Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4579–4583, June 1989
Genetics

Juxtaposition of domains homologous to protein kinases and histidyl-tRNA synthetases in GCN2 protein suggests a mechanism for coupling *GCN4* expression to amino acid availability

(translational control/aminoacyl-tRNA synthetase)

RONALD C. WEK, BELINDA M. JACKSON, AND ALAN G. HINNEBUSCH

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

Communicated by Igor B. Dawid, February 22, 1989

ABSTRACT **The GCN2 protein of *Saccharomyces cerevisiae* stimulates the expression of amino acid biosynthetic genes under conditions of amino acid starvation by derepressing *GCN4*, a transcriptional activator of these genes. GCN2 contains sequences homologous to the catalytic domain of protein kinases. We show here that substitution of a highly conserved lysine in the presumed ATP-binding site of this domain impairs the derepression of histidine biosynthetic genes under GCN4 control. This result supports the idea that protein kinase activity is required for GCN2 positive regulatory function. Determination of the nucleotide sequence of the entire *GCN2* complementation unit, and measurement of the molecular weight of GCN2 protein expressed *in vivo*, indicate that GCN2 is a $M_r \approx 180{,}000$ protein and contains a $M_r \approx 60{,}000$ segment homologous to histidyl-tRNA synthetases (HisRSs) juxtaposed to the protein kinase domain. Several two-codon insertion mutations in the HisRS-related coding sequences inactivate *GCN2* regulatory function. Based on these results, we propose that the GCN2 HisRS domain responds to the presence of uncharged tRNA by activating the adjacent protein kinase moiety, thus providing a means of coupling GCN2-mediated derepression of *GCN4* expression to the availability of amino acids.**

Protein phosphorylation is an important posttranslational modification involved in regulating many cellular processes, including signal transduction, growth control, carbon catabolite repression, and protein synthesis (1, 2). Protein kinases are often regulated by ligands that bind to regulatory domains or subunits to enhance or inhibit catalytic activity. Examples of this phenomenon are cyclic nucleotide-regulated protein kinases, diacylglycerol activation of protein kinase C, and calmodulin-mediated calcium regulation of phosphorylase kinase and myosin light-chain kinase.

A protein kinase has been implicated in the general amino acid control of the yeast *Saccharomyces cerevisiae* (3). In this system, starvation for any one of at least 10 amino acids, or a defective aminoacyl-tRNA synthetase, leads to increased transcription of 30 or more genes encoding amino acid biosynthetic enzymes in nine different pathways (reviewed in ref. 4). The transcriptional activator GCN4 directly mediates this derepression response. Expression of *GCN4* itself is regulated by amino acid availability, but at the level of translation initiation. Trans-acting positive factors encoded by *GCN1*, *GCN2*, and *GCN3* are required to stimulate translation of *GCN4* mRNA in response to starvation, presumably by antagonism of negative-acting *GCD* factors (4). A portion of the predicted amino acid sequence of GCN2 is homologous to the catalytic domain of eukaryotic protein kinases and evidence was presented that *GCN2* either encodes or regulates a protein that has kinase activity *in vitro* (3).

In this report we show that a highly conserved lysine in the presumptive ATP-binding site of the GCN2 kinase domain is required for derepression of genes under the general control, supporting the idea that GCN2 kinase activity is required for its role as an activator of gene expression. In addition, reexamination of the entire *GCN2* nucleotide sequence indicates that the carboxyl-terminal region of GCN2 is closely related to histidyl-tRNA synthetases (HisRSs) from *S. cerevisiae*, humans, and *Escherichia coli*. The juxtaposition of sequences homologous to HisRS with the catalytic domain of protein kinases raises the possibility that the HisRS portion of GCN2 monitors the concentration of aminoacyl-tRNA in the cell and activates the adjacent protein kinase moiety when uncharged tRNA accumulates.

MATERIALS AND METHODS

Plasmid pC102-2 (5) contains *GCN2* on a 7.0-kilobase (kb) *Sau*3AI fragment in the *Bam*HI site of YCp50. The complete nucleotide sequence of *GCN2* was determined by the dideoxy chain-termination method (6). Regions where discrepancies exist between our sequence and that reported previously (3), and the entire region downstream from +4261 not heretofore analyzed, were sequenced multiple times on both strands. Oligonucleotide-directed mutations were generated as described (7, 8) and verified by DNA sequence analysis of the restriction fragments that were subcloned into pC102-2. To construct insertions of the kanamycin-resistance gene (*kan*^{r}) in *GCN2*, the oligonucleotide 5′-CGAGCT-3′ was ligated to a 1.3-kb *Sac* I fragment containing *kan*^{r} isolated from plasmid pUC4-KISS (Pharmacia). This fragment was ligated to pC102-2 DNA linearized by partial digestion with *Taq* I (9). To obtain two-codon insertions, *kan*^{r} insertion plasmids were digested with *Sac* I and religated, leaving a *Sac* I site between the cytosine and guanine nucleotides of each original *Taq* I site. Plasmids were transformed into yeast strain H1149 (*MATα ura3-52 ino1 leu2-3 leu2-112 gcn2::LEU2 HIS4-lacZ*) by the lithium acetate method (10). The *HIS4-lacZ* fusion was described previously (11); *gcn2::LEU2* contains a 2.8-kb *LEU2* fragment in place of the *GCN2* sequences located between the +63 *Eco*RI site and the +3284 *Hin*dIII site (Fig. 1).

A *trpE-GCN2* fusion was constructed by inserting the 1524-base-pair (bp) *Hin*dIII *GCN2* restriction fragment from the middle of *GCN2* (Fig. 1) into the *Hin*dIII site in plasmid pATH2 (12). The insoluble *trpE-GCN2* fusion protein was purified from *E. coli* (12) and used for antiserum production in rabbits. GCN2 was immunoprecipitated from extracts made

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HisRS, histidyl-tRNA synthetase; 3-AT, 3-aminotriazole.

from transformants of strains H1149 and H1153 (*MATa ura3-52 leu2-3 leu2-112 HIS4-lacZ*). Five-milliliter cultures of each strain were grown in repressing conditions for 6 hr (13), pulsed for 20 min with 300 μCi of [^{35}S]methionine (specific activity, 1300 Ci/mmol; 1 Ci = 37 GBq), and chased for 10 min with unlabeled methionine at 1 mM. Preparation of extracts and immunoprecipitations were performed as described (14).

RESULTS

Characterization of the *GCN2* Gene and Its Protein Product. An interesting feature of the previously reported nucleotide sequence of *GCN2* is a long 3′-untranslated region that is required for *GCN2* function (3). We further investigated the importance of this region by introducing insertions and deletions into a plasmid-borne copy of *GCN2* and testing the effects of these mutations on genetic complementation of a chromosomal *gcn2::LEU2* deletion. Two different kinds of sequences were inserted at random into various *Taq* I sites in the *GCN2* region: a bacterial kanamycin-resistance gene (*kan*^{r}) or the 6 bp (two codons) that constitutes a *Sac* I restriction site. *GCN2* function was assayed in transformants containing the mutant plasmids by measuring resistance to 3-aminotriazole (3-AT), a competitive inhibitor of the *HIS3* product that induces histidine starvation. (*gcn2::LEU2* strains are defective for *HIS3* derepression under starvation conditions and hence exhibit increased 3-AT sensitivity compared to wild-type strains.) The results (Fig. 1) showed that the *GCN2* complementation unit extends ≈1.7 kb downstream from the previously reported stop codon (3). Although sequences downstream from the 3′-proximal *Bgl* II site are not needed for complementation activity, they are required for expression of *GCN2* mRNA of the correct size (data not shown), suggesting that the 3′ end of *GCN2* mRNA maps downstream from this *Bgl* II site. Our complementation data are consistent with the position of the 5′ end of the *GCN2* transcript (+1 in Fig. 1) reported by Roussou *et al.* (3).

The nucleotide sequence of the *GCN2* gene was reexamined and extended to include the entire complementation unit (Fig. 2). Our data indicate that the *GCN2* open reading frame is coextensive with the genetic complementation unit, terminating just upstream from the 3′-proximal *Bgl* II site (Fig. 1). The deduced GCN2 protein sequence (Fig. 2) is 1590 amino acids in length and has a M_r of 182,000, a value greater by 64,000 than that suggested previously (3).

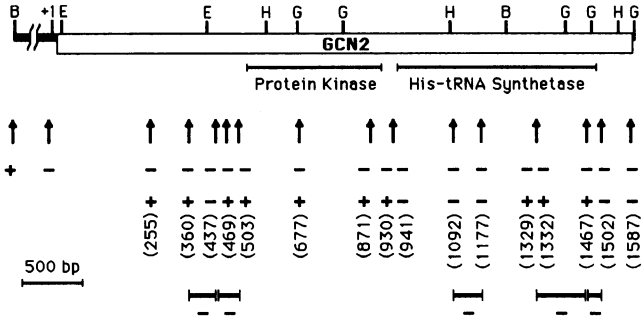

FIG. 1. Functional map of *GCN2*. The open box designates the GCN2 protein-coding sequence oriented 5′ → 3′, with domains homologous to protein kinases and HisRSs indicated below. The +1 marks the 5′ end of *GCN2* mRNA (3). Arrows mark the sites of insertion mutations; "+" and "−" indicate complementing activity of *kan*^{r} (top row) and codon (bottom row) insertion alleles; "+" indicates wild-type growth 3 days after replica-printing to medium containing 30 mM 3-AT (15); "−" indicates poor growth under the same conditions. Positions of codon insertions are given in parentheses. *Sac* I insertions 1092, 1177, and 1587 introduce Glu-Leu codons; 255, 469, and 941 introduce Ala-Arg codons; 360, 437, 503, 677, 871, 930, 1329, 1332, 1467, and 1502 introduce Ser-Ser codons. The extent of noncomplementing in-frame deletions constructed from pairs of *Sac* I insertion mutations is shown by bars at the bottom of the figure. From left to right, the codons removed by the deletions are 360–436, 437–502, 1092–1176, 1332–1466, and 1467–1501. Letter designations for restriction sites: B, *Bam*HI; E, *Eco*RI; G, *Bgl* II; H, *Hin*dIII.

We used antibodies to measure the size of the GCN2 protein expressed *in vivo*. Rabbit antiserum was raised against a *trpE*-GCN2 fusion polypeptide and used to immunoprecipitate radiolabeled proteins extracted from yeast transformants containing single or high copy-number plasmids bearing *GCN2*. The results in Fig. 3 show that a protein with M_r of ≈180,000 was detected in strains containing one copy of *GCN2*, absent from the *gcn2::LEU2* deletion strain, and present in larger amounts in a strain transformed with a high copy-number plasmid containing *GCN2*. These data indicate that GCN2 protein has a molecular weight in good agreement with the value predicted from our nucleotide sequence.

The GCN2 Protein Kinase Domain. The GCN2 protein sequence between residues 530 and 910 has significant homology with the catalytic domain of protein kinases (3). Hanks *et al.* (2) recently identified 11 conserved subdomains shared among 65 protein kinases, including 15 nearly invariant amino acids plus 18 conserved residues of similar chemical structure. All 33 of these highly conserved residues are present in GCN2 (Fig. 2). One of the best characterized regions in the kinase catalytic domain is the ATP-binding site, including the sequence Gly-Xaa-Gly-Xaa-Xaa-Gly-Xaa-Val followed 13–18 residues downstream by Ala-Xaa-Lys. The lysine residue is thought to be directly involved in the phosphotransfer reaction, and amino acid substitutions at this position invariably abolish kinase activity (1, 2). According to the sequence alignments of Hanks *et al.* (2), this important residue in GCN2 is expected to be Lys-559 (Fig. 2). We altered the coding sequence of *GCN2* by site-directed mutagenesis to replace Lys-559 with a valine or arginine residue. A third mutation was constructed to replace the adjacent Lys-560 with a valine residue. The ability of the resulting substitution alleles to complement *gcn2::LEU2* for its failure to derepress *HIS3* and *HIS4* expression was determined as described in Table 1. The Val-559 substitution almost completely abolished GCN2-mediated derepression of these *HIS* genes under starvation conditions. Even in the strain containing the conservative Arg-559 substitution, derepression was impaired, whether the *gcn2-K559R* allele was present on a single-copy (Table 1) or multicopy plasmid (unpublished observations). By contrast, the Val-560 substitution of the adjacent nonconserved lysine residue had no effect on derepression. These results support the idea that protein kinase activity is required for the positive regulatory function of GCN2.

The GCN2 HisRS Domain. Comparison of the GCN2 amino acid sequence with the GenBank sequence data base (16) revealed that the carboxyl-terminal region of GCN2 (codons 920–1450) is closely related to HisRS from *S. cerevisiae* (17). As illustrated in Fig. 4, GCN2 and the cytoplasmic form of yeast HisRS are 22% identical over the entire length of HisRS. A statistical analysis aimed at determining the probability that two random sequences of the same amino acid composition would show this degree of sequence identity (21) indicated that the similarity between GCN2 and yeast HisRS has a significance level of 16 standard deviation units above the mean. The carboxyl terminus of GCN2 also shows similarity to the sequences of HisRS from humans and *E. coli* (19, 20), with regions of strong similarity (e.g., codons 982–1032, 1069–1095, and 1251–1270) interspersed with regions of considerable divergence. The three HisRSs are identical at 61 positions (17%) and of these, 23 positions are shared by GCN2 (Figs. 2 and 4).

Two *Sac* I insertions that abolish GCN2 function map in the HisRS-related domain (1092 and 1177, Fig. 1), suggesting the importance of this region for GCN2 positive regulatory function. In particular, the 1092 mutation inserts Glu-Leu in

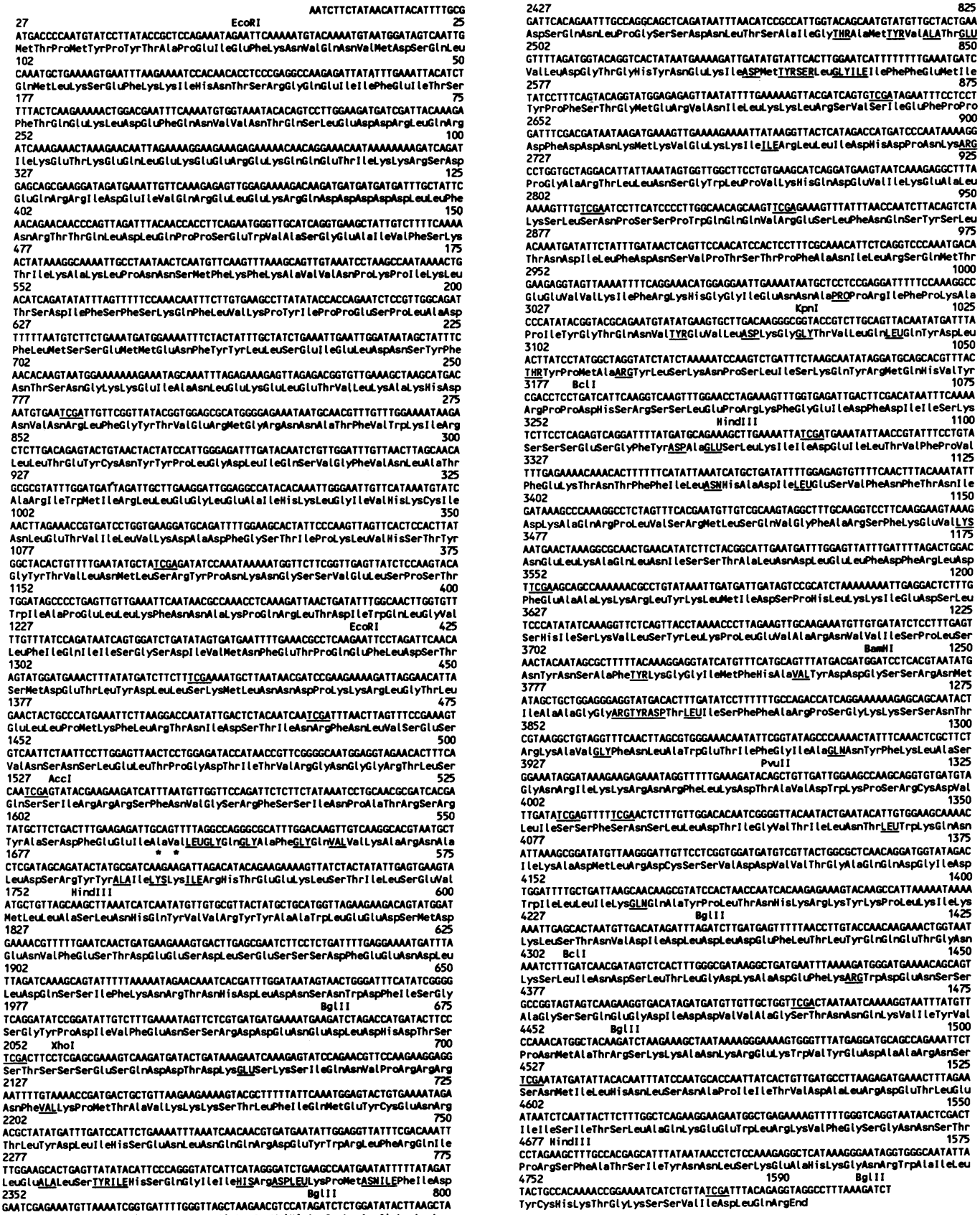

FIG. 2. Nucleotide sequence and deduced amino acid sequence of *GCN2*. The sequence begins at the 5′ end of *GCN2* mRNA (+1) and extends to the 3′-proximal *Bgl* II site shown in Fig. 1. The numbers on the left and right are nucleotide and codon positions, respectively. Underlined and capitalized amino acids are highly conserved residues in the protein kinase (codons 536–900) or HisRS domain (codons 993–1444). *Taq* I sites in which insertions were made are underlined. Asterisks mark the two AAG codons that were mutated to substitute lysine residues in the putative kinase ATP-binding site.

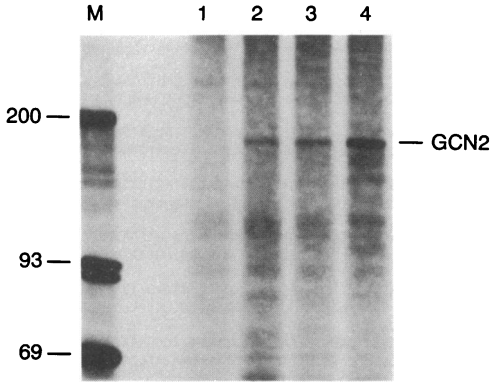

FIG. 3. Immunoprecipitation of GCN2 protein. Antiserum prepared against a *trpE*-GCN2 fusion protein was used to immunoprecipitate GCN2 from [^{35}S]methionine-labeled total protein extracts made from the following strains. Lane 1, *gcn2*::*LEU2* deletion strain H1149 transformed with plasmid YCp50, the parent plasmid of pC102-2, containing no *GCN2* gene; lane 2, H1149 transformed with pC102-2, a low copy-number plasmid containing *GCN2*; lane 3, *GCN2*$^+$ strain H1153 transformed with YEp13, the parent vector of pAH15, containing no *GCN2* gene; lane 4, strain H1153 transformed with high copy-number plasmid pAH15 (15) containing *GCN2*. Immunoprecipitates were collected and analyzed by electrophoresis on a 5–12.5% gradient SDS/polyacrylamide gel, followed by fluorography. Lane M contained size markers with the molecular weights ($\times 10^{-3}$) indicated on the left.

a region highly conserved between GCN2 and the yeast and human HisRS sequences. (The *gcn2-1092* allele provides no detectable GCN2 function even when present on a multicopy plasmid.) The two *Sac* I insertions in the HisRS domain (1329 and 1332) with a Gcn2$^+$ phenotype insert Ser-Ser into a serine-rich region that is less well-conserved between GCN2 and HisRS sequences. No significant similarity was detected between GCN2 and other aminoacyl-tRNA synthetases, consistent with the fact that HisRS appears to be unrelated in sequence to these other enzymes (17, 19, 20).

Table 1. Effect of amino acid substitutions in the GCN2 kinase domain on derepression of *HIS3* and *HIS4* expression in response to histidine starvation

Plasmid-borne allele	*HIS4-lacZ* enzyme activity,* nmol/min per mg		*HIS3* derepression
	R	DR	
GCN2	190	970	+
gcn2-K559V	140	260	−
gcn2-K559R	130	310	−
GCN2-K560V	140	960	+
None	140	220	−

Transformants of *gcn2*::*LEU2 HIS4-lacZ* strain H1149 containing the designated *GCN2* alleles on low copy-number plasmids were analyzed. The *GCN2* point mutations are designated by the wild-type amino acid, codon position, and the substituting amino acid, in that order. The strain carrying no *GCN2* allele was transformed with YCp50, the parent vector of pC102-2. Strains were tested for growth sensitivity to 3-AT as a measure of *HIS3* expression under starvation conditions, as described in the legend to Fig. 1. *HIS4-lacZ* expression was measured in the same strains grown for 6 hr under non-starvation conditions (minimal medium with the required nutrients; repressing, R) or in the same medium containing 10 mM 3-AT to induce histidine starvation (derepressing, DR) (13).

*Enzyme activities are expressed as nmol of *o*-nitrophenyl β-D-galactopyranoside hydrolyzed per min/mg of protein. Values shown are the averages of assays done on two or three independently derived transformants. The result of each assay varied from the mean by 30% or less.

FIG. 4. Sequence alignment of GCN2 and HisRSs from yeast, human, and *E. coli*. Pairwise alignments made using the BestFit alignment program (18) were incorporated manually into this multiple sequence alignment. Dots indicate gaps introduced to maximize similarities. Boxes enclose residues that are identical among the aligned sequences. Asterisks (*) indicate positions that are conserved in all four sequences. Exclamation points (!) signify positions that are conserved in the three HisRS sequences only. The numbers above the sequence indicate codon positions in the GCN2 sequence. GCN2 amino acids 920–1448 from GCN2 were aligned with the entire cytoplasmic form of yeast HisRS (17), 508 residues of human HisRS (19) beginning with residue 22, and the complete *E. coli* HisRS sequence of 424 residues (20). Percentages of sequence identity in pairwise comparisons are as follows: GCN2 vs. yeast HisRS (22%); GCN2 vs. human HisRS (22%); GCN2 vs. *E. coli* HisRS (15%); yeast HisRS vs. human HisRS (47%); yeast HisRS vs. *E. coli* HisRS (25%); human HisRS vs. *E. coli* HisRS (23%).

DISCUSSION

A fundamental question regarding general amino acid control is how derepression of amino acid biosynthetic genes is coupled to the intracellular level of charged tRNA. GCN4 functions directly as a transcriptional activator of these genes under conditions of amino acid starvation. GCN2 activates gene expression indirectly by stimulating GCN4 synthesis in response to starvation. Our mutational analysis supports the idea that GCN2 acts as a positive regulator of *GCN4* expression by functioning as a protein kinase (3). Furthermore, we present the remarkable finding that GCN2 protein contains a domain closely related in sequence to HisRS. Given that aminoacyl-tRNA synthetases bind uncharged tRNA as a substrate, we propose that the HisRS-related domain of GCN2 can monitor the concentration of uncharged tRNA and

activate the adjacent protein kinase moiety under starvation conditions when uncharged tRNA accumulates.

Derepression of *GCN4* and its target genes occurs in response to limitation for any one of several amino acids in addition to histidine. Therefore, it is not obvious how a domain related to HisRS could monitor the charging levels of these other tRNAs. One possibility is that GCN2 has diverged sufficiently from HisRS that it now lacks the ability to discriminate between different tRNAs. In this way, any uncharged tRNA could bind to GCN2 and stimulate protein kinase activity. Unfortunately, it is not well understood what regions in aminoacyl-tRNA synthetases bind tRNA and confer substrate specificity (22). The amino acid sequence Lys-Met-Ser-Lys-Ser is implicated in binding the 3′ end of tRNA in certain synthetases, but even this limited motif is absent from HisRS sequences. If this model is correct, then those residues that are invariant among the three HisRSs but absent from GCN2 may be important determinants of $tRNA^{His}$ binding specificity. A second model is that the HisRS-related domain of GCN2 is specific for $tRNA^{His}$ (or a subset of tRNAs), whose charging level is somehow affected by starvation for other amino acids. In these two models, there is no requirement that GCN2 be competent for aminoacylation of tRNA; rather, binding of uncharged tRNA is expected to be sufficient for kinase activation. In either case, it seems likely that GCN2 would have a smaller binding constant for uncharged tRNA than aminoacyl-tRNA synthetases so that activation of GCN2 protein kinase would require a higher concentration of uncharged tRNA than is needed for efficient aminoacylation by the *bona fide* synthetases.

Another important question is how activation of GCN2 protein kinase activity under starvation conditions leads to increased *GCN4* expression. GCN2 stimulates GCN4 synthesis at the translational level by overcoming the inhibitory effects of short open reading frames present in the leader of *GCN4* mRNA. Translational repression by these sequences depends on the functions of trans-acting factors encoded by *GCD* genes. Mutations in *GCD* genes lead to constitutive derepression of *GCN4* expression even in the absence of functional *GCN2*, prompting the suggestion that GCN2 acts indirectly as a positive effector by antagonism of GCD factors (4). Thus, one or more GCD proteins could be substrates of GCN2 protein kinase activity. Alternatively, since GCN1 and GCN3 are required in addition to GCN2 for negative regulation of GCD factors under starvation conditions (4), GCN1 or GCN3 could be substrates of GCN2 kinase activity. In this view, phosphorylation by GCN2 would activate GCN1 or GCN3 to become antagonists of GCD function.

We thank Ernest Hannig and Paul Miller for thoughtful reading of the manuscript and Charles Moehle and Ernest Hannig for helpful advice on immunoprecipitations and computer analysis of sequences.

1. Edelman, A. M., Blumenthal, D. K. & Krebs, E. G. (1987) *Annu. Rev. Biochem.* **56,** 567–613.
2. Hanks, S. K., Quinn, A. M. & Hunter, T. (1988) *Science* **241,** 42–52.
3. Roussou, I., Thireos, G. & Hauge, B. M. (1988) *Mol. Cell. Biol.* **8,** 2132–2139.
4. Hinnebusch, A. G. (1988) *Microbiol. Rev.* **52,** 248–273.
5. Driscoll-Penn, M., Thireos, G. & Greer, H. (1984) *Mol. Cell. Biol.* **4,** 520–528.
6. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
7. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 488–492.
8. Taylor, J. W., Ott, J. & Eckstein, F. (1986) *Nucleic Acids Res.* **13,** 8764–8785.
9. Barany, F. (1985) *Gene* **37,** 111–123.
10. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. (1983) *J. Bacteriol.* **153,** 163–168.
11. Harashima, S. & Hinnebusch, A. G. (1986) *Mol. Cell. Biol.* **6,** 3990–3998.
12. Dieckmann, C. L. & Tzagoloff, A. (1985) *J. Biol. Chem.* **260,** 1513–1520.
13. Hinnebusch, A. G. (1985) *Mol. Cell. Biol.* **5,** 2349–2360.
14. Klionsky, D. L., Banta, L. M. & Emr, S. D. (1988) *Mol. Cell. Biol.* **8,** 2105–2116.
15. Hinnebusch, A. G. & Fink, G. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5374–5378.
16. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
17. Natsoulis, G., Hilger, F. & Fink, G. R. (1986) *Cell* **46,** 235–243.
18. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
19. Tsui, F. W. L. & Siminovitch, L. (1987) *Nucleic Acids Res.* **15,** 3349–3367.
20. Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S., Parker, J. & Schimmel, P. (1985) *J. Biol. Chem.* **260,** 10063–10068.
21. Kanehisa, M. (1982) *Nucleic Acids Res.* **10,** 183–196.
22. Schimmel, P. (1987) *Annu. Rev. Biochem.* **56,** 125–158.