Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 2822–2826, April 1996
Biochemistry

The single Cys_2–His_2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding

(gene expression/DNA–protein interaction)

PAOLO V. PEDONE*, RODOLFO GHIRLANDO*†, G. MARIUS CLORE‡, ANGELA M. GRONENBORN‡, GARY FELSENFELD*¶, AND JAMES G. OMICHINSKI‡

Laboratories of *Molecular Biology and ‡Chemical Physics, Building 5, National Institutes of Health, Bethesda, MD 20892-0530

Contributed by Gary Felsenfeld, December 6, 1995

ABSTRACT **Specific DNA binding to the core consensus site GAGAGAG has been shown with an 82-residue peptide (residues 310–391) taken from the *Drosophila* transcription factor GAGA. Using a series of deletion mutants, it was demonstrated that the minimal domain required for specific binding (residues 310–372) includes a single zinc finger of the Cys_2–His_2 family and a stretch of basic amino acids located on the N-terminal end of the zinc finger. In gel retardation assays, the specific binding seen with either the peptide or the whole protein is zinc dependent and corresponds to a dissociation constant of $\approx 5 \times 10^{-9}$ M for the purified peptide. It has previously been thought that a single zinc finger of the Cys_2–His_2 family is incapable of specific, high-affinity binding to DNA. The combination of an N-terminal basic region with a single Cys_2–His_2 zinc finger in the GAGA protein can thus be viewed as a novel DNA binding domain. This raises the possibility that other proteins carrying only one Cys_2–His_2 finger are also capable of high-affinity specific binding to DNA.**

The *Drosophila* GAGA factor was originally identified on the basis of its specific binding to GA(CT)-rich regions in the Ultrabithorax promoter (1). Isolation of the cDNA for the GAGA protein (2) revealed that it is 519 amino acids long and that it contains an N-terminal "POZ/BTB" domain (3, 4) of 120 amino acids and a single Cys_2–His_2 zinc finger at residues 343–366 followed by several polyglutamine stretches in the C-terminal region. The known GAGA protein DNA binding sites vary considerably in length but they almost always contain the core consensus sequence GAGAGAG and they are generally located near transcription start sites (for review, see ref. 5). Several *Drosophila* genes have been identified as potential targets of GAGA and many possible functional roles have been proposed for this transcription factor (5). Both *in vivo* (6) and *in vitro* (7, 8) experiments suggest that GAGA can act as an "anti-repressor" by generating nucleosome free DNA in the promoter regions. Genetic analysis in *Drosophila* confirms a possible role of the GAGA factor in chromatin modeling. It has been shown that the gene Trithorax-like, which is required for normal expression of the homeotic genes and is a dominant enhancer of position effect variegation, encodes the GAGA factor (9). This result suggests that the function of the protein in chromatin organization might not be restricted solely to the promoter region.

Although the specific DNA binding of the GAGA protein to the consensus sequence GAGAGAG has been well documented (5), the domain responsible for the recognition has not yet been identified. Based on studies with GAGA protein and related "POZ/BTB"-containing proteins, it appears that the POZ/BTB domain is not a DNA binding motif; in some cases it may act to inhibit DNA binding (4). The polyglutamine stretch, which is found in a number of eukaryotic transcription factors, is presumably important for transactivation. Thus, the section of the GAGA protein most likely responsible for DNA recognition is the region encompassing the single classical Cys_2–His_2 zinc finger. Whereas the GAGA protein contains only a single zinc finger, in the typical DNA binding protein that contains Cys_2–His_2 zinc fingers the number of tandemly repeating finger domains varies from 2 to 37 with a minimum of two units required for specific high-affinity DNA binding (for review, see ref. 10). In this manuscript, a series of peptides containing the single Cys_2–His_2 zinc finger of the GAGA protein and adjacent amino acids either N-terminal or C-terminal of the finger have been overexpressed in *Escherichia coli* and purified. Gel retardation experiments demonstrate that a 63-residue segment of the protein containing the single finger domain along with a stretch of basic residues located just N-terminal to it are sufficient for specific DNA binding to the sequence GAGAGAG. It appears that the zinc finger domain of the GAGA protein recognizes DNA in a manner that differs considerably from that seen with the classical Cys_2–His_2 zinc finger DNA binding protein. These results with a single zinc finger domain show that the Cys_2–His_2 motif can be used in a wide variety of structural combinations for DNA recognition and that its use is not limited to the simple tandem repeat pattern with which it is commonly associated.

MATERIALS AND METHODS

Cloning and Purification of the Peptides. DNA fragments encoding the different peptides were generated by PCR from the plasmid parGAGA (2). Oligonucleotides were synthesized on the basis of the published sequence (2). The following oligonucleotides have been used as primers: primer 1, 5′-ACATGCCATGGGCAATACAAGCGGCGTCC-3′; primer 2, 5′-CGGGATCCTTACTCCATGCTGGAGTCTAGGG-3′; primer 3, 5′-ACATGCCATGGAACAACCTGCTACTTGCCC-3′; primer 4, 5′-CGGGATCCTTACACGCCGGGTTTGGCAAAATG-3′; primer 5, 5′-ACATGCCATGGGAACAGAGAAACCACGTTC-3′. For cloning the coding sequence for the peptide GAGA 1, we used primers 1 and 2, for the peptide GAGA 2 primers 1 and 4, for the peptide GAGA 3 primers 3 and 2, and for the peptide GAGA 4 primers 5 and 4. The PCR products were digested with the restriction enzymes *Nco* I and *Bam*HI and cloned into a *Nco* I/*Bam*HI-digested pET-11D (Novagen) expression vector.

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.

†Present address: Centre de Biochimie Structurale, Faculté de Pharmacie, Centre National de la Recherche Scientifique Unité Mixte de Recherche 9955-UM 1-Institut National de la Santé et de la Recherche Médicale U 414, 34060 Montpellier cedex, France.
¶To whom reprint requests should be addressed.

Peptides were expressed in BL21 (DE3) cells. Cells were grown overnight at 37°C, after which protein expression was induced for 4 hr with 0.5 mM isopropyl β-D-thiogalactoside. The cells were harvested and resuspended in 50 mM Tris buffer, pH 8.0/5 mM EDTA/5 mM benzamidine/5 mM dithiothreitol (DTT). The cells were then lysed by passage through a French press and centrifuged at 100,000 × *g* for 1 hr. The supernatant was purified over two ion-exchange columns as described (11). The fractions containing the desired peptide were pooled and subjected to further purification on a C_4 reversed-phase (Vydac) high-performance liquid chromatography (HPLC) column with a 1–100% acetonitrile gradient in 0.05% aqueous trifluoroacetic acid. All the peptides were characterized by electrospray mass spectroscopy. The lyophilized proteins were reconstituted with 1.1 equivalent of zinc, and the final pH was slowly adjusted to 6.0 with NaOH.

Purification of the GAGA Protein. Full-length GAGA protein was expressed using BL21 (DE3) *E. coli* transformed with pAR-GAGA (2). The protein extract was loaded on a fast flow SP-Sepharose (Pharmacia) column (100 ml) equilibrated with 0.1 M KCl/25 mM Hepes, pH 7.6/1 mM EDTA/1 mM DTT/0.01% Triton X-100. GAGA protein was eluted using a 0.1–1.0 M KCl gradient. The protein, as evidenced by electrophoretic mobility-shift assays, eluted at ≈0.6 M KCl. The GAGA protein was further purified by DNA affinity chromatography using sequence-specific DNA. The DNA affinity column was prepared in a manner similar to that described by Kadonaga and Tjian (12) using the following double-stranded DNA oligonucleotide: 5′-GATCGTAGAGAGAGAGAA-GAGAAGAGAGAGAACGTGC-3′. Analysis of the DNA affinity-purified GAGA protein by denaturing SDS/PAGE shows a single band having an apparent molecular mass of 57 KDa.

Gel Mobility-Shift Analysis. Unless otherwise specified the purified GAGA protein or the different peptides were incubated for 15 min on ice with 10 fmol of the labeled 20-bp oligonucleotide h3/h4 GAGA (5′-AAACCCGAGAGAG-TACGAAC-3′) in the presence of 20 mM Hepes, pH 7.9/50 mM KCl/6.25 mM $MgCl_2$/5% glycerol/200 ng of poly(dI-dC) (binding buffer). After incubation, the mixture was loaded onto a 5% polyacrylamide gel (19:1, acrylamide/bisacrylamide) and run in 1× TBE at 4°C. Unlabeled competitor DNA was added at 100-fold molar excess of the amount of labeled DNA in the reaction mixture. As a nonspecific competitor for competition experiment the oligonucleotide 5′-ACGGCTG-CAGGTCATGACCAGCCCCACGCC-3′ (oligonucleotide NS) was used. The sequence of the mutated h3/h4 GAGA oligonucleotide (mut h3/h4) is 5′-AAACCCGATTTAGTA-CGAAC-3′.

RESULTS

To determine whether the single putative zinc finger domain present in the GAGA protein was able to mimic the DNA binding activity of the entire protein, a fragment encoding 82 amino acids of the protein (residues 310–391, GAGA 1; Fig. 1) was cloned and overexpressed in *E. coli*. A gel mobility-shift assay was used to study the interaction between the peptide and a DNA oligonucleotide containing the GAGAGAG site (oligonucleotide h3/h4 GAGA). The oligonucleotide was derived from the sequence of the intergenic region located between the promoters of the *Drosophila* histone genes *his3* and *his4* (13). This site has been shown to bind the GAGA protein both *in vivo* and *in vitro* (5, 13). As shown in Fig. 2*A* (lane 2), the peptide interacts with its target DNA to produce a single complex. The binding specificity of the purified peptide was demonstrated by competition experiments with unlabeled oligonucleotides; the retarded complex is competed by addition of a 100-fold excess of the unlabeled h3/h4 GAGA oligonucleotide (Fig. 2*A*, lane 3) but not by a 100-fold excess of an oligonucleotide with an unrelated sequence (lane 4). Incubation of the protein with a high concentration of either EDTA (10 mM and 50 mM; Fig. 2*B*, lanes 2 and 3) or the zinc chelating agent 1,10-phenanthroline (5 mM; lane 4) drastically reduces the DNA binding activity of GAGA 1, indicating that the binding is dependent on the presence of the zinc ion. The affinity of the GAGA 1 peptide for the DNA was measured

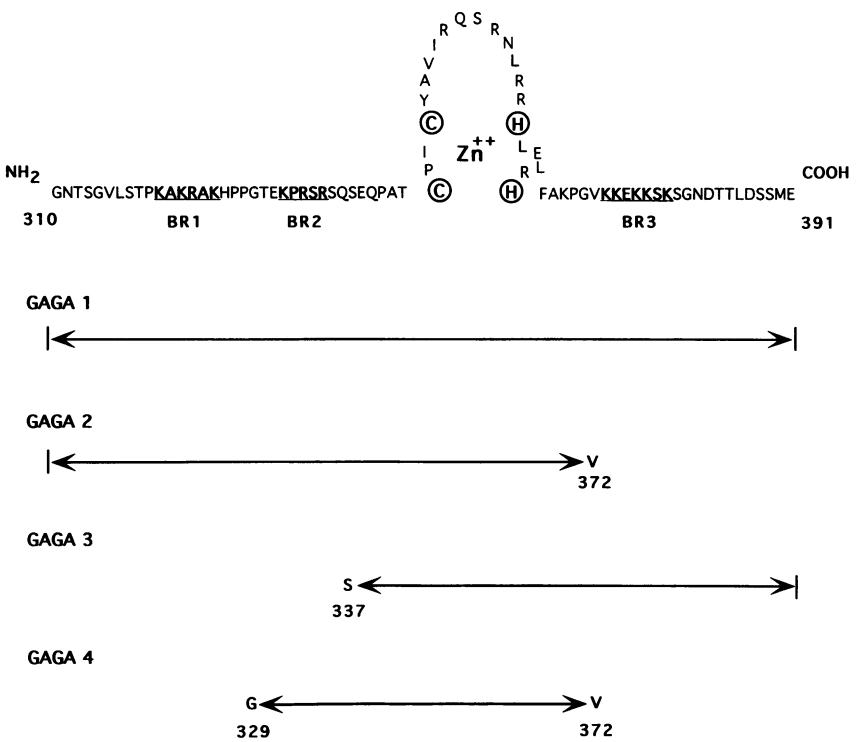

FIG. 1. Schematic representation of the peptides used in these studies. Underlined sequences indicate the basic regions flanking the zinc finger motif (BR1, BR2, and BR3).

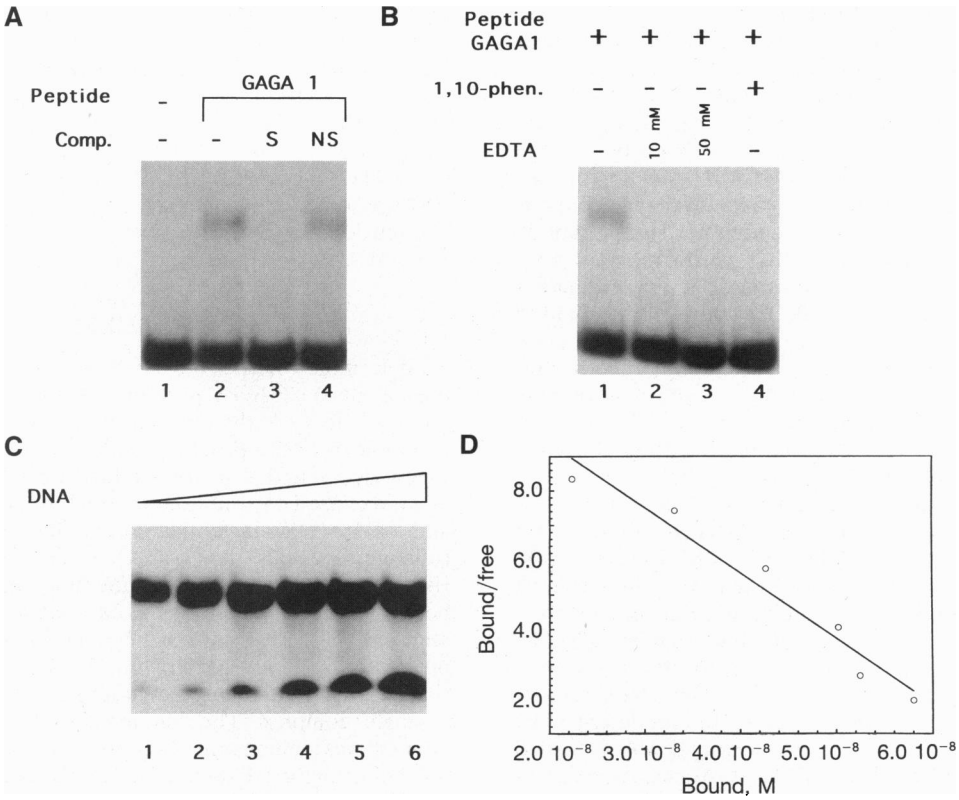

FIG. 2. Gel mobility-shift analysis of binding of the GAGA 1 peptide. (*A*) Purified GAGA 1 peptide (440 fmol) (lanes 2–4) was incubated with 10 fmol of the labeled 20-bp oligonucleotide h3/h4 GAGA in the presence of binding buffer. Where shown, unlabeled h3/h4 GAGA oligonucleotide (lane 3) or an oligonucleotide with an unrelated sequence (oligonucleotide NS) (lane 4) was added as competitor in a 100-fold excess. Comp., competitor; S, specific; NS, nonspecific. (*B*) Zinc ion requirement of the GAGA 1 peptide for DNA binding. The GAGA 1 peptide was incubated 40 min on ice in binding buffer with or without 10 mM EDTA, 50 mM EDTA, or 5 mM 1,10-phenanthroline and then subjected to gel-shift analysis. phen., Phenanthroline. (*C*) Gel mobility-shift titrations of the GAGA 1 peptide. Lanes 1–6, titration of peptide GAGA 1 with the h3/h4 GAGA oligonucleotide. In lanes 1–6, each sample contained, in a vol of 20 μl, 4.4 pmol of peptide and 0.5, 0.75, 1, 1.25, 1.5, and 1.75 pmol of duplex, respectively. The reactions were performed as described but in the absence of poly(dI-dC). (*D*) Scatchard analysis of gel-shift binding data shown in *C*. The ratio of bound to free DNA is plotted vs. the molar concentration of bound DNA in the reaction mixture. All numerical values were obtained by computer quantitation of the image using a Molecular Dynamics PhosphorImager.

using a gel mobility-shift assay (Fig. 2*C*). Scatchard analysis (Fig. 2*D*) leads to an apparent dissociation constant of $5.3 \pm 0.7 \times 10^{-9}$ M.

Consistent with the GAGA 1 results, the binding of the full-length GAGA protein to the h3/h4 GAGA oligonucleotide is specific and dependent upon the presence of zinc ions (Fig. 3*A*). To prove that both the peptides and the protein recognize the GAG core sequence on the DNA, a mutated h3/h4 GAGA (the core GAG motif was substituted with the sequence TTT; oligo mut h3/h4 GAGA) was used as a competitor in the gel mobility-shift assay. The mutated oligonucleotide was not able to compete for the binding of either the full-length protein or the GAGA 1 peptide (Fig. 3*B*) even when used in a 100-fold molar excess. Furthermore, neither the full-length protein nor the GAGA 1 peptide bind to the mut h3/h4 GAGA oligonucleotide when used as a probe (data not shown).

The zinc finger motif present in the GAGA 1 peptide is flanked on both termini by highly basic stretches of amino acids (BR-1, BR-2, BR-3; see Fig. 1). It has previously been demonstrated that a single zinc finger of cGATA-1 (a protein in which zinc ion is chelated by four cysteines in the finger instead of two cysteines and two histidines) is capable of high-affinity specific binding but that this binding requires a C-terminal basic region to stabilize the interaction (11, 14). Using a series of deletion mutants (GAGA 2–4; see Fig. 1), the role of the adjacent basic regions in GAGA 1 binding was investigated. Removal of the C-terminal basic region (BR-3) by deletion of 19 amino acids yielded a peptide (GAGA 2) that binds to the h3/h4 GAGA oligonucleotide with the same specificity (Fig. 4, lanes 1–3) and affinity (data not shown) as GAGA 1. In contrast, removal of the N-terminal 27 residues of GAGA 1 resulted in a peptide (GAGA 3) that failed to display specific DNA binding (lane 4). These results suggest that the zinc finger requires the amino acids at its N terminus to bind the DNA. To determine whether BR-2 alone was sufficient to stabilize the binding of the finger to the DNA, an additional peptide (GAGA 4) was prepared in which BR-1 was deleted. GAGA 4 failed to demonstrate specific binding to the GAGA DNA motif (lane 5). All the peptides utilized in these experiments were found by one-dimensional ^{1}H NMR (data not shown) to form an ordered structure in the presence of zinc. The lack of binding by GAGA 3 and 4 is therefore unlikely to result from incorrect folding of the peptide.

Based on these experiments, it can be concluded that the single Cys_2–His_2 zinc finger domain from the GAGA protein, along with a stretch of amino acids (>14 and <33) at the N terminus of the finger, is sufficient for specific DNA binding.

DISCUSSION

The classical Cys_2–His_2 zinc finger domain consists of 30 amino acids folded around a central zinc ion, which is chelated by 2 cysteines and 2 histidines in a tetrahedral geometry (10). This zinc finger is common in eukaryotic DNA binding proteins and several NMR and x-ray studies have shown that it consists structurally of a β-sheet (including the two chelating cysteines) and an α-helix (which usually includes the two chelating

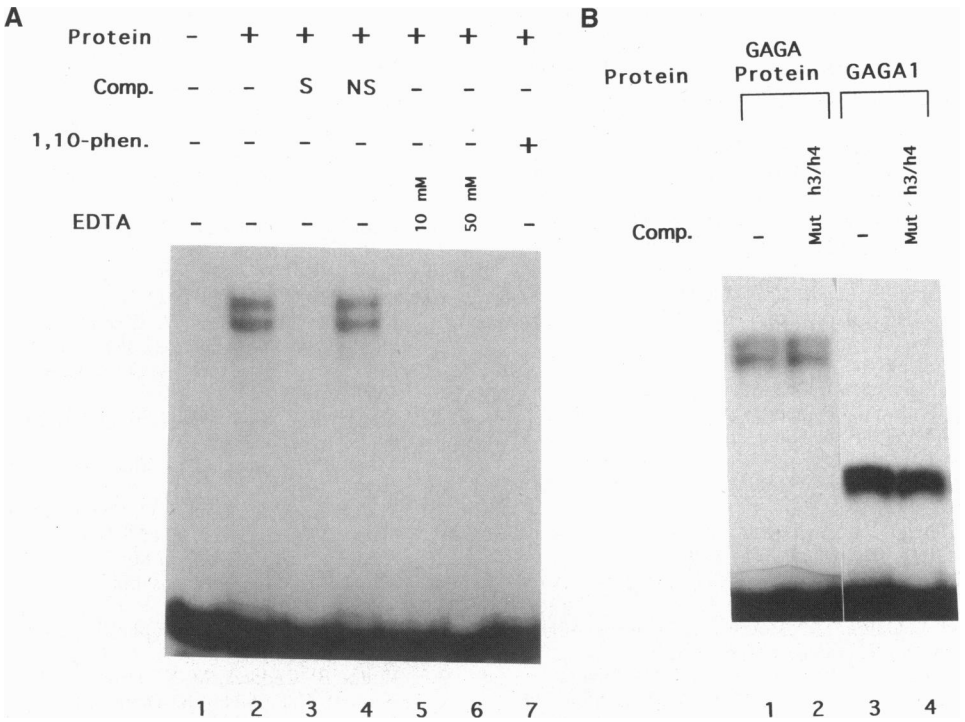

FIG. 3. Analysis of binding of the purified GAGA protein. (*A*) The h3/h4 GAGA oligonucleotide was used as probe in gel mobility-shift assays with the purified GAGA protein. The protein generates two complexes of different mobility, as observed by Soeller *et al.* (2). Where shown, unlabeled h3/h4 GAGA oligonucleotide (lane 3) or the oligonucleotide NS with an unrelated sequence (lane 4) was added as competitor in a 100-fold excess. To demonstrate the zinc ion requirement for DNA binding activity, the protein was treated with 10 or 50 mM EDTA (lanes 5 and 6) or 5 mM 1,10-phenanthroline (lane 7) for 40 min on ice. The treated protein was then subjected to gel-shift analysis. Comp., competitor; phen., phenanthroline; S, specific; NS, nonspecific. (*B*) The unlabeled mutated h3/h4 GAGA oligonucleotide (mut h3/h4) in a 100-fold molar excess was used to inhibit binding of either the GAGA protein (lane 2) or the peptide GAGA 1 (lane 4) to the h3/h4 GAGA probe.

histidines) (15–21). In most DNA binding proteins that contain this classical zinc finger motif, it is found that such a motif occurs in a tandem array of canonically spaced units, with the number of zinc finger repeating units varying from 2 to 37 (10, 22). The recognition of the target DNA site is often modular in nature, with each unit typically recognizing consecutive sets of 3 bp within the DNA (19–21, 23, 24). Prior to these experiments with single finger peptides from the GAGA protein, the common perception had been that there must be a minimum of two zinc finger domains for high-affinity DNA binding (10, 22).

The use of a single zinc finger in combination with flanking amino acids has been shown to occur in the GATA family of DNA binding proteins (25, 26), where the finger has four cysteines and a different three-dimensional structure. Within this family, the vertebrate derived members contain two zinc finger domains (for review, see ref. 27), while the fungal derived members have only one (28–31). Studies with the GATA proteins containing two zinc fingers indicated that the C-terminal finger was capable of binding independent of the N-terminal finger but the reverse situation was not true (25). Further biochemical and structural analysis showed that the C-terminal finger alone was not sufficient as an isolated unit but that it required several basic residues C-terminal to the finger to stabilize the recognition (14). The results with the GAGA protein are somewhat analogous to this situation but differ in that the flanking sequence is located N-terminal to the finger domain.

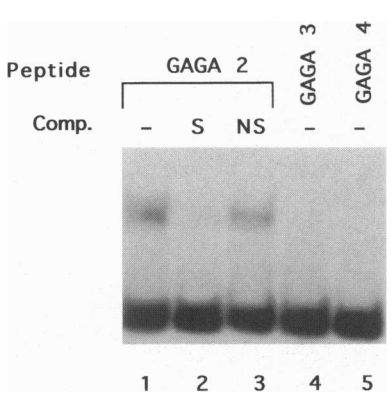

FIG. 4. Gel mobility-shift analysis of binding of the GAGA 2, 3, and 4 peptides. Where shown, unlabeled h3/h4 GAGA (lane 2) or the oligonucleotide NS with an unrelated sequence (lane 3) was added as competitor at a 100-fold excess.

It is interesting to note that in several other Cys_2–His_2 zinc finger-containing proteins, residues located N-terminal to the finger have been shown to be important for structure formation and DNA binding (18, 32, 33). In all these cases, however, two finger units with canonical spacing between the fingers are essential for high-affinity DNA binding. In the proteins tramtrack and SWI5, the N-terminal residues are involved in formation of an extra strand of β-sheet not seen in the typical Cys_2–His_2 zinc finger domain (18, 21). The residues involved in this extra β-strand formation, which are important for DNA binding, are located within the first 10 residues preceding the zinc finger unit. In contrast, addition of 14 residues to the N terminus of the GAGA finger (GAGA 4) was insufficient for high-affinity binding. However, when 33 residues were added (GAGA 2), the peptide mimicked the binding of the full-length protein. Based on the results with the single finger domain peptides of GAGA, one might speculate that in the numerous proteins containing either a single zinc finger or multiple fingers with noncanonical spacing (especially long spacing patterns) each finger has the potential to recognize a DNA target site as an independent unit. This single unit binding

would most likely differ considerably from the previously described tandem repeat pattern of binding. Thus, the single zinc finger of GAGA protein could well exemplify a new mechanism by which proteins recognize DNA.

We thank Dr. Adam Bell and Dr. David Clark for critical reading of the manuscript and Dr. Cecelia Trainor for helpful discussion. We thank Dr. Kazuyasu Sakaguchi and Dr. Hiroshi Sakamoto for mass spectral analysis. The cDNA clone of GAGA protein was kindly provided by Dr. Thomas Kornberg.

1. Biggin, M. D. & Tjian, R. (1988) *Cell* **53,** 699–711.
2. Soeller, W. C., Oh, C. E. & Kornberg, T. B. (1993) *Mol. Cell. Biol.* **13,** 7961–7970.
3. Zollman, S., Godt, D., Prive, G. G., Couderc, J. L. & Laski, F. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10717–10721.
4. Bardwell, V. J. & Treisman, R. (1994) *Genes Dev.* **8,** 1664–1677.
5. Granok, H., Leibovitch, B. A., Shaffer, C. D. & Elgin, S. C. (1995) *Curr. Biol.* **5,** 238–241.
6. Lu, Q., Wallrath, L. L., Howard, G. & Elgin, S. C. R. (1993) *Mol. Cell. Biol.* **13,** 2802–2814.
7. Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R. & Kadonaga, J. T. (1991) *Science* **251,** 643–649.
8. Tsukiyama, T., Becker, P. B. & Wu, C. (1994) *Nature (London)* **367,** 525–532.
9. Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H. & Karch, F. (1994) *Nature (London)* **371,** 806–808.
10. Klug, A. & Schwabe, J. W. (1995) *FASEB J.* **9,** 597–604.
11. Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J. & Gronenborn, A. M. (1993) *Science* **261,** 438–446.
12. Kadonaga, J. T. & Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5889–5893.
13. Gilmour, D. S., Thomas, G. H. & Elgin, S. C. (1989) *Science* **245,** 1487–1490.
14. Omichinski, J. G., Trainor, C., Evans, T., Gronenborn, A. M., Clore, G. M. & Felsenfeld, G. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1676–1680.
15. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245,** 635–637.
16. Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990) *Biochemistry* **29,** 9324–9334.
17. Klevit, R. E., Herriott, J. R. & Horvath, S. J. (1990) *Proteins* **7,** 215–226.
18. Neuhaus, D., Nakaseko, Y., Schwabe, J. W. & Klug, A. (1992) *J. Mol. Biol.* **228,** 637–651.
19. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252,** 809–817.
20. Pavletich, N. P. & Pabo, C. O. (1993) *Science* **261,** 1701–1707.
21. Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366,** 483–487.
22. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4,** 1609–1614.
23. Desjarlais, J. R. & Berg, J. M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 2256–2260.
24. Nardelli, J., Gibson, T. & Charnay, P. (1992) *Nucleic Acids Res.* **20,** 4137–4144.
25. Martin, D. I. & Orkin, S. H. (1990) *Genes Dev.* **4,** 1886–1898.
26. Yang, H. Y. & Evans, T. (1992) *Mol. Cell. Biol.* **12,** 4562–4570.
27. Orkin, S. H. (1992) *Blood* **80,** 575–581.
28. Minehart, P. L. & Magasanik, B. (1991) *Mol. Cell. Biol.* **11,** 6216–6228.
29. Cunningham, T. S. & Cooper, T. G. (1991) *Mol. Cell. Biol.* **11,** 6205–6215.
30. Kudla, B., Caddick, M. X., Langdon, T., Martinez-Rossi, N. M., Bennett, C. F., Sibley, S., Davies, R. W. & Arst, H. N., Jr. (1990) *EMBO J.* **9,** 1355–1364.
31. Fu, Y. H. & Marzluf, G. A. (1990) *Mol. Cell. Biol.* **10,** 1056–1065.
32. Thukral, S. K., Eisen, A. & Young, E. T. (1991) *Mol. Cell. Biol.* **11,** 1566–1577.
33. Fairall, L., Harrison, S. D., Travers, A. A. & Rhodes, D. (1992) *J. Mol. Biol.* **226,** 349–366.