Biochemical and Biophysical Research Communications **283,** 644–647 (2001)
doi:10.1006/bbrc.2001.4826, available online at http://www.idealibrary.com on IDEAL®

A House Fly Gene Homologous to the Zinc Finger Proto-oncogene *Gfi*-1

Shinji Kasai and Jeffrey G. Scott[1]
Department of Entomology, Comstock Hall, Cornell University, Ithaca, New York 14853-0901

Received March 28, 2001

***Gfi*-1 was first cloned from rats (and subsequently from mice, chickens, and humans) and was found to be a 55-kDa protein that bound to DNA in a sequence-specific manner to act as a transcriptional repressor and proto-oncogene. Using PCR, a *Gfi*-1 homologous cDNA (*mdGfi*-1) was cloned from the house fly, *Musca domestica*. Comparison of the *mdGfi*-1-deduced amino acid sequence with those of vertebrates indicates only moderate overall homology (40.9–43.0% identity). However, these proteins are highly conserved when the zinc finger domains are compared, with *mdGfi*-1 having 81.0–82.2% identity to the vertebrate homologues. Within each of the six zinc finger domains there are three amino acids that are predicted to contact the DNA and these amino acids are 100% identical for all six domains for all species. Given that *Gfi*-1 is highly conserved from insects to vertebrates suggests this may be an important transcription factor in many taxa.** © 2001 Academic Press

Key Words: **gene repressor; zinc finger protein; Insecta; *Musca domestica; Drosophila melanogaster*.**

The proto-oncogene *Gfi*-1 was first cloned from rats, where it was shown to be expressed at high levels in Moloney murine leukemia virus induced T cell lymphomas selected for growth on interleukin-2-free media (1). *Gfi*-1 also contributes to the induction and progression of several types of rodent hematopoietic neoplasms. The *Gfi*-1 ORF encoded a putative zinc finger DNA-binding protein with six finger domains of the C_2-H_2 type at its carboxy-terminal region. *Gfi*-1 expression in adults was limited to the thymus, spleen and testis (1). Subsequent studies showed that *Gfi*-1 was a 55-kDa protein that bound to DNA in a sequence-specific manner to act as a transcriptional repressor (2). The *Gfi*-1 binding site was defined, confirmed and a consensus sequence for the DNA binding sequence was derived. A database search with the *Gfi*-1 consensus binding sequence revealed numerous mammalian genes having a homologous sequence in their 5′ flanking sequence, suggesting that *Gfi*-1 may have a role in regulating the expression of numerous genes (2). *Gfi*-1 homologous genes were subsequently reported from mouse (3), chicken (4), and human (5), as well as a related gene from *Caenorhabditis elegans* (6).

Low, constitutive levels of CYP6D1 (a cytochrome P450 from house fly) that can be induced by phenobarbital (7, 8) have been associated with the presence of a *Gfi*-1 homologous binding sequence within the promoter region of this gene (9). A strain that could not be induced with phenobarbital, and had constitutively high levels of CYP6D1 expression, had a 15 bp insert which interrupts this putative *Gfi*-1 binding domain, suggesting that a *Gfi*-1-like protein may have a role in strain specific and/or phenobarbital inducible expression of this gene.

To understand if a *Gfi*-1-like protein could have a role in the regulating *CYP6D1* (or other genes) we needed to determine if *Gfi*-1 existed in insects. Herein, we report on the cloning and sequencing of a *Gfi*-1-like cDNA (*mdGfi*-1) from house flies. The expression of *mdGfi*-1 and the relative conservation of *Gfi*-1 across taxa are discussed.

MATERIALS AND METHODS

Cloning of a partial Gfi-1 cDNA using degenerate primers. Poly(A)$^+$ RNA was isolated from adult house fly abdomens (0.1 g) using a Quick Prep Micro mRNA purification kit (Amersham Pharmacia Biotech). First-strand cDNA was synthesized with isolated mRNA and an oligo(dT) primer (taatacgactcactatagggagatttttttttttttt). The following degenerate primers, designed based on the homologous region of vertebrate *Gfi*-1 genes from mouse (P70338), rat (Q07120), human (NP004179), and chicken (CAA71836), were synthesized (Cornell BioResource Center) for PCR: 5′-GA-(C/T)AC(A/C/G/T)(A/C)G(A/C/G/T)CC(A/C/G/T)TA(C/T)CC(A/C/G/T)-TG-3′ (GfiF primer for DTRPYPC) and 5′-AA(A/C/G/T)GG(C/T)TT(A/G)AA(A/C/G/T)CC(A/C/G/T)GT(A/G)TG-3′ (GfiR primer for HTG-FKPF). *Gfi*-1 cDNA was amplified by RT-PCR under the following conditions. The reaction mixtures (100 μl) were first kept at 95°C for 1 min, then 40 cycles of PCR (95°C for 30 s, 47°C for 30 s and 72°C for 40 s) were done, and the samples were finally kept at 72°C for 5 min. The sample solution (1 μl) was used as template for the 2nd

[1] To whom correspondence should be addressed. Fax: (607) 255-0939. E-mail: JGS5@CORNELL.EDU.

0006-291X/01 $35.00

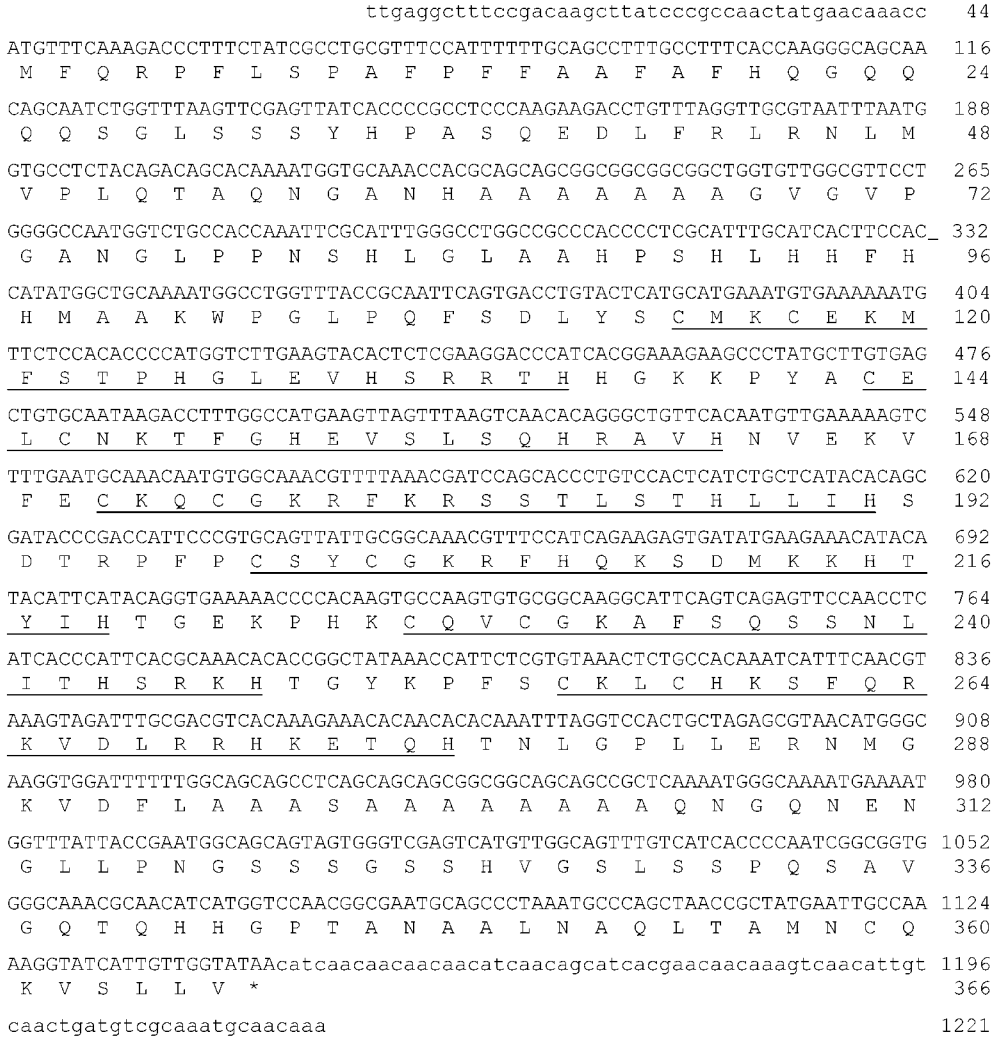

FIG. 1. The cDNA and deduced amino acid sequences of *mdGfi*-1 from the CS strain of house fly (Accession No. AF339860). The stop codon is indicated by an asterisk. Sequences of the six zinc finger motifs were underlined.

PCR which was performed under the same conditions as above. PCR products were separated by agarose gel electrophoresis (4.0%). Those products of expected length (183 bp) were excised and purified using a QIAquick Gel Extraction Kit (QIAGEN) and cloned into TA-cloning vector (Invitrogen) as described by the manufacturer. Plasmid DNA from putative positive clones (i.e., those having the correct size insert) were isolated and sequenced (Cornell Biotechnology Facility).

3′ and 5′ RACE. 3′-RACE was performed to determine the 3′-region of *Gfi*-1. First-strand cDNA was synthesized with oligo(dT) C2 primer as described above. First PCR was performed with Gfi5 (5′-cgtttccatcagaagagtg-3′) and C2 primers (5′-taatacgactcactat-agggaga-3′) with cDNA as template. The reaction mixtures (100 μl) were first kept at 95°C for 1 min, then 40 cycles of PCR (95°C for 30 s, 47°C for 30 s and 72°C for 40 s) were done, and the samples were finally kept at 72°C for 5 min. The second PCR was then performed using the first PCR (1 μl) as a template with Gfi6 (5′-aca-agtgccaagtgtgc-3′) and C2 primers. The PCR conditions were the same as the first reaction. PCR products were separated by agarose gel electrophoresis and a clear band (600 bp) was excised, gel purified, TA-cloned and sequenced as described above.

The 5′ end of *Gfi*-1 cDNA was amplified using 5′-RACE System (Gibco). First strand cDNA was synthesized using house fly mRNA (3 μg) and the GfiX primer (5′-atgttacgctctagcagtgg-3′) designed based on the 3′ sequence. First strand cDNA was purified by GlassMAX spin column chromatography (Gibco) and tailed with dCTP and TdT followed by the first PCR using synthesized cDNA as a template, and GfiY primer (5′-tgtggcagagtttacacgag-3′) and an anchor primer from the kit. The reaction mixture was first kept at 95°C for 5 min and then 40 cycles of PCR (95°C for 40 s, 55°C for 40 s, and 72°C for 40 s) were done, and the samples were finally kept at 72°C for 5 min. The second PCR was run with GfiZ primer (cggaattcgaggttggaactctgactg) and the Abridged Universal Amplification Primer (from the kit) under the same conditions as the first PCR. The products were separated by agarose gel, DNA of approximately 800–1000 bp was excised, purified, TA-cloned and sequenced.

Sequence analysis. Sequences were aligned and compared using the MegAlign (3.06b) program (Clustal method) from DNASTAR (Madison, WI). Searches for similar sequences were carried out using a Blast search (10) of GenBank.

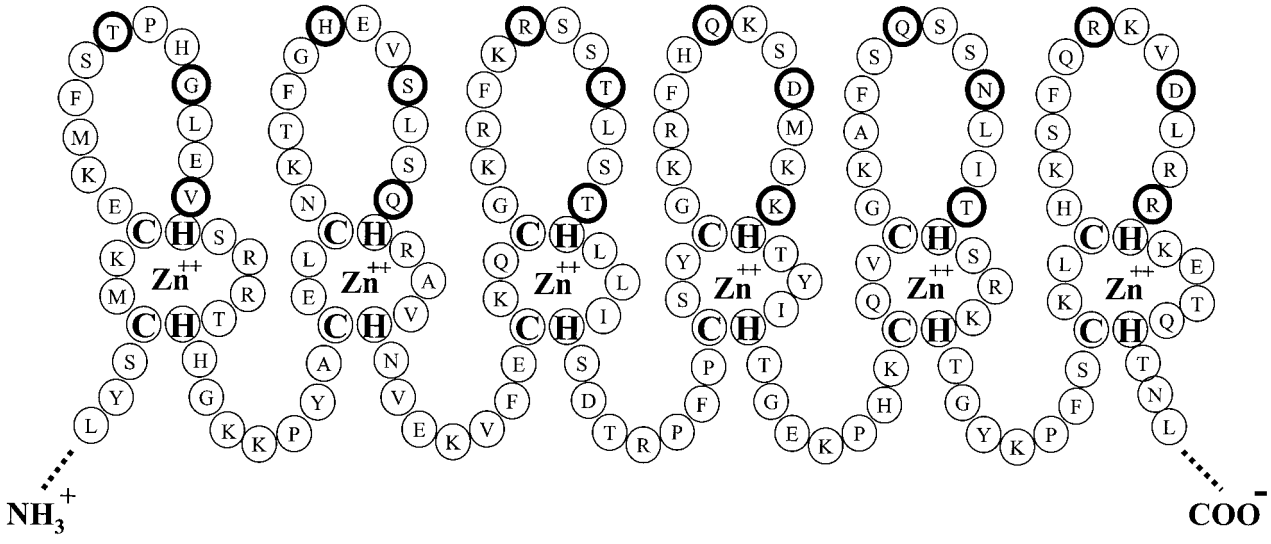

FIG. 2. Diagram of zinc finger domains of *mdGfi*-1. The *mdGfi*-1 deduced amino acid sequence has a series of six zinc fingers that are characteristic of vertebrate *Gif*-1s. Each finger is formed with cysteine and histidine residues that constitute the zinc-binding site. Amino acids in bold circles are predicted to contact DNA (2).

RESULTS AND DISCUSSION

The cDNA and deduced amino acid sequence of the house fly *Gfi*-1 like gene (*mdGfi*-1) are shown in Fig. 1. *mdGfi*-1 codes for a typical C_2H_2 type zinc finger protein having six zinc finger domains (as illustrated in Fig. 2) which is characteristic of vertebrate *Gfi*-1s. Comparison of the house fly *Gfi*-1 like protein with those of vertebrates (Table 1) indicates only moderate overall homology (40.9–43.0% identity). However these proteins are highly conserved when the zinc finger domains are compared (Fig. 3), with the house fly *Gfi*-1 having 81.0-82.2% identity to the vertebrate homologues. Within each of the six zinc finger domains there are three amino acids that are predicted to contact the DNA (2). These amino acids are 100% identical for all six domains for all species. Further, in domain 5 the house fly sequence is 100% identical to the vertebrate sequences (Fig. 3), suggesting this domain may be critical for gene regulatory processes in both insects and vertebrates. The high similarity of the linker sequences between domains 4 and 5 (and perhaps 3 and 4) suggests that these regions may serve some role in the function of *Gfi*-1 beyond spacing of the zinc finger domains. The observation that zinc finger domains 3, 4, and 5 were the most conserved across all species is consistent with results that indicated these same domains were required for DNA binding (whereas domains 1, 2, and 6 were dispensable) (2). Another feature of vertebrate *Gfi*-1 is the presence of an amino-terminal domain of 20 amino acids known as a SNAG (nuclear localization) domain (11). House fly *Gfi*-1 does not have a homologous sequence at its amino terminus.

There is generally a high degree of sequence conservation for proteins involved in transcription (TATA binding protein, etc.) across eukaryotic taxa. Thus, the high degree of homology of *mdGfi*-1 with its vertebrate counterparts suggests this may be an important transcription factor in animals. In vertebrates *Gfi*-1 represses transcription either directly (2) or indirectly (12). Given the high degree of homology between vertebrate and *mdGfi*-1, it seems reasonable to assume *Gfi*-1-like proteins may also act as transcriptional repressors in insects.

Although there is very limited information about the promoter regions of genes in the house fly, we were able to identify one gene, *CYP6A8,* which had a region in its 5′ flanking sequence with a putative *Gfi*-1 binding site (9 of 12 nucleotides matched the mammalian *Gfi*-1 consensus site). However, lack of a *Gfi*-1 binding site consensus sequence in insects prevents a meaningful attempt at identification of the genes that could be regulated by *Gfi*-1.

TABLE 1

Percent Identity of Insect and Vertebrate *Gfi*-1 Homologous Proteins

	House fly	Fruit fly	Chicken	Mouse	Rat	Human
House fly	—	66.2	43.0	40.9	40.9	40.3
Fruit fly	98.2	—	44.8	39.2	39.5	40.3
Chicken	82.2	83.4	—	51.0	51.3	50.1
Mouse	81.6	81.6	92.0	—	96.9	85.1
Rat	81.6	81.6	92.0	100	—	85.1
Human	81.0	81.0	91.4	99.4	99.4	—

Note. Figures in upper right and lower left are percent identity of full-length and zinc finger domain sequences, respectively.

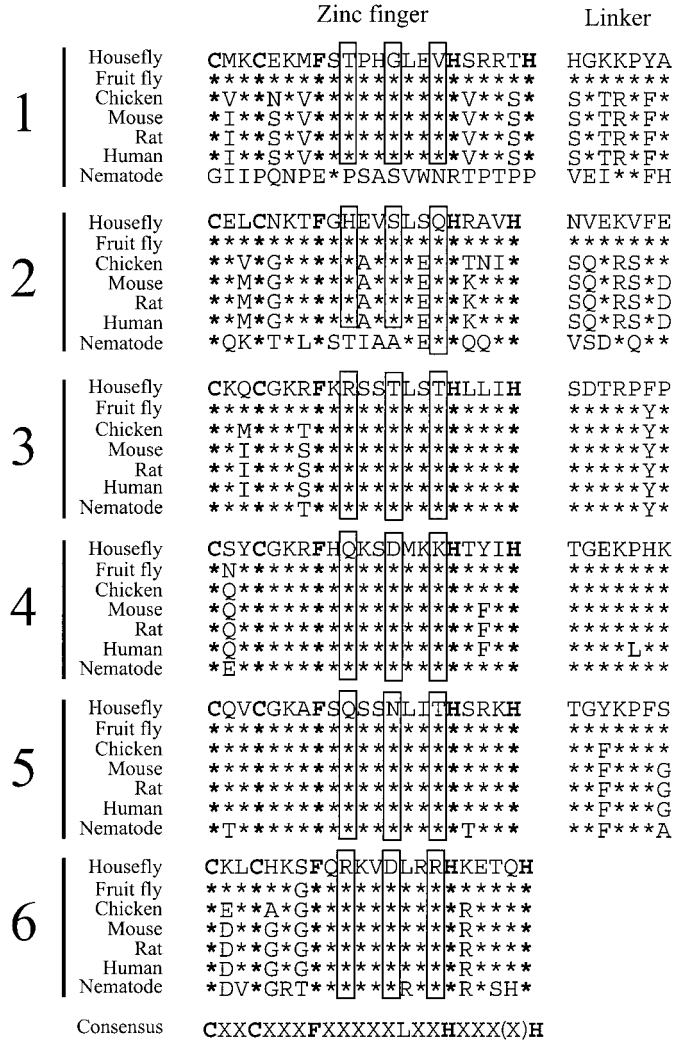

FIG. 3. Alignment of six zinc finger domains and linker regions of *Gfi*-1 homologous proteins. Asterisks denote identical amino acids to those of house fly. Amino acids predicted to contact DNA are boxed (2). Conserved amino acids are indicated in bold letters. The C_2H_2-type zinc finger consensus sequences are shown at the bottom. Accession numbers of amino acid sequences are as follows: house fly (AF339860), fruit fly (AAF52446), chicken (CAA71836), mouse (P70338), rat (Q07120), human (NP_004179), and nematode PAG-3 (U63996).

Another *Gfi*-1 homologue, PAG-3 (pattern of gene expression abnormal-3), was isolated from *C. elegans* (6) and is involved in controlling neuron-specific gene expression (6). Unlike *Gfi*-1, this protein has only five zinc fingers and the region outside of the zinc finger domain is not conserved, however, the last four fingers are highly conserved with vertebrate *Gfi*-1 protein and this domain is 88% identical to human *Gfi*-1 (6). Furthermore, PAG-3 binds the *Gfi*-1 consensus binding site tightly (McDermott and Aamodt, unpublished results, cited in (13)) suggesting possible ability of *mdGfi*-1 to recognize and bind to *Gfi*-1 consensus binding site. The greater similarity of the insect and vertebrate sequences compared to *C. elegans* is consistent with current hypotheses of the evolutionary relatedness of these groups (14).

We also carried out a Blast search to determine if a *mdGfi*-1 homologous gene existed in the genome of *Drosophila melanogaster*. An homologous protein was identified (Accession No. AAF52446), which we termed *dmGfi*-1, and it is 66.2% identical overall to *mdGfi*-1 and 98.8% identical across the six zinc finger domains (Table 1). Given that *Gfi*-1 exists in the two insects and several vertebrate species suggests this gene is widely conserved across taxa.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (GM47835) and Hatch Project 414.

REFERENCES

1. Gilks, C. B., Bear, S. E., Grimes, H. L., and Tsichlis, P. N. (1993) *Mol. Cell. Biol.* **13,** 1759–1768.
2. Zweidler-McKay, P. A., Grimes, H. L., Flubacher, M. M., and Tsichlis, P. N. (1996) *Mol. Cell. Biol.* **16,** 4024–4034.
3. Zornig, M., Schmidt, T., Karsunky, H., Grzeschiczek, A., and Moroy, T. (1996) *Oncogene* **12,** 1789–1801.
4. Fuchs, B., Wagner, T., Rossel, N., Antoine, M., Beug, H., and Niessing, J. (1997) *Gene* **195,** 277–284.
5. Roberts, T., and Cowell, J. K. (1997) *Oncogene* **14,** 1003–1005.
6. Jia, Y., Xie, G., McDermott, J. B., and Aamodt, E. (1997) *Dev. Biol.* **124,** 2063–2073.
7. Liu, N., and Scott, J. G. (1997) *Insect Mol. Biol.* **6,** 77–81.
8. Kasai, S., and Scott, J. G. (2001) *Insect Mol. Biol.*, in press.
9. Scott, J. G., Liu, N., Wen, Z., Smith, F. F., Kasai, S., and Horak, C. E. (1999) *Gene* **226,** 347–353.
10. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
11. Grimes, H. L., Chan, T. O., Zweidler-McKay, P. A., Tong, B., and Tsichlis, P. N. (1996) *Mol. Cell. Biol.* **16,** 6263–6272.
12. Rodel, B., Tavassoli, K., Karsunky, H., Schmidt, T., Bachmann, M., Schaper, F., Heinrich, P., Shuai, K., Elsasser, H. P., and Moroy, T. (2000) *EMBO J.* **19,** 5845–5855.
13. Aamodt, E., Shen, L., Marra, M., Schein, J., Rose, B., and McDermott, J. B. (2000) *Gene* **243,** 67–74.
14. Brusca, R. C., and Brusca, G. J. (1990) Invertebrates, Sinauer Associates, Sunderland, MA.