Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxyl-terminal region of a protein coded for by the oncogene *jun*

(amino acid sequence/transcriptional activator)

PETER K. VOGT*, TIMOTHY J. BOS*, AND RUSSELL F. DOOLITTLE†

*Department of Microbiology, University of Southern California School of Medicine, 2011 Zonal Avenue, Los Angeles, CA 90033; and †Department of Chemistry, University of California, San Diego, La Jolla, CA 92093

Contributed by Peter K. Vogt, December 24, 1986

ABSTRACT   The product of the recently described oncogene *jun* shows significant amino acid sequence homology with the GCN4 yeast transcriptional activator protein. The similarity is restricted to the 66 carboxyl-terminal amino acids, thought to be the DNA-binding domain of the GCN4 protein. In these α-helix-permissive regions of the *jun* and *GCN4* products there is also a lesser but still significant amino acid resemblance to the fos protein and a marginal degree of similarity to myc proteins. The amino acid sequence homology between *GCN4* and *jun* gene products suggests that the jun protein may bind to DNA in a sequence-specific way and exert a regulatory function.

jun is a cell-derived genetic insert identified recently in avian sarcoma virus 17; it probably functions as the oncogenic determinant of this virus (1, 2). In contrast to the oncogenes of other avian sarcoma viruses, *jun* does not show any sequence relationship to tyrosine-specific protein kinases. Nor does nucleic acid hybridization detect homology between *jun* and other known oncogenes. In ASV 17-transformed cells *jun* appears to be expressed as a *gag–jun* fusion product.

We have compared the amino acid sequence of jun protein with a large collection of computer-stored amino acid sequences of proteins and found significant similarity between the jun protein and the yeast regulatory protein GCN4. GCN4 is a component of the yeast general control system that regulates the expression of amino acid-synthesizing enzymes in response to extracellular amino acid concentrations (3–6). The system consists of several interacting genes that code for positive and negative trans-acting effector proteins. GCN4 is a DNA-binding protein that functions as a transcriptional activator.

MATERIALS AND METHODS

The amino acid sequence of jun protein has been published in ref. 2. Other amino acid sequences were taken from either the NEWAT sequence bank (7) or release 9.0 of the National Biomedical Research Foundation Protein Identification Resource (8). The programs used for searching and alignment have been described (9, 10). Secondary structure predictions were made using the systems of Chou and Fasman (11) and Garnier *et al.* (12).

RESULTS

Homology Between Amino Acid Sequences of GCN4 and jun Proteins Exists in the 66 Carboxyl-Terminal Residues of

GCN4. Fig. 1 shows an alignment of the amino acid sequences of the GCN4 and jun proteins. The GCN4 protein is 281 amino acids long. The *jun* sequence can code for 296 amino acids. The alignment introduces gaps in both sequences and was computer generated (9). Of the 66 carboxyl-terminal amino acids of the GCN4 protein, 31 are identical with residues in the jun protein. The similarity of nucleotide sequence in these regions, when optimal amino acid alignment is used, is 37%. Homology between the jun and GCN4 proteins is also demonstrated by the diagonal plot shown in Fig. 2. Significant homology is confined to the carboxyl-terminal fourth of the proteins. Matching sections in the remainders of the sequences cannot be distinguished from random.

The GCN4 protein has two functions located in different domains of the protein: DNA binding and transcriptional activation (13). The DNA-binding function depends on the integrity of the carboxyl-terminal segment of 60 amino acids (13). This highly basic region is also the domain showing homology to jun protein. Transcriptional activation by GCN4 requires a 19-amino acid segment from the middle of the molecule (residues 106 to 125 counted from the amino terminus). This segment is located in the center of a very acidic region of the GCN4 protein (13). The jun protein does not contain a comparable concentration of acidic residues in this region of the molecule.

Proteins Coded for by the Oncogenes *fos* and *myc* also Have Segments That Are Similar to Segments of the GCN4 and jun Proteins. Fig. 3 presents a computer-generated alignment of the amino acid sequences of *GCN4, jun, fos*, and human c-*myc* gene products. The 71 carboxyl-terminal amino acids of GCN4 that show homology to the jun sequence contain 11 matches with human c-myc. The incidence of these matches is on the borderline of significance. However, in the same carboxyl-terminal segment there are 18 amino acids (22%) identical between jun and fos and 20 (28%) between GCN4 and fos, highly suggestive of homology (Figs. 3 and 4). Thus, *fos, jun*, and *GCN4* may be evolutionarily related. We also searched for similarity among other nuclear DNA-binding oncogene product. The myb, ski, and E1A proteins do not show any detectable relatedness to jun.

Conformational Predictions of jun, GCN4, and fos Proteins Suggest α-Helical Structure in the Areas of Homology. Secondary structure predictions for the GCN4, fos, and myc proteins indicate that the four similar segments are all highly helix-permissive. The two predictive systems used (11, 12) give reasonably concordant but not identical results (Fig. 5). It has been pointed out that the terminal 60 amino acids of the GCN4 protein are helix-permissive (13). The homologous region of the jun protein is also strongly helix-permissive by these same criteria. The same predicted α-helical structure is also seen in the suggested area of similarity between jun, fos, and myc.

```
(jun)            V P P L R G L C S M S   A K M E   P T F Y E D A L N  24
(GCN4)  M S E Y Q P S L F A L N P M G F S P L D G S K S T N E N V S   30

        A S F A P P E S G G Y G Y N N A D I L T S P D V G L L K L A   54
        A S T S T A K P M V G Q L I F D K F I K T E E D P I I K Q D   60

        S P E L E R L I I Q S S N G L I T T T P T P T Q F L C P K N   84
        T P     S N L D F D F A L P Q T A T A P D A K T V L P I P E   88

        V T D E   Q E G F A E G F V R A           L A E L H N Q   N  107
        L D D A V V E S F F S S S T D S T P M F E Y E N L E D N S K  118

        T L P S V T S A A Q P V S G G M A P V S S M A G G G S F N T  137
        E W T S L F D N D I P V T T D D V S L A D K A I E S T E E V  148

        S L H S E P P V Y A N L S N F N P N A L N S A P N Y N A N R  167
        S L V P     S N L E V S T T S F L P T P V L E D A K L T Q T R 177

        M G Y A P Q H H I N P Q M P V Q H P R L Q A L K E E P Q T V  197
        K V K K P N S V V K K S H H V G K D D E S R L D H L G V V A  207

        P E M P G E T P P L F P I D M E S Q E R I K A E R K R M R N  227
        Y N R K Q R S I P L S P I V P E S S D P     A A L K R A R N  235

        R I A A S K S R K R K L E R I A R L E E K V K T L K A Q N S  257
        T E A A R R S R A R K L Q R M K Q L E D K V             E    258

        E L A S T A N M L R E Q V A Q L K Q K V M N H V N S G C Q L  287
        E L L S K N Y H L E N E V A R L K K L V G E R              281

        M L T Q Q L Q T F                                            296
```

FIG. 1.   Amino acid sequence alignment of *jun* and *GCN4* gene products. The standard single-letter code is used. Dots mark identical residues.

DISCUSSION

The properties of the GCN4 protein have been studied in detail over the past few years (9, 11, 13). Because of the significant similarity between the carboxyl-terminal segments of the GCN4 and jun proteins, some observations made with GCN4 may be useful as predictors for functions of
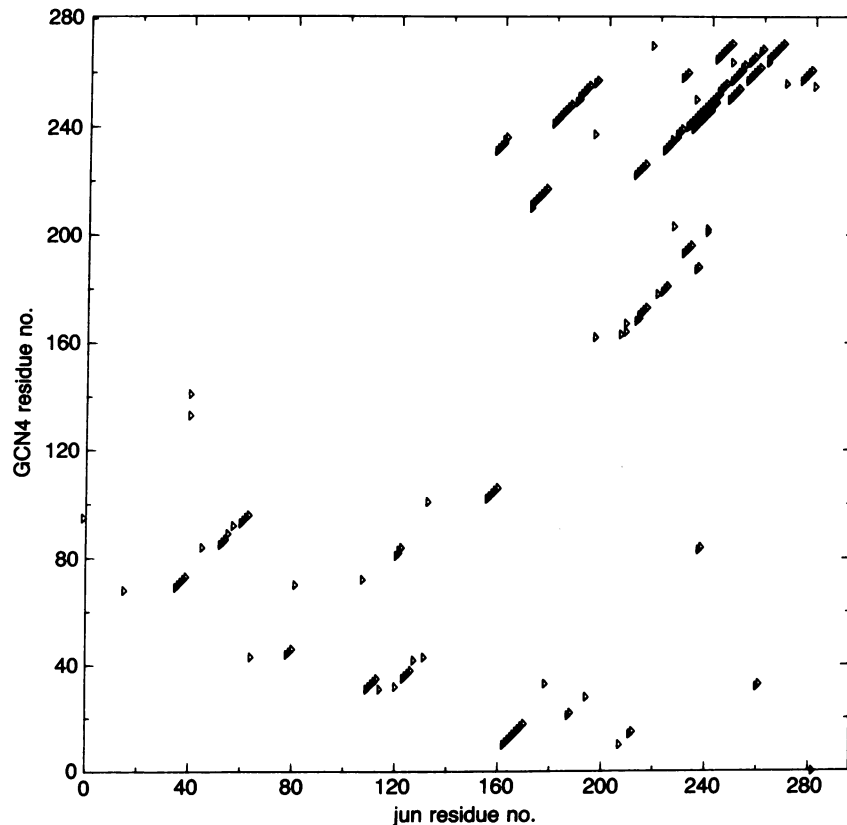

FIG. 2.   Diagonal plot of amino acid matches. Numbering of residues starts at the amino terminus of the proteins.

```
juna  ..P G E T P P L F P I D M E S Q E R I K A E R K R M R N R I A A S K S R
rgby  . K Q R S I P L S P I V P E S S D   P A A L K R A R N T E A A R R S R
vfos  ..I G R R G K V E Q L S P E E E E   K R R I R R E R N K M A A A K C R
cmyh  . K R R T H N V L E R Q R R N E L   K R S F F A L R D Q I P E L E N N
```

```
juna  K R K L E R I A R L E E K V K T L K A Q N S E L A S T A N M L R E Q V
rgby  A R K L Q R M K Q L E D K V   E E L L S K N Y H L E N E V
vfos  N R R R E L T D T L Q A E T   D Q L E D K K S A L Q T E I
cmyh  E K A P K V V I L K K A T   A Y I L S V Q A E E Q K L I S E E
```

```
juna  A Q L K Q       K V M N H V N S G C Q ..
rgby  A R L K K       L V G E R
vfos  A N L L K E K E K L E F I L A A H R P A   C K ..
cmyh  D L L R K R R E Q L K H K L E Q L R N S   C A
```

FIG. 3. Alignment of amino acid sequences comparing products of *jun* (juna), *GCN4* (rgby), v-*fos*, and human c-*myc* (cmyh). Circles denote amino acid matches.

	jun (n = 86)	*GCN4* (n = 71)	*v-fos* (n = 83)	*c-myc* (n = 85)
jun		31	18	11
GCN4	44%		20	11
v-fos	22%	28%		17
c-myc	13%	15%	20%	

Percent identical

(right axis: Matches of amino acid residues)

FIG. 4. Summary of comparisons between the amino sequences of *jun*, v-*fos*, c-*myc*, and *GCN4* gene products. The upper right of the matrix represents the number of identical amino acid residues in the compared segments. In the lower left these matches are expressed as percent identity.

the jun protein. The GCN4 protein has two functions that reside in different domains of the molecule: DNA binding, for which the carboxyl-terminal 60 amino acids are needed, and transcriptional activation, which requires the intactness of a 19-residue peptide in the acidic middle of the molecule (13).

DNA binding of the GCN4 protein is sequence specific and thought to be mediated through a helix–turn–helix structure (13–16). Schemes for the prediction of secondary structure also reveal a strong preference for α-helical conformation in the carboxyl-terminal region of jun, which is homologous to the GCN4 DNA-binding domain. We suggest that the jun protein binds to DNA, possibly in a sequence-specific way. The target sequences to which the GCN4 protein binds have been determined (14–16). Similar target sequences may exist for the jun protein and conceivably could be defined by using large quantities of purified jun protein produced by an expression vector. It might even be possible that clues as to possible targets of jun might also be provided by using the DNA sequence to which the GCN4 protein binds as a probe to search regulatory sequences of vertebrate genes in computer data banks. The targets of jun are likely not amino acid synthesizing enzymes. Unlike yeast cells, animal cells do not synthesize their own amino acids *de novo* (17). That the jun proteins could interact with target genes quite different from those of the GCN4 system is also suggested by a lack of detectable similarity outside the carboxyl-terminal one-fourth of the molecule. Possible interactions with different RNA polymerases may also account for the divergence between the jun protein and GCN4 outside the carboxyl-terminal portion. Nonetheless, jun could be an important regulatory protein in animal cells. Such a function would be in accord with oncogenic potential. The relatedness between

```
A  PGETPPLFPIDMESQERIKAERKRMRNRIAASKSRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQ*
   t    ttttxxxxxxxxxxxxxxxxxxxxxxxxxttttxxxxxxxxxxxxxxxxxxxxxxxxttttxxxxxxxxxxxxxxxxxxxxxxttt
        t xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx..


B  KQRSIPLSPIVPESSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER*
   t    tttt    ttttxxxxttttxxxxttttxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
        ...t xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx


C  IGRRGKVEQLSPEEEEKRRIRRERNKMAAAKCRNRRRELTDTLQAETDQLEDKKSALQTEIANLLKEKEKLEFILAAHRPACK*
   t xxxxxxxttttxxxxxxxxttttxxxxxxxxttttxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxttt
       x  xxxxxxxxxxxxxxxxxxxxttttt      xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx


D  KRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA*
   t  xxxxxxxxxxxxxxxxxxxxxxxxttttxxx.................xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       xxxxxt  ttxxxxxt    xxxxxxxxxxxxxxxxxxx...xxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

FIG. 5. Predicted secondary structure of the carboxyl-terminal segments of the jun protein (*A*), GCN4 protein (*B*), the v-fos protein (*C*), and human c-myc protein (*D*). x denotes helix-permissive structure; dots denote β-sheet; and t denotes turn, with the most likely turn positions underlined. The upper line of conformational predictions was generated by the program of Chou and Fasman (11), the lower line, by the program of Garnier *et al.* (12).

jun and *fos* is significant, and the matches between *myc* and *fos* and *myc* and *jun*, taken together, suggest that the myc protein may also belong in this same family of regulatory proteins.

1. Cavalieri, F., Ruscio, T., Tinoco, R., Benedict, S., Davis, C. & Vogt, P. K. (1985) *Virology* **143**, 680–683.
2. Maki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2848–2852.
3. Hinnebusch, A. G. & Fink, G. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5374–5378.
4. Penn, M. D., Galgoci, B. & Greer, H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2704–2708.
5. Lucchini, G., Hinnebusch, A. G., Chen, C. & Fink, G. R. (1984) *Mol. Cell. Biol.* **4**, 1326–1333.
6. Fink, G. R. (1986) *Cell* **45**, 155–156.
7. Doolittle, R. F. (1981) *Science* **214**, 149–159.
8. George, D. G., Barker, W. C. & Hunt, L. T. (1986) *Nucleic Acids Res.* **14**, 11–15.
9. Feng, D. F., Johnson, M. S. & Doolittle, R. F. (1985) *J. Mol. Evol.* **21**, 112–125.
10. Fong, H. K. W., Hurley, J. B., Hopkins, R. S., Miake-Lye, R., Johnson, M. S., Doolittle, R. F. & Simon, M. I. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2162–2166.
11. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245.
12. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
13. Hope, I. A. & Struhl, K. (1986) *Cell* **46**, 885–894.
14. Hinnebusch, A., Lucchini, G. & Fink, G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 498–502.
15. Hope, I. A. & Struhl, K. (1985) *Cell* **43**, 177–188.
16. Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. (1986) *Science* **234**, 451–457.
17. Eagle, H. & Levintow, L. (1965) in *Cells and Tissues in Culture*, ed. Willmer, E. N. (Academic, New York), pp. 277–296.