Ga-ga over GAGA factor

Recent results suggest that the *Drosophila* transcriptional activator known as GAGA factor functions by influencing chromatin structure.

It is well established that chromatin structure is an active participant in the process of eukaryotic gene expression (reviewed in [1,2]). Recent results have helped to elucidate the roles of some of the structural components of chromatin that are involved in this process [3]. These results have, however, generally come from studies *in vitro*; our limited ability to identify and maintain organisms that are mutant for genes encoding critical chromatin proteins has hindered complementary studies *in vivo*. This situation should change following the recent, and very welcome, discovery [4] that a previously identified chromatin factor from the fruit fly *Drosophila melanogaster*, the so-called GAGA factor, is encoded by the *Trithorax-like* (*Trl*) gene.

The *Trl* gene is required for normal expression of several homeotic genes in *Drosophila*. *Trl* mutations additionally enhance 'position effect variegation' [4] — the clonally inherited pattern of gene inactivation observed when heterochromatic and euchromatic regions of the chromosome are juxtaposed by a chromosomal rearrangement. This begins to bear out the suggestion that position effect variegation in *Drosophila* provides an excellent genetic tool for identifying structural components of chromatin [5]. The discovery that *Trl* encodes GAGA factor further suggests that factors involved in regulating the chromatin structure of individual genes also participate in packaging large segments of the chromosome into active, euchromatic or inactive, heterochromatic domains.

GAGA factor was originally identified as a transcriptional activator that binds to several $(GA)_n$-rich sites in the *Ultrabithorax* (*Ubx*) promoter [6]. A similar factor was isolated by virtue of its ability to bind $(GA)_n$ sequences in the promoter regions of two heat-shock genes and the genes for histones H3 and H4 [7]. As the number of identified genes requiring $(GA)_n$ sites for efficient expression increased (Table 1), several laboratories attempted to isolate complementary (c)DNAs for GAGA factor. Soeller *et al.* [8] isolated cDNAs encoding a 519 amino-acid protein that binds to $(GA)_n$ sites in the *engrailed* and *hsp26* [9] promoters (Fig. 1a). From the open-reading frame identified in the cDNA sequences, GAGA factor is predicted to have a single 'zinc-finger' DNA-binding domain. The carboxyl terminus of the protein contains several polyglutamine stretches, a feature typical of many transcriptional activators.

Table 1. Putative *Drosophila* GAGA-factor target genes.

Gene type	Gene	Reference
Inducible/regulated	*hsp26*	[7]
	hsp70	[7]
	H3/H4	[7]
	Adh	[20]
	E74	[8]
Pattern formation	*engrailed*	[8]
	even-skipped	[21]
	fushi tarazu	[8]
	Krüppel	[16]
	Ultrabithorax	[6]
Housekeeping/ constitutive	*actin-5C*	[22]
	α1-tubulin	[23]

One of the more interesting features of GAGA factor is the 'BTB' or 'POZ' domain [10,11], a 120 amino-acid sequence at the amino terminus of GAGA factor that is named after the proteins in which similar domains were first found: bric-á-brac, tramtrack and Broad Complex; and the Poxvirus zinc finger (Fig. 1). This domain, which is highly conserved among a large number of eukaryotic and viral proteins (listed in [10,11]), was shown to mediate *in vitro* both homodimerization — of the protein tramtrack, for example — and heterodimerization — of GAGA factor and tramtrack. Interestingly, when present in proteins that also contain a DNA-binding motif, the BTB/POZ domain appears to inhibit DNA binding [11].

Multiple, developmentally regulated GAGA factor messenger (m)RNAs have recently been identified ([8] and C. Benyajati, personal communication). As the gene encoding GAGA factor is present as a single copy in *Drosophila* [4,8], the suggestion is that these different transcripts must be generated by alternative RNA splicing. Screening of cDNA expression libraries with $(GA)_n$ probes has identified only products of the known GAGA factor gene ([9] and C. Benyajati, personal communication). At least one of the identified cDNAs is predicted to encode a protein with a carboxyl terminus different to that of the GAGA factor described above (Fig. 1b; C. Benyajati, personal communication). Post-translational modifications ([8] and references therein) may further contribute to the large number of GAGA factor isoforms observed to date [6–8].

Most identified GAGA-factor-binding sites are located within 200 base pairs (bp) upstream of a gene's transcription start site. These sites vary considerably in length and sequence (Table 2). The longer sequences, such as that upstream of *hsp26*, probably represent multiple binding sites, but this has not been tested. Although most of these

sequences contain a GAGAGAG core, it is clear that notable exceptions to this consensus exist (such as in the *Krüppel* promoter). In order to provide a rigorous definition of GAGA-factor binding sites, experimental approaches such as binding-site selection and methylation interference assays will need to be employed.

The large number of genes identified to date that appear to be targets of GAGA factor is probably just the tip of the iceberg. Immunofluorescent staining of polytene chromosomes with antibodies against GAGA factor has shown [3] that the protein is distributed in many bands of varied intensity within the euchromatic part of the genome, with little or no staining in heterochromatic regions (Fig. 2). Raff *et al.* [12], however, have observed that GAGA factor is associated with specific regions of heterochromatin at all stages of the cell cycle in pre-blastoderm *Drosophila* embryos. This difference in staining patterns may be due to the under-representation of heterochromatic sequences in polytene chromosomes.

Centromere-associated staining is also seen in older (3–4 hour) embryos, when mitotic synchrony has been lost. The staining pattern on metaphase chromosomes parallels the distribution of the simple satellite repeats AAGAG and AAGAGAG ([12] and references therein), although binding of GAGA factor to these repeats has not been directly shown. The binding of GAGA factor to both gene-regulatory sites and simple-sequence repeats would be reminiscent of interactions between the yeast

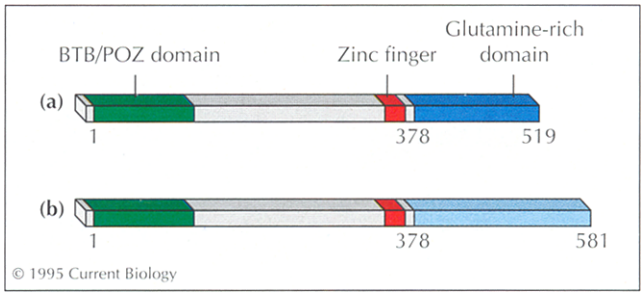

Fig. 1. GAGA factor protein domains. **(a)** the 519 amino-acid protein predicted by cDNAs isolated by Soeller *et al.* [8]. **(b)** the 581 amino-acid protein predicted by additional cDNAs isolated by C. Benyajati and colleagues (personal communication). The two proteins have identical sequences up to residue 378, at which point they diverge. The carboxy-terminal regions (light and dark blue) of both proteins are glutamine rich. Other features include the 'BTB' or 'POZ' domain (green, see text) and a single zinc finger (red), classified as being of the C2/H2 type from the characteristic cysteine and histidine pairs.

protein RAP1 (repressor/activator protein 1) and specific genes and telomeric repeat sequences (for review, see [13]). It is not known whether the association of GAGA factor with centromeres has structural consequences, or whether it is simply part of a mechanism that ensures the protein is evenly distributed when cells divide (for review, see [14]).

What does GAGA factor do in the nucleus? Recent studies suggest a role for GAGA factor in establishing

Table 2. Representative DNA sequences bound by GAGA factor.

Gene	Sequence
**Adh-1*	....GAAAGT GCACTGAGAA ACAAATAA.. ....
**Adh-2*	....GACAACT GCACTGAGAC AATAATGCAT GAC..... ..
**H3/H4-1*	....A AGAGAGAGAG AGAGAGAGAG A..... ....
**H3/H4-2*	....A CCCGAGAGAG TACG..... ....
**hsp26-1*	..... GAGTAGAGAG AGAGAAGAGA AGAGAGAGAA CGTGCACAGA GAGAAAA... ..
hsp26-2	....TG AGTGAGAGAG CC..... ....
†hsp26-3	....T CTAGAAAGAG CGCAAAAGAA A..... ..
hsp70-1	.... .AAAAGAGAG AATAA..... ....
hsp70-2	....G AGTGAGAGAG CATTAGTGCA GAGAGGGAGA GA
hsp70-3	....G TTCGAGAGAG CGCGC..... ....
hsp70-4	....AAGAGCG CCGG..... ....
hsp70-5	....GAGAGAG C..... ....
Kr-1	....CGTGTG AGCGGGAGAG CCAATT.... ....
Kr-2	.... .GAGTGAGAA TCT..... ....
Kr-3	....CG AGACAGAGCG T..... ....
Cons27/48	---------- ---------- ---------- ---GAGAGAG ---------- ---------- --

The sequences shown are those protected by GAGA factor in *in vitro* DNase I footprinting assays. Separate sites within a single gene are indicated by consecutive numbers. A total of 48 sequences, including 33 not shown here, were aligned (using the GCG package, University of Wisconsin) to facilitate visual inspection, and a possible consensus sequence (bases present in at least 27 of the 48 binding sites) was determined (Cons27/48). * Indicates sites for which protection against nuclease digestion has been observed *in vivo*. † Indicates a site from an *hsp26/lacZ* transgene [9]; the native sequence contains an additional 21 base pairs.

Fig. 2. Distribution of GAGA factor along *Drosophila* larval polytene chromosomes. Polytene chromosomes were stained with antibodies against GAGA factor, followed by a fluorescein-labelled secondary antibody. Note the large number of stained euchromatic sites and the unstained heterochromatic chromocenter (arrow).

and/or maintaining DNase I hypersensitive sites, the nucleosome-free regions associated with important gene regulatory sequences ([9] and references therein). Substitution or deletion of the $(GA)_n$ sequences upstream of the *hsp26* transcription start site results in loss of the ordered chromatin structure *in vivo*, accompanied by a proportional loss of heat-shock inducible expression. The sequence alterations that result in loss of chromatin structure *in vivo* eliminate GAGA factor binding *in vitro* [9]. These results imply that GAGA factor, by binding to the $(GA)_n$ elements, directs the formation of DNase I hypersensitive sites.

Another role for GAGA factor might be in recruiting RNA polymerase to the promoter. Lis and co-workers used nuclear 'run-on' experiments — which assay the level of RNA polymerase loading on a gene — to show that RNA polymerase II is transcriptionally paused at both the *hsp70* and *hsp26* promoters before heat shock (for review, see [15]). GAGA-factor binding sites are required to generate the paused polymerase configuration at the *hsp70* promoter; gene activity is substantially reduced in the absence of this paused transcription complex. It is not known whether RNA polymerase is recruited to the promoter by GAGA factor, or merely binds to a promoter that has been made nucleosome-free by a mechanism dependent on this protein.

Several studies *in vitro* suggested that GAGA factor itself might be sufficient to model or remodel chromatin structure. Kadonaga and colleagues showed that GAGA factor counteracts the effects of a transcriptional repressor, histone H1, *in vitro* ([16] and references therein). It appears that the high levels of *in vitro* activation achieved with GAGA factor in reconstitution/transcription assays are due primarily to this antirepression activity, rather than 'true activation' (transcriptional activation in the absence of histones) [16]. Other investigators have observed transcriptional activation by GAGA factor in transient transfection assays ([8] and C. Benyajati, personal communication); however, little is known about the chromatin structure of the reporter plasmids in the transfected cells.

More recently, Wu and colleagues [3] have used a *Drosophila* chromatin assembly system to reconstitute nucleosomes over the *hsp70* promoter. They found that addition of GAGA factor, either before or after nucleosome assembly, leads to a local disruption of chromatin structure and the formation of DNase I hypersensitive sites. This disruption is ATP-dependent and appears to be confined to the region immediately adjacent to the GAGA-factor binding sites. It is not known whether GAGA factor itself disrupts chromatin, or whether it directs a disrupting activity present in the assembly extract. Future experiments using this system will undoubtedly shed light on the regions of the GAGA factor molecule responsible for this effect, and on the activities responsible for chromatin remodelling.

Homozygous lethal mutations in the gene encoding GAGA factor have now been described by Karch and colleagues [4]. The gene, *Trl*, was originally identified as one of the *Trithorax* group of genes; these genes work in opposition to the *Polycomb* group, which maintain homeotic loci such as *Ubx* or *Abdominal-B* (*Abd-B*) in a repressed state [17]. Such repression has been inferred to involve chromatin packaging [17,18]. *Trl* mutant alleles enhance the enlarged haltere phenotype of *Ubx* mutants, and some alleles lead to homeotic transformations of certain abdominal segments in a manner resembling the effects of *Abd-B* mutations. Thus, a loss of *Trl* function apparently results in lower levels of expression of these homeotic genes in tissues where they are normally active.

Significantly, Karch and colleagues [4] have found that all of these *Trl* mutant alleles act as dominant enhancers of position effect variegation. As mentioned above, position effect variegation refers to the clonally inherited pattern of gene inactivation observed when regions of heterochromatin and euchromatin are juxtaposed by a chromosomal rearrangement — in this particular study, the inactivation of *white* caused by the inversion $In(1)w^{m4}$. Because this inactivation appears to be due to the position of the gene in the genome, rather than to any alteration of the gene's sequence, it is thought to reflect the packaging of DNA into inactive heterochromatin or active euchromatin. Thus the identification of *Trl* mutant alleles as enhancers of position effect variegation points to a broad role for GAGA factor in maintaining an active, euchromatic structure. Interestingly, another enhancer of position-effect variegation, *E(var)3-93D*, encodes a protein that contains a BTB/POZ domain similar to that found in GAGA factor [19].

Experiments *in vitro* to identify regions of GAGA factor required for nucleosome disruption, DNA binding and interactions with other proteins can now be complemented by genetic experiments to assess the function of

these regions *in vivo*. Although modifications in such an important protein might be lethal when homozygous, the assay of enhancement or suppression of position effect variegation will allow the effects of such mutations to be discerned in the heterozygous state (see [5]). The identification of other factors with which GAGA factor interacts will undoubtedly be greatly facilitated by using the dual approaches of genetics and biochemistry. GAGA factor is a remarkable protein in its ability to drive the formation of specific chromatin structures required for gene activation. Understanding how GAGA factor — or related proteins with similar activities — accomplishes this will provide us with new insights into the mechanics of both constitutive and inducible gene expression.

Acknowledgements: We thank other members of our laboratory for their comments and suggestions, and Carolyn Craig for Figure 2. We apologize to the many authors whose work could not be cited due to space limitations. Work from the Elgin laboratory reported here was supported by NIH grants GM 31532 and HD 23844 to S.C.R.E.

References

1. Felsenfeld G: **Chromatin as an essential part of the transcriptional mechanism.** *Nature* 1992, **355**:219–224.
2. Wallrath LL, Lu Q, Granok H, Elgin SCR: **Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures.** *Bioessays* 1994, **16**:165–170.
3. Tsukiyama T, Becker PB, Wu C: **ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor.** *Nature* 1994, **367**:525–532.
4. Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F: **The *Trithorax-like* gene encodes the *Drosophila* GAGA factor.** *Nature* 1994, **371**:806–808.
5. Reuter G, Spierer P: **Position effect variegation and chromatin proteins.** *Bioessays* 1992, **14**:605–612.
6. Biggin MD, Tjian R: **Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts.** *Cell* 1988, **53**:699–711.
7. Gilmour DS, Thomas GH, Elgin SCR: ***Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters.** *Science* 1989, **245**:1487–1490.
8. Soeller WC, Oh CE, Kornberg TB: **Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor.** *Mol Cell Biol* 1993, **13**:7961–7970.
9. Lu Q, Wallrath LL, Granok H, Elgin SCR: **$(CT)_n$.$(GA)_n$ repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the *Drosophila hsp26* gene.** *Mol Cell Biol* 1993, **13**:2802–2814.
10. Zollman S, Godt D, Prive GG, Couderc JL, Laski FA: **The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*.** *Proc Natl Acad Sci USA* 1994, **91**:10717–10721.
11. Bardwell VJ, Treisman R: **The POZ domain — a conserved protein–protein interaction motif.** *Genes Dev* 1994, **8**:1664–1677.
12. Raff JW, Kellum R, Alberts B: **The *Drosophila* GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle.** *EMBO J* 1995, **13**:5977–5983.
13. Shore D: **RAP1: a protean regulator in yeast.** *Trends Genet* 1994, **10**:408–412.
14. Earnshaw WC, Mackay AM: **Role of nonhistone proteins in the chromosomal events of mitosis.** *FASEB J* 1994, **8**:947–956.
15. Lis J, Wu C: **Protein traffic on the heat shock promoter: parking, stalling, and trucking along.** *Cell* 1993, **74**:1–4.
16. Croston GE, Kerrigan LA, Lira LM, Marshak DR, Kadonaga JT: **Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription.** *Science* 1991, **251**:643–649.
17. Paro R: **Mechanisms of heritable gene repression during development of *Drosophila*.** *Curr Opin Cell Biol* 1993, **5**:999–1005.
18. Shaffer CD, Wallrath LL, Elgin SCR: **Regulating genes by packaging domains: bits of heterochromatin in euchromatin?** *Trends Genet* 1993, **9**:35–37.
19. Dorn R, Krauss V, Reuter G, Saumweber H: **The enhancer of position-effect variegation of *Drosophila*, *E(var)3-93D*, codes for a chromatin protein containing a conserved domain common to several transcriptional regulators.** *Proc Natl Acad Sci USA* 1993, **90**:11376–11380.
20. Benyajati C, Ewel A, McKeon J, Chovav M, Juan E: **Characterization and purification of *Adh* distal promoter factor 2, Adf-2, a cell-specific and promoter-specific repressor in *Drosophila*.** *Nucleic Acids Res* 1992, **20**:4481–4489.
21. Read D, Nishigaki T, Manley JL: **The *Drosophila even-skipped* promoter is transcribed in a stage-specific manner *in vitro* and contains multiple, overlapping factor-binding sites.** *Mol Cell Biol* 1990, **10**:4334–4344.
22. Chung YT, Keller EB: **Positive and negative regulatory elements mediating transcription from the *Drosophila melanogaster actin 5C* distal promoter.** *Mol Cell Biol* 1990, **10**:6172–6180.
23. O'Donnell KH, Wensink PC: **GAGA factor and TBF1 bind DNA elements that direct ubiquitous transcription of the *Drosophila alpha-1-tubulin* gene.** *Nucleic Acids Res* 1994, **22**:4712–4718.

H. Granok, B.A. Leibovitch, C.D. Shaffer and S.C.R. Elgin. Department of Biology, Box 1229, Washington University, One Brookings Drive, St Louis, Missouri 63130, USA.