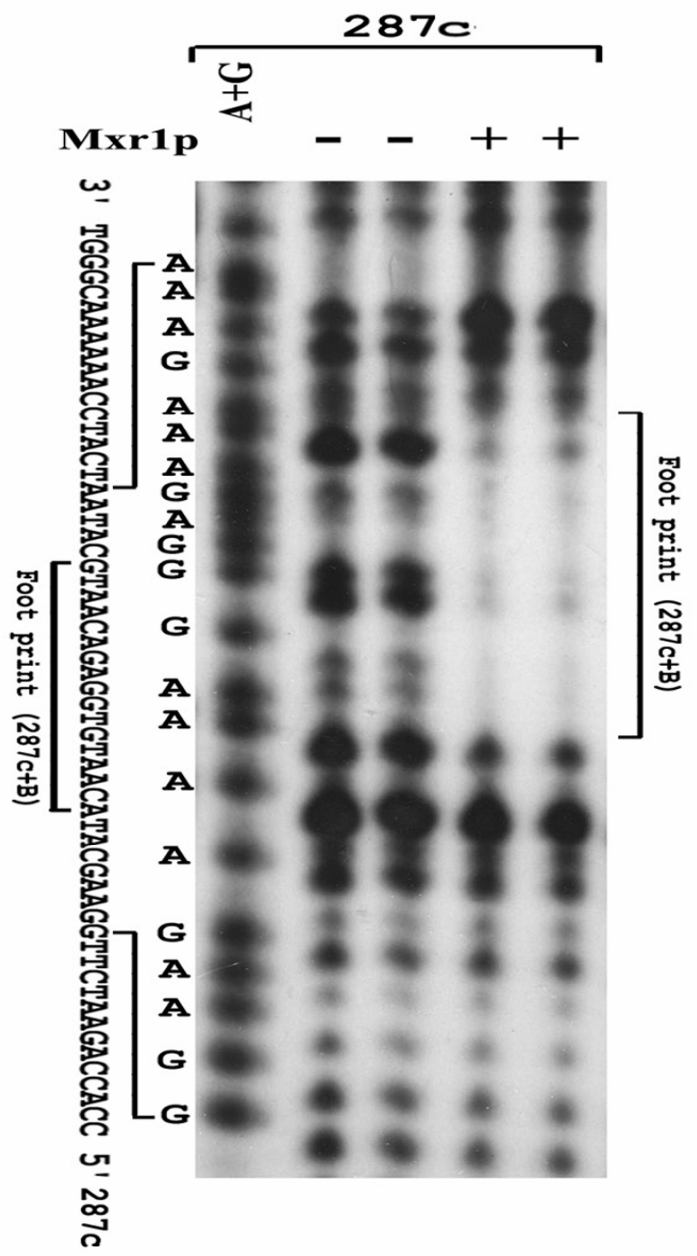
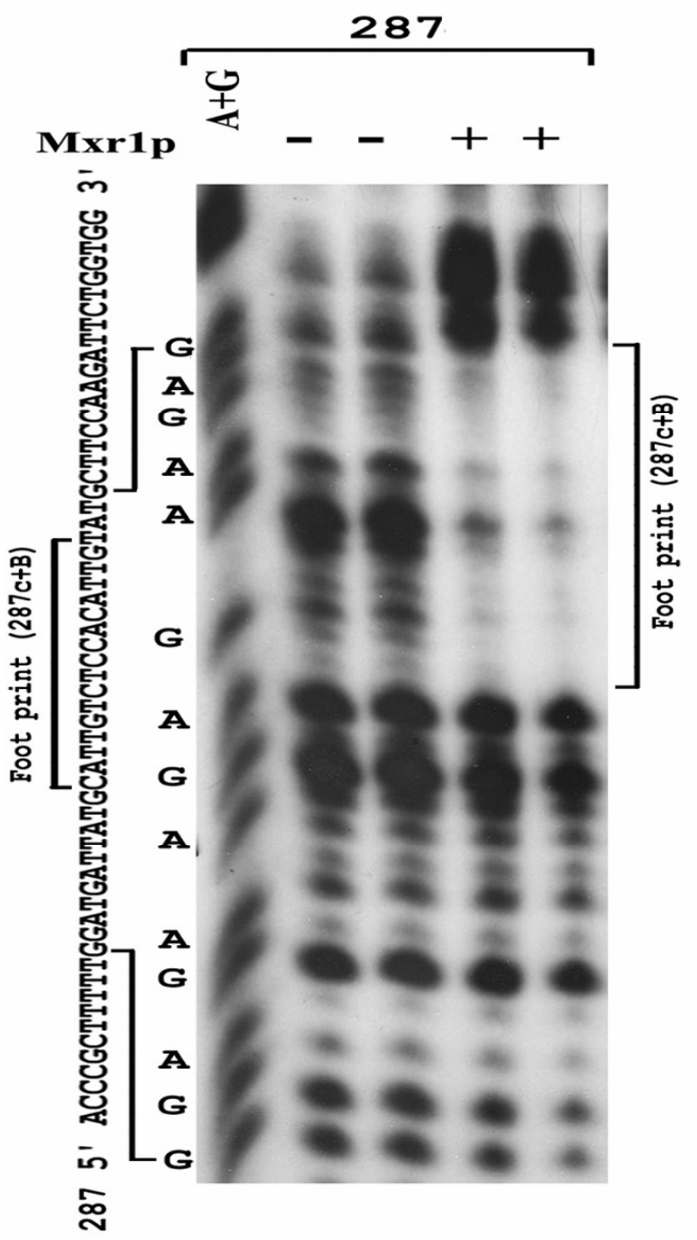
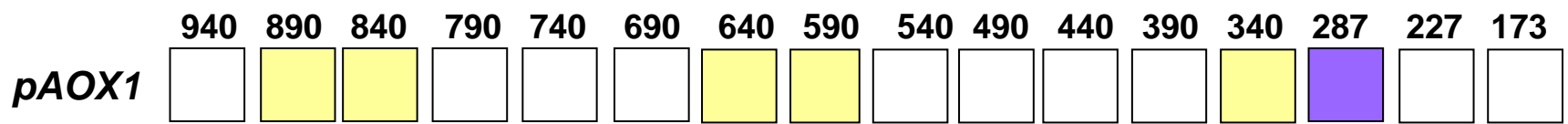


Mxr1p-DBD binding sites in the *AOX1* promoter



890 (-890 to -831)	5' ACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAGGAGGGGATACTAGC 3'
840 (-840 to -781)	5' ATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCTCTTCTCCTCAACACCC 3'
640 (-640 to -581)	5' TTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAGATGAGGGCTTTCTG 3'
590 (-590 to -531)	5' GGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCAAATGGCCCAAACACTGACAG 3'
340 (-340 to -281)	5' CTCTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAAACACCCGCT 3'
287 (-287 to -228)	5' ACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTATGCTTCCAAGATTCTGGTGG 3'

Mxr1p-DBD binds to 6 different regions in the *AOX1* promoter.



MXRE1: 5' ACAGGAGGGGATACACTAGC 3'
3' TGTCCTCCCCTATGTGATCG 5'

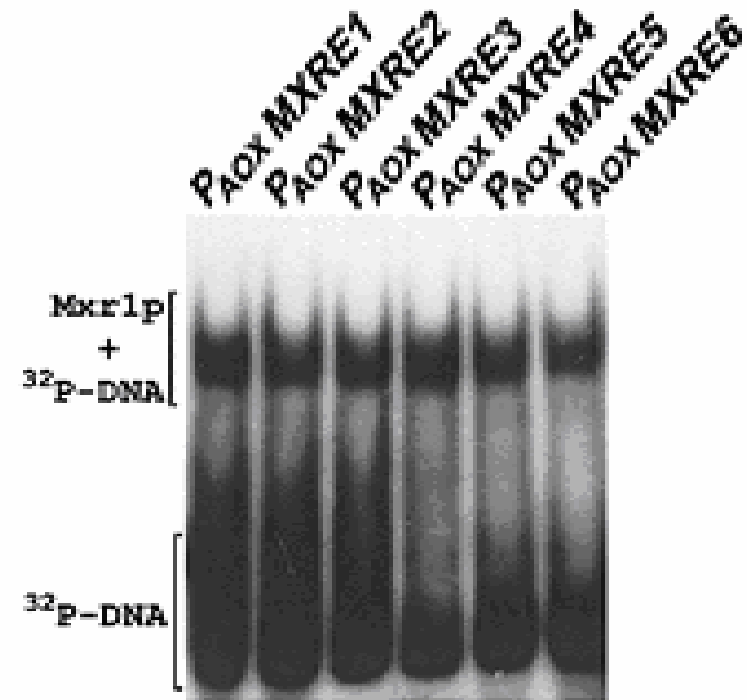
MXRE2: 5' TCCTCTTCTCCTCAACACCC 3'
3' AGGAGAAGAGGAGTTGTGGG 5'

MXRE3: 5' CAACAAGCTCCGCATTACAC 3'
3' GTTGTTGAGGCGTAATGTG 5'

MXRE4: 5' TGAGTGTGGGGTTCAAATAGT 3'
3' ACTCACACCCCAAGTTTATCA 5'

MXRE5: 5' CGCAAATGGGGAAACACCCG 3'
3' GCGTTTACCCCTTTGTGGGC 5'

MXRE6: 5' TGCATTGTCTCCACATTGTA 3'
3' ACGTAACAGAGGTGTAACAT 5'



Mxr1p binds to all the six P_{AOXI} Mxr1p response elements each of which is ~20bp in length

The sequence 5' CYCC/GGRG 3' is conserved in all the six P_{AOXI} Mxr1p response elements

5' CYCC 3'
3' GRGG 5'

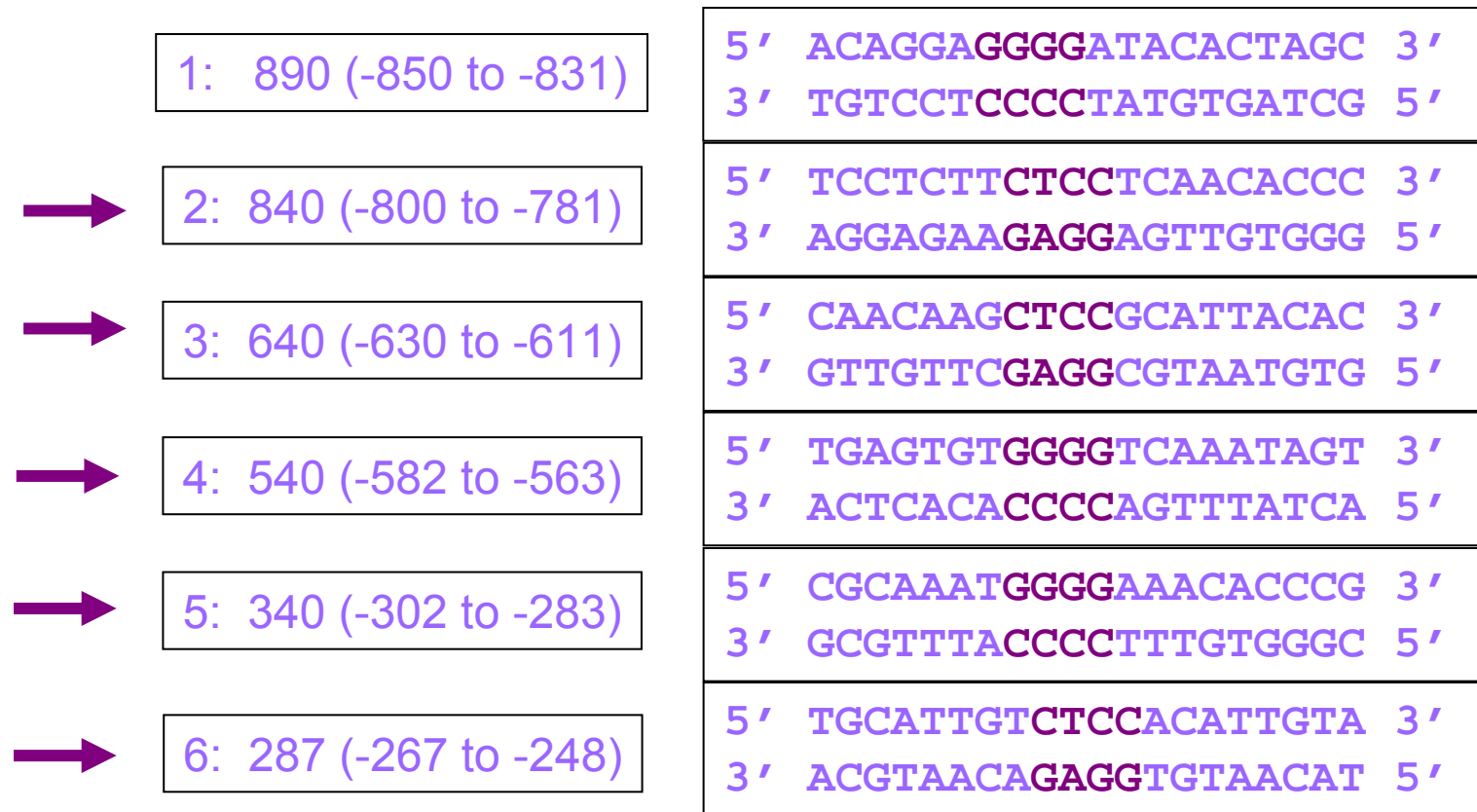
Y: C/T, R: G/A

DO THE MXR1P BINDING SITES IDENTIFIED IN OUR STUDY HAVE A FUNCTIONAL ROLE *IN VIVO*?

**Promoter library designed for fine-tuned gene expression in
*Pichia pastoris***

**Franz S. Hartner Claudia Ruth, David Langenegger, Sabrina
N. Johnson, Petr Hyka, Geoffrey P. Lin-Cereghino, Joan Lin-
Cereghino, Karin Kovar, James M. Cregg and Anton Glieder**

Nucleic Acids Res. 2008 Jul;36(12):e76. Epub 2008 Jun 6.



Deletion mutant (Hartner et al.,)	Relative AOX1 promoter activity	deleted region in AOX1 promoter	Mxr1p binding site identified in our study
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

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCGACATCCACAGGTCCATTCTC

Mxr1p BINDING SITE-1

ACACATAAGTGCCAAACGCAACAGGAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCC

Mxr1p BINDING SITE-2

ACTCCTCTTCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCTTGATTGGAGCT

dHSF1 (-40%)

CGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCATGACTTTATTAGCCTGTCTATCCTGGCCCCCT

Mxr1p BINDING SITE-3

GGCGAGGTTTCATGTTTGTATTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAGATG

Mxr1p BINDING SITE-4

d2 (-67%)

AGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCAAATGGCCCAAACACTGACAGTTTAAACGC

dAdr1 (-70%)

TGTCTTGGAACCTAATATGACAAAAGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGC

TAACGGCCAGTTGGTCAAAAAGAACTTCCAAAAGTCGCCATACCGTTTGTCTTGTTTGGTATTGATTGA

CGAATGCTCAAAAATAATCTCATTAAATGCTTAGCGCAGTCTCTCTATCGCTTCTGAACCCCGGTGCACCT

Mxr1p BINDING SITE-5

Mxr1p BINDING SITE-6

GTGCCGAAACGCAAATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGCTCCACATTGTATGCTTC

d5 (-60%)

dMat1-MC (-58%)

CAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTTCATGATCAAATTTAACTGTTCTAACCCCTACT

d6 (-55%)

TGACAGCAATATATAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTAC

TTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAAACGACTTTTAAACGACAACCTTGAGAAGA

TCAAAAACAACCTAATTATTCGAAACGATG

[Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris.](#)

Kranthi BV, Kumar R, Kumar NV, Rao DN, Rangarajan PN.

Biochim Biophys Acta. 2009 Jun-Aug;1789(6-8):460-8. Epub 2009 May 18.

[Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris.](#)

Kranthi BV, Vinod Kumar HR, Rangarajan PN.

Yeast. 2010 Mar 2. [Epub ahead of print]