The EMBO Journal vol.6 no.9 pp.2781–2784, 1987

GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA

Ian A.Hope[1] and Kevin Struhl

Department of Biological Chemistry, Harvard Medical School, Boston, MA 02115, USA

[1]Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Communicated by I.Grummt

The eukaryotic transcriptional activator protein, GCN4, synthesized *in vitro* from the cloned gene, binds specifically to the promoters of yeast amino acid biosynthetic genes. Previous analysis of truncated GCN4 derivatives localized the DNA binding domain to the C-terminal 60 amino acids and revealed that the size of the GCN4 derivative and the electrophoretic mobility of the protein–DNA complex were inversely related. This observation was utilized here to develop a novel method for determining the subunit structure of DNA binding proteins. A mixture of wild-type GCN4 protein and a smaller GCN4 derivative generated three complexes with DNA, two corresponding to those observed when the proteins are present individually and one new complex of intermediate mobility. This extra complex results from the heterodimer of the two GCN4 proteins of different sizes, demonstrating that GCN4 binds DNA as a dimer. The contacts sufficient for dimerization were localized to the 60 C-terminal amino acid, DNA binding domain, suggesting that dimerization of GCN4 is a critical aspect of specific DNA binding. Furthermore, stable GCN4 dimers were formed in the absence of target DNA. These observations suggest a structural model of GCN4 protein in which a dimer binds to overlapping and non-identical half-sites, explaining why GCN4 recognition sites act bidirectionally in stimulating transcription.

Key words: DNA binding proteins/promoters/protein structure/*Saccharomyces cerevisiae*/transcriptional activation

Introduction

In general, prokaryotic activator proteins involved in transcriptional induction have a symmetric subunit structure and bind to DNA elements of palindromic sequence (McKay and Steitz, 1981; Pabo and Lewis, 1982). Such structural information has not yet been obtained for a eukaryotic transcriptional activator protein. The GCN4 protein of the yeast *Saccharomyces cerevisiae* binds specifically to the promoters of amino acid biosynthetic genes and coordinately induces their transcription (Hope and Struhl, 1985). Deletion analysis of GCN4 protein indicates that the 60 C-terminal residues are sufficient for specific DNA binding and that a separate 19-amino acid region of acidic character is critical for transcriptional activation (Hope and Struhl, 1986). Saturation mutagenesis of a target sequence within the *HIS3* promoter defines a 9-bp region as the major determinant for GCN4 binding, with optimal binding to DNA containing the palindrome ATGA(C/G)TCAT (Hill *et al.*, 1986). This palindrome also represents the consensus of presumptive regulatory sites from 15 genes subject to coordinate induction by GCN4 (Hill *et al.*, 1986).

In previous work, we have described the use of *in vitro* synthesized proteins for DNA binding studies (Hope and Struhl, 1985). Specifically, radioactively pure GCN4 protein (labeled with [^{35}S]methionine) is synthesized by *in vitro* translation of mRNA produced by *in vitro* transcription of cloned DNA using bacteriophage SP6 RNA polymerase. To detect DNA binding activity, the labeled protein is incubated with a DNA fragment from the *HIS3* promoter region, and protein–DNA complexes are separated from unbound protein and DNA by electrophoresis in non-denaturing polyacrylamide gels.

In the course of experiments involving deleted versions of

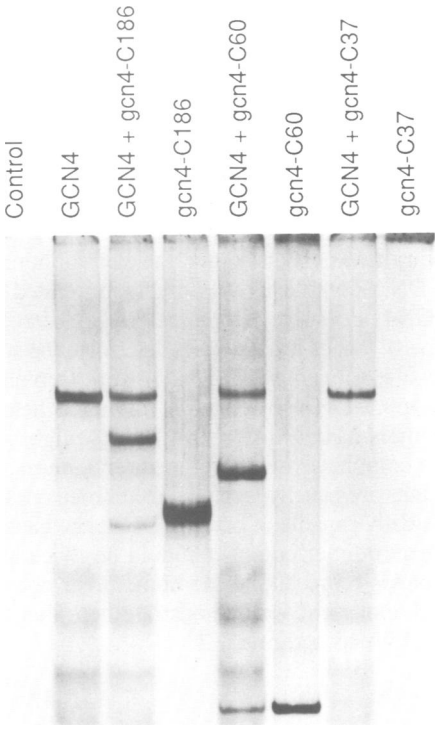

Fig. 1. GCN4 protein–DNA complexes. The protein(s) indicated above each track were generated by transcription and translation *in vitro*. Translation products were incubated with the *Taq*I DNA fragment from the upstream region of *HIS3* prior to non-denaturing PAGE and autoradiography. The relative molarities of the complexes were determined by scanning the autoradiogram with a DU6 spectrophotomer (Beckman), and normalizing the band intensities to the number of methionine codons for each protein derivative. For the GCN4 + gcn4-C186 lane, the relative band intensities of the $(GCN4)_2$, (GCN4/gcn4-C186), and $(gcn4\text{-}C186)_2$ complexes were 5.6:8.8:2.3. Therefore, the molar ratio of these complexes was 1.1:2.2:0.8 because the GCN4 and gcn4-C186 monomers contain five and three methionine residues respectively. The control track was generated using pSP64 vector DNA (no GCN4 coding region present) to program the transcription plus translation reactions. The bands observed in each track were not observed if the *HIS3* DNA fragment was omitted from the incubation (data not shown).

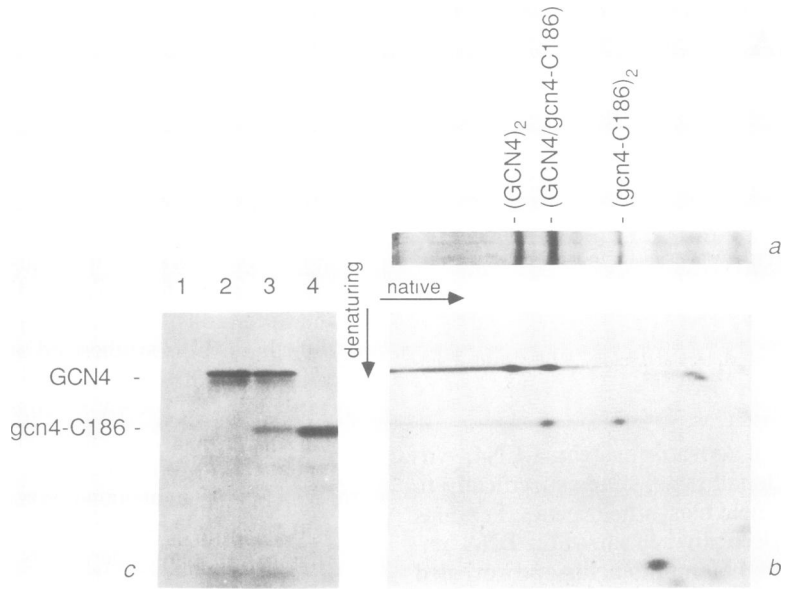

Fig. 2. Two-dimensional gel electrophoresis of protein−DNA complexes. (**a**) The complexes generated from the GCN4 + gcn4-C186 protein mixture and the *HIS3* DNA fragment were detected as described in Figure 1. The positions of the homodimer and heterodimer protein−DNA complexes are indicated. The direction of the non-denaturing PAGE in the first dimension is indicated by the horizontal arrow. (**b**) A gel track equivalent to that shown in panel (a) was excised and subject to SDS−PAGE in the second dimension (direction indicated by the vertical arrow) on a 15% gel. Protein positions were revealed by autoradiography. The different intensities observed for GCN4 and gcn4-C186 when separated from the heterodimer complex occur because gcn4-C186 contains fewer methionine residues than the full-length protein. (**c**) The proteins used to generate the tracks in Figure 1 marked control, GCN4, GCN4 + gcn4-C186 and gcn4-C186 are examined in tracks 1−4, respectively, by SDS−PAGE on a 15% gel, with autoradiography, for comparison with the two-dimensional gel in panel (b). The positions of GCN4 and gcn4-C186 are indicated.

GCN4 protein, it was observed that the electrophoretic mobility of protein−DNA complexes was inversely related to the mol. wt of the GCN4 derivative (Hope and Struhl, 1986). In principle, this property makes it possible to determine the subunit structure of GCN4. For example, if GCN4 were to bind DNA as a dimer, a protein−DNA complex involving a heterodimer of GCN4 monomers having different mol. wts might migrate between the two complexes involving the homodimers. When both monomeric units are equally labeled and all three possible dimeric species are equally capable of binding DNA, the band corresponding to the heterodimer complex should be twice as intense as each of the bands representing the homodimer complexes. Using this novel approach, we demonstrate here that GCN4 does indeed bind DNA as a dimer.

Results

GCN4 binds DNA as a dimer

To determine the stoichiometry of binding, we generated a mixture of wild-type GCN4 protein and gcn4-C186, a derivative that lacks the N-terminal 95 amino acids but retains full GCN4 function *in vivo* (Hope and Struhl, 1986). The two proteins were synthesized simultaneously in the same reaction mixture by combining the DNA templates prior to transcription and translation *in vitro*. When incubated with the *HIS3* DNA fragment and examined by non-denaturing polyacrylamide gel electrophoresis (PAGE), the GCN4 and gcn4-C186 mixture generated three strong bands resulting from complex formation (Figure 1). The upper and lower bands respectively have the same mobilities as those of complexes involving GCN4 alone and gcn4-C186 alone. The third band has an intermediate mobility, suggesting that complexes were formed with heterodimers of GCN4 and gcn4-C186. Spectrophoretic scanning of the autoradiogram indicates that the three complexes are present in the molar ratio of 1.1:2.2:0.8, in excellent agreement with the prediction for GCN4 binding as a dimer.

Two-dimensional gel electrophoresis was used to demonstrate that the intermediate band does indeed result from a DNA−protein complex involving a heterodimer between GCN4 and gcn4-C186. For the first dimension, complexes generated by the GCN4 and gcn4-C186 protein mixture were electrophoretically separated in a non-denaturing gel as shown in Figure 1. The gel track was excised and the proteins that compose the three complexes were examined in the second dimension by SDS-PAGE (Figure 2). The upper complex band contains only GCN4, the lower complex band contains only gcn4-C186 and the intermediate complex band contains both GCN4 and gcn4-C186, precisely as predicted.

The observation of three bands whose relative intensities are in a 1:2:1 ratio is completely consistent with GCN4 binding as a dimer to the single site in the *HIS3* target DNA, and inconsistent with GCN4 binding as a monomer or as a multimer with a larger number of subunits. The unlikely possibility that GCN4 exists as an oligomer but extra intermediate bands were not detected because particular complexes lack DNA binding activity is excluded by the relative band intensities. For example, if GCN4 exists as a tetramer but only mixed 2:2 (but not 3:1 or 1:3) tetramers are capable of binding, the expected ratio of heterotetramers to homotetramers would be 6:1. In addition, as the experiments are performed under conditions where most of the protein is driven into complexes with DNA (Hope and Struhl, 1985, 1986), the existence of non-binding species would result in an overall decrease in band intensities when protein mixtures are used; this was not observed. Therefore GCN4 interacts specifically with *HIS3* DNA as a dimer.

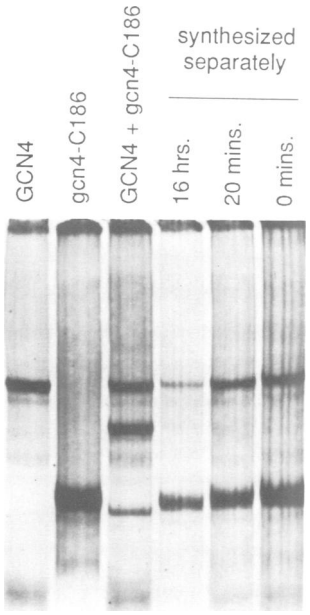

Fig. 3. Stability of GCN4 dimers. Protein–DNA complexes were examined as described in Figure 1. For the first three tracks the protein preparation used in the incubation is indicated above each track. For the last three tracks GCN4 and gcn4-C186 (10 000 acid-precipitable c.p.m. of each) had been synthesized separately and were: (i) mixed and stored for 16 h at 4°C prior to incubation with DNA, or (ii) mixed immediately before, and so could interact during, the 20-min incubation with the DNA, or (iii) mixed after incubation with the DNA and immediately before loading on to the gel (0 min).

Dimerization occurs through the DNA binding domain

To determine what part of GCN4 was sufficient for dimer formation, smaller derivatives were examined in the same way, gcn4-C60 and gcn4-C37 are derivatives consisting respectively of just the 60 and 37 C-terminal amino acids of GCN4 (Hope and Struhl, 1986). In the DNA binding assay, GCN4 plus gcn4-C60 generated three distinct protein–DNA complexes (Figure 1) indicating that the 60 C-terminal amino acids are capable of dimer formation. In contrast, gcn4-C37, which does not bind DNA (Hope and Struhl, 1986), failed to generate a heterodimer complex with GCN4. These observations show that GCN4 can dimerize through the C-terminal 60 amino acids and suggest that dimerization requires an intact DNA binding domain. Although subunit interactions involving other segments of GCN4 protein cannot be ruled out, such interactions, if present, are not necessary for dimerization.

Stable GCN4 dimers are formed in the absence of DNA

GCN4 could be a stable dimer when free in solution and or dimerize only upon binding to DNA. This was investigated by synthesizing GCN4 and gcn4-C186 in separate reactions prior to combining them and assaying for heterodimer complex formation (Figure 3). Even when the two proteins were combined and then stored for 16 h at 4°C before performing the DNA binding assay, complexes due to heterodimers were barely detectable. In other words, the heterodimer complex was detected only when GCN4 and gcn4-C186 were synthesized in the same reaction. The apparent absence of subunit interchange even after 16 h of incubation indicates that stable GCN4 dimers are formed in the absence of target DNA.

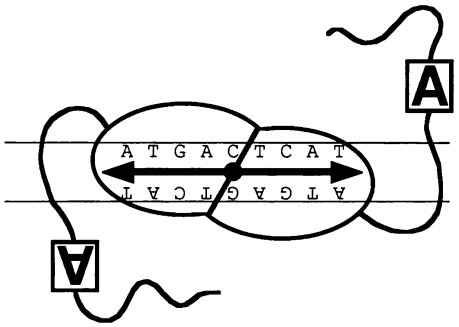

Fig. 4. Structural model for the GCN4 dimer interacting with target DNA. The DNA is depicted by the thin lines, along with the sequence of the optimal binding site and arrows emanating from the central C:G base pair (closed circle) to indicate the symmetry of the site. The bold lines repesent GCN4 protein. The dimerized DNA binding domains are depicted by the fused ovals, the transcriptional activation regions by the boxes marked A, and the rest of the protein by the wavy lines. The GCN4 monomers are drawn as interacting with overlapping and non-equivalent half-sites (see text). Precise details about which nucleotides are contacted by each monomer are not yet known and so the relationship indicated here should only be taken as one possibility.

Discussion

From the data presented here as well as the functional dissection of the 281-amino acid GCN4 protein (Hope and Struhl, 1986) and the target DNA site in the *HIS3* promoter (Hill *et al.*, 1986), we propose a structural model for GCN4 interacting with DNA (Figure 4). The 60 C-terminal amino acids of GCN4 constitute an independently structured domain that dimerizes and associates strongly with DNA of a specific sequence. Interestingly, although dimer formation appears to be critical for DNA binding, stable dimers can be formed even in the absence of target DNA. The remainder of the protein is viewed as two flexible 'arms' lacking functionally important tertiary structure because large regions of the N-terminal 221 amino acids can be deleted without affecting GCN4 function *in vivo* or *in vitro* (Hope and Struhl, 1986; J.P.Macke and K.Struhl, unpublished results). A 19-amino acid region of acidic character in the centre of these arms (amino acids 107–125) is critical for transcriptional activation (Hope and Struhl, 1986).

Previous results, obtained by direct DNaseI footprinting and analysis of a large set of point mutations of the *HIS3* regulatory site, indicate that interactions between GCN4 and target DNA are confined primarily to a 9-bp sequence (Hope and Struhl, 1985; Hill *et al.*, 1986). Optimal GCN4 binding was observed to *HIS3* DNA segments with the palindromic sequence ATGA(C/G)TCAT, suggesting that the GCN4 dimer is arranged symmetrically with respect to the target DNA. However, several lines of evidence suggest that protein–DNA recognition involves non-identical contacts between GCN4 monomers and half-sites in the target DNA. First, neither that native *HIS3* site nor any of the presumptive regulatory sequences in 14 other promoters activated by GCN4 are perfectly symmetric (Hill *et al.*, 1986). Second, some symmetrical changes of the *HIS3* regulatory site do not have equivalent effects on DNA binding affinity or transcriptional activation (Hill *et al.*, 1986). Third, GCN4 binding is reduced significantly when the central C of the *HIS3* site is changed to any other base including G, its symmetric counterpart (Hill *et al.*, 1986). This suggests that the central base pair is part of a half-site recognized by a GCN4 monomer, and given the odd number of base pairs in the palindrome, it follows that the protein–DNA interactions at the half-sites cannot possibly

be identical, even for the optimal sequence. These considerations also suggest that the half-sites overlap at the central base pair, and the overlap might conceivably be more extensive (Figure 4). In this regard, it is interesting that the 9-bp GCN4 binding site is relatively small in comparison to the 15- to 20-bp target sites recognized by other DNA binding proteins of similar mol. wt (Anderson *et al.*, 1981; McKay and Steitz, 1981; Pabo and Lewis, 1982; Joachimiak *et al.*, 1983).

A striking property of eukaryotic upstream promoter elements is their ability to bidirectionally activate mRNA initiation, an inherently unidirectional process. In the case of transcriptional activation by GCN4, the situation is now clarified by the observations that the target DNA sequences have palindromic character (Hill *et al.*, 1986) and that stable GCN4 dimers can form in the absence of target DNA prior to binding. Thus, when bound to a regulatory element, GCN4 would be symmetrically arranged with respect to the DNA template and hence be capable of activating transcription equally in either direction, even though the protein–DNA contacts cannot be identical for the half-sites of a single promoter. If, for example, GCN4 stimulates transcription through contacts to RNA polymerase II and/or to proteins binding to the TATA promoter element, these interactions would not depend on the orientation of the GCN4 binding sites. Flexibility in the DNA (Ptashne, 1986) and in GCN4 protein (Hope and Struhl, 1986) would make such interactions possible over the long and variable distances that have been observed (Struhl, 1982; Hinnebusch *et al.*, 1985; Hill *et al.*, 1986). Thus, bidirectional activation of eukaryotic mRNA initiation may reflect the DNA binding properties of activator proteins rather than the actual mechanism of transcriptional activation.

Materials and methods

In vitro transcription and translation

The DNA templates used to generate GCN4 and the N-terminally deleted gcn4 derivatives have been previously described (Hope and Struhl, 1986) and consist of the coding region of the GCN4 gene deleted using BAL31 and cloned into pSP64 (Melton *et al.*, 1984). Transcription *in vitro* was as previously described (Melton *et al.*, 1984) with 0.5 μg DNA template in a 25-μl reaction containing ATP, CTP, TTP and G-5′ppp5′-G at 500 μM, GTP at 50 μM and with 5 units of SP6 RNA polymerase. For the protein mixtures 0.5 μg of each of the two template DNAs were used. The RNA produced was extracted with phenol and precipitated with ethanol prior to translation *in vitro* using a wheat germ extract (not nuclease treated) as directed by the manufacturers (Bethesda Research Laboratories). The total RNA from a transcription reaction was translated in 30 μl with 16 μCi of [^{35}S]methionine (1400 Ci/mmol). To measure protein synthesis, 1 μl of the translation products plus 50 μl of 0.1 M NaOH was incubated for 15 min at 37°C, precipitated in 1 ml 10% trichloroacetic acid for 15 min on ice, the precipitate was collected on a glass fibre filter and the incorporated [^{35}S]methionine was measured by scintillation counting using aquasol (New England Nuclear).

Electrophoresis

DNA–protein complex formation was examined by non-denaturing PAGE as previously described (Hope and Struhl, 1985); 20 000 trichloroacetic acid-precipitable c.p.m. of translation products (without further purification) were incubated with 30 ng of the 294-bp *TaqI* DNA fragment from the upstream region of *HIS3* (Hope and Struhl, 1985) which had been rendered blunt ended by treatment with the large fragment of *Escherichia coli* DNA polymerase I. Incubation was for 20 min at 25°C in 15 μl of 20 mM Tris (pH 7.4), 50 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 100 μg/ml gelatin and 50 μg/ml sonicated salmon sperm DNA (binding buffer). The samples were mixed with 5 μl binding buffer containing 20% glycerol, 1 mg/ml xylene cyanol FF and 1 mg/ml bromophenol blue and loaded onto a 5% polyacrylamide gel (30 parts acrylamide:0.8 parts bisacrylamide) prepared and run in 90 mM Tris–borate buffer (pH 8.3). Electrophoresis was at 400 V until the samples had entered the gel and then 175 V until the bromophenol blue had migrated the length of the gel (20 cm). The gel was fixed, treated with en^{3}Hance and autoradiographed.

For two-dimensional electrophoresis the track excised from the non-denaturing gel was washed twice for 15 min in SDS-PAGE sample buffer at 25°C and then loaded across the top of a gel for SDS-PAGE. SDS-PAGE was carried out on a 15% gel using the discontinuous buffer system. After electrophoresis, proteins were detected by fixation, treatment with en^{3}Hance and autoradiography.

Acknowledgements

We thank Phil Youdarian and Jay Hirsh for useful suggestions on the experiments. This work was supported by grants to K.S. from the National Institutes of Health (GM30186) and from the Chicago Community Trust (Searle Scholars Program).

References

Anderson,W.F., Ohlendorf,D.H., Takeda,Y. and Matthews,B.W. (1981) *Nature*, **290**, 754–758.

Hinnebusch,A., Lucchini,G. and Fink,G. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 498–502.

Hill,D.E., Hope,I.A., Macke,J.P. and Struhl,K. (1986) *Science*, **234**, 451–457.

Hope,I.A. and Struhl,K. (1985) *Cell*, **43**, 177–188.

Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885–894.

Joachimiak,A.J., Kelley,R.L., Gunsalus,R.P. and Yanofsky,C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 668–672.

McKay,D.B. and Steitz,T.A. (1981) *Nature*, **290**, 744–749.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.

Pabo,C.O. and Lewis,M. (1982) *Nature*, **298**, 443–447.

Ptashne,M. (1986) *Nature*, **322**, 697–701.

Struhl,K. (1982) *Nature*, **300**, 284–287.

Received on April 15, 1987; revised on May 29, 1987