GAGA factor down regulates its own promoter

Ana Kosoy, Sara Pagans, Mª Lluïsa Espinás, Ferran Azorín, and Jordi Bernués

Departament de Biologia Molecular i Cel·lular
Institut de Biologia Molecular de Barcelona
Consell Superior d'Investigacions Científiques
Jordi Girona, 18-26
08034 Barcelona, Spain

To whom correspondence should be addressed:
Jordi Bernués
Dept. Biologia Molecular i Cel·lular,
Institut de Biologia Molecular de Barcelona, CSIC
Jordi Girona, 18-26
08034 Barcelona, Spain
Tel: +34 934 006 177
Fax: +34 932 095 904
e-mail: jbmbmc@ibmb.csic.es

Running title: GAGA factor represses Trl

Keywords: activation, GAGA, repression, RNAi, transcription, Trithorax-like promoter
Summary

GAGA factor is involved in many nuclear transactions, notably in transcription acting as an activator in *Drosophila*. The genomic region corresponding to *Trl* promoter has been obtained and a minimal version of a fully active *Trl* promoter has been defined using transient transfection assays in S2 cells. DNase I footprinting analysis has shown that this region contains multiple GAGA binding sites suggesting a potential regulatory role of GAGA on its own promoter. The study has shown that GAGA down regulates *Trl* expression. The repression does not depend on the GAGA isoform but binding to DNA is absolutely required. A fragment of the *Trl* promoter can mediate repression to an heterologous promoter only upon GAGA overexpression in transiently transfected S2 cells. Chromatin immunoprecipitation analysis of S2 cells confirmed that GAGA factors are bound to the *Trl* promoter over a region of 1.4 kbp. Using a dsRNA interference approach, we show that endogenous GAGA factors are in fact limiting *Trl* expression in S2 cells. Our results open the possibility to observe similar GAGA repressive effects on other promoters.
Introduction

GAGA is a *Drosophila* nuclear factor that is encoded by the Trithorax-like gene (*Trl*) and presents maternal effects (1). At least two isoforms, GAGA519 and GAGA581, are known to be produced by alternative splicing (2) and both the sequence of the isoforms and their splicing patterns are highly conserved between *D. melanogaster* and the distant species *D. virilis* (3). The GAGA factors present an overall modular structure that is formed by a N-terminal BTB/POZ domain involved in protein oligomerization (4,5), a DNA binding domain composed of a single Zn-finger and three adjacent basic regions that confers sequence specific DNA binding (6-8), and a C-terminal domain (Q-domain) that in both isoforms is glutamine-rich, although amino acid sequence is different, and is the activation domain ((9-11) and 2 ). There is some evidence that Q-domain may also be involved in GAGA aggregation (12,13). Between the POZ/BTB domain and the DNA-binding domain there is a fourth domain, that here will be referred as X-domain, about 200 amino acids long of unknown structure and properties. To date, both isoforms have been found to be rather equivalent in general terms and only some distinct although largely overlapping functions have been noticed in transgenic flies (11).

GAGA binds to DNA sequences with a consensus (GAGAG) (7) but neither this is a strict consensus nor a single site is enough for efficient GAGA binding. In fact, a GAG or CTC trinucleotide sequence has been described to be sufficient for specific binding (8). However, as pointed by others GAGA binding sites on natural promoters show a clear tendency to cluster and give rise to strong composite GAGA binding sites (14). This is in agreement with the finding that GAGA is oligomeric and can interact with specific sequences that are variably spaced by other irrelevant DNA sequences which are neither recognized nor bound (4,5). The need for the clustering of GAGA binding sites apparently arises from the fact that GAGA activation is very weak (if not negligible) when only single or double sites are present (4).

Since its discovery, GAGA was scored as a transcription activation factor because upon binding to (GA)_n-rich sequences it stimulated transcription of genes like Ultrabithorax (*Ubx*) and engrailed (*en*) (15,16). This role has been frequently reported for many other promoters either in vitro and/or in vivo including actin5C, krüppel, fushi-tarazu (*ftz*), etc. (recapitulated in (17)). Nevertheless, several early observations suggested that GAGA was a complex nuclear factor. For instance, unexpectedly during mitosis GAGA remains bound to chromatin while transcription factors are usually displaced (18,19). Also at
metaphase, GAGA is redirected from its interphasic locations on the promoters (20-22) to heterochromatic GA-rich regions located in the vicinity of Drosophila centromeres (18,19). These features appear to be common to both GAGA isoforms (2,11). This mitotic location might explain the phenotypes observed for the hypormorphic mutant \( Trl^{13C} \) that include serious defects on the synchrony in cleavage cycles, failure in chromosome condensation, abnormal chromosome segregation and chromosome fragmentation (23). Other explanations invoke the evidence that GAGA co-operates with chromatin remodeling machines like NURF to open several promoters (22,24). Also GAGA and ISWI, the catalytic subunit shared by several chromatin remodeling complexes, present some limited co-localization at some loci on polytenic chromosomes whereas they do not co-localize at all on mitotic chromosomes (25). Recently, a direct interaction has been observed between GAGA and the largest NURF subunit, NURF301. The interaction involves two regions in NURF301 and a GAGA region around the DNA binding domain. On the other hand, the chromatin remodelling activity and the stimulatory activity of GAGA have been shown to be independent in transgenic flies (26).

The existence of GAGA binding sites on several polycomb responsive elements (PRE)\(^1\) and its co-localization with some proteins of the Polycomb group also indicates a connection of GAGA with gene silencing (27,28). The interaction of GAGA with SAP18, a Sin3-associated polypeptide, and the presence of GAGA in a complex containing the RPD3 deacetylase and the polycomb group factors Polyhomeotic and Posterior Sex Combs indirectly suggested a connection with repression too (21,29). A functional requirement for both GAGA and pleiohomeotic has been established to account for the silencing activity of the iab-7 PRE and the MCP silencer of Abdominal-B (30,31).

Here, we have analyzed the \( Trl \) promoter and provide direct evidence that both GAGA isoforms can mediate repression of its own promoter. Repression is promoter-specific and GAGA DNA binding activity is required while transactivation and oligomerization domains are dispensable.

**Experimental Procedures**

*DNA constructs*

The sequences corresponding to the 5' upstream region to the \( Trl \) coding sequence, and released in Flybase before the whole Drosophila genome was available, were used to isolate the sequences putatively corresponding to the
Trl promoter. A ~4.3 kbp fragment was PCR amplified from P1 clone ds01921 (kindly provided by the European Drosophila Genome Project) using 5'-TAC ATG GAT AAG ATT CTG ACG G-3' and 5'-GGG AGG CGG CGC ACA GG-3' as primers and cloned into pGEM-T (Promega). This fragment comprised about 3.5 kbp of putative promoter region, the three 5’ end sequences of the cDNAs reported before (2) and the first 168 bp of GAGA coding sequence. Automated sequencing exactly confirmed the genomic sequence deposited later on by Celera Genomics. For transfection assays subclones were generated using PCR in a way that coding sequences were omitted while leaving intact a long 5’ UTR region. Promoter deletions were introduced either by restriction at unique sites whenever possible or by PCR with specific primers. All of them were inserted between MluI and HindIII sites in pGL3 vector (Promega).

The eve-luciferase reporter was obtained by inserting a 1.8 kbp fragment of the eve promoter that directs the embryonic expression of the stripe 2 in the embryo (from position -1759 to +102, kindly provided by Dr. M. Levine, UCSF) into pGL3 vector (Promega).

Constructs for expressing GAGA isoforms and mutants in S2 cells were all prepared using Act5PPA vector. Constructs for GAGA519, GAGAΔQ, GAGAΔQ-VP16 and GAL4BD have already been described (9). Expressing construct for GAGA581 was generated by PCR cloning the C-terminus of GAGA581 and replacing the C-terminal region of GAGA519 construct. Expression construct for GAGAΔPOZ was generated by subcloning the previously prepared ΔPOZ122 construct into Act5PPA vector (4). The construct expressing the mutated DNA binding domain (GAGA519H361P) was generated by PCR across the DNA binding domain using appropriate primers that introduced a his361→pro change at the Zn-finger. This mutant was initially prepared in a bacterial expression vector (pET14b) and then subcloned into Act5PPA for expression in S2 cells.

Constructs containing the CMV promoter and a luciferase reporter gene were fused to fragments FI, FII, FII+FIII, FIII, FIV and FV derived from the Trl promoter using either a MluI site located upstream the CMV promoter for the FI to FIII fragments and a HindⅢ site located downstream of the CMV promoter for fragments FIV and FV in an attempt to mimic as much as possible the relative location of the Trl-derived fragments with respect to the transcription start sites.

**Transient cell transfections**

S2 cells were grown and transfected as described (9). Each transfection included 3 µg of CMV-βgal, variable amounts of reporter constructs fused to luciferase gene and expression vectors driven by the Actin 5C promoter of Drosophila, and 7 µg of pGL3 vector (Promega). Final amount of DNA was
adjusted to 20 µg by the addition of Act5PPA empty vector. After 48 h of incubation with the DNA, cell lysates were prepared according to manufacturer’s instructions. Luciferase and β-Galactosidase activities to correct for transfection efficiency were determined using commercial kits (Promega and Roche, respectively). Each point is the average of three to five independent experiments, each performed in duplicate.

**DNase I footprinting assays**

The region of the **Trl** promoter to be studied was divided into five different regions and subcloned into pBS(-) (Stratagene). These five regions contained sequences between positions -679/-462, -462/-204, -204/-49, -49/+347, and +347/+737 respectively, according to the numbering indicated in Fig.3A. These fragments were end-labeled and were purified on native polyacrylamide gels following standard protocols. Recombinant GAGA factor used in these experiments corresponded to GAGA519 and was expressed and purified in *E. coli* BL21(DE3) as described (32). DNase I footprinting reactions were carried out as described (33).

**RNase Protection Assays**

RNase protection analysis was performed using the Direct Protect Lysate RPA kit following manufacturer’s instructions (Ambion). Total RNA was isolated from 4x10^7 *Drosophila* S2 cells and for analysis of **Trl** transcripts 25-65 µg of total RNA was mixed with 1x10^4 cpm of ^32^P-labeled antisense riboprobes. Protected RNAs were analyzed on denaturing 5% polyacrylamide gels containing 8 M urea. Gels were dried, exposed at -80°C for 24 h with intensifier screens and developed. Uniformly radiolabeled RNA probes R1 (from position -49 to +347) and R2 (from position +347 to +737) were generated by *in vitro* transcription using either T3 (for R1) or T7 (for R2) RNA polymerases from the respective plasmids bearing these sequences in pBS(-). Riboprobes were purified on polyacrylamide gels.

**Chromatin Immunoprecipitation analysis (ChIP)**

*Drosophila* Schneider S2 cells were cross-linked and chromatin immunoprecipitated as described (34) with the modification in the PCR amplification protocol subsequently described (35). Polyclonal antibody raised in rabbits against bacterial expressed GAGA519 isoform was affinity purified using standard protocols. DNA amplified from immunoprecipitations was analyzed by slot-blot hybridization. Radiolabeled DNA probes corresponding to
regions -679/-462 (FI), -462/-204 (FII), -204/-49 (FIII), -49/+347 (FIV), and +347/+737 (FV) were prepared using the Ready-to-Go labeling kit (Amersham Biosciences). Hybridizations were carried out following standard protocols. Blots were exposed at -80°C, and films were quantified using a Molecular Dynamics laser microdensitometer. All the results were checked to be in the linear response range as indicated by comparison with the corresponding sets of dilution input standards (not shown).

**RNA interference experiments**

In order to generate the dsRNA an *EcoRI-BamHI* fragment of GAGA (718 bp long) encoding a protein sequence common to both isoforms of GAGA and including most of the POZ/BTB domain and part of the adjacent sequence was subcloned into pBS(-). The RNAs were prepared by *in vitro* transcription as described above and resuspended in water. To prepare dsRNAs equimolar amounts of sense and antisense RNAs were mixed, heated for 5 minutes at 90°C, cooled down slowly to room temperature and stored at -20°C until use. For transient transfection assays, dsRNA (0.5 µg to 5 µg) was added to the transfection mix and proceeded as described. Luciferase and β-galactosidase activities were assayed after 48 or 72 h as described above. A control dsRNA for the β-galactosidase mRNA was similarly prepared from a lacZ*SstI-PvuII* fragment about 0.7 kbp long cloned into pBS(-).

**Results**

**Isolation and analysis of the *Trl* promoter**

A 4.2 kbp DNA fragment was obtained by PCR amplification from P1 clone ds01921 (obtained from the European *Drosophila* Genome Project) using primers specific for the *Trl* gene. The complete sequence was fully consistent with the sequence deposited by Celera Genomics in the Berkeley *Drosophila* Genome Project Database. The data available suggested the existence of at least three different transcription start sites according to the different 5′ ends of the cDNAs isolated so far (2). Here, we assigned position +1 to the hypothetical start site the most upstream from the coding sequence. Our clone obtained contained about 0.73 kbp of 5′ UTR sequence of the *Trl* gene and ~3.4 kbp of upstream sequence.

The study of the *Trl* promoter elements included transient transfection in Schneider’s S2 cell line. As a first step, *Trl* transcription start sites used in S2 cells were mapped. RNase protection assays (RPA) with total RNA from S2
cells revealed at least three start sites (Fig.1B). From its position with respect to the coding sequence they were named, distal, intermediate and proximal. These three sites reasonably correspond with the three 5' cDNA ends already described (2). Other start sites may exist because an intense band appeared in the RPA analysis a few nucleotides upstream the position of the proximal start site (marked *1, Fig. 1A left panel, and B). Also, another band appeared some 50 nt upstream of the distal site (marked *2, Fig. 1A right panel, and B) which is clearly different from the undigested R1 probe. From the previously reported results, the proximal start site accounts for most of the GAGA519 isoform and the distal for the GAGA581 isoform in embryos (2). In S2 cells and since the RPA technique is quantitative, we observed that the proximal start site is also the most frequently used as indicated by the high intensity of the protected band.

The Trl promoter fragment described above was linked to a luciferase reporter gene and studied in S2 cells. Transcriptional activity from this construct reached values above 2600-fold with respect to empty pGL3 vector and was taken as 100% of Trl promoter activity in our analysis (not shown). A deletion analysis of the Trl promoter showed that near-full activity could be observed in serial 5’ deletions from position -3470 down to position -345 (Fig.2). Deletion to position -270 resulted in a drop of about 50% of the promoter activity. Deletion to position -204 could only support about 20% of total activity and activity was completely lost by deletion to position -49. Deletion of 400 bp at the 3’ end of the fragment resulted in about 50% of total activity. Since this deletion removed the proximal start site, it is reasonable to assume that the two other sites were functional and accounting for reasonable levels of mRNAs. We concluded that the Trl promoter in S2 cells is small and extends to about position -345.

Computer analysis of the promoter sequence lighted up several potential GAGA factor sites and some potential binding sites for other regulatory proteins (not shown). Since GAGA binds to sequences showing a considerable degree of heterogeneity and a good binding site has shown to be of a composite nature rather than of a single linear sequence (4,5,8) a DNase I footprinting analysis using recombinant GAGA519 factor was performed. Since full promoter activity in S2 cells was confined to position -345 we limited our assay to a slightly larger fragment (up to position -680) and included the 737 bp of the 5’UTR. Because the region to be analyzed was rather long (~1.4 kbp), it was subdivided by convenience into five fragments (denoted FI to FV, Fig.3A) covering the entire length of the region studied. Fig.3B shows that a large number of GAGA sites were present on the Trl promoter (see Fig.3A for a
general scheme and summary of results). Specially rich in GAGA binding sites were regions FII, FIV and FV with 5, 8, and 8 sites, whereas regions FI and FIII showed only 2 and 3 sites, respectively. Relative affinities of the binding sites were not identical and some showed a high affinity (e.g., FIII no. 1, FIV nos. 6 and 7, FV nos. 6, 7 and 8) whereas others were of low affinity (e.g., FII no.1, FIV nos. 1 and 2, FV nos. 1 and 3, and results not shown). These differences suggested that the relative occupancies of the sites could be different in vivo. Some hypersensitive sites that might suggest a stressed DNA conformation upon GAGA binding were also noted (indicated by *).

GAGA represses the Trl promoter

GAGA factor has always been regarded as an activator because in the fly it stimulated the transcriptional activity of genes under its control (e.g. ftz, en, Ubx, etc.) (1,11,17,23). GAGA also stimulated transcription of many reporter genes when assayed using transient transfection in Drosophila cell lines (2,9,14) and in transgenic flies (26). From our results, an appealing possibility was that GAGA could regulate the expression of its own promoter. This point was studied by performing transient transfection experiments in which GAGA factors were overexpressed and their activity on the Trl promoter analyzed. Unexpectedly, overexpression of either GAGA519 or GAGA581 resulted in a strong repression of the Trl promoter. Repression was efficient and dose-dependent (Fig.4A), reaching a minimal level of activity about 15% with respect to controls without GAGA overexpression. Overexpression of larger amounts of GAGA factor did not result in further repression (results not shown). On the contrary, expression of increasing amounts of GAGA519 stimulated transcription of the eve stripe 2 promoter up to ~15-fold in a dose-dependent manner (Fig.4B). Overexpression of an unrelated factor -the yeast GAL4BD- showed no effect on the activity of both promoters at any dose (Fig.5 and results not shown).

RNA interference assays were performed to assess that stimulation of eve and repression of Trl were due to GAGA overexpression. Both repression of Trl (light grey columns) as well as stimulation of eve (dark grey columns) were efficiently abolished by co-transfection of constructs and specific RNAiGAGA at low doses (Fig.4C). At the highest dose, Trl expression was stimulated ~2-fold above the starting level clearly suggesting a GAGA-mediated down regulation of Trl promoter even in the absence of expressed GAGA (Fig.4C, compare columns at 0 and 5 µg RNAiGAGA and see also Fig.7B). As expected, β-
galactosidase activity was not affected by co-transfection of RNAiGAGA, and *vice versa*, co-transfection of the a RNAiβ-gal did not affect luciferase levels (results not shown).

**While BTB/POZ domain and Q-domain are dispensable for GAGA-mediated repression of *Trl*, DNA-binding domain is required.**

In order to define the domain(s) of GAGA required for the repression of the *Trl* promoter, the two GAGA isoforms, GAGA519 and GAGA581, which are identical in sequence except for the glutamine-rich C-terminal domain (Q-domain) were initially assayed. Both isoforms similarly repressed *Trl*, much the same as GAGAΔQ did, confirming that Q-domains were not required for repression (Fig.5). On the contrary, both isoforms stimulated transcription of *eve stripe 2* promoter (Fig.4B and results not shown). POZ/BTB domain was not required either because its deletion did not prevent repression (Fig.5, GAGA519ΔPOZ). However, a single point mutation at his361 (GAGA519(H361P)) which disrupted the unique zinc-finger domain and rendered GAGA completely unable to recognize GAGA sites (as shown by DNase I footprinting experiments, results not shown) significantly abolished repression of *Trl*. Replacement in GAGA of the Q domain by the VP16 activation domain (GAGAΔQVP16) reverted the situation and could even stimulate *Trl* transcription ~2.5-fold. Expression of an irrelevant factor (GAL4BD) had no effect at all on *Trl* expression. In all cases, analysis by western blot analysis indicated that expression of all the constructs resulted in proteins of the expected sizes that were produced in similar amounts (results not shown).

**A region of the *Trl* promoter is required to mediate GAGA repression**

Because the DNA binding activity of GAGA was required, it was of interest to identify the DNA sequence of the *Trl* promoter that mediated repression. As shown before, there were so many GAGA binding sites on this promoter that assignment of the sites involved in repression and discrimination of those that might not take part became rather complex. Even more because GAGA does not simply operate on a single site basis but binds clusters of relatively close binding sites (2,4,5,14,15,17). Therefore, we decided to identify regions rather than sites that could confer GAGA-mediated repression. The region required for GAGA to repress *Trl* was delimited, using the promoter deletion constructs described before, to position -345 in S2 cells since
deletions from position -3470 showed no effect on repression (Fig. 6A). Deletions to position -270 reduced the repression levels slightly. Further deletions could not be assayed because of the very low transcriptional activity (see Fig. 2). These results indicated that the promoter element involved in GAGA-mediated repression had to be located between positions -345 and +737. To better define and assay which region was required we considered that repression might be transferred to heterologous promoters. To perform such an assay, a prerequisite was that the recipient promoter had to be insensitive to GAGA overexpression. CMV was found to be insensitive (Fig. 6C) and then, fragments FI to FV which spanned from -679 to +737 positions were inserted into the CMV-luciferase reporter. Fragments FI to FIII were inserted upstream of the core CMV promoter whereas fragments FIV and FV were inserted downstream in an attempt to mimic as much as possible the relative situation of the Trl promoter regions with respect to the transcription start site of the core CMV promoter. Transfection of these constructs in S2 cells resulted in transcription levels of the luciferase reporter gene higher than observed for the parental CMV construction, suggesting they contained some stimulatory sequence elements (Fig. 6B). The sole exception was FI (spanning positions -679 to -462) that did not show any effect likely because positions upstream of -345 were not required for full activity of the Trl promoter (Fig. 2). Upon GAGA overexpression, fragments FII and FIII could mediate partial repression of CMV promoter with respect to their corresponding controls (Fig. 6C, columns FII-CMV and FIII-CMV). This repression, however, still left about 70% of activity while the values observed with the complete Trl promoter left only 15-20%. The very weak repression activity observed with fragment FI was not considered significant. On the other hand, GAGA overexpression clearly had no effect on CMV-FIV construct and transcription from CMV-FV construct was stimulated ~4-fold (Fig. 6C, columns CMV-FIV and CM-FV). Since these results indicated that a region (perhaps two) of the Trl promoter could direct partial GAGA-mediated repression when transported to an unrelated heterologous promoter and the Trl regions were defined and inserted in the CMV promoter for convenience, it was possible that the partial repression observed could be due to disruption of a genuine repressive element of the Trl promoter. Thus, insertion of a fragment encompassing the complete FII and FIII regions upstream of the CMV promoter resulted in higher levels of repression only upon GAGA overexpression (50% of activity left, Fig. 6C, FII+FIII-CMV). In the absence of GAGA overexpression, insertion of this region resulted in a weak stimulation of CMV transcription as observed for the individual fragments (Fig. 6B, FII+FIII-
In S2 cells endogenous GAGA is bound to Trl promoter and down regulates its transcription

In order to get some insight into the state of the genomic Trl promoter in intact cells, chromatin immunoprecipitation analysis were performed. Using as probes the same fragments described for the footprinting experiments (FI to FV, Fig.3), the presence of GAGA across the minimal Trl promoter region presenting full activity was revealed (Fig.7A). GAGA could be detected bound to all of them although a clear accumulation of GAGA factors on fragments FII and FIV was detected. Accumulation of GAGA factors was specially intense at region FII (100-fold), one of the regions able to mediate repression to a heterologous CMV promoter. Fragment FIV where binding of GAGA was shown to have no effect in the same assay was highly enriched in GAGA factors as well (65-fold). Enrichment of GAGA was clearly similar but lower for the other three fragments (FI, FIII and FV, 8-fold, 14-fold and 10-fold, respectively) (Fig.7A, see lower panel for quantification). Overall the data supported the results obtained by DNase I footprinting in vitro and suggested that GAGA could have a regulatory role on its own promoter. Accordingly, the repressive effect on the Trl promoter could also be observed in S2 cells by depletion of the endogenous GAGA content using RNA interference assays. Co-transfection of the Trl luciferase reporter construct and increasing amounts of RNAi directed against GAGA mRNA (RNAiGAGA) resulted in a dose-dependent stimulation of transcription that reached a ~4-fold increase at the highest dose tested (Fig.7B). Since in this experiment GAGA was not overexpressed but RNAi was knocking down the endogenous content of GAGA factors we can conclude that normal physiological levels of GAGA factors were limiting Trl transcription in S2 cells so that a reduction in the GAGA content resulted in a derepression of Trl promoter.

Discussion

A functional promoter of the Trl gene has been isolated and the regions relevant for expression in S2 cells have been defined. A region of about 0.35 kbp upstream of the distal transcription start site was sufficient to support full transcription in S2 cells. Transcription dropped to a minimum when only 0.2 kbp were left and completely disappeared when only 49 bp were left (Fig.2). Preliminary results in transgenic flies confirm these results and indicate that
both the -3470 to +737 promoter and also the minimal -345 to +737 promoter appear to be functional and direct similar embryonic expression of a reporter gene while the -49 to +737 can not (results not shown).

Three transcription start sites of the Trl gene have been mapped in S2 cells that closely correlate with the start sites of the three different cDNA clones from a Drosophila ovarian cDNA library previously reported (2). In addition, the possible existence of other two start sites has been described. Notably, the most distal site observed (Fig.1, marked *2) matches reasonably well with one of the start sites predicted by GadFly for Trl gene (entry number CG9343). The existence of other start sites can not be discarded at this point, especially one slightly upstream to the proximal one (Fig.1, marked *1).

Inspection of the Trl promoter sequence immediately suggested several potential GAGA factor binding sites that were confirmed by DNase I footprinting analysis. In fact, as many as 27 sites were identified in a ~1.5 kbp region around the transcription start sites, and there may likely be some more that could not be resolved in our analysis. It was not our purpose to make an extensive cartography of the GAGA binding sites on this promoter and in fact, this high abundance of sites was not expected. The analysis of the sites showed that 18 of them presented GAGA sequences of high affinity (7). For the remaining sites, eight presented at least a GAG/CTC trinucleotide sequence that in principle would suffice to bind GAGA factor (8) and one did not present any sequence motif reminiscent of a GAGA site. From all of them there is no correlation that can be established with the relative affinity estimated from the gels because while some look weak, others look much stronger irrespective of the motif recognized in each case. These results suggest that depending on the context and likely due to the oligomeric nature of GAGA factor some potentially weak sites may become better than expected and be clearly observable if surrounded by high affinity sites and vice versa, some sites showing weak binding can contain a high affinity GAGA binding site, which per se would not be sufficient as already observed before (4,36). Our results are in agreement with a distribution of GAGA footprints centered around 14-15 bp (14). Some of the sites appear to be double as they often contain two high affinity GAGA binding sites and in general, protections larger than the average tend to be observed on these sites suggesting that both sequences are similarly bound. Similar results showing an extensive GAGA binding have been reported for other natural promoters (5,15,37).

These results were suggestive of a potential autoregulation of Trl promoter by GAGA factors. So far, GAGA factors were never reported to
repress transcription of any promoter but to stimulate transcription much like we have shown for the eve stripe 2 promoter. The presence of several GAGA binding sites on the eve promoter was previously shown and therefore was suspected to be under GAGA regulation (5,38). Unexpectedly however, co-transfection of either GAGA519 or GAGA581 led to a dose-dependent down regulation of Trl that reduced transcription levels to 15%. The results indicated that whereas neither Q-domain nor POZ/BTB domain were required for repression a single amino acid change affecting the Zinc-finger DNA binding domain largely abolished repression. As in the absence of POZ/BTB domain GAGA binds as a monomer and can not impose any three-dimensional structure to the promoter (4,5) architectural arguments can not explain GAGA repressive effect. On the other hand, the fact that this repression can be changed to a weak activation by replacing the activation domain may be indicative that sterical hindrance to other factors may not be the only event, unless GAGA is unable to stimulate Trl transcription because of some basal promoter selectivity. Neither it is a simple switch of the Q-domain from activation to repression because it is dispensable. On the other hand, the weak stimulation observed with GAGAΔQVP16 can not be explained by the basal promoter selectivity that VP16 can present (39) because a similar weak activation (9-fold) was previously observed for this mutant using a basal promoter that responds robustly to activation by Gal4VP16 (more than 1000-fold, not shown). Thus, the basis of its weak activation potential are likely to reside on the particular features of the chimerical construct itself (9). Our results suggest that the GAGA DNA binding domain on its own or in combination with the X-domain may be responsible for the repression of the Trl promoter. In addition, a double deletion mutant has shown that POZ/BTB domain and Q-domains do not act as redundant repressive domains that might separately be sufficient for repression in combination with the DNA binding domain (results not shown).

The negative regulation of Trl gene by its products GAGA519 and GAGA581 suggests that intranuclear concentration of GAGA is presumably kept at a constant level in S2 cells. In fact, endogenous GAGA depletion experiments indicated that Trl gene is actively kept at a submaximal level of transcription. Nevertheless, repression is not complete and intact S2 cells express normal levels of GAGA factors while a high amount of GAGA can be detected on the Trl promoter, and especially on the region that has been shown to mediate repression. The occupancy of a low affinity GAGA site which would be sensitive to changes in GAGA concentration might be a mechanism to explain it.
Nevertheless, the oligomeric nature of GAGA factor and its simultaneous binding to several sites question this simple mechanism. In any case, it is very unlikely that GAGA is acting alone but rather in combination with other(s) factor(s) which should be Trl specific. Two models that are not mutually exclusive and involve the presence of at least an additional factor can be envisaged.

In the first model, upon interacting with some factor(s) GAGA would trigger the repressive effect. Using synthetic promoters in which bona fide GAGA binding sites were inserted near to a TATA box, or in minimal natural promoters in which this situation is reproduced, GAGA only stimulated transcription in vitro and in vivo (2,4,9,14,26,33). From these results GAGA should essentially be considered as a transcriptional activator. On the other hand, the hypothetical factor that is driving GAGA-mediated repression of Trl promoter must be specific for this promoter and require GAGA binding to DNA. This hypothesis implicitly requires the existence of a DNA binding site for the unknown factor(s) in the vicinity of the GAGA sites since the silencing observed can be transported to CMV promoter by the FII+FIII region whereas it is not observed with other promoters (for instance eve which was assayed here). The observation that each of these two Trl promoter regions can still mediate GAGA silencing on their own, albeit partially, might suggest either the presence of one binding site -or more- for the unknown factor in each region or the existence of two -or more- factors which co-operate and bind to either one or the other region. In this scheme, GAGA would bind to both regions and would act as a necessary factor for eliciting repression. In agreement, other fragments from the Trl promoter despite having in some cases several good GAGA binding sites could not support repression, suggesting that while GAGA factors (either GAGA519 or GAGA581) are necessary they are not sufficient. This model has a precedent in Dorsal that although is an activator, its binding to a silencer region of the zernknüllt gene results in recruitment of the co-repressor Groucho. This does not take place with Dorsal sites alone but requires the presence of an AT2 DNA element that by binding cut and dead ringer factors collaborates with Dorsal in the recruitment of Groucho, and results in its conversion to a repressor. (40,41).

The second model would suggest that GAGA is displacing an activating factor from the Trl promoter, in such a way that transcription is not actually repressed but deactivated. The model also requires another factor, in this case an activator different to GAGA, bound close enough to some GAGA sites in the FII+FIII region so that it can be gradually displaced by increasing amounts of
GAGA. To explain why GAGA binding does not result in some activation it has to consider that the unknown activator is largely more active than GAGA itself and/or that *Trl* basal promoter is not responsive to GAGA. This model has a precedent in SV40 virus large T antigen which initially stimulates transcription from the late promoter of the virus. As the viral cycle progresses large T antigen accumulates and results in its binding to the early promoter with a displacement of the stimulatory Sp1 factor from this promoter that leads to a repression of its own transcription unit (42,43).

The existence of a GAGA site on a certain promoter can not be directly correlated to an activation of transcription. Whether GAGA binding to a promoter results in activation or repression it must be encoded in the specific DNA sequence of the promoter itself and depend on factor(s) other than GAGA. Although peculiar in the sense that GAGA factors usually are activators, this activation-to-repression conversion is not unprecedented.

Acknowledgements
A.K. and S.P. are recipients of doctoral fellowships of the "Comissió Interdepartamental de Recerca i Innovació Tecnològica" (CIRIT) of the Generalitat de Catalunya. This work was supported by grants BMC2000-0898 from the spanish "Ministerio de Ciencia y Tecnología", and SGR97-55 from the CIRIT of the Generalitat de Catalunya. This work was carried out in the context of the "Centre de Referència en Biotecnologia" of the CIRIT of the Generalitat de Catalunya.

References
1. Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., and Karch, F. (1994) *Nature* **371**, 806-808
2. Benyajati, C., Mueller, L., Xu, N., Pappano, M., Gao, J., Mosammaparast, M., Conklin, D., Granok, H., Craig, C., and Elgin, S. (1997) *Nucleic Acids Res.* **25**, 3345-3353
3. Linterman, K.-G., Roth, G. E., King-Jones, K., Korge, G., and Lehmann, M. (1998) *Dev. Genes Evol.* **208**, 447-456
4. Espinás, M. L., Jiménez-García, E., Vaquero, A., Canudas, S., Bernués, J., and Azorín, F. (1999) *J. Biol. Chem.* **274**, 16461-16469
5. Katsani, K. R., Hajibagueri, M. A. N., and Verrijzer, C. P. (1999) *EMBO J.* **18**, 698-708
6. Pedone, P. V., Ghirlando, R., Clore, G. M., Gronenborn, A. M., Felsenfeld, G., and Omichinski, J. G. (1995) Proc. Natl. Acad. Sci. U.S.A. 93, 2822-2826
7. Omichinski, J. G., Pedone, P. V., Felsenfeld, G., Gronenborn, A. M., and Clore, G. M. (1997) Nat. Struct. Biol. 4, 122-132
8. Wilkins, R. C., and Lis, J. T. (1998) Nucleic Acids Res. 26, 2672-2678
9. Vaquero, A., Espinás, M. L., Azorín, F., and Bernués, J. (2000) J. Biol. Chem. 275, 19461-19468
10. Vaquero, A., Espinás, M. L., and Bernués, J. (2002) submitted
11. Greenberg, A. J., and Schedl, P. (2001) Mol. Cell. Biol. 21, 8565-8574
12. Agianian, B., Leonard, K., Bonte, E., Van der Zandt, H., Becker, P. B., and Tucker, P. A. (1999) J. Mol. Biol. 285, 527-544
13. Wilkins, R. C., and Lis, J. T. (1999) J. Mol. Biol. 285, 515-525
14. Soeller, W. C., Oh, C. E., and Kornberg, T. B. (1993) Mol. Cell. Biol. 13, 7961-7970
15. Biggin, M. D., and Tjian, R. (1988) Cell 53, 699-711
16. Soeller, W. C., Poole, S. J., and Kornberg, T. (1988) Genes Dev. 2, 68-81
17. Wilkins, R. C., and Lis, J. T. (1997) Nucleic Acids Res. 25, 3963-3968
18. Raff, J. W., Kellum, R., and Alberts, B. (1994) EMBO J. 13, 5977-5983
19. Platero, J. S., Csík, A. K., Quintanilla, A., and Henikoff, S. (1997) J. Cell Biol. 140, 1297-1306
20. Granok, H., Leibovitch, B. A., Shaffer, C. D., and Elgin, S. C. R. (1995) Curr. Biol. 5, 238-241
21. Espinás, M. L., Canudas, S., Fanti, L., Pimpinelli, S., Casanova, J., and Azorín, F. (2000) EMBO Rep. 1, 253-259
22. Tsukiyama, T., Becker, P. B., and Wu, C. (1994) Nature 367, 525-532
23. Bhat, K. M., Farkas, G., Karch, F., Gyurkovics, H., Gausz, J., and Schedl, P. (1996) Development 122, 1113-1124
24. Okada, M., and Hirose, S. (1998) Mol. Cell. Biol. 18, 2455-2461
25. Deuring, R., Fanti, L., Armstrong, J. A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S. L., Berloco, M., Tsukiyama, T., Wu, C., Pimpinelli, S., and Tamkun, J. W. (2000) Mol. Cell 5, 355-365
26. Pile, L. A., and Cartwright, I. L. (2000) J. Biol. Chem. 275, 1398-1404
27. Strutt, H., Cavalli, G., and Paro, R. (1997) EMBO J. 16, 3621-3632
28. Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000) Mol. Cell. Biol. 20, 3187-3197
29. Poux, S., Melfi, R., and Pirrotta, V. (2001) Genes Dev. 15, 2509-2514
30. Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., and Sakonju, S. (2001) Development **128**, 2163-2173
31. Mishra, R. K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S. E., and Schedl, P. (2001) Mol. Cell. Biol. **21**, 1311-1318
32. Tsukiyama, T., and Wu, C. (1996) Methods Enzymol. **274**, 291-299
33. Jiménez-García, E., Vaquero, A., Espinás, M. L., Soliva, R., Orozco, M., Bernués, J., and Azorín, F. (1998) J. Biol. Chem. **273**, 24640-24648
34. Orlando, V., Strutt, H., and Paro, R. (1997) in Methods: A Companion to Methods in Enzymology Vol. 11, pp. 205-214
35. Cavalli, G., Orlando, V., and Paro, R. (1999) in Chromosome Structural Analysis: A practical Approach (Bickmore, W. A., ed), pp. 20-37, Oxford University Press
36. Vaquero, A. (2000) PhD Thesis. University of Barcelona.
37. Weber, J. A., Taxman, D. J., Lu, Q., and Gilmour, D. S. (1997) Mol. Cell. Biol. **17**, 3799-3808
38. Read, D., Nishigaki, T., and Manley, J. L. (1990) Mol. Cell. Biol. **10**, 4334-4344
39. Emami, K. H., Navarre, W. W., and Smale, S. T. (1995) Mol. Cell. Biol. **15**, 5906-5916
40. Dubnicoff, T., Valentine, S. A., Chen, G., Shi, T., Lengyel, J. A., Paroush, Z., and Courey, A. J. (1997) Genes Dev. **11**, 2952-2957
41. Valentine, S. A., Chen, G., Shandala, T., Fernandez, J., Mische, S., Saint, R., and Courey, A. J. (1998) Mol. Cell. Biol. **18**, 6584-6594
42. Rio, D. C., and Tjian, R. (1983) Cell **32**, 1227-1240
43. Jones, K. A., Myers, R. M., and Tjian, R. (1984) EMBO J. **3**, 3247-3255

Footnotes
1: GTF: general transcription factors; PRE: Polycomb responsive element; PCR: Polymerase Chain Reaction; RPA: Rnase protection assay; UTR: Untranslated region; RNAi: interference RNA; ChIP: Chromatin immunoprecipitation
2: S. Pagans, F.A., and J.B., unpublished observations
Figure Legends

Figure 1
Mapping of the transcription start sites of Trl gene in S2 cells.
A. Relative positions of the transcription start sites for Trl. Probes R1 and R2 used for the RPA experiments are indicated.
B. Mapping of the Trl transcription start sites in S2 cells by RPA. Lanes 1 and 3 show the riboprobes R2 and R1 respectively. Lane 2 shows the start sites detected with probe R2 and lane 4 with probe R1 (indicated by arrows). Start sites indicated Pr (proximal), In (intermediate) and Di (distal), correspond within error to the 5' ends of the cDNAs previously reported and are summarized in panel A (2). Sites marked *1 and *2 correspond to additional start sites not indicated before. Markers are included on the left to show the approximate length of the protected RNAs. In the left panel arrow labeled Di+In indicates that R2 probe could not distinguish between them for technical reasons.

Figure 2
Functional determination of the Trl promoter in S2 cells.
On the left, schematic representation of the constructs used to delineate the functional Trl promoter in S2 cells. On the right, quantification of the activity corresponding to the deletion mutants shown on the left upon transient transfection in S2 cells. Positions respect to the distal start site are indicated.

Figure 3
Determination of GAGA binding sites on the Trl promoter by DNase I footprinting analysis.
A. The sequence of the Trl promoter analyzed is shown. Promoter fragments (FI to FV) used for the DNase I footprinting analysis and other assays throughout this work are boxed. The regions bound by GAGA519 factor and protected to DNase I are underlined. Superscripts refer to footprint number for each fragment and correspond to those indicated on the autoradiographs below. Core-sequence elements inside GAGA footprints are framed and correspond to either the sites described in (7) or to GAG/CTC triplets (8). Arrows indicate the position of the three transcription start sites described before (see Fig.1 and (2)).
B. DNase I footprinting results of regions FI to FV obtained with increasing amounts of recombinant GAGA519 factor. The first lane in each panel displays a G+A sequencing ladder of the fragment. The second lane shows a DNase I
digestion of the naked DNA fragment. Lanes 3 and 4 show protection to DNase I digestion at two amounts of GAGA. Footprints are indicated by bars on the right of each panel and are correlatively numbered. The 5' to 3' direction is indicated on the right of each panel. The positions covered by each fragment are indicated at the bottom. Due to the length of fragments FIV and FV and to allow the mapping of most, if not all the footprints present on these fragments, three sets of footprinting gels that were run for increasing time are shown for each fragment. All these results were obtained using 0.5 µg of competitor DNA except for FII that used 1 µg. The amounts of GAGA used were: 0.5 and 0.8 µg for FI, 0.1 and 0.3 µg for FII, 0.5 and 0.8 µg for FIII, 0.1 and 0.3 µg for FIV, and 0.3 and 0.6 µg for FV.

Figure 4
Overexpression of GAGA519 leads to repression of Trl promoter and activation of eve promoter in S2 cells
A. Transient transfection of increasing amounts of a plasmid expressing GAGA519 produces strong repression of Trl expression in S2 cells. The long promoter fragment of Trl (-3470 to +737) was used as a reporter for the experiments. Full activity (100%) was assigned to the reporter under no GAGA519 overexpression.
B. Transient transfection of increasing amounts of a plasmid expressing GAGA519 produces activation of eve stripe 2 promoter expression in S2 cells. A long promoter fragment (about 1.5 kbp) directing expression of stripe 2 was used as a reporter for the experiments. Unit activity was assigned to the reporter under no GAGA519 overexpression.
C. Activation of eve and repression of Trl can be abolished by blocking GAGA overexpression. Effects of increasing RNAiGAGA dosage on the expression of Trl (light grey columns) and eve (dark grey columns) promoters in conditions of GAGA519 overexpression. 4 µg of GAGA519 expression construct was cotransfected with the reporter plasmids and the indicated amounts of RNAiGAGA. Luciferase and β-galactosidase activities were measured at 48 h post-transfection.

Figure 5
The repressive effect of GAGA on Trl is dependent on its DNA binding activity. Several GAGA mutants as outlined in the lower panel were tested for their activity on the long promoter fragment of Trl (-3470 to +737) at the same amount (4 µg). Full activity (100%) was assigned to the reporter under no
overexpression of any protein.

**Figure 6**
A DNA region is responsible for the specific GAGA-mediated repression of *Trl* promoter
A. Transient transfection analysis of GAGA519 overexpression on several *Trl* promoter constructs. *Trl* promoters encompassing positions -3470 to +737 (white bars), -679 to +737 (light grey bars), -345 to +737 (dark grey bars) and -270 to +737 (black bars) were assayed with increasing doses of GAGA519 expression vector. Full activity (100%) was assigned to each reporter under no overexpression of any protein to allow cross-comparisons. Note that this value is almost identical for all of them except for the shorter promoter (-270 to +737, black bars, see Fig.2).
B. Transient transfection analysis of *Trl* promoter fragments inserted into a CMV-luciferase reporter construct without GAGA519 overexpression. CMV (at the leftmost column) is the parental reporter and is taken as 100%. Fragments FI, FII, FIII and FII+FIII were inserted 5' upstream and fragments FIV and FV 3' downstream of the CMV promoter.
C. Transient transfection analysis of *Trl* promoter fragments inserted into a CMV-luciferase reporter construct upon GAGA519 overexpression. The activity of each reporter without GAGA519 overexpression (white columns, taken as 100%) and upon co-transfection of 4 µg of expression plasmid for GAGA519 (grey columns) is shown.

**Figure 7**
Endogenous GAGA factor down regulates the expression and is bound to the *Trl* genomic promoter in S2 cells.
A. Chromatin Immunoprecipitation analysis (ChIP) for GAGA in S2 cells. In the upper panel, results are shown in the form of slot-blot analysis for samples incubated without the addition of any antibody (upper row, (-)), using anti-GAGA519 antibody (middle row, (αGAGA antibody)) or an unrelated antibody (bottom row, (unrelated antibody)). Each sample was analyzed for the five regions studied (columns labeled FI to FV). Relative enrichment for GAGA is indicated below for each fragment and was calculated as the ratio of signal in antibody vs. mock immunoprecipitations. In the bottom panel relative enrichment is plotted vs. promoter position to give a visual idea of GAGA promoter occupancy in vivo.
B. Transcription of *Trl* promoter encompassing positions -3470 to +737 is
stimulated by depletion of endogenous GAGA factors in S2 cells. Stimulatory effect of increasing RNAiGAGA dosage on the expression of Tlr promoter in the absence of GAGA overexpression. Luciferase and β-galactosidase activities were measured at 72 h post-transfection.
A

+2 DISTAL → INTERMEDIATE → PROXIMAL +1

-49 UNCHANGED  +347 R1probe

+347 UNCHANGED  +337 R2probe

B

1 2
527 404 190 180
Di+In

3 4
527 404 190 180
Di

*2

In
GAGA factor down regulates its own promoter
Ana Kosoy, Sara Pagans, Mª Lluïsa Espinás, Ferran Azorín and Jordi Bernués
J. Biol. Chem. published online August 27, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M207505200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts