2672–2678 *Nucleic Acids Research, 1998, Vol. 26, No. 11*

GAGA factor binding to DNA via a single trinucleotide sequence element

R. Chris Wilkins and John T. Lis[1,*]

Section of Genetics and Development and [1]Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received February 9, 1998; Revised and Accepted April 16, 1998

ABSTRACT

GAGA transcription factor (GAF) is an essential protein in *Drosophila*, important for the transcriptional regulation of numerous genes. GAF binds to GA repeats in the promoters of these genes via a DNA-binding domain containing a single zinc finger. While GAF binding sites are typically composed of 3.5 GA repeats, the *Drosophila hsp70* gene contains much smaller elements, some of which are as little as three bases (GAG) in length. Interestingly, the binding of GAF to more distant trinucleotide elements is relatively strong and not appreciably affected by the removal of larger GA arrays in the promoter. Moreover, a simple synthetic GAG sequence is sufficient to bind GAF *in vitro*. Here we directly compare the affinity of GAF for different sequence elements by immunoprecipitation and gel mobility shift analysis. Furthermore, our measures of the concentration of GAF *in vivo* indicate that it is a highly abundant nuclear protein, prevalent enough to occupy a sizable fraction of correspondingly abundant trinucleotide sites.

INTRODUCTION

The repressive nature of chromatin has been demonstrated, and those factors capable of alleviating this repression have become a major target of investigation (for review see 1). Not only must a mechanism exist to assemble nucleosomes on genes and facilitate the condensation and formation of higher order chromatin structures, but a means to selectively disrupt chromatin is needed to predicate the subsequent expression of these genes. One factor that plays an important role in promoter architecture in *Drosophila* is the GAGA transcription factor (GAF). Originally identified by virtue of its ability to stimulate transcription from the engrailed (*en*) and ultrabithorax (*Ubx*) promoters *in vitro* (2,3), GAF binds to GA-rich sequences, $(CT{\cdot}GA)_n$, found in the promoters of numerous *Drosophila* genes (for reviews see 4,5). Interestingly, GAF appears to exert its effect not by activating or repressing the transcriptional machinery directly, but by mitigating the repressive effects of histones (6,7).

GAF has specifically been shown to play a critical role in the expression of the well-characterized heat shock genes (for review see 4). On the *hsp26* gene, the GA-rich sequences that bind GAF have been shown to be important for the maintenance of DNase I hypersensitive sites in the promoter, sites which are subsequently critical for heat-induced and developmental expression of *hsp26* (8–10). It therefore appears that GAF can keep the promoter in a chromatin conformation which presumably allows for the association of additional factors necessary for transcriptional competence. GAF elements have also been shown to be critical in the regulation of the *hsp70* gene. Mutations in GAF elements in the promoter of *hsp70* have been shown *in vivo* to specifically affect the formation of promoter-paused RNA polymerase II (pol II), the accessibility of heat shock factor (HSF) to its target sequences and transcriptional activation (11,12). Further analysis of *hsp70* sequences indicates that recombinant GAF can disrupt nucleosomes on *in vitro* assembled templates (13), and can facilitate the ‘remodeling’ of chromatin in an ATP-dependent fashion with the help of a remodeling complex termed NURF (14). It is therefore reasonable to conclude that GAF stimulates gene expression by opening chromatin and maintaining the promoter in a conformation which sets the stage for the binding of other sequence-specific factors and pol II.

GAF does not appear to limit its interactions to the promoter. *In vivo* crosslinking on *hsp70* and *hsp26* indicates that GAF association is restricted to promoter sequences prior to heat shock, but that after induction it associates with sequences throughout the body of the gene (15). Interestingly, within the gene itself there are very few large GA arrays, which are characteristic of GAF regulatable promoters; however, there are several smaller elements of 3–5 nucleotides which occur every 75 bases on average (R.C.Wilkins and J.T.Lis, unpublished observation). The nature of the association and the function of GAF over the length of the transcriptional unit therefore remains an open question.

Originally, according to sequence comparison, GAF elements were believed to be comprised of ~3.5 (GA) repeats (5). However, the binding domain of GAF contains a single zinc finger and may require no more than a single trinucleotide sequence for GAF specific binding, given that most zinc finger peptides require only a three base pair consensus per finger (16–18). Recently, Omichinski *et al*. (19) have published the NMR structure of the GAF DNA binding domain bound to a pentamer sequence (GAGAG). The zinc finger itself is responsible for base contacts at the first three nucleotides, and basic regions flanking the finger, BR1 and BR2, contact the last two bases of

*To whom correspondence should be addressed. Tel: +1 607 255 2442; Fax: +1 607 255 2428; Email: jtl10@cornell.edu

the pentamer, indicating that a GAGAG pentamer may indeed be the consensus binding site for GAF (19). Interestingly, most single base pair substitutions within this pentamer, except the central G, only moderately affect the affinity of GAF for these sequences.

Here we examine the interaction of GAF with both native and synthetic binding sites, and given the predicted prevalence of low affinity GAF elements in the genome, we also examine the abundance of GAF in the nucleus. While GAF strongly interacts with its known high affinity binding sites, it also interacts more extensively with lower affinity sequences in the *hsp70* promoter, around the start site, and into the beginning of the transcribed region. Low affinity elements as small as GAG trinucleotides appear to be targets of GAF binding as assayed by DNase I footprinting. Filling of more distant minimal sites appears largely uncooperative with the occupancy of other GAF binding sites. In addition, a trinucleotide sequence is sufficient to confer GAF binding to synthetic sites with an affinity only a few fold less than that of a larger 'high affinity' site. Measurements of GAF abundance indicate that GAF is plentiful, and is present at a concentration in the nucleus sufficient to bind minimal trinucleotide sites *in vivo*.

MATERIALS and METHODS

Purification of GAF fusion proteins

Histidine-tagged GAGA protein (His-GAF) was overproduced and purified from *Escherichia coli* BL21/DE3pLysS cells using His-bind resin (Novagen) according to the manufacturer's recommendations. Cells were grown to an OD_{600} between 0.5 and 1.0, and induced by addition of 1 mm IPTG (Sigma). Cells were grown for an additional 3 h and harvested by centrifugation. Prior to use, purified product was dialyzed into buffer containing 100 mM NaCl, 10 mM Tris pH 8, 10% glycerol, 0.1 mM EDTA and 1 mM DTT. The His-GAF fusion protein was coded for by the p10His-GAF plasmid which contains amino acids 5–519 of GAF fused in frame to the 10 histidine residues of pET16b. This construct was generated by ligating an *Nde*I–*Hin*dIII fragment from the pAR-GAGA plasmid (20) to the pET16b vector treated with the same enzymes.

The maltose binding protein–zinc finger fusion protein (MBP–Zn) was overproduced and purified from *E.coli* BL21 cells using Amylose resin (NEB) according to the manufacturer's recommendations. Cells were grown, induced and harvested as above. The MBP–Zn fusion protein was coded for by the pMal-Zn plasmid which contains amino acids 244–407, inclusive of the zinc finger and the basic regions BR1 and BR2 of the GAF, generated by dropping a *Bam*HI–*Sal*I cut PCR product into the pMalC2 vector (NEB). The PCR product included bases 1–1220 of GAF, and was generated with the following oligos: RCW3 (5′-GGCGGGAATTCATGTCGCTGCCAATGAACTCGCTG-3′) and Zn_{ex} (5′-TCCATCACTGTCGACTGTGTTGTCGCCCTC-3′).

In vitro footprinting

Preparation of end-labeled DNA for DNase I footprinting was performed by PCR-amplification of fragments from plasmids containing the *hsp70* gene with primers described previously (21). Either the top or bottom DNA strand was ^{32}P-labeled by kinasing the appropriate strand primer before the amplification reaction (21). Approximately 30 fmol of gel purified, end labeled DNA was incubated in a 25 µl binding reaction with His-GAF protein in GAGA binding buffer (GBB): 80 mM KCl, 10 mM HEPES pH 8.3, 5 mM $MgCl_2$, 10% Glycerol, 1 mM DTT, 0.1% NP40. The upper level of His-GAF used in all footprinting experiments was 75 ng, and was sequentially diluted 3-fold for all subsequent lanes. After a 30 min incubation at room temperature, samples were treated with 1 µl of a 20 µg/ml solution of DNase I (Worthington) for 1 min at room temperature. The reaction was stopped by adding 30 µl 50 mM EDTA and 1% SDS; the DNA was phenol–chloroform extracted and ethanol-precipitated, and run on a 6% polyacrylamide–urea sequencing gel.

In vitro immunoprecipitation

In vitro immunoprecipitations were done using identical levels of His-GAF protein and the *hsp70* promoter fragments described for DNase I footprinting. Binding reactions were for 30 min on ice in GBB, at which point GAF/Zn+Q specific antibody (15) was added and incubated for an additional 1 h at 4°C. Complexes were then precipitated with Protein A–Sepharose beads (Sigma) pre-blocked with salmon sperm DNA. The beads were then washed four times with GBB, and the DNA was extracted by boiling the samples in 1% SDS and analyzed on a 6% polyacrylamide–urea sequencing gel.

Gel mobility shifts

Probes for band-shifts were generated by kinasing the indicated oligonucleotides with [γ-^{32}P]ATP. The complimentary unlabeled oligo was mixed with the kinased oligo and annealed by heating the mix to 95°C for 5 min., cooling to room temperature in a 75°C heat block, and finally cooling to 4°C on ice. Approximately 30 fmol of double-strand probe was incubated for 30 min at room temperature in a 25 µl binding reaction with the indicated protein. As with footprinting and immunoprecipitation experiments, binding reactions were in GBB with a maximal protein mass of 75 ng. In addition, 50 ng poly(dG–dC) was added to gel shift experiments to reduce non-zinc-finger specific binding. Samples were run out for analysis on a 1% agarose gel in 1× TBE buffer.

Determination of relative binding constants

The relative binding constants of the various DNA fragments analyzed were determined as described previously (22,23). Typically, oligonucleotides were mixed together with a limiting amount of purified MBP–Zn or His-GAF in a single binding reaction as described above (~10 ng of either protein was used). Bound and unbound DNA species were separated by electrophoresis on a 1% TBE agarose gel and eluted by crushing the gel slice in Gilbert's buffer and incubating overnight. Samples were analyzed side-by-side on a 10% denaturing acrylamide gel. Gels were analyzed on a Molecular Dynamics STORM 840, and species were quantitated using the Molecular Dynamics ImageQuant™ program. Relative binding constants of different oligonucleotides n and m can therefore be calculated by using the equation $K_{relative} = K_n/K_m = [C_n/D_n]/[C_m/D_m]$, where C is the concentration of labeled DNA in the bound fraction and D is the concentration of the same labeled species in the unbound fraction (22,23).

Western analysis

Purified His-GAF and S2 nuclear proteins were size-fractionated by electrophoresis on 10% SDS–polyacrylamide gels. Nuclei from *Drosophila* S2 cells were prepared as described previously (24). Gels were electroblotted onto nitrocellulose membranes, probed with GAF specific antibodies (15) and detection was carried out using an ECL detection reagent kit (Amersham). Protein species in nuclei samples were quantified by comparison to known amounts of recombinant GAF protein, that itself was quantified by comparison to BSA standards by Coomassie and OD_{260}.

Immunodepletion

Nuclei from *Drosophila* Kc cells were prepared as described previously (24) and extract was prepared by sonication. The extract was incubated on a rotary wheel for 1 h at 4°C in the presence or absence of GAF specific antibody. Antibody–protein complexes were then precipitated by centrifugation after incubation for an additional 30 min at 4°C with protein A–Sepharose beads. The remaining supernatant was size-fractionated by electrophoresis on 10% SDS–polyacrylamide gels and analyzed by silver staining or western blotting with an ECL detection reagent kit (Amersham).

RESULTS

GAF protects trinucleotides on the *hsp70* promoter

GAF has been shown to interact with a variety of different sequence targets on the *hsp70* gene (15). Here we examine in detail the affinity of GAF for different promoter sequences using a His-GAF construct coding for residues 5–519 of GAF. This recombinant source lacks bulky fusion moieties and includes the entire GAF sequence, except the first five amino acids which are dispensible for DNA binding (25). This recombinant protein is stable and easily purifiable by its histidine tag as a full length product (Fig. 4A).

Figure 1 shows GAF protection of *hsp70* sequences from DNase I digestion as a function of protein concentration. At low levels of GAF, high affinity sites within the promoter are readily protected, as represented by open ovals. As protein concentration increases, this protection spreads to sequences flanking these strong sites and additional distinct sites of protection become evident, as indicated by closed circles in Figure 1. Spreading of the GAF footprint is more prominent on the bottom stand of *hsp70* (Fig. 1), where it can be seen to extend from the lower affinity sites flanking the TATA box down towards the transcription start site, consistent with recent reports of GAF binding in this region (26). In contrast, at higher GAF concentrations the top strand of *hsp70* shows several regions of hypersensitivity. These regions of increased accessibility to DNase I, indicated by arrows, also map to key regions of the promoter. Hypersensitive bands are seen between GAF binding sites in and around the TATA box, heat shock elements (HSEs) and the start site of transcription. This hypersensitivity appears to be strand-specific, as no such hyper-reactive sequences are seen on the bottom strand with increasing GAF concentrations. Interestingly, regions of hypersensitivity on the top strand of hsp70 and extended protection on the bottom strand coincide with the presence of low affinity sites. Sequence inspection of these sites reveals GA repeats of only 3–4 bases in length, with a subset of these comprised of GAG trinucleotides. The presence of low affinity sites within these key regulatory regions suggests their possible importance for factor accessibility. When compared with the frequency of sites within the gene itself, the density of low affinity sites around the promoter is significantly higher.

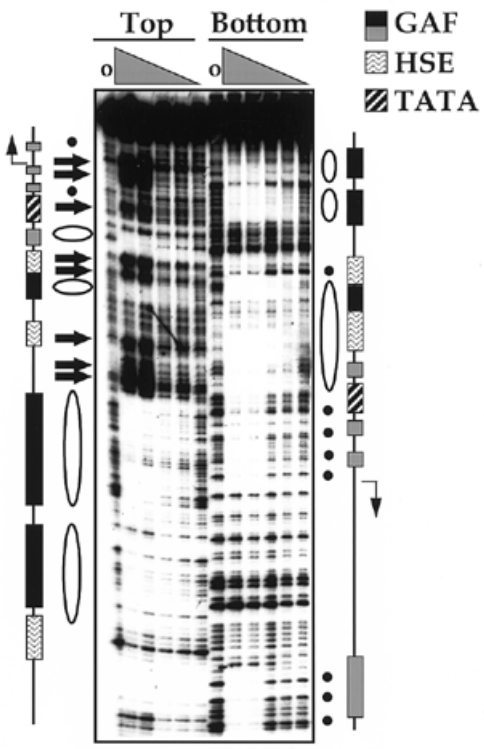

Figure 1. GAF protein binding to high and low affinity sites on the *hsp70* gene as assayed by DNase I digestion. Lanes 1 and 7 are the DNase I patterns in the absence of protein on the top and bottom strands of *hsp70* respectively. Lanes 2–6 and 8–12 are DNase I digestions performed in the presence of 3-fold decreases in the concentration of purified His-GAF protein. High affinity GAF sites are delineated with solid black boxes, and low affinity GAF sites are indicated by gray boxes, corresponding regions of protection are shown as open ovals and filled in circles, respectively. The location of HSEs and the TATA box are shown as gray stippled and black stripped boxes respectively.

When sequences at weak and strong GAF binding sites are compared, it is found that strong sites contain GA repeats of 5 bases or longer, whereas weaker sites contain a simpler GAG motif. At the highest concentration of GAF, all GAG sequences in the fragment are occupied. Since the concentration of the DNA used in these experiments is extremely low relative to that of GAF, we were able to estimate the apparent dissociation constant as equal to the protein concentration giving half occupancy of a particular site. The difference in affinities of GAF for strong and weak sites is <10-fold, with high affinity sites having a K_d of ~1.3 $\times 10^{-9}$ M.

Occupancy of trinucleotides appears to be independent of high affinity sites

For many DNA binding proteins, the filling of low affinity protein binding sites is highly cooperative and dependent upon nearby

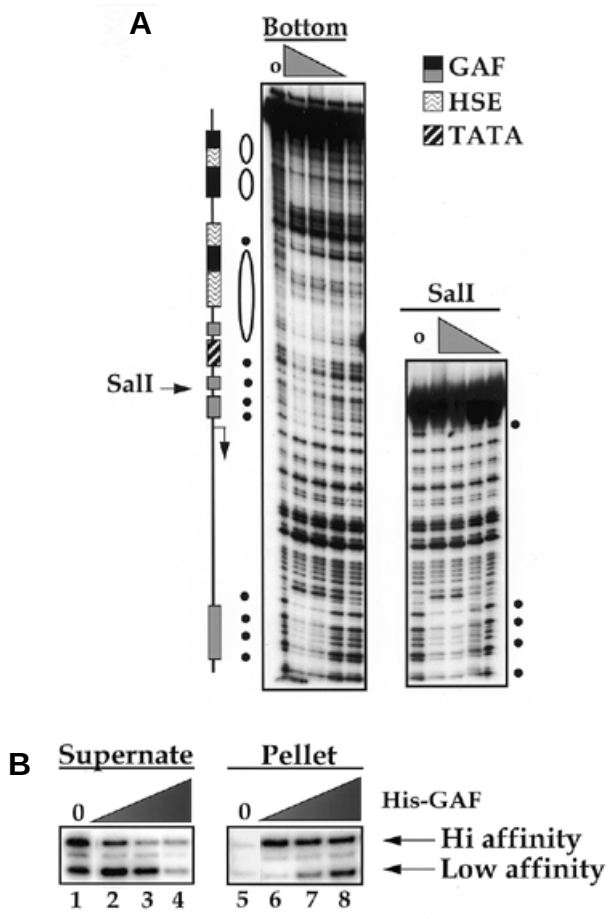

Figure 2. (**A**) GAF binding to low affinity sites on *hsp70* in the presence and absence of high affinity sites in cis. Lanes 1 and 6 are the DNase I pattern in the absence of protein on the bottom strand of *hsp70*. Lanes 2–5 and 7–10 are DNase I digestions performed in the presence of 3-fold decreases in the concentration of purified His-GAF protein on the intact *hsp70* promoter, and the same fragment digested with *Sal*I prior to GAF addition. Site of digestion with *Sal*I is indicated. (**B**) Immunoprecipitation (IP) of high and low affinity sites of *hsp70* digested with *Sal*I. Lanes 2–4 and 6–8 represent the fragments found in the supernatant and pellet of IPs in the presence of increasing concentrations of His-GAF. Lane 1 is the supernatant of a protein-free control, and lane 5 is the pellet. IPs were performed using a GAF specific polyclonal antibody to the C-terminal two-thirds of the protein [Zn+Q (15)]. High and low affinity site fragments are labeled.

stronger binding sites. However, with respect to GAF association on *hsp70*, this appears not to be the case. When high and low affinity sites are separated before GAF binding by restriction digestion, occupancy of these trinucleotides persists. High and low affinity sites were separated by cleavage of *hsp70* promoter DNA sequences with a restriction endonuclease, and their DNase I footprints compared with the same promoter fragments prior to cleavage as a function of GAF concentration (Fig 2A). The occupancy of more distant, low affinity sites is not appreciably affected by the presence and absence of high affinity sites.

Immunoprecipitation assays show that GAF is able to efficiently bind DNA fragments containing either high or low affinity sites (Fig. 2B). High and low affinity sites were separated as before by restriction digestion, incubated with His-GAF and immunoprecipitated with GAF specific antibodies. Three-fold increases in protein concentrations show that GAF can bind lower affinity sites at higher concentrations with a <10-fold difference in affinity when compared with high affinity sites, in agreement with DNase I footprinting results for these fragments.

Relative affinity of GAF for sites of varying length

The NMR structure of the GAF DNA binding domain complexed with DNA shows base contacts over an entire GAGAG pentamer sequence. The zinc finger itself is responsible for base contacts over the first three nucleotides, with basic regions flanking the finger (BR1 and BR2) contacting the last two bases of the pentamer (19). Most single base pair substitutions within this pentamer, except the central guanine, only moderately affect the affinity of GAF for these sequences, perhaps because the zinc finger contacts alone are sufficient for GAF specific binding, and therefore only require a single intact GAG.

To determine if a trinucleotide was indeed sufficient for GAF specific binding, several synthetic oligos were generated (Fig. 3A) and their affinities for an MBP fusion with the zinc finger binding domain of GAF (MBP–Zn) were analyzed by band-shift analysis (Fig. 3B). Sequences for oligos were derived from actual *hsp70* sequences, with minor changes. MBP–Zn binds quite strongly to a high affinity site [$(GA)_4$], moderately to a pentamer, and less so to a trinucleotide (GAG), but is nevertheless still capable of binding all three. When the trinucleotide is changed to TGT or AGA, GAF binding is undetectable.

In order to more quantitatively assess GAF sequence requirements for binding, the relative affinities of GAF for a high affinity site [$(GA)_4$], a pentamer consensus site (GAGAG) and a trinucleotide site (GAG) were examined directly (Fig. 3C). The method of Liu-Johnson *et al.* (22) was used to determine the relative binding affinities within the context of a single binding reaction in the presence of either limiting His-GAF or the GAF DNA binding domain fusion MBP–Zn. Bound and free DNA sequences were separated by electrophoresis, as in a standard band-shift, and the DNAs were eluted and resolved on a denaturing sequencing gel for quantitation (Fig. 3D). The affinity of either GAF source for high [$(GA)_4$] and low affinity (GAG) sites differed by 3.5-fold, with the pentamer consensus (GAGAG) having an intermediate affinity, only 1.5-fold greater than that of a trinucleotide site alone (GAG).

GAF is very abundant in the nucleus

The ability of GAF to bind strongly to trinucleotides indicates that there are a large number of potential binding sites. To determine the abundance of GAF, we examined its prevalence in *Drosophila* S2 nuclei by western analysis, comparing with a recombinant GAF standard (Fig. 4A). For the most abundant GAF species (the 67 kDa isoform), we estimate roughly 3.0×10^5 molecules/cell with a relatively large standard deviation of 1.4×10^5. This number is the average of 10 independent experiments. When totalling all immunoreactive species, we roughly estimate the total number of GAF isomers to be in the order of one million. Taking into consideration the diameter of a *Drosophila* nucleus of 2 µm (27), we calculate a nuclear volume of 3.4×10^{-12} l, and an approximate concentration of the GAF 67 kDa isomer at 130 nM. Assuming one nucleosome per 200 base pairs, and a total genome size of 1.7×10^5 kb, on average one molecule of GAF[67] is present for every 2.5 nucleosomes.

Quantitative western analysis indicates GAF is a very abundant nuclear protein. One might expect a protein this abundant to be

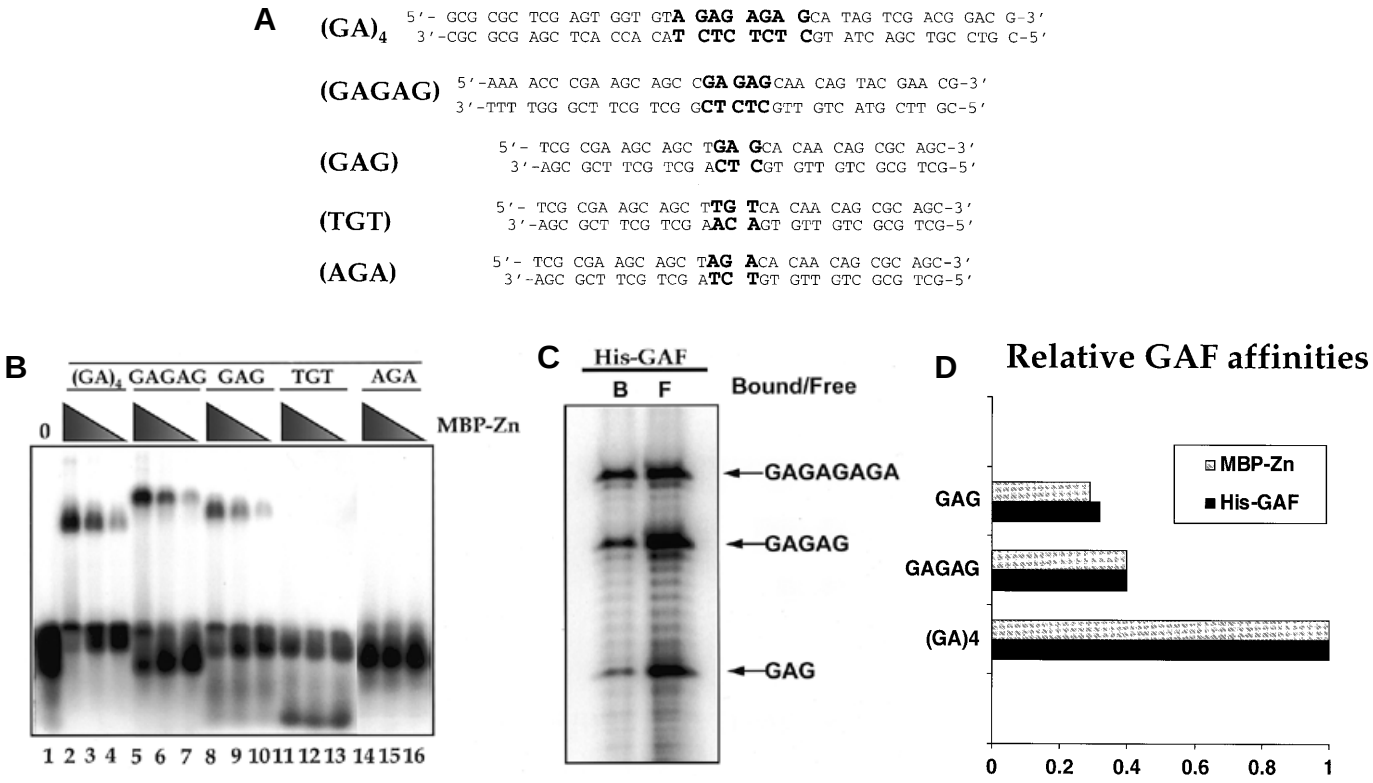

Figure 3. (**A**) Sequences of synthetic oligos used. (**B**) MBP–Zn binding to various synthetic oligos. MBP–Zn binding was tested against oligos containing either an idealized binding element [$(GA)_4$], a pentamer (GAGAG), a trinucleotide (GAG) or alternate trinucleotide sequences (AGA or TGT). (**C**) Comparison of the relative affinities of His-GAF for an idealized binding element [$(GA)_4$], a pentamer consensus, and a trinucleotide (GAG). Sequences present in the bound or free fractions are indicated. (**D**) Graph representing the relative affinities of both His-GAF and MBP–Zn for each of the above sequence elements for three independent experiments of the type shown in (C). Values were normalized against binding to the [$(GA)_4$] element which was assigned a value of one.

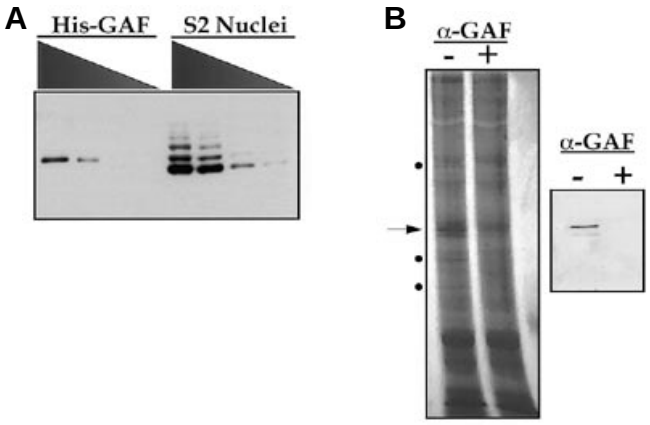

Figure 4. (**A**) Quantitation of GAF molecules per cell. Representative gel of known quantities of *Drosophila* S2 nuclei compared with known concentrations of recombinant His-GAF, analyzed by western blotting with GAF specific antibody. (**B**) Immunodpeletion of the GAF 67 kDa isoform. Representative silver staining and western blotting of *Drosophila* Kc nulcear extract supernatants after a mock or GAF specific immunoprecipitation. The 67 kDa isoform is marked by an arrow, and several other interesting changes in total protein are indicated by filled circles.

easily detected by analyzing total nuclear protein, and indeed it is (Fig. 4B). Immunoprecipitation with GAF specific antibodies causes a change in total nuclear protein as assayed by silver staining with respect to several bands, indicated by closed circles in Figure 4B, and most strikingly in the range of the most prevalent GAF isoform at 67 kDa, indicated by an arrow. Western analysis of duplicate samples run on the same gel verify the identity and depletion of the 67 kDa isoform upon immunoprecipitation with GAF specific antibody, indicating the relative abundance of GAF when compared with total nuclear protein.

DISCUSSION

GAF is an essential nuclear factor in *Drosophila* and has been ascribed a variety of functions. GAF has a global role in chromatin structure and is a critical component of numerous promoters (for review see 4). Although sequence comparison of known GAF elements shows a typical length of 3.5 GA repeats, and recent NMR studies indicate that the binding consensus is a pentamer (GAGAG) (19), here we show that a single trinucleotide (GAG) appears sufficient to dictate GAF specific binding. Most zinc fingers have a core three base pair recognition sequence with only minor contacts outside this core contributing to specificity of the interaction (16–18). GAF contains only one zinc finger of the C_2H_2 variety, which, in addition to some basic N-terminal regions, is sufficient for GAF specific binding (19,25). These N-terminal basic regions, BR1 and BR2, have been implicated in sequence recognition, and have been shown to interact with the remaining two bases of the pentamer consensus;

however, the importance of these regions has not been fully addressed. Single base pair mutations in the pentamer consensus, except the central G (GA**G**AG), have only moderate effects on GAF binding, indicating the relatively minor contribution of BR1 and BR2 to the binding of GAF. With the central G in the pentamer left untouched, any single mutation would still result in a remaining trinucleotide for the zinc finger of GAF to bind. Surprisingly, not only can GAF bind a single trinucleotide site, but it can also do so by means of a single zinc finger which binds with near nanomolar affinity.

Association of GAF with trinucleotides is evident on both the *hsp70* and *hsp26* genes in vitro (data not shown). All GAG trinucleotides in the promoter bind GAF *in vitro* when assayed by DNase I footprinting. The frequency of GAG trinucleotides at the promoter is higher than that seen in the body of the gene, and these triplets are found clustered around important regulatory regions. HSEs, the TATA box, and the transcription start site, become hypersensitive at high GAF concentrations, as a result of increased GAF binding to the promoter at low affinity sites. This perturbation of DNA in these regions may have a role in providing access to key regulatory factors in a chromatin context, as hypersensitive promoter regions correlate with inducibility of many genes. Indeed, mutations in the larger GAF elements have been shown to affect *in vivo* occupancy of HSEs by HSF, the association of TBP with the TATA box, and the establishment of a promoter-paused RNA pol II molecule (11,12,28). Likewise, GAG trinucleotides may also be functionally aiding factor association by increasing DNA accessibility in these regions.

GAF has also been shown to interact *in vivo* with internal *hsp70* sequences after heat shock. No large binding sites for GAF appear in these sequences, and presumably heat shock induced association of GAF with these sequences is due to the ability of GAF to interact with lower affinity DNA binding sites during transcription (15). Trinucleotides (GAG) can be predicted to occur approximately once every 83 bp given base frequencies in the *Drosophila* genome, and could easily explain GAF crosslinking in the body of several genes. In fact, GAF elements of 3–5 bp occur, on average, once every 75 bp within *hsp70* (R.C.Wilkins and J.T.Lis, unpublished observation). If accessible during transcription, these sites might provide the means by which GAF associates with the body of active genes.

When high and low affinity sites are presented on the same, or separate fragments, there is little difference in the relative occupancy of the more distant low affinity sites. GAF association with GAG trinucleotides appears largely non-cooperative with occupancy of high affinity sites. While our footprinting pattern agrees with results of GAF binding to the hsp70 gene by Weber *et al*. (26), this lack of, or modest, cooperativity is in apparent contrast. The basis of this difference is not obvious, but could be due to differences in recombinant GAF preparations, or technical factors that are augmented by the steepness of the GAF gradient used in those experiments. We believe that any cooperative interactions between more distant sites must be weak *in vitro*, since measures of the affinity of GAF for promoter sites *in cis* by footprinting is in agreement with the relative affinities of GAF sites assayed *in trans*. Similar types of non-cooperative loading have been shown with zeste mediated activation *in vivo* (29).

When looking at the capacity of single synthetic sites to bind GAF, it is once again apparent that a trinucleotide alone is sufficient for binding. GAF specifically recognizes a GAG trinucleotide, as opposed to an AGA or TGT trinucleotide sequence, and can do so by means of a single zinc finger. Upon comparison of the relative affinities of GAF for various sites, there is only a modest increase in affinity of GAF for a pentamer consensus over a single trinucleotide. The affinity of GAF for larger repeats [$(GA)_4$] is, however, greater than both pentamer and trinucleotide sites. The strong binding of GAF to larger sites may reflect a cooperative binding of GAF when sites are immediately adjacent. This is reminiscent of HSF binding to its targets where strong cooperativity exists for adjacent sites, but little cooperativity is evidenced between separated sites (30,31). Clearly, while GAF does prefer a pentamer sequence, the difference in affinity between pentamer and trinucleotide sites is less than a factor of two.

Differences in GAF binding to high and low affinity sites as assayed by footprinting and immunoprecipitation indicate a <10-fold difference in affinity between fragments with multiple binding sites. More precise measures of GAF affinity to single sites shows an ~3.5-fold difference between elements of differing length. Binding to all sites assayed is in the nanomolar range. GAF is extremely abundant and sufficient to produce an approximate nuclear concentration of 150 nM. Though some percentage of GAF molecules may be multimeric (32), insoluble or not available for binding, this high nuclear concentration of GAF makes it possible for high affinity sites, and a substantial fraction of the low affinity sites, to be occupied *in vivo*.

Not all low affinity sites may be accessible for GAF binding due to competition with other DNA binding proteins, chromatin structure and nuclear organization. Such may be the case with the *hsp70* gene, where GAF is only bound at the promoter prior to heat shock. After heat shock, GAF gains access to internal sequences and is seen to crosslink throughout the body of the gene (15). It is therefore possible that GAF gains access to these weaker internal sites only during transcription when the gene is in an open conformation. Conversely, GAF may also be dependent upon other factors for this interaction. Gal4 protein has recently been shown to bind and activate transcription from its own high affinity sites within nucleosomal templates, but requires SWI/SNF for occupancy and activation through lower affinity sites (33). Similarly, GAF may not be able to access low affinity sites in the body of the gene without the help of additional factors, or may be impeded by the same barriers that generate paused polymerase on these genes. Clearly, internal sites are not available for GAF binding prior to induction, and their accessibility may be dependent upon cooperative interactions with nucleosome remodeling factors or the transcriptional machinery itself.

ACKNOWLEDGEMENTS

We thank the members of the Lis laboratory for critical comments on this manuscript. This work was supported by National Institutes of Health grant GM25232 to J.T.L., and by National Institutes of Health Predoctoral Training grant 5T32 GM07617 to R.C.W.

REFERENCES

1 Elgin, S. C. R. (1995) In Hames, B. D. and Glover, D. M. (Eds), *Frontiers in Molecular Biology* **8**, 1st edn., Oxford University Press, New York, NY.
2 Biggin, M. D. and Tijan, R. (1988) *Cell*, **53**, 699–711.
3 Soeller, W. C. and Kornberg, T. (1987) *Genes Dev.*, **2**, 68–81.
4 Wilkins, R. C. and Lis, J. T. (1997) *Nucleic Acids Res.*, **25**, 3963–3968.

5 Granok, H., Leibovitch, B. A., Shaffer, C. D. and Elgin, S. C. R. (1995) *Curr. Biol.*, **5**, 238–241.
6 Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R. and Kadonaga, J. T. (1991) *Science*, **251**, 643–649.
7 Kerrigan, L. A., Croston, G. E., Lira, L. M. and Kadonaga, J. T. (1990) *J. Biol. Chem.*, **266**, 574–582.
8 Lu, Q., Wallrath, L. L., Glaser, R. L., Lis, J. T. and Elgin, S. (1992) *J. Mol. Biol.*, **225**, 985–998.
9 Lu, Q., Wallrath, L. L., Granok, H. and Elgin, S. C. R. (1993) *Mol. Cell Biol.*, **13**, 2802–2814.
10 Glaser, R. L., Thomas, G. H., Siegfried, E., Elgin, S. C. R. and Lis, J. T. (1990) *J. Mol. Biol.*, **211**. 751–761.
11 Lee, H.-s., Kraus, K. W., Wolfner, M. F. and Lis, J. T. (1992) *Genes Dev.*, **6**, 284–295.
12 Shopland, L., Hirayoshi, K., Fernandes, M. and Lis, J. T. (1995) *Genes Dev.*, **9**, 2756–2769.
13 Tsukiyama, T., Becker, P. and Wu, C. (1994) *Nature*, **367**, 525–532.
14 Tsukiyama, T. and Wu, C. (1995) *Cell*, **83**, 1011–1020.
15 O'Brien, T., Wilkins, R. C., Giardina, C. and Lis, J. T. (1995) *Genes Dev.*, **9**, 1098–1110.
16 Nardelli, J., Gibson, T., Vesque, C. and Charnay, P. (1991) *Nature*, **349**, 175–178.
17 Greisman, H. A. and Pabo, C. O. (1997) *Nature*, **275**, 657–660.
18 Kim, J.-S., Kim, J., Cepek, K. L., Sharp, P. A. and Pabo, C. O. (1997) *Proc. Natl. Acad. Sci.* USA, **94**, 3616–3620.
19 Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M. and Clore, G. M. (1997) *Nature Struc. Biol.*, **4**, 122–132.
20 Soeller, W. C., Oh, C. E. and Kornberg, T. B. (1993) *Mol. Cell Biol.*, **13**, 7961–7970.
21 Giardina, C., Perez-Ribs, M. and Lis, J. T. (1992) *Genes Dev.*, **6**. 2190–2200.
22 Liu-Johnson, H.-N., Gartenberg, M. R. and Crothers, D. M. (1986) *Cell*, **47**, 995–1005.
23 Fernandes, M., Xiao, H. and Lis, J. T. (1994) *Nucleic Acids Res.*, **22**, 167–173.
24 Rougvie, A. E. and Lis, J. T. (1988) *Cell*, **54**, 795–804.
25 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.
26 Weber, J. A., Taxman, D. J., Lu, Q. and Gilmour, D. S. (1997) *Mol. Cell Biol.*, **17**, 3799–3808.
27 Dernburg, A. F., Broman, K. W., Fung, J. C., Marshall, W. F., Philips, J., Agard, D. A. and Sedat, J. W. (1996) *Cell*, **85**, 745–759.
28 Shopland, L. S. and Lis, J. T. (1996) *Chromosoma*, **105**, 158–171.
29 Laney, J. D. and Biggin, M. D. (1997) *Proc. Natl. Acad. Sci.* USA, **94**, 3602–3604.
30 Xiao, H., Perisic, O. and Lis, J. (1991) *Cell*, **64**, 585–593.
31 Fernandes, M., Xiao, H. and Lis, J. T. (1995) *Nucleic Acids Res.*, **23**, 4799–4804.
32 Benyajati, C., Mueller, L., Xu, N., Pappano, M., Gao, J., Mosammaparast, M., Conklin, D., Granok, H., Craig, C. and Elgin, S. C. R. (1997) *Nucleic Acids Res.*, **25**, 3345–3353.
33 Burns, L. G. and Peterson, C. L. (1997) *Mol. Cell Biol.*, **17**, 4811–4819.