

IDENTIFICATION OF *SCHIZOSACCHAROMYCES POMBE* GENES THAT ENCODE PUTATIVE HOMOLOGUES OF *SACCHAROMYCES CEREVISIAE* MEDIATOR COMPLEX SUBUNITS*

M. SIPICZKI

Department of Genetics and Molecular Biology, University of Debrecen,
Research Group for Microbial Developmental Genetics, Hungarian Academy of Sciences, Debrecen,
P.O.Box 56, H-4010 Debrecen, Hungary

The mediator complexes transduce regulatory information from upstream regulatory elements to the transcription machinery in organisms ranging from yeasts to humans. By a genome-wide search we identified 14 ORFs and genes in the genome of the fission yeast *Schizosaccharomyces pombe* that encode putative homologues of *Saccharomyces cerevisiae* mediator subunits. The *Sch. pombe* proteins are smaller and appear to form a mediator of lower complexity, which is consistent with the hypothesized ancient origin of fission yeasts.

Keywords: fission yeast, mediator, transcription, BLAST, phylogenesis

Introduction

All eukaryotes share a highly conserved mechanism of transcription regulation. The transcription initiation of protein-encoding genes requires interactions between cis-acting promoter elements and trans-acting factors [for a review, see 1]. The promoter consists of core elements, which include the TATA box and other DNA sequences that define transcription start sites and upstream regulatory elements, which either enhance or repress transcription in a gene-specific manner. The cis-acting factors include RNA polymerase II, general transcription factors, transcription activators and coactivators.

* This paper was written to commemorate to the fiftieth anniversary of the foundation of the Hungarian Society for Microbiology.

The general transcription factors assemble at the core promoter and include the TATA-binding protein (TBP), TFIIB, TFIIE, TFIIF and TFIIH. Transcription activators – which bind sequence-specific regulatory elements located upstream of the core promoter – stimulate the rate of transcription in response to physiological or developmental stimuli. A third class of transcription factors, known as coactivators, facilitates the interaction between the sequence-specific activators and the general RNA polymerase II machinery. They are not gene-specific and usually mediate activation by a broad spectrum of activators. Coactivators include the TAF (TBP-associated factor) component of the general transcription factor TFIID, the mediator complex(es) (e.g. SRB/MED) that associate(s) with RNA polymerase II, TFIIA, the histone acetyltransferase complexes (e.g. SAGA), and the chromatin remodelling complexes (e.g. SWI/SNF) [for recent reviews, see 1, 2, 3].

The mediator complex (SRB/MED) discovered in *Saccharomyces cerevisiae* interacts with RNA polymerase II, forming a "holoenzyme" complex. It binds tightly to the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II and can physically interact with at least some activators. The *S. cerevisiae* mediator complex includes the Srb proteins, the Med proteins and the proteins Cse2, Gal11, Nut2, Rgr1, Rox3 and Sin4 [4, 5, 6, 7]. The components of the complex assemble into functional modules that regulate the transcription of distinct groups of genes [8, 9].

Mediator complexes have also been identified in various metazoans [reviewed in 3], indicating that the function of mediator as a coactivator for gene-specific transcription regulation is evolutionary conserved. These complexes contain homologues of several, but not all, components of the mediator of *S. cerevisiae* and also contain proteins that do not have counterparts in the budding yeast. Very recently, ten mediator proteins were also identified in the fission yeast *Schizosaccharomyces pombe* [10, 11]. Since these yeasts are almost as different from each other as from animals [12], a detailed comparison of their mediators might shed more light on the evolution of transcription regulation in eukaryotes. Here we report on the identification of homologues of the *S. cerevisiae* mediator genes in the Sanger Center database that contains the recently completed genomic *Sch. pombe* sequence (http://www.sanger.ac.uk/Projects/S_pombe). We also present an initial analysis of the data, describing some of the insight that can be obtained from the sequences.

Methods

The *S. cerevisiae* mediator sequences were used as query sequences to conduct basic BLAST searches, using the blastp (basic local alignment search technique for protein sequences) search algorithm [13], to look for proteins (translated sequences)

that are similar in *Sch. pombe* in the non-redundant database at the Sanger Center, Cambridge. Peptide sequences were compared with entries in other nonredundant proteins and nucleotide databases with the use of the National Center for Biotechnology Information Advanced BLASTP programme [14]. For pairwise comparison of sequences we used the National Center for Biotechnology Information "Blast 2 sequences" programme [15]. The phylogenetic relationships among the proteins were investigated by the PROTPARS (Protein Parsimony) programme of the PHYLIP (Phylogeny Inference Package) package, version 3.573c [16].

Results and Discussion

The *S. cerevisiae* mediator complex is composed of the Srb family of proteins (Srb2, Srb4, Srb5, Srb6 and Srb7), the Med proteins (Med1, Med2, Med4, Med6, Med8, Med9, Med10 and Med11), Cse2, Gal11, Rgr1, Rox3 and Sin4 (Table I) [4, 5, 6, 7]. Recently, two independent lines of research demonstrated that a similar complex exists in *Sch. pombe*. One of the cell separation genes, *sep15⁺*, turned out to encode a protein with high degree of sequence similarity to Med8 [11]. Independently, the purification and partial characterisation of RNA polymerase II holoenzyme from *Sch. pombe* revealed a multiprotein complex that contains homologues to Rgr1, Srb4, Med4, Med7 and Nut2 proteins, representing each of the three subgroups of mediator proteins of *S. cerevisiae* [10]. It also contained four proteins denoted Pmc2, Pmc3, Pmc5 and Pmc6 with no discernible similarity to *S. cerevisiae* subunits. Its total mass was about 400 kDa, significantly less than 1000 kDa, the mass of the *S. cerevisiae* mediator. Interestingly, no homologues were found for 17 *S. cerevisiae* mediator components. Since the composition of the isolated complex is critically dependent on the methods applied to purification, the described proteins may represent only a subset of the subunits of the *Sch. pombe* mediator complex. Different purification strategies may lead to the identification of additional components.

An alternative approach to the identification of more components can be the search of the *Sch. pombe* sequence database for genes that encode amino acid sequences similar to those of the *S. cerevisiae* subunits. With the completion of the *Sch. pombe* genome project, the whole genome has become available for homology search. In order to identify putative homologues we performed a genome-wide Blastp search in the *Sch. pombe* genome database based at the Sanger Center (<http://www.sanger.ac.uk>). As shown in Table II, matching sequences were found for all but two *S. cerevisiae* proteins. For most of them more than one *Sch. pombe* sequence producing high-scoring segment pairs was found, but for further analysis we selected

only the most similar sequence from each group. We identified all genes previously described to encode homologues of *S. cerevisiae* mediator proteins such as Med7, Med8, Nut2, Pmc1 and Srb4. We also found fourteen uncharacterized ORFs whose putative products showed various degrees of similarity to additional *S. cerevisiae* mediator proteins. In principle, these might be the genes that encode the *Sch. pombe* counterparts. We were unable to identify sequences homologous to Med1 and Med2.

Table 1

Saccharomyces cerevisiae Mediator subunits¹

Subunit	Size (amino acid)	Gene
Cse2	149	<i>CSE2, MED9</i>
Gal11	1081	<i>GAL11, SPT13, SDS4, RAR3</i>
Med1	566	<i>MED1</i>
Med2	431	<i>MED2</i>
Med3	431	<i>PGD1, HRS1</i>
Med4	284	<i>MED4</i>
Med6	295	<i>MED6, MTR32</i>
Med7	222	<i>MED7</i>
Med8	223	<i>MED8</i>
Med11	131	<i>MED11</i>
Nut2	157	<i>NUT2, MED10</i>
Rgr1	1082	<i>RGR1</i>
Rox3	220	<i>ROX3, SSN7</i>
Sin4	974	<i>SIN4, SSN4, TSF3</i>
Srb2	210	<i>SRB2</i>
Srb4	687	<i>SRB4</i>
Srb5	307	<i>SRB5</i>
Srb6	121	<i>SRB6</i>
Srb7	140	<i>SRB7</i>
Srb8	1427	<i>SRB8, SSN5, ARE2</i>
Srb9	1420	<i>SRB9, SSN2, UME2</i>
Srb10	555	<i>SRB10, SSN3, UME5, ARE1</i>
Srb11	323	<i>SRB11, SSN8, UME3</i>

¹see references [2, 5, 6]

To verify the homologies revealed, each pair of sequences was subjected to a Blast 2 sequence comparison. This analysis did not confirm the homology between Cse2, Med3, Med11, Sin4, Srb2, Srb5, Srb9 of *S. cerevisiae* and the *Sch. pombe* sequences identified in the database search. Most probably, these proteins do not have counterparts in *Sch. pombe*. The rest of the proteins shared regions of homology with 24 to 53 % sequence identity. In certain proteins the similarity of amino acid sequences was confined to rather short regions. For example, Gal11 (1018 aa) and its putative

Sch. pombe homologue (653 aa) share only a very short (93 aa) homologous stretch (Fig. 1A). On the other hand, proteins like Med6, Med8 and Srb7 show similarity over the entire sequence (Fig. 1B, D and E). Altogether, 14 out of the 23 known *S. cerevisiae* mediator proteins seem to have homologues in *Sch. pombe*.

Table II

Schizosaccharomyces pombe Mediator subunits¹

<i>Sch. Pombe</i>			Homology with
Subunit	Size	Gene or ORF	<i>S. cerevisiae</i> subunits
Pmc4	239	SPBC1105.06	Med4
SpMed7	376	<i>med7</i>	Med7
	200	<i>sep15</i>	Med8
SpNut2	138	SPBC31F10	Nut2
Pmc1	879	<i>pmc1</i> (SPBC1A4.10C)	Rgr1
SpSrb4	545	<i>srb4</i>	Srb4
Pmc2		<i>pmc2</i> (SPAC2F7.04)	
Pmc3			
Pmc5			
Pmc6			

¹see references [10, 11]

Remarkably, 12 out of the 14 *Sch. pombe* proteins are smaller than their *S. cerevisiae* counterparts, which fits in fairly well with the earlier finding that the *Sch. pombe* mediator is smaller [10]. The smaller size of the *Sch. pombe* proteins might be attributed to the hypothesized ancient origin of the fission yeasts [12]. One can speculate that these proteins contain little more than the domains necessary for the basic activity, whereas the *S. cerevisiae* proteins have acquired longer sequence extensions required for specific tuning of the activity. With the exception of Med6, all *S. cerevisiae* proteins have these non-homologous extensions at their ends (examples are shown in Fig. 1). In Med6 the region of homology is split up in four parts by a shorter and two longer unique stretches that are absent in the *Sch. pombe* sequence (Fig. 1B). Med7 represents an exception to the rule that the *S. cerevisiae* proteins are longer. Its counterpart in *Sch. pombe* has a unique internal block and a unique terminal stretch that make the whole amino acid sequence longer than that of the *S. cerevisiae* Med7. It seems likely that the homologous regions separated by the internal block in *Sch. pombe* correspond to two specific protein domains.

Table III

Mediator homologs identified in the Sanger database of the *Sch. pombe* genome project

<i>Saccharomyces cerevisiae</i>		<i>Schizosaccharomyces pombe</i> homologs in Sanger database ¹			Sequence homology <i>Sch. pombe</i> / <i>S. cerevisiae</i>	
Subunit	Size (amino acid)	Gene or ORF	BLAST P value ²	Size (amino acid)	identity	positives
Cse2	149	SPAC10F6.09C	0.66	1234	no significant similarity	
Gal11	1081	SPBP35G2.15	9.8e-06	653	28/93 (30%)	46/93 (49%)
Med1	566	no hit				
Med2	431	no hit				
Med3	431	SPBC800.06	0.9992	305	no significant similarity	
Med4	284	SPBC1105.06	1.3e-06	239	34/125 (27%)	64/125 (51%)
Med6	295	SPAC1002.15c	1.7e-09	216	63/214 (29%)	99/214 (45%)
Med7	222	SPBC14F5.08	1.2e-18	376	82/314 (26%)	127/314 (40%)
Med8	223	<i>sep15</i>	2.1e-11	200	58/191 (30%)	100/191 (51%)
Med11	131	SPAC6G9.06C	0.0075	1208	no significant similarity	
Nut2	157	SPBC31F10	2.7e-13	138	40/150 (26%)	76/150 (50%)
Rgr1	1082	SPBP23A10.01C	1.2e-13	879	96/357 (26%)	158/357 (43%)
Rox3	220	SPCC1450.05C	0.0042	138	24/60 (40%)	37/60 (61%)
Sin4	974	SPBC1347.01C	0.9992	935	no significant similarity	
Srb2	210	<i>Lys1</i>	0.98	1415	no significant similarity	
Srb4	687	SPBC31F10.04C	1.4e-08	545	47/137 (34%)	78/137 (56%)
Srb5	307	SPAC1327.01C	0.11	977	No significant similarity	
Srb6	121	SPAC29A4.07	0.061	136	25/100 (25%)	44/100 (44%)
Srb7	140	SPBC1604.10	0.019	138	41/129 (31%)	67/129 (51%)
Srb8	1427	SPAC688.08	6.8e-09	1233	52/214 (24%)	91/214 (42%)
Srb9	1420	SPBC30D10.15	0.83	516	No significant similarity	
Srb10	555	SPAC23H4.17C	2.6e-92	352	160/299 (53%)	215/299 (71%)
Srb11	323	SPBC12D12.06	4.4e-17	228	63/163 (38%)	99/163 (60%)

¹<http://www.sanger.ac.uk>

²smallest sum probability

Fig. 1. Sequence comparison of *S. cerevisiae* mediator subunits and their putative *S. pombe* homologues. Comparison of Gal11Sp with Gal11Sc (A), Med6Sp with Med6Sc (B), Med7Sp with Med7Sc (C), Med8Sp with Med8Sc (D), Srb7Sp with Srb7Sc (E), Srb10Sp with Srb10Sc (F), and Srb11Sp with Srb11Sc (G). Symbols of amino acids are A: alanine, C: cysteine, D: aspartic acid, E: glutamic acid, F: phenylalanine, G: glycine, H: histidine, I: Isoleucine, L: leucine, K: lysine, M: methionine, N: asparagine, P: proline, Q: glutamine, R: arginine, S: serine, T: threonine, W: tryptophane, Y: tyrosine, V: valine

Fig. 1. (continued)

Fig. 1 (continued)

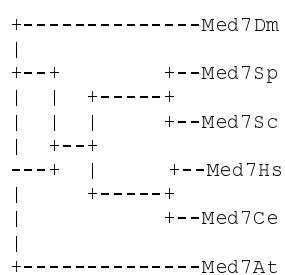


Fig. 2. Maximum parsimony tree for the *S. cerevisiae* Med7 protein (Med7Sc) and its homologues in *Arabidopsis thaliana* (Med7At), *Cenorhabditis elegans* (Med7Ce), *Drosophila melanogaster* (Med7Dm), *Homo sapiens* (Med7Hs) and *Sch. pombe* (Med7Sp)

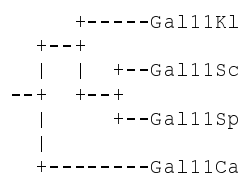


Fig. 3. A maximum parsimony tree inferred from the sequences of the *S. cerevisiae* Gal11 protein (Gal11Sc) and its homologues in *Candida albicans* (Gal11Ca), *Kuyveromyces lactis* (Gal11Kl) and *Sch. pombe* (Gal11Sp)

Recent studies indicate that mediator complexes are organized in a modular fashion [reviewed in 3]. In *S. cerevisiae* the core mediator subunits form two submodules, Cse2-Med1-Med2-Med4-Med7-Med8-Nut2-Rgr1-Srb7 and Med6-Rox3-Srb2-Srb4-Srb5-Srb6 [17], that interact with the modules composed of Gal11-Med3-Sin4 and Srb8-Srb9-Srb10-Srb11 [18]. Combined biochemical and genetic studies demonstrated that specific functions can be attributed to each subcomplex [19]. For example, *SRB10* and *SRB11* encode a kinase-cyclin pair involved in the phosphorylation of CTD in a cell cycle-dependent manner [20, 21]. We found putative *Sch. pombe* homologues of components of all these modules, indicating that the *Sch. pombe* mediator performs functions similar to those of the *S. cerevisiae* mediator. However, none of the modules are completely represented in *Sch. pombe*, which suggests that the two yeasts may also differ in the organisation and activity of their mediators. The Pmc proteins (Pmc2, Pmc3, Pmc5 and Pmc6), for which no homology was found in *S. cerevisiae* [10], might be substitutes for the missing homologues in these modules.

Numerous mammalian mediator-like complexes have been identified recently, including SMCC/TRAP, NAT, DRIP, ARC, murine Mediator and CRSP [for a review, see 3]. Their molecular composition has not been fully defined but they all appear to contain homologues of Med6, Med7, Rgr1, Srb10 and Srb11. *Cenorhabditis elegans* also has genes that code for homologues of Med6, Med7, Rgr1, Srb7 [22]. We found the counterparts for these genes in *Sch. pombe*, too, which indicates that these proteins may constitute the key mediator components that are evolutionarily conserved from yeasts to mammals. Remarkably, none of the mammalian and *C. elegans* complexes described to date contain readily apparent homologues of yeast Srb2, Srb4, Srb5 or Srb6. These proteins constitute a distinct submodul in *S. cerevisiae* and are critical for activator interactions [23]. *Sch. pombe* appears to have two members of this group, Srb2 and Srb5, demonstrating that it may possess a similar submodul. This finding further indicates that the *S. cerevisiae* mediator shares more features in common with the *Sch. pombe* mediator complex than with the corresponding metazoan complexes. We hypothesized that a closer relationship should also be reflected in a higher degree of sequence similarity between mediator proteins of these yeasts. To test this idea, we tried to find homologues for them in as many species as possible by searching various sequence data bases. Srb10 showed high degree of homology to a wide spectrum of cyclin dependent protein kinases of a great number of species. Many proteins with highly similar sequences were also found for Srb11, which were mostly cyclins. However, the limited information available about the exact functions of many of these kinases and cyclins did not allow the identification of the real homologues. Therefore we did not use them for multiple sequence comparison. Instead, we choose Med7, for

which single highly similar homologues were found in three Metazoans and one plant species. The most parsimonious tree shown in Fig. 2. indicates a close relationship between the yeast proteins.

A search of the National Center for Biotechnology Information data base revealed Gal11 homologues only in yeasts. In addition to *S. cerevisiae* and *Sch. pombe*, genes encoding Gal11-like proteins were also identified in *Candida albicans* and *Kluyveromyces lactis*. Previously, it was suggested that Gal11 is absent in *Sch. pombe* because the RNA polymerase II holoenzyme does not contain a protein with molecular mass similar to that of *S. cerevisiae* Gal11 [10]. Its absence seemed to correlate with the inability of *Sch. pombe* to grow on galactose. In *S. cerevisiae*, the Gal11 protein is needed to support Gal4-dependent induction of genes involved in galactose metabolism [16]. Interestingly, we could find an ORF encoding an amino acid sequence that shares high degree of sequence similarity with Gal11 over a 93-amino acid region (Fig. 1A). This hypothetical protein is significantly shorter (653 aa) than Gal11 (1018 aa), which might account for the failure of its identification in SDS-PAGE analysis of the purified polymerase II holoenzyme [10]. In spite of the considerable size difference, it appears to be more closely related to Gal11 than their homologues in the other yeast species.

In this analysis we identified numerous *Sch. pombe* genes that code for putative homologues of *S. cerevisiae* mediator subunits. The conclusion one may draw from these results is that the components of the putative *Sch. pombe* mediator complex show a higher degree of similarity with those of *S. cerevisiae* mediator than with the components of the metazoan mediators. However, the revealed sequence homologies do not necessarily coincide with functional homologies, and certain similarities may be only accidental. Future experimental studies will test which of the identified genes code for *Sch. pombe* mediator subunits and how these proteins work together to regulate transcription.

Acknowledgements. I thank Ida Miklós and Zsolt Szilágyi for comments on the manuscript. This work was supported by grants provided by OTKA and FKFP.

References

1. Hampsey, M.: Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62**, 465 (1998).
2. Wu, W.-H., Hampsey, M.: Common cofactors and cooperative recruitment. *Current Biol* **9**, R606 (1999)
3. Malik, S., Roeder, R.G.: Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem Sci* **25**, 277 (2000).

4. Hampsey, M., Reinberg, D.: RNA polymerase II as a control panel for multiple coactivator complexes. *Curr Opin Genet Dev* **9**, 132 (1999).
5. Gustafsson, C.M., Myers, L.C., Beve, J., Spahr, H., Lui, M., Erdjument-Bromage, H., Tempts, P., Kornberg, D.: Identification of new mediator subunits in the RNA polymerase II holoenzyme from *Saccharomyces cerevisiae*. *J Biol Chem* **273**, 30851 (1998).
6. Han, S.J., Lee, Y.C., Gim, B.S., Ryu, G.-H., Park, S.J., Lane, W.S., Kim, Y.J.: Activator-specific requirement of yeast mediator proteins for RNA polymerase II transcriptional activation. *Mol Cell Biol* **19**, 979 (1999).
7. Bjorklund, S., Almouzi, G., Nightingale, K.P., Weiss, K.: Global transcription regulators of eukaryotes. *Cell* **96**, 759 (1999).
8. Myers, L.C., Gustafsson, C.M., Hayashibara, K.C., Brown, P.O., Kornberg, R.D.: Mediator protein mutations that selectively abolish activated transcription. *Proc Natl Acad Sci USA* **96**, 67 (1999).
9. Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., Young, R.A.: Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717 (1998).
10. Spahr, H., Beve, J., Larsson, T., Bergstrom, J., Karlsson, K.-A., Gustafsson, C.M.: Purification and characterization of RNA polymerase II holoenzyme from *Schizosaccharomyces pombe*. *J Biol Chem* **275**, 1351 (2000).
11. Zilahi, E., Miklós, I., Sipiczki, M.: The *Schizosaccharomyces pombe sep15⁺* gene encodes a protein homologous to the Med8 subunit of the *Saccharomyces cerevisiae* transcriptional mediator complex. *Curr Genet* **38**, 227 (2000).
12. Sipiczki, M.: Where does fission yeast sit on the tree of life? *Genome Biology* **1**:1011.1 (2000).
13. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: Basic local alignment search tool. *J Mol Biol* **215**, 403 (1990).
14. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389 (1997).
15. Tatusova, T.A., Madden, T.L.: Blast 2 sequences – a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**, 247 (1999).
16. Felsenstein, J.: PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.
17. Lee, Y.C., Kim, Y.J.: Requirement for a functional interaction between mediator components Med6 and Srb4 in RNA polymerase II transcription. *Mol Cell Biol* **18**, 5364 (1998).
18. Myer, V.E., Young, R.A.: RNA polymerase II holoenzymes and subcomplexes. *J Biol Chem* **273**, 27757 (1998).
19. Lee, Y.C., Park, J.M., Min, S., Han, S.J., Kim, Y.J.: An activator binding module of yeast RNA polymerase II holoenzyme. *Mol Cell Biol* **19**, 2967 (1999).
20. Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., Young, R.A.: Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell* **2**, 43 (1998).
21. Liao, S.M., Zhang, J.H., Jeffrey, D.A., Koleske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., Vanvuuren, H.J.J., Young, R.A.: A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* **374**, 193 (1995).

22. Kwon,J.Y., Park,J.M., Gim,B.S., Han,S.J., Lee,J., Kim,Y-J.: *Caenorhabditis elegans* Mediator complexes are required for developmental-specific transcriptional activation. Proc Natl Acad Sci USA **96**, 14990 (1999).
23. Koh,S.S., Ansari,A.Z., Ptashne,M., Young,R.A.: An activator target in the RNA polymerase II holoenzyme. Mol Cell **1**, 895 (1998).