Nucleic Acids Research, Vol. 20, No. 14 3721–3724

Role of the conserved leucines in the leucine zipper dimerization motif of yeast GCN4

Willem J.van Heeckeren, Joan W.Sellers and Kevin Struhl*
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

Received March 16, 1992; Revised and Accepted June 12, 1992

ABSTRACT

Yeast GCN4 belongs to the class of eukaryotic transcription factors whose bZIP DNA-binding domains dimerize via a leucine zipper motif that structurally resembles a coiled coil. The leucine zipper contains 4–5 highly conserved leucine residues spaced exactly 7 residues apart that are located within the α-helical hydrophobic interface between protein monomers. Here, we investigate the role of the four canonical leucines in the GCN4 leucine zipper by analyzing a series of mutated derivatives for their ability to activate transcription *in vivo* and to bind DNA *in vitro*. The GCN4 leucine zipper is surprisingly tolerant of mutations, with a wide variety of single substitutions at any of the four leucines including basic and acidic amino acids behaving indistinguishably from wild-type GCN4. Moreover, some derivatives containing two leucine substitutions display detectable though reduced function. These results indicate that other residues within the coiled coil are crucial for efficient dimerization, and they suggest that some eukaryotic transcriptional regulatory proteins lacking the conserved leucine repeat will dimerize through a structurally homologous motif. Interestingly, our results differ in several respects from those obtained by analyzing mutations in the GCN4 leucine zipper in the context of a λ repressor-GCN4 zipper hybrid protein. These apparent differences may reflect a functional interrelationship between the leucine zipper and basic region subdomains for DNA-binding by bZIP proteins.

INTRODUCTION

GCN4 protein binds to the promoters of many yeast amino acid biosynthetic genes and activates their transcription during conditions of amino acid starvation (1, 2). GCN4 binds as a dimer to a 9-bp dyad, ATGA(C/G)TCAT, such that the two protein monomers interact with overlapping half-sites (3–6). The DNA-binding domain is localized to the 56 C-terminal amino acids (7, 8) and contains the bZIP structural motif found in many eukaryotic transcription factors including the Jun and Fos oncoproteins (9). The bZIP motif consists of a leucine zipper that mediates dimerization (10–12), and an adjacent basic region that directly interacts with DNA (13, 14). Analysis of mutants that alter the spacing between these sub-domains indicates that the leucine zipper symmetrically positions the adjacent basic regions along the surface of the DNA half-sites (15).

The leucine zipper motif was initially defined by the presence of 4 or 5 leucine residues spaced exactly 7 amino acids apart (9). From this conserved feature, the leucine zipper was initially proposed to be an α-helical dimer formed primarily by interdigitation of leucine residues between monomer subunits (9). Indeed, a GCN4 leucine zipper peptide forms stable α-helical dimers in solution (12), and the same region exists as a dimeric α-helical structure in the context of a functional DNA-binding domain (8, 16). However, the parallel association of the α-helices and the presence of an additional heptad repeat of hydrophobic residues interspersed between the leucine repeat strongly suggests that the leucine zipper is a coiled coil (12). Moreover, X-ray scattering studies demonstrate that the GCN4 leucine zipper is similar to the coiled coil structure found in muscle filament proteins (17, 18), and an artificial coiled coil can substitute for the GCN4 leucine zipper to provide the dimerization function necessary for DNA binding (16). In the coiled coil model, the dimerization interface is not formed by leucine interdigitation, but rather by pairwise interaction of the leucines (position d) with alternate hydrophobic residues (position a) in the opposing α-helix. This model helps explain why the various leucine zippers have distinct dimerization specificities (10, 11, 19) even though they share the canonical leucine residues.

Although a variety of hydrophobic residues are found at positions a and d of standard coiled coils, leucine zipper sequences in eukaryotic DNA-binding proteins containing bZIP domains show an extremely strong preference for leucine residues at the d positions. However, initial mutational analyses of the Fos, Jun, and C/EBP leucine zippers (20–24) suggest that the leucine residues are important, but not essential, for dimerization. Individual leucines can be replaced by conservative residues such as valine and isoleucine with minimal decrease in the dimerization ability, but multiple changes generally cause a severe functional defect. More recently, the role of the conserved leucines has been investigated in an artificial protein in which the dimerization domain of λ repressor was replaced by the GCN4 leucine zipper

* To whom correspondence should be addressed

(25). The results indicate that any one of the 4 leucine residues can be replaced, but functionally acceptable substitutions are generally restricted to hydrophobic amino acids. However, these experiments do not address the importance of the leucine residues in the context of intact GCN4 functioning in a biologically meaningful situation.

In this paper, we randomize the codons for each of the four leucines in the GCN4 zipper and examine many of the resulting proteins for transcriptional activity in yeast cells and DNA-binding activity *in vitro*. We find that a surprisingly wide variety of residues can substitute for any individual leucine with no detectable functional effect *in vitro* or *in vivo*. In addition, some proteins containing two leucine substitutions confer weak, but detectable activity. The mechanistic and evolutionary implications of these results are discussed.

MATERIALS

Mutagenesis of the conserved leucines

Mutant GCN4 derivatives were constructed essentially as described previously (26) using the following degenerate oligonucleotides (N indicates an equimolar mixture of all four nucleotides).

100 AGAATGAAACAANNNGAAGACAGGGTT
101 AGGGTTGAAGAANNNCTTTCGAAAAAT
102 AAAAATTATCACNNNGAAAATGAGGTT
103 GAGGTTGCCAGANNNAAGAAATTAGTT
122 AGAATGAAACAAATNGAAGACAGGGTT
123 AGGGTTGAAGAAATNCTTTCGAAAAAT

5′-phosphorylated oligonucleotides were annealed to uracil-substitued mp18-Sc4380 DNA, extended with the Klenow fragment of DNA polymerase I, and ligated. Sc4380 is a derivative of the 1.1 kb *Hin*dIII-*Eco*RI GCN4 fragment from YCp88-GCN4 (7) that contains an artificial *Sac*I site 5 bp beyond the translational termination codon. The reaction mixtures were introduced into *E.coli* cells that select against the uracil-substituted strand, and DNAs containing prospective mutations in the leucine zipper were screened directly by DNA sequencing. For phenotypic analysis of the mutations *in vivo* or *in vitro*, mutant DNAs were subcloned as 800 bp *Kpn*I-*Eco*RI fragments into YCp88-GCN4.

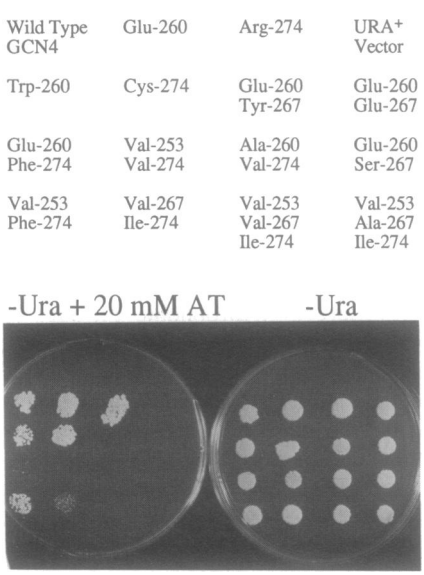

Figure 1. *In vivo* GCN4 activity of representative derivatives mutated at indicated positions from leucine to the indicated residues. Approximately 500 cells containing the indicated derivatives were spotted on plates containing minimal medium with appropriate supplements in the presence or absence of 20 mM AT. The Ura^+ vector represents a control plasmid lacking a functional GCN4 gene.

DNA binding experiments

In vitro synthesis of ^{35}S-labelled GCN4 protein and analysis of DNA-protein complexes were carried out essentially as described (1), except that the binding buffer contained poly (dI·dC) at a final concentration of 100 μg/ml instead of salmon sperm DNA. Roughly equivalent amounts of the various ^{35}S-proteins (estimated by SDS-PAGE) were incubated with 15 nM of a ^{32}P-labelled 631 bp *Pvu*II-*Rsa*I DNA fragment from pUC8-Sc4251 that contains the optimal binding site ATGACTCAT (3). The resulting protein-DNA complexes were separated by acrylamide gel electrophoresis and characterized as follows: +++, indistinguishable from wild-type; ++, slightly worse than wild-type; +, weak but detectable; −, no detectable binding. The conditions for the binding assay are such that the intensities of the bands are roughly proportional to the binding constants (1, 3, 7).

Phenotypic analysis

YCp88 DNA molecules containing the mutant GCN4 genes were introduced into KY803 cells (relevant genotype *ura3*-52, *gcn4*-Δ1) (7). Ura^+ transformants were selected on glucose minimal medium and the resulting strains were examined for GCN4 function by growth in the presence of 20 mM aminotriazole, a competitive inhibitor of the *his3* gene product. As shown previously (3, 7), growth under these conditions requires that the GCN4 derivative activate transcription of *his3* and other amino acid biosynthetic genes with the degree of AT resistance being related directly to the level of *his3* expression. Phenotypes are designated as follows: +++, indistinguishable from wild-type; ++, slightly worse than wild-type; +, weak but detectable; −, no growth.

RESULTS

Phenotypic analysis of single leucine substitutions

The C-terminal 33 amino acids of GCN4 (corresponding to residues 259−281) are sufficient for dimerization as an isolated peptide (12), when fused to the helix-turn-helix DNA-binding domain of λ repressor (25), or in the context of a heterologous bZIP domain (10, 11). However, insertion of two or four amino acids after Lys^{277}, which should disrupt a continuous α-helical coiled coil, does not affect GCN4 function *in vivo* (27). This suggests that the last 4 residues of GCN4 are not required for a functional leucine zipper.

In order to test the importance of each canonical leucine in the GCN4 zipper (positions 253, 260, 267, and 274), we obtained a wide variety of amino acid substitutions using a set of degenerate oligonucleotides. Mutated derivatives were identified by DNA sequencing and tested for GCN4 function *in vivo* by the standard complementation assay which involves growth in the presence of aminotriazole, a competitive inhibitor of the *his3* gene product. Complementation requires that the GCN4 derivative activate

transcription of *his3* and other amino acid biosynthetic genes, with the degree of aminotriazole resistance being directly related to *his3* mRNA levels (3, 7). The GCN4 derivatives are expressed constitutively from the *ded1* promoter at levels that are roughly comparable to those achieved during conditions of amino acid starvation (7).

Surprisingly, nearly all single substitution proteins tested appear functionally indistinguishable from wild-type GCN4 (Fig. 1; Table 1). Each of the four leucine residues can be changed, and a wide variety of substitutions are permitted including basic (arg^{267}) and acidic (glu^{260} and glu^{274}) amino acids. The gly^{267} derivative displays a reduced but clearly detectable level of function, probably a consequence of the α-helix destabilizing nature of glycine residues. In this regard, the set of single mutants does not include proline substitutions that are predicted to more seriously disrupt the α-helical nature of the leucine zipper. Nevertheless, it is clear that most (and probably nearly all) changes of any individual conserved leucine in the GCN4 zipper do not significantly affect the function of the protein *in vivo*.

Phenotypic analysis of multiple leucine substitutions

We also generated substitutions of two or three leucines by carrying out reactions with multiple oligonucleotides or with mutant templates containing single substitutions (Fig. 1; Table 1). Analysis of the resulting derivatives indicates that most of the double mutants and all of the triple mutants conferred no detectable GCN4 function *in vivo*. In particular, the double mutants $val^{253}val^{274}$ and $met^{253}ile^{274}$ fail to support growth in the presence of aminotriazole even though the individual substitutions confer no detectable phenotype. However, the double mutant $val^{253}phe^{274}$ displays moderate functional activity and the double substitutions $val^{267}ile^{274}$ and $glu^{260}phe^{274}$ confer detectable, but weak, GCN4 activity. As GCN4 function *in vivo* appears to be mediated by homodimers (15), this suggests that changing 4 out of the 8 leucines in the dimeric GCN4 zipper does not abolish functional activity.

Table 1. Properties of mutant proteins.

Protein	DNA-binding *in vitro*	Transcription *in vivo*
GCN4	+++	+++
val-253	NT	+++
phe-253	+++	+++
met-253	+++	+++
amber-260	NT	−
glu-260	+++	+++
ala-260	+++	+++
trp-260	++	++
val-260	NT	+++
amber-267	NT	−
val-267	+++	+++
arg-267	+++	+++
gly-267	++	++
phe-274	+++	+++
cys-274	+++	+++
val-274	+++	+++
ala-274	NT	+++
arg-274	+++	++
his-274	NT	+++
ile-274	+++	+++
val-253, val-274	−	−
val-253, phe-274	+	++
met-253, ile-274	NT	−
glu-260, ala-267	−	−
glu-260, tyr-267	NT	−
glu-260, glu-267	NT	−
glu-260, ser-267	NT	−
glu-260, phe-274	+	+
ala-260, val-274	−	−
val-260, arg-274	NT	−
val-267, ile-274	++	+
val-253, val-267, ile-274	NT	−
val-253, ala-267, ile-274	NT	−
val-253, lys-267, ile-274	NT	−

+++, indistinguishable from wild-type; ++, slightly lower than wild-type activity; +, weak but detectable activity; −, no detectable activity; NT, not tested.

DNA-binding *in vitro* by mutant GCN4 proteins

A variety of leucine substitution mutants were analyzed for DNA-binding activity using ^{35}S-labeled proteins synthesized *in vitro* and a DNA fragment containing the optimal GCN4 binding site ATGACTCAT (3). As shown in Fig. 2, DNA-binding ability of these and other (not shown) derivatives are in good accord with the predictions from the phenotypes *in vivo*. Nearly all the single mutants tested bind the target site with high affinity similar to that observed with wild-type GCN4. However, the Gly^{267} derivative, which shows a non-wild-type phenotype *in vivo*, binds with lower affinity. As expected, double mutants that confer some GCN4 activity *in vivo* bind weakly to DNA, whereas derivatives lacking GCN4 function do not bind detectably.

DISCUSSION

Mutational analyses of the Fos, Jun, and C/EBP leucine zippers (20−24) suggest that the leucine residues are very important for dimerization. Although individual leucines can sometimes be replaced by conservative residues (e.g. valine and isoleucine) with minimal decrease in the dimerization ability, non-conservative or multiple changes generally cause a severe functional defect. Our results indicate that in comparison to these other proteins, the GCN4 leucine zipper tolerates considerably more variation at amino acid positions corresponding to the canonical leucine residues. Nearly all the single substitution proteins behave similarly to wild-type GCN4 in terms of DNA-binding *in vitro*

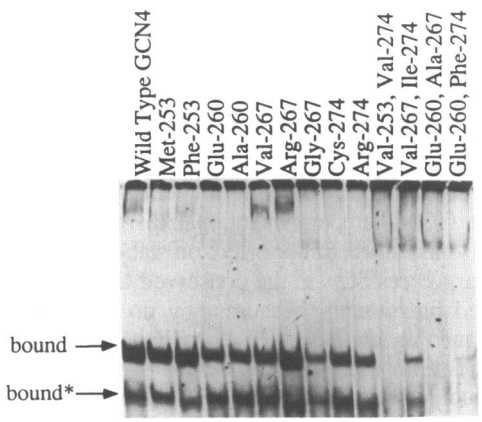

Figure 2. DNA-binding activities of representative derivatives mutated at indicated positions from leucine to indicated residues. *In vitro* synthesized proteins with the indicated amino acid substitution were incubated with the optimal GCN4 binding site, and protein-DNA complexes were separated from unbound DNA. The complex indicated as bound* represents a heterodimer between the full-length GCN4 derivative and an artifactual translation product containing the 175 C-terminal residues that is generated by aberrant initiation at an internal AUG codon.

and transcriptional activation *in vivo*. Functionally acceptable substitutions can occur at any of the four leucines and can involve residues that are acidic, basic, large, or small. Moreover, some of the double substitution proteins function detectably *in vivo* and *in vitro*. As both functional assays require DNA-binding by GCN4 homodimers (4, 15), 4 out of the 8 canonical leucines in the GCN4 zipper can be replaced without abolishing activity.

The role of the conserved leucines in the GCN4 zipper has been examined previously by measuring repression in *E.coli* mediated by a λ repressor-GCN4 leucine zipper fusion protein (25). Although our results are in general agreement with these earlier studies, three of the 'phenotypically silent' mutations described in this paper (arg^{267}, arg^{274}, and his^{274}) eliminate function of the λ repressor-GCN4 fusion protein (25). One possibility for this apparent discrepancy is that the artificial repression assay carried out in bacterial cells is more sensitive than the physiologically relevant of transcriptional activation assay in yeast cells. Aside from the obvious differences between these two assays (especially intracellular protein and target DNA concentrations), repression assays require high binding site occupancy *in vivo* and hence are typically more sensitive than activation assays. A second possibility is that structural (8, 16, 28) and functional (13, 15) interactions between the leucine zipper and basic region subdomains in GCN4 compensate for the effects of mutations that are inherently destabilizing in autonomous leucine zippers (i.e. those not associated with bZIP domains). We favor this possibility, particularly because it account for why the arg^{267}, arg^{274}, and his^{274} derivatives of GCN4 display wild-type DNA-binding activity *in vitro*. Thus, in terms of biological and biochemical significance, the GCN4 leucine zipper has a higher tolerance towards individual leucine substitutions than implied from the analysis of the λ repressor-GCN4 hybrid.

The surprising tolerance of non-leucine residues at the conserved positions does not exclude leucines as being important for dimerization, but rather indicates that other interactions within the zipper motif are crucial for efficient dimerization. In this regard, our results are consistent with the coiled coil model (12, 17, 18) in which the leucines do not interact with each other but rather with the interspersed hydrophobic residues (position a) of the opposing a helix. Indeed, mutational analysis of the GCN4 zipper in the context of a λ repressor-GCN4 hybrid protein indicates that these hydrophobic residues are functionally important (25). Although non-essential, individual leucines in the GCN4 zipper clearly make functional contributions because double substitutions strongly reduce or eliminate GCN4 function even though the individual mutations have little or no effect. As a consequence of the extensive set of interactions within the coiled coil of the GCN4 zipper, mutational disruption of individual interactions has only a minor effect on stability. In a coiled coil model, charged residues at the conserved leucine positions can be tolerated presumably because they do not disrupt α-helix formation and they are not in close proximity, thereby precluding electrostatic replusion.

If the leucine residues are not essential for a high degree of biological function, why are they so highly conserved? One possibility is that the GCN4 zipper is exceptionally stable because of multiple interactions involving the non-leucine residues; more typical bZIP proteins would have fewer such interactions, thus increasing the relative importance of the individual leucines. Another possibility is that the leucines might be conserved not for their inherent functional importance but rather for their compatibility with other residues in the homologous or heterologous zippers. In this sense, the conserved residues would increase the number of homodimeric and heterodimeric proteins and hence contribute to evolutionary and regulatory flexibility. Finally, as leucine zipper proteins are defined by the presence of the leucines, the apparent conservation might be an artifact due to circular reasoning. It seems likely that proteins lacking the conserved leucine repeat will utilize a structurally and functionally homologous zipper motif for dimerization. In this view, the current list of leucine zipper proteins would represent a subfamily of related proteins.

ACKNOWLEDGEMENTS

This work was supported by an N.I.H. postdoctoral fellowship to J.W.S., a Howard Hughes Medical Institute predoctoral fellowship to W.J.v.H. and by a grant to K.S. from the National Institutes of Health (GM 30186).

REFERENCES

1. Hope, I. A. & Struhl, K. (1985) *Cell* **43**, 177–188.
2. Arndt, K. & Fink, G. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8516–8520.
3. Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. (1986) *Science* **234**, 451–457.
4. Hope, I. A. & Struhl, K. (1987) *EMBO J.* **6**, 2781–2784.
5. Oliphant, A. R., Brandl, C. J. & Struhl, K. (1989) *Mol. Cell. Biol.* **9**, 2944–2949.
6. Sellers, J. W., Vincent, A. C. & Struhl, K. (1990) *Mol. Cell. Biol.* **10**, 5077–5086.
7. Hope, I. A. & Struhl, K. (1986) *Cell* **46**, 885–894.
8. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990) *Nature* **347**, 575–578.
9. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
10. Kouzarides, T. & Ziff, E. (1989) *Nature (Lond).* **340**, 568–571.
11. Sellers, J. W. & Struhl, K. (1989) *Nature* **341**, 74–76.
12. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989) *Science* **243**, 538–542.
13. Agre, P., Johnson, P. F. & McKnight, S. L. (1989) *Science* **246**, 922–926.
14. Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* **249**, 769–771.
15. Pu, W. T. & Struhl, K. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6901–6905.
16. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* **249**, 774–778.
17. Rasmussen, R., Benvegnu, D., O'Shea, E. K., Kim, P. S. & Alber, T. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 561–564.
18. O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991) *Science* **254**, 539–544.
19. O'Shea, E. K., Rutkowski, R., Stafford, W. F. I. & Kim, P. S. (1989) *Science* **245**, 646–648.
20. Kouzarides, T. & Ziff, E. (1988) *Nature* **336**, 646–651.
21. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1989) *Science* **243**, 1681–1688.
22. Turner, R. & Tjian, R. (1989) *Science* **243**, 1689–1694.
23. Gentz, R., Rauscher, F. I., Abate, C. & Curran, T. (1989) *Science* **243**, 1695–1699.
24. Schuermann, M., Neuberg, M., Hunter, J. B., Jenuwein, T., Ryseck, R.-P., Bravo, R. & Muller, R. (1989) *Cell* **56**, 507–516.
25. Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. (1990) *Science* **250**, 1400–1403.
26. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Meth. Enzymol.* **154**, 367–382.
27. Pu, W. T. & Struhl, K. (1991) *Mol. Cell. Biol.* **11**, 4918–4926.
28. Patel, L., Abate, C. & Curran, T. (1990) *Nature* **347**, 572–575.