



basic region of dMyc is more similar to its counterpart in human c-Myc (69% identity) than to other human and *Drosophila* bHLH proteins (11). Notably, all but one of the basic region residues predicted to contact DNA (13, 14) are identical to those in human c-Myc. The HLH region of dMyc is more divergent (35% identity to human c-Myc). Interestingly, dMyc resembles all known Myc proteins in that the Myc-box II is NH<sub>2</sub>-terminal, whereas the bHLHZip region is at the extreme COOH-terminus.

Furthermore, the *dmec* gene is comprised of three exons with the protein-coding region spanning exons 2 and 3 as in mammalian myc. The positions of the exon-intron junctions are also similar to those in vertebrate myc (1, 15).

In a directed yeast interaction assay, the bHLHZip region of dMyc showed no detectable association with the bHLH domains of the *Drosophila* proteins E(spl)m3, Da, Hairy, and Emc, nor did dMyc form homodimers (15). We therefore hypothe-

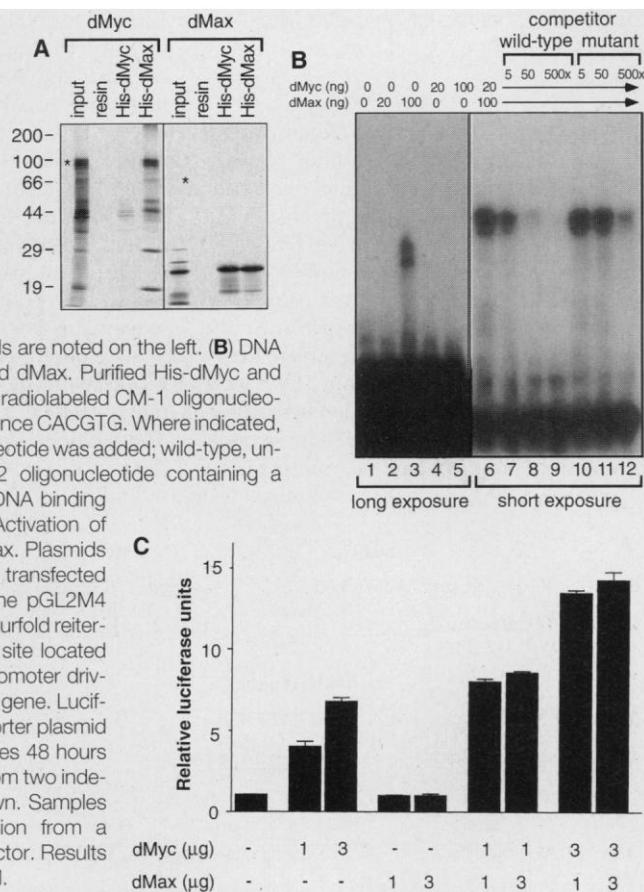
sized that *Drosophila* may encode a Max-related dimerization partner for dMyc. To identify this protein, we performed a second yeast two-hybrid screen using the bHLHZip of dMyc as bait. These experiments yielded a *Drosophila* bHLHZip protein related to vertebrate Max, which we called dMax (10) (Fig. 1B). The clones contained a complete 161 amino acid ORF preceded by several in-frame stop codons. A computer search of the protein database with the dMax ORF identified vertebrate Max proteins as the most closely related (11). The greatest sequence similarity is within the bHLHZip domain, which shows 67% identity (52% overall identity in the alignment shown in Fig. 1B). Importantly, all residues contacting DNA are conserved. Furthermore, the NH<sub>2</sub>-termini of dMax and human Max are highly conserved (52% identity) and include two casein kinase II phosphorylation sites that negatively regulate DNA binding (16) (Fig. 1B).

We next assessed the interaction of dMyc and dMax in vitro by incubating histidine-tagged dMyc or dMax (His-dMyc or His-dMax), linked to a nickel-agarose resin, with in vitro-translated [<sup>35</sup>S]methionine-labeled proteins. Labeled proteins bound to the resin after detergent washes were detected by SDS-polyacrylamide gel electrophoresis (PAGE) and autoradiography (17). His-dMax interacted strongly with dMyc and dMax, whereas His-dMyc associated with dMax but not dMyc (Fig. 2A). These binding properties are consistent with those of vertebrate Myc and Max (4, 5).

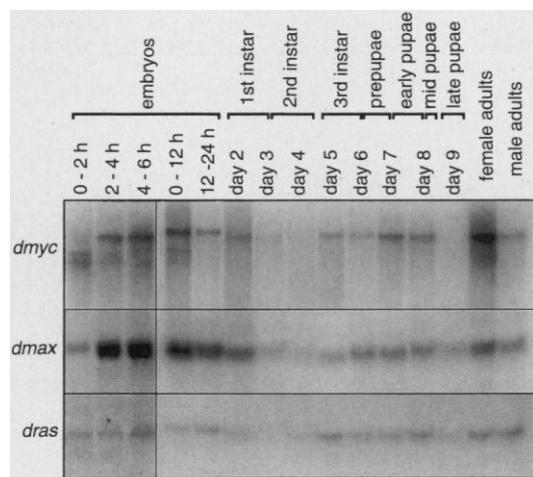
To determine whether the bHLHZip domains of dMyc and dMax exhibit E box binding specificity, we used bacterially expressed proteins in electrophoretic mobility shift assays (EMSA) with a labeled oligonucleotide probe (CM-1) containing one copy of the CACGTG site (18). The dMyc protein alone did not bind CM-1 even at relatively high concentrations of protein, whereas dMax alone bound the probe weakly only at high protein concentrations (Fig. 2B). By contrast, dMyc and dMax together generated a strong band shift with CM-1 under conditions in which neither protein alone bound. The specificity of dMyc-dMax binding to CACGTG was demonstrated by competition experiments. A fivefold and 50-fold excess of unlabeled CM-1 reduced dMyc-dMax binding to the labeled CM-1 probe by two- and eightfold, respectively. However, unlabeled B1/B2 oligonucleotide containing a MyoD E-box binding site CACCTG (18) reduced binding to labeled CM-1 only when present at a 500-fold excess over labeled CM-1. We then examined whether dMyc-dMax heterodimers could transactivate a CACGTG-

**Fig. 2.** Biochemical properties of dMyc and dMax.

(A) In vitro association was assayed by incubation of bacterially expressed histidine-tagged dMyc and dMax bound to a nickel resin (His-dMyc; His-dMax) with the [<sup>35</sup>S]methionine-labeled in vitro-translated proteins indicated at the top (17). Asterisks indicate the positions of dMyc (left) and dMax (right). Molecular size standards are noted on the left. (B) DNA binding properties of dMyc and dMax. Purified His-dMyc and His-dMax were incubated with radiolabeled CM-1 oligonucleotide probe containing the sequence CACGTG. Where indicated, unlabeled competitor oligonucleotide was added; wild-type, unlabeled CM-1; mutant, B1/B2 oligonucleotide containing a MyoD binding site CACCTG. DNA binding was analyzed as in (18) (C) Activation of transcription by dMyc and dMax. Plasmids encoding dMyc or dMax were transfected into 293 cells together with the pGL2M4 reporter. pGL2M4 contains a fourfold reiteration of the CACGTG binding site located proximal to a minimal SV40 promoter driving expression of the luciferase gene. Luciferase activity produced by reporter plasmid was determined from cell lysates 48 hours later (19). Duplicate samples from two independent experiments are shown. Samples were normalized for expression from a β-galactosidase expression vector. Results are presented as mean ± SEM.



**Fig. 3.** Expression of *dmec* and *dmec* during *Drosophila* development. Poly(A)<sup>+</sup> RNA was prepared from Oregon R flies at the indicated stages and analyzed by Northern blotting. The radiolabeled probes indicated at the left were used successively on the two blots shown (20). The *dras* mRNA served as a loading control because it is expressed at nearly constant levels throughout *Drosophila* development (28); note that several lanes are underloaded. The numbers above the lanes indicate days after egg-laying; each sample represents a successive 24-hour period.



containing reporter gene in mammalian cells (19). A cytomegalovirus (CMV) vector encoding dMyc stimulated the reporter gene in a concentration-dependent manner (Fig. 2C). Expression of dMax alone produced no transactivation over background levels; however, cotransfection of dMax with higher concentrations of dMyc vector significantly enhanced the transactivation potential of dMyc. These findings demonstrate that the ability of Max and Myc to recognize the CACGTG sequence and activate transcription has been conserved between vertebrates and invertebrates.

We examined the temporal and spatial patterns of *dmyc* and *dmax* expression during development. As shown by Northern blot analysis (20) (Fig. 3), *dmax* is expressed as a single ~1.2-kb transcript at essentially constant levels throughout development. In contrast, a *dmyc* transcript of ~6 kb is expressed at the highest levels during early embryogenesis and in adult females. It is barely detectable during the larval and pupal stages. In addition, early embryos and adult females contain smaller *dmyc* transcripts that appear to differ primarily in their untranslated regions (15).

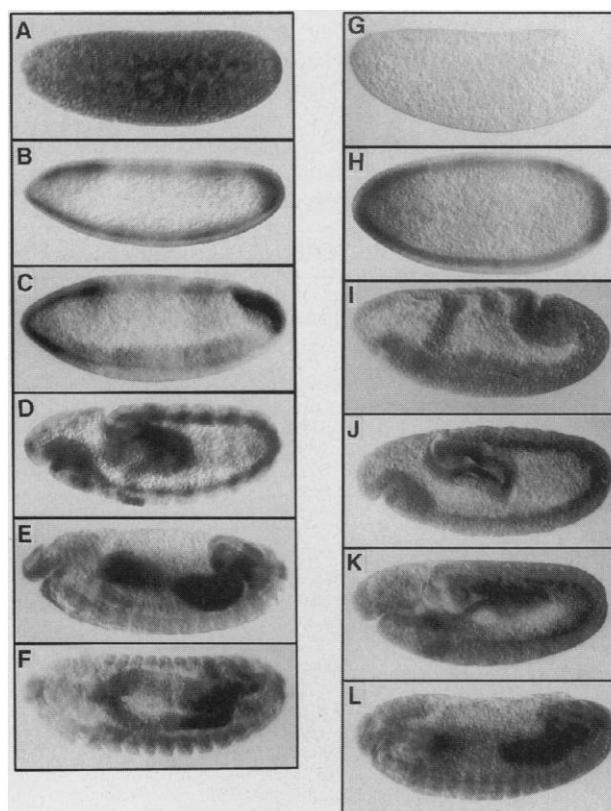
The spatial distribution of *dmyc* and *dmax* transcripts was examined in early embryos by in situ hybridization (21). The *dmyc* transcripts, presumably maternal in origin, can be detected ubiquitously from the earliest stages (Fig. 4). Later the zygotically derived transcripts accumulate in a dynamic pattern in various tissues throughout embryogenesis. In preblastoderm embryos, *dmyc* transcripts are present throughout the embryo, with highest levels at anterior and posterior termini, but are absent from the pole cells. In early gastrulation, additional intense *dmyc* staining can be detected in the presumptive mesoderm along the ventral midline. This mesodermal staining intensifies during germband extension and remains until late embryogenesis. The terminally located *dmyc* expression follows the posterior and anterior midgut primordia during the germband extended stages. Additional intense *dmyc* staining is found transiently in structures presumably corresponding to salivary placodes. These results indicate that *dmyc* is expressed in endoreplicating as well as in mitotically dividing tissues. The *dmax* transcript is less abundant than *dmyc*, particularly during the earliest stages. In addition, *dmax* is expressed in tissues with undetectable levels of *dmyc*, such as the developing central nervous system. Hence, whereas tissues containing the highest levels of *dmyc* and *dmax* transcripts are undergoing DNA replication, not all actively proliferating tissues have detectable levels of *dmyc*.

By means of chromosome in situ hybrid-

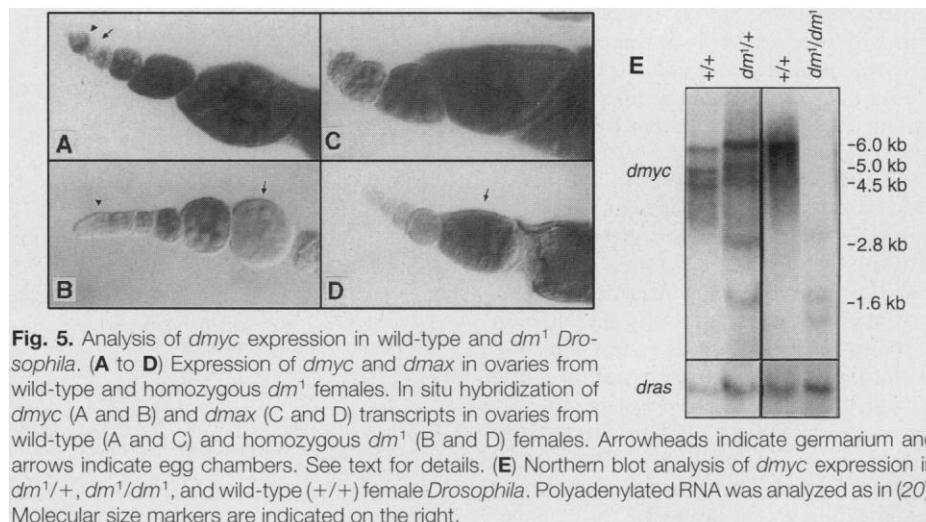
ization, we localized the *dmyc* gene to the X chromosome at polytene band 3D5 (15). Preliminary experiments with existing mutations in the region (15) pointed to *diminutive* (*dm*) as a strong candidate for *dmyc*. The *dm* gene has not yet been cloned, but *dm*<sup>1</sup> is thought to be a hypomorphic mutation resulting from insertion of a gypsy transposable element. This supposition is based on a genetic interaction of *dm*<sup>1</sup> with *su(Hw)*, as well as hybridization of a gypsy probe to the polytene band 3D5 of the *dm*<sup>1</sup>, but not the *dm*<sup>+</sup>, chromosome (22, 23). Phenotypically, *dm*<sup>1</sup> homozygous flies are

viable but of smaller body size, and *dm*<sup>1</sup> females are sterile as a result of defective oogenesis. Defects are observed in both germ cell nuclei and follicle cells, which fail to migrate and undergo the transition to columnar epithelium (23). This phenotype suggests dysfunction in follicle or nurse cells, or in communication between these two cell types.

To determine whether *dmyc* and *dm* are related, we analyzed the *dmyc* genes in Oregon R wild-type flies and in flies heterozygous for the *dm*<sup>1</sup> mutation. In chromosomes bearing the *dm*<sup>1</sup> mutation, the *dmyc* gene



**Fig. 4.** Expression of *dmyc* and *dmax* during embryogenesis. In situ hybridization of *dmyc* (A to F) and *dmax* (G to L) transcripts in wild-type embryos. Embryo stages are as follows: (A) stage 2; (B) stage 5; (C) stage 6; (D) stage 11; (E) stage 12; (F) stage 16; (G) stage 2; (H) stage 4; (I) stage 7; (J) stage 10; (K) stage 11; (L) stage 14. See text for details.



**Fig. 5.** Analysis of *dmyc* expression in wild-type and *dm*<sup>1</sup> *Drosophila*. (A to D) Expression of *dmyc* and *dmax* in ovaries from wild-type and homozygous *dm*<sup>1</sup> females. In situ hybridization of *dmyc* (A and B) and *dmax* (C and D) transcripts in ovaries from wild-type (A and C) and homozygous *dm*<sup>1</sup> (B and D) females. Arrowheads indicate germarium and arrows indicate egg chambers. See text for details. (E) Northern blot analysis of *dmyc* expression in *dm*<sup>1</sup>/+, *dm*<sup>1</sup>/*dm*<sup>1</sup>, and wild-type (+/+) female *Drosophila*. Polyadenylated RNA was analyzed as in (20). Molecular size markers are indicated on the right.

contains a gypsy transposon insertion in the first intron, 418 nucleotides upstream of the translation initiation site (15). The gypsy insertion occurs within a six-nucleotide sequence identical to other known gypsy integration sites (24). The location of this element within the *dmyc* coding region strongly supports the hypothesis that *dm* codes for dMyc.

Because homozygous *dm*<sup>1</sup> flies exhibit female sterility due to defective ovarian development (23), we analyzed the expression pattern of *dmyc* by RNA in situ hybridization of ovaries (Fig. 5, A to D) (25). The *dmyc* gene shows a dynamic expression pattern. At early stages, high levels of *dmyc* transcripts accumulate in the germarium (Fig. 5A, arrowhead), with lower levels in stage 1 or stage 2 egg chambers (Fig. 5A, arrow). By stage 3, *dmyc* transcripts can be detected in all cell types of the chamber: the nurse cells, oocyte, and follicle cells. This expression pattern is maintained throughout oogenesis. In contrast, in ovaries isolated from females homozygous for *dm*<sup>1</sup>, *dmyc* expression is undetectable in the germarium (Fig. 5B, arrowhead). The *dmyc* transcripts are present from stage 3 onward in *dm*<sup>1</sup> ovaries, as in wild-type flies, but at a lower level. The expression of *dmyc* is not maintained throughout oogenesis in *dm*<sup>1</sup> ovaries. The *dm*<sup>1</sup> ovaries are morphologically normal until stage 8, but *dmyc* expression is not detected beyond stage 6 (Fig. 5B, arrow). The absence of *dmyc* transcripts in stage 8 egg chambers is not due to degeneration of the ovaries, as *dmax* expression can still be detected in *dm*<sup>1</sup> ovaries at these later stages (Fig. 5D, arrow). Hence, *dmyc* expression ceases prior to the onset of ovarian degeneration in *dm* mutant ovaries. Analysis of *dmyc* expression in embryos derived from crosses between *dm*<sup>1</sup>/Y males and *dm*<sup>1</sup>/+ females (15) revealed few embryos with an altered pattern of *dmyc* staining. However, Northern blot analysis of mRNA isolated from *dm*<sup>1</sup>/*dm*<sup>1</sup> flies revealed substantially reduced levels of the major *dmyc* transcript as well as additional *dmyc*-derived transcripts. Smaller transcripts are also present in *dm*<sup>1</sup>/+ females but not in wild-type females (Fig. 5E).

Our results suggest that the altered expression of *dmyc* in the *dm*<sup>1</sup> mutant flies is caused by insertion of the gypsy element into the *dmyc* gene. It is formally possible that the gypsy insertion influences expression of other genes; however, the location of the insertion and the alterations in *dmyc* expression argue strongly that *dmyc* is the critical target gene. Interestingly, the degeneration of the egg chamber in *dm*<sup>1</sup>/*dm*<sup>1</sup> females occurs at about stage 8, when cell division does not occur. In the mutant ovaries, the follicle cells lining the egg chamber

grow abnormally and degenerate before their transition to columnar epithelium surrounding the oocyte. We speculate that a stage-specific downregulation of *dmyc* expression in *dm*<sup>1</sup> due to the gypsy insertion results in a loss of the capacity of the follicle cells to grow and migrate. A possibly related effect has been observed in mice, where a hypomorphic N-myc mutation resulted in a lethal loss of induction of tissue-specific differentiation (26). In both cases diminished Myc expression in progenitor cells may result in their inability to respond to inductive signals. In addition, the smaller size of the *dm*<sup>1</sup>/*dm*<sup>1</sup> flies (23) may also result from partial loss of dMyc function in other tissues. Recent studies in mice demonstrate that alterations in cell cycle regulatory genes can significantly influence overall body size (27). It will clearly be of interest to determine the phenotypes of flies carrying null mutations in *dm(dmyc)* or that ectopically overexpress *dm(dmyc)*.

## REFERENCES AND NOTES

- C. V. Dang, *Biochim. Biophys. Acta* **1072**, 103 (1991); A. Meichle, A. Philipp, M. Eilers, *ibid.* **1114**, 129 (1992); G. C. Prendergast, and E. B. Ziff, *Trends Genet.* **8**, 91 (1992); M. Henriksson and B. Luscher, *Adv. Cancer Res.* **68**, 109 (1996).
- I. Magrath, *Cancer Res.* **55**, 134 (1990); C. A. Spencer and M. Groudine, *ibid.* **56**, 1 (1991); R. A. DePinho, N. Schreiber-Agus, F. W. Alt, *ibid.* **57**, 1 (1991); K. B. Marcu, S. A. Bossone, A. J. Patel, *Annu. Rev. Biochem.* **61**, 809 (1992).
- B. R. Stanton, A. S. Perkins, L. Tessarollo, D. A. Sassoon, L. F. Parada, *Genes Dev.* **6**, 2235 (1993); J. Charron *et al.*, *ibid.* p. 2248; A. C. Davis, M. Wims, G. D. Spotts, S. R. Hann, A. Bradley, *ibid.* **7**, 671; S. Sawai *et al.*, *Development* **117**, 1445 (1993).
- E. M. Blackwood and R. N. Eisenman, *Science* **251**, 1211 (1991).
- G. C. Prendergast, D. Lawe, E. B. Ziff, *Cell* **65**, 395 (1991).
- B. Amati *et al.*, *Nature* **359**, 423 (1992).
- L. Kretzner, E. M. Blackwood, R. N. Eisenman, *ibid.* **359**, 426 (1992); C. Amin, A. J. Wagner, N. Hay, *Mol. Cell. Biol.* **13**, 383 (1993); W. Gu, K. Cechova, V. Tassi, R. Dalla-Favera, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2935 (1993); C. Bello-Fernandez, G. Packham, J. L. Cleveland, *ibid.*, p. 7804; L. Desbarats, S. Gaubatz, M. Eilers, *Genes Dev.* **10**, 447 (1996); K. Galaktionov, X. Chen, D. Beach, *Nature* **382**, 511 (1996); C. Grandori, J. Mac, F. Siebelt, D. E. Ayer, R. N. Eisenman, *EMBO J.* **15**, 4344 (1996).
- B. Mukherjee, S. D. Morgenbesser, R. A. DePinho, *Genes Dev.* **6**, 1480 (1992); B. Amati *et al.*, *Cell* **72**, 233 (1993); B. Amati, T. D. Littlewood, G. I. Evan, H. Land, *EMBO J.* **12**, 5083 (1993).
- W. R. Atchley and W. M. Fitch, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10217 (1995).
- A yeast two-hybrid screen was performed as in T. Durfee *et al.* [*Genes Dev.* **7**, 555 (1993)], using the yeast strain Y190. As bait we used a human *max* cDNA fused in frame with the DNA binding domain of GAL4 in the yeast vector GBT9 [C. D. Laherty, unpublished data; P. L. Bartel, C. Chien, R. Sternglanz, S. Fields, in *Cellular Interactions in Development: A Practical Approach*, D. Hartley, Ed. (Oxford Univ. Press, Oxford, UK, 1993), pp. 153-179]. The cDNA library was from *Drosophila* third-instar larvae in the vector pACT/pSE 1107. From  $1.8 \times 10^6$  primary transformants, seven clones strongly interacted with human Max, all of which were derived from a single gene, *dmyc*. A cDNA encoding the complete dMyc protein was isolated from an embryonic cDNA library
- [N. H. Brown and F. C. Kafatos, *J. Mol. Biol.* **203**, 425 (1988)]. To identify dMax, we used GBT9-dMyc (amino acids 579 to 717) as bait. Clones strongly interacting with GBT9-dMyc were shown to be derived from one gene, *dmax*. Larger *dmax* cDNAs, were isolated from a 0- to 2-hour embryo library (Novagen, Madison, WI). Genomic *dmyc* clones were isolated from a cosmid library [J. W. Tamkun *et al.*, *Cell* **68**, 561 (1992)]. For analysis of the *dm*<sup>1</sup> mutant *dmyc* gene, genomic DNA was prepared from *dm*<sup>1</sup>/+ flies as in S. M. Parkhurst and V. G. Corces [*Cell* **41**, 429 (1985)]. Inserts from clones hybridizing with probes derived from *gypsy* and from *dmyc* were analyzed further.
- The deduced dMyc and dMax sequences were used to search GenBank and EMBL databases using the program BLASTX [S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990)]. Full-length dMyc had the greatest homology to human c-Myc ( $P = 6.1 \times 10^{-10}$ ), and the 84 most homologous sequences were all derived from viral or cellular *myc* family genes. For the bHLHZip domain of dMyc, the 83 most closely related proteins belonged to the Myc family. Full-length Max was most similar to chicken Max [ $P = 1.9 \times 10^{-39}$ ], followed by 15 other Max proteins. For additional sequence analysis (Fig. 1) we used programs from the GCG package (Genetics Computer Group, University of Wisconsin).
- D. E. Brough, T. J. Hofmann, K. B. Ellwood, R. A. Townley, M. D. Cole, *Mol. Cell. Biol.* **15**, 1536 (1995).
- C. V. Dang, C. Dolde, M. L. Gillison, G. J. Kato, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 599 (1992).
- A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff, S. K. Burley, *Nature* **363**, 38 (1993).
- P. Gallant, Y. Shio, P. F. Cheng, S. Parkhurst, R. N. Eisenman, unpublished data.
- S. J. Berberich and M. D. Cole, *Genes Dev.* **6**, 166 (1992); P. J. Koskinen, I. Väström, T. P. Mäkelä, R. N. Eisenman, K. Allitalo, *Cell Growth Diff.* **5**, 313 (1994).
- Histidine-tagged dMyc and dMax proteins were expressed from pET19b vectors (Novagen) containing a dMyc cDNA fragment (amino acids 534 to 717) or a dMax cDNA fragment (amino acids 26 to 161), and purified. In vitro transcription and translation were performed as in (4) using pRC/CMV vectors (Invitrogen, San Diego, CA). For in vitro interaction assays, 4  $\mu$ l of each translated protein was incubated with 2  $\mu$ l of His-dMyc or His-dMax attached to a nickel resin or 2  $\mu$ l of bovine serum albumin (BSA)-blocked nickel resin in 400  $\mu$ l of L buffer (phosphate-buffered saline containing 0.4% NP-40) for 1 hour at 4°C. After three washes with L buffer, the bound proteins were eluted with SDS-containing sample buffer and subjected to SDS-PAGE and autoradiography.
- His-dMyc and/or His-dMax was incubated with 1 ng of radiolabeled CM-1 oligonucleotide probe (4) in a buffer containing 25 mM Hepes (pH 7.9), 50 mM KCl, 5 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 5% glycerol, 10 mM dithiothreitol, 0.1% NP-40, 0.5 mg/ml BSA, and 1  $\mu$ g of sheared salmon sperm DNA for 30 min at room temperature. For competition experiments, increasing amounts of unlabeled CM-1 or B1/B2 (4) oligonucleotide were included in the binding reaction. The DNA-protein complexes were resolved on a 4% polyacrylamide gel containing 2.5% glycerol and subjected to autoradiography.
- Human embryonic kidney 293 cells ( $\sim 5 \times 10^5$  per 6-cm dish) were transfected with 1  $\mu$ g of CMV- $\beta$ gal, 2  $\mu$ g of pGL2M4 [the vector pGL2 (Promega, Madison, WI) containing a fourfold reiteration of the sequence CACGTG] and the indicated amount of pRC/CMV-dMyc, pRC/CMV-dMax, and pRC/CMV vector to a total of 9  $\mu$ g. After 48 hours, luciferase activity was determined [F. M. Ausubel *et al.*, Eds., *Current Protocols in Molecular Biology* (Wiley, New York, 1995) Supplement 29, pp. 9.7.11-9.7.21] and normalized for expression from CMV- $\beta$ gal. The data represent duplicate samples from two independent experiments.
- Polyadenylated [poly(A)<sup>+</sup>] RNA was analyzed as in S. M. Parkhurst and V. G. Corces [*Cell* **41**, 429 (1985)]. For the developmental analysis shown in Fig. 3, wild-type flies were kept at 25°C.
- For RNA in situ hybridizations [D. Tautz and C.

- Pfeifle, *Chromosoma* **98**, 81 (1989)], digoxigenin-labeled probes were obtained by random-primer labeling of cDNA fragments comprising the *dmax* ORF, and nucleotides 1 to 2083 for *dmyc*, respectively.
22. J. Modolell, W. Bender, M. Meselson, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1678 (1983).
23. D. L. Lindsley and G. G. Zimm, *The Genome of Drosophila melanogaster* (Academic Press, New York, 1992).
24. R. L. Marlor, S. M. Parkhurst, V. G. Corces, *Mol. Cell. Biol.* **6**, 1129 (1986).
25. Genotypes of flies used in this study: wild-type refers to our isogenic Oregon R stock; RNA and genomic DNA were isolated from *Df(1)Pgd-kz/FM6, In(1)sc<sup>8</sup> y<sup>31d</sup> dm<sup>1</sup> B<sup>1</sup> and Tp(3;1)N<sup>264-6</sup>/y<sup>1</sup> w<sup>1</sup> dm<sup>1</sup>* flies; the latter line was also used for RNA in situ hybridization to ovaries. Both *dm<sup>1</sup>*-containing lines were obtained from the Bloomington stock center.
26. C. B. Moens, A. B. Auerbach, R. A. Conlon, A. L. Joyner, J. Rossant, *Genes Dev.* **6**, 691 (1992).
27. K. Nakayama *et al.*, *Cell* **85**, 707 (1996); H. Kiyokawa *et al.*, *ibid.* p. 721; M. L. Fero *et al.*, *ibid.* p. 733.
28. B. Mozer, R. Marlor, S. M. Parkhurst, V. Corces, *Mol. Cell. Biol.* **5**, 885 (1985).
29. We thank N. H. Brown, R. Davis, J. C. J. Eeken, S. Elledge, J. Kiger, C. D. Laherty, A. P. Mahowald, M. Meyer, G. Poortinga, J. W. Tamkun, C. Thummel, P. P. Tolias, M. Watanabe, and the Bloomington Stock Center for reagents; D. Gottschling, S. Henikoff, T. Neufeld, and J. Roberts for critical readings of the manuscript; P. Goodwin for help with image analysis; J. Torgerson for assistance with the manuscript; and members of the Parkhurst, Eisenman, and Edgar laboratories for advice and support. Funded by National Institutes of Health-National Cancer Institute (NCI) grants RO1CA57138 (to R.N.E.), and RO1GM47852 (to S.M.P.), a Postdoctoral Fellowship from the Fonds National Suisse (to P.G.) and an NCI Japanese Foundation for Cancer Research Training Program in the U.S.-Japan Cooperative Cancer Committee (to Y.S.).

23 September 1996; accepted 1 November 1996

## Association of Anxiety-Related Traits with a Polymorphism in the Serotonin Transporter Gene Regulatory Region

Klaus-Peter Lesch,\* Dietmar Bengel, Armin Heils, Sue Z. Sabol, Benjamin D. Greenberg, Susanne Petri, Jonathan Benjamin, Clemens R. Müller, Dean H. Hamer, Dennis L. Murphy

Transporter-facilitated uptake of serotonin (5-hydroxytryptamine or 5-HT) has been implicated in anxiety in humans and animal models and is the site of action of widely used uptake-inhibiting antidepressant and anti-anxiety drugs. Human 5-HT transporter (5-HTT) gene transcription is modulated by a common polymorphism in its upstream regulatory region. The short variant of the polymorphism reduces the transcriptional efficiency of the 5-HTT gene promoter, resulting in decreased 5-HTT expression and 5-HT uptake in lymphoblasts. Association studies in two independent samples totaling 505 individuals revealed that the 5-HTT polymorphism accounts for 3 to 4 percent of total variation and 7 to 9 percent of inherited variance in anxiety-related personality traits in individuals as well as sibships.

Anxiety-related traits are fundamental, enduring, and continuously distributed dimensions of normal human personality (1–3). Although twin studies have indicated that individual variation in measures of anxiety-related personality traits is 40 to 60% heritable (4), none of the relevant genes has yet been identified. Variance in personality traits, including those related to anxiety, is thought to be generated by a complex interaction of environmental and experiential

factors with a number of gene products involving distinct brain systems such as the midbrain raphe serotonin (5-HT) system (4). Neurotransmission mediated by 5-HT contributes to many physiologic functions such as motor activity, food intake, sleep, and reproductive activity, as well as to cognition and emotional states including mood and anxiety (5). By regulating the magnitude and duration of serotonergic responses, the 5-HT transporter (5-HTT) is central to the fine-tuning of brain serotonergic neurotransmission and of the peripheral actions of 5-HT. In the brain, 5-HTT expression is particularly abundant in cortical and limbic areas involved in emotional aspects of behavior (5). The human 5-HTT is encoded by a single gene (*SLC6A4*) on chromosome 17q12 (6–8). Although 5-HTT has long been suspected to play a role in behavioral and psychiatric disorders, previous studies did not reveal any common, replicated 5-HTT gene sequence variation in either

neuropsychiatric patients or healthy individuals (9).

Recently, we reported a polymorphism in the transcriptional control region upstream of the 5-HTT coding sequence (10). Initial experiments demonstrated that the long and short variants of this 5-HTT gene-linked polymorphic region (5-HTTLPR) had different transcriptional efficiencies when fused to a reporter gene and transfected into human placental choriocarcinoma (JAR) cells (10). The 5-HTTLPR is located ~1 kb upstream of the 5-HTT gene transcription initiation site and is composed of 16 repeat elements. The polymorphism consists of a 44-base pair (bp) insertion or deletion involving repeat elements 6 to 8 (Fig. 1A). In the present study, polymerase chain reaction (PCR)-based genotype analysis of 505 subjects revealed allele frequencies of 57% for the long (*l*) and 43% for the short (*s*) allele (11). The 5-HTTLPR genotypes were distributed according to Hardy-Weinberg equilibrium: 32% *l/l*, 49% *l/s*, and 19% *s/s*.

Because appropriate cell models for human serotonergic neurons do not exist and JAR cells are monozygotic for the 5-HTTLPR, we studied 5-HTT gene expression in human lymphoblastoid cell lines. Like 5-HT neurons and JAR cells, lymphoblasts constitutively express functional 5-HTT and exhibit adenosine 3',5'-monophosphate (cAMP)-dependent and protein kinase C (PKC)-dependent 5-HTT gene regulation, but they do not express dopamine or norepinephrine transporters (12). Cell lines with the complete range of different 5-HTTLPR genotypes can readily be obtained (13).

Lymphoblast cell lines with different genotypes were first transfected with constructs in which a luciferase reporter gene was fused to ~1.4 kb of the 5'-flanking promoter sequence containing the *l* or *s* form of the 5-HTTLPR (11, 13, 14). The basal activity of the *l* variant was more than twice that of the *s* form of the 5-HTT gene promoter (Fig. 1B). Stimulation of PKC by phorbol 12-myristate 13-acetate (PMA) or activation of adenylyl cyclase with forskolin-induced transcriptional activity was observed in both the *l* and *s* promoter variants, but the dose-dependent increases remained proportionally smaller in the *s* variant (Fig. 1B).

Although transfection experiments with reporter gene constructs are useful in assaying the transcriptional competence of a promoter sequence, they could conceivably give spurious results because of the absence of distant control elements or chromatin effects. Therefore, we next studied the expression of the native 5-HTT gene in lymphoblast cell lines cultured from subjects with different 5-HTTLPR genotypes (15). Cells homozygous for the *l* form of the

K.-P. Lesch, A. Heils, S. Petri, Department of Psychiatry, University of Würzburg, Fuchsleinstrasse 15, 97080 Würzburg, Germany.

D. Bengel, B. D. Greenberg, J. Benjamin, D. L. Murphy, Laboratory of Clinical Science, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA.

S. Z. Sabol and D. H. Hamer, Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA.

C. R. Müller, Institute of Human Genetics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany.

\*To whom correspondence should be addressed. E-mail: kplesch@rzbox.uni-wuerzburg.de