Regulation of Gene Expression by a Metabolic Enzyme

David A. Hall,[1] Heng Zhu,[2*] Xiaowei Zhu,[3] Thomas Royce,[1] Mark Gerstein,[1] Michael Snyder[1,2†]

Gene expression in eukaryotes is normally believed to be controlled by transcriptional regulators that activate genes encoding structural proteins and enzymes. To identify previously unrecognized DNA binding activities, a yeast proteome microarray was screened with DNA probes; Arg5,6, a well-characterized mitochondrial enzyme involved in arginine biosynthesis, was identified. Chromatin immunoprecipitation experiments revealed that Arg5,6 is associated with specific nuclear and mitochondrial loci in vivo, and Arg5,6 binds to specific fragments in vitro. Deletion of Arg5,6 causes altered transcript levels of both nuclear and mitochondrial target genes. These results indicate that metabolic enzymes can directly regulate eukaryotic gene expression.

Although gene products with enzymatic and structural functions have been known to regulate gene expression through indirect mechanisms (*1*, *2*), whether enzymatic proteins can directly control gene expression has not been extensively investigated. Butow and co-workers found that a protein involved in isoleucine synthesis, Ilv5, is associated with mitochondrial DNA in vivo (*3*, *4*). However, it is not known whether this enzyme was directly associated with DNA, bound specific sequences, or had a direct role in affecting the expression of specific genes. Zheng *et al*. (*5*) reported that glyceraldehyde-3-phosphate dehydrogenase (GAPDH) serves as a coactivator to regulate the expression of histone H2B. GAPDH does not directly associate with DNA and may have an indirect role; moreover, its ability to regulate other loci and thus serve as a global regulator was not investigated.

To further understand mechanisms of gene regulation, we used proteome arrays and chromatin immunoprecipitation (ChIP/chip) to identify previously unrecognized DNA binding activities (Fig. 1) (*6*). Protein microarrays containing nearly all of the proteins of yeast were spotted in duplicate onto microscope slides (*7*) and probed with either single-stranded or double-stranded yeast genomic DNA labeled with Cy3; 6 and 10 arrays were probed with single-stranded DNA and double-stranded DNA probes, respectively. Of the >200 total identified DNA binding proteins that reproducibly bound double- or single-stranded DNA, 84 proteins bound single-stranded DNA, 58 bound double-stranded DNA, and 131 bound both (see table S1 for a detailed list).

Many (~50%) of the proteins identified are expected to bind DNA based on their known functions (e.g., transcription factors). However, a large number were either not known or would not be suspected to have DNA binding activity based on available information. These latter proteins could (i) represent previously unrecognized DNA binding proteins, (ii) be associated with DNA binding proteins that copurified with the protein present on the proteome array, or (iii) bind DNA nonspecifically in vitro but not in vivo, and thus be artifacts.

We tested whether eight proteins (Arg5,6, Dig2, Mtw1, Yrb2, Akl1, Yer152c, Lrg1, and Rub1), which have not previously been reported to bind DNA, were specifically associated with chromosomal DNA in vivo by immunoprecipitation of the protein and probing of genomic DNA microarrays [ChIP/chip (*8*)]. Each candidate DNA binding protein was first tagged at its C terminus with 13 copies of c-Myc epitopes (13Xmyc) by insertion of the epitope coding sequences at the endogenous locus. Immunoblot analysis revealed that each protein migrated at its expected size. The exception was Arg5,6, which is normally cleaved into two peptides, an N-terminal N-acetyl-gamma-glutamyl phosphate reductase (Arg6) and C-terminal acetylglutamate kinase (Arg5); the tagged protein migrated at ~70 kD, the size expected of Arg5::13Xmyc (Fig. 2A). The cells were treated with formaldehyde to cross-link protein and DNA, and the cells lysed and sonicated. The epitope-tagged protein was immunoprecipitated and the associated DNA purified and labeled with Cy5. As a control, DNA from an untagged strain was prepared in parallel and labeled with Cy3. The two probes were mixed and used to probe a yeast genomic DNA array containing the intergenic regions and the entire mitochondrial genome. To ensure lack of bias, experiments were performed blindly with respect to the particular protein analyzed. Immunoprecipitates of five epitope-tagged yeast proteins (Yrb2, Akl1, Yer152c, Lrg1, and Rub1) did not show enrichment of any chromosomal loci over the untagged strains. These proteins either do not associate with DNA in cells grown in rich medium or they associate with DNA nonspecifically; they were not pursued further.

Three proteins—Mtw1, Dig2, and Arg5,6—were found to be associated with specific DNA regions in vivo. Mtw1, a kinetochore protein immunoprecipitated nearly all of the yeast centromeres (*9*), and Dig2, a negative regulator of the Ste12 transcription factor (*10*) bound a number of loci. These proteins either bind DNA directly or are associated with a DNA binding component and will be pursued as part of a separate study. In four independent ChIP/chip experiments, Arg5 was found to be reproducibly associated with a number of DNA fragments [with the use of a high stringency cut-off ($P < 0.001$) (Fig. 2B) (*6*)]. Arg5,6 encodes two mitochon-

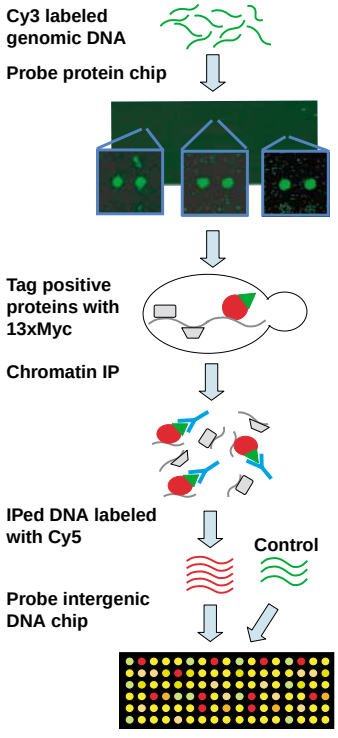


Fig. 1. A microarray containing 6500 protein preparations of 5800 different yeast proteins spotted in duplicate (*7*) was probed with genomic yeast double-stranded DNA labeled with Cy3. Positives were identified, and eight were tested for association with specific chromosomal and mitochondrial regions with the use of ChIP/chip.

[1]Department of Molecular Biophysics and Biochemistry, [2]Department of Molecular, Cellular, and Developmental Biology, [3]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520–8005, USA.

*Present address: Department of Pharmacology and the High Throughput Biology Center, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

†To whom the correspondence should be addressed. E-mail: michael.snyder@yale.edu

drial enzymes, Arg6 and Arg5, which mediate two key steps in the biosynthesis of ornithine, which is a precursor to arginine. Neither of the Arg5 or Arg6 gene products was known previously to regulate gene expression, but nitrogen metabolism is known to be closely linked with mitochondrial function and ornithine synthesis (*11*, *12*).

Arg5 bound 22 fragments which contained primarily mitochondrial loci, including the 15*S* ribosomal DNA (rDNA), *COX3*, three regions of *COB1*, and 10 regions within *COX1* (*13*–*15*); the latter two genes contain introns (table S2 and Fig. 2C). Arg5 is also associated with several nuclear loci including regions adjacent to *PUF4*, *PHO23*, and *THI13*, indicating that this protein likely resides in both cellular compartments.

We next performed ChIP experiments with tagged and untagged strains and tested for enrichment of specific loci in the tagged strain by standard polymerase chain reaction (PCR) assays (*8*). Fragments were specifically enriched for many nuclear and mitochondrial loci immunoprecipitated from the tagged strain relative to those from the untagged strain; these include four *COX1* regions, *COB1*, *COX3*, *YOR352w*, *PHO23*, *THI13*, and *PUF4* (Fig. 2C). Control fragments (*SUN4*, *CDC7*, and *CHS5*) were not enriched in the Arg5-tagged immunoprecipitates. Thus, Arg5 is associated with mitochondrial and nuclear DNA in vivo.

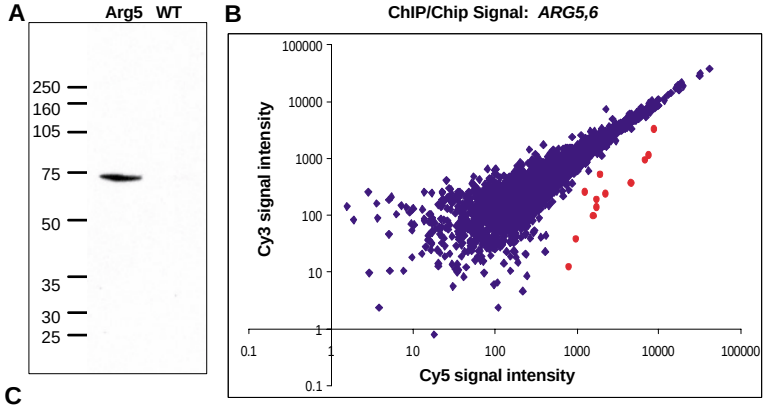
To determine whether Arg5,6 can bind DNA directly, Arg5,6 was overproduced in yeast as a glutathione *S*-transferase (GST) fusion, purified with the use of stringent conditions (in the presence of 0.5 M NaCl) and tested for binding to two *COX1* targets, a fragment at the 5′ end *COX1* and another in the first exon, with the use of in vitro "gel-shift" assays. Increasing amounts of Arg5,6 were mixed with the *COX1* targets or negative control DNAs [Epstein-Barr virus nuclear antigen (EBNA) or three mitochondrial fragments that do not bind Arg5,6] and separated in a native polyacrylamide gel. As an additional control, GST alone was added to the *COX1* DNA. As shown in Fig. 3A, *COX1* DNA forms a slower migrating complex in the presence of Arg5,6, but not in the presence of GST. The negative control DNAs (EBNA and other mitochondria segments) do not form a complex (Fig. 3C) (*16*). Thus, Arg5,6 can specifically associate with *COX1* DNA in vitro.

To determine the site bound by Arg5,6, the sequences of the in vivo targets were compared and a strong common motif was identified ($P < 10^{-5}$; Fig. 3B) (*17*). Incubation of the Arg5,6 protein with a labeled double-stranded oligonucleotide containing this sequence from the *COX1* region resulted in the formation of a complex. The complex was not formed with a randomized version of this sequence or in the presence of unlabeled competitor oligonucleotide. Thus, Arg5,6 associates with a specific DNA sequence. This sequence is 78% GC, unlike mitochondrial DNA, which is 83% AT. Whether Arg5 or Arg6 binds this sequence in vitro is not clear; however, because Arg5 associates with specific loci in vivo, it is the most probable candidate for this activity.

Mutant analysis revealed no morphological or growth defects in *arg5,6*Δ cells grown in rich medium, nitrogen-limiting medium, and medium containing glycerol as the primary carbon source. The distribution and intensities of mitochondrial DNA staining in cells incubated in the same conditions were also identical in wild-type and *arg5,6*Δ cells.

To determine whether Arg5,6 has a role in regulating gene expression, we used real-time PCR (RT-PCR) to quantify the level of *COX1*, *COB1*, and *COX3* mRNAs and the mRNA levels of nuclear targets *PUF4*, *YOR352w*, and *YHL045w* in wild-type and *arg5,6*Δ strains. Cells were examined in rich medium and in medium limited for nitrogen or amino acids; in the absence of amino acids, the levels of many mitochondrial enzymes are induced (*18*). *arg5,6*Δ strains grown in medium lacking amino acids exhibited significantly lower levels (3.5- to 6-fold) of the *COX1* mitochondrial and the *YOR352W* nuclear transcripts as compared with those of wild-type cells; 2.2-fold lower *PUF4* mRNA levels were observed in the same conditions (Fig. 4). Similar but weaker effects were observed when cells were grown in medium limited for nitrogen. Only a modest effect was observed for different messages when cells are grown in rich medium. The mRNA levels of three genes—*ACT1*, *COX2*, and 21*S* ribosomal RNA (the latter two are mitochondrial)—that are not targets of Arg5,6 are not affected. Thus, Arg5,6 is important for controlling the levels of specific mitochondrial and nuclear transcripts in cells lacking amino acids or limited for nitrogen; whether Arg5 or Arg6 is responsible for this activity is not resolved by these experiments.

The presence of Arg5,6 binding sites in the intronic regions of *COX1* and *COB1* raises the possibility that Arg5,6 might also play a role in RNA processing. We therefore designed real-time RT-PCR probes to monitor the levels of unprocessed (i.e., intronic) *COX1* and *COB1* messages in *arg5,6*Δ cells grown in the media described above. As shown in Fig. 4, unprocessed transcript levels for *COX1* are slightly affected by the presence of Arg5,6, whereas those for *COB1* are significantly affected. Thus, these data indicate that Arg5,6 affects the levels of unprocessed RNA for *COB1*, as expected for a transcriptional regulator; processed *COX1* levels are more significantly affected than unprocessed messages, raising the possibility that Arg5,6 might also directly or indirectly participate in *COX1* RNA processing. Because Arg5,6 regulates the *COB1* locus, which encodes a maturase in its precursor RNA that processes *COX1*



C

Fragment	15S rRNA2	PC1 COB	PC2 COX1	PC4 COX1	PC6 COX1	PC8 COX1	PC2 YOR352W	PC2 PHO23	PC1 THI13	PC YHL045W	PC PUF4	Negative Control
Fold Enrichment	10.1±8.6	3.9±2.0	22.0±7.9	3.7±2.2	13.8±12.0	4.5±0.2	8.5±3.6	8.4±4.1	>30	2.0±0.4	3.8±2.3	1.0±0.1

Fig. 2. (**A**) Immunoblot analysis of Arg5. Proteins were prepared from untagged wild-type cells and an Arg5,6 13Xmyc strain containing 13 copies of Myc coding sequences at the C terminus. Immunoblots were prepared and probed with antibodies to Myc. Left lane (+): the Arg5::13Xmyc strain. Right lane (–): the untagged strain. WT, wild type. Migration of molecular weight markers in kD is indicated on the left. (**B**) Identification of loci bound by Arg5. Arg5::13Xmyc-bound DNA was purified by immunoprecipitation and labeled with Cy5; control samples purified from an untagged strain were labeled with Cy3. The DNA samples were mixed and used to probe a DNA microarray containing the intergenic regions of yeast. Fragments exhibiting a strong enrichment ($P < 0.001$) in the tagged strain relative to that of the untagged strains in four separate experiments were identified with the use of ExpressYourself (*8*). The plot of Cy5/Cy3 ratios for each fragment is present; significantly enriched fragments are indicated in red. (**C**) Table summarizing the PCR confirmation of 11 regions that bind Arg5. Fold-enrichments of three separate experiments are presented. Enrichment was observed for *COX1*, *COB1*, *YOR352w*, *PHO23*, *THI3*, *PUF4*, and *YH045w* loci but not control loci. Fold enrichment data are shown as mean ± standard error.

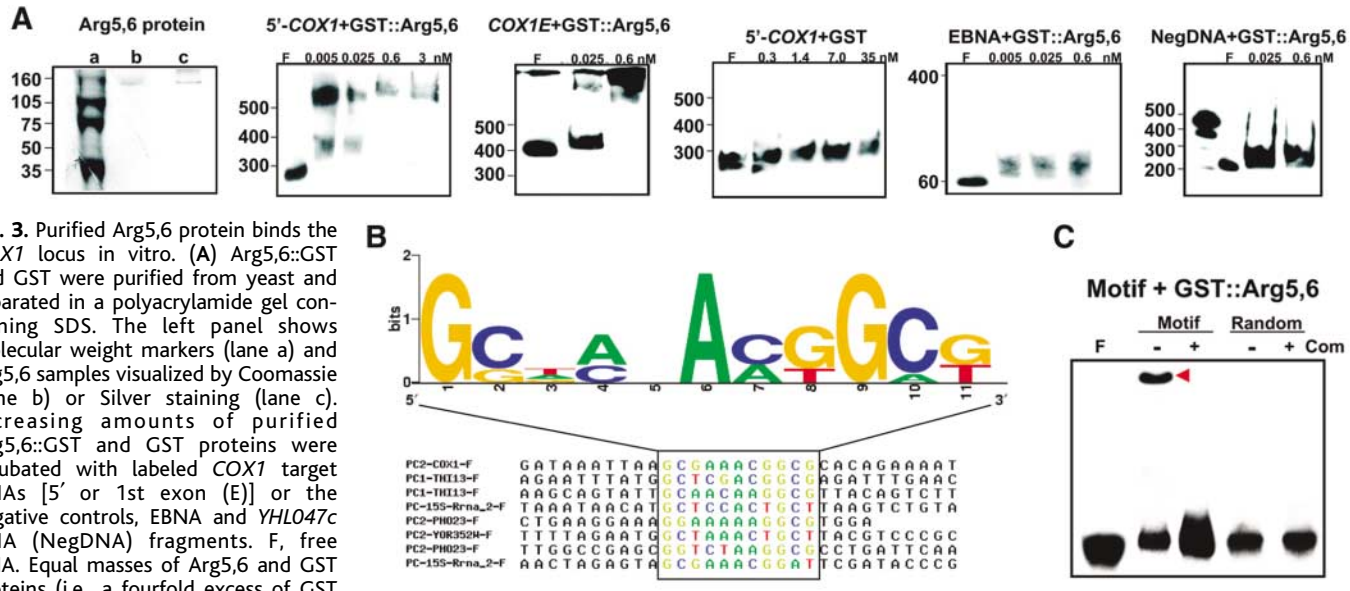


Fig. 3. Purified Arg5,6 protein binds the *COX1* locus in vitro. (**A**) Arg5,6::GST and GST were purified from yeast and separated in a polyacrylamide gel containing SDS. The left panel shows molecular weight markers (lane a) and Arg5,6 samples visualized by Coomassie (lane b) or Silver staining (lane c). Increasing amounts of purified Arg5,6::GST and GST proteins were incubated with labeled *COX1* target DNAs [5′ or 1st exon (E)] or the negative controls, EBNA and *YHL047c* DNA (NegDNA) fragments. F, free DNA. Equal masses of Arg5,6 and GST proteins (i.e., a fourfold excess of GST molecules) were tested at the indicated concentrations. Arg5,6 fusion proteins specifically bound to *COX1* DNAs, whereas GST did not. Arg5,6 did not bind to the EBNA DNA or other control DNA fragments (*16*). (**B**) Consensus motif derived from targets confirmed by PCR. (**C**) Complex formation of Arg5,6 in the presence of the labeled *COX1* double-stranded oligonucleotide containing three copies of the motif (lane 2), but not in the presence of 200-fold excess of unlabeled competitor double-stranded oligonucleotide or in the presence of a randomized *COX1* motif double-stranded oligonucleotide. Com, unlabeled competitor DNA.

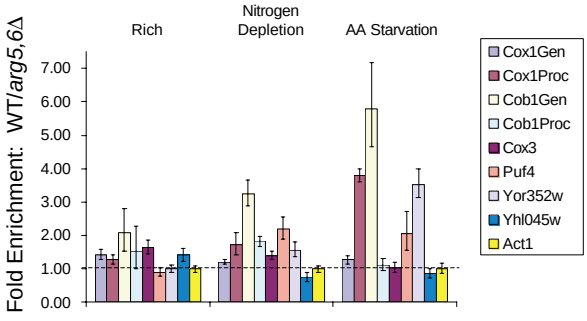


Fig. 4. Reduced levels of specific mitochondrial and nuclear transcripts in cells lacking Arg5,6. Wild-type and *arg5,6*Δ cells were grown in either rich medium or media limited for nitrogen or lacking amino acids (AA). The transcripts levels of several Arg5,6 targets were quantified with the use of real-time RT-PCR. The transcript ratios were normalized to actin and relative transcript levels in wild-type relative to *arg5,6*Δ cells are presented.

RNA (*19*), the reduced level of *COB1* is expected to decrease *COX1* RNA processing.

The association of Arg5 within the *COX1*, *COB1*, *COX3*, and 15*S* rDNA genes, rather than just at the initiation sites, raises the possibility that Arg5,6 plays a role in either transcription elongation or RNA processing. However, it is likely that Arg5,6 also affects transcription because many of its target loci lack introns (e.g., *COX3*, *PUF4*, *PHO23*, and *THI13*). Thus, Arg5,6 may have a role in the regulation of both transcriptional and post-transcriptional processes. Coupling of transcription and RNA processing has been reported in eukaryotes previously (*20*, *21*).

Although the presence of Arg5 in the mitochondria has been well described (*22*), this protein has not been previously found in the nucleus by ChIP. We presume that the nuclear levels of Arg5 are low because we could not detect Arg5::13Xmyc by indirect immunofluorescence.

Nitrogen metabolism has been linked to mitochondrial function in yeast (*11*, *12*). Discovery of a protein involved in ornithine biosynthesis, which is linked to nitrogen metabolism, which regulates the activity of mitochondrial gene expression, provides a direct link for how this might occur.

Several examples of multifunctional proteins have been reported previously. Butow and co-workers identified proteins associated with mitochondrial DNA and found tricarboxylic acid cycle proteins and several involved in amino acid biosynthesis (*23*). These proteins were demonstrated to be important for mitochondrial stability. It is plausible that many proteins have multiple roles in vivo and unbiased proteomic approaches such as those used here (proteome chips and chIP/chip) will be required to fully identify the plethora of functions of eukaryotic proteins.

References and Notes

1. B. M. Turner, *Nature Cell Biol.* **5**, 390 (2003).
2. D. L. Black, *Annu. Rev. Biochem.* **72**, 291 (2003).
3. J. M. Bateman, P. S. Perlman, R. A. Butow, *Genetics* **161**, 1043 (2002).
4. J. M. Bateman, M. Iacovino, P. S. Perlman, R. A. Butow, *J. Biol. Chem.* **277**, 47946 (2002).
5. L. Zheng, R. G. Roeder, Y. Luo, *Cell* **114**, 255 (2003).
6. Materials and methods are available as supporting material on *Science* Online.
7. H. Zhu *et al.*, *Science* **293**, 2101 (2001).
8. C. E. Horak *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2924 (2002).
9. B. A. Pinsky, S. Y. Tatsutani, K. A. Collins, S. Biggins, *Dev. Cell* **5**, 735 (2003).
10. A. B. Kusari, D. M. Molina, W. Sabbagh, Jr., C. S. Lau, L. Bardwell, *J. Cell Biol.* **164**, 267 (2004).
11. A. Abadjieva, K. Pauwels, P. Hilven, M. Crabeel, *J. Biol. Chem.* **276**, 42869 (2001).
12. J. M. Guillamon, N. A. van Riel, M. L. Giuseppin, C. T. Verrips, *FEMS Yeast Res.* **1**, 169 (2001).
13. S. Naithani, S. A. Saracco, C. A. Butler, T. D. Fox, *Mol. Biol. Cell* **14**, 324 (2003).
14. F. H. MacIver, I. W. Dawes, C. M. Grant, *Curr. Genet.* **31**, 119 (1997).
15. B. J. Hicke, E. L. Christian, M. Yarus, *EMBO J.* **8**, 3843 (1989).
16. D. A. Hall *et al.*, data not shown.
17. T. L. Bailey, M. Gribskov, *J. Comput. Biol.* **5**, 211 (1998).
18. K. Natarajan *et al.*, *Mol. Cell. Biol.* **21**, 4347 (2001).
19. W. M. Schmidt, R. J. Schweyen, K. Wolf, M. W. Mueller, *J. Mol. Biol.* **243**, 157 (1994).
20. Y. Hirose, J. L. Manley, *Genes Dev.* **14**, 1415 (2000).
21. S. H. Ahn, M. Kim, S. Buratowski, *Mol. Cell* **13**, 67 (2004).
22. A. Sickmann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13207 (2003).
23. B. A. Kaufman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7772 (2000).

24. We thank G. Jona, D. Gelperin, M. Smith, and L. Kung for manuscript comments and J. Ptacek and L. Meng for experimental assistance. D.H. was supported by an NIH training grant and H.Z. by a fellowship from the Damon Runyon–Walter Winchell Cancer Research Foundation. This research was supported by grants from the NIH.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5695/482/DC1

Materials and Methods

Figs. S1 and S2

Tables S1 and S2

13 February 2004; accepted 5 August 2004