MOLECULAR AND CELLULAR BIOLOGY, Dec. 1988, p. 5206–5215
0270-7306/88/125206-10$02.00/0

Vol. 8, No. 12

The Brown Protein of *Drosophila melanogaster* Is Similar to the White Protein and to Components of Active Transport Complexes

THOMAS D. DREESEN,* DANIEL H. JOHNSON,† AND STEVEN HENIKOFF
Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104

Received 28 June 1988/Accepted 6 September 1988

The brown gene of *Drosophila melanogaster* is required for deposition of pteridine pigments in the compound eye and other tissues. We isolated a ca. 150-kilobase region including brown by microdissection and chromosome walking using cosmids. Among the cDNAs identified by hybridization to the cosmids, one class hybridized to a genomic region that is interrupted in two brown mutants, *bw* and In(2LR)CK, and to 2.8- and 3.0-kilobase poly(A)$^+$ RNAs which are altered in the mutants. Nucleotide sequencing of these cDNAs revealed that the two transcripts differ as a consequence of alternative poly(A) addition and that both encode the same predicted protein of 675 amino acids. Searches of available databases for amino acid sequence similarities detected a striking overall similarity of this predicted protein to that of the *D. melanogaster* white gene. The N-terminal portion aligned with the HisP family of membrane-associated ATP-binding proteins, most of which are subunits of active transport complexes in bacteria, and to two regions of the multidrug resistance P-glycoprotein. The C-terminal portion showed a structural similarity to integral membrane components of the same complexes. Taken together with earlier biochemical evidence that brown and white gene products are necessary for uptake of a pteridine precursor and genetic evidence that brown and white proteins interact, our results are consistent with suggestions that these proteins are subunits of a pteridine precursor permease.

The screening pigments of the dipteran compound eye are present to optically isolate each ommatidium (facet) from its neighbor. In *Drosophila* species, there are two major classes of screening pigments: ommochromes, which are brown, and pteridines, which include the red and yellow pigments. The ommochromes are found in the primary pigment cells near the surface of the eye and in the secondary pigment cells which surround the photoreceptors, while the pteridines are found in the secondary pigment cells (46). The two pigment classes are biochemically distinct, as they are derived from unrelated precursors: the ommochromes from 3-hydroxykynurenine and the pteridines from GTP.

Numerous mutations are known that affect the production or deposition of either or both of the eye pigments (for a review, see reference 46). For example, the eyes of white (*w*) null mutants contain neither of these pigments. The eyes of scarlet (*st*) null mutants lack the ommochromes, and those of brown (*bw*) null mutants lack the pteridines.

In this paper, we report the isolation, identification, transcriptional analysis, and sequencing of cDNA of the brown gene. Two very rare mRNAs were found to differ by alternative sites of poly(A) addition. Both were disrupted in null mutants. A computer search of protein and DNA databases revealed homology of the predicted brown protein with the white protein and the HisP family of ATP-binding proteins, many of which are components of transmembrane active transport complexes (19). (We refer to the family of ATP-binding subunits of the binding-dependent protein active transport system simply as the HisP family to avoid specific implications of mechanism for this diverse group of proteins.) Our work strongly supports the hypothesized role of brown in the uptake of pteridine precursors (44) and the hypothesized interaction of brown and white based on genetic evidence (7, 8).

* Corresponding author.
† Present address: Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, FL 33101.

MATERIALS AND METHODS

***Drosophila* strains.** The mutations and chromosomes utilized in this study are described by Lindsley and Grell (29) except for In(2LR)CK, 59E2;30A, *b cn bw*^{85f4}, which was kindly provided by D. Nash. It is a homozygous, viable, X-ray-induced brown null mutation associated with an inversion of the second chromosome with breaks at 59E2 and 30A.

Microdissection. The 59E1-2 region of salivary gland polytene chromosomes was excised and microcloned utilizing bacteriophage λ607 essentially as described by Scalenghe et al. (39) and Pirrotta et al. (35).

In situ hybridization. In situ hybridization of DNA probes biotinylated by nick translation was carried out on salivary gland polytene chromosomes as described by Simon et al. (40).

Chromosome walk. Restriction fragments were purified on agarose gels (27), labeled by nick translation, and used to probe a *Drosophila* total genomic DNA library in the cosmid vector cosPneo (43) (kindly provided by M. Champe and C. D. Laird). The direction and extent of each step in the walk was assayed by in situ hybridization to Amherst (wild type) and In(2LR)CK, 59E2;30A, *b cn bw*^{85f4} salivary gland polytene chromosomes. To obtain consistently strong hybridization signals, we used probes of 50 to 100 nucleotides (16).

Isolation of RNA and DNA. Total RNA was isolated from the heads of newly eclosed (<24 h) flies by the guanidine isothiocyanate-CsCl method (30). Poly(A)$^+$ RNA was purified by a batch method utilizing oligo(dT)-cellulose (Sigma Chemical Co., St. Louis, Mo.).

Total genomic DNA was isolated from adult flies by the high salt-protease method (25). Restricted DNA fragments were purified from low-melting-temperature agarose gels by the quaternary ammonium salt-butanol method (27).

Northern and Southern gel analysis. Southern blots were prepared and probed by the method of Wahl and co-workers (49). Northern (RNA) blots were prepared with glyoxylated

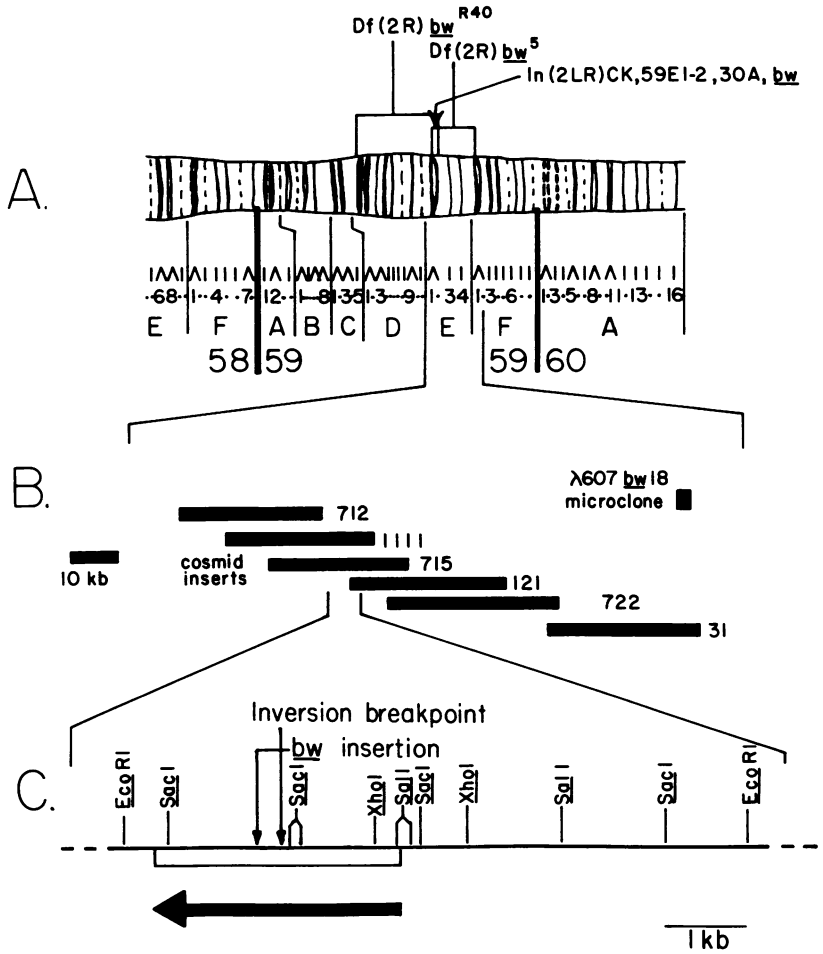

FIG. 1. Cytological and restriction maps of the brown locus. (A) Chromosome arm 2R after Bridges (see reference 29). Brackets indicate the approximate regions removed by the indicated deficiencies. The arrow at 59E2 indicates the right breakpoint of the *bw* inversion mutant. (B) Genomic inserts used in this study. (C) Restriction map of the brown locus. The transcription unit is indicated by the open rectangle, and the direction of transcription is indicated by the arrow. The approximate locations of mutant disruptions in the transcription unit are indicated by the vertical arrows.

RNA by the method of Thomas (48). Hybridization probes were either nick-translated or randomly primed double-stranded DNA (9), using [^{32}P]dATP at a specific activity of 3,000 Ci/mmol (Dupont, NEN Research Products, Boston, Mass.).

Isolation and sequencing of cDNAs. A cDNA library in λgt10 made from late *Drosophila* pupae (36) and a head-specific cDNA library in λZap (prepared by Stratagene Cloning Systems and kindly supplied by B. Tempel) were screened with nick-translated gel-purified DNA fragments. The resulting positive clones were further screened by restriction enzyme and genomic Southern analysis. cDNA inserts were subcloned in both orientations into pVZ1, a phagemid vector (15). The apparently full-length insert, c914, was sequenced on both strands by the dideoxy method (37), utilizing full-length clones and targeted deletions generated by exonuclease III (14). Other cDNA inserts were sequenced at their 3′ ends by the dideoxy method. Cloning methods not specifically described followed the procedures of Maniatis et al. (30).

Computer methods. The open reading frame of c914 was used as query sequence in searches of NBRF-PIR 15, GenBank 54.0, and EMBL 14 with Genepro 4.1 software (Riverside Scientific Enterprises, Seattle, Wash.) as described previously (17). The DNA databases were translated "on-the-fly" and searched with a window of 90. Alignments were performed by using Genepro, followed by minor manual adjustment. Hydropathy plots were done with Genepro by the algorithm of Kyte and Doolittle (26), using a window of 12.

RESULTS

Molecular cloning of genomic DNA and cDNA sequences of the brown locus. The brown gene has been localized cytologically to the 59E1-2 region on 2R by failure of *bw* mutations to complement the eye color phenotype of two deficiencies, Df(2R)*bw*^{R40} (41), which deletes 59D1 to 59E1-2, and Df(2R)*bw*^{5} (29), which deletes 59E1-2 to 59F1 (Fig. 1A). Therefore, to isolate brown genomic DNA, we microcloned the vicinity of 59E1-2 from salivary gland polytene chromosomes. Of the approximately 1,000 recombinant clones obtained, 12 were randomly selected for further analysis. DNA from these bacteriophage was examined by in situ hybridization. One microclone, λ607*bw*18, which labeled the centromere-proximal side of 59F1 (data not shown) was chosen as the starting point for a chromosome walk (Fig. 1B). The insert of λ607*bw*18 was used to screen a

genomic cosmid library of total *Drosophila* DNA. The direction and extent of each step in the walk was determined by in situ hybridization to both Amherst (wild type) and In(2LR)CK, 59E2;30A, *b cn bw*^{85f4} (brown inversion) salivary gland polytene chromosomes. Cosmid DNA of cPn1111 obtained from step 4 of the walk was hybridized in situ to salivary chromosomes of the brown inversion strain. Two bands of staining were observed, one at each of the two breakpoints of the inversion (Fig. 2A). Only one band is seen at 59E in wild-type chromosomes (data not shown). DNA from cPn1111 was used to probe late pupal and head-specific cDNA libraries for cDNAs from the area around the brown inversion breakpoint. The 30 cDNAs isolated could be divided into four groups based on restriction enzyme and Southern analysis.

To determine whether the coding region represented by any of the cDNA inserts was disrupted by the inversion, we hybridized the largest of each group in situ to brown inversion salivary gland chromosomes. Only one, c914, exhibited bands of hybridization at both of the breakpoints (Fig. 2B). Hybridization of a cDNA probe to wild-type salivary gland chromosomes demonstrated that this class of cDNAs is derived from 59E2 (Fig. 2C). To confirm that the coding region represented by c914 was interrupted by the inversion, we conducted genomic Southern analysis using ^{32}P-labeled c914 in vector pVZ1 (pc914) as a hybridization probe of Amherst, the brown inversion, *b cn* and *cn bw* genomic DNAs, and cosmid cPn1111 (Fig. 3). The *b cn* stock is the parent stock of the inversion mutant. The *cn bw* genomic DNA was included both because it was extracted from a stock containing the original spontaneous *bw* allele that genetically defines the locus (29) and because spontaneous mutations are often caused by mobile element insertions in *Drosophila* species (3, 23, 31, 51).

The analysis revealed that the Amherst, *b cn*, and cosmid cPn1111 hybridization patterns were identical when these DNAs were digested with *Eco*RI, *Xho*I, and *Sac*I and probed with pc914 (Fig. 3A to C). However, the brown inversion and *cn bw* patterns exhibited alterations. With *Eco*RI, an 8.4-kilobase (kb) band of hybridization was detected in Amherst, *b cn*, and cosmid cPn1111 DNA. In contrast, the brown inversion exhibited a single 5.7-kb band and *cn bw* exhibited two bands of 9.0 and 3.9 kb (Fig. 3A). Amherst, *b cn*, and cosmid cPn1111 cut with *Xho*I exhibited two bands of hybridization of 6.8 and 1.1 kb. The brown inversion and *cn bw* each exhibited three bands of hybridization, with only the 1.1-kb band in common with the others (Fig. 3B). With *Sac*I, the analysis revealed that Amherst, *b cn*, and cosmid cPn1111 each had two bands of hybridization of 1.8 and 1.6 kb. The brown inversion and *cn bw* each exhibited two bands, with only the 1.8-kb band in common with the others (Fig. 3C). These results indicate that both the brown inversion and *cn bw* are disrupted within the region that specifies c914. A restriction map of the region including the locations of the brown inversion breakpoint and the *bw* disruption (evidently an insertion of approximately 8 kb) is shown in Fig. 1C.

Transcriptional analysis. Since the brown gene product is required for pteridine production in the eye and eye pigments are deposited during the late pupal stage and in the young adult (46), poly(A)$^+$ RNA was isolated from the heads of newly eclosed flies of Amherst, In(2LR)CK, 59E2;30A, *b cn bw*^{85f4}, and *cn bw* strains. Northern blots of this RNA were hybridized with pc914. Two poly(A)$^+$ RNA species of 3.0 and 2.8 kb were detected in heads from newly eclosed Amherst flies (Fig. 4, lane 1). In heads from newly eclosed flies with the brown inversion, a single 2.1-kb less abundant RNA species was detected (Fig. 4, lane 2), while in heads from newly eclosed *cn bw* flies, two more abundant RNA species of 1.8 and 1.5 kb were detected with pc914 as the hybridization probe (Fig. 4, lane 3).

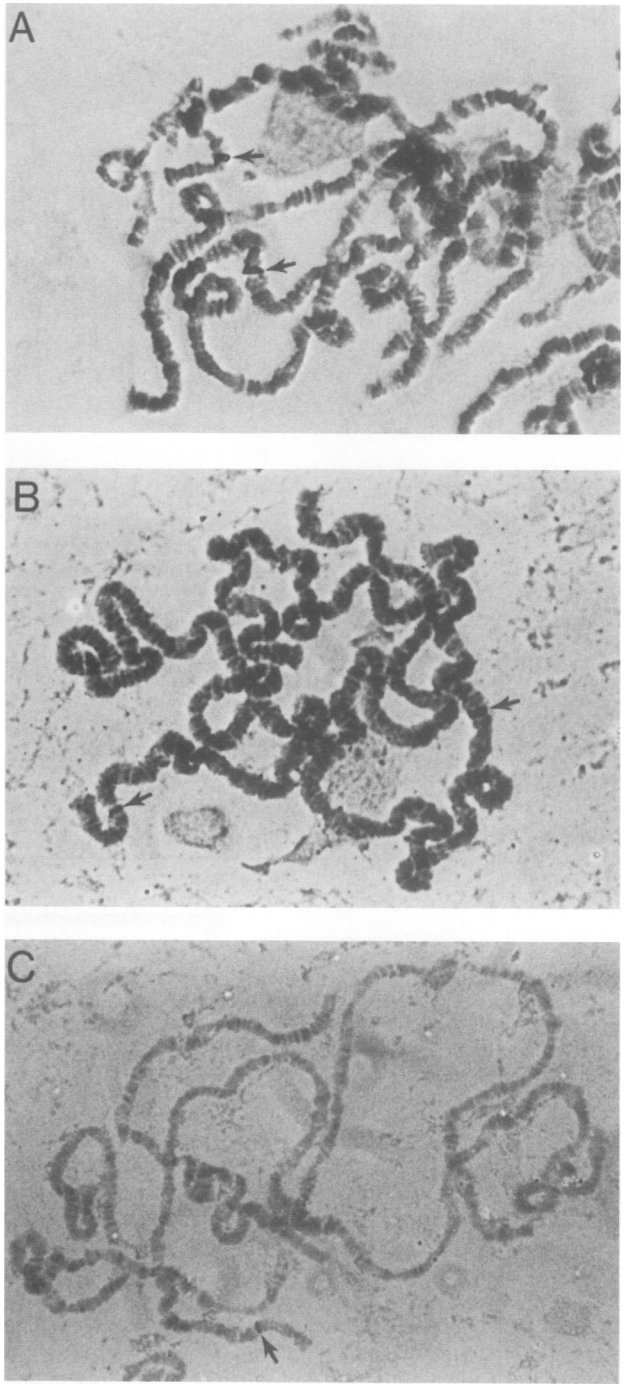

FIG. 2. In situ hybridizations. (A) Biotinylated DNA derived from cosmid cPn1111 was hybridized to In(2LR)CK, 59E2;30A, *b cn bw*^{85f4} salivary gland chromosomes. Two signals are detected, one at each breakpoint at 59E2 and 30A. (B) Plasmid pc914 was used as probe of the brown inversion mutant. Two signals are detected. (C) A cloned cDNA that includes a subset of c914 sequences was used as a probe of wild-type salivary gland chromosomes. A single signal at 59E2 is indicated.

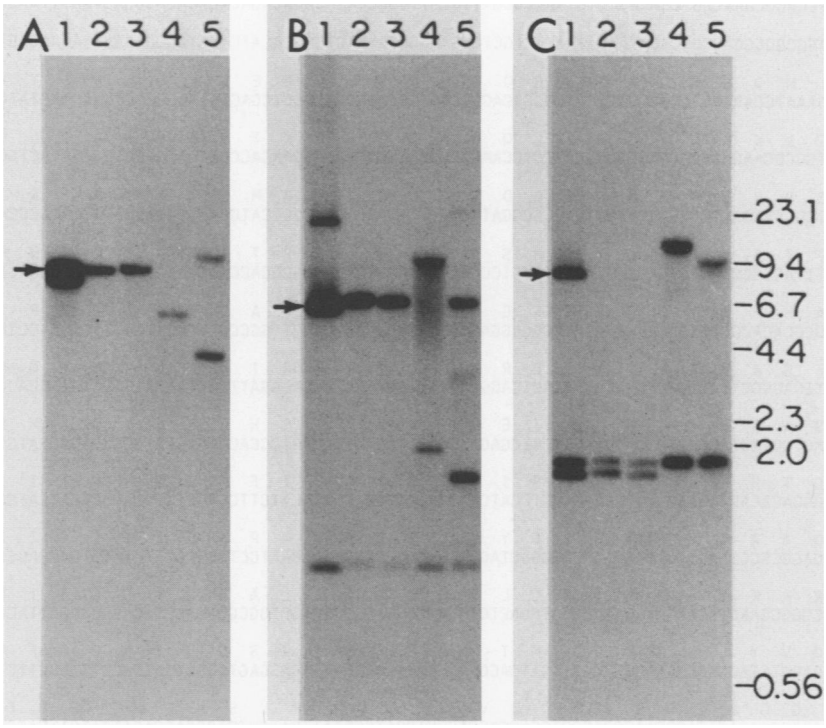

FIG. 3. Southern blot analysis of various brown alleles. DNA was digested with *Eco*RI (A), *Xho*I (B), or *Sac*I (C). The DNAs were extracted from bacteria containing cosmid cPn1111 (lane 1), and adult flies of Amherst (lane 2), *b cn* (lane 3), In(2LR)CK, 59E2;30A, *b cn* bw^{85f4} (lane 4), and *cn bw* (lane 5). The resulting blot was hybridized with randomly primed ^{32}P-labeled pc914. Because of the pVZ1 vector present in the hybridization probe, cosmid cPn vector was detected (arrow). Numbers on right show kilobases.

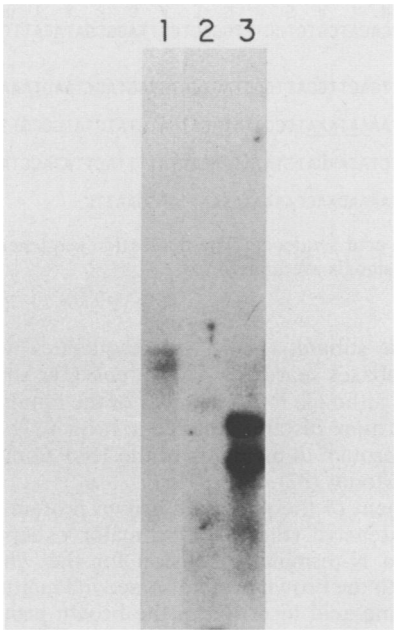

FIG. 4. Northern blot analysis. Poly(A)$^+$ RNA was isolated from the heads of Amherst (lane 1), In(2LR)CK, 59E2;30A, *b cn* bw^{85f4} (lane 2), or *cn bw* (lane 3) flies less than 24 h after eclosion. A total of 20 μg of RNA was loaded per lane, and the resulting blot was hybridized with ^{32}P-labeled randomly primed pc914.

The entire 2,740-base-pair (bp) c914 cDNA insert was sequenced (Fig. 5). There was a single extensive open reading frame of 675 amino acids beginning with an ATG (Fig. 5), and only this frame exhibited codon usage typical of *Drosophila* species (data not shown). The orientation of this coding sequence is from telomere to centromere (arrow in Fig. 1C) based on restriction mapping of genomic and cDNAs.

In addition, the poly(A) addition sites of five other cDNAs were determined by cDNA sequencing (data not shown). Based on sequence differences, at least four independent mRNA molecules are represented. Three of the cDNAs, including c914, diverge from one another in sequence 15 to 20 bp downstream of the AATAAA site at nucleotide 2707 (Fig. 5). The other cDNA diverges from the c914 sequence 37 bp downstream of the AATAAA site at nucleotide 2482. Assuming the c914 is a nearly full-length cDNA and assuming a poly(A) tail of 200 nucleotides at each 3′ end, these two classes of cDNAs would have been derived from mRNAs of 2.8 and 3.0 kb, nearly identical in size with the two poly(A)$^+$ RNAs detected with the c914 cDNA.

Comparison with database sequences. The amino acid sequence deduced from c914 was used to search the NBRF-PIR protein and the GenBank and EMBL (translated) DNA databases. Striking homologies with several proteins were detected. The best match was to the *Drosophila melanogaster* white protein (33). The other matches were to members of the HisP family of ATP-binding proteins, most of which are subunits of bacterial membrane-associated active transport complexes. These are *Salmonella typhimurium* oligopeptide permease subunit OppD (21); *S. typhimurium* histi-

```
GAATTCTGGAGTTGCCGCTCAGTCTTCCACGCGAACAGTCGACGGCGTGTGACCGCGATTGCGTTCGCCCAGACAGACAGACAGACAGACGGACAGATTTAAGGTTTCCCCCGAAAGATC 120

CCGAGCAGATATGCAAATACATTGTAGGCGGCGGACCCTCAGATAAGATAAGCCTGCTCCGAAGGCGATAAGTCGGCACATCACATAGCCAGAATTATAAGAGTTCGCCGGTGAGGTGAC 240

                                M  Q  E  S  G  G  S  S  G  Q  G  G  P  S  L  C  L  E  W  K  Q  L  N  Y  Y  V  P  D  Q  E  Q
TGAACCCGTCGAACTGACATAGCTCGAAATGCAAGAATCCGGCGGCTCGTCGGGCCAGGGCGGTCCATCGCTGTGCCTCGAGTGGAAGCAGCTGAACTACTATGTGCCCGACCAGGAGCA 360

 S  N  Y  S  F  W  N  E  C  R  K  K  R  E  L  R  I  L  Q  D  A  S  G  H  M  K  T  G  D  L  I  A  I  L  G  G  S  G  A  G
GAGCAACTACAGTTTCTGGAACGAATGCCGCAAGAAGCGCGAGCTGCGGATCCTCCAAGACGCCAGCGGGCACATGAAGACCGGCGACCTCATCGCCATACTGGGCGGATCCGGAGCGGG 480

 K  T  T  L  L  A  A  I  S  Q  R  L  R  G  N  L  T  G  D  V  V  L  N  G  M  A  M  E  R  H  Q  M  T  R  I  S  S  F  L  P
GAAGACCACGCTGCTGGCGGCGATTTCGCAACGACTGCGTGGTAACCTGACCGGGGATGTGGTTTTGAACGGCATGGCCATGGAGCGGCATCAGATGACGCGCATCTCCAGCTTTCTGCC 600

 Q  F  E  I  N  V  K  T  F  T  A  Y  E  H  L  Y  F  M  S  H  F  K  M  H  R  R  T  T  K  A  E  K  R  Q  R  V  A  D  L  L
GCAGTTCGAGATCAACGTGAAGACGTTTACGGCCTACGAGCATCTGTACTTTATGTCCCATTTCAAGATGCACCGGCGCACCACCAAGGCGGAGAAGCGCCAGAGGGTGGCGGATCTCCT 720

 L  A  V  G  L  R  D  A  A  H  T  R  I  Q  Q  L  S  G  G  E  R  K  R  L  S  L  A  E  E  L  I  T  D  P  I  F  L  F  C  D
CCTGGCAGTGGGCCTGCGCGACGCCGCCCACACCCGCATCCAGCAGCTGTCGGGCGGAGAGCGAAAGCGACTCAGCTTGGCCGAGGAGCTGATCACCGATCCCATATTCCTGTTCTGCGA 840

 E  P  T  T  G  L  D  S  F  S  A  Y  S  V  I  K  T  L  R  H  L  C  T  R  R  R  I  A  K  H  S  L  N  Q  V  Y  G  E  D  S
TGAACCCACCACAGGCTTGGACAGCTTCAGCGCTTATTCGGTGATCAAAACACTCAGGCACTTGTGCACCAGGCGACGGATTGCCAAACATTCCTTGAACCAGGTCTACGGCGAGGACTC 960

 F  E  T  P  S  G  E  S  S  A  S  G  S  G  S  K  S  I  E  M  E  V  V  A  E  S  H  E  S  L  L  Q  T  M  R  E  L  P  A  L
GTTTGAGACGCCGAGTGGCGAGAGCAGCGCCAGTGGCAGTGGCAGCAAGTCCATCGAGATGGAGGTCGTGGCCGAGTCCCACGAGAGTCTGCTGCAGACGATGCGGGAGCTCCCAGCCCT 1080

 G  V  L  S  N  S  P  N  G  T  H  K  K  A  A  I  C  S  I  H  Q  P  T  S  D  I  F  E  L  F  T  H  I  I  L  M  D  G  G  R
GGGAGTCCTAAGCAACAGCCCCAACGGCACACACAAGAAGGCGGCCATCTGCTCCATCCATCAGCCGACGTCGGACATCTTCGAGCTCTTCACCCACATCATCCTCATGGATGGCGGCAG 1200

 I  V  Y  Q  G  R  T  E  Q  A  A  K  F  F  T  D  L  G  Y  E  L  P  L  N  C  N  P  A  D  F  Y  L  K  T  L  A  D  K  E  G
GATCGTCTACCAGGGACGCACCGAGCAGGCGGCCAAGTTCTTTACAGACCTGGGCTACGAACTCCCGCTGAACTGCAATCCTGCGGACTTCTACCTGAAGACCCTGGCGGACAAGGAGGG 1320

 K  E  N  A  G  A  V  L  R  A  K  Y  E  H  E  T  D  G  L  Y  S  G  S  W  L  L  A  R  S  Y  S  G  D  Y  L  K  H  V  Q  N
CAAAGAGAACGCGGGGGCTGTGCTCCGGGCGAAGTACGAGCACGAGACGGATGGACTCTACAGCGGGAGCTGGCTGCTGGCGCGGAGCTACAGCGGGGATTACCTGAAGCACGTCCAGAA 1440

 F  K  K  I  R  W  I  Y  Q  V  Y  L  L  M  V  R  F  M  T  E  D  L  R  N  I  R  S  G  L  I  A  F  G  F  F  M  I  T  A  V
CTTCAAGAAAATAAGGTGGATATATCAGGTGTACCTCCTGATGGTGCGCTTCATGACCGAGGATTTGCGGAATATCAGGAGTGGGCTCATCGCCTTTGGCTTCTTCATGATCACCGCTGT 1560

 T  L  S  L  M  Y  S  G  I  G  G  L  T  Q  R  T  V  Q  D  V  G  G  S  I  F  M  L  S  N  E  M  I  F  T  F  S  Y  G  V  T
AACACTATCCCTCATGTACTCCGGCATTGGAGGACTCACGCAACGGACGGTGCAGGACGTCGGTGGATCCATATTCATGCTCAGCAACGAGATGATCTTCACGTTCAGCTATGGCGTGAC 1680

 Y  I  F  P  A  A  L  P  I  I  R  R  E  V  G  E  G  T  Y  S  L  S  A  Y  Y  V  A  L  V  L  S  F  V  P  V  A  F  F  K  G
GTATATATTTCCCGCCGCCCTTCCCATCATCAGGCGGGAGGTGGGCGAGGGCACCTACAGCCTGTCCGCCTACTACGTGGCCCTGGTGCTCTCTTTTGTGCCCGTGGCCTTCTTCAAGGG 1800

 Y  V  F  L  S  V  I  Y  A  S  I  Y  Y  T  R  G  F  L  L  Y  L  S  M  G  F  L  M  S  L  S  A  V  A  A  V  G  Y  G  V  F
CTACGTCTTCCTGTCGGTGATATATGCCTCCATATACTACACGCGCGGCTTCCTTCTGTACCTCAGCATGGGCTTCCTGATGAGCCTGTCCGCGGTGGCGGCTGTTGGCTATGGGGTCTT 1920

 L  S  S  L  F  E  S  D  K  M  A  S  E  C  A  A  P  F  D  L  I  F  L  I  F  G  G  T  Y  M  N  V  D  T  V  P  G  L  K  Y
CCTCTCCAGCCTCTTCGAGTCGGATAAGATGGCCTCCGAGTGCGCGGCGCCCTTCGATCTGATCTTCCTGATCTTTGGCGGCACCTACATGAATGTGGACACAGTGCCTGGACTAAAGTA 2040

 L  S  L  F  F  Y  S  N  E  A  L  M  Y  K  F  W  I  D  I  D  N  I  D  C  P  V  N  E  D  H  P  C  I  K  T  G  V  E  V  L
CCTCTCGTTGTTCTTCTACTCCAACGAGGCGCTGATGTACAAGTTCTGGATCGACATCGATAATATCGACTGCCCCGTCAACGAGGATCATCCGTGCATCAAGACCGGAGTGGAGGTGCT 2160

 Q  Q  G  S  Y  R  N  A  D  Y  T  Y  W  L  D  C  F  S  L  V  V  V  A  V  I  F  H  I  V  S  F  G  L  V  R  R  Y  I  H  R
GCAGCAGGGATCCTATCGCAACGCCGACTACACCTATTGGCTGGACTGCTTCAGTCTGGTGGTGGTGGCCGTCATCTTCCACATCGTGTCCTTTGGGCTGGTTAGGCGATACATTCACCG 2280

 S  G  Y  Y  *
CAGTGGTTACTATTGATTGGGGCGCTTTTATTATACTAATTTCAGCCCACACTGTTACAGAACTTGCTTCAGCTTAATTTCAGTTGCATTGGGTACTGGTGCGGTAGTTAAGTAAAATTT 2400

GTAATTTTAAGCACAAACTGTTCTTAATACATCTAAACTTCAGTGGAAAATATTGCATTCTCGTTTGCCGATACGACATAAAATAAATCCCATATAGATCATGTATGTATGCCGTTACTA 2520

AAAGATATTTAACTTGCATTTGAACTACATTTGAGCTCAAGAAAAAGTCGGGAGTCTTGCGAGTCGAAGAAGAGGCGATCTATAAGATCTATATCAGATTATTTTACTTATACCTATTAT 2640

TATACGATTGTCTGTCATTATATATAAATGGTTTGAAGGGGTTTAAAGAGACCTGACAGTTGAATTAATAAATTAAGCTAAAACAACCAAAAAATAATGGAGGAATTC 2748
```

FIG. 5. Nucleotide sequence of (Oregon R-derived) c914 and the deduced amino acid sequence. The nucleotide sequence begins and ends with *Eco*RI sites resulting from the cloning procedure. The two polyadenylation signals are underlined.

dine transport subunit HisP (18); *Escherichia coli* phosphate transport subunit PstB (47); *E. coli* maltose transport subunit MalK (13); *E. coli* ribose transport subunit RbsA (2); *E. coli* molybdenum transport subunit ChlD (24); *Rhizobium leguminosarum* nodulation protein NodI (6); *E. coli* hemolysin secretion protein HlyB (10); liverwort chloroplast predicted protein MbpX (34); the human *mdr1* (multiple drug resistance) P-glycoprotein (4); and an unrecognized 177-amino-acid open reading frame in the vicinity of the *Bacillus subtilis purA* gene (38). Except for the last sequence, each of these proteins is a previously identified member of the HisP family. Amino acid sequence similarities between various of these sequences have been reported by others (19). Other known members of the family not detected by sequence similarity to the brown protein include the *E. coli* FtsE cell division protein (12) and the *S. typhimurium* OppF oligopeptide permease subunit (21). These sequences were absent from the databases searched. The *E. coli* UvrA protein was not detected, although it is a member of the family (5). UvrA appears to be more distantly related to brown. The similarity of the white protein to members of the HisP family was first reported by Mount (32).

The alignment of the predicted brown protein with white was quite extensive (Fig. 6). One major exception was a 72-amino-acid N-terminal extension for the white protein compared with the brown protein. A second major exception was a 66-amino-acid insertion in the brown protein (amino acids 216 to 381) compared with the white protein. Otherwise, these two proteins were identical at 179 (29%) of the 609 aligned amino acids.

Alignment of the predicted brown protein with HisP, a typical member of the family, is presented in Fig. 6. HisP

FIG. 6. Sequence alignments of HisP (18), the predicted brown protein (bw), and the white protein (w) (33). Identical residues are indicated by colons, and conserved changes are shown by single dots. Matches of the HisP protein to HisP family consensus residues (19) are indicated by asterisks above the HisP sequence.

aligned with the N-terminal 342 amino acids of the brown protein, except for the same 66-amino-acid insertion noted in the comparison with the white protein. Otherwise, these two proteins were identical at 51 (21%) of the 243 aligned amino acids. Consensus residues of HisP compiled for the bacterial family by Higgins and co-workers (19) are indicated by asterisks above the sequence (Fig. 6). Overall, 43 of these 90 consensus amino acids (48%) were identical residues or were conservative substitutions in the brown protein. There were two regions of extensive conservation. The first extended from amino acid 65 to amino acid 74 of the brown sequence and contained a consensus ATP-binding site (50) (consensus, GXXXXGKT/S; brown, GGSGAGKT). The second region extended from amino acid 168 to amino acid 199 of the brown sequence and matched the HisP family consensus at 20 of 32 aligned residues. This region also is thought to be involved in the binding of ATP (50).

Hydropathy comparison. The brown and white proteins were examined for regions of hydrophobicity and hydrophilicity (Fig. 7A and B). The N-terminal portion of each protein displayed a pattern typical of soluble proteins and was similar to that of HisP (Fig. 7C), with which they share homology. However, the C-terminal one-third of the predicted brown and white proteins contained a large proportion of hydrophobic residues. Similar regions are seen in the integral membrane subunits of the same permease complexes that also include HisP family components. As an example, we present the hydropathy profile of HisQ, one of the integral membrane subunits of the histidine permease complex (Fig. 7C). According to Hiles et al. (21), there are five potential transmembrane segments in each of the integral membrane subunits (Fig. 7C). A rather similar set of potential transmembrane segments also was seen for the corresponding region of the brown and white proteins (Fig.

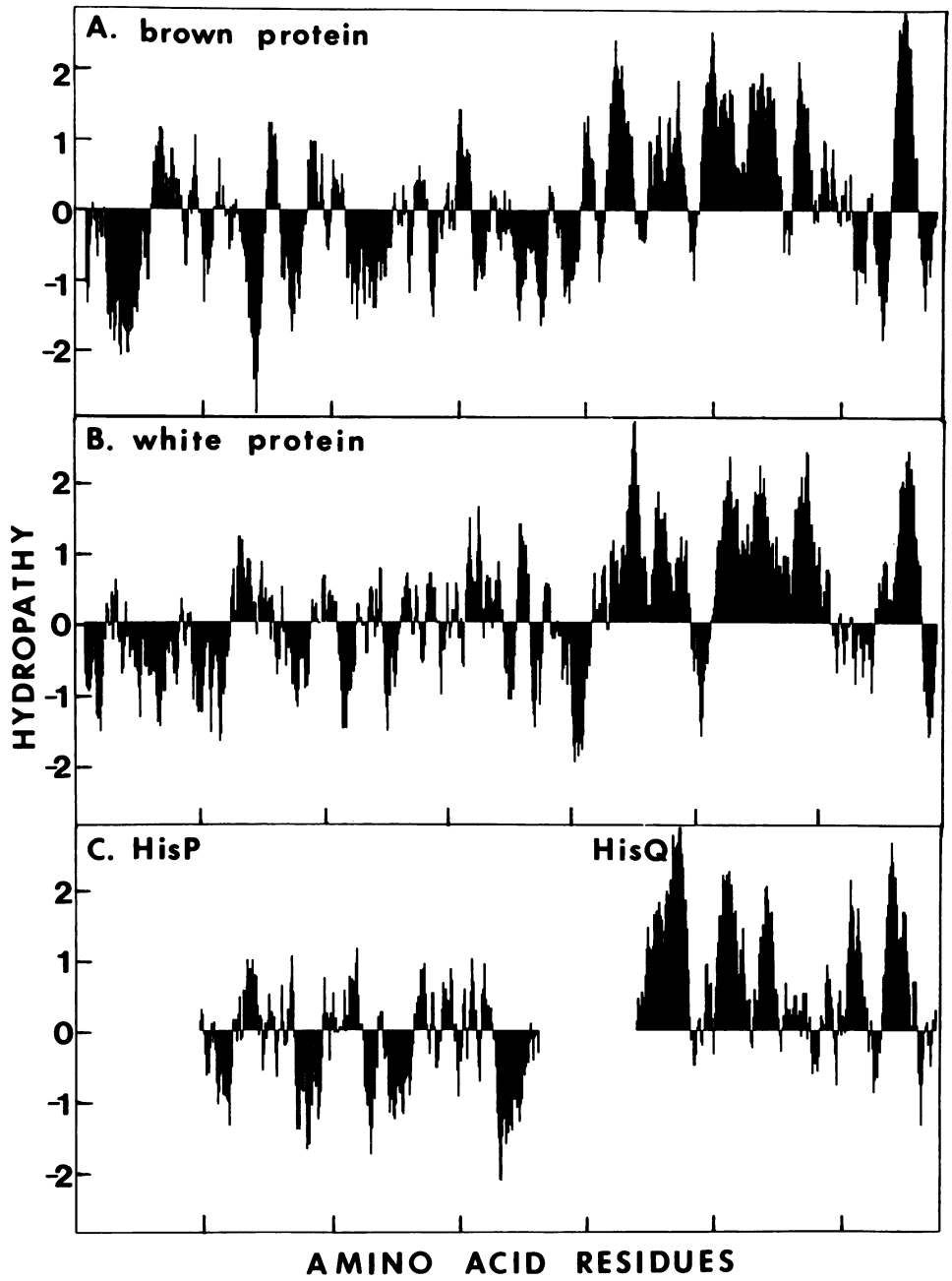

FIG. 7. Hydropathy plots of the predicted brown protein (A), the predicted white protein (33) (B), and HisP (18) and HisQ (18) (C). The scale of the abscissa is 100 amino acid residues per interval.

7A and B) when the C-terminal residues were precisely aligned. These potential transmembrane segments in the white protein have been previously noted by O'Hare et al. (33).

DISCUSSION

Identification of brown gene product. By microcloning and chromosome walking in the 59E region of *D. melanogaster*, we isolated DNA that includes the brown gene. The precise location of brown was determined by using two *bw* mutants: the original spontaneous *bw* mutant and the euchromatic *bw* inversion mutant. In situ hybridization (Fig. 2) and genomic Southern (Fig. 3) and Northern (Fig. 4) analyses showed that a single transcription unit is interrupted in the two *bw* mutants. We therefore conclude that this transcription unit encodes the brown protein.

One reason for interest in the brown gene is its unusual behavior when brought into the vicinity of heterochromatin. Genes that are juxtaposed in this way often show mosaic expression, called position-effect variegation (for a review, see reference 42). In nearly all cases, such as for white, the effect of the heterochromatin is manifested only in *cis*, leading to a recessive phenotype. In contrast, the effect of heterochromatin on brown is manifested in both *cis* and *trans*, leading to a dominant phenotype. The similarity between the protein products of brown and white suggests

that this sensitivity to heterochromatin in *trans* is not likely to result from an unusual feature of the gene product. Also, the brown transcription unit does not appear to be unusual in size. Our largest cDNA insert was 2,740 bp, a size consistent with it being full length, whereas the corresponding genomic region encompassed approximately 3.2 kb of DNA. This small size should aid in the molecular study of dominant variegation.

Northern blot analysis detected the presence of two RNA species of 2.8 and 3.0 kb in the heads of newly eclosed wild-type flies (Fig. 4). Examination of brown cDNAs revealed two possible polyadenylation signals (AATAAA). The most 5′ site was apparently used in processing an mRNA corresponding to at least one of the cDNAs. The other site, 225 bp upstream, was apparently used in processing mRNAs corresponding to at least three of the cDNAs. The close correspondence of the two classes of cDNAs to the sizes of the two brown mRNAs is evidence that these mRNAs result from the use of alternative poly(A) addition sites with approximately equal frequency. Multiple polyadenylation signals within noncoding sequence are not an uncommon feature of eucaryotic transcription units (28), although no functional explanation has yet been reported.

The brown mRNAs appear to be very rare. We were unable to detect transcripts in whole pupae using procedures able to detect rare housekeeping transcripts estimated to be present at 1 part in 100,000 of the poly(A)$^+$ RNA (data not shown). Detection was successful only when we used heads from newly eclosed adults and probes with a specific activity of $>10^9$ cpm/μg. No brown mRNA was detected in bodies from newly eclosed adults (data not shown). Head-specific expression at low levels is consistent with suggestions that the brown protein is a transport system component in the developing eye, as discussed below.

Possible function of brown protein. Many of the *Drosophila* eye pigment mutants are also deficient in larval malpighian tubule pigment, suggesting that the genes are expressed in this organ (29). In culture, developing eyes and malpighian tubules from white and certain other ommochrome-deficient mutants take up much reduced amounts of 3-hydroxykynurenine, an ommochrome precursor (45). Similarly, malpighian tubules take up reduced amounts of guanine in *w* and *bw* mutants (44, 45). These results led to the hypothesis that these loci encode permease proteins involved in the transport of pigment precursors across cell membranes of the developing eye and malpighian tubules (44). Thus, the white protein would facilitate uptake of both ommochrome and pteridine precursors, whereas the brown protein would facilitate uptake of a pteridine precursor. Consistent with this hypothesis is the cell autonomy of *w* and *bw* mutants (29) and the in situ localization of white mRNA to cells in which pigments are produced (11).

Of the 11 bacterial members of the HisP family homologous to white and brown proteins, 7 are known components of active transport transmembrane complexes (19). These complexes are composed of a periplasmic substrate-binding component, two hydrophobic integral membrane components, and the conserved ATP-binding subunit (1, 19, 21). HisP family members OppD, MalK, and HisP are known to bind ATP or ATP analogs (20, 22) and probably hydrolyze them to supply the energy needed for transport of their substrates. As pointed out by Mount (32), the similarity of the white protein to members of this family strongly supports its hypothesized role in the uptake of ommochrome and pteridine precursors. Likewise, our finding of similarity between the brown protein and members of this family strongly supports its hypothesized role in uptake of a pteridine precursor.

It has been proposed that proteins of the HisP family function as dimers or higher-order multimers (19, 21). The *E. coli* oligopeptide transport system requires two ATP-binding subunits, OppD and OppF (21). In addition, the *E. coli* ribose transport ATP-binding subunit, RbsA, contains two homologous domains in tandem (2). The mammalian multidrug-resistant (*mdr1*) P-glycoprotein also contains two homologous ATP-binding domains (4). The homologous brown and white proteins might provide a corresponding pair of ATP-binding domains for the hypothesized pteridine precursor permease.

In addition to the two ATP-binding domains, some and perhaps all of the permease complexes require two integral membrane domains. The operons encoding the bacterial permeases for oligopeptides (21), phosphate (47), maltose (13), and histidine (18) each encode two homologous integral membrane subunits. These do not show any significant sequence similarities when proteins encoded at different operons are compared. However, several potential transmembrane segments are predicted for the various permease complex transmembrane subunits (21). Both the C-terminal one-third of the brown protein and the C-terminal one-third of the white protein share this predicted structural feature (Fig. 7). This suggests that both the brown and white proteins contain the ATP-binding and membrane-spanning domains fused into a single polypeptide.

In at least five of the bacterial operons containing a member of the HisP family (*nod* [6], *pst* [47], *opp* [21], *his* [18], and *chl* [24]), translational coupling on a polycistronic mRNA is thought to ensure equal numbers of various of the active transport components. For the brown and white proteins, fusion of the ATP-binding domain to the multiple membrane-spanning domain guarantees an equal numbers of components.

The hypothesized heteromultimeric structure of transport system family members is particularly interesting in light of the genetic evidence for an interaction between white and brown proteins. Farmer and Fairbanks (7, 8) have reported an interaction of a particular allele of white, w^{co2}, and a particular allele of brown, bw^6. The pteridine content of eyes was much reduced in $w^{co2};bw^6$ mutants, but was at nearly normal levels if a wild-type allele of either was present. They proposed a model in which the white and brown protein products interact to produce the active protein, suggesting that w^{co2} and bw^6 interfere with the quaternary structure of a dimer or multimer. This proposed interaction is consistent with the similarity between these two proteins and the permease components, which are thought to interact as heterodimers or heteromultimers. In this regard, it is interesting that the human multidrug resistance protein is composed of four domains: two integral membrane domains alternating with two domains homologous to the HisP family (4).

Thus, several lines of evidence lead to a consistent picture for brown protein function. The homology between brown and white proteins, together with genetic evidence for their interaction (7, 8), argues that the two proteins are part of a complex that closely resembles well-studied permeases. These findings support earlier suggestions that brown and white proteins are part of a precursor uptake system required for the synthesis of pteridines in the secondary pigment cells of the *Drosophila* eye and malpighian tubules (44, 45). Similar suggestions for the involvement of scarlet and white proteins in the kynurenine uptake system are

supported by the finding that the scarlet protein is also a member of the same family (A. Howells, personal communication; R. Tearle, J. Belote, M. McKeown, B. Baker, and A. Howells, manuscript in preparation). Thus, the uptake of ommochrome and pteridine precursors appears to be mediated by similar heteromultimers involving a common subunit, the white protein, and one of two related subunits, either the scarlet or the brown protein.

ACKNOWLEDGMENTS

We thank M. Champe and B. Tempel for providing clone libraries that were important for this work.

This work was supported by Institutional Cancer grant IN-26Y from the American Cancer Society, Public Health Service grant GM29009 from the National Institutes of Health, and grant DCB-8717937 from the National Science Foundation.

LITERATURE CITED

1. **Ames, G. F.-L., and C. F. Higgins.** 1983. The organization, mechanism of action, and evolution of periplasmic transport systems. Trends Biochem. Sci. **8**:97–100.
2. **Bell, A. W., S. D. Buckel, J. M. Groarke, J. N. Hope, D. H. Kingsley, and M. A. Hermodson.** 1986. The nucleotide sequences of the *rbs*D, *rbs*A, and *rbs*C genes of *Escherichia coli* K12. J. Biol. Chem. **261**:7652–7658.
3. **Campuzano, S., L. Carramolino, C. V. Cabrera, M. Ruíz-Gómez, R. Villares, A. Boronat, and J. Modolell.** 1985. Molecular genetics of the *achaete-scute* gene complex of *D. melanogaster*. Cell **40**:327–338.
4. **Chen, C., J. E. Chin, K. Ueda, D. P. Clark, I. Pastan, M. M. Gottesman, and I. B. Roninson.** 1986. Internal duplication and homology with bacterial transport proteins in the *mdr*1 (P-glycoprotein) gene from multidrug-resistant human cells. Cell **47**:381–389.
5. **Doolittle, R. F., M. S. Johnson, I. Husain, B. Van Houten, D. C. Thomas, and A. Sancar.** 1986. Domainal evolution of a prokaryotic DNA repair protein and its relationship to active-transport proteins. Nature (London) **323**:451–453.
6. **Evans, I. J., and J. A. Downie.** 1986. The *nod*I gene product of *Rhizobium leguminosarum* is closely related to ATP-binding bacterial transport proteins; nucleotide sequence analysis of the *nod*I and *nod*J genes. Gene **43**:95–101.
7. **Farmer, J. L.** 1977. An allele-specific suppressor of *white-coral* in *Drosophila melanogaster*. Heredity **39**:297–303.
8. **Farmer, J. L., and D. J. Fairbanks.** 1986. Interaction of the *bw* and *w* loci in *D. melanogaster*. Dros. Info. Serv. **63**:50–51.
9. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
10. **Felmlee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163**:94–105.
11. **Fjose, A., L. C. Polito, U. Weber, and W. J. Gehring.** 1984. Developmental expression of the *white* locus of *Drosophila melanogaster*. EMBO J. **3**:2087–2094.
12. **Gill, D. R., G. F. Hatfull, and G. P. C. Salmond.** 1986. A new cell division operon in *Escherichia coli*. Mol. Gen. Genet. **205**: 134–145.
13. **Gilson, E., H. Nikaido, and M. Hofnung.** 1982. Sequence of the *mal*K gene in *E. coli* K12. Nucleic Acids Res. **10**:7449–7458.
14. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
15. **Henikoff, S., and M. K. Eghtedarzadeh.** 1987. Conserved arrangement of nested genes at the Drosophila *Gart* locus. Genetics **117**:711–725.
16. **Henikoff, S., and M. Meselson.** 1977. Transcription at two heat shock loci in *Drosophila*. Cell **12**:441–451.
17. **Henikoff, S., and J. Wallace.** 1988. Detection of protein similarities using nucleotide sequence databases. Nucleic Acids Res. **16**:6191–6204.
18. **Higgins, C. F., P. D. Haag, K. Nikaido, F. Ardeshir, G. Garcia, and G. F.-L. Ames.** 1982. Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. Nature (London) **298**:723–727.
19. **Higgins, C. F., I. D. Hiles, G. P. C. Salmond, D. R. Gill, J. A. Downie, I. J. Evans, I. B. Holland, L. Gray, S. D. Buckel, A. W. Bell, and M. A. Hermodson.** 1986. A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature (London) **323**:448–450.
20. **Higgins, C. F., I. D. Hiles, K. Whalley, and D. J. Jamieson.** 1985. Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. EMBO J. **4**:1033–1040.
21. **Hiles, I. D., M. P. Gallagher, D. J. Jamieson, and C. F. Higgins.** 1987. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. J. Mol. Biol. **195**:125–142.
22. **Hobson, A. C., R. Weatherwax, and G. F.-L. Ames.** 1984. ATP-binding sites in the membrane components of histidine permease, a periplasmic transport system. Proc. Natl. Acad. Sci. USA **81**:7333–7337.
23. **Jack, J. W.** 1985. Molecular organization of the *cut* locus of Drosophila melanogaster. Cell **42**:869–876.
24. **Johann, S., and S. M. Hinton.** 1987. Cloning and nucleotide sequence of the *chlD* locus. J. Bacteriol. **169**:1911–1916.
25. **Kavenoff, R., and B. H. Zimm.** 1973. Chromosome-sized DNA molecules from *Drosophila*. Chromosoma **41**:1–27.
26. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
27. **Langridge, J., P. Langridge, and P. L. Bergquist.** 1980. Extraction of nucleic acids from agarose gels. Anal. Biochem. **103**:264–271.
28. **Leff, S. E., M. G. Rosenfeld, and R. E. Evans.** 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. Annu. Rev. Biochem. **55**:1091–1117.
29. **Lindsley, D. L., and E. H. Grell.** 1968. Genetic variations of *Drosophila melanogaster*. Publication no. 627. Carnegie Institution of Washington, Washington, D.C.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **McLachlan, A.** 1986. *Drosophila forked* locus. Mol. Cell. Biol. **6**:1–6.
32. **Mount, S. M.** 1987. Sequence similarity. Nature (London) **325**: 487.
33. **O'Hare, K., C. Murphy, R. Levis, and G. M. Rubin.** 1984. DNA sequence of the *white* locus of *Drosophila melanogaster*. J. Mol. Biol. **180**:437–455.
34. **Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, and H. Ozeki.** 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature (London) **322**:572–574.
35. **Pirrotta, V., H. Jackle, and J.-E. Edström.** 1983. Microcloning of microdissected chromosome fragments. Genet. Eng. Principles Methods **5**:1–17.
36. **Poole, S. J., L. M. Kauvar, B. Drees, and T. Kornberg.** 1985. The engrailed locus of Drosophila: structural analysis of an embryonic transcript. Cell **40**:37–43.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Sargent, M. G., and M. F. Bennett.** 1986. Identification of a specific membrane-associated DNA sequence in *Bacillus subtilis*. J. Bacteriol. **166**:38–43.
39. **Scalenghe, F., K. E. Turco, J.-E. Edström, V. Pirrotta, and M. Melli.** 1981. Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. Chromosoma **82**:205–216.
40. **Simon, J. A., C. A. Sutton, R. B. Lobel, R. L. Glaser, and J. T. Lis.** 1985. Determinants of heat shock-induced chromosome puffing. Cell **40**:805–817.
41. **Slatis, H. M.** 1955. Position effects at the brown locus in

Drosophila melanogaster. Genetics **40**:5–23.

42. **Spofford, J. B.** 1976. Position-effect variegation in *Drosophila*, p. 955–1018. *In* M. Ashburner and E. Novitski (ed.), The genetics and biology of *Drosophila*, vol. 1C. Academic Press, Inc. (London), Ltd., London.
43. **Steller, H., and V. Pirrotta.** 1985. A transposable P vector that confers selectable G418 resistance to *Drosophila* larvae. EMBO J. **4**:167–171.
44. **Sullivan, D. T., L. A. Bell, D. R. Paton, and M. C. Sullivan.** 1979. Purine transport by malpighian tubules of pteridine-deficient eye color mutants of *Drosophila melanogaster*. Biochem. Genet. **17**:565–573.
45. **Sullivan, D. T., and M. C. Sullivan.** 1975. Transport defects as the physiological basis for eye color mutants of *Drosophila melanogaster*. Biochem. Genet. **13**:603–613.
46. **Summers, K. M., A. J. Howells, and N. A. Pyliotis.** 1982. Biology of eye pigmentation in insects. Adv. Insect Physiol. **16**:119–166.
47. **Surin, B. P., H. Rosenberg, and G. B. Cox.** 1985. Phosphate-specific transport system of *Escherichia coli*: nucleotide sequence and gene polypeptide relationships. J. Bacteriol. **161**: 189–198.
48. **Thomas, P. S.** 1983. Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol. **100**: 255–259.
49. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA **76**:3683–3687.
50. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
51. **Zachar, Z., and P. M. Bingham.** 1982. Regulation of *white* locus expression: the structure of mutant alleles at the *white* locus of Drosophila melanogaster. Cell **30**:529–541.