BAS1 Has a Myb Motif and Activates *HIS4* Transcription Only in Combination with BAS2

KIMBERLY TICE-BALDWIN, GERALD R. FINK, KIM T. ARNDT

The BAS1 and BAS2 proteins are both required for activation of GCN4-independent (basal) *HIS4* transcription in yeast. BAS1 has an NH_2-terminal region similar to those of the *myb* proto-oncogene family. BAS1 and BAS2, which contains a homeo box, bound to adjacent sites on the *HIS4* promoter. The joint requirement of BAS1 and BAS2 for activation is probably not due to cooperative binding or the transcriptional control of one of the genes by the other. Although BAS1 and BAS2 were both required for activation of *HIS4* transcription, BAS1 was not required for BAS2-dependent expression of the secreted acid phosphatases. The transcriptional activators of *HIS4* have DNA binding domains that are conserved in evolution (BAS1 = Myb, BAS2 = homeo box, GCN4 = Jun). Their interactions, therefore, may be relevant to the control of gene expression in more complex systems.

THE *HIS4* GENE OF *Saccharomyces cerevisiae* is regulated by two control systems: general and basal. General control is a global response to amino acid starvation; limiting a single amino acid increases the transcription of *HIS4* and other amino acid biosynthetic genes (*1*). Transcriptional activation by general control is mediated by the trans-acting GCN4 protein that binds to 5′-TGACTC-3′ sequences in the promoters of target genes (*2*). In the absence of amino acid starvation or in strains deleted for *GCN4*, *HIS4* is still transcribed at a high basal level. Two trans-acting proteins, BAS1 and BAS2, are required for this GCN4-independent transcription of *HIS4* (*3*). Genetic analysis revealed a cis-acting region of *HIS4*, from −246 to −215, which is required for the basal transcription. Strains that lack both basal and general control (*bas1-2 bas2-2 gcn4-2* strains) show little *HIS4* transcription and require histidine for growth.

BAS1 and BAS2, like GCN4, are global regulatory proteins. Strains carrying deletions of either *BAS1* or *BAS2* require adenine for normal growth, presumably because these proteins also activate purine biosynthesis (*3*). In addition, BAS2 [also known as PHO2 (*4*)] is required for expression of the secreted acid phosphatases that enable yeast to use organic phosphates in the growth medium as a phosphate source. BAS2 binds to the *HIS4* promoter, as demonstrated with crude extracts of yeast in gel shift assays (*3*); however, BAS1-dependent DNA binding was not detected in a similar assay. We now show that BAS1, like BAS2, is a DNA binding protein and identify the binding sites in the *HIS4* promoter for both.

In the absence of adenine BAS1 and BAS2 jointly stimulated *HIS4* transcription about 100 times more than either alone (Fig. 1A). The presence of adenine reduced this joint stimulation of transcription from 115 units to 13 units. Thus *HIS4* transcription is under both adenine and phosphate control (*3*). The adenine and phosphate regulation of *HIS4* transcription may be related to the roles of BAS1 and BAS2 in purine and phosphate regulation. Regulation of *HIS4* transcription by extracellular adenine could reflect an interconnection of the histidine and purine biosynthesis pathways (*5*).

In addition to activating *HIS4* basal transcription, BAS2 is required for expression of the secreted acid phosphatases (*6*) (Fig. 1B). In vitro BAS2 binds directly to the promoter of *PHO5*, which encodes the major secreted acid phosphatase (*3, 7*). In contrast, a *bas1 BAS2* strain has virtually identical levels of the secreted acid phosphatases as does an isogenic *BAS1 BAS2* strain (Fig. 1B). Therefore, BAS2 activates the expression of the secreted acid phosphatases independently of BAS1.

The DNA sequence of *BAS1* predicts an 89.6-kD protein of 811 amino acids (Fig. 2). Amino acids 85 to 215 of BAS1 are similar to a motif found in Myb proteins of higher organisms (Fig. 2). The Myb motif was first identified in the v-*myb* gene of avian myeloblastosis virus (*8*). Cellular homologs were then found in chicken (*9*), *Drosophila* (*10*), mouse (*11*), humans (*12*), and corn (*13*). The Myb motif of higher organisms contains three repeats of a sequence that has three regularly spaced tryptophan residues. Within each repeat, the tryptophans are separated by 18 or 19 amino acids, and the third tryptophan of a repeat is separated by 12 amino acids from the first tryptophan of the next repeat. The register of sequence repeats indicated in Fig. 2 may not be the functional Myb repeat.

The BAS1 protein contains almost three Myb repeats (Fig. 2) (*14*). The second Myb repeat of BAS1 contains a tyrosine in place of the third tryptophan. In the C1 protein of corn, an isoleucine replaces the first tryptophan in a Myb repeat. Perhaps the substitution of a hydrophobic amino acid for a tryptophan is acceptable for function. BAS1 contains additional conserved Myb motifs; the second Myb repeat of BAS1 contains the amino acid sequence PGRT (at positions 146 to 149), which occurs at the similar position in the third repeat of other Myb proteins. BAS1 also contains the sequence $GPGSKGX_{(12)}LISK$ (where X is any amino acid), starting at position 161 and overlapping the junction of the second and third Myb repeats. This sequence conforms to a purine nucleotide binding motif, $GXGXXGX_{(11-17)}HfHfXK$, where Hf is a hydrophobic amino acid (*15*). The purine binding motif and the involvement of purines in *HIS4* basal regulation could be

A

Strains			Adenine (m*M*)		
GCN4	*BAS1*	*BAS2*	0	0.1	0.3
−	+	+	115	86	13
−	+	−	1.3	0.8	0.8
−	−	+	0.3	0.3	0.2

B

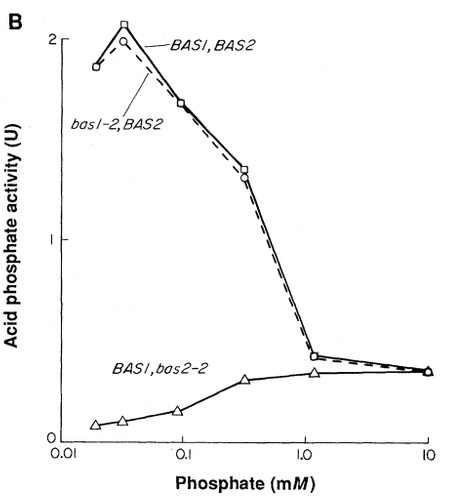

Fig. 1. BAS1 and BAS2 regulation of *HIS4* and secreted acid phosphatase levels. (**A**) *HIS4* basal level transcription as measured by β-galactosidase levels. Isogenic *ura3-52* strains (*3*), containing either wild type (+) or deletion (−) alleles of *GCN4*, *BAS1*, or *BAS2*, were transformed to Ura^+ with a low copy number YCp50 centromere plasmid containing a *HIS4-lacZ* fusion, pFN8 (*17*). Extracts of these strains were assayed for β-galactosidase activity (*32*). A control *CYC1-lacZ* fusion gave β-galactosidase values that varied by less than 20% as the adenine levels were varied. (**B**) BAS1 and BAS2 regulation of acid phosphatase activity. Strains of the indicated genotype were assayed for secreted acid phosphatase activity (*33*). The *PHO5* gene encodes the major inducible secreted acid phosphatase (*7*).

K. Tice-Baldwin and K. T. Arndt, Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724.
G. R. Fink, Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142.

Fig. 2. BAS1 contains a Myb domain. The predicted amino acid sequence of BAS1 is compared with *Drosophila* c-Myb (*10*), chicken c-Myb (*9, 15*), mouse c-Myb (*11*), and *Zea mays* C1 protein (*13*). Asterisks indicate the conserved tryptophan residues. Residues that are identical between BAS1 and the other Myb proteins are boxed. The numbers on the left give the position for the NH_2-terminal amino acid on each line. The *BAS1* DNA sequence has been submitted to GenBank.

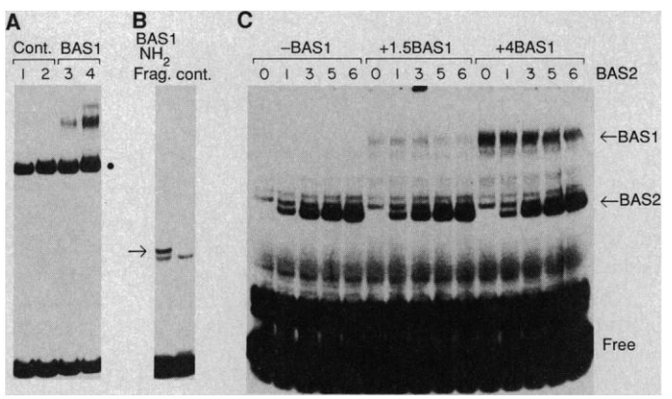

Fig. 3. BAS1 and an NH_2-terminal fragment of BAS1 bind to the *HIS4* promoter. (**A**) Gel shift assay using full-length BAS1. Extracts were prepared from yeast (*gcn4-2 bas1-2 bas2-2 ura3-52*) that contained either the high copy number 2μ *GAL1* promoter expression plasmid with no insert (pAB477, *URA3* selectable marker; control lanes) or the same plasmid with a Spe I–Mlu I DNA fragment containing *BAS1* inserted downstream of the *GAL1* promoter (pCB159; BAS1 lanes). Assays were performed as described (*34*). The control lanes 1 and 2 and the BAS1 lanes 3 and 4 are extracts from cultures of independent transformants and show the variability in the level of BAS1 binding activity (possibly because the high levels of BAS1 kill the yeast cells). The filled circle on the right side indicates a complex due to a yeast protein [which is not GCN4, BAS1, or BAS2, see (*2*)] that binds strongly to the *HIS4* promoter (sequences −261 to −172, relative to *HIS4* ATG). (**B**) Gel shift assay of an NH_2-terminal fragment of BAS1. Reactions (25 μl total volume) included 2 μl of crude *E. coli* lysate (0.70 mg per milliliter of protein) that contained either no BAS1 (control lane) or an NH_2-terminal fragment of BAS1. The preparation of these cells and extracts is in (*35*). The arrow indicates the complex with the amino terminal fragment of BAS1. An *E. coli* protein, present in both the control and BAS1 amino terminal fragment containing lysates, also bound to the *HIS4* promoter fragment (sequences −261 to −172, relative to the *HIS4* ATG). (**C**) BAS1 and BAS2 do not show cooperative binding in vitro. All gel shift assays had 10 μl of yeast extract. Left third, no BAS1; middle third, intermediate levels of BAS1 (1.5 μl of BAS1 extract); right third, high levels of BAS1 (4 μl of BAS1 extract). These extracts were prepared from a *gcn4-2 bas1-2 bas2-2* strain that had a high copy number *URA3* 2μ *GAL1* expression plasmid with either no insert (pAB477; control extract) or the *BAS1* gene inserted downstream of the *GAL1* promoter (pCB159; BAS1 extract). The volume (in microliters) of BAS2 extract was varied as shown. These extracts were prepared from a *gcn4-2 bas1-2 bas2-2* strain that had a high copy number 2μ *URA3* plasmid with either no insert (YEp24) or a 6.2-kb Cla I fragment of the *BAS2* gene (pAB291). All extracts were diluted to 1:0 mg of protein per milliliter before mixing. The order in which these extracts were added had no effect on the assay. An Eco RI–Hind III DNA fragment of a plasmid that contained *HIS4* promoter sequences −268 to −209 cloned into the Sal I site of pUC18 was used.

explained if BAS1 function were modulated by the binding of a purine nucleotide to BAS1.

When the *BAS1* coding sequences were transcribed in yeast from the strong *GAL1* promoter, BAS1-dependent binding activity to the *HIS4* promoter was detectable in crude yeast extracts by gel shift analysis (Fig. 3A). The specific DNA binding activity of BAS1 was contained in an NH_2-terminal fragment (amino acids 1 to 378) that contains the Myb motif (Fig. 3B). Since insufficient quantities of BAS1 are available from yeast for DNase I footprint analysis, BAS1 was prepared from an overexpressing strain of *Escherichia coli*. The *E. coli*–derived BAS1 bound to *HIS4* sequences that are required for BAS1-BAS2–dependent transcription of *HIS4* (Fig. 4A). The *HIS4* promoter sequences required for BAS1-BAS2–dependent transcription were defined by linker insertion-deletion analysis and are indicated by the vertical bar in Fig. 4A (*3, 16*).

BAS2 derived from *E. coli* also bound sequences in the *HIS4* promoter that are required for BAS1-BAS2–dependent *HIS4* transcription (Fig. 4B). Within the sequences protected from deoxyribonuclease I (DNase I) digestion by BAS2 is a 16-base motif that differs from $(TTAA)_n$ by a single nucleotide (Fig. 4C). The homeo box of BAS2 is most closely related to that of En, the protein product of the *engrailed* gene of *Drosophila* (*17*). Interestingly, En binds tightly to the *HIS4* promoter and gives a protection pattern almost identical to that of BAS2 (Fig. 4B).

BAS1 and BAS2 prepared from overexpressing strains of *E. coli* can bind simultaneously to the *HIS4* promoter (Fig. 4B). The sequences protected from DNase I by the simultaneous binding of BAS1 and BAS2 (Fig. 4C) closely match the sequences required for BAS1-BAS2–dependent transcription (*18*). No evidence for cooperative binding of *E. coli*–derived BAS1 and BAS2 was observed in this footprint analysis (Fig. 4B) or in gel shift analysis (*19*). One objection to these experiments is that BAS1 and BAS2 prepared from *E. coli* may lack a post-translational modification required for a cooperative interaction. Furthermore, both proteins produced in *E. coli* are partially proteolyzed [see (*20*)]. These problems were avoided by producing BAS1 and BAS2 in yeast, where the correct protein modifications would be made and proteolysis was not evident. We did not detect any evidence for cooperative binding to the *HIS4* promoter. As the amount of BAS2 increased, the presence or absence of BAS1 did not alter the intensity of the BAS2-DNA complex (Fig. 3C). In addition, even at the highest amounts of BAS2, the presence of

BAS1 did not result in a higher molecular weight BAS1-BAS2–DNA complex expected from cooperative binding. The reciprocal experiment, where the amount of BAS1 is varied in the presence or absence of BAS2, also gave no evidence of cooperative binding (*19*).

The requirement of both BAS1 and BAS2 for the activation of *HIS4* transcription was not a consequence of a regulatory interaction between BAS1 and BAS2 at the level of transcription. There is no difference in the steady state *BAS1* or *BAS2* RNA levels when wild type *BAS1 BAS2* strains are compared to *bas1-1* or *bas2-1* mutant strains (*3*). Therefore, under our culture conditions (synthetic complete medium), BAS1 does not regulate *BAS1* or *BAS2* transcription and BAS2 does not regulate *BAS2* or *BAS1* transcription. That a *BAS1-lacZ* fusion produces similar amounts of β-galactosidase in isogenic *BAS2* and *bas2-2* strains indicates that BAS2 does not regulate the translation of *BAS1* mRNA (*21*). Also, extracts prepared from isogenic *BAS1* and *bas1-2* strains contain similar amounts of BAS2 binding activity to the *HIS4* promoter (*3*). Furthermore, increasing the levels of BAS1 protein in a *bas2-2* strain or increasing the levels of BAS2 protein in a *bas1-2* strain does not relieve the requirement for both proteins in the activation of *HIS4* transcription (*3, 22*).

We did not detect a cooperative interaction between BAS1 and BAS2 for binding to the *HIS4* promoter. It is possible that proteins prepared differently would show cooperative binding. Another possibility is that the adjacent binding of BAS1 and BAS2 is a prerequisite for activation. Adjacent binding could (i) create a composite BAS1 + BAS2 surface that activates transcription; (ii) cause a conformational change in one or both of the proteins that would activate transcription; or (iii) result in the modification of one of the proteins to a transcriptionally active form by the other protein while both proteins are bound on the *HIS4* promoter.

The BAS1 and BAS2 proteins share similarities with two classes of DNA binding proteins. BAS1 has a pattern of tryptophans that is a common sequence motif for proteins of the Myb family. Within the sequences bound by BAS1 is the sequence TCCGGTA, which is similar to the concensus motif TCCGTTA located in sequences to which chicken v-Myb binds in vitro (*23*). The BAS2 protein contains a region related to the homeo box of proteins that control development of *Drosophila* (*24*); the En protein bound tightly to the *HIS4* promoter at the same sequences, $(TTAA)_n$, bound by BAS2. In addition to the similarity in their homeo box regions, both BAS2 and En contain a stretch of glutamine residues near the amino terminus. Interestingly, certain homeo box proteins, including En, also bind to another simple repeated sequence,

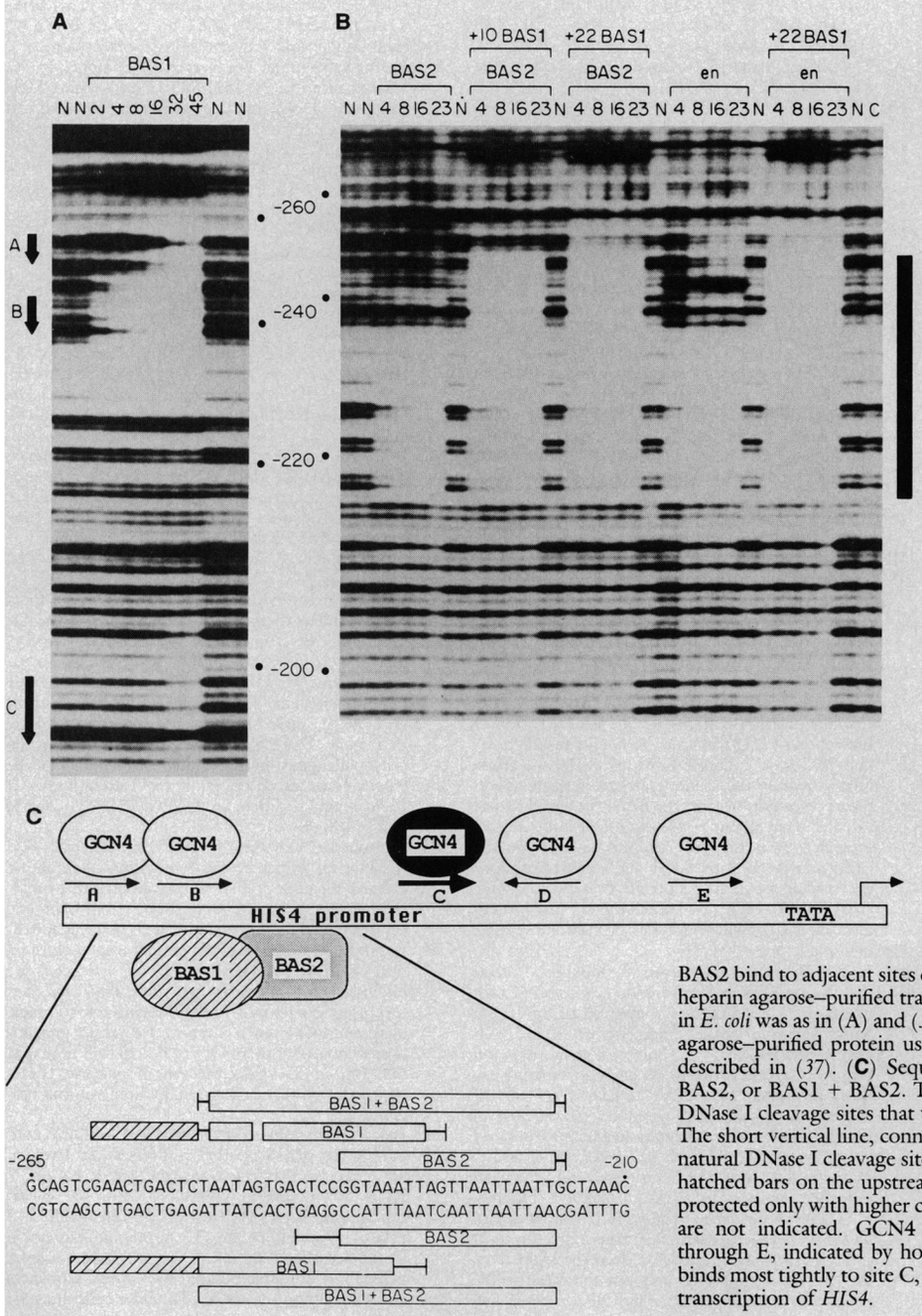

Fig. 4. *HIS4* sequences bound by BAS1 and BAS2. (**A**) The DNase I footprint on the *HIS4* promoter produced by BAS1 from *E. coli*. The amounts of heparin-agarose purified BAS1 are given in arbitrary units. Lanes marked by an N are controls that have no added protein. At high BAS1 concentrations, BAS1 protects bases upstream of −250 (probably a second BAS1 binding) and also protects bases around −198. These additional protected regions contain the GCN4 binding site 5′-TGACTC-3′ (indicated by vertical arrows) but probably did not represent physiological binding of BAS1; these sequences could be deleted without lowering the BAS1-BAS2–dependent *HIS4* basal transcription. Strains, protein preparation methods, and conditions for the DNase I footprint assays are in (*36*). (**B**) BAS1 and BAS2 bind to adjacent sites on the *HIS4* promoter. DNase I footprint analysis of heparin agarose–purified transcription, BAS1, BAS2, and En proteins expressed in *E. coli* was as in (A) and (*37*). BAS1 was from the same preparation of heparin agarose–purified protein used in (A). The lane marked by C is a control lane described in (*37*). (**C**) Sequences of the *HIS4* promoter protected by BAS1, BAS2, or BAS1 + BAS2. The bars represent the contiguous region of natural DNase I cleavage sites that were protected by BAS1, BAS2, or BAS1 + BAS2. The short vertical line, connected to the bar by a horizontal line, shows the next natural DNase I cleavage site that was not protected by the bound proteins. The hatched bars on the upstream side of the BAS1 protection are bases that were protected only with higher concentrations of BAS1. DNase I hypersensitive sites are not indicated. GCN4 binds to five 5′-TGACTC-3′ motifs (labeled A through E, indicated by horizontal arrows) in the *HIS4* promoter (*2*). GCN4 binds most tightly to site C, and this site mediates most of the GCN4-dependent transcription of *HIS4*.

$(TAA)_n$ (*25*). Perhaps such simple repeated sequence motifs represent ancestral DNA binding sites for homeobox proteins.

The fact that BAS1 contains a Myb domain and BAS2 contains a homeo box may shed some light on the activation properties of these homologs in more complex systems. The c-Myb family proteins might cooperate with certain homeo box proteins, such as En, to activate transcription. For example, the c-*myb* gene of *Drosophila* is expressed during embryogenesis (*26*). Expression of c-*myb* genes in vertebrates occurs primarily in immature hematopoietic cells (*27*) and the constitutive expression of c-*myb* cDNA blocks erythroleukemia cell differentiation in vitro (*28*). The chicken v-*myb* gene is able to transform only myeloid cell lineages in vivo or in vitro (*29*). Perhaps only myeloid cell lineages contain a second protein, like BAS2, that cooperates with the v-Myb protein to activate the transcription of genes that cause a transformed phenotype.

The inability of the En protein in *Drosophila* culture cells to activate transcription from a promoter that contains multiple copies of a DNA sequence to which En binds in vitro (*30*) could also reflect the requirement for a second protein. En probably binds to this promoter in vivo because En can inhibit Ftz-activated transcription from the same promoter element when both En and Ftz are coexpressed in the *Drosophila* cells (*31*). One explanation for discrepancy between En binding and En activation may be that the En protein, like BAS2, may require a second DNA binding protein for activation of transcription.

The molecular basis for the joint requirement of BAS1 and BAS2 for activation of *HIS4* transcription is not yet known. Although both BAS1 and BAS2 are required to activate *HIS4* transcription, BAS2 activates transcription of the secreted acid phosphatases independently of BAS1. In this role of activating transcription of the secreted acid phosphatases, it is not known if BAS2 requires a second DNA binding protein to take the place of BAS1. The ability of a DNA binding protein to activate some promoters by itself and other promoters only in concert with a second DNA binding protein could be important for developmental systems. For example, two regulatory proteins, A and B, localized at either end of the *Drosophila* embryo might form a gradient during embryogenesis—A, A + B, B—from anterior to posterior. Complex patterns of transcriptional expression could be generated with only these two activators if different promoters responded to the exact concentrations of each activator or combination of activators.

REFERENCES AND NOTES

1. E. W. Jones and G. R. Fink, in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983), pp. 181–299; A. G. Hinnebusch, *Microbiol. Rev.* **52**, 248 (1988).
2. I. A. Hope and K. Struhl, *Cell* **43**, 177 (1985); K. T. Arndt and G. R. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8516 (1986).
3. K. T. Arndt *et al.*, *Science* **237**, 874 (1987).
4. C. Sengstag and A. Hinnen, *Nucleic Acids Res.* **15**, 233 (1987).
5. The first step in histidine biosynthesis requires adenosine triphosphate. Also, a by-product of the histidine biosynthetic pathway, phospho-ribosyl-aminoimidazole-carboxamide (AICAR), is an intermediate in purine biosynthesis.
6. A. Toh-e *et al.*, *J. Bacteriol.* **113**, 727 (1973).
7. It is not known if BAS2 activates *PHO5* transcription by binding directly to the *PHO5* promoter. See K. Vogel *et al.*, *Mol. Cell. Biol.* **9**, 2050 (1989).
8. P. H. Duesberg, K. Bister, C. Moscovici, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5120 (1980); J. H. Chen, W. S. Hayward, C. Moscovici, *Virology* **110**, 128 (1981); T. J. Gonda *et al.*, *Cell* **23**, 279 (1981).
9. K.-H. Klempnauer, T. J. Gonda, J. M. Bishop, *Cell* **31**, 453 (1982).
10. C. W. B. Peters, A. E. Sippel, M. Vingron, K.-H. Klempnauer, *EMBO J.* **6**, 3085 (1987).
11. T. J. Gonda, N. M. Gough, A. R. Dunn, J. de Blaguiere, *EMBO J.* **4**, 2003 (1985); G. J. C. Shenong, H. C. Morse III, M. Potter, J. F. Mushinski, *Mol. Cell. Biol.* **6**, 380 (1986).
12. D. G. Bergmann, L. M. Souza, M. A. Baluda, *J. Virology* **40**, 450 (1981); W. J. Boyle, J. S. Lipsick, M. A. Baluda, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4685 (1986); K.-H. Klempnauer, C. Bonifer, A. E. Sippel, *EMBO J.* **5**, 1903 (1986).
13. J. Paz-Ares, D. Ghosal, U. Wienand, P. A. Peterson, H. Saedler, *EMBO J.* **6**, 3553 (1987).
14. BAS1 does not have a small sequence motif located towards the carboxyl terminus that is conserved between the *Drosophila* and vertebrate Mybs (*10*). Myb from vertebrates, including chicken, mouse, and human, are similar over most of their length.
15. P. Argos and R. Leberman, *Eur. J. Biochem.* **152**, 651 (1985).
16. F. Nagawa and G. R. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8557 (1985).
17. A. Fjose *et al.*, *Nature* **313**, 284 (1985).
18. The extent of overlap between the BAS1 and BAS2 binding sites on the *HIS4* promoter is not completely defined because there is a difference in the protection pattern between the heparin-agarose–purified BAS2 and BAS2 in the crude lysates. BAS2 in crude lysates of *E. coli* gives a smaller region of DNase I protection (extending from −225 through −217) than the protection in Fig. 4. The smaller region of protection of BAS2 in crude *E. coli* lysates might be due to lower concentrations of BAS2 in crude lysates because the crude lysates do not give complete protection. Because the BAS1 protected region extends from about −249 to −228, the region protected by both BAS1 and BAS2 could be the additive protection by BAS1 and an adjacent, nonoverlapping single BAS2 bound at −225 to −217. Crude lysates of *E. coli* containing the En protein give the same smaller region of protection as BAS2 in crude lysates.
19. K. Tice-Baldwin and K. T. Arndt, unpublished data.
20. Analysis on SDS–polyacrylamide gels of BAS1 produced in *E. coli* indicates a mixture of full-length BAS1 and two smaller fragments of BAS1. Full-length BAS1 from *E. coli* migrates at the identical position as yeast BAS1 in gel shift assays using the *HIS4* promoter. One of the smaller fragments of BAS1 binds to the *HIS4* promoter (about 50% of total binding activity) and migrates at approximately twice the rate of full length BAS1 in gel shift assays. These fragments of BAS1 (expressed from the heat-inducible lambda P_L-*BAS1* expression plasmid, pCB114) are produced in an *htpR* cI_{857} strain and an *htpR lon*::Tn10 cI_{857} strain. Furthermore, the heat shock–induction required for overexpression from the lambda P_L promoter was not the cause of the proteolysis because preparation of BAS1 from *E. coli* (using *lon*::Tn10, *htpR*, or *lon*::Tn10 *htpR* strains with an inducible T7 RNA polymerase gene, grown at 30°C or 37°C) that contained a plasmid with the *BAS1* gene under control of the T7 promoter (*35*) also yielded partially proteolyzed BAS1 protein.
21. A *BAS1-lacZ* fusion, integrated at the *URA3* locus, gave similar levels of β-galactosidase in isogenic *BAS1 BAS2* (5.5 units) or *BAS1 bas2-2* (4.9 units) strains grown on synthetic complete medium. Also, the β-galactosidase from the *BAS1-lacZ* fusion in the *BAS1 BAS2* strain did not vary when the amount of adenine in the growth medium was varied from 0 to 0.3 m*M*. Culture preparation and conditions for β-galactosidase assays were the same as those used for Fig. 1.
22. *BAS1 bas2-2 gcn4-2* strains that contained the *BAS1* gene on a high copy number plasmid (*3*) or *BAS1* expressed from the strong *GAL1* promoter (grown in 2% raffinose or 2% galactose medium; growth on galactose medium eventually kills the cells, however) grow extremely slowly on medium without histidine compared to medium containing histidine. Also, a *bas1-2 BAS2 gcn4-2* strain containing the *BAS2* gene on a high copy number plasmid grows extremely slowly in the absence of histidine (*3*). Therefore, when over-expressed, BAS1 by itself or BAS2 by itself only poorly activated *HIS4* transcription.
23. H. Biedenkapp, U. Borgmeyer, A. E. Sippel, K.-H. Klempnauer, *Nature* **335**, 835 (1988). The v-Myb concensus DNA binding site was obtained by DNase I footprint analysis of random fragments of chicken DNA that bind to v-Myb in vitro.
24. T. R. Burglin, *Cell* **53**, 339 (1988).
25. C. Desplan, J. Theis, P. H. O'Farrell, *ibid.* **54**, 1081.
26. A. L. Katzen *et al.*, *ibid.* **41**, 449 (1985).
27. T. J. Gonda, D. K. Sheiness, J. M. Bishop, *Mol. Cell. Biol.* **2**, 617 (1982); E. H. Westin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2194 (1982).
28. M. F. Clarke *et al.*, *Mol. Cell. Biol.* **8**, 884 (1988); K. Todokoro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8900 (1988).
29. W. J. Boyle, J. S. Lipsick, E. P. Reddy, M. A. Baluda, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4265 (1983); K.-H. Klempnauer *et al.*, *Cell* **33**, 345 (1983); K. E. Rushlow *et al.*, *Science* **216**, 1421 (1982).
30. Both En and Ftz bind to the DNA sequence TCAATTAAAT, which occurs within the promoter of the *engrailed* gene (*25*). This DNA sequence differs by two nucleotides from the $(TTAA)_n$ motif protected by BAS2 and En on the *HIS4* promoter.
31. J. Jaynes and P. H. O'Farrell, *Nature* **336**, 744 (1988).
32. The β-galactosidase activities are the average from two separate experiments. In each experiment, duplicate cultures for each of two separate transformants were assayed. Strains were grown to an optical density (OD) of 1 at 600 nm in yeast minimal medium (2% glucose) containing 0.5 m*M* arginine, 0.3 m*M* histidine, and the indicated level of adenine. Extracts were prepared by breaking the cells with glass beads and were assayed for β-galactosidase as described in G. Lucchini, A. G. Hinnebusch, C. Chen, G. R. Fink, *Mol. Cell. Biol.* **4**, 1326 (1984).
33. Phosphatase activities (*6*) are the average from two separate experiments (each consisting of duplicate cultures for each point). Yeast were grown to mid-exponential phase in a phosphate-free synthetic complete medium (*3*) to which KH_2PO_4 was added.
34. Gel shift assays (25 μl total volume) were performed using 5 μl of yeast extract (1.0 mg of protein per milliliter) prepared as in (*3*). Overnight yeast cultures were grown in 2% raffinose synthetic complete medium lacking uracil. Cultures for extract preparation were grown to an OD of 0.8 (6 to 8 hours) at 600 nm in 2% galactose (to induce the *GAL1* promoter) synthetic complete medium without uracil.
35. Lysates were prepared from strain BL21(DE3) containing the pLysS plasmid [expresses low levels of T7 lysozyme that inactivates the low levels of T7 RNA polymerase in uninduced cells, see F. W. Studier and B. A. Moffatt, *J. Mol. Biol.* **189**, 113 (1986)] and either the T7 promoter expression plasmid without an insert (control lane, plasmid pCB67) or the same plasmid with BAS1 sequences encoding amino acids 1 to 378 under control of the

T7 promoter (plasmid pCB112). Cultures were grown at 37°C to an OD at 600 nm of 0.5, IPTG was added to 0.5 m*M*, and incubated 130 min with shaking at 37°C. Extracts were prepared by sonication of cells in 0.05*M* KCl buffer A (25 m*M* Hepes, *p*H 7.7, 10% glycerol, 0.5 m*M* EDTA, 10 m*M* 2-mercaptoethanol, 1 m*M* phenylmethylsulfonyl fluoride, and 0.5 μg/ml each of leupeptin, antipain, chymostatin, and pepstatin A, Sigma), centrifugation, and dialysis of the resulting supernatant against 0.05*M* KCl buffer A. Dialyzed extracts were centrifuged and analyzed in gel shift assays as in (*3*) except that the final concentration of KCl was 10 m*M* and of NACl was 40 m*M*.

36. The final buffer for the DNase I protection assays was 40 m*M* Hepes, *p*H 7.8, 40 m*M* KCl, 5 m*M* $MgCl_2$, 0.1 m*M* EDTA, and sonicated calf thymus DNA (14 μg/ml). The $MgCl_2$ was added with the DNase I. BAS1 was prepared from an overproducing strain of *E. coli* (AR68; *htpR*, λ cI_{857}) that contains a plasmid with the *BAS1* gene controlled by the inducible λ P_L promoter (pCB114). The BAS1 expression plasmid (pCB114) was an inframe fusion of the ATG start codon of the expression vector [pOT5, a derivative of pAS1; M. Rosenberg, Y. Ho, A. Shatzman, *Methods Enzymol.* **101**, 123 (1983)] with the second codon of the *BAS1* coding sequences. Heat induced cells containing the BAS1 protein were sonicated into 0.05*M* KCl buffer A [see (*35*)]. This extract was centrifuged and two volumes of 4*M* ammonium sulfate (*p*H 7.8) was added to the supernatant. After centrifugation, the pellet was resuspended in 0.05*M* KCl buffer A, dialyzed, and loaded onto a heparin-agarose column equilibrated in 0.05*M* KCl buffer A. The column was rinsed with 0.2*M* KCl buffer A; BAS1 was eluted with 0.7*M* KCl buffer A and dialyzed against 0.05*M* KCl buffer A. The BAS1 protein prepared by this method from *E. coli* binds very well to DNA even though it is partially proteolyzed (*20*).

37. Heparin-agarose–purified BAS2 protein [0.10 mg per milliliter of total protein (*36*)] was prepared from BL21(DE3) bacteria [with pLysS plasmid; see (*35*)] that contained a plasmid with the *BAS2* gene downstream of the T7 promoter (plasmid pCB262). Heparin-agarose–purified En protein [0.10 mg per milliliter of total protein, see (*36*)] was prepared from BL21(DE3) bacteria (without pLysS plasmid) that contained a plasmid with the *engrailed* gene downstream of the T7 promoter [plasmid pAR3040; T. Hoey and M. Levine, *Nature* **332**, 858 (1988)]. As determined by PAGE and gel shift analysis, the preparation of En protein was not proteolyzed while the preparation of BAS2 protein, like BAS1 (*20*), contained proteolytic fragments (about 50% of total binding activity, migrating almost twice as fast as full-length BAS2) that bound to the *HIS4* promoter. Expression of BAS2 from the lambda P_L promoter in strain AR68 or from the T7 promoter in other protease deficient strains did not prevent proteolysis. The full-length BAS1 and BAS2 from bacteria, migrated during gel shift analysis to the same position as yeast-derived BAS1 and BAS2. The lane marked by C contains 45 μl of heparin agarose–purified control extract (0.10 μg per milliliter of total protein, prepared identically as for BAS1, BAS2, and En extracts) from control-induced BL21(DE3) bacteria that had the pLysS plasmid and the T7 polymerase expression plasmid with no insert. A similar control (*19*) for the BAS1 extract with an extract prepared from a heat induced control lambda P_L expression strain had results identical to lane C.

38. We thank J. Jaynes and P. O'Farrell for plasmid pAR3040 containing the *engrailed* gene under control of the T7 promoter, M. Goebl for pointing out that BAS2 is a homeodomain protein, and D. Shevell and A. Sutton for comments on the manuscript. G.R.F. is an American Cancer Society Research Professor of Genetics. Supported by NIH grant GM39892 to K.T.A. and GM35010 to G.R.F.

26 July 1989; accepted 10 October 1989

Vaccination with a Synthetic Zona Pellucida Peptide Produces Long-Term Contraception in Female Mice

SARAH E. MILLAR, STEVEN M. CHAMOW, ANNE W. BAUR, CONSTANCE OLIVER, FRANK ROBEY, JURRIEN DEAN

The zona pellucida surrounding mouse oocytes is an extracellular matrix composed of three sulfated glycoproteins, ZP1, ZP2, and ZP3. It has been demonstrated that a monoclonal antibody to ZP3 injected into female mice inhibits fertilization by binding to the zona pellucida and blocking sperm penetration. A complementary DNA encoding ZP3 was randomly cleaved and 200– to 1000–base pair fragments were cloned into the expression vector λgt11. This epitope library was screened with the aforementioned contraceptive antibody, and the positive clones were used to map the seven–amino acid epitope recognized by the antibody. Female mice were immunized with a synthetic peptide containing this B cell epitope coupled to a carrier protein to provide helper T cell epitopes. The resultant circulating antibodies to ZP3 bound to the zona pellucida of immunized animals and produced long-lasting contraception. The lack of ovarian histopathology or cellular cytotoxicity among the immunized animals may be because of the absence of zona pellucida T cell epitopes in this vaccine.

THERE IS CURRENTLY MUCH INTEREST in the development of a safe and effective contraceptive vaccine for population control. An ideal vaccine should have an effect that is long-lasting and highly specific and should inhibit fertilization as a contraceptive agent rather than disrupt early development as an abortifacient. In addition, the immunogen must induce an immunological response to an endogenous antigen that is effective as a contraceptive without eliciting a cytotoxic response that might result in abnormal reproductive function or other damage.

The mammalian zona pellucida (zona), which surrounds growing oocytes and ovulated eggs, is a potential immunogen for a contraceptive vaccine (*1, 2*). The zona of the mouse is composed of three sulfated glycoproteins (ZP1, ZP2, and ZP3) (*3*). Sperm initially bind to ZP3 via *O*-linked oligosaccharide chains, and continued binding involves ZP2 as a secondary sperm receptor. These two zona proteins form filaments that are crossed linked by ZP1 in the extracellular zona pellucida (*4*). The zona is unique to the ovary, highly antigenic, and accessible to circulating antibody during the intraovarian oocyte growth phase prior to meiotic maturation and ovulation (*1, 2*).

Passive immunization of mice with sera against the zona produces reversible contraception without obvious side effects (*1, 5*). In more recent experiments, rat monoclonal antibodies against ZP2 and ZP3 were injected into female mice. The antibodies bound specifically to the zonae surrounding intraovarian oocytes and produced long-term (more than 8 weeks), reversible contraception by preventing sperm penetration of the zona pellucida (*6, 7*). However, the epitopes recognized on mouse ZP2 and ZP3 by five different rat monoclonal antibodies are not present on other mammalian zonae pellucidae (*6, 7*), limiting their usefulness as contraceptive agents.

The recent cloning of the ZP3 gene and the characterization of its transcript and protein product (*8, 9*) have provided sufficient molecular detail of the zona proteins to suggest an alternative contraceptive strategy based on active immunization with a zona peptide. We have made use of the specificity of a monoclonal antibody to ZP3 known to block fertilization to identify a zona pellucida peptide for testing as a contraceptive vaccine. A 1.0-kb cDNA that contains sequences encoding the epitope recognized by the monoclonal antibody to ZP3 (*8*) was cut into random fragments, which were size selected (200 to 1000 bp) and cloned into the λgt11 expression vector (*10*). This epitope library was screened with a monoclonal antibody to ZP3 (*7*) and the nucleic acid sequence of the cDNA insert from eight positive clones was determined (Fig. 1A).

S. E. Millar, S. M. Chamow, A. W. Baur, J. Dean, Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892.
C. Oliver, Laboratory of Microbiology and Immunology, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20892.
F. Robey, Laboratory of Cellular Development and Oncology, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20892.