Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 5189–5194, May 1997
Genetics

An essential cell division gene of *Drosophila*, absent from *Saccharomyces*, encodes an unusual protein with tubulin-like and myosin-like peptide motifs

(orphan gene/transgenic organisms/protein structures/chromosomal nondisjunction)

GEORGE L. GABOR MIKLOS*†, M.-T. YAMAMOTO‡, R. G. BURNS§, AND R. MALESZKA¶

*The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121; ‡Department of Applied Biology, Kyoto Institute of Technology, Matsugasaki, Sakyo, Kyoto, Japan; §Biophysics Section, Imperial College of Science, Technology and Medicine, London SW7 2BZ, United Kingdom; and ¶Research School of Biological Sciences, Australian National University, Canberra, ACT 0200, Australia

Communicated by M. M. Green, University of California, Davis, CA, March 17, 1997 (received for review December 2, 1996)

ABSTRACT **Null mutations at the *misato* locus of *Drosophila melanogaster* are associated with irregular chromosomal segregation at cell division. The consequences for morphogenesis are that mutant larvae are almost devoid of imaginal disk tissue, have a reduction in brain size, and die before the late third-instar larval stage. To analyze these findings, we isolated cDNAs in and around the *misato* locus, mapped the breakpoints of chromosomal deficiencies, determined which transcript corresponded to the *misato* gene, rescued the cell division defects in transgenic organisms, and sequenced the genomic DNA. Database searches revealed that *misato* codes for a novel protein, the N-terminal half of which contains a mixture of peptide motifs found in α-, β-, and γ-tubulins, as well as a motif related to part of the myosin heavy chain proteins. The sequence characteristics of *misato* indicate either that it arose from an ancestral tubulin-like gene, different parts of which underwent convergent evolution to resemble motifs in the conventional tubulins, or that it arose by the capture of motifs from different tubulin genes. The *Saccharomyces cerevisiae* genome lacks a true homolog of the *misato* gene, and this finding highlights the emerging problem of assigning functional attributes to orphan genes that occur only in some evolutionary lineages.**

At the present time, approximately a third of the genes in both *Saccharomyces cerevisiae* and the archaebacterium *Methanococcus jannaschii* can be classified as orphans—i.e., either they have no known biochemical properties or they have no obvious relatives in protein and DNA databases (1, 2). Many other proteins have been assigned to families on the basis of a single domain, and hence their biochemical attributes are, at best, only partially catalogued. Extrapolation of these results to other organisms implies that

(*i*) combinatorial shuffling in different evolutionary lineages has produced a pool of orphan genes which characterize individual lineages—e.g., the immunoglobulin loci of vertebrates, or

(*ii*) some orphans may be more apparent than real; proteins in different phyla are known to have similar three-dimensional structures, but their relationships sometimes remain unrecognized owing to extensive divergence at the sequence level—e.g., human interleukin-1β and soybean trypsin inhibitor (3).

In this report we focus on a particular orphan gene in *Drosophila,* which we have named *misato* (beautiful dawn). It codes for a protein with a surprising combination of peptide motifs, each of which is characteristic of a different part of a conventional tubulin. It also shares a motif with myosin heavy chain proteins. The gene occurs in a region of the fly X chromosome that has been well studied genetically (4), and it is a representative of a class of genes termed "late lethals," a subset of which is almost diagnostic for essential cell cycle genes (5). Some mutant organisms carrying null alleles at this locus survive up until the early third-instar larval stage but are unable to complete morphogenesis because cell cycle defects lead to under-developed imaginal disks. The additional finding that the *S. cerevisiae* genome lacks a true homolog of this essential *Drosophila* gene highlights the limitations of comparative genomics. It reinforces the necessity for both conventional and unusual screens to analyze phenotypes in individual phyla (6–9) and the absolute necessity for crystallographic data at the protein level (10).

MATERIALS AND METHODS

Chromosome Preparations and Crossing Programs. Neuroblast cells from larvae of different genotypes were prepared as previously described without the use of colchicine (11). The lethal alleles tested at the *misato* locus (previously designated as *LB20*), were *LB20*, $LB20^{C27}$, $LB20^{DA618}$, and $LB20^{SK1}$ (ref. 4; M.-T.Y., unpublished work); the lethal allele tested at the *A112* locus was $A112^{17-62}$ (4). To determine the lethal stage of hemizygous or homozygous mutants, versus the lethal stage of mutant/*deficiency* individuals, $LB20^{SK1}/Y$ male larvae were compared with $LB20^{SK1}/JA117$ female larvae. To examine the mitotic consequences of mutations at the *misato* locus, and to distinguish the sex chromosomes more easily from the large autosomes, mutant male larvae of the genotypes *misato/0* and *misato/Y*, or their control sibs *FM7/0* and *FM7/Y*, were generated from crosses of the type *misato/FM7* females to $XY^L.Y^S$, y^2 *sc cv v f/0* males, or from *misato/FM7* females to *FM7/Y* males. Male larvae of the genotypes *FM7/0* and *FM7/Y* have white malpighian tubules and can thus be readily distinguished from their mutant *misato/0* and *misato/Y* sibs, which have yellow malpighian tubules. The mutants develop very slowly and die in the second or early third larval instars. They have almost no imaginal disk tissue and have significantly reduced neuroblast tissue and a nearly transparent body.

***P*-Element Constructs and Germ-Line Transformations.** Four parts of the *misato* region (denoted transforming fragments, A, B, C, and D; see Fig. 1) were cloned into the vector pW8 by previously described procedures (12–15) and used for transgenic analysis. Construct A consists of a 10.9-kb *Bam*HI genomic fragment; construct B is a 4.8-kb *Spe*I–*Pst*I fragment; construct C is an 11.6-kb *Bgl*II fragment; and construct D is a

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/945189-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. U80043).
†To whom reprint requests should be addressed.

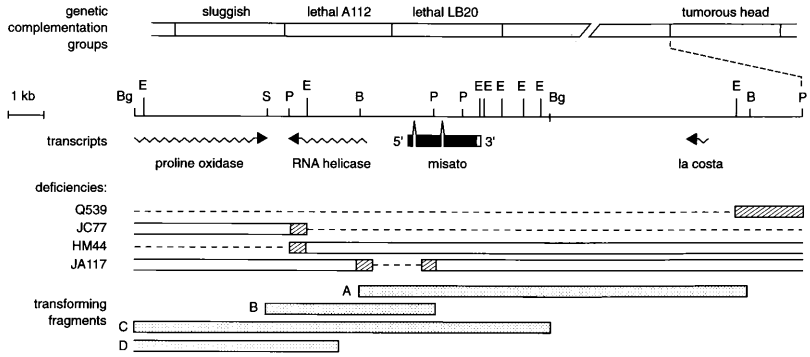

FIG. 1. Genetic and molecular properties of genes, chromosomal deficiencies, and transforming fragments in the *misato* region. The genetic complementation groups within the region are as shown. The four transcription units are aligned on the genomic DNA; the intron–exon structure of only *misato* is shown. Restriction sites are E, *Eco*RI; P, *Pst*I; Bg, *Bgl*II; S, *Spe*I; and B, *Bam*HI. The breakpoints of the four chromosomal deficiencies are indicated by hatched boxes. The four transforming fragments A, B, C, and D are as indicated.

15.3-kb *Eco*RI–*Xho*I genomic fragment used previously to rescue the *sluggish* gene. Germ-line transformations using constructs A, C, and D were carried out as previously described (12–14). Germ-line transformation using construct B was implemented using embryos of strain *yw*; +/+; *Sb, P*[*ry*$^{+}$, Δ*2–3*]/*TM6, Ubx* (15). Three independent genomic insertions were obtained for fragment A, six for fragment B, three for fragment C, and three for fragment D.

Molecular Protocols. DNA and RNA extractions, blotting, hybridization, molecular isolation and cloning of cDNAs from standard *Drosophila* cDNA libraries, and genomic and cDNA sequencing were carried out as previously described (12–17). A number of *misato* cDNAs were incomplete at their 5′ and 3′ ends as determined from comparison to our genomic sequence. Reverse transcription (RT)–PCR analyses and the isolation of longer cDNAs (D. Slifka, A. B. Kasprzak, J. Cotsell, D. Hall, H. Campbell, and G.L.G.M., unpublished work), confirmed the position of the ATG codon and the position of the first intron, which was inferred from bioinformatic analysis of our genomic sequence data (S. Karlin, R. Maleszka, and G.L.G.M, unpublished work). Genomic and cDNA sequences were determined on both strands using Sequenase (United States Biochemical) and by manual as well as automated methods using the Pharmacia ALF DNA Sequencer. Database searches were carried out at the National Center for Biotechnology Information (NCBI) by employing the BLAST network service. CLUSTAL W and MACAW were used to generate multiple sequence alignments.

RESULTS AND DISCUSSION

Genetic and Molecular Analysis of the *misato* Region. The molecular genetics of the chromosomal segment under study is shown in Fig. 1. This region contains two essential loci, denoted lethals *A112* and *LB20,* and is flanked distally by the *sluggish* locus and proximally by the *tumorous head* locus, neither of which is essential for organismal viability. These loci have all been ordered by deficiency analysis. Deficiency *Q539* uncovers the mutant phenotypes of *sluggish*, *A112*, and *LB20*, but not *tumorous head*; deficiency *JC77* uncovers *A112*, *LB20*, and *tumorous head*; deficiency *HM44* uncovers *sluggish* and *A112*; and deficiency *JA117* uncovers both *A112* and *LB20* (4).

We found four transcription units in this region, the most distal of which, *sluggish*, codes for proline oxidase, an enzyme involved in glutamate biosynthesis (14). The next transcription unit codes for a member of the DEAD-box family of ATP-dependent RNA helicases (13), and we have denoted it *Helicase* (*Hlc*). This is followed by the *misato* transcription unit and then by a transcription unit which we have called *la costa (lcs)*, because it contains a *l*eader sequence, a *c*ollagen-related repetitive motif, and a *s*erine/threonine-rich motif found in many different proteins.

To align the genetic and molecular maps, we (*i*) determined the approximate molecular breakpoints of the four deficiencies; (*ii*) isolated all available cDNAs from the 35-kb region; (*iii*) sequenced the genomic DNA except for a small region near the proximal breakpoint of deficiency *Q539*; and (*iv*) analyzed the phenotypic consequences of transgenic organisms carrying one each of the wild-type portions of the regions denoted A, B, C, and D (see *Materials and Methods*).

The breakpoints of the four chromosomal deficiencies were determined by Southern analysis of the genomic DNA of individuals heterozygous for rearranged and normal chromosomes, using radiolabeled probes from across the entire region (data not shown). We found that deficiency *Q539* has its breakpoint in the most proximal *Eco*RI–*Pst*I fragment (Fig. 1), and this breakpoint anchors the centromeric end of the region, without interrupting any transcription unit. Deficiencies *JC77* and *HM44* both break in a 0.5-kb *Pst*I–*Eco*RI fragment, but the deficiencies extend in opposite directions. At the molecular level, *JC77* interrupts the *Helicase* transcription unit and deletes *misato* and *la costa. HM44* also interrupts the *Helicase* transcription unit, and it deletes the *sluggish* transcription unit and further genes distal to this breakpoint. Unlike deficiencies *Q539, JC77*, and *HM44*, which extend for hundreds of kilobases, deficiency *JA117* is a small deletion of approximately 1.5 kb which largely removes the 5′ regions of the *Helicase* and *misato* transcription units (Fig. 1). These deficiency-based data, in combination with the positions of the four transcription units and the genetic data described above, make it likely that the *Helicase* transcription unit corresponds to the lethal *A112* locus and that the *misato* transcription unit corresponds to the lethal *LB20* locus.

Rescue of the Mutant Phenotypes Associated with Alterations in the *misato* Region. To obtain further evidence that the *Helicase* and *misato* transcription units correspond to the *A112* and *LB20* loci, respectively, we constructed transgenic organisms that carried different combinations of mutant alleles, deficiencies, and transforming fragments for the region (Fig. 1). Transgenic organisms containing a single copy of each of these fragments were crossed to individuals carrying mutant alleles at the *A112, LB20*, and *sluggish* loci, as well as to individuals carrying deficiency *JA117*, which interrupts both the *Helicase* and *misato* transcription units.

Three independent transgenic lines containing genomic fragment A, (which carries wild-type *misato* and *la costa* transcription units), each rescue the lethality caused by mutations at the *LB20* locus, namely the mutations *LB20* and *LB20*^{C27}. However, fragment A-containing individuals do not rescue a mutation at the *A112* locus, namely the mutation

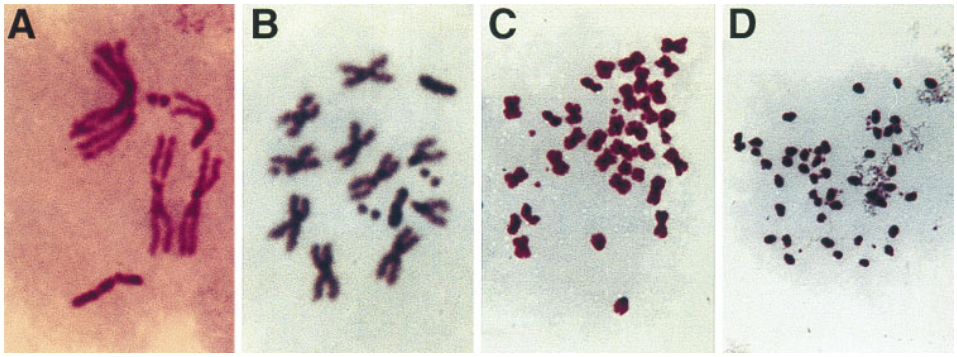

FIG. 2. Chromosome complements in neuroblast tissue of mutant *misato* males of the genotype *LB20/Y*. (*A*) Diploid cell. (*B*) Tetraploid cell. (*C* and *D*) Aneuploid complements with hypercondensed chromosomes. (All panels are at the same magnification.)

$A112^{17-62}$, nor do they rescue the inviability of deficiency *JA117/Y* males.

The two tested transformant lines carrying fragment B (which carries a wild-type *Helicase* transcription unit) both rescue the lethality caused by a mutation at the *A112* locus (namely the mutation $A112^{17-62}$). However, neither of these transformant lines rescues the mutations at the adjacent *LB20* locus, namely the mutations *LB20,* $LB20^{C27}$, and $LB20^{SK1}$.

All three transformants carrying fragment C (which have wild-type copies of both the *Helicase* and *misato* transcription units), rescue the lethality of individuals carrying mutations at the *A112* and *LB20* loci. Fragment C-bearing transformants also rescue the lethality of deficiency *JA117/Y* males, a deficiency that uncovers lethal alleles at both the *A112* and *LB20* loci.

All three transformant lines carrying fragment D (which carries a wild-type *sluggish* transcription unit) rescue the *sluggish* phenotype but not the lethality caused by a mutation at the adjacent *A112* locus, namely the $A112^{17-62}$ allele.

These transgenic, mutant, and deficiency combinations provide strong evidence that the *A112* and *LB20* loci correspond to the *Helicase* and *misato* transcription units, respectively, and henceforth we refer to them by these designations. Finally, since deficiency *Q539* is not deficient for the *tumorous head* gene, the *la costa* transcription unit is unlikely to correspond to the *tumorous head* locus.

Cell Biology of Null Alleles at the *misato* Locus. In attempting to determine the primary defect caused by lesions in the gene, we carried out germ-line clone analysis of three *misato* alleles (*LB20,* $LB20^{C27}$, and $LB20^{DA618}$), and we found that germ cells are unable to survive when homozygous for these alleles (4). The *misato* protein is clearly an essential one for normal functioning of the female germ line. We also examined homozygous mutant *misato* embryos derived from heterozygous *misato*/wild type mothers, and we found that none of the mutant embryos had any central or peripheral nervous system defects (4). It is likely therefore, that a significant maternal contribution of wild-type product from heterozygous mothers allows homozygous mutant embryos to develop apparently normally and that the lethality in the larval instars occurs as a result of the depletion of wild-type maternal product. Furthermore, the stage of lethality is similar whether larvae are homozygous for mutant *misato* alleles or are heterozygous for a mutant allele and deficiency *JA117*. Therefore the alleles used in this study are likely to be null, or nearly so.

In looking hard at the phenotype of mutant larvae, we found that they had very little imaginal disk tissue and significantly reduced neural tissue, a finding considered almost diagnostic for essential cell cycle genes (5). We therefore carefully examined cell division, chromosome numbers, and morphologies in larval brains of mutant *misato* males, and we found that more than 60% of the cells had aneuploid or polyploid chromosome complements. These findings are illustrated in Fig. 2: a normal diploid XY cell is shown in *A*, a tetraploid cell (4A XXYY) is shown in *B*, and cells with aneuploid complements, in which the chromosomes are typically hypercondensed, are shown in *C* and *D*. Except in cases of extreme chromatin condensation, we were able to score for the presence of the dot-like chromosome 4s, the X and Y chromosomes, and the number of large autosomes in a cell.

The distribution of chromosome numbers per cell in mutant individuals is illustrated in Fig. 3. Our data from 1,538 cells reveal that minimally 60% of the cells have aneuploid chromosome complements, and that the distribution of chromosome numbers is highly skewed above the diploid value. Any cell deficient for either an X chromosome or both copies of an autosome will be inviable. We presume that those cells that have a normal chromosome complement still retain sufficient maternal product for normal chromosomal segregation.

We also find from a detailed analysis of the chromosomal constitutions of individual cells that chromosome number (for any given chromosome, X, Y, 2, 3, or 4) increases approximately linearly with the number of chromosomes in a cell. Thus the absence of the normal *misato* product does not differentially affect the probability of nondisjunction of autosomes versus sex chromosomes, or of small versus large chromosomes. Finally, examination of metaphase and anaphase figures reveals that cytokinesis is normal and that we are dealing

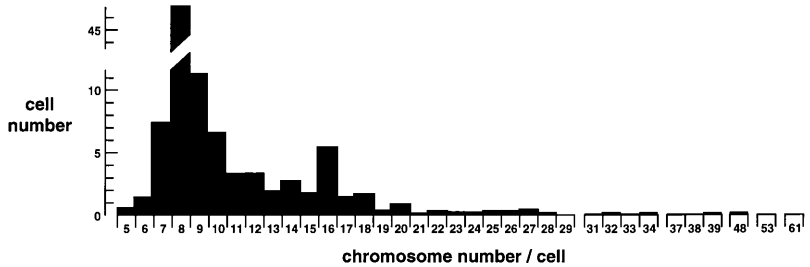

FIG. 3. Histogram of the variation in chromosome number in neuroblast cells from second- and early-third-instar mutant *misato* males. Pooled data from *LB20/Y*, $LB20^{C27}/Y$, $LB20^{DA618}/Y$, and *LB20/O*.

with a mitotic defect. The mutant phenotype at the cellular level is thus similar to a class of mutants affecting chromosome condensation previously described in *Drosophila* (5), but the underlying processes responsible for these condensation defects are unknown.

We also examined cell division in larval brains of transgenic organisms that carried a wild-type copy of *misato* on an autosome, together with a lethal allele of *misato* (*LB20*) on the X chromosome. These organisms exhibit normal cell division, have normal-sized imaginal disks and normal brain tissues, and reveal no evidence of aneuploidy (data not shown). It is clear that the consequences of being mutant at the *misato* locus are irregular mitotic segregation and hypercondensed chromatin, ultimately leading to cell death and the inability to complete morphogenesis. What, then, are some of the potential properties of the gene product that lead to these phenotypes?

The *misato* Gene and Its Predicted Protein. We sequenced a number of *misato* cDNAs as well as the underlying genomic DNA, aligned the cDNAs on the genomic sequence, and deposited the cDNA and genomic sequences in public databases. The alignment revealed that *misato* contains two small introns of 55 and 60 bp. Database searches showed that the top 50 matches to the predicted *misato* product were tubulins, with sequence similarities being highest in the N-terminal half of *misato*, in two regions that we have have designated as segments I and III (Fig. 4*a*). In addition, database searches revealed a region with excellent sequence similarity to a motif found in myosin heavy chain proteins, which we have denoted segment II (Fig. 4 *a* and *d*).

A comparison of the predicted *misato* product with that of the highest scoring database relative, namely a β-tubulin from *Dictyostelium discoideum*, reveals a number of regions of excellent sequence similarity. In addition, when segment I of *misato* is compared with γ-tubulins from *Zea mays* and humans, there is 60% sequence similarity and nearly 50% amino acid identity (Fig. 4*b*).

When segment III of *misato* is compared with α-tubulins from *D. discoideum* and *S. cerevisiae*, there is 60% sequence similarity and 44% sequence identity (Fig. 4*c*). While the sequence similarity of segment III to the α-tubulins is striking, nothing is known of the properties of this segment in either the α- or β-tubulins. It lies between two highly conserved peptides which are presumptive components of the β-tubulin GTP-binding site, namely KGHYTEG and GGGTGSG (10), yet neither of these motifs is conserved within the *misato* sequence. This difference between *misato* and the β-tubulins extends to a peptide lying within segment I. In this tubulin peptide, a highly conserved β-tubulin cysteine (Cys-12) can be photocrosslinked to the guanine base of the bound GTP (10), yet this cysteine is not present in segment I of *misato*, and the marked corruption of the KGHYTEG and GGGTGSG motifs indicates that *misato* is not a GTP-binding protein. In addition, the MREI motif in segment I, which is known to contribute to the autoregulation of tubulin expression, is disrupted in the *misato* sequence, as it is in the γ-, δ-, and ε-tubulins. The δ-tubulin of *Caenorhabditis elegans*, for example, is missing the MREI motif altogether (18).

Segment II of *misato* is a major contributor to the difference in length between it and the various tubulins (Fig. 4*a*). The most significant sequence similarities to this 54-amino acid region are from a motif found in the heavy chains of myosins from different organisms (Fig. 4*d*). A comparison of segment II with the vertebrate myosin heavy chains reveals that it is homologous to a myosin peptide in the hinge region linking the S2 and LMM domains. Segment II also contains heptad repeats which are characteristic of the myosin tail α-helical coiled-coils. However, unlike the myosins, the *misato* sequence contains three helix-breaking proline residues, which are positioned toward the N and C termini and at the middle of segment II (Fig. 4*d*). This suggests that the tertiary structure of segment II consists of two short antiparallel α-helices separated by a proline-containing turn, and that they interact to form a short coiled-coil which projects from the tubulin-like domain.

The N-terminal half of the *misato* protein thus has a curious mixture of peptide motifs held in common with, and in the same order as, the conventional tubulins. This analysis is based upon a database of all available tubulins, now totalling approximately 400 sequences (R.G.B, unpublished work). This database has been inspected to determine the fraction of

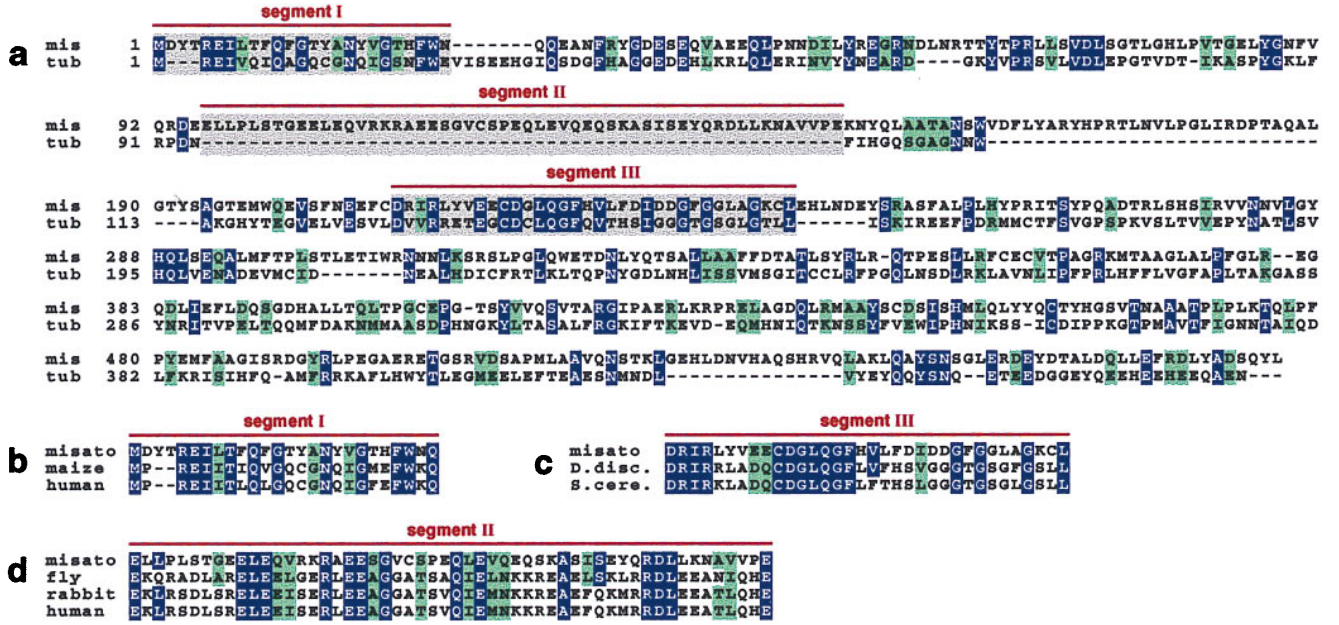

FIG. 4. Amino acid sequence alignments. (*a*) *misato* and β-tubulin from *Dictyostelium discoideum* (SwissProt accession no. SP P32256). (*b*) Segment I, N-terminal portion of *misato*, maize γ-tubulin [Protein Identification Resource (PIR) S43329], and human γ-tubulin (SP P23258). (*c*) Segment III of *misato* and α-tubulin from *D. discoideum* (SP P32225) and *S. cerevisiae* (PIR S50871). (*d*) Segment II of *misato* and myosin heavy chain sequences from *D. melanogaster* (PIR C35815), rabbit (SP P04460), and human (PIR S12458). Identical residues are in solid boxes, and conservative changes are in lightly shaded boxes.

sequences with a specific amino acid at each residue position, and the *misato* sequence has been aligned with the α-, β-, and γ-tubulins. Plots of the summed fractional values at each residue position against the corresponding *misato* residue reveal the regions with the highest sequence similarities. These are all in the N-terminal 280 amino acids of each of the three tubulin sequences, and they are separated either by *misato* inserts or by peptides in which the *misato* sequence shares only weak sequence similarity with the tubulin sequences. By contrast, the C-terminal sequence of *misato* exhibits a weak sequence similarity of approximately 15% to the conventional tubulins.

Putative *misato* Relatives from Other Organisms. We have searched the complete *S. cerevisiae* genome for relatives of *misato* at the sequence level and have found five predicted proteins, parts of which yield significant similarity scores. Four of these proteins are tubulins, namely TUB1, TUB2, TUB3, and TUB4, whereas the fifth, designated YM8261.05, is a predicted open reading frame which has weak sequence similarity to members of the tubulin superfamily. Since the BLAST score for YM8261.05 is higher than for each of the tubulins, we aligned *misato* and YM8261.05 to determine the regions of greatest sequence similarity. This alignment of the 574-amino acid *misato* sequence and the 479-amino acid YM8261.05 sequence is shown in Fig. 5. The highest density of identical residues is in the N-terminal half, in segments I, II, and III. By contrast, the C-terminal halves of these two proteins are totally unrelated, and in addition, there are no other predicted proteins in this yeast that contain both tubulin- and myosin-like motifs. We conclude that *misato* and YM8261.05 are not true homologs, and that *S. cerevisiae* lacks a comparable protein that contains both tubulin and myosin motifs. The presence of the tubulin motifs in both *misato* and YM8261.05 may indicate that the two predicted proteins share some functional properties. It is also possible that the *S. cerevisiae* genome has been streamlined by deletion, one of the losses being that of a *misato* homolog. Finally, it is entirely possible that *misato* is an invention restricted to some multicellular lineages. If this is the case, then appropriate information may come from other metazoa.

We have found that four human expressed sequence tags (ESTs) show excellent sequence similarity to *misato* at its N terminus (R17341, H12093, and Z43579). They are 63% similar and 46% identical at the amino acid level, and their alignments involve no gaps. The fourth human EST (N23629) is 46% similar and 32% identical, again without gaps, and its sequence extends into segment II of *misato*. These ESTs provide a starting point for analyses of human genes that may belong to the *misato* family.

The *misato* transcript is a relatively rare one as assayed by Northern blotting of embryos, larvae, pupae, and adults, and our reverse transcriptase–PCR analysis reveals that the highest levels of this message are in the female germ line. So far, however, we have been unsuccessful in producing antibodies to the *misato* protein and analyzing its cellular distribution, and hence in carrying out a phenotypic analysis in greater depth. Further transgenic and comparative analysis of *misato* relatives in vertebrates, using the available human ESTs, for example, may yield alternative approaches to assaying its biological significance, since the overlap at the gene level between *Drosophila* and vertebrates is large.

Homologies and Multifunctionalities. The major stumbling block to understanding the properties of the tubulins, and in particular their tertiary structures, has been the inability to crystallize them (18, 19). This is compounded by the findings that the folding of the nascent β-tubulin molecule requires a multisubunit chaperonin, TCP1, as well as additional cofactors. Hence, in the absence of crystallographic data, almost all comparisons between tubulin and tubulin-like molecules have depended on site-directed mutagenesis and crosslinking analyses. To date, however, the mutagenesis approach has not yielded potential peptides that interact with the guanine base, and the crosslinking results have yet to provide a coherent picture of the properties of β-tubulin (10). Without crystallographic structures, inferences from sequence comparisons, site-directed mutagenesis, and crosslinking studies are limited.

In the case of sequence comparisons, for example, *misato* has a higher overall sequence similarity to tubulins than does the bacterial FtsZ protein, which has been proposed to be a tubulin homolog on the grounds that it is a GTPase, forms microtubule-like structures, and has some sequence similarity to tubulins (20). However, some enzymes can have properties strikingly similar to those of cytoskeletal proteins, even though they are not thought to be related to tubulins in an evolutionary sense. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) has as much sequence similarity to tubulins as does FtsZ, and GAPDH can bind GTP, actin filaments, microtubules, and colchicine. Furthermore, GAPDH can polymerize to form cytoskeleton-like oligomers (21). These data highlight the difficulties of assigning homologies, *sensu strictu*, to one of the best-characterized cytoskeletal protein families (22). They also put into perspective the emerging multifunctionality of proteins that have so far been considered to have only a single conserved property.

The existence of the *misato* gene, with its intriguing combination of motifs and potential secondary structures, highlights the paucity of functional data on unconventional proteins that go undetected when relatives are sought by means of hybridization, PCR assays, or database searches. The finding that the *S. cerevisiae* genome lacks a true *misato* homolog demonstrates that this yeast does not automatically provide the entire foundation for cell division processes in the metazoa, and it has yet to be demonstrated that cell division processes

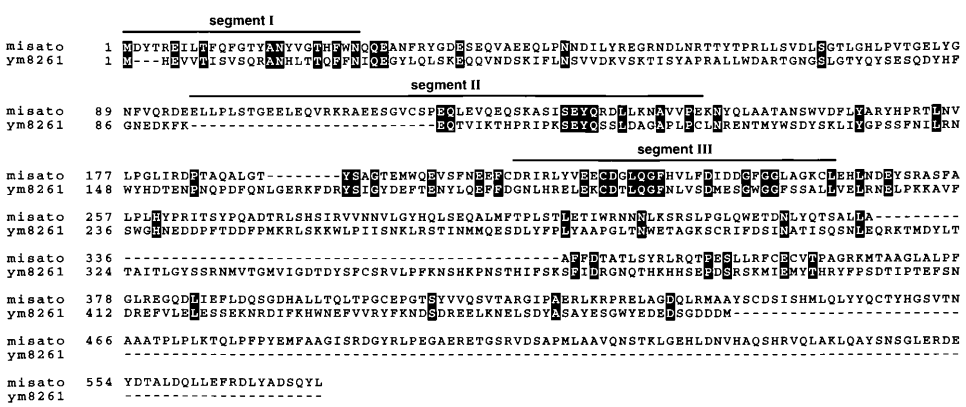

FIG. 5. Amino acid sequence alignment of *misato* and the YM8261.05 predicted protein of *S. cerevisiae.* The extents of segments I, II, and III are the same as in Fig. 4, and only the identical residues are highlighted.

in this yeast are representative of those in other fungi. In addition, it is already clear that *S. cerevisiae* is missing other genes known to be present in *Drosophila* and in *Homo* (S. Karlin, R. Maleszka, and G.L.G.M., unpublished work). While the genome project data from the model organisms and humans will reveal the highly conserved core of proteins in different metazoan phyla (23), it is probable that the crystallographic analysis of orphan proteins will be essential to understanding the origins of evolutionary novelties in different lineages.

We thank J. Cotsell, D. Hall, M. Kai, A. Kasprzak, J. Mason, and D. Slifka for technical help and H. G. de Couet, K. Crossin, G. M. Edelman, E. Raff, and S. Karlin for advice. G.L.G.M. was funded by Neurosciences Research Foundation and R.M. by the Australian National University.

1. Dujon, B. (1996) *Trends Genet.* **12,** 263–270.
2. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., *et al.* (1996) *Science* **273,** 1058–1073.
3. Priestle, J. P., Schar, H.-P. & Grutter, M. G. (1988) *EMBO J.* **7,** 339–343.
4. Perrimon, N., Smouse, D. & Miklos, G. L. G. (1989) *Genetics* **121,** 313–331.
5. Gatti, M. & Baker, B. S. (1989) *Genes Dev.* **3,** 438–453.
6. Simon, M. A., Bowtell, D. D. L., Dodson, G. S., Laverty, T. R. & Rubin, G. M. (1991) *Cell* **67,** 701–716.
7. Matthews, K. A., Rees, D. & Kaufman, T. C. (1993) *Development (Cambridge, U.K.)* **117,** 977–991.
8. Edwards, K. A., Montague, R. A., Shepard, S., Edgar, B. A., Erikson, R. L. & Kiehart, D. P. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 4589–4593.
9. Fackenthal, J. D., Hutchens, J. A., Turner, F. R. & Raff, E. C. (1995) *Genetics* **139,** 267–286.
10. Burns, R. G. & Farrell, K. W. (1996) *Trends Cell Biol.* **6,** 297–303.
11. Yamamoto., M.-T., Mitchelson, A., Tudor, M., O'Hare, K., Davies, J. A. & Miklos, G. L. G. (1990) *Genetics* **125,** 821–832.
12. Delaney, S. J., Hayward, D. C., Barleben, F., Fischbach, K.-F. & Miklos, G. L. G. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7214–7218.
13. Schuppler, U. (1992) Ph.D. thesis (Australian National Univ., Canberra, Australia).
14. Hayward, D. C., Delaney, S. J., Campbell, H. D., Ghysen, A., Benzer, S., Kasprzak, A. B., Cotsell, J. N., Young, I. G. & Miklos, G. L. G. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 2979–2983.
15. Maleszka, R., Hanes, S. D., Hackett, R. L., de Couet, H. G. & Miklos, G. L. G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 447–451.
16. Campbell, H. D., Schimansky, T., Claudianos, C., Ozsarac, N., Kasprzak, A. B., Cotsell, J. N., Young, I. G., de Couet, H. G. & Miklos, G. L. G. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11386–11390.
17. de Couet, H. G., Fong, K. S. K., Weeds, A. G., McLaughlin, P. J. & Miklos, G. L. G. (1995) *Genetics* **141,** 1049–1059.
18. Burns, R. G. (1995) *Cell Motil. Cytoskeleton* **31,** 255–258.
19. Burns, R. G. (1995) *J. Cell Sci.* **108,** 2123–2130.
20. Erickson, H. P., Taylor, D. W., Taylor, K. A. & Bramhill, D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 519–523.
21. Gupta, R. S. & Soltys, B. J. (1996) *Biochem. Mol. Biol. Int.* **38,** 1211–1221.
22. Burns, R. G. (1996) *Trends Cell Biol.* **6,** 376.
23. Miklos, G. L. G. & Rubin, G. M. (1996) *Cell* **86,** 521–529.