Mechanisms of Gene Regulation in the General Control of Amino Acid Biosynthesis in \textit{Saccharomyces cerevisiae}

ALAN G. HINNEBUSCH

Unit on Molecular Genetics of Lower Eukaryotes, Laboratory of Molecular Biology, National Institute of Child Health and Human Development, Bethesda, Maryland 20892

INTRODUCTION

A large number of genes encoding enzymes in multiple amino acid biosynthetic pathways are under a common control in the yeast \textit{Saccharomyces cerevisiae}. Expression of these genes increases in response to starvation for any one of several amino acids. The cross-pathway character of this response has led to its designation as the general control of amino acid biosynthesis. The derepression of enzyme activity brought about by the general control system is essential for survival since nonderepressible mutants containing \textit{gcn} mutations are unable to grow in starvation conditions to which wild-type cells are resistant. In many pathways, amino acid-specific regulatory systems operate to override derepression by the general control system when the end products of those pathways are not limiting.

Molecular cloning and analysis of structural genes subject to the general control demonstrated early on that their expression is regulated at the transcriptional level. Since these genes are unlinked in the yeast genome, the general control system emerged as a useful model for studying the coordinate regulation of unlinked genes in a eucaryotic organism. Thanks to the efforts of workers in several different laboratories, it is now clear that transcriptional regulation by the general control system is dependent upon a short nucleotide sequence found repeated upstream from each coregulated structural gene. This sequence functions as a positive regulatory site and is known to be the recognition sequence for the \textit{GCN4} protein.

The \textit{GCN4} gene was first identified by genetic analysis and was considered to encode the most direct \textit{trans}-acting positive regulator in the general control system. This prediction was confirmed by in vitro studies of the interaction of \textit{GCN4} protein with the \textit{5}'-noncoding deoxyribonucleic acid (DNA) of structural genes subject to the general control. \textit{GCN4} contains a very ancient and highly conserved DNA-binding domain; another domain of the protein is required to stimulate transcription once it is bound at the correct site in the promoter. The \textit{GCN4} protein is now one of the best-characterized eucaryotic transcription factors.

Genetic analysis showed that \textit{GCN4} is not the only \textit{trans}-acting positive regulator in the general control system.
At least three additional GCN gene products are needed for the derepression response. In addition, a collection of negative-acting GCD gene products are required to maintain repression in nonstarvation conditions. The genetic data suggest that many of the GCD gene products act indirectly to regulate gene expression, functioning as negative regulators of GCN4. GCN1, GCN2, and GCN3 also appear to act indirectly by repression or antagonism of GCD gene products. It is now clear that the level of GCN4 activity in the cell is controlled by GCN and GCD gene products at the translational level. The GCN4 transcript has an unusual structure, containing four AUG triplets located upstream from the initiation codon for GCN4 protein synthesis. Upstream AUG codons are extremely rare in yeast messenger ribonucleic acids (mRNAs) and generally act to inhibit translation of downstream coding sequences. This property forms the basis for repression of GCN4 protein synthesis in nonstarvation conditions. The novel aspect of GCN4 mRNA is that the inhibitory effect exerted by its upstream AUG codons can be modulated in response to amino acid availability. This modulation requires a particular combination of the upstream AUG codons as well as GCN and GCD regulatory factors. GCN4 mRNA is the first known example of translational control exerted by upstream initiation codons in an eucaryotic organism.

Mutations in most GCD genes are pleiotropic and some lead to conditional lethality. This phenotype suggests that GCD gene products carry out an essential function in addition to regulation of GCN4 expression. Given their involvement in translational control of GCN4 mRNA, perhaps GCD gene products contribute to general protein synthesis. In this view, the role of GCN1, GCN2, and GCN3 would be to modify certain key components of the translational apparatus to suppress the effects of the upstream AUG codons in GCN4 mRNA without inhibiting general protein synthesis. Recent genetic analysis suggests that the GCN3 product plays a central role in reducing GCD1 and GCD2 activity in starvation conditions.

This review begins with a consideration of the nature of the starvation signal, the scope of the derepression response, and the interplay between the general control and various pathway-specific regulatory mechanisms. Discussion of these topics is followed by a detailed description of general control regulatory sites at several target structural genes and the relationships between these sites and basal expression and pathway-specific control elements present at the same genes. The next section focuses on the trans-acting factors involved in the general control. GCN4 as a DNA-binding protein and transcriptional activator is discussed in detail, followed by an in-depth analysis of the role of other GCN and GCD factors in translational control of GCN4 mRNA. The requirement of the GCD gene products for cell division and the participation of the GCN3 protein in this essential function conclude the treatment of the general control system in S. cerevisiae. The final section provides a brief description of general amino acid control in other microorganisms.

GENERAL AMINO ACID CONTROL IN S. CEREVISIAE

Derepression Signal

Starvation of yeast cells for any one of at least 10 amino acids leads to increased expression of more than 30 enzymes in nine different amino acid biosynthetic pathways. It is possible that starvation for any amino acid is sufficient to elicit this response. In addition, the involvement of other pathways has not been ruled out. In the laboratory, one commonly used method of imposing starvation is to culture wild-type cells in the presence of inhibitors of amino acid biosynthesis. For example, 3-amino triazole inhibits the activity of the HIS3 gene product and can be used to produce histidine starvation. An alternative method is to culture a strain containing a leaky mutation in a structural gene for an amino acid biosynthetic enzyme on medium lacking the required amino acid. A lesser degree of derepression has been observed in wild-type cells grown in the presence of certain amino acid supplements. These supplements probably lead to starvation for another amino acid as the result of regulatory interactions between related biosynthetic pathways. For example, a leucine supplement to minimal medium results in partial derepression of enzymes under the general control. This derepression can be reversed by simultaneous addition of valine and isoleucine (82). The biosynthesis of these three amino acids involves several shared enzymatic steps, and it seems likely that one or more enzymes common to the three pathways are repressed or inhibited by leucine.

Although depletion of an amino acid pool is sufficient to derepress enzyme expression, this condition is not required to elicit the general control response. Derepression can be observed in the absence of amino acid starvation in lsl1 mutants that contain a temperature-sensitive isoleucyl-tRNA synthetase (tRNA synthetase). At 36°C, these mutants cannot survive because protein synthesis is blocked. At lower temperatures, the cells continue to divide but exhibit reduced synthetase activity and low levels of isoleucyl-tRNA. Under the latter conditions, derepression of enzymes in at least four amino acid biosynthetic pathways has been observed (66, 80). Similar results were obtained with mesl1 mutants that produce a temperature-sensitive methionyl-tRNA synthetase (15). Derepression of enzyme expression in lsl1 mutants is blocked by a mutation in the GCN1 gene, one of several positive regulatory genes whose products are required for general control-mediated derepression (80). This result shows that a reduced tRNA charge leads to derepression by the same mechanism that operates in response to actual amino acid starvation imposed by one of the regimens described above. Therefore, either the presence of uncharged tRNA or a reduced rate of protein synthesis is a more proximal signal for derepression than the depletion of an amino acid pool.

Expression of a number of amino acid biosynthetic enzymes increases in response to a nutritional shiftdown from rich to minimal medium. In S. cerevisiae, the new steady-state level of expression is established in <1 h (37). It may be best to distinguish this effect from derepression brought about by the severe starvation imposed by enzyme inhibitors or by culturing bradytrophs on minimal medium. In the latter two conditions, derepression is strongly dependent upon the positive regulatory factors encoded by GCN genes. By contrast, derepression in response to a shiftdown from rich to minimal medium occurs with similar efficiency in wild-type strains and gen mutants. In addition, no change in GCN4 expression occurs in a nutritional shiftdown, whereas GCN4 expression is induced by severe amino acid starvation (37). This observation is important because GCN4 encodes the most direct positive regulator in the general control system and derepression is controlled primarily by regulating the amount of GCN4 protein present in the cell (see below). That GCN4 expression is unaffected by a nutritional shiftdown suggests that...
enzyme derepression in these conditions is the result of a regulatory mechanism distinct from general amino acid control.

**Amino Acid Biosynthetic Enzymes Subject to General Amino Acid Control**

Table 1 lists the amino acid biosynthetic enzymes that have been examined for regulation under the general control system. For each enzyme listed as subject to general control, derepression was demonstrated by starvation for an amino acid whose biosynthesis does not depend upon that enzyme. This condition is required to distinguish between general and pathway-specific controls. In some cases, regulation was demonstrated differently by comparing enzyme expression in wild-type and gcd mutant strains. gcd mutants are unable to maintain repression of enzymes subject to general control in nonstarvation conditions. By using these approaches, 32 enzymes belonging to nine different pathways were shown to be regulated by the general control system, with derepres-

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Enzyme*</th>
<th>Genetic locus</th>
<th>Derepression observed in response to:</th>
<th>mRNA examined</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp</td>
<td>Anthranilate synthase</td>
<td>TRP2</td>
<td>+ +</td>
<td>18, 70, 94, 118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anthranilate PR-transferase</td>
<td>TRP4</td>
<td>+ +</td>
<td>70, 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indoleglycerol-P-synthase</td>
<td>TRP5</td>
<td>+ +</td>
<td>70, 121</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PR-anthranilate isomerase</td>
<td>TRP1</td>
<td>+ +</td>
<td>70, 80</td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>DAHP synthetase</td>
<td>ARO3</td>
<td>+ +</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>ARO4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Acetylglutamate synthase</td>
<td>ARG2</td>
<td>+ +</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acetylglutamate kinase</td>
<td>ARG6</td>
<td>+ +</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acetylglutamyl-P reductase</td>
<td>ARG5</td>
<td>+ +</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acetylornithine aminotransferase</td>
<td>ARG8</td>
<td>+ +</td>
<td>94, 117</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OTCase</td>
<td>ARG3</td>
<td>+ +</td>
<td>18, 47, 64, 67, 74, 91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Argininosuccinate-lyase</td>
<td>ARG4</td>
<td>+ +</td>
<td>18, 20, 64, 67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbamoyl-P synthetase</td>
<td>CPA1</td>
<td>+ +</td>
<td>68, 86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbamoyl-P aminotransferase</td>
<td>CPA2</td>
<td>+ +</td>
<td>67, 68, 86</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>ATP PR-transferase</td>
<td>HIS1</td>
<td>+ +</td>
<td>18, 38, 118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PR-AMP cyclohydrase</td>
<td>HIS4A</td>
<td>+ +</td>
<td>19, 20, 67, 98, 118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histidinol dehydrogenase</td>
<td>HIS4C</td>
<td>+ +</td>
<td>19, 20, 67, 94, 98, 118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Imidazoleglycerol-P dehydratase</td>
<td>HIS3</td>
<td>+ +</td>
<td>33, 100, K. Struhl*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histidinol-P aminotransferase</td>
<td>HIS5</td>
<td>+ +</td>
<td>83, 94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Histidinol-P phosphatase</td>
<td>HIS2</td>
<td>+ +</td>
<td>94, 118</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Homocitrate synthetase</td>
<td>LYS1</td>
<td>+</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Saccharopine dehydrogenase</td>
<td>LYS13</td>
<td>+</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Saccharopine reductase</td>
<td>LYS9</td>
<td>+</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>α-Aminoadipate reductase</td>
<td>LYS14</td>
<td>+</td>
<td>89, 118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LYS2</td>
<td>+ +</td>
<td>89, 118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LYS3</td>
<td>+ +</td>
<td>89, 118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LYS4</td>
<td>+ +</td>
<td>89, 118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LYS5</td>
<td>+ +</td>
<td>89, 118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ilv</td>
<td>Transaminase B</td>
<td>ILV1</td>
<td>+ +</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Threonine deaminase</td>
<td>ILV2</td>
<td>+ +</td>
<td>J. G. L. Petersen and S. Holmberg*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acetohydroxyacid synthase</td>
<td>ILV3</td>
<td>+ +</td>
<td>S. C. Falco*</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>α-IPM synthase</td>
<td>LEU4</td>
<td>+ +</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LEU5</td>
<td>+ +</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>α-IPM isomerase</td>
<td>LEU1</td>
<td>+ +</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-IPM dehydrogenase</td>
<td>LEU2</td>
<td>+ +</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive regulatory protein</td>
<td>LEU3</td>
<td>+ +</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Glutamine synthetase</td>
<td>GLN1</td>
<td>+ +</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAD-GDH</td>
<td>+ +</td>
<td>18, 72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glt</td>
<td>NADP-GDH</td>
<td>GDH1</td>
<td>+ +</td>
<td>18, 118</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>Sulfite reductase</td>
<td>MET10</td>
<td>- -</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MET15</td>
<td>- -</td>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MET18</td>
<td>- -</td>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MET19</td>
<td>- -</td>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MET20</td>
<td>- -</td>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ILSI</td>
<td>+ +</td>
<td>69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* PR, Phosphoribosyl; DAHP, 3-deoxy-D-arabinohexulosonic acid 7-phosphate; OTCase, ornithine carbamoyltransferase; ATP, adenosine triphosphate; AMP, adenosine monophosphate; AHA, acetohydroxyacid.

**References:**

1. Holmberg, L. Petersen
2. Falco, S. C.
3. Struhl, K.
4. P. Petersen and S. Holmberg
5. Holmberg, L. Petersen
6. Falco, S. C.
7. Personal communication.

*Table 1. Enzymes subject to general amino acid control*
sion ratios ranging from 2- to 10-fold. In numerous cases, it was demonstrated that the increase in enzyme expression observed in starvation conditions can be accounted for by an increase in the steady-state amount of the corresponding mRNA (Table 1).

Interplay between Pathway-Specific Regulation and General Amino Acid Control

Enzymes in certain amino acid biosynthetic pathways are subject to pathway-specific regulation in addition to general amino acid control. In the absence of starvation for the amino acid end product of the pathway, these specific mechanisms can partially override general control-mediated derepression in response to starvation for other amino acids. In such cases, efficient derepression in response to starvation for an amino acid other than the pathway end product can only be observed in a mutant lacking the specific repression mechanism.

Arginine biosynthetic enzymes. ARG3 and CPAI are two genes encoding enzymes in the arginine biosynthetic pathway that are subject to dual control. Expression of these genes is derepressed in response to starvation for several amino acids in addition to arginine (18, 86, 118). At least for ARG3, this derepression requires the GCN2 (118) and GCN4 gene products (16) and is thus mediated by general amino acid control. In addition to regulation by the general control, ARG3 and CPAI expression is repressed by the addition of arginine to minimal medium. Moreover, an arginine supplement partially suppresses general control-mediated derepression of these genes in response to starvation for other amino acids (18, 86). A mutation in the argRII (= arg81) gene eliminates the repression of ARG3 enzyme expression by arginine. As a result of this mutation, derepression of ARG3 in response to starvation for histidine, lysine, or valine becomes comparable to that observed in response to starvation for arginine (18). The cis-acting mutation caseO-8 eliminates arginine repression of CPAI, and in strains containing this mutation CPAI expression is more substantially derepressed by starvation for histidine, lysine, and valine (86).

Unlike the general control which regulates transcription of ARG3 and CPAI, arginine-specific repression of these genes appears to operate posttranscriptionally. The presence of arginine in the medium does not reduce the steady-state amounts of ARG3 and CPAI mRNA relative to their levels in medium lacking arginine (67, 68). Consistent with a posttranscriptional control mechanism, it was recently demonstrated that a short coding sequence in the leader of CPAI mRNA is required for the arginine-specific regulation of this gene (114). By contrast, two cis-acting mutations that abolish ARGr-mediated repression of ARG3 map upstream of the presumed initiation site of ARG3 transcription. To reconcile this fact with posttranscriptional control of ARG3 expression, it was proposed that the ARGr regulatory factors bind to ARG3 5'-noncoding DNA and become associated with the ARG3 transcript during its synthesis by RNA polymerase (16).

Lysine biosynthetic enzymes. Enzymes in the lysine biosynthetic pathway are also subject to dual control. Addition of lysine to minimal medium leads to repression of lysine biosynthetic enzymes. This repression is impaired by a mutation known as lys80. Because the lys80 mutation does not affect expression of ARG4, a gene regulated solely by the general control, it is unlikely that lys80 encodes a general control regulatory factor. General control-mediated derepression of LYS genes in response to starvation for amino acids besides lysine is much more evident in lys80 mutants than in wild-type cells. This is particularly true for saccharopine reductase, that shows no derepression in response to arginine or isoleucine-valine starvation in a lys80 strain, but exhibits two- to fourfold derepression under the same circumstances in a lys80 mutant. Thus, the lys80 gene product can override general control-mediated derepression of lys gene expression when lysine is not a limiting amino acid (89). The step in lys gene expression controlled by lys80 remains to be determined.

Leucine biosynthetic enzymes. The LEU1 and LEU2 genes, encoding two enzymes in leucine biosynthesis, are subject to repression by excess leucine (reviewed in reference 49). This regulation has been shown to operate at the translational level (2, 46). Mutational analysis of the LEU2 5'-noncoding region has identified a guanine-plus-cytosine-rich palindrome sequence that is required for efficient expression and probably also mediates leucine repression of LEU2 (63). Related sequences occur upstream of LEU1 (46).

Expression of LEU1 and LEU2 does not derepress in response to starvation for tryptophan, lysine, or arginine or in response to a mutation in the general control-negative regulatory gene Gcn2 (45). From these results it was suggested that LEU1 and LEU2 are not under the general control. However, in the 5'-noncoding region of the LEU1 gene is a sequence (9) that closely resembles the consensus sequence for the binding site of GCN4 protein, the direct transcriptional activator in the general control system. This fact raises the possibility that LEU1 is subject to dual control but that leucine-specific repression overrides general control-mediated derepression when cells are starved for amino acids besides leucine. Sequences with lesser homology to the GCN4-binding site are also found in the 5'-noncoding DNA of the LEU2 gene (3).

In contrast to LEU1 and LEU2, LEU4 expression is clearly governed by the general control but does not appear to be subject to leucine repression (45). In fact, addition of leucine leads to increased LEU4 expression, presumably as the result of general control-mediated derepression in response to isoleucine-valine starvation. Interestingly, the LEU4 5’-noncoding region contains both the GCN4-binding site and sequences with substantial homology to the guanine-plus-cytosine-rich sequences implicated in leucine repression (9). This observation raises the possibility that LEU4 is also subject to dual control. Perhaps leucine-specific repression of LEU4 is overridden by general control-mediated derepression in response to isoleucine-valine starvation when cells are grown in the presence of leucine. To test this idea, it would be of interest to look for leucine repression of LEU4 in the presence of isoleucine and valine or in a gcn mutant. Both conditions should eliminate derepression of LEU4 by the general control system. In fact, unpublished data have been cited suggesting that in such circumstances LEU4 expression is indeed repressed by leucine (11).

The product of the reaction carried out by the LEU4 gene product, α-isopropylmalate (α-IPM), is a positive effector of both LEU1 and LEU2 expression (6). The LEU4 gene product is feedback inhibited by leucine (93, 111). Therefore, leucine repression of LEU1 and LEU2 could be mediated by reducing the level of the inducer α-IPM. This suggestion is consistent with the idea that the leucine-specific regulatory sequence at LEU2 functions as a site for positive control of transcription (63). Recessive mutations in the LEU3 gene lead to constitutively low levels of the LEU1 and LEU2 gene products, suggesting that LEU3 encodes a positive regulatory factor responsible for leucine-specific regulation of
these genes (6, 11). It seems likely that the regulatory function of LEU3 is stimulated by α-IPM (or a metabolic derivative of this compound). Given that LEU4 is subject to the general control, derepression of LEU1 and LEU2 should occur in response to leucine starvation as the result of increased synthesis of α-IPM and consequent greater activation of LEU3 function. Interestingly, it now appears that LEU3 expression is also governed by the general control (122). This finding provides a second reason to suspect that LEU1 and LEU2 expression derepress in leucine starvation conditions via the general control system. This response could be indirect and result strictly from increased LEU3 and LEU4 expression, or, as suggested above, it may also involve a direct interaction between the GCN4 protein and regulatory sequences at LEU1 and LEU2.

Isoleucine-valine biosynthetic enzymes. Expression of four of the five enzymes involved in the biosynthesis of isoleucine and valine are subject to both the general control (Table 1) and multivalent repression exerted by a combination of leucine, isoleucine, and valine (reviewed in reference 49). It is interesting that two of the structural genes for the ILV pathway, ILV2 (23) and ILV5 (85), contain sequences in their 5′-noncoding regions closely related to the guanine-plus-cytosine-rich palindrome implicated in leucine-specific repression. This finding raises the possibility that these genes are under leucine-specific control. If so, the multivalent nature of the repression of these genes might simply reflect the need for simultaneous addition of isoleucine and valine to prevent derepression by the general control system in conditions of leucine excess.

Glutamine biosynthetic enzymes. Regulation of the structural gene for glutamine synthetase (GLN1) illustrates a different sort of interaction between specific and general control mechanisms from those just described. Glutamine synthetase is subject to dramatic repression by glutamine, mediated in part by inactivation of a positive regulator encoded by GLN3 (22, 71, 72). In the absence of glutamine in the medium, glutamine synthetase is derepressed, and neither starvation for histidine nor a gcd1 general control regulatory mutation leads to additional derepression (72, 118). Ostensibly, this result suggests that GLN1 expression is not subject to the general control. However, in glutamine-supplemented medium, in which glutamine synthetase is fully repressed, higher levels of enzyme activity are observed in response to either histidine starvation or a gcd1 mutation. The derepression produced by histidine starvation is dependent upon GCN4 and is thus mediated by the general control system (72). Therefore, unlike the examples cited above, the general control-mediated derepression of GLN1 occurs only under conditions in which amino acid-specific repression operates. Derepression of glutamine synthetase in histidine-starved cells is partially dependent upon the function of GLN3 as well as GCN4. This suggests that at least one component of GLN1 derepression in histidine starvation conditions is indirect and results from increased GLN3 positive regulatory activity brought about by the general control response. It was proposed that GLN3 activation in histidine-starved cells occurs as a by-product of a reduction in the glutamine pool (72).

Nicotinamide adenine dinucleotide (NAD)-dependent glutamate dehydrogenase (GDH) is another enzyme subject to GLN3 positive control (71). In contrast to the regulation of glutamine synthetase expression, this enzyme is derepressible by the general control system in both the presence and the absence of glutamine (18, 72). As in the case of glutamine synthetase, general control-mediated derepression of NAD-GDH in the presence of glutamine seems to be partially dependent upon GLN3 function (72). That NAD-GDH expression is derepressible under the general control in the presence or absence of glutamine makes it likely that the structural gene encoding this enzyme is regulated directly by the products of both GCN4 and GLN3.

Histidine biosynthetic enzymes. Addition of histidine to minimal medium does not repress histidine biosynthetic enzymes (118), suggesting that the HIS genes are regulated only by general amino acid control. However, HIS gene expression occurs at a considerable level even in cells lacking the GCN4 protein, the direct positive regulator in the general control system. Recent work has shown that two additional trans-acting factors encoded by the genes BAS1 and BAS2 are required for the GCN4-independent component of HIS4 basal expression. The two BAS gene products act synergistically as positive effectors of HIS4. The BAS2 product binds to positive regulatory sequences in the HIS4 promoter distinct from the binding sites for GCN4 protein (see below). Interestingly, BAS2 is identical to PHO2, a gene required for derepression of the acid phosphatase activities encoded by PHO5 and PHO11. As would be expected from this finding, the BAS2 product also binds to the 5′-noncoding region of the PHOS gene. Moreover, HIS4 expression is derepressed by low levels of inorganic phosphate, the same condition that elicits derepression of PHO5 and PHO11. This derepression of HIS4 in response to phosphate starvation requires both the BAS2 gene product and sequences in the HIS4 promoter that bind the BAS2 protein in vitro. Unlike BAS2, BAS1 is not involved in the regulation of PHO gene expression; however, mutations in both BAS1 and BAS2 impair adenine biosynthesis in addition to their effects on histidine production (5). These results demonstrate the existence of a second cross-pathway regulatory mechanism operating at HIS4. Unlike the interactions described above between general and specific amino acid control mechanisms, the effects of BAS-mediated gene regulation and the general control system appear to be completely additive.

Consequences of General Control-Mediated Derepression of Enzyme Activity on Amino Acid Pools

The levels of some enzymes in amino acid biosynthetic pathways subject to the general control derepress more than others. The significance of this variation is not obvious; however, it is generally the case that the enzymes that carry out the rate-limiting steps in the pathway exhibit the largest derepression ratios (reviewed in reference 37). This situation obviously tends to ensure that general control-mediated derepression results in an increased flux through the regulated pathways.

An effect of enzyme derepression on the sizes of amino acid pools has been observed for tyrosine, phenylalanine, lysine, arginine, and histidine. These pools increase 2- to 10-fold in response to starvation for any one of several amino acids. The leucine and glutamate pools show smaller increases between 50 and 100%, and these changes are not observed consistently. The tryptophan and methionine pools may also increase in starvation conditions; however, their small size makes these measurements difficult (18, 24, 64). Therefore, with the exception of isoleucine, valine, and glutamine, amino acid pool sizes are found to increase in starvation conditions for each of the pathways in which enzyme expression is regulated by the general control system. The absence of a consistent increase in the leucine, isoleucine, and valine pools is probably explained by the fact
that either enzymes in these pathways are not subject to the general control or their response to starvation for other amino acids is completely overridden by pathway-specific repression mechanisms.

It is uncertain whether the large increases in the amounts of basic and aromatic amino acids that result from general control-mediated derepression play any role in the cell's response to starvation for other amino acids. One possibility is that the accumulation of basic and aromatic amino acids changes the partitioning of other amino acids between the vacuolar and cytoplasmic compartments. For example, it was shown that a lysine supplement can increase the cytoplasmic concentration of arginine and ornithine. In addition, arginine and histidine supplements increase the concentration of several other amino acids in the cytoplasm. When these basic amino acids are added to the culture medium, they make up more than half of the total amino acids in the cell and are localized predominantly in the vacuole. Similarly, in histidine-starved cells, the concentrations of arginine, ornithine, and lysine become very large and make up almost three-fourths of the total vacuolar pool. This vacuolar accumulation of basic amino acids in starved cells might lead to larger cytoplasmic fractions of other, limiting amino acids (65).

An alternative explanation for the evolution of the general control system is its mechanistic simplicity: a single signal for starvation is generated in response to starvation for any amino acid (e.g., decylated tRNA) and a single trans-acting regulator of transcription is mobilized when the starvation signal appears (the GCN4 protein). The trans-acting factor can increase the expression of all structural genes subject to the general control through recognition of a single cis-acting regulatory sequence located upstream from each coregulated gene. This system should be very efficient in starvation conditions in which many amino acids become limiting at once. Presumably, situations involving more limited amino acid deprivation also occur frequently in nature, since it was necessary to superimpose upon the general control system amino acid-specific repression mechanisms for those pathways in which overproduction of the end product is either too expensive or toxic.

cis-ACTING REGULATORY ELEMENTS AT STRUCTURAL GENES SUBJECT TO THE GENERAL CONTROL

The structural genes whose transcription is subject to the general control are unlinked in the yeast genome. This suggests that a common nucleotide sequence exists at each gene to mediate the coordinate transcriptional response to amino acid starvation. In fact, the 6-base-pair (bp) sequence 5' TGACTC 3' has been found repeated upstream from every structural gene under the general control examined thus far (Fig. 1). Evidence reviewed below from mutational analyses of the HIS4, HIS3, and TRP5 genes demonstrates that this hexanucleotide forms the highly conserved core of a 12-bp element that is both necessary and sufficient for general control-mediated regulation of transcription. As discussed in greater detail in a subsequent section, the TGA CTC sequence is the binding site for the GCN4 positive regulatory protein.

**HIS4 Regulatory Region**

The HIS4 gene exhibits substantial basal expression in nonstarvation conditions. In wild-type cells subjected to amino acid starvation, or in gcn4 mutant cells grown in any condition, HIS4 expression increases three- to fourfold above this basal level. This derepression is completely dependent upon GCN4 function. About two-thirds of the HIS4 basal expression is also dependent upon GCN4 function, suggesting that the same sequences that mediate derepression in response to starvation also contribute to HIS4 transcription in nonstarvation conditions (61). By examining the effects of deletion mutations in both wild-type and gcn4 cells, it has been possible to determine the contributions of various sequences to the GCN4-dependent and GCN4-independent components of HIS4 basal expression.

Sequences upstream of position -181 (relative to the 5' end of the HIS4 transcription unit) can be deleted with little or no effect on HIS4 basal expression or its efficiency of derepression in response to starvation (5, 19, 61, 78). By contrast, an internal deletion of sequences between -202 and -179 leads to greatly diminished HIS4 basal expression. The drastic effect of the latter deletion is observed in a gcn4 mutant as well as in wild-type cells, indicating that a GCN4-independent basal control element is impaired by this mutation. Taken together, these results place the 5' boundary of a basal control element in the vicinity of -180 (5, 78). The 3' boundary of this element was placed at -154 from the fact that a variety of internal deletions with 5' junctions downstream of -154 have little or no effect on HIS4 basal expression (5, 78) (Fig. 1). As discussed below, the product of the BAS2 gene is required in trans for efficient HIS4 basal expression, in both wild-type and gcn4 mutant strains. The BAS2 protein has been shown to bind to sequences in the region -180 to -150 interval (5).

The deletions just described that impair the basal control element lead to only about twofold-lower HIS4 expression in amino acid starvation conditions compared with the wild-type HIS4 gene. As a result, these deletion alleles display a greatly increased derepression ratio (19, 61, 78). The -202 to -179 interval contains two copies of the TGACTC sequence. Therefore, these copies of the repeat are not absolutely required for general control-mediated derepression, at least in the presence of the remaining TGACTC sequences located downstream; however, they probably contribute to the efficiency of derepression.

Deletion of all HIS4 sequences between -588 and -138 leads to the same phenotype as small internal deletions that remove the basal control element in the -179 to -150 region: reduced basal expression and an elevated derepression ratio. By contrast, deletion of only two additional bases, to position -136, further reduces basal expression to almost background levels and, more importantly, completely abolishes derepression. The deletion to -136 removes the 5'-proximal T residue of the third copy of the TGACTC repeat, numbering from the 5' end (Fig. 1). This result strongly suggests that the repeat at -136 is absolutely required for derepression in the absence of the TGACTC sequences located upstream (19, 61). It also indicates that the repeat at -136 contributes to GCN4-dependent basal expression. In fact, the residual basal expression observed in all deletions that remove the upstream basal control element is completely GCN4-dependent (61).

An elegant demonstration of the importance of the TGACTC sequence for derepression at HIS4 was provided by the isolation of cis-acting suppressor mutations of the -136 deletion. These mutations restore both the GCN4-dependent basal expression and the efficient derepression in starvation conditions observed for deletion -138 and other deletions that leave the -136 TGACTC sequence intact.
These suppressor mutations were found to restore a copy of the TGACTC sequence by either of two single-base substitutions. In one class of revertants, the T residue normally found at -136 was recovered; in the second class, a substitution occurred 20 bp downstream that changes the sequence at -119 to -114 from 5'-TGACTC to 5'-TGAGTC 3', a sequence containing five of the six residues in the repeat core. In fact, the substitution in this second class of revertants creates a perfect copy of the TGACTC sequence on the opposite strand. (As discussed below, the repeat functions in both orientations.) In a third class of revertants, multiple mutations occurred at the -136 deletion junction that do not restore a copy of the repeat. In this class, the basal level of HIS4 expression is elevated relative to the parental deletion allele; however, derepression remains completely defective and there is no dependence for expression upon GCN4 function (19, 61). Taken together, these results strongly suggest that the TGACTC sequence is essential for GCN4-mediated derepression in starvation conditions.

Internal deletions of 20 to 30 bp that remove the HIS4 TGACTC sequence at -136 have been constructed. Two such mutations reduce expression under derepressing conditions to about one-third of the wild-type level. One of these mutations also removes the fourth repeat found in an inverted orientation at -110 (78) (Fig. 1). These results suggest that the -136 copy of the repeat is required for a wild-type level of derepressed expression, even in the presence of the upstream copies of the repeat at -192 and -180. From a comparison of the effects of deletions of different copies of the TGACTC sequence, it seems likely that the -136 copy is more potent as a regulatory site than the other copies at HIS4. On the other hand, no internal deletion of any single copy of the repeat reduces HIS4 expression in derepressing conditions to the low constitutive level observed for the wild-type gene in a gcna4 mutant (61, 78). This result implies that multiple copies of the TGACTC sequence at HIS4 are functional as GCN4-dependent regulatory sites. This conclusion is supported by additional data reviewed below.

The mutations described thus far have no effect on the 5' end of HIS4 mRNA. Therefore, they affect only the rate and not the site of transcription initiation (19). By contrast, deletions in the region between -70 and -23 eliminate production of the wild-type HIS4 transcript. All of these deletions remove the two TGACTC TATAA 3' -20 (Fig. 1) that is closely related to the "TATA" element known to be involved in mRNA 5'-end selection in higher eukaryotes. The effects of such deletion on the level of HIS4 transcription were largely reversed by replacing the deleted sequences with a synthetic oligonucleotide containing the sequence 5' TATAA 3'. This exact sequence does not occur in the -70 to -23 interval of the normal HIS4 promoter. Unlike the wild-type gene, transcription initiation does not occur at a single site for the TATAAAA replacement allele; however, transcripts are produced with discrete 5' ends that map near the normal 5' end (78). These results strongly suggest that a TATA box is the crucial sequence element in the -70 to -23 interval required for transcription initiation at the correct site.

The TATA box deletions do not completely abolish HIS4 expression; instead, a constitutive level of expression several-fold higher than that seen for deletions in the upstream basal control element is observed. However, since HIS4 transcripts with discrete 5' ends are not detectable, it appears that a multiplicity of transcripts with a variety of 5' endpoints must be generated, each at a low level. Interestingly, expression from such deletion alleles is unresponsive to amino acid starvation and appears to be completely GCN4 independent (5, 78). When the deleted sequences were replaced by an oligonucleotide containing the sequence 5' TATAA 3' (as opposed to the 5' TATAAAA 3' insertion discussed above), expression was boosted three- to fourfold. However, as with the parental deletion, transcripts with discrete 5' ends remained undetectable and HIS4 expression remained unregulated by the general control system (78).

That expression from the TATAA insertion and its parental deletion is constitutive suggests that regulation of HIS4 transcription may require an interaction between the GCN4 protein, bound to TGACTC sequences upstream, and a factor that recognizes a functional TATA sequence downstream (103). Since the sequence TATAAA can be substituted for the wild-type HIS4 TATA box without destroying regulation by GCN4, there seems to be some flexibility in the sequence requirements for this downstream element.

This flexibility was further demonstrated by the construction of hybrid promoters containing upstream sequences from the HIS4 promoter and downstream sequences from the yeast CYC1 promoter. To produce these hybrid constructs, a variety of small HIS4 fragments containing different numbers of the TGACTC repeats and different amounts of flanking nucleotides were used to replace the upstream activation sequence (UAS) of the CYC1 gene. The downstream CYC1 sequences that specify the sites of transcription initiation were left intact. Expression of CYC1 does not normally derepress in response to amino acid starvation. By contrast, every hybrid promoter examined (except one), containing at least one copy of the TGACTC sequence, exhibited significant derepression in response to histidine starvation. Like the wild-type HIS4 promoter, derepression was dependent upon GCN4 function, and repression of the hybrid promoters in non-starvation conditions required the GCD1 gene product (40).

The wild-type CYC1 gene utilizes multiple transcription initiation sites mapping over a -600-bp region. The HIS4-CYC1 hybrid promoters use exactly the same 5' ends under both repressing and derepressing conditions (40). Selection of the 5' ends of HIS4 mRNA has been shown to require three different TATA boxes (31). Two of these TATA boxes have the same sequence as the HIS4 TATA box, 5' TA TATAA 3', but the third has a slightly different sequence, 5' TAAATA 3'. These observations support the idea that the GCN4 protein may regulate transcription initiation events directed by a family of related TATA boxes.

FIG. 1. 5'-Noncoding regions of structural genes subject to general amino acid control. The scale at the top gives the distance in base pairs from the 5'-proximal transcription initiation site at each gene. The labeled boxes designate the protein-coding sequences, I's designate the transcription initiation sites; the arrows symbolize TGACTC sequences found in either orientation. Arrows with asterisks indicate TGACTC elements for which data exist to certify their probable function as regulatory elements, either mutational data indicating a requirement for normal derepression in vivo or an in vitro demonstration of GCN4 binding. The hatched boxes indicate basal control elements. For HIS4, the basal control element is the BAS2 protein-binding site. For the other three genes, the nature of the protein, if any, that binds to the basal control sequences is unknown. The identical hatching used for all four basal control elements is not meant to imply any sequence homology among them. The o mutations at ARG3 inactive arginine repression of this gene; the stippled boxes are sequence elements homologous to the leucine-specific regulatory element defined by mutational analysis of LEU2.
Two of the HIS4 fragments that confer general control over the CYC1 downstream promoter elements are only 14 and 24 bp in length. These two fragments contain, respectively, the −136 copy and the 5′-proximal pair of TGACTC sequences. That these fragments share no obvious sequence homology beyond their TGACTC sequences strongly suggests that this sequence is sufficient to confer transcriptional regulation by the general control system. Ostensibly at odds with this conclusion is the finding that a 7-bp fragment containing the TGACTC sequence is incapable of conferring general control upon the CYC1 promoter (40). However, as discussed more fully below, this result is more likely to indicate a requirement for a particular sequence context surrounding the TGACTC sequence for efficient binding of the GCN4 protein. A second illustration of the importance of sequences surrounding the TGACTC element arises from a comparison of the 24-bp fragment containing the two 5′-proximal TGACTC sequences and a tandem insertion of the 14-bp fragment containing the −136 TGACTC sequence: Two copies of the 14-bp fragment lead to fivefold-greater activation of CYC1 expression under derepressing conditions than the 24-bp fragment (40).

The tandem 14-bp insertion just described exhibits a level of derepressed expression comparable to that conferred by an 89-bp fragment containing the three TGACTC sequences at −192, −180, and −136 plus all of the intervening HIS4 nucleotides. By contrast, expression under repression conditions from the tandem 14-bp insertion is 50-fold lower than that conferred by the 89-bp fragment. Moreover, the basal expression conferred by the 14-bp fragment is GCN4 dependent, whereas that conferred by the 89-bp fragment is largely GCN4 independent (40). These findings are in accord with the idea that sequences located between the 5′-proximal pair and the −136 TGACTC sequences are required for efficient basal HIS4 expression. Bisection of the 89-bp fragment at position −173 has little effect on the ability of the resulting downstream half-fragment to confer efficient basal expression upon the CYC1 promoter (40). This result suggests that the BAS2 binding domain begins downstream from −173.

The HIS4 fragments used in constructing the HIS4-CYC1 hybrid promoters regulate transcription in either orientation relative to the direction of transcription (40). In the case of the GCN4 binding site, this feature probably reflects the fact that the TGACTC sequence in its optimum flanking sequence context shows twofold rotational symmetry (33). It is also evident from the results summarized above that a precise spacing is not required between the HIS4 upstream regulatory sequences and downstream promoter elements. Both the bidirectionality and the position independence exhibited by HIS4 regulatory sequences are characteristics reminiscent of the enhancer elements of mammalian transcriptional control elements (52).

These similarities between the HIS4 regulatory region and mammalian enhancers are also illustrated by the effects of a Ty insertion mutation on HIS4 expression. The his4-9128 allele has a cold-sensitive His" phenotype that is due to a 333-bp insertion of the terminal direct repeat of a Ty element (delta) at HIS4 position −97 (25). This allele produces a low level of wild-type HIS4 mRNA whose expression is subject to the general control (97). Recalling that HIS4 regulatory sequences are located upstream from the δ insertion site, this result shows that the TGACTC sequences can modulate transcription from the normal HIS4 initiation site at a distance >300 bp upstream from their normal location in the wild-type HIS4 gene. A larger HIS4 transcript is produced by his4-9128 with a 5′ end located in the δ sequence.

Expression of this transcript is also subject to the general control, showing once again that the HIS4 TGACTC sequences can activate transcription from a heterologous TATA box and transcription initiation site (97).

**HIS3 Regulatory Region**

HIS3 expression derepresses about fivefold in wild-type cells subjected to starvation conditions, and this derepression is GCN4 dependent. Similar to the HIS4 gene, 30 to 60% of HIS3 basal expression is also GCN4 dependent (106). There are two perfect copies of the TGACTC sequence in the HIS3 5′-noncoding region, at −258 and −99, and several additional copies that deviate from the consensus sequence at one position located at −225, −216, −181, and −142 (Fig. 1). Deletion of all HIS3 sequences between positions −447 and −142 has no effect on either basal expression or the extent of derepression of HIS3 transcription in starvation conditions. This result suggests that the four 5′-proximal copies of the TGACTC element are dispensable for HIS3 derepression, at least in the presence of the remaining copies downstream. By contrast, several internal deletions of the −99 TGACTC sequence, including one of only a single base pair that removes the 5′-proximal T of the repeat, result in the complete loss of derepression without affecting basal HIS3 expression (100, 106). Therefore, the −99 copy of the repeat is essential for regulation of HIS3 transcription even in the presence of other copies of the TGACTC sequence. None of the copies of the repeat seem to be uniquely required for normal HIS3 basal expression. However, in the absence of all other repeat sequences upstream, deletion of the TGACTC sequence at −99 leads to a substantial reduction in basal expression as well as a complete loss of regulation. This result suggests that the TGACTC element contributes to basal HIS3 expression. Surprisingly, this contribution can be observed in a gcncd mutant as well as in wild-type cells, suggesting that its effect is mediated by a transcriptional factor other than GCN4 (106).

The HIS3 repeat at −99 is necessary for derepression but does not appear to be sufficient for a wide-type degree of regulation. Deletions from −442 to −137 and −442 to −130 which leave the −99 repeat intact reduce the derepression ratio from five- to twofold without affecting basal expres-

sion. These deletions remove one of the imperfect copies of the repeat found at −142: 5′ TGCCCTC 3′. As discussed below, the substitution of the A residue with a C residue in the TGACTC element is expected to abolish the ability of the repeat to mediate transcriptional control by GCN4 (33). Moreover, no binding of GCN4 protein to the −142 sequence has been observed in vitro (40, 42). To explain the requirement for this imperfect copy of the repeat for maximal derepression, it has been suggested that a protein other than GCN4 binds to the TGCCCTC sequence and contributes to transcriptional activation of HIS3. This hypothetical protein is also proposed to explain the GCN4-independent contribution of the −99 TGACTC sequence to HIS3 basal expression, just mentioned above (106). Since a gcncd mutation or removal of the −99 TGACTC repeat is each sufficient to completely abolish HIS3 regulation, it appears necessary that the hypothetical protein is able to influence expression of the HIS3 gene at the −142 TGCCCTC sequence only when the GCN4 product also acts at the −99 repeat. By contrast, the hypothetical protein must be capable of acting at the −99 TGACTC sequence to affect basal expression in the absence of GCN4 protein.

Saturation mutagenesis of the HIS3 −99 TGACTC sequence has been conducted with the result that 15 of the
possible 18 single-base-pair substitution mutations in this sequence severely impair or completely abolish HIS3 derepression (33). Some of these mutations are shown in Fig. 2. In agreement with the above observations that different copies of the TGACTC sequence function to different extents as general control regulatory elements, it was found that certain mutations in sequences flanking the −99 repeat have detrimental effects on regulation that are almost as serious as mutations in the TGACTC core (Fig. 2). Based on analysis of many different mutations in the HIS3 −99 sequence, plus a compilation of TGACTC sequences at multiple genes under the general control, an expanded consensus sequence was proposed: 5′ RRTGACTCAATTT 3′, where R designates a purine. The three positions immediately downstream of the TGACTC core appear to be the most important among the nucleotides flanking the core. In fact, a mutation that altered the HIS3 −99 sequence from 5′ GATGACT CTTTT 3′ to 5′ GATGACTCAATT 3′ was the only mutation in this regulatory element that led to greater derepression of HIS3 expression than the wild-type element (Fig. 2). This substitution increases the twofold rotational symmetry of the −99 repeat, prompting the suggestion that GCN4 protein binds as a dimer to two half-sites of sequence 5′ ATGA 3′ separated by a single C nucleotide (33). The effects of these substitution mutations on HIS3 regulation in vivo are completely correlated with their effects on the binding affinity of GCN4 protein in vitro (33).

Similar to the case of the HIS4 gene, deletions that remove the sequence 5′ TATAA 3′, beginning at position −46, abolish HIS3 derepression as effectively as a deletion of the TGACTC sequence at −99. Some of the TATA box deletions have little effect on HIS3 basal expression (100). Assuming that the TATA box mutations alter the mechanism for HIS3 mRNA 5′-end selection, these results add support to the idea that transcriptional regulation by the GCN4 protein is dependent upon the presence of a TATA box downstream. Another finding that supports this general idea is that two transcripts are made from the HIS3 gene: one initiates at +1; the second initiates at +12. Surprisingly, only the +12 transcript is subject to the general control (103, 106). Sequences between −52 and −32 are required for the production of both transcripts (101); however, it is not known at present whether initiation of these two transcripts is mediated by different TATA sequences, only one being GCN4 responsive.

Deletions of a T-rich sequence between −130 and −110 at HIS3 lead to fivefold-lower HIS3 basal expression and leave regulation by the general control intact. Expression of both the +1 and +12 transcripts is reduced by such deletions (102, 106). This finding has led to the idea that the T-rich sequence functions as a GCN4-independent basal control element in the HIS3 promoter (Fig. 1). A deletion of the T-rich sequence also reduces expression of the PET56 gene which is transcribed divergently from HIS3 beginning at a site 195 bp upstream of the initiation site of the HIS3 +1 transcript (102). Thus, this sequence appears to act as a bidirectional control element to activate expression of both HIS3 and PET56. There is evidence that a similar sequence plays the same role in the expression of the yeast DED1 gene (102) and also in the increased promoter activity that results from certain cis-acting mutations at the yeast ADIR2 gene (92). It remains to be determined whether the polydeoxothyrimidylate sequence can function autonomously as a UAS element. Although the BAS2 binding site at HIS4 is very rich in adenine plus thymine, there is little primary sequence homology between this basal control element and the polydeoxythymidylic tract at HIS3 (5). It will be of interest to determine what role, if any, BAS2 plays in HIS3 basal expression.

**TRP5 Regulatory Region**

Deletion analysis of the TRP5 promoter has led to conclusions about the requirements for transcriptional regulation by GCN4 very similar to those reached from the work done on HIS4 and HIS3. The TRP5 5′-noncoding region contains no perfect copies of the TGACTC sequence; however, there are two copies of the most highly conserved portion of the core, 5′ TGACTC 3′, at −187 and −62 (121) (Fig. 1). Moreover, the −62 copy matches the expanded consensus sequence mentioned above at all other positions flanking the TGACTC core. Deletions from −454 to −188 leave all three copies of the repeat intact and have no effect on TRP5 expression in either repressing or derepressing conditions. By contrast, a deletion extending only 5 bp more downstream that removes the TGACTC sequence at −187 results in twofold-lower expression under both growth conditions. Insertion of an oligonucleotide containing the sequence 5′ GTGACTC 3′ at the junction of the −183 deletion restores expression to wild-type levels. These results suggest that the −189 repeat contributes to the efficiency of expression in both repressing and derepressing conditions. A deletion extending further downstream to position −149 has a dramatic effect on TRP5 expression under both growth conditions, but more so in repressing conditions, resulting in a derepression ratio greater than that observed for the wild-type gene (73, 74). The −183 to −149 interval contains no TGACTC sequences, suggesting that an important basal promoter element maps there (Fig. 1). This interval contains no T-rich sequence like the one found at HIS3, nor does it show any obvious homology to the BAS2 binding site at HIS4. BAS2 dependence may only apply to the HIS genes.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Sequence</th>
<th>Derepression Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>−162</td>
<td>GGA</td>
<td>TGACTC</td>
</tr>
<tr>
<td>−163</td>
<td>GTA</td>
<td>TGACTC</td>
</tr>
<tr>
<td>−164</td>
<td>GGC</td>
<td>TGACTC</td>
</tr>
<tr>
<td>−165</td>
<td>GGA</td>
<td>AGACTC</td>
</tr>
<tr>
<td>−166</td>
<td>GGA</td>
<td>TATGTCT</td>
</tr>
<tr>
<td>−169</td>
<td>GGA</td>
<td>TACGCTC</td>
</tr>
<tr>
<td>−168</td>
<td>GGA</td>
<td>TACGCTC</td>
</tr>
<tr>
<td>−169</td>
<td>GGA</td>
<td>TGACTA</td>
</tr>
<tr>
<td>−188</td>
<td>GGA</td>
<td>TGACTC</td>
</tr>
<tr>
<td>−189</td>
<td>GGA</td>
<td>TGACTA</td>
</tr>
<tr>
<td>−172</td>
<td>GGA</td>
<td>TGCTT</td>
</tr>
<tr>
<td>−173</td>
<td>GGA</td>
<td>TGACTT</td>
</tr>
<tr>
<td>−186</td>
<td>GGA</td>
<td>TGCC</td>
</tr>
<tr>
<td>−Δ63</td>
<td>GGA</td>
<td>GAGCTC</td>
</tr>
<tr>
<td>−145</td>
<td>GGA</td>
<td>TGACTC</td>
</tr>
<tr>
<td>−161</td>
<td>GGA</td>
<td>TGACTC</td>
</tr>
</tbody>
</table>

**FIG. 2.** Selected mutant alleles from saturation mutagenesis of the −99 HIS3 TGACTC sequence. his3−162 contains the wild-type −99 element. Base changes from this sequence are indicated by underlining. The derepression ratio is the ratio of the +12 (constitutive) HIS3 transcript in a gpd1 mutant.
since histidine is the only amino acid requirement observed for gcn4 bas2 double mutants (5).

When a deletion was constructed that removes the -62 copy of the TGACTC sequence in addition to all other TRP5 sequences upstream, TRP5 derepression was completely abolished. A 23-bp internal deletion of the -62 repeat lowers expression under both repressing and derepressing conditions, the result being a 50% drop in the derepression ratio from 3.6 to 2.4 (73). Therefore, both the -189 and the -62 TGACTC sequences appear to be necessary for a wild-type level of derepression at TRP5. Each sequence is sufficient for regulation by the general control system at a reduced efficiency.

As for the HIS4 and HIS3 genes, deletion of a downstream sequence with TATA box homology abolishes TRP5 derepression without completely destroying gene expression. The sequence 5′ ATATATA 3′ is found in the deletion interval -49 to -29 required for both normal TRP5 basal expression and derepression by the general control system (73). However, it should be noted that the -49 to -29 interval contains a sequence related to the TGACTC element in addition to a TATA box homology (Fig. 1). The contribution of the former to regulation of TRP5 expression cannot be assessed at present.

**HIS5 Regulatory Region**

HIS5 mRNA levels increase roughly threefold in response to amino acid starvation. This response, as well as the amount of HIS5 basal expression, is unaffected by deletion of the 5′-noncoding DNA between -579 and -346. By contrast, expression under both conditions is reduced by extending the deletion interval further to -241. The -346 to -241 segment contains three sequences with close homology to the TGACTC sequence (Fig. 1). Deletion further downstream to -172 leads to very low, unregulated expression. The -241 to -172 interval also contains a TGACTC sequence (Fig. 1). Finally, a linker insertion into a TGACTC sequence at -165 results in a promoter with almost normal basal expression but one incapable of derepression in starvation conditions (83). These results strongly suggest that the -165 copy of the TGACTC sequence at HIS5 is absolutely required for derepression, even in the presence of upstream copies of the repeat. The data are consistent with the idea that the upstream copies of the repeat at HIS5 contribute to the efficiency of derepression.

**ARG3 Regulatory Region**

ARG3 is an interesting case because this gene is subject to dual regulation by the general control and by the ARG genes that mediate arginine-specific repression. Addition of arginine to minimal medium results in a 10-fold repression of ARG3 enzyme activity. Starvation for arginine leads to 10-fold derepression of ARG3 expression. The latter response is GCN4 dependent; the former is unaffected by general control regulatory mutations. Deletion analysis reveals that the two regulatory mechanisms controlling ARG3 expression depend upon sequences located in different parts of the ARG3 5′-noncoding region. Deletion of sequences upstream from position -282 abolishes derepression mediated by GCN4 in response to arginine starvation. This undoubtedly occurs because the only two copies of the TGACTC sequence at ARG3 are located upstream of position -282 (Fig. 1). By contrast, basal expression and arginine-specific repression remain intact in the absence of all ARG3 sequences upstream from -170. As mentioned in an earlier section, two cis-acting mutations that abolish arginine repression have been shown to be substitutions in the 5′-noncoding sequences at -80 and -46, between the putative TATA box and the 5′ end of the ARG3 transcription unit (Fig. 1). Interestingly, these mutations seem to affect translation of ARG3 mRNA, not its synthesis or 5′ end (16, 67).

**Other Structural Genes**

Every other structural gene whose transcription is known to be governed by the general control contains at least one sequence closely related to the expanded consensus: 5′ RRT GACTCATTT 3′ (Fig. 1). Like TRP5, the ILV2 gene has no sequences that conserve all six positions of the TGACTC core; however, it contains a sequence that matches the 12-bp expanded consensus at every other position. The ILV2 sequence is known to be an efficient binding site for GCN4 protein in vitro (4). Based on the results of the mutational analysis summarized above and the in vitro GCN4-binding analysis discussed below, it appears that the ARG3 regulatory system can be divided into two or three potential GCN4-binding sites can be found. At present, there is no simple relationship between the number of binding sites and the efficiency of expression, although analysis of in vitro constructed promoter elements suggests that multiple copies of the repeat lead to greater expression in derepressing conditions (40, 74). Finally, it is of interest that potential GCN4-binding sites occur at LEU1 and LEU2, genes not believed to be governed by the general control. As discussed above, these two genes are subject to leucine-specific repression, raising the possibility that GCN4-mediated derepression can only be observed in conditions of leucine starvation.

**trans-ACTING REGULATORY FACTORS IN GENERAL AMINO ACID CONTROL**

**GCN4 Encodes the Most Direct Positive Regulator in the General Control System**

Mutations have been isolated in multiple unlinked genes that impair the regulation of enzymes subject to general amino acid control. Based on their phenotypes, these mutations fall into two classes. A recessive mutation in any one of the nine GCN genes blocks the derepression of enzymes regulated by the general control (Table 2). As a result, gcn mutations are more sensitive than wild-type strains to a variety of conditions in which amino acid biosynthesis is inhibited. These conditions include (i) growth in the presence of amino acid analogs; (ii) growth on minimal medium in the presence of a leaky mutation in a structural gene for a biosynthetic enzyme or a leaky defect in an aminoacyl-tRNA synthetase; (iii) growth in the presence of an amino acid supplement that inhibits the biosynthesis of another amino acid because a shared step in their biosynthetic pathways is negatively regulated by the supplement. In many cases, some genes appear to have more than one of these conditions at once. Even so, the last of these conditions, involving the effects of certain amino acid imbalances, is probably closest to the situation in nature in which the general control response plays an important role. Because the gcn mutations are recessive, their phenotype indicates that the GCN products are positive regulators in the general control system. All nine GCN genes have been shown to control expression of structural genes under the general control by regulating the
TABLE 2. Mutations in trans-acting regulatory genes involved in general amino acid control

<table>
<thead>
<tr>
<th>Locus</th>
<th>Current designation</th>
<th>Previous designation</th>
<th>Phenotype&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Map position (reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCN1</td>
<td>gcn1-1</td>
<td>ndr1-1 (94)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gcn1-2</td>
<td>ndr1-2 (94)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gcn1-201</td>
<td>aas103-13 (20)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td>GCN2</td>
<td>gcn2-1</td>
<td>ndr2-1 (94)</td>
<td>Nonderepressible</td>
<td>IV, R (39)</td>
</tr>
<tr>
<td></td>
<td>gcn2-101</td>
<td>aas1-1 (118)</td>
<td>Nonderepressible</td>
<td>XI, R (39)</td>
</tr>
<tr>
<td></td>
<td>gcns-101</td>
<td>aas3-1 (39)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gcd2-1</td>
<td>cdrl-1 (70, 81)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td>XV, R&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>gcd2-1</td>
<td>cdrl-1 (79, 81)</td>
<td>Constitutively derepressed, Slg&lt;sup&gt;−&lt;/sup&gt;</td>
<td>VII, R (85)</td>
</tr>
<tr>
<td>GCN4</td>
<td>gcn4-1</td>
<td>aas101-1 (20)</td>
<td>Nonderepressible</td>
<td>V, R (20, 39)</td>
</tr>
<tr>
<td></td>
<td>gcn4-101</td>
<td>aas3-1 (39)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gcn4-102</td>
<td>aas3-2 (39)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gcd4-600</td>
<td>arg9 (39)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td>GCN5</td>
<td>gcns-201</td>
<td>aas104-1 (20)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td>GCN6</td>
<td>gcns-201 (30)</td>
<td>aas1-2 (30)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td>GCN7</td>
<td>gcns-201 (30)</td>
<td>aas1-2 (30)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td>GCN8</td>
<td>gcns-201 (30)</td>
<td>aas1-2 (30)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td>GCN9</td>
<td>gcns-201 (30)</td>
<td>aas1-2 (30)</td>
<td>Nonderepressible</td>
<td></td>
</tr>
<tr>
<td>GCD1</td>
<td>gcd1-101</td>
<td>tra3-1 (118)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD2</td>
<td>gcd2-1</td>
<td>cdr1-1 (70, 81)</td>
<td>Constitutively derepressed, Slg&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD3</td>
<td>gdc3-1</td>
<td>cd2-1 (79, 81)</td>
<td>Constitutively derepressed, Slg&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD4</td>
<td>gdc4-1</td>
<td>gdc4-1 (81)</td>
<td>Constitutively derepressed, Slg&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD5</td>
<td>gdc5-1</td>
<td>gdc5-1 (81)</td>
<td>Constitutively derepressed, Slg&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD6</td>
<td>gdc6-1</td>
<td>gdc6-1 (81)</td>
<td>Constitutively derepressed, Slg&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD7</td>
<td>gdc7-201</td>
<td>gdc2-201 (77)</td>
<td>Constitutively derepressed</td>
<td></td>
</tr>
<tr>
<td>GCD8</td>
<td>gdc8-201</td>
<td>gdc5-201 (30)</td>
<td>Constitutively derepressed</td>
<td></td>
</tr>
<tr>
<td>GCD10</td>
<td>gcd10-101</td>
<td>gdc10-502 (32)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD11</td>
<td>gcd11-501 (32)</td>
<td>gdc10-504 (32)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gdc11-504 (32)</td>
<td>gdc10-504 (32)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gdc11-508 (32)</td>
<td>gdc10-508 (32)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD12</td>
<td>gdc12-501 (32)</td>
<td>gdc12-501 (32)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GCD13</td>
<td>gdc13-501 (32)</td>
<td>gdc13-502 (32)</td>
<td>Constitutively derepressed, Tsm&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Tsm<sup>−</sup> = Temperature sensitive; Slg<sup>−</sup> = slow growth.

<sup>b</sup> Hinnebusch, unpublished observations. The GCN4-501 allele is a deletion similar to GCN4<sup>Δ1</sup> in which sequences of the GCN4 mRNA 5' leader containing the short ORFs have been omitted. The endpoints of the deletions are the same as those described in reference 35 for GCN4A-lacZ.


abundance of their mRNA products (20, 21, 30, 40, 61, 106; E. M. Hannig and A. G. Hinnebusch, unpublished observations).

A recessive mutation in any of 12 GCD genes results in constitutive derepression of enzymes subject to the general control (Table 2). Therefore, GCD gene products are negative effectors in the general control system. As described above, derepression of enzymes under the general control can also be produced by leaky mutations that interfere with amino acid biosynthesis or aminoacylation of tRNA. These conditions are unlikely to be explained by the Gcd<sup>−</sup> phenotype of most gcd mutants for the following reasons. First, the gcd1-101 mutation leads to derepression of enzyme levels even in amino acid-complete medium (98, 106, 118). The same observation has been made for mutations in GCD2-GCD6 (82), GCD7 (first described as GCD3) (77), GCD8 (first described as GCD5) (30), and GCD10-GCD13 (A. Hinnebusch, unpublished observations). Second, levels of all 20 aminoacylated tRNAs were found to be normal in a gcd1-101 strain (34). Third, the derepressing effect of the ils1-1 RNA synthetase mutation is reversed by a gcn1 mutation (80). By contrast, as discussed more fully below, most gcd mutations are not suppressible by a gcn1 mutation. Taken together, these observations suggest that the majority of the GCD genes are not involved in amino acid biosynthesis or RNA aminoacylation. Rather, they appear to function in a more direct fashion in the regulation of gene expression. GCD1, GCD7, and GCD10-GCD13 have been shown to regulate expression of amino acid biosynthetic genes by controlling the steady-state amounts of their mRNA products (19, 32, 34, 40, 69, 77). By contrast, it was proposed that GCD8 regulates structural gene expression through a posttranscriptional mechanism (30). The latter makes it questionable whether GCD8 is a regulatory factor in the general control system.

The pet18 mutation, also known as tra5 (118), leads to derepression of enzymes under the general control, as do mitochondrial p<sup>−</sup> mutations (G. Lucchini and G. R. Fink, unpublished observations). In addition, incubation at 37°C has been reported to derepress expression of two arginine
biosynthetic enzymes under the general control (64). It is possible that these conditions lead to enzyme derepression indirectly by interfering with the efficiency of amino acid biosynthesis or aminoclaylation of tRNA. If so, their derepression effects should be dependent upon the functions of GCN1, GCN2, and GCN3. This dependence has not yet been examined.

The GCN genes were divided into two classes based upon interactions between gen and ged mutations. All ged gdc double mutants examined thus far have a Gcn− phenotype; i.e., they are unable to derepress enzymes subject to the general control (32, 39, 77, 81). This result implies that GCN4 positive regulatory function is required for derepression even in the absence of the GCD negative regulatory factors. That the phenotype of ged mutations is completely dependent upon the presence of a functional allele of GCN4 suggests that GCD gene products act as negative effectors indirectly by antagonism or repression of the GCN4 protein (Fig. 3). This relationship was demonstrated for all known GCD genes except GCD5 and GCD8. The ged gdc5 double mutation appears to be lethal (81).

In contrast to the interactions observed between gen4 and ged mutations, it was shown that gen1, gen2, and gen3 mutations are suppressed by mutations in GCD1, GCD2, GCD3, GCD4, GCD6, and GCD7 (32, 39, 77, 81, 118). These interactions suggest that the GCN1, GCN2, and GCD3 gene products function indirectly as positive regulators by antagonism or repression of GCD gene products (Fig. 3). Mutations in GCN6-GCN9 suppress the ged8-201 mutation and thus exhibit the same interaction seen between GCN4 and other GCD genes (30). However, the relationship between mutations in GCD8 and GCN6-GCN9 with mutations in GCN4 and all other known GCD genes has not been determined. Therefore, it is not clear whether GCN6 through GCN9 function more directly than any other GCD factors besides GCD8. It is possible that the ged8-201 mutation can be suppressed by any gen mutation.

The interactions between gen and ged mutations suggest that the GCN4 product is a direct positive regulator of genes subject to the general control and that GCN1, GCN2, and GCN3 function indirectly by positive regulation of GCN4 (39) (Fig. 3). This conclusion is supported by several lines of evidence. First, it was shown that an increase in the dosage of the GCN4 gene can partially bypass the requirement for the GCN1, GCN2, and GCN3 gene products for derepressed expression of structural genes under the general control (39, 61). Second, it was demonstrated that expression of β-galactosidase activity from a GCN4-lacZ translational fusion derepresses ~10-fold in response to amino acid starvation in wild-type cells (35, 108). This derepression is blocked in gen2 and gen3 mutants (Table 3), suggesting that the products of GCN2 and GCN3 are positive regulators of GCN4 expression (35, 36). Also in accord with the model shown in Fig. 3, it was demonstrated that gcd1 and gcd10-gcd13 mutations lead to constitutively derepressed expression of GCN4-lacZ enzyme activity in gen2 and gen3 mutants (e.g., Table 3) (32, 36). As described in great detail below, GCN4 expression is regulated by these trans-acting factors at the translational level.

One way in which the regulatory model in Fig. 3 may be incomplete is suggested by the result that the expression of the GCN2 transcript in starvation conditions requires the GCN4 gene product (21). This finding implies that the general control regulatory hierarchy is actually a closed circuit. Derepression of GCN2 mRNA might occur to provide self-amplification of GCN4 expression. Alternatively, the response may be necessary to maintain a constant level of GCN2 protein synthesis in the face of increasing amino acid starvation. Given that GCN4 acts directly as a transcriptional factor to derepress expression of structural genes subject to the general control, it seems likely that it also stimulates expression of GCN2 mRNA by binding in the 5′-noncoding sequences at GCN2. In fact, several copies of the GCN4-binding site are present at GCN2 (B. Hauge, personal communication).

### GCN4 Protein Binds to the TGACTC Regulatory Sequences

The prediction that the GCN4 protein acts directly to control the transcription of structural genes subject to general control was confirmed at the biochemical level by showing that GCN4 is a DNA-binding protein that interacts specifically with TGACTC regulatory sequences found in the 5′-noncoding regions of genes under its control. This was first demonstrated with GCN4 protein synthesized in vitro from a cell-free translation system prepared from rabbit reticulocytes. By an electrophoretic gel mobility shift assay, it was shown that DNA fragments containing the 5′-noncoding sequences of HIS3, ARG4, TRP5, and HIS4 are each capable of altering the electrophoretic mobility of radiolabeled GCN4 protein, whereas similar fragments from five genes not subject to the general control have no such effect on GCN4 electrophoretic mobility. The HIS3 sequences responsible for the mobility shift were mapped to the ~104 to ~83 interval of the HIS3 promoter. The deoxyribonucleic acid "footprinting" technique was used subsequently to demonstrate that the GCN4 protein binds to the TGACTC sequence in this interval. Binding to other copies of the 99 TGACTC sequence found in the HIS3 5′-noncoding region was not detected, suggesting that nucleotides flanking the

---

**Table 3. Expression of GCN4-lacZ fusions in general control regulatory mutants**

<table>
<thead>
<tr>
<th>Strain</th>
<th>GCN4-lacZ</th>
<th>GCN4-lacZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>DR</td>
</tr>
<tr>
<td>Wild type</td>
<td>10</td>
<td>105</td>
</tr>
<tr>
<td>gcd2-1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>gcd3-102</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>gcd1-101</td>
<td>360</td>
<td>270</td>
</tr>
<tr>
<td>gcd2-1 gcd1-101</td>
<td>250</td>
<td>240</td>
</tr>
</tbody>
</table>

* GCN4-lacZ is a fusion construct that differs from the wild-type GCN4-lacZ fusion shown here by a deletion of ~240 bp in the leader of the GCN4 transcription unit. The deleted sequences contain the four upstream ORFs. R and DR refer, respectively, to repressing (minimal medium) or derepressing (minimal medium containing 10 mM 3-aminotriazole) conditions.

---

**Figure 3.** Hierarchy of regulatory factors in general amino acid control.
TGACTC repeat contribute to the efficiency of GCN4 binding (42).

These findings were extended by using a partially purified preparation of GCN4 protein synthesized in Escherichia coli. Deoxyribonucleic acid footprinting was used to examine the interactions between GCN4 protein and the TGACTC sequences found at HIS4, HIS3, ILV1, and ILV2. With one exception, all sequences in the 5'-noncoding regions of these genes containing at least four of the five nucleotides in the TGACTC core sequence bind GCN4 protein if it is present at a high enough concentration. The relative binding affinities for the different sequences span a 50-fold range (Fig. 4). In agreement with the effects of mutations in the HIS3 -99 TGACTC sequence on HIS3 derepression in vivo and GCN4 binding in vitro (33), it was generally found that those sequences that conform most closely to the consensus sequence 5' RRTGACTCATT 3' exhibit the strongest binding to GCN4 protein. It is noteworthy that, for each of the four genes tested, at least one copy of the repeat exhibited very efficient binding of GCN4 protein (42). This is of particular importance for ILV2 since only a single sequence related to the TGACTC repeat occurs at this gene.

The amino acid residues of the GCN4 protein involved in sequence-specific binding to DNA and activation of transcription have been defined by functional analysis of the protein products of gcn4 deletion alleles. Deletion derivatives synthesized in vitro translation that retain 60 or more carboxyl-terminal amino acids exhibit DNA binding in the electrophoretic mobility shift assay mentioned above. Derivatives containing 114 or more of the carboxyl-terminal amino acids bind to DNA and appear to be identical to the wild-type protein in the specificity of their interaction with the TGACTC sequence. Thus, the carboxyl-terminal one-third of the protein appears to be sufficient for highly specific binding of GCN4 protein to the TGACTC sequence (43) (Fig. 5).

The mobility shift assay was also used to demonstrate that GCN4 protein probably binds to DNA as a dimer. This conclusion was reached by showing that a mixture of wild-type and an amino-terminal truncated GCN4 protein that is competent for DNA binding gives rise to three DNA-protein complexes with different electrophoretic mobilities. The fastest- and slowest-moving complexes contain only wild-type or truncated GCN4 protein, respectively. The complex with intermediate mobility is a mixture of the two proteins. The molar ratios of the three complexes (fast/intermediate/slow, 1.1:2.2:0.8) are consistent with the idea that GCN4 normally binds to DNA as a dimer. The carboxyl-terminal 60 residues minimally required for DNA binding in this assay are also required for formation of the intermediate-mobility complex (44). This result suggests a close relationship between the amino acid residues required for protein dimerization and DNA binding.

Interestingly, the carboxyl-terminal 66 amino acids of GCN4 are homologous to the carboxyl terminus of an avian viral oncoprotein known as jun. A lesser homology is observed between the same segment of GCN4 and the fos oncoprotein product. The homologous segments in the three proteins are each predicted to be a-helix-permissive domains (113). It was recently shown that the homologous segment of the jun protein could partially substitute for the carboxyl terminus of GCN4 and restore a degree of derepression of HIS3 expression in a gcn4 deletion strain. The amount of derepression conferred depended on the exact sequences present.
at the HIS3 -99 GCN4-binding site in a fashion similar to that observed for the corresponding construct containing the authentic GCN4 carboxyl terminus (104). These results suggest that the carboxyl-terminal domains of these proteins contain an ancient and highly conserved structural motif for DNA binding.

Removal of the amino-terminal 106 residues of the GCN4 protein does not eliminate its ability to complement a gcna chromosomal deletion for derepression of HIS3 expression. Although it is not known whether the steady-state amount of this truncated protein is comparable to that of the native GCN4 protein, it seems likely from this result that the carboxyl-terminal 175 residues of GCN4 are sufficient for transcriptional activation in vivo. By contrast, the carboxyl-terminal 163 residues, or fewer, are insufficient for GCN4 positive regulatory function when assayed by the same complementation assay. The requirement for 175 carboxyl-terminal residues for in vivo function suggests that the DNA-binding activity of GCN4 is insufficient for its role as a positive regulator of transcription. In fact, smaller derivatives containing 141 or fewer amino acids appear to actually inhibit expression of amino acid biosynthetic genes in the gcna deletion strain. Apparently, the binding of nonfunctional GCN4 protein molecules to TGCAC sequences interferes with the function of basal promoter elements at amino acid biosynthetic genes (43). Similar effects have been observed for mutations in other positive regulatory proteins in yeasts and bacteria (41, 50). The inhibitory effect on amino acid biosynthesis conferred by these nonfunctional gcna alleles is the only evidence, thus far, to indicate that the loss of function observed for deletions of more than 175 residues from the amino terminus of the GCN4 protein is not simply the result of protein instability or a failure to localize in the nucleus.

The deletion alleles described above locate the amino-terminal boundary of a domain required for transcriptional activation, at least in the absence of all other amino acids upstream, between the 107th and 119th amino acids. To map the 3' boundary of this domain, sequences encoding the lexA DNA-binding domain were fused to the amino-terminal end of GCN4 and deletions were made from the 3' end of the GCN4 coding sequences. The resulting constructs were tested in yeast cells for positive regulation of a lacZ gene containing a TATA box and a lexA binding site upstream. Removal of the first 10 GCN4 amino acids from the carboxyl terminus of the lexA-GCN4 hybrid protein reduced positive regulation of lacZ expression by about fivefold. Deletions of an additional 150 amino acids had little further effect on function. More extensive deletions that remove residues upstream of amino acid 120 in the GCN4 protein sequence completely abolished positive regulation. These results suggest that two distinct domains contribute to positive regulation of transcription by GCN4. One is located at the extreme carboxyl terminus and is coincident with residues required for DNA binding. The other is located upstream of amino acid 120. Combining these results with those obtained from the amino-terminal deletions suggests that a positive regulatory domain maps in the vicinity of residues 107 to 120 (43) (Fig. 5). It remains to be determined whether this small segment is sufficient to confer positive regulation of transcription when attached to a DNA-binding domain.

Five of the 14 amino acids in the 107 to 120 interval are acidic and only 1 is basic. In addition, this segment is contained within a larger interval of 60 amino acids made up of roughly two-thirds acidic residues and only two basic residues (Fig. 5). The composition of this segment raises the possibility that acidic residues are an important feature of the positive regulatory domain of GCN4. In support of this idea, it was recently shown that the yeast GAL4 transcriptional activator protein also contains two acidic domains, each of which is sufficient to confer positive regulatory function when fused to the DNA-binding domain (62). There is no obvious primary sequence homology between these acidic regions, suggesting that negative charge or a structural feature promoted by acidic amino acids is the important feature for their ability to stimulate transcription. Presumably, the acidic domains make contacts with RNA polymerase or some other component of the transcriptional complex, such as a TATA box-binding factor (103).

**GCN and GCD Factors That Function as Translational Regulators of GCN4 Expression**

Results from a variety of experiments suggest that GCN4 expression is regulated at the translational level in response to amino acid availability. One line of evidence comes from the lack of correlation between the effects of gcen and gcdn mutations on the expression of GCN4 protein and GCN4 mRNA. gcen2 and gcdn3 mutations lead to increased amounts of GCN4 and GCN4-lacZ mRNA in starvation conditions relative to the wild-type GCN2 and GCD3 genes (21, 35, 36, 75). In the same conditions, these mutations block the derepression of GCN4-lacZ enzyme activity observed in wild-type cells (35, 36). This discrepancy suggests that the GCN2 and GCD3 gene products stimulate translation of GCN4 mRNA in starvation conditions.

In ged1-101 strains, GCN4 and GCN4-lacZ mRNA levels are constitutively elevated relative to their amounts in wild-type cells. Measurements of GCN4-lacZ enzyme activity reveal an increase in the fusion protein level five- to tenfold higher than the increase observed in the fusion transcript level. When galactose is used in place of glucose as the carbon source in such experiments, the ged1-101 mutation has little effect on the amount of GCN4-lacZ fusion transcript; however, fusion enzyme activity remains derepressed at a level ~20-fold higher in a ged1-101 strain than in GCD cells. Similar results were obtained for ged10-gcd13 mutations (36, 75). These discrepancies suggest that GCD1 and GCD10-GCD13 act, at least in part, to repress translation of GCN4 mRNA in nonstarvation conditions.

Another indication of increased translational efficiency of GCN4 mRNA in gcd1 mutants came from a comparison of the amount of GCN4 protein produced in a GCD strain containing the GCN4 gene in high copy versus gcd1 mutants containing a single copy of the GCN4 gene. GCN4 mRNA level in the GCD strain is several-fold higher than in the ged1 and gcd10-gcd13 mutants; however, more GCN4 protein is produced in the gcd1 mutants than in the GCD strain. (The amount of GCN4 protein produced was inferred from the level of derepression of the HIS4 gene, a target of GCN4 transcriptional activation.) The implication drawn from this result was that GCN4 mRNA has a higher translational efficiency in the gcd mutant cells than in wild-type cells (75). These observations help to explain why overexpression of GCN4 mRNA is insufficient to completely bypass the requirement for GCN2 and GCN3 function for derepression of genes under the general control (61). The GCN4 products are needed for efficient translation, not efficient production of the GCN4 transcript.

If GCN4 expression is regulated by GCN and GCD factors posttranscriptionally, then substitution of the GCN4 promoter that of another yeast gene might not interfere
with the regulation of GCN4 expression. To address this possibility, the transcriptional activation site (UAS) of the GAL1 gene was inserted 53 nucleotides upstream from the 5′ end of the GCN4-lacZ transcription unit in place of the GCN4 5′-noncoding sequences. As expected, the resulting construct is transcribed only in medium containing galactose as a carbon source. In spite of certain differences in the initiation site and steady-state levels of the mRNAs produced by the wild-type and GAL1 UAS-driven fusion constructs, expression of fusion enzyme activity from the two genes is very similar: Enzyme expression is derepressible in wild-type cells in response to starvation, nonderepressible in gal2 cells, and constitutively derepressed in gcl1 cells (36). Positive regulation by GCN3 and negative regulation by GCD10-GCD13 are also retained following the replacement of the GCN4 promoter with the GAL1 UAS (75; Hannig and Hinnebusch, unpublished observations).

The key to understanding the translational control of GCN4 expression was provided by the sequence organization of GCN4 mRNA. The GCN4 transcript is ~1,500 nucleotides in length. A protein-coding sequence of 281 codons is located near the 3′ end of this transcript unit, leaving ~600 bp in the 5′ leader. Within the leader are four AUG codons, each of which is followed by one or two sense codons before an in-frame termination codon is reached (Fig. 6). Deletion analysis has demonstrated that the 3′ half of the 281-amino-acid coding sequence is required for complementation of a gcn4 mutation. By contrast, the entire 5′ leader can be deleted without loss of GCN4 function. No processing of the 1,500-nucleotide transcript has been detected in wild-type, gal2, and gcl1 cells in starvation or nonstarvation conditions (35, 36, 108). Taken together, these findings strongly suggest that the GCN4 protein is translated from a mRNA containing four upstream AUG codons.

The presence of upstream AUG codons is rare in yeast mRNAs (15a), and it was demonstrated that introduction of such sequences into the untranslated leader of the yeast CYC1 transcript leads to substantial reduction in expression of the CYC1 gene product (95, 123). Similar observations have been made for several mammalian mRNAs (48, 55, 57, 59, 60), leading to the idea that translation initiation at the 5′-proximal AUG codon is greatly preferred in eucaryotic cells and that reinitiation at downstream AUG codons occurs very inefficiently (56, 57, 59).

In view of these considerations, it was of great interest to find that an ~250-nucleotide deletion of GCN4 mRNA leader sequences encompassing the four upstream open reading frames (ORFs) results in constitutive derepression of structural genes under the general control. Using the GCN4-lacZ fusion to measure the effects of this deletion on GCN4 expression revealed a ~50-fold increase in fusion enzyme activity in wild-type cells grown in nonstarvation conditions. Moreover, amino acid starvation results in no further increase in the fusion enzyme level expressed from this fusion construct (Table 3). Because the deletion has only a small effect on the steady-state level of GCN4 and GCN4-lacZ transcripts, it was concluded that the leader sequences act in cis to repress translation of the GCN4 protein-coding sequences in nonstarvation conditions (35, 36, 108).

When introduced upstream from the authentic GCN4 protein-coding sequences, the above-mentioned leader deletion leads to derepression of genes under the general control in the absence of the positive regulators encoded by GCN1, GCN2, and GCN3 (108). The deletion also bypasses the requirement for these factors for efficient expression of GCN4-lacZ enzyme activity (Table 3). These findings are consistent with the idea that the products of GCN1, GCN2, and GCN3 function as positive regulators by antagonism of the negative effects of the leader sequences on GCN4 expression. After removal of the leader sequences, mutations in GCD1 and GCD10-GCD13 also have little effect on GCN4-lacZ expression (36, 75). The simplest explanation for the nonadditivity of the effects of the leader deletion and the gcn mutations is that the GCD factors promote the negative effects of the leader sequences on GCN4 expression.

An independent confirmation of the importance of the mRNA leader sequences in controlling GCN4 expression was provided by the demonstration that these sequences are sufficient to confer regulation typical of GCN4 upon a heterologous yeast transcript. Four GCN4 mRNA leader segments of different lengths, each containing the four upstream ORFs, were inserted into the mRNA leader of a yeast GAL1-lacZ fusion gene. The resulting hybrid transcripts, expressed under the control of the GALI promoter, lead to a pattern of β-galactosidase expression similar to that observed for a GCN4-lacZ fusion under the same conditions. All four leader segments greatly reduce GALI-lacZ enzyme expression under all conditions; however, for each hybrid transcript, expression of enzyme activity is derepressed 10-fold in wild-type cells in response to starvation and constitutively derepressed in a gcl1 strain (Fig. 7). Derepression in starvation conditions was blocked in GCD2 and GCD2 gcn3 strains and was constitutively elevated in gcd gcd2 gcd3 triple mutants containing mutations in GCD1 or GCD10-GCD13 (75).

Expression of the GAL1-GCN4 hybrid transcripts containing GCN4 leader segments of different lengths showed a similar degree of regulation. The only differences observed among these constructs were in the absolute levels of GALI-lacZ enzyme activity. The construct containing the largest leader segment, with ca. 475 of the 591 nucleotides of the GCN4 mRNA leader, gave rise to roughly four times as much expression as the construct bearing a 240-nucleotide segment containing the upstream ORFs and only a few flanking nucleotides (75). These results suggest that the flanking leader sequences of ORFs contribute to the efficiency of GCN4 expression and not to regulation. That the 240-nucleotide segment is sufficient for regulation narrows the target of the trans-acting factors involved in translational control to this small leader segment. We can also rule out a unique requirement for the 5′ end of the GCN4 transcript, the distance between the 5′ end and the upstream ORFs, the distance between the upstream ORFs and the GCN4 protein

![FIG. 6. GCN4 transcription unit and point mutations constructed in the upstream AUG codons. The wavy line represents the GCN4 transcript and the numbers indicate distance from the 5′ end. Solid boxes designate ORFs in the GCN4 transcription unit. The sequences of the upstream ORFs and the mutations generated in their AUG codons are indicated below the schematic.](http://mmbr.asm.org/DownloadedFrom/3633274848735/3633274848735.html)
initiation codon, the initiation codon itself, the protein-coding sequences, and the 3'-untranslated regions of GCN4 mRNA.

Translational Control of GCN4 Expression Requires the Upstream AUG Codons

Direct proof that the AUG codons in the GCN4 mRNA leader are required for translational control was provided by the construction of point mutations in these sequences. Removal of all four AUG codons by substitution mutations has the same effect on GCN4 expression as a deletion of all four upstream ORFs: expression is constitutively derepressed and independent of the GCN2 and GCDI gene products (Fig. 8). Consecutive removal of the upstream AUG codons beginning at the 3'-proximal upstream ORF (no. 4) and ending with the 5'-proximal upstream ORF (no. 1) results in a stepwise increase in GCN4-lacZ expression. In wild-type cells, this increase is more pronounced in derepressing conditions than in derepressing conditions. It is also more pronounced in gcn2 cells than in gedi cells. As a result, the derepression ratio decreases as each upstream AUG codon is eliminated until, upon removal of all four AUG codons, no derepression remains (Fig. 8, top). Removal of the upstream AUG codons has little or no effect on the steady-state level of the GCN4 and GCN4-lacZ transcripts. It can thus be concluded that the upstream AUG codons are required for translational control of GCN4 expression (76).

The stepwise increase in expression resulting from consecutive removal of the upstream AUG codons from the 3' end indicates that each AUG codon has the potential to inhibit GCN4 expression. However, comparison of the amount of GCN4-lacZ expression observed in different triple mutants containing only a single upstream AUG codon reveals that AUG codons 3 and 4 are far more inhibitory than AUG codons 1 and 2 (Fig. 9). In fact, AUG codon 4 is sufficient to repress GCN4-lacZ expression to the repressed level observed for the wild-type fusion gene containing all four upstream AUG codons. By contrast, the combination of AUG codons 1 and 2 permits a level of expression under repressing conditions that is 30-fold higher than that observed for the wild-type fusion (76).
The interactions between the upstream AUG codons in the GCN4 mRNA leader are best illustrated by comparing triple mutant alleles containing only single upstream AUG codons with the double mutants containing AUG codon 1 paired with AUG codon 3 or 4 (Fig. 9). As stated above, the triple mutant alleles containing only one AUG codon 3 or 4 exhibit low GCN4-lacZ expression in repressing conditions and, in this respect, closely resemble the wild-type allele. However, unlike the wild-type gene, these two alleles also exhibit low expression in derepressing conditions. Addition of the first AUG codon upstream of AUG codon 3 or 4 restores the ability to derepress GCN4-lacZ expression. In fact, these double mutant alleles containing only two of the four upstream AUG codons give a pattern of expression nearly identical to that of the wild-type gene (Fig. 9). Thus, the presence of the first AUG codon somehow provides a mechanism to overcome in starvation conditions the strong inhibitory effect exerted by AUG codons 3 and 4 (76).

It is important to note that, for those alleles containing only a single upstream AUG codon in the mRNA leader, GCN4-lacZ expression is almost completely unresponsive to a gcd1 mutation. For example, the strong inhibitory effect exerted by AUG codons 3 or 4 when present as single upstream AUG codons is virtually independent of GCD1 function (Fig. 9). The GCD1 product is needed to repress GCN4-lacZ expression only when multiple AUG codons are present in the leader. The requirement for GCD1 is strongest when AUG codon 1 is present upstream from AUG codon 3 or 4 (Fig. 9). These observations suggest that the primary role of GCD1 is to modulate the interaction between the upstream AUG codons. More specifically, GCD1 function seems to be required to prevent AUG codon 1 from suppressing the inhibitory effect of AUG codons 3 and 4 in nonstarvation conditions. According to the interactions between gen2 and gcd1 mutations described above, it is expected that GCD1 is prevented from carrying out its function in starvation conditions as the result of repression or antagonism by GCN2 (Fig. 10) (76).

Like mutations in GCD1, mutations in GCD10-GCD13 have much less effect on GCN4-lacZ expression when only a single upstream AUG codon is present in the leader compared with their effects on the wild-type GCN4-lacZ gene. This result implies that all five of these GCD factors function by modulating interactions between the multiple upstream AUG codons in the GCN4 mRNA leader (75). If, instead, one or more of these GCD gene products acted by a different mechanism, e.g., by repressing the amount of GCN4 mRNA produced in nonstarvation conditions, the magnitude of the effect of a mutation in that gene on GCN4 expression should be independent of the number of AUG codons present in the GCN4 mRNA leader.

Deletion mutations that remove different numbers of the upstream AUG codons from the GCN4-lacZ mRNA leader have also been constructed with somewhat different results. As in the case of point mutations in AUG codons 3 and 4, two different deletions of these sequences were found to increase GCN4 expression in wild-type cells under repressing conditions and in gcen2 mutants grown in repressing or derepressing conditions. This finding supports the idea that this segment of the leader contains a negative control element needed to maintain GCN4 expression at low levels in repressing conditions. At odds with the findings obtained from the point mutations is the result that deletions of AUG codons 3 and 4 lead to lower expression in a gcdl mutant compared with a wild-type strain. This finding prompted the suggestion that a positive control site required for maximum derepression also exists in the vicinity of AUG codons 3 and 4. Because expression from these deletion alleles is completely unresponsive to a gcen2 mutation, it was suggested that the positive site is acted upon by the GCN2 product (110). One difficulty with this explanation is that the deletions clearly remove a negative control site and, in the process, render expression in wild-type cells constitutively derepressed. Thus, the lack of GCN2 responsiveness could simply result from the absence of the negative control sequences that are normally antagonized by the GCN2 product in starvation conditions. In addition, that the deletions reduce GCN4 expression in a gcdl mutant but not in wild-type cells grown under derepressing conditions suggests the possibility that multiple repression mechanisms are impaired in the gcdl mutant, not all of which are modulated by amino acid availability in amino acid-starved wild-type cells.

A deletion of AUG codon 1 was found to reduce expression in both wild-type and gcdl cells (110). This finding supports the conclusion stated above that the first AUG codon has a positive effect on GCN4 expression, particularly under derepressing conditions. However, whereas the deletion of ORF 1 almost completely destroys regulation, the point mutation in the first AUG codon reduces the derepression ratio but leaves regulation largely intact. The residual regulation observed for the point mutation in the first AUG codon led to the idea that the second AUG codon can partially substitute for the first in promoting derepression (76). Perhaps the deletion of ORF 1 leads to greater inhibition of GCN4 expression because of an added inhibitory effect of novel junction sequences generated in its construction.

The GCN4 mRNA leader segment containing all four upstream ORFs has been replaced by 115 bp from the 3' end of a sea urchin alpha-tubulin complementary DNA containing a single AUG codon, followed by eight sense codons and a termination codon. The resulting allele exhibits constitutively repressed GCN4 expression comparable to that observed for the wild-type gene in repressing conditions. A similar result was obtained for an inversion of the GCN4 segment containing the four upstream AUG codons that substitutes the normal sequences with two short ORFs of 8 and 12 codons (110). These observations suggest that heterologous upstream AUG codons can supply the repressing function exerted by the GCN4 upstream AUG codons; however, it appears that specific sequences are required to regulate the inhibitory effects of upstream AUG codons on GCN4 expression.
Translational Control of GCN4 and the Scanning Hypothesis for Translation Initiation

It is unlikely that the upstream ORFs in the leader of GCN4 mRNA function as sensors of amino acid availability in the manner described for leader peptide-coding sequences in prokaryotic amino acid biosynthetic operons (reviewed in reference 54). The main argument against this possibility is that derepression of GCN4 expression occurs in response to starvation for amino acids not encoded in the upstream ORFs. Instead, the regulatory function of the GCN4 leader is probably best understood in the context of the scanning hypothesis for translation initiation in eucaryotes. This hypothesis proposes that the 40S ribosomal subunit associates with the 5' end of the mRNA and scans downstream until an AUG codon suitable for initiation is encountered. If initiation occurs efficiently at an AUG codon found in the leader, then initiation at the protein-coding sequences downstream will occur at a low level. This hypothesis explains the tendency for eucaryotic mRNAs to be monocistronic and for the initiation codon to occur on the mRNA as the 5'-proximal AUG triplet. At least for mammalian transcripts, when the latter condition does not prevail, upstream AUG codons are generally found in an unfavorable sequence context for initiation. In such mRNAs, it is reasonable to expect that a large fraction of scanning 40S subunits bypass the upstream AUG codon and initiate efficiently at the correct AUG triplet downstream (56).

Consistent with the scanning model is the demonstration that insertion of an AUG codon into the leader of a mammalian transcript often leads to a substantial reduction in the translation of protein-coding sequences located downstream (48, 57, 58, 59, 84). At least in some cases, the strength of the blockade was correlated with the presence of a favorable sequence context surrounding the upstream AUG codon (48, 57, 58). Moreover, the blockade could be weakened if translation from the upstream AUG codon was terminated in the vicinity of the downstream initiation codon. The latter observation led to the suggestion that the scanning process can resume following a termination event, allowing reinitiation to occur at a second AUG codon located downstream. Presumably, reinitiation events occur inefficiently since the presence of a termination codon does not completely overcome the blockade exerted by an upstream AUG codon (48, 57, 58, 84). As in higher eucaryotes, the 5'-proximal AUG codons in yeast transcripts generally serve as translation initiation sites for the protein-coding sequences on the mRNA (15a). In addition, it was shown that the presence of an upstream AUG codon in the leader of the CYC1 and HHS4 transcripts greatly attenuate initiation at downstream AUG codons (95; M. A. Cigan and T. F. Donahue, personal communication).

In view of these considerations, it seems likely that the upstream AUG codons in GCN4 mRNA are efficiently recognized as initiation codons by scanning 40S subunits prior to recognition of the GCN4 initiation codon. In repressing conditions, initiation at the upstream AUG codons attenuates translation of the GCN4 protein-coding sequences because reinitiation following translation of the upstream ORFs is inefficient. To simplify our consideration of the process, recall that a GCN4 allele containing only AUG codons 1 and 4 is regulated very similarly to the wild-type gene. When the first AUG is present as the sole upstream AUG codon in the leader, GCN4 expression is reduced by more than 50% in repressing conditions (gcen2 cells; Fig. 9). This observation suggests that ~50% of the ribosomes that scan the GCN4 mRNA leader initiate at the first AUG codon and fail to reinitiate downstream. The 50% that move beyond the first ORF either failed to initiate at this site or are capable of reinitiation following translation of ORF 1. When the fourth AUG codon is present downstream from the first AUG codon, GCN4 expression is reduced to <1% of the value observed in the absence of any upstream AUG codons (Fig. 9). Thus, in repressing conditions, essentially all ribosomes scanning the GCN4 mRNA leader initiate at one or the other of these two upstream AUG codons and fail to reinitiate at the GCN4 protein-coding sequences downstream.

The novel aspect of the GCN4 transcript is that the strong blockade exerted by the fourth AUG codon can be overcome in derepressing conditions. In gcen1 cells, about one-half of the ribosomes that move beyond ORF 1 also make it past ORF 4 and initiate at the GCN4 coding sequences. In the absence of ORF 1, the blockade exerted by AUG codon 4 cannot be overcome, even in a gcen1 mutant (Fig. 8). The key question, then, is how recognition of the first AUG codon allows ribosomes to escape the blockade exerted by the fourth AUG codon.

One possibility is that initiation at the fourth ORF is suppressed by translation of ORF 1 in derepressing conditions, thereby allowing a substantial number of ribosomes (or 40S subunits) to scan past the fourth AUG codon and initiate at the GCN4 coding sequences downstream. A different sort of model is based on the idea that ORF 4 acts as a blockade because ribosomes become arrested in the course of its translation, thus preventing other ribosomes from moving through the region. According to the latter model, translation of ORF 1 somehow stimulates translation of ORF 4, thereby permitting a stream of scanning ribosomes to reach the GCN4 initiation codon. A third model suggests that translation of ORF 4 occurs constitutively and that the regulated parameter is the efficiency of reinitiation following translation of ORF 4.

For each of these models, several molecular mechanisms can be imagined whereby translation of ORF 1 influences events at ORF 4. One possibility is that ribosomes translating ORF 1 physically interact with ribosomes at ORF 4. Such interactions could be stimulatory or inhibitory depending on the model under consideration. A second possibility is that translation of ORF 1 alters the secondary structure of the mRNA in the vicinity of ORF 4 and thereby changes the efficiency of translation events at that site. These two models require no translational reinitiation events whatsoever. A third mechanism proposes that ORF 1 is uniquely designed to permit reinitiation events following its translation; its role in the regulatory mechanism is simply to ensure that many initiation events at the fourth AUG codon involve reinitiating ribosomes. The central idea behind this third mechanism is that ribosomes involved in reinitiation events at the fourth AUG codon may behave differently, at least in derepressing conditions, from ribosomes involved in "primary" initiation events at this site. They may be more likely to scan over the fourth AUG codon or they may reinitiate at a higher frequency following translation of ORF 4. An important assumption of this mechanism is that many 40S subunits initiate translation at ORF 1 and are capable of reinitiation events downstream. By contrast, primary initiation events at ORF 4 are almost never followed by reinitiation events downstream.

The final aspect of the control mechanism that must be explained is how the effect of ORF 1 on ORF 4 is regulated by amino acid availability. One possibility is that translation
of ORF 1 is itself regulated by the trans-acting factors. GCD gene products would act to repress initiation at the first AUG codon; GCD gene products would antagonize this effect and stimulate initiation at the first AUG codon in starvation conditions. In this way, the trans-acting factors indirectly influence events at ORF 4 via their effects on translation of ORF 1. One argument against this model is that the inhibitory effect on GCN4 expression exerted by ORF 1, when it is present as the sole upstream ORF in the leader, is very similar in repressing and derepressing conditions (Fig. 10). Unless the reinitiation efficiency following translation of ORF 1 is also regulated, this result implies that the initiation frequency at the first AUG codon is constitutive.

The alternative is to propose that translation of ORF 1 occurs constitutively and is necessary but not sufficient for overcoming the blockade exerted by ORF 4. Only when GCD factors are antagonized by GCN gene products can translation of ORF 1 influence events at ORF 4. Depending on which of the above-mentioned molecular mechanisms is operative, antagonism of the GCD factors by GCN products in starvation conditions would permit the ribosome-ribosome interactions between ORFs 1 and 4, or lead to alterations in the behavior of reinitiating ribosomes at ORF 4.

Clearly, there are many possibilities for the translational regulatory mechanism involving the upstream AUG codons in GCN4 mRNA. The requirement for sequences flanking the upstream ORFs for proper regulation should provide valuable clues about whether initiation, elongation, or termination of translation in a particular upstream ORF is the critical parameter for its role as a regulatory element. In addition, analysis of the biochemical properties of the GCN and GCD gene products should yield important information about how these factors modulate the effect of the upstream ORFs on translation initiation at the GCN4 AUG codon.

Comparison of Translational Control Mechanisms for GCN4 and CPAI

The CPAI gene of S. cerevisiae is the second well-documented example of an eucaryotic gene whose expression is regulated at the translational level by short protein-coding sequences in the 5' leader of its mRNA. Unlike the GCN4 upstream ORFs that are only two or three codons in length, the CPAI transcript contains a single upstream ORF of 25 codons. CPAI encodes an enzyme in the arginine biosynthetic pathway whose expression is repressed by arginine about fivefold. Removal of the AUG codon of the CPAI upstream ORF leads to constitutive expression of CPAI enzyme expression while having less than a twofold effect on the CPAI transcript level. This finding suggests that the CPAI upstream ORF acts as a negative regulatory element at the translational level to mediate arginine control over CPAI expression. Interestingly, a variety of other mutations throughout the upstream ORF, both missense and nonsense, also lead to derepressed CPAI expression. These mutations leave the upstream AUG codon intact. While it cannot be ruled out definitively, it seems unlikely that all of these mutations interfere with recognition of the upstream AUG codon as a site for translation initiation. Therefore, it appears that the upstream AUG codon is not sufficient for translational control of CPAI expression. One possibility is that the peptide product of the 25-codon upstream ORF is involved in the regulatory mechanism. Because the phenotypes of mutations in the upstream ORF are only observed when the mutations are located in cis with CPAI, the leader peptide would necessarily be confined in its action to regulating the translation of its own or nearby mRNAs (114). An alternative possibility is that the primary or secondary structure of the leader region encoding the short peptide is required in addition to an upstream AUG codon. These sequences might serve as a binding site for CPAI, a trans-acting negative effector in arginine regulation of CPAI expression (109). Since the missense and nonsense mutations that derepress CPAI expression map throughout the upstream ORF, this putative recognition site would have to be nearly as large as the upstream ORF itself, or 75 nucleotides. To help distinguish between these two models, it would be useful to know the phenotype of substitution mutations in the upstream ORF that do not change the amino acid sequence of the predicted peptide.

One important difference between these two regulatory mechanisms is that the GCN4 transcript contains multiple upstream AUG codons whereas CPAI mRNA has only a single leader ORF. Moreover, the GCN4 upstream AUG codons (or the ORFs they initiate) appear to be functionally distinct, with the 3'-proximal sequences acting as strong negative elements and the 5'-proximal sequences acting as positive regulatory elements in translational control. At least one of each type of upstream ORF is needed for a wild-type pattern of GCN4 expression, and the available data suggest that interactions between these ORFs occur at some level to regulate the number of scanning 40S subunits that reach the GCN4 initiation codon. This model envisions a strictly cis-acting function for the GCN4 upstream ORFs. In the case of CPAI, a limited trans-acting function for the peptide product of the upstream ORF has been postulated wherein the short peptide marks the ribosome as a target for arginine-mediated repression of initiation at the CPAI coding sequences downstream (114).

Pleiotropy of gcd Mutations

With one exception, all mutations in GCD genes whose products regulate GCN4 activity are pleiotropic and lead to conditional lethality at 36°C or unconditional slow growth in both repressing and derepressing conditions (32, 81, 118). The only exception is the gcd7 mutation, first described as gcd3 (76), for which no effect on growth rate was reported. The growth defects associated with other gcd mutations do not result simply from overexpression of GCN4 because they remain evident in gcd gcdn double mutants (32, 81). However, inactivation of GCN4 does improve the growth of gcdl mutants, showing that a toxic effect associated with overexpression of GCN4 is one component of the growth defect in gcdl cells. In fact, this component formed the basis for isolation of gcdn mutations as weak suppressors of the temperature-sensitive phenotype (Tsm-) of a gcdl-101 strain (39). Overexpression of GCN4 also results in a slightly reduced growth rate in otherwise wild-type cells (Hannig and Hinnebusch, unpublished observations).

Mutations in GCD1 and GCD12 cause cells to cease growth at 36°C and accumulate predominantly as un budded cells (118). This phenotype is characteristic of cell cycle arrest in the G1 phase (see reference 88). As expected for a mutation that leads to G1 arrest, it was shown that diploids homozygous for the gcd1-101 mutation fail to sporulate at the restrictive temperature (96). In both gcd1 and gcd12 mutants, there is a substantial reduction in the rates of incorporation of radioactive precursors into DNA, RNA, and protein following a shift to the restrictive temperature (118; S. Harashima and A. G. Hinnebusch, unpublished...
observations). Therefore, it appears that the products of GCD1 and GCD12 contribute to an essential function required for entry into the cell cycle. Only in the case of GCD1 has it been demonstrated that a deletion of the gene is unconditionally lethal (D. E. Hill and K. Struhl, personal communication).

The nature of the essential function affected by GCD gene products remains to be determined. Given that at least some of these factors are involved in translational control of GCN4, it is tempting to consider that the GCD products are components of the protein synthesis machinery. Other mutations in S. cerevisiae are known that directly affect protein synthesis and also lead to G1 cell cycle arrest. These include mes1 and iles1, mutations that impair two aminoacyl-tRNA synthetases (80, 112), and cdc63-1, a mutant allele of the PRT1 gene whose product is involved in the initiation of protein synthesis (26). If GCD factors are involved in general protein synthesis, then the role of the GCN factors would be to modify GCD-encoded components of the translational apparatus in starvation conditions so as to alter the behavior of the upstream AUG codons in the GCN4 mRNA leader without blocking general protein synthesis. According to this model, the properties of the GCN4 upstream ORFs have evolved to exploit minor alterations in the translational process that occur in amino acid-starved cells. If this hypothesis is correct, there may be additional mRNAs similarly constructed that also respond to changes in the translational machinery in starvation conditions. An alternative explanation for the effects of gcd mutations on cell growth and division is that GCD factors not only regulate GCN4 expression but also are involved in coordinating the yeast cell cycle with the availability of amino acids (118).

The above hypothesis concerning the role of the GCD factors in general protein synthesis is consistent with the relatively large number of GCD genes identified thus far. In fact, there is no reason to assume that the gcd class of regulatory mutations has been saturated. There may be many components of the translational process whose functions are required to maintain repression of GCN4 mRNA in nonstarvation conditions. All such factors identified genetically need not be involved directly in the general control regulatory mechanism. Some gcd mutations may lead to a general perturbation of the translational machinery that mimics the more specific changes brought about in amino acid-starved wild-type cells as the result of GCN function. For this reason, evidence of interactions between GCN and GCD gene products should be valuable for identifying GCD factors that play a central role in the translational control of GCN4.

Genetic evidence has been obtained for such an interaction between the products of GCN3 and three GCD genes, GCD1, GCD12, and GCD11. Certain mutant alleles of GCN3 were isolated as partial suppressors of the temperature-sensitive phenotype of a gcdl-101 strain (39). These mutations, typified by the gcd3-102 allele, also suppress the regulatory defect associated with the gcd1-101 mutation, leading to reduced GCN4 expression compared to a gcdl-101 GCN3 strain (36). Therefore, the gcd3-102 allele acts to promote better GCD1 function in gcdl-101 mutant strains. When separated from the gcdl-101 mutation, the gcd3-102 allele blocks derepression of GCN4 as efficiently as a null allele of GCN3 (31a). By contrast with the suppressive effect of the gcd3-102 mutation, the gcd3-101 allele exacerbates the growth defect in gcdl-101 cells and leads to unconditional lethality (39). The gcd3-101 mutant was isolated strictly on the basis of its nonderepressible phenotype (118). The observation of markedly different growth phenotypes for gcng-102 gcdl-101 and gcng-102 gcdl-101 double mutants was taken as an indication that the GCN3 and GCD1 proteins might physically interact to execute an essential function (39). Interestingly, it is now known that a null mutation in GCN3 has no detectable effect on growth rate in an otherwise wild-type strain (31a), whereas deletion of GCD1 is unconditionally lethal (Hill and Struhl, personal communication).

Additional instances of allele-specific interactions between gcdl and gcng mutations and similar, more dramatic interactions between gcng and gcdl2 mutations were reported recently (31a). gcdl and gcdl2 mutants were isolated as suppressors of a gcng-101 gcng-101 double mutation. When recombined with the wild-type GCN3 allele, it was discovered that both the general control-derepressed and temperature-sensitive phenotypes associated with the gcdl mutations were partially suppressed. The same phenotypes associated with the gcdl2 mutations were completely suppressed. Thus, the gcdl2 mutations have no phenotype whatsoever in a GCN3 background. In contrast to these interactions, the growth defects associated with several mutant alleles of GCD11 were exacerbated by replacement of gcng-101 with the wild-type allele of GCN3. The allele state of GCN2 appeared to have no effect on expression of the mutant phenotypes produced by any of these gcd mutations (31a).

Replacement of the gcng-101 mutation with a gcng insertion/deletion allele in the gcdl and gcdl2 mutants did not suppress the mutant phenotypes associated with the gcd mutations. This observation suggested that the necessary condition for expression of the Gcd" phenotype is the absence of the GCN3 gene product. As mentioned above, the gcd3-102 allele is completely defective for its positive regulatory function; nevertheless, it behaves like the wild-type GCN3 allele and efficiency suppresses the gcdl and gcdl2 mutations. These results suggest that the GCN3 product carries out multiple functions. One function is positive regulation of GCN4 expression in amino acid starvation conditions. The other function, evident from these genetic data, appears to involve the promotion of GCD function in nonstarvation conditions (31a).

Two explanations have been advanced to explain the interactions that occur among gcng, gcdl, and gcdl2 mutations. One hypothesis is that the GCN3 protein exists in a complex with the GCD1 and GCD12 proteins and can stabilize thermolabile products encoded by mutant alleles of these genes. In a gcng deletion strain or in the gcng-101 cells in which the gcd mutations were isolated, GCN3 protein is expected to be absent from the complex and the thermolability of the gcd proteins is expressed. By contrast, the products of GCN3 and gcng-102 are capable of complex formation and can stabilize gcd mutant proteins. Complex formation between these proteins is attractive because it suggests a mechanism whereby the GCN3 protein can antagonize GCD function: modification of the GCN3 protein in starvation conditions could lead to a structural alteration and loss of activity in the GCD proteins with which GCN3 interacts. In this model, the gcng-102 product is capable of complex formation but cannot undergo the appropriate modification in starvation conditions needed to influence the GCD factors (31a).

An alternative explanation is that the GCN3 protein can substitute for the GCD gene products to carry out the functions needed for entry into the cell cycle and to maintain repression of GCN4 expression in nonstarvation conditions.
Given that GCN3 is believed to antagonize GCD function in starvation conditions (Fig. 3), this hypothesis suggests the interesting possibility that GCN3 antagonizes GCD factors by competing with them. In the absence of the GCD1 and GCD12 products, the GCN3 protein would maintain repression of GCN4 expression in nonstarvation conditions. In starvation conditions, the repressor function of the GCN3 product would be inactivated and, consequently, competition with the GCD factors by inactive GCN3 protein would allow derepression of GCN4 expression to occur. In this model, the gen3-102 mutation would block inactivation of GCN3 repressor function in starvation conditions (31a).

Regardless of which model is correct, the interactions among mutations in GCN3, GCD1, and GCD12 suggest that the products of these genes have closely related functions in the general control regulatory hierarchy. Thus, the GCN3 protein emerges as a good candidate for a GCN factor that operates at the interface between the positive and negative regulators of GCN4 expression. GCN1 and GCN2 might then function by regulating the expression or activity of the GCN3 protein in response to amino acid starvation. The roles of other GCD factors involved in the translational control remain unknown. The GCD1 and GCD12 proteins might physically interact with additional GCD proteins in a larger complex. Alternatively, a pathway of GCD functions may be involved in translational repression of GCN4 mRNA.

Transcriptional Control of GCN4

It was suggested that the GCD7 gene product regulates the amount of GCN4 mRNA in addition to, or instead of, regulating the translational efficiency of GCN4 mRNA (76). This conclusion is based on the finding that the steady-state amount of GCN4 mRNA is greater in a gcd7 mutant than in a closely related GCD strain. It should be noted that mutations in GCD1 and GCD10-GCD13 also lead to increased amounts of GCN4 mRNA. As described earlier, even greater increases in the GCN4 transcript level brought about by increased GCN4 gene dosage are insufficient to derepress HIS4 expression in a GCD strain to the same extent observed in gcd1 and gcd10-gcd13 mutants (75). This and other results described above strongly suggest that GCD1 and GCD10-GCD13 have a major effect on the translational efficiency of the GCN4 transcript. The manner in which these GCD factors also affect the steady-state level of GCN4 mRNA is not well understood. It could be that the stability of the GCN4 transcript is coupled to its translational efficiency. At odds with this suggestion is the finding that removal of all four upstream ORFs from GCN4-lacZ mRNA increases expression of GCN4-lacZ fusion enzyme by the same amount as the gcd mutations but without increasing the fusion transcript level (36, 75). An alternative explanation is that the GCD factors control the rate of GCN4 transcription as well as its translation.

Evidence was presented recently that the GCN6 and GCN7 gene products act as positive effectors of the GCN4 transcript level (30). As in the case of the gcd mutations, the effects of gcn6 and gcn7 mutations on GCN4 mRNA levels could be secondary to their influence on translation; however, it is significant that these gcn mutations lead to lower amounts of GCN4 mRNA, whereas gcn2 and gcn3 mutations produce higher levels of the GCN4 transcript. The latter phenotype is one of the indications that GCN2 and GCN3 stimulate GCN4 expression at the translational level (36, 75). In fact, a gcn4 mutation also leads to increased GCN4 transcript levels in starvation conditions, raising the possibility that either the transcription rate or the stability of GCN4 mRNA is increased by severe starvation. Perhaps the GCN6 and GCN7 gene products are responsible for this transcriptional effect. The GCD gene products might then influence GCN4 transcript levels indirectly by repressing expression of GCN6 and GCN7.

GENERAL AMINO ACID CONTROL IN OTHER FUNGI AND IN BACTERIA

General control of amino acid biosynthetic enzyme expression has been detected in two other fungi of the Ascomycete class, Neurospora crassa and Aspergillus nidulans. In N. crassa, starvation for a variety of amino acids results in two- to tenfold derepression of enzymes in the histidine, tryptophan, arginine, and lysine biosynthetic pathways (7, 12, 13, 51). Similar observations have been made for A. nidulans (87). The derepression response in N. crassa was shown to result from increased expression of the mRNAs encoding the regulated enzymes (27, 28).

Mutations have been isolated that block derepression of enzymes subject to the general control in N. crassa. These mutations map to a single gene known as cpl-l (7, 17). The cpl-l mutation prevents derepression of arg-l2 expression at the level of mRNA accumulation (28). Interestingly, it appears that the abundance of as many as 20% of the total mRNA species in this organism is affected by a cpl-l mutation in amino acid starvation conditions (27). This observation raises the possibility that the general control in N. crassa extends beyond enzymes involved in amino acid biosynthesis. At present, the cpl-l gene product is the best candidate to carry out the role of the GCN4 protein in N. crassa. In fact, it was recently found that the cpl-l gene encodes a protein with a domain having considerable homology to the DNA-binding domain of GCN4 (J. Paluh, M. Orbach, T. Legerton, and C. Yanofsky, personal communication). This finding makes it very likely indeed that the cpl-l gene product is the direct regulator of transcription in the general control system of N. crassa.

A general control of amino acid biosynthesis also operates in the enteric bacteria E. coli and Salmonella typhimurium and is mediated by the “alarmone” guanosine tetraphosphate (ppGpp). In these organisms, the main effects of the general control are observed when cells are subjected to a “shiftdown” from amino acid-complete medium to media lacking one or more amino acids. In such shifts, expression of the his operon of Salmonella typhimurium derepresses severalfold, even in the presence of exogenous histidine. This response occurs independent of the his attenuator and requires the product of the relA gene, a ribosome-associated protein that carries out the synthesis of ppGpp. In vivo, the activity of the relA gene product is stimulated in response to a shiftdown and the amount of ppGpp increases (10, 14, 99, 116). His operon expression can be elevated independently of a shiftdown to minimal medium by inactivation of the spoT gene product, a protein capable of ppGpp degradation. spoT mutants exhibit elevated levels of ppGpp in these conditions (91).

The effect of spoT mutations on his operon expression is known to depend on DNA sequences in the his promoter (91). ppGpp can stimulate transcription of several amino acid biosynthetic operons in vitro (99, 119), and in the case of the his operon, this effect was shown to require sequences in the −10 region of the promoter (90). The actual target of ppGpp action in the transcriptional complex of these bacteria remains to be determined.
ACKNOWLEDGMENTS

I am grateful to my colleagues in the field for communication of unpublished results and helpful discussions about general amino acid control. I thank Angela Stewart for her skill and patience in preparing the manuscript.

LITERATURE CITED


115. Wolfer, M., D. Yef, F. Messunguy, and G. R. Fink. 1975. Integration of amino acid biosynthesis into the cell cycle of


