8-1997

Multiple isoforms of GAGA factor, a critical component of chromatin structure

C Benyajati

L Mueller

N Xu

M Pappano

J Gao

See next page for additional authors

Nucleic Acids Research, 1997, Vol. 25, No. 16 **3345–3353**

Multiple isoforms of GAGA factor, a critical component of chromatin structure

Cheeptip Benyajati*, Loretta Mueller, Nong Xu, Michelle Pappano, Jing Gao, Mana Mosammaparast, Dawn Conklin, Howard Granok[1,2], Carolyn Craig[1] and Sarah Elgin[1,2]

Department of Biology, University of Rochester, Rochester, NY 14627, USA, [1]Department of Biology and [2]Program in Molecular Genetics, Division of Biology and Biomedical Sciences, Washington University, St Louis, MO 63130, USA

Received March 27, 1997; Revised and Accepted June 20, 1997

DDBJ/EMBL/GenBank accession nos U16728, U18386, U68563

ABSTRACT

The GAGA transcription factor of *Drosophila melanogaster* is ubiquitous and plays multiple roles. Characterization of cDNA clones and detection by domain-specific antibodies has revealed that the 70–90 kDa major GAGA species are encoded by two open reading frames producing GAGA factor proteins of 519 amino acids (GAGA-519) and 581 amino acids (GAGA-581), which share a common N-terminal region that is linked to two different glutamine-rich C-termini. Purified recombinant GAGA-519 and GAGA-581 proteins can form homomeric complexes that bind specifically to a single GAGA sequence *in vitro*. The two GAGA isoforms also function similarly in transient transactivation assays in tissue culture cells and in chromatin remodeling experiments *in vitro*. Only GAGA-519 protein accumulates during the first 6 h of embryogenesis. Thereafter, both GAGA proteins are present in nearly equal amounts throughout development; in larval salivary gland nuclei they colocalize completely to specific regions along the euchromatic arms of the polytene chromosomes. Coimmunoprecipitation of GAGA-519 and GAGA-581 from crude nuclear extracts and from mixtures of purified recombinant proteins, indicates direct interactions. We suggest that homomeric complexes of GAGA-519 may function during early embryogenesis; both homomeric and heteromeric complexes of GAGA-519 and GAGA-581 may function later.

INTRODUCTION

GAGA factor binds specific GA-rich sequences in both euchromatic and heterochromatic regions of the *Drosophila melanogaster* genome (reviewed in 1,2). GAGA factor is localized to numerous bands along the euchromatic arms of salivary gland polytene chromosomes (3; Fig. 8). Several of these bands include housekeeping or developmentally regulated genes with GAGA factor binding sites in their promoter regions (4). GAGA factor is also associated with centromeric heterochromatin of preblastoderm embryos and in embryo-derived tissue culture cells (5). Consistent with these patterns of distribution, mutations in the gene encoding GAGA factor, Trithorax-like (*Trl*), affect both the expression of several developmental loci (6,7) and chromosome segregation and nuclear division (6). *Trl* mutations also behave as dominant enhancers of position effect variegation (PEV; the clonal inactivation of translocated euchromatic sequences by adjacent heterochromatin) (7). Thus, GAGA factor has both local and global roles in chromosome structure and function.

GAGA factor can activate promoters containing GAGA binding sites in transient cotransfection assays (4,8) and in *in vitro* transcription reactions (8,9). This transactivation may result from an anti-repression activity of GAGA factor in the presence of general DNA-binding transcriptional repressors, such as histone H1 (10,11), rather than a 'true activation' involving interaction with RNA polymerase. *In vivo*, GAGA factor binding at the *Drosophila hsp26* promoter is necessary to establish and maintain nucleosome-free, DNase I hypersensitive sites at the heat shock elements (HSEs) (12). Heat shock transcription factor binding to chromatin of the *hsp70* promoter *in vivo* depends on a promoter architecture established by GAGA factor, TFIID and RNA polymerase II (13). GAGA factor may also play a role during RNA polymerase elongation through chromatin *in vivo* (14). A recombinant GAGA factor can bind and cause local rearrangement of nucleosomes over the GAGA binding sites of the *hsp70* promoter, creating DNase I hypersensitive sites at the adjacent TATA box and HSEs during chromatin assembly *in vitro* (3). Thus, GAGA factor may function by organizing and/or remodeling the nucleosome array.

Purification of GAGA factor by biochemical fractionation and sequence-specific DNA affinity chromatography yields several polypeptides detected by SDS–PAGE (4,9,11,15,16). The patterns of bands are slightly different between preparations from *Drosophila* embryos and tissue culture cells. This multiplicity of GAGA factors could reflect multiple primary protein sequences, or could indicate post-translational modification. Purified GAGA proteins can be modified by galactosyl transferase, indicating that they bear O-linked *N*-acetylglucosamine monosaccharide residues (17), which may alter electrophoretic mobility of proteins on SDS–PAGE. It is not known whether heterogeneity in GAGA factor would have any consequences for the various nuclear processes in which GAGA factor may be involved.

*To whom correspondence should be addressed. Tel: +1 716 275 8040; Fax: +1 716 275 2070; Email: tip@cb.biology.rochester.edu

Here we have dissected the multiplicity of GAGA factor by characterizing several distinct full-length GAGA factor cDNA clones; immunodetection using antibodies against specific domains of the GAGA proteins has identified the specific products. Two major primary isoforms for GAGA factor have been identified, GAGA-519 and GAGA-581, which share an N-terminal protein–protein interaction domain and a DNA binding domain, but differ in their C-termini. Both GAGA factor isoforms exhibit a full spectrum of biological activity.

MATERIALS AND METHODS

Sequence accession numbers

The nucleotide accession numbers of the cDNA sequences are U16728 (GAGA-581), U18386 (GAGA-519a) and U68563 (GAGA-519b).

GAGA factor cDNA clones

The first GAGA cDNA clone recovered in the study, GAGA-519c, was obtained by screening an ovarian cDNA expression library in λgt11 (gift of Laura Kalfayan) using a ^{32}P-end-labelled multimer of the 36 bp double-stranded oligonucleotide containing the Adf-2a binding site, from –215 to –184 of the distal ADH transcription start site (18). Additional cDNA clones were identified by hybridization and analyzed by restriction mapping; representative clones of each class were sequenced.

Anti-GAGA factor antibodies

To generate anti-GAGA$^{\text{BTB-ZF-C-ter 519}}$, rats were immunized with purified maltose binding protein-GAGA-519 [amino acids (aa) 38–519] a fusion protein expressed from pMAL-c2-GAGA-519 (New England Biolabs). Isoform-specific rabbit antisera, anti-GAGA$^{\text{C-ter 519}}$ (aa 419–519) and anti-GAGA$^{\text{C-ter 581}}$ (aa 381–581) were obtained by immunization with isoform-specific polypeptides fused in frame to glutathione-*S*-transferase (GST), expressed from pGEX-KT (20). Antibody affinity purification was performed by first depleting anti-GST antibodies by adsorption of the crude antiserum to nitrocellulose-bound purified GST. The residual antibodies were then adsorbed to filters bound with the respective purified fusion proteins and the specifically bound antibody eluted with 0.2 M glycine pH 2.8.

Expression and purification of GAGA factors

Recombinant proteins were expressed from pAR-GAGA-519 (4) and pAR-GAGA-581. Both fusion proteins lack the five N-terminal aa and substitute 13 aa from T7 gene 10. GAGA factors were purified from *Escherichia coli* BL21 pLysS lysates by chromatography over heparin agarose (21).

Competition gel mobility shift assay

Purified recombinant GAGA factors (250 ng) were incubated on ice for 20 min with ~1 ng of ^{32}P-labelled 200 bp DNA fragment containing the Adf-2b site (from –136 to +64 of the distal ADH transcription start site), 500 ng of poly (dI-dC), with or without unlabelled competitor DNA in 20 µl reaction mixtures (25 mM Tris–HCl pH 7.6, 50 mM KCl, 0.5 mM EDTA, 2.5 mM DTT, 5% glycerol, 1 mg/ml BSA and 0.05% NP-40). Competitor DNA fragments were gel-isolated from plasmids containing the multimerized 6mer of the 32 bp wild-type Adf-2a site (from –215 to –184), or the multimerized 6mer of the 29 bp mutated Adf-2a site with a 3 bp deletion of nucleotides between –198 and –200 (15). Samples were electrophoresed through 1% agarose gels in 0.5× Tris–borate–EDTA buffer. The agarose gels were dried and exposed to X-ray film.

Drosophila expression vectors

Expression plasmids were constructed by inserting the GAGA factor ORFs downstream of the *Drosophila* metallothionein (*Mt*) gene promoter at the *Sma*I site in the vector pRmHa3 (22). cDNA inserts containing the intact ORFs for GAGA-519 and GAGA-581 (2.25 and 3.3 kb, respectively) were removed from the pNB 40 expression vector (19) by *Hin*dIII–*Not*I double digests. The ends were filled in using DNA polymerase (Klenow fragment) and the fragments were cloned in both the sense and the antisense orientations in pRmHa3.

Transient transfection of *Drosophila* cell lines

ADH-non-expressing S2 and ADH-expressing 1006-2 cells were transfected by calcium phosphate coprecipitation (23,24). After 24 h, induction of the GAGA factor expression vector was performed by adding $CuSO_4$ to a final concentration of 0.7 mM. At 48 h, cells were harvested; 25% of the cells were lysed in a nuclear buffer (15). Proteins in the pelleted nuclei were analyzed by Western blot analysis. The rest of the cells were assayed for ADH and β-galactosidase activity (23,24).

Chromatin assembly and MNase digestions

Regularly spaced nucleosomal arrays were assembled on the plasmid pdhspXX3.2 containing the –1.5 to +1.8 kb region of the *Drosophila hsp70* gene using the *Drosophila* embryo extract (3,26). A typical reaction contained 300 ng pdhspXX3.2 and 700 ng RF ΦX174 DNA with or without addition of purified recombinant GAGA factor and was incubated for 6 h at 26°C. MNase digestions, agarose gel electrophoresis of nucleosomal DNA and Southern blotting were as described (3).

Developmental Western analysis

Staged wild-type (Canton-S) *Drosophila* embryos, larvae, pupae, adult flies and hand-dissected mid-third instar larval fat body and salivary gland were homogenized directly in SDS–protein gel sample buffer, the samples boiled and frozen in liquid nitrogen. Samples were separated by SDS–PAGE, transferred to nitrocellulose and the filters probed with either rat or rabbit affinity-purified primary antibodies directed against a GAGA factor. Secondary antibody detection was performed with Amersham's ECL (Enhanced Chemiluminescence) kit.

Double immunofluorescence staining

Polytene chromosomes from third instar larval salivary glands were fixed with 2% formaldehyde and squashed in 45% acetic acid; the chromosomes were then simultaneously incubated with affinity-purified rat anti-GAGA$^{\text{BTB-ZF-C-ter 519}}$ and affinity-purified rabbit anti-GAGA$^{\text{C-ter 581}}$ antibodies. The rat antibody was detected using FITC-conjugated goat anti-rat IgG; the rabbit antibody was detected using Texas Red-conjugated goat anti-rabbit IgG (27).

Immunoprecipitation assays

Purified recombinant GAGA proteins, or their equal mixtures, or crude KcO nuclear extracts (15) were incubated at 4°C for 1 h with various rabbit sera in 100 µl IP buffer (50 mM Tris–HCl pH 8.0, 0.15 M NaCl and 1% NP-40). Aliquots (100 µl) of a 50% slurry of protein A–agarose (IPA 300, Repligen) was added and after incubation at 4°C for 1 h with rotation, the beads were recovered by centrifugation and washed four times with RIPA (50 mM Tris–HCl pH 8.0, 0.15 M NaCl, 1% NP-40, 0.1% SDS and 0.1% DOC). The proteins were eluted by boiling in SDS–protein gel sample buffer and analyzed by Western blot analysis.

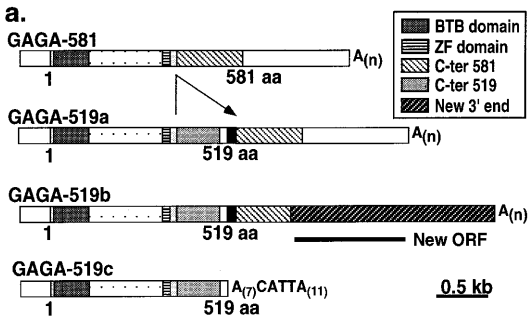
RESULTS

Complexity of GAGA factor revealed by cDNA clones

Several polypeptides are present in preparations of DNA affinity-purified GAGA factor from *Drosophila* embryos and tissue culture cells (4,9,11,15,16). Western blot analyses using antisera against the polypeptide encoded by the GAGA-519c cDNA clone (Fig. 1a) also detect several polypeptides, ranging from ~70 to ~160 kDa, in varying abundance (3–6,14; Fig. 2). By screening a cDNA library we have recovered the same clone and further identified four different GAGA factor cDNA classes (Fig. 1a); three of these encode a 519 aa ORF beginning with a methionine residue and one encodes a 581 aa ORF (Fig. 1b). Clones GAGA-519a, GAGA-519b and GAGA-581 (but not GAGA-519c, see legend to Fig. 1a) possess long poly-A tails and likely represent GAGA factor mRNAs. Comparison of various GAGA factor cDNA structures indicates that they most likely arose from alternative splicing of transcripts from the *Trl* locus. The sizes of these cDNA clones are in general consistent with the sizes of the multiple GAGA factor poly-A^+ RNAs observed (from ~2.4 to 4.4 kb); these RNAs are displayed in complex patterns during development (4,6).

The 519 and 581 aa ORFs share the N-terminal amino acids 1–377, but differ in their C-termini (Fig. 1b). The predicted shared region includes the BTB or POZ domain, a 110 aa block sharing significant homology (~50% identity) with the N-terminal regions of several *Drosophila* and vertebrate zinc finger proteins, as well as the poxvirus proteins (2). *In vitro* studies have shown that this domain can mediate both homomeric and heteromeric protein interactions (28,29,30). Both ORFs also share a single Cys_2–His_2 zinc finger flanked by basic residues; together, these elements constitute a high affinity sequence-specific DNA binding domain (31). The two different C-termini are rich in glutamine, 23% (33 of 142 aa) in GAGA-519 and 42% (85 of 204 aa) in GAGA-581. The most striking difference between the two ORFs is the length and number of the polyglutamine stretches (two stretches of seven and six residues in GAGA-519, three stretches of 16, 13 and 19 residues in GAGA-581).

Unexpectedly, the additional 1.8 kb cDNA sequence in GAGA-519b contains a 1 kb long ORF following the termination codon of the GAGA-519 ORF (Fig. 1a) that encodes several amino acid repeats and stretches of polyglutamine. The significance of this new ORF is not yet known; it might code for an additional C-terminal domain of GAGA factor.



b. GAGA-519 and GAGA-581

```
  1 MSLPMNSLYS LTWGDYGTSL VSAIQLLRCH GDLVDCTLAA GGRSFPAHKI  50
 51 VLCAASPFLL DLLKNTPCKH PVVMLAGVNA NDLEALLEFV YRGEVSVDHA 100
101 QLPSLLQAAQ CLNIQGLAPQ TVTKDDYTTH SIQLQHMIPQ HHDQDQLIAT 150
151 IATAPQQTVH AQVVEDIHHQ GQILQATTQT NAAGQQQTIV TTDAAKHDQA 200
201 VIQAFLPARK RKPRVKKMSP TAPKISKVEG MDTIMGTPTS SHGSGSVQQV 250
251 LGENGAEGQL LSSTPIIKSE GQKVETIVTM DPNNMIPVTS ANAATGEITP 300
301 AQGATGSSGG NTSGVLSTPK AKRAKHPPGT EKPRSRSQSE QPATCPICYA 350
351 VIRQSRNLRR HLELRHFAKP GVKKEKKTTS GKKSSSGSSG SGSGALSSSG 400
                              SKS GNDTTLDSSM EMNTTAEGDN
401 SVPQVQTVQS LHTLQGVQVK KDPDAQQQQQ QQQQQQQQQQ QAMTVSGATG 450
    TVGSDGAGGA GSAGGQSSGT TPTRVISNAP QAAGAPGILA QGVLPQQQQQ
451 GQVQQQVQQV QQQVQQQQQQ QQQQQQQLQH HQIIDSSGNI TTATTSAQAA 500
    QQLQQQHQQH LTATLAGGGQ AYIKHEGGGG GGTGQQQQQQ AAQQQGMQNV
501 QLAQQQAAGQ QQQLVAQSDG SESGAPLSIA QVQTLQGHQI IGNLNQVNMT 550
    IHIVGDQVFI PQQQQPQPQ 519
551 DFQQQQPQQQ QQQQQQQQQQ QQQQQQTQQT L 581
```

Figure 1. Comparison of the GAGA factor cDNA clones. (**a**) Four different cDNA clones have been completely sequenced and their relationship is indicated in blocks of different patterns. BTB is a conserved 110 aa block (see text) ZF is a single Cys_2-His_2 zinc finger flanked by basic residues. Clone GAGA-581 has a 581 aa ORF. Clones GAGA-519a, GAGA-519b and GAGA-519c have a 519 aa ORF. Clone GAGA-519c does not have an authentic poly-A tail; the A-rich stretch at the 3′ end is present in clones GAGA-519a and GAGA-519b (black box), indicating that the GAGA-519c clone was probably the result of oligo-dT priming during synthesis of the cDNA library. Clone GAGA-581 is likely to have been generated by alternative splicing of the transcript that gave rise to clone GAGA-519a (removing the $GAGA\text{-}519^{C\text{-}ter}$ domain coding sequence, as shown by an arrow) 2.3 kb of clone GAGA-519b is identical to GAGA-519a with an additional 1.8 kb DNA containing a 322 aa ORF (New ORF). (**b**) GAGA-519 and GAGA-581 ORFs share the BTB and the zinc finger domains (in bold), but they differ in their glutamine-rich C-termini (italicized for GAGA-519) The high affinity DNA binding domain (31) is underlined.

Complexity of GAGA factor revealed by domain-specific immunodetection

To verify the existence of multiple GAGA factor C-terminal domains and to dissect the relationship among GAGA factors, we generated antisera against specific epitopes in the C-terminal domains of GAGA-519 (aa 419–519) and GAGA-581 (aa 381–581) using recombinant fusion proteins as immunogens. Figure 2 shows Western blot analysis of a nuclear extract from *Drosophila* tissue culture cells fractionated on the same SDS–polyacrylamide

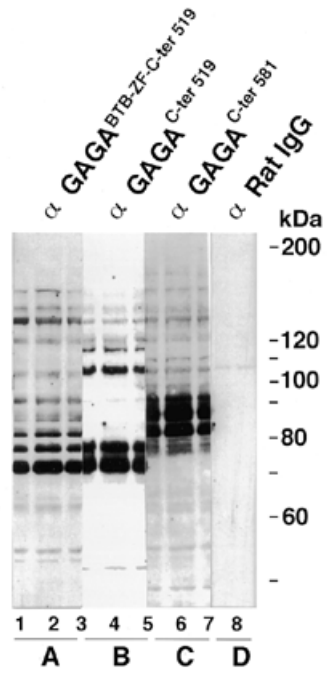


Figure 2. Western blot analysis using domain-specific antibodies. Each lane (1–8) contained a nuclear extract (40 µg) from KcO cells that was fractionated on a 15 cm 7.5% SDS–PAGE gel, transferred to nitrocellulose, the filter cut (through lanes 3, 5 and 7) into four sections (A–D) and each section reacted with various affinity-purified rat (A) and rabbit (B and C) primary antibodies as indicated. Use of a control goat anti-rat secondary antibody alone is shown (D); goat anti-rabbit secondary antibody yielded identical results to D (not shown)

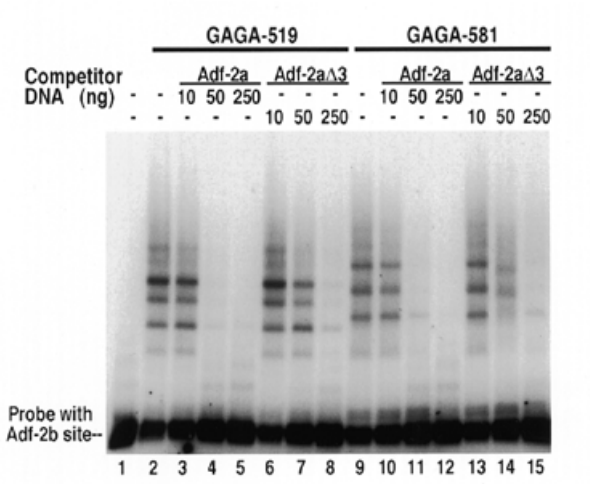


Figure 3. GAGA-519 and GAGA-581 factors bind a single GAGA sequence forming multiple-related nucleoprotein complexes. Purified recombinant GAGA factor (250 ng) was incubated with a ^{32}P-labelled DNA fragment (from –136 to +64) from the *Adh* distal promoter that contains the Adf-2b site with the 10 bp Adf-2 motif, 5′-TCTCAGTGCA-3′. Unlabelled competitor DNAs were either the multimerized wildtype Adf-2a site, from –215 to –184, with the identical 10 bp Adf-2 motif but different flanking sequences, or the multimerized mutated Adf-2a site with 3 bp deleted from the 10 bp Adf-2 motif. Nucleoprotein complexes were fractionated on a 1% agarose gel.

gel and probed separately with three different affinity-purified polyclonal primary antibodies. At least 10 cross-reacting polypeptides were detected with a primary antibody against a near full-length GAGA-519 protein (αGAGA$^{BTB\text{-}ZF\text{-}C\text{-}ter\ 519}$) (Fig. 2A). They ranged from ~70 to 160 kDa in size and were present in varying abundance. Subsets of these polypeptides could be identified using antibodies specific to the individual C-terminal domains. Anti-GAGA$^{C\text{-}ter\ 519}$ antibody (Fig. 2B) detected the major polypeptides of 70 and 75 kDa, whereas the major 80, 85 and 90 kDa polypeptides were seen with anti-GAGA$^{C\text{-}ter\ 581}$ antibody (Fig. 2C). We conclude that these polypeptides are the translation products of GAGA-519 and GAGA-581 mRNAs, respectively (Figs 4 and 5). Clearly each polypeptide has been modified post-translationally. There is a near-equal mixture of the major 70 and 75 kDa GAGA-519 and the 80 and 90 kDa GAGA-581 forms (Figs 2A, 7a and b: detection by the common antibody, αGAGA$^{BTB\text{-}ZF\text{-}C\text{-}ter\ 519}$). A few of the less abundant 105–160 kDa GAGA cross-reacting polypeptides were detected by all three affinity-purified antibodies (A, B and C); this might indicate the existence of additional large GAGA factors that could result from alternative splicing of *Trl* transcripts in tissue culture cells. However, no corresponding cDNA clones have been recovered.

GAGA-519 and GAGA-581 factors form multiple nucleoprotein complexes on a single GAGA sequence

The availability of full-length recombinant GAGA factor cDNA clones has allowed us to compare DNA binding activities and functions of the major GAGA factor isoforms. GAGA factor binding sites typically contain multiple repeats of GAGA sequence of various lengths, clustered in the promoter regions within a few hundred base pairs upstream of the transcription start sites (2). We recovered the GAGA-519c cDNA clone (Fig. 1a) in binding site screening experiments using the atypical 10 bp Adf-2 motif from the *Drosophila Adh* distal promoter, 5′-TCTCAGTGCA-3′, with only a single GAGA sequence. Using a competitive gel mobility shift assay, we have shown that recombinant GAGA proteins can indeed interact with either Adf-2a (located at –202) or Adf-2b (located at +8) sites (Fig. 3). Unexpectedly, purified recombinant GAGA-519 and GAGA-581 factors, each typically one major single polypeptide detected by SDS–PAGE (Fig. 9a and b), produced multiple nucleoprotein complexes when bound to a ^{32}P-labelled DNA probe containing the Adf-2b site (Fig. 3). From each GAGA factor isoform, at least four major nucleoprotein complexes formed, generating a discrete, regularly spaced ladder on an agarose gel; their relative mobilities are consistent with binding of additional GAGA factor molecules. GAGA factor binding was competed away by the 32 bp DNA sequence containing the wildtype Adf-2a site; this competition was ~5-fold more efficient than that by a 29 bp mutated Adf-2a site, with part of the single GAGA sequence deleted (Adf-2aΔ3). These results indicate specific and similar binding of either GAGA factor to the single GAGA sequence in the 10 bp Adf-2 motif, consistent with the presence of an identical DNA binding domain in both protein isoforms. Interestingly, multiple, clearly related GAGA factor–DNA complexes, were detected concomitantly on the Adf-2b site (at several GAGA protein concentrations, not shown) and all complexes could be competed similarly (Fig. 3), which suggests that purified recombinant GAGA-519 and GAGA-581 may exist as several alternative homomeric complexes, each of which has similar DNA binding activity.

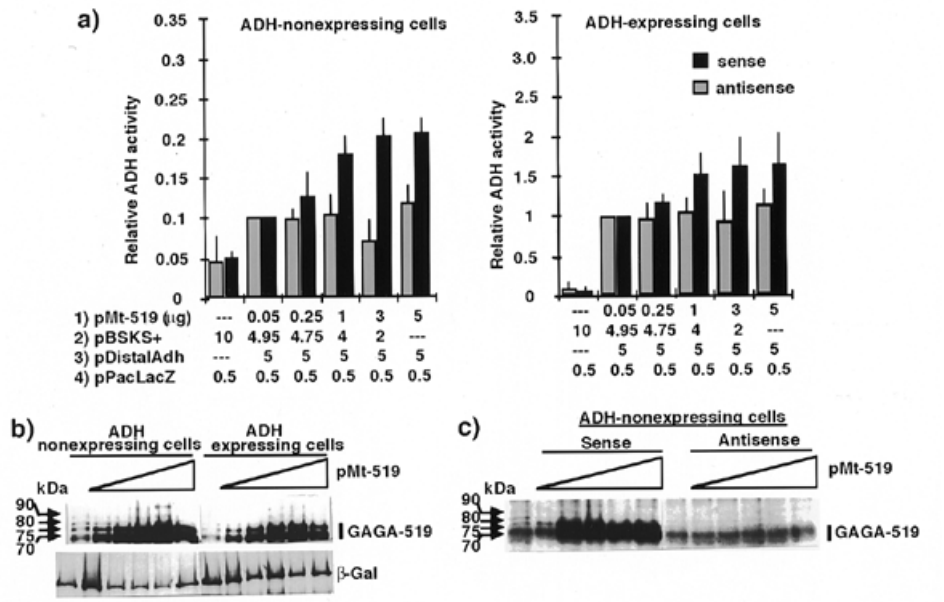


Figure 4. Activation of the *Adh* distal promoter by GAGA-519. pMt-GAGA-519, in sense and antisense orientations, was transiently transfected into S2 ADH-non-expressing (ADH$^-$) and 1006-2 ADH-expressing (ADH$^+$) cell lines. (**a**) The metal-inducible expression vector was transfected, along with pDistalAdh, which contains the wildtype *Adh* gene (with 660 bp of upstream distal promoter elements and distal enhancer) as a reporter (24), pPacLacZ (containing the *Drosophila* constitutive *act5C* exon 1-proximal promoter fused in frame to *LacZ*; 25) and pBSKS$^+$. The ADH activity relative to β-galactosidase activity (averages of six experiments) was measured and is presented as fold increase over the activity seen in the presence of 0.05 μg of the pMt-GAGA-519 expression vector in ADH-expressing cells. The activity of pDistalAdh is ~10-fold higher in ADH$^+$ than in ADH$^-$ cells (23,32) Error bars indicate ± one standard deviation. (**b**) Western blot analysis of nuclear proteins from cells transfected with pMt-GAGA-519 (sense) and probed with affinity-purified rat anti-GAGA$^{\text{BTB-ZF-C-ter 519}}$ antibody. Transfected whole cell lysates were also probed with mouse monoclonal anti-β galactosidase antibody (bottom panel) (**c**) Western blot analysis of ADH$^-$ cells transfected with pMt-GAGA-519 sense or pMt-GAGA-519 antisense.

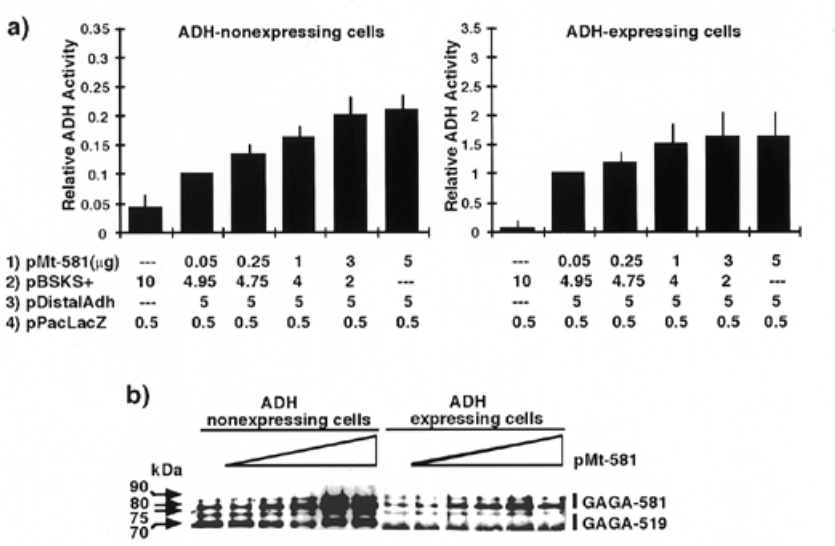


Figure 5. Activation of the *Adh* distal promoter by GAGA-581. (**a**) pMt-GAGA-581 was transfected, as was done with pMt-GAGA-519 (Fig. 4 legend). (**b**) Western blot analysis of nuclear proteins from cells transfected with pMt-GAGA-581 and probed with affinity-purified anti-GAGA$^{\text{BTB-ZF-C-ter 519}}$ antibody.

Either GAGA-519 or GAGA-581 activates expression from the *Drosophila Adh* distal promoter

We compared the capability of GAGA isoforms to *trans*-activate the *Drosophila Adh* distal promoter by overexpressing them in tissue culture cells together with an *Adh* reporter gene, pDistalAdh, which contains an intact *Adh* gene in its natural context (24). Cotransfection experiments were performed in two different cell lines, ADH-non-expressing cells (ADH$^-$) and ADH-expressing cells (ADH$^+$) that possess different combinations of *trans*-acting factors necessary for *Adh* distal promoter transcription (15,23,32). Increasing amounts of pMt-GAGA-519 in the sense orientation

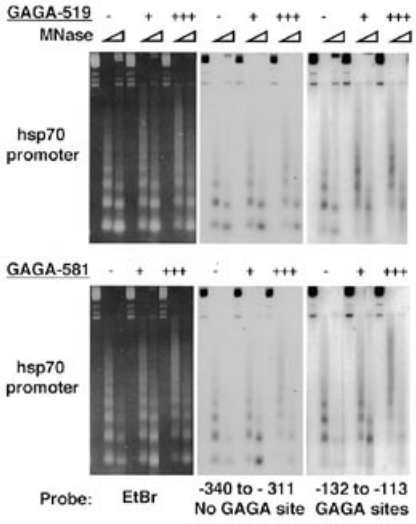


Figure 6. GAGA-519 and GAGA-581 both mediate specific chromatin rearrangement at the *hsp70* promoter *in vitro*. Chromatin reconstituted on the *hsp70*-plasmid, without or with addition (0.5 and 1 µg) of recombinant GAGA-519 (top) and GAGA-581 (bottom), was digested with MNase and the digestion patterns analyzed on agarose gels. Ethidium bromide (EtBr) stains of MNase digestion patterns are shown on the left; DNA blots were hybridized sequentially with two oligonucleotide probes: –340 to –311 in the *hsp70* promoter, which contains no GAGA binding site and –132 to –113, which contains several GAGA binding sites (3).

led to gradual increases in the 70/75 kDa GAGA-519 protein above the endogenous levels in nuclei of both ADH$^-$ and ADH$^+$ cells (Fig. 4b and c). Similar levels of overexpressed GAGA-519 proteins were seen in both cell types. In parallel, there were also gradual increases in ADH expression in a GAGA-519 dosage-dependent manner; however, the increase was more pronounced in ADH$^-$ cells, ~2-fold compared to 1.5-fold in ADH$^+$ cells (Fig. 4a). *Trans*-activation was due to overexpressed GAGA-519; pMt-519 expression vector in the antisense orientation did not produce any 70/75 kDa GAGA-519 above the endogenous level (Fig. 4c) and did not systematically alter ADH expression from pDistal *Adh* (Fig. 4a). In both cell types, overexpression of GAGA-519 did not affect a second *Drosophila* promoter, the *act5C* exon 1-proximal promoter, which drove the expression of a transfected *LacZ* gene (Fig. 4b). This is consistent with previous findings that transcription from the *act5C* exon1-proximal promoter is not dependent on GAGA binding sites (33). Overexpression of the 80/90 kDa GAGA-581 (Fig. 5b) also stimulated ADH activity in a dosage-dependent manner, by 2-fold in ADH$^-$ cells and by ~1.5-fold in ADH$^+$ cells (Fig. 5a). The relatively low level of stimulation suggests that GAGA factor is not limiting in these cells. In both experiments, transcripts from the single cDNA clone resulted in the synthesis of multiple isomers, consistent with the results of Figure 2; this result supports the inference that these sub-groups of isoforms are the consequence of post-translational modification.

GAGA-581, like GAGA-519, can direct rearrangement of nucleosomes during chromatin assembly *in vitro*

GAGA factor may function by influencing chromatin structure (reviewed in 1,2). A recombinant 70 kDa GAGA-519 protein has

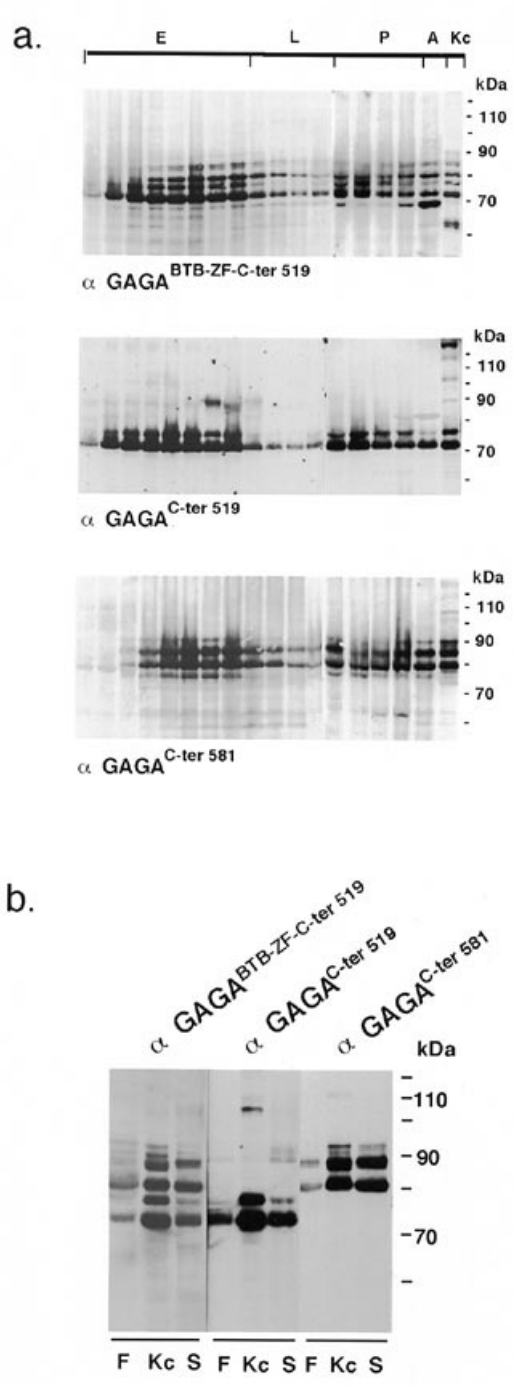


Figure 7. Western blot analysis of GAGA factors during development and in differentiated larval tissues. (**a**) Equal amounts of protein extracts from staged animals and 20 µg KcO nuclear extract were fractionated and analyzed using affinity-purified primary antibodies as indicated. E, embryos (0–3, 3–6, 6–9, 9–12, 12–15, 15–18, 18–21 and 21–24 h); L, larvae (first, second, third and late third instar); P, pupae (24 h apart); A, adults. (**b**) Protein extracts from larval fat body (F, from ~3 larvae/lane), from salivary gland (S, from ~3 larvae/lane), and from a nuclear extract (Kc, 20 µg) were analyzed using affinity-purified primary antibodies as indicated.

been shown to direct rearrangement of nucleosomes within the regulatory regions of *hsp70* and *hsp26* during physiological chromatin assembly *in vitro* (3,34). We have confirmed and extended these results using the same ATP-dependent chromatin

assembly extract from *Drosophila* embryos to reconstitute chromatin on an *hsp70* plasmid (3). Micrococcal nuclease digestion patterns indicate the assembly of regularly spaced nucleosomes on the 6 kb plasmid as a whole with or without addition of either recombinant GAGA protein (Fig. 6). Nucleosome rearrangement was observed specifically and locally over the GAGA binding sites of the *hsp70* promoter only when recombinant GAGA proteins (either isoform) were present. This is indicated by the disarray of the oligonucleosomal DNA ladder and by the decrease of mononucleosomal DNA when the Southern blots were probed with an oligonucleotide containing the GAGA binding sites. These experiments indicate that recombinant GAGA-519 and GAGA-581 can function similarly to modulate chromatin structure *in vitro*.

Distribution of GAGA factor isoforms during *Drosophila* development and differentiation

We have examined GAGA factor isoforms at various developmental stages and in certain differentiated larval tissues by Western blot analyses using the domain-specific antibodies described above. GAGA factors are present at all stages (Fig. 7a); however, combinations of specific isoforms differ significantly through development. During early embryogenesis the 70/75 kDa GAGA-519 proteins are the predominant species. At the start of organogenesis in 6 h embryos, the 80/90 kDa GAGA-581 proteins begin to accumulate. They become as abundant as the 70/75 kDa forms by 12 h of embryogenesis and throughout development; this pattern is similar to that seen for tissue culture cell nuclei (Fig. 7a). Variations are observed in differentiated larval tissues (Fig. 7b). This may indicate tissue-specific regulation of post-translational modification.

In larval salivary gland cells, GAGA factors have been localized along polytene chromosomes to a large number of sites within the euchromatic part of the genome; no staining is observed in the heterochromatic chromocenter (3; reviewed in 2). These earlier experiments detected all GAGA species. To distinguish the locations of the 70/75 kDa GAGA-519 and the 80/90 kDa GAGA-581 isoforms, we performed indirect immunofluorescent costaining of polytene chromosomes with affinity-purified rat anti-GAGA$^{BTB\text{-}ZF\text{-}C\text{-}ter\ 519}$ antibody, which detects all GAGA species (Fig. 8a) and rabbit anti-GAGA$^{C\text{-}ter\ 581}$ antibody, which detects only the GAGA-581 species (Fig. 8b). The merged micrograph (Fig. 8c) reveals, within the limits of detection, a complete overlap of the specific staining patterns; there is no difference in the relative intensity of various bands and of no staining at the chromocenter. Identical observations were made when affinity-purified rat anti-GAGA$^{BTB\text{-}ZF\text{-}C\text{-}ter\ 519}$ and rabbit anti-GAGA$^{C\text{-}ter\ 519}$ antibodies were used (not shown). Strikingly, neither GAGA-581 or GAGA-519 signals were seen alone (Fig. 8c). These observations indicate that the major GAGA-581 and GAGA-519 proteins, which coexist in near-equal amounts in salivary gland nuclei as shown by Western blot analysis (Fig. 7b), colocalize completely to specific chromosomal regions. Our observations raise the significant possibility that in differentiated tissues such as salivary gland, GAGA-519 and GAGA-581 might bind chromatin as heteromeric protein complexes. However, the absence of 80/90 kDa GAGA-581 during the first 6 h of embryogenesis (Fig. 7a) indicates that GAGA-519 can also function without GAGA-581 proteins.

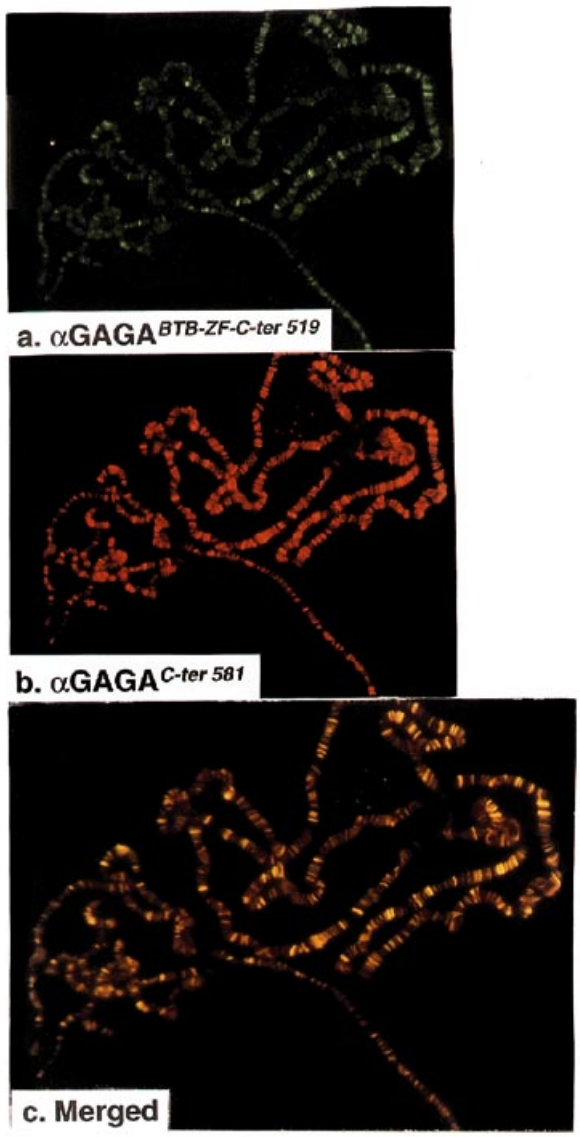


Figure 8. Colocalization of GAGA-519 and GAGA-581 along polytene chromosomes. (**a**), (**b**) (**c**) Fluorescent micrographs of a single set of polytene chromosomes costained with affinity-purified anti-GAGA$^{BTB\text{-}ZF\text{-}C\text{-}ter\ 519}$ (green) and anti-GAGA$^{C\text{-}ter\ 581}$ (red) antibodies. (c) Double exposure (merged) showing regions of overlap (yellow). Note the absence of chromocenter staining (upper right).

Direct heteromeric protein interactions between GAGA factor isoforms

To investigate whether GAGA-519 and GAGA-581 proteins physically interact, we performed immunoprecipitation assays (Fig. 9). Under conditions where pre-immunization rabbit sera failed to precipitate purified recombinant GAGA factors, both isoform-specific antibodies precipitated only their antigens and specifically detected these antigens in western blot analysis (Fig. 9a). When recombinant GAGA factors were mixed together prior to immunoprecipitation (Fig. 9b), anti-GAGA$^{C\text{-}ter\ 519}$ antibody specifically precipitated both GAGA-519 (lanes 2 and 6) and GAGA-581 (lanes 10 and 14). Conversely, anti-GAGA$^{C\text{-}ter\ 581}$

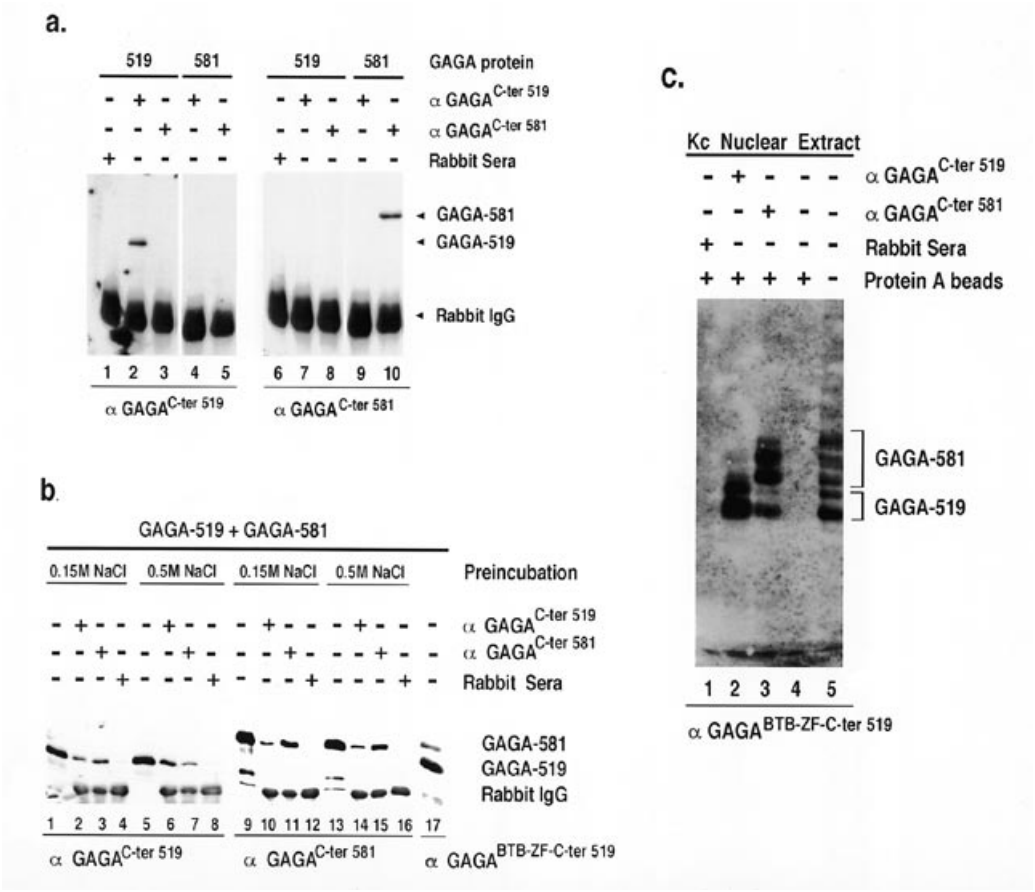


Figure 9. Heteromeric interactions of GAGA factor isoforms. (**a**) Specificity of anti-GAGA antibodies. Purified recombinant GAGA-519 or GAGA-581 (250 ng each) was incubated with the indicated rabbit sera (αGAGA$^{C\text{-ter }519}$, αGAGA$^{C\text{-ter }581}$ and pre-immunized rabbit serum; see Materials and Methods) Proteins in each immunoprecipitate were divided into two halves, fractionated on an 8% SDS–PAGE gel and each half was analyzed by Western blotting with affinity-purified primary antibodies as indicated. Goat anti-rabbit IgG was used as the secondary antibody in the chemiluminescence detection; it recognized precipitating rabbit antibodies bound to protein A beads, as expected. (**b**) Heteromeric interactions between GAGA-519 and GAGA-581. Mixtures of purified recombinant GAGA-519 and GAGA-581 (1 µg each) were pre-incubated in a solution containing either 0.15 M NaCl (lanes 1–4, 9–12 and 17), or 0.5 M NaCl (lanes 5–8 and 13–16) for 1 h at 4°C. The immunoprecipitation and Western blot assays were then performed on equal aliquots of these samples as described above. Lanes 1, 5, 9 and 13 contained 125 ng each of GAGA-519 and GAGA-581 loaded directly onto the gel. Lane 17 shows the reprobing of proteins in lane 1 with rat anti-GAGA$^{BTB\text{-}ZF\text{-}C\text{-}ter\ 519}$ antibody. (**c**) Heteromeric complexes of GAGA-519 and GAGA-581 in tissue culture cell nuclear extract. 1.5 mg of crude KcO nuclear extract was pre-incubated with protein A beads; equal aliquots were then immunoprecipitated (with 2.5 µl of the indicated rabbit sera, lanes 1–3) and analyzed by Western blotting with rat anti-GAGA$^{BTB\text{-}ZF\text{-}C\text{-}ter\ 519}$ antibody. Lane 4, protein A beads after pre-incubation with the crude nuclear extract. Lane 5, 20 µg of KcO nuclear extract.

antibody specifically precipitated GAGA-519 (lanes 3 and 7) as well as GAGA-581 (lanes 11 and 15). GAGA factor associations appear unaffected by salt concentrations [range from 0.5 M NaCl (Fig. 9b) to 0.05 M NaCl (result not shown)]. In addition, GAGA-519 and GAGA-581 proteins, including their *in vivo* modified subtypes, could be specifically coimmunoprecipitated from crude tissue culture cell nuclear extracts (Fig. 9c). Note that the levels of GAGA isoform association seen in Figure 9c, lanes 2 and 3 may be an underestimate, as the specific precipitating polyclonal antibodies may interfere with GAGA protein interactions. Together, our observations in Figures 8 and 9 indicate strongly that heteromeric complexes of GAGA-519 and GAGA-581 exist *in vivo*.

DISCUSSION

GAGA factor is a remarkable protein. It binds with high affinity to a discrete set of specific DNA sequences; it can direct specific modulation of the nucleosome array and it has been shown by genetic analysis to have an effect on higher order chromatin structure (reviewed in 1,2). Multiple isoforms of GAGA factor have been recognized for some time, but the source of this variation had not been identified. We have presented a set of experiments indicating that the GAGA factors consist of several related nuclear proteins, all encoded by the *Trithorax-like* gene. The GAGA proteins share a common N-terminal region that is linked to two different C-terminal domains. Alternative splicing of *Trl* transcripts is the most likely mechanism to yield such polypeptides both sharing and differing in primary sequences. Both primary GAGA species are apparently modified post-translationally to generate subsets with altered electrophoretic mobility. Both splicing of *Trl* RNA and GAGA protein modification are regulated processes, since combinations and amounts of various GAGA species differ during development and in specific cell types.

We have identified two major forms of GAGA factor, the 70/75 kDa GAGA-519 and the 80-90 kDa GAGA-581; there may be additional minor GAGA proteins as well. Both GAGA-519 and

GAGA-581 are relatively abundant and likely to have significant functional roles. Only GAGA-519 proteins accumulate during the first 6 h of embryogenesis, thereafter both GAGA-519 and GAGA-581 are present in near-equal amounts throughout development (Fig. 7a and b). Thus, GAGA-519 proteins are likely to be the GAGA factor involved in transcriptional regulation and nuclear division in early embryos (6). However, within the limits of detection, GAGA-519 and GAGA-581 completely colocalize to specific regions along euchromatic arms of salivary gland polytene chromosomes (Fig. 8c). This observation, coupled with tests indicating similar functional capabilities, suggests that GAGA-519 and GAGA-581 both function in later developmental stages; they may function together as heteromeric protein complexes (Fig. 9b and c) when they coexist in near equal amounts (Figs 2, 7a and b) in the same nucleus (Fig. 8).

GAGA proteins can form both homo- and heteromeric complexes. We detected similar multiple-related large GAGA factor–DNA complexes when either purified recombinant GAGA- 519 or GAGA-581, each a single polypeptide, bound specifically to a single GAGA sequence within the Adf-2 motif (Fig. 3). Our results suggest stable homomeric interactions of GAGA proteins. Further experiments (Figs 8 and 9) indicate that GAGA-519 and GAGA-581 form heteromeric complexes *in vivo* and *in vitro*. Such stable protein homo- and heteromultimerization might be mediated by either the BTB domain (28–30) and/or by the glutamine repeats (35–38). Whatever the mechanisms responsible, the complete colocalization of GAGA-519 and GAGA-581 suggests that the heteromeric GAGA complexes may be the preferred stable structures when both isoforms coexist *in vivo*.

Two different assays have shown that GAGA-519 and GAGA-581 can function similarly. In transient cotransfection assays in tissue culture cells, overproduction of either GAGA-519 or GAGA-581 stimulated expression from the *Adh* distal promoter (Figs 4 and 5). Given that the endogenous factor was present in both forms (Fig. 2), the active species could be a homomer or heteromer. In chromatin remodeling experiments where purified recombinant GAGA proteins were used (Fig. 6), the active structures were probably homomeric complexes. These two functional assays confirmed that GAGA-519 and GAGA-581 have similar sequence-specific DNA binding activities and interact similarly with other *trans*-acting factors and the general transcriptional machinery *in vivo*, as well as with nucleosomes and nucleosome remodeling factors *in vitro*.

In summary, it appears that complexity of GAGA factors exists at several levels: multiple primary amino acid sequences, varying levels of post-translational modification of the same GAGA polypeptide and potentially distinct homo- and heteromeric complexes. It appears possible that the multimerization of GAGA factor into higher-order complexes may play a role in the organization of higher-order chromatin structure (7), for example by generating stable chromatin loops between distant regulatory regions that contain GAGA factor binding sites. In addition, GAGA factor interactions may be involved in the recruitment of additional GAGA factor to active transcription units and in the spreading of GAGA factor to open the chromatin structure during RNA polymerase elongation *in vivo* (14). Such protein–protein interactions may well be critical for the mechanism(s) by which this specific DNA-binding protein affects higher order chromatin structure.

ACKNOWLEDGEMENTS

We thank Yuli Chen for sequencing the GAGA-519c cDNA clone and Toshi Tsukiyama, Peter Becker and Carl Wu for the heat shock plasmid and for their advice on the chromatin assembly experiments. This work was supported by grants from the National Institutes of Health to C.B. (GM 34850) and to S.C.R.E. (GM 31532).

REFERENCES

1 Becker,P.B. (1995) *Seminars in Cell Biol.* **6,** 185–190.
2 Granok,H., Leibovitch,B.A., Shaffer,C.D. and Elgin,S.C.R. (1995) *Current Biol.* **5**, 238–241.
3 Tsukiyama,T., Becker,P.B. and Wu.C. (1994) *Nature* **367,** 525–531.
4 Soeller,W.C., Oh,C.E. and Kornberg,T.B. (1993) *Mol. Cell. Biol.* **13,** 7961–7970.
5 Raff,J.W., Kellum,R. and Alberts,B. (1994) *EMBO J.* **13,** 5977–5983.
6 Bhat,K.M., Farkas,G., Karch,F., Gyurkovics,H., Gausz,J. and Schedl,P. (1996) *Development* **122,** 1113–1124.
7 Farkas,G., Gausz,J., Galloni,M., Reuter,G., Gyurkovics,H. and Karch,F. (1994) *Nature* **371,** 806–808.
8 Soeller,W.C., Poole S.J. and Kornberg,T.B. (1988) *Genes Dev.* **2,** 68–81.
9 Biggin,M.D. and Tjian,R. (1988) *Cell* **53,** 699–711.
10 Croston,G.E., Kerrigan,L.A. Lira,L.M., Marshak,D.R. and Kadonaga,J.T. (1991) *Science* **251,** 643–649.
11 Kerrigan,L.A., Croston,G.E., Lira,L.M. and Kadonaga,J.T. (1991) *J. Biol. Chem.* **266,** 574–582.
12 Lu,Q., Wallrath,L.L., Granok,H. and. Elgin,S.C.R. (1993) *Mol. Cell. Biol.* **13,** 2802–2814.
13 Shopland,L.S., Hirayoshi,K., Fernandes,M. and Lis,J.T. (1995) *Genes Dev.* **9,** 2756–2769.
14 O'Brien,T., Wilkins,R.C., Giardina,C. and Lis,J.T. (1995) *Genes Dev.* **9,** 1098–1110.
15 Benyajati,C., Ewel,A., McKeon,J., Chovav,M. and Juan,E. (1992) *Nucleic Acids Res.* **20,** 4481–4489.
16 Gilmour,D.S., Thomas,G.H. and Elgin,S.C.R. (1989) *Science* **245,** 1487–1490.
17 Jackson,S.P. and Tjian.R. (1988) *Cell* **55,** 125–133.
18 Vinson,C.S., LaMarco,K.L., Johnson,P.F., Landschulz,W.H. and McKnight,S.L. (1988) *Genes Dev.* **2,** 801–806.
19 Brown,N.H. and Kafatos,F.C. (1988) *J. Mol. Biol.* **203,** 425–437.
20 Hakes,D.J. and Dixon,J.E. (1992) *Anal. Biochem.* **202,** 293–298.
21 Tsukiyama,T. and Wu,C. (1995) *Cell* **83,** 1011–1020.
22 Bunch,T.A., Grinblat,Y. and Goldstein,L.S.B. (1988) *Nucleic Acid Res.* **16,** 1043–1061.
23 Ayer,S. and Benyajati,C. (1990) *Mol. Cell. Biol.* **10**, 3512–3523.
24 Ayer,S., Walker,N., Mosammaparast,M., Nelson,J.P., Shilo,B. and Benyajati.C. (1993) *Nucleic Acids Res.* **21,** 1619–1627.
25 Driever,W. and Nusslein-Volhard,C. (1989) *Nature* **337,** 138–143.
26 Becker,P.B., Tsukiyama,T. and Wu.C. (1994) *Methods Cell Biol.* **44,** 207–223.
27 Clark,R.F., Wagner,C.R., Craig,C.A. and Elgin,S.C.R. (1991) *Methods Cell Biol.* **35,** 203–228.
28 Bardwell,V.J. and Treisman,R. (1994) *Genes Dev.* **8,** 1664–1677.
29 Chen,W., Zollman,S., Couderc,J-L. and Laski,F.A. (1995) *Mol. Cell Biol.* **15,** 3424–3429.
30 Dong,S., Zhu,J., Reid,A., Strutt,P., Guidez,F., Zhong,H.J., Wang,J.-W., Licht,J., Waxman,J., Chomienne,C., Chen,Z., Zelent,A. and Chen.S-J. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 3624–3629.
31 Pedone,P.V., Ghirlando,R., Clore,G.M., Gronenborn,A.M., Felsenfeld,G. and Omichinski,J.G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2822–2826.
32 Ayer,S. and Benyajati.C. (1992) *Mol. Cell. Biol.* **12,** 661–673.
33 Bond-Matthews,B.J. and Davidson.N. (1988) *Gene* **62,** 289–300.
34 Wall,G., Varga-Weisz,P.D., Sandaltzopoulos,R. and Becker.P.B. (1995) *EMBO J.* **14,** 1727–1736.
35 Emili,A., Greenblatt,J. and Ingles.C.J. (1994) *Mol. Cell Biol* **14,** 1582–1593.
36 Gill,G., Pascal,E., Tseng,Z.H. and Tjian,R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 192–196.
37 Pascal,E. and Tjian.R. (1991) *Genes Dev.* **5,** 1646–1656.
38 Stott,K., Blackburn,J.M., Butler,P.J.G. and Perutz.M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6509–6513.