The EMBO Journal Vol.2 No.6 pp.853–860, 1983

Molecular analysis of large transposable elements carrying the *white* locus of *Drosophila melanogaster*

Renato Paro[1], Michael L. Goldberg[2] and Walter J.Gehring*

Department of Cell Biology, Biozentrum der Universität Basel, Klingelbergstr. 70, CH-4056 Basel, Switzerland

Communicated by W.J.Gehring
Received on 18 February 1983; revised on 25 March 1983

A large transposable element *(TE)* comprising the *white-apricot* and *roughest* genes has been found to transpose to well over a hundred sites scattered over the *Drosophila* genome. We report the cloning of the essential parts of several *TEs*. *TE98* and *TE28* sequences were cloned by 'walking' along the chromosome from the previously cloned heat-shock genes. The ends of the *TEs* are characterized by dispersed repetitive elements belonging to the foldback *(FB)* family. *FB* elements are also associated with two independently isolated transposable elements originating from the *white* locus, *Tp* w^c*-1* and *Tp* w^{+IV}. The strong correlation between *FB* elements and large composite transposons suggests that a pair of these elements can mobilize large intermediary DNA segments. One particular *FB* family member, *FB-NOF*, is associated with *TE28*, the *white-crimson* (w^c) mutant, the w^c-derived *Tp* w^c*-1* and probably also with *Tp* w^{+IV}. A unique sequence located close to the *white* end of *TE28* was used to clone the borders of *TE77* and the surrounding sequences in the *bithorax* region, indicating that the *TE* can be used as a probe for gene isolation. Some evolutionary implications of the large composite transposons are discussed.

Key words: Drosophila/white locus/transposable elements

Introduction

In *Drosophila* several different classes of elements which have the potential to change their position in the genome and to induce chromosomal rearrangements have been described (reviewed by Spradling and Rubin, 1981). The *copia*-like elements represent a major class of moderately repetitive sequences and share some common structural features; they measure ~5–10 kb in size and are flanked by direct repeats of 200–500 bp. Approximately 20–100 copies are dispersed over the entire genome, and their chromosomal locations differ considerably between different stocks. At the site of insertion they generate a duplication of 4 or 5 bp in the target DNA depending on the family. Structurally these elements closely resemble integrated retroviruses, and recently circular DNA copies of *copia* and retrovirus-like particles containing *copia* RNA and reverse transcriptase activity have been found (Flavell and Ish-Horowicz, 1981; Shiba and Saigo, 1983). There may be as many as 30 different families of *copia*-like elements but the individual members of a given family are similar in sequence.

[1]Present address: University of Edinburgh, Department of Molecular Biology, King's Building, Mayfield Road, Edinburgh EH9 3JR, UK.
[2]Present address: Cornell University, Section of Genetics and Development, Bradfield Hall, Ithaca, NY 14853, USA.
*To whom reprint requests should be sent.

This is in contrast to the foldback DNA sequences, designated as the *FB* family, whose members are much more variable (Potter *et al.*, 1980; Goldberg *et al.*, 1982). The common structural organization of the *FB* family consists of a pair of inverted repeats of variable length (300–1500 bp) that share extensive sequence homology especially at their ends. The inverted repeats in turn are composed of tandem repeats of short sequences similar to satellite DNA (Truett *et al.*, 1981; Potter, 1982a). The inverted repeats may flank a central core segment, but the core segments of different family members are not homologous and may be completely absent. There are some 20–60 copies of *FB* per genome whose location differs among different stocks. Upon integration, a 9-bp duplication is generated in the target DNA.

The P-factor family is also characterized by inverted repeats (Rubin *et al.*, 1982), but in this case they are only 31 bp in length (O'Hare and Rubin, cited in Spradling and Rubin, 1982). In addition to the complete P-factor which is ~3 kb in size, a large fraction of defective family members are found, most likely representing internal deletions. It has been shown directly that the P-factor functions as a transposon; upon injection of the cloned complete P-factor into syncytial embryos it becomes integrated into the recipient chromosomes at a high frequency (Spradling and Rubin, 1982). Only those sequences flanked by the inverted repeats, but not the bacterial vector sequences, become integrated, emphasizing the importance of these termini for transposition. In addition, the intact P-factor can mobilize defective P-elements *(in trans)* provided that they have the proper termini (Rubin and Spradling, 1982).

Transposable elements carrying genetic markers had been discovered in genetic experiments prior to the advent of recombinant DNA technology. Green (1967) described a highly unstable *white* mutant, *white-crimson* (w^c), which subsequently was found to transpose from the X chromosome to several different sites on the third chromosome (Green, 1969a, 1969b). On the basis of the genetic data, Green postulated a 'controlling element' causing both the instability and the transposition. Another transposition of the *white* locus has been described by Rasmuson and collaborators (Rasmuson and Green, 1974; Rasmuson *et al.*, 1980).

The most extensively studied transposing element *(TE)* in *Drosophila* has been isolated and analyzed by G.Ising and his co-workers. A large DNA segment containing the *white-apricot* (w^a) allele and the adjacent *roughest* (rst^+) gene was shown to have left its normal position on the X chromosome and to have transposed to the second chromosome (Ising and Ramel, 1976). Subsequently, new transpositions have been identified both genetically and cytologically at well over 100 insertion sites scattered over the entire *Drosophila* genome (Ising and Block, 1981). The *TE* consists of several chromosomal bands which represent a few hundred kilobases of DNA. The transposition events are accompanied by considerable genetic variation: w^a may revert to a near wild-type eye color, it may become duplicated or lost from the element and genes adjacent to the *TE* insertion sites may 'hitch hike'

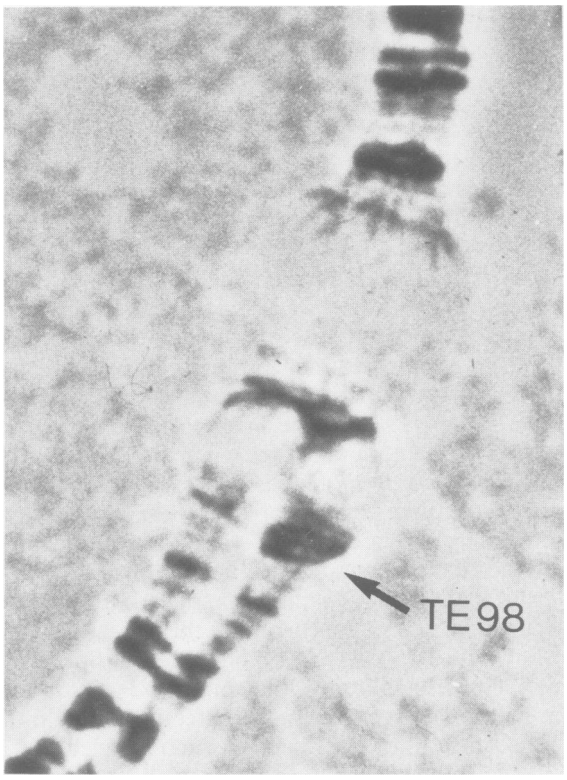

Fig. 1. An Orcein-stained preparation of polytene chromosomes of the strain *TE98/D³-CxF*. Larvae were heat-shocked for 30 min at 37°C prior to the squashing of the chromosomes. The two puffed areas at 87A7 and 87C1 reflect the gene activity of the 70-kd heat-hock protein clusters. *TE98* (arrow) is visible as a dark banded chromomere structure compared with the homologous balancer chromosome lacking the *TE.TE98* derived from a transposition of the red form of *TE-1*, and is therefore a revertant of *white-apricot* due to the loss of the *copia* insertion.

with the *TE* during subsequent transpositions. In an earlier study we showed, by *in situ* hybridization to polytene chromosomes, that many *TEs* contain sequences homologous to *copia*, and that *copia* is associated with the *white-apricot* mutation (Gehring and Paro, 1980). Since the transposing property can be dissociated from the w^a and rst^+ genes, we proposed that a pair of *copia* or similar elements may be involved in mobilizing these two marker genes. To test this hypothesis we have now cloned the relevant parts of the *TE*. Three different *TEs* have been analyzed: *TE98* which has inserted in close proximity to the heat-shock genes *(hsp70)* at 87A on the third chromosome, *TE28* which is located a few bands more proximal to *TE98* at 86E/F and finally *TE77*, which has been mapped to the region of the *bithorax* complex at 89DE. *TE98* and *TE77* are derived directly from the original *TE1* on the second chromosome and confer a red near wild-type eye color, whereas *TE28* is a secondary transposition originating from *TE8* on the X chromosome and retains the *white-apricot* marker. We report the main structural features of the *TE* and the presence of flanking *FB* elements which appear to be responsible for its mobility.

Results

Strategy for cloning TE98 and isolation of the target site

The observation that *TE98* has inserted at 87A5-6, very close to the heat-shock genes, which previously have been cloned in our laboratory (Schedl *et al.*, 1978), offered the possibility of cloning the *TE* by 'walking along the chromosome'. Salivary gland preparations of heat-shock induced polytene chromosomes show the close proximity of *TE98* to the heat-shock puffs (Figure 1). The distance between the heat-shock genes and *TE98* was judged to be sufficiently small so as to be crossed by isolating overlapping chromosomal segments, starting from a previously cloned entry point in the heat-shock genes until the *TE* has been reached.

As starting DNA segment we used the clone λ122 which contains the two 70-kd heat-shock genes from 87A7 (Goldschmidt-Clermont, 1980). The orientation of the cloned DNA on the chromosome was known from the work of Ish-Horowicz and Pinchin (1980), who showed that the proximal breakpoint of the large, cytologically visible deletion ($Df(3R)kar^{D1}$) is located in clone λ122 (see Figure 2).

Our strategy was to isolate the target site of *TE98* first. This would allow us to obtain both ends of *TE98* by screening with the bordering target fragments. We therefore attempted a 'walk' in a wild-type Canton-S library (Maniatis *et al.*, 1978). The fragment 122.2 (see Figure 2), proximal to the heat-shock genes, was used as entry point. Overlapping DNA segments were subsequently isolated until the *TE98* insertion site was reached (Figure 2). Evidence that we had cloned the target site was given by the *in situ* hybridization of the λ-clone M100 to polytene chromosomes carrying the *TE98* insertion. As shown in Figure 3, M100 hybridizes to both sides of the *TE*. Therefore, this clone spans the DNA into which *TE98* has integrated. By whole genome Southern analysis (data not shown) the site of integration was narrowed down to the distal part of fragment *F8*.

Isolation of TE98 sequences from the distal end

A *TE98* library was constructed by partial digestion of *TE98/Cxf,D³* DNA with *Eco*RI and insertion into the cloning vector λSep6 (see Materials and methods). The fragment *F7* (Figure 2) was used to screen the library for the distal border sequences of *TE98*. Among the various positive clones obtained, those derived from the balancer chromosome were eliminated by backscreening with fragment *F9*. The remaining clones were shown to contain *TE98* sequences. The restriction enzyme pattern of one example, 98/2, was in agreement with the results obtained by whole genome blotting analysis of the *TE98* chromosome DNA. Its right part has the same restriction map as the wild-type DNA, whereas the left part contains new sequences (Figure 2). These new sequences were analyzed by various means. Hybridization of ^{32}P-labeled total genomic DNA to various digests of 98/2 revealed two regions containing repetitive DNA flanked by single copy DNA. The single copy fragment T2B.2 was hybridized *in situ* to wild-type polytene chromosomes and a single site of labeling was found at 3C5-7 the *roughest* locus (Figure 4). This indicates that the cloned distal end contains the *rst* side of the *TE*. Further analysis revealed that the proximal (internal) of the two repetitive sequences, designated as *NEB* (Nebuchadnezzar), is most probably a *copia*-like element. *NEB* occurs at multiple dispersed sites in the genome which differ among different stocks. Cloning and characterization of several *NEB* sequences located at other sites in the genome indicate that the pattern of restriction sites is conserved among related sequences. Fragments containing the two ends cross-hybridize with each other, indicating reiterated termini, characteristic for *copia*-like elements. No indication of inverted repeat structures could be seen by electron microscopy.

Most emphasis was placed on the other repetitive sequence found directly at the border of *TE98* since, in all cases

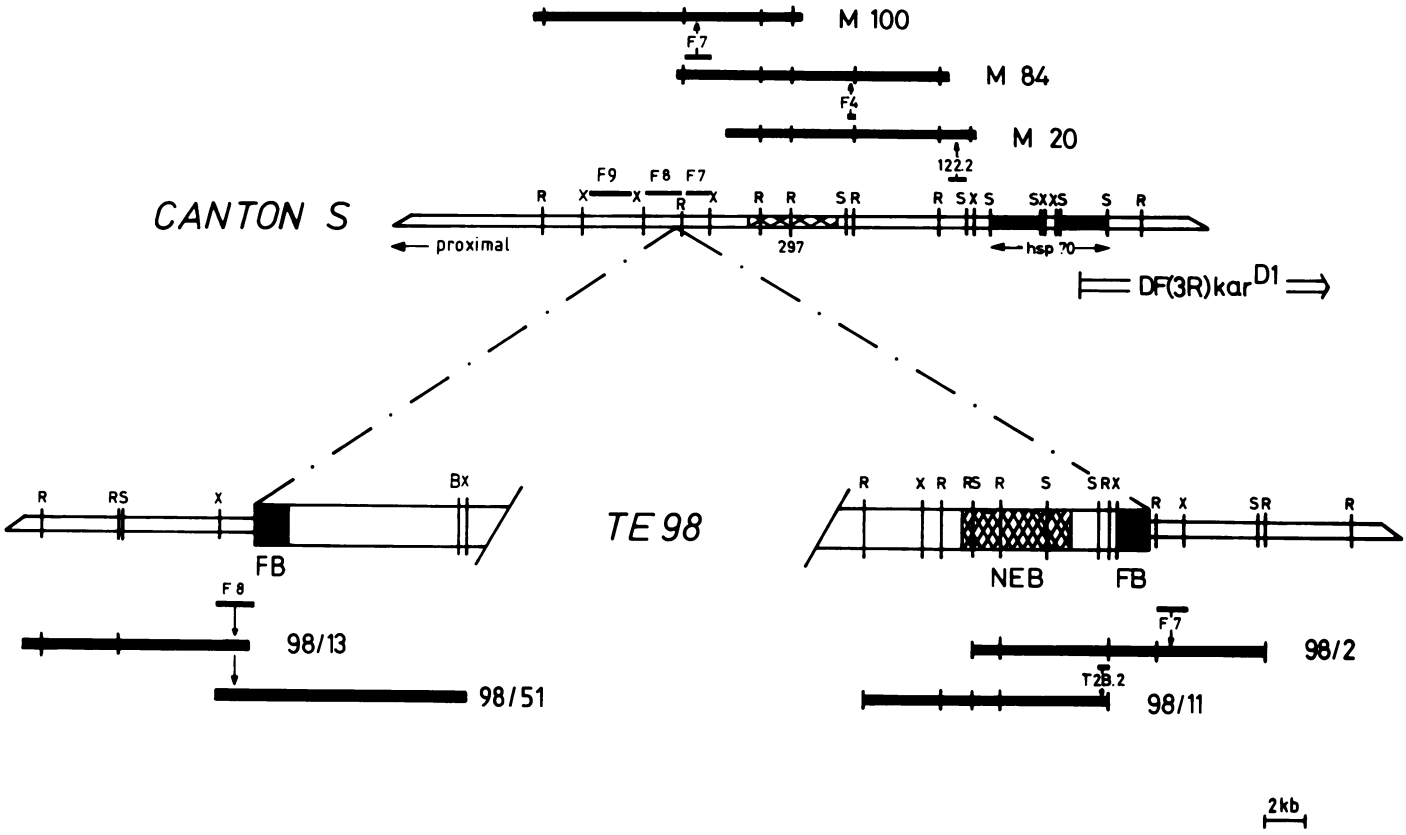

Fig. 2. Strategy used to clone *TE98* border sequences. The initial 'chromosomal walk' was carried out on a wild-type chromosome (Canton-S). As an entry fragment the previously cloned sequence 122.2 near the 70-kd heat-shock genes was used. *Df(3R)kar*^{D1} was used to map the proximal-distal orientation of the DNA (see text for details). Overlapping segments (M20, M84 and M100) were then isolated until the *TE* target site at *F8* was reached. The flanking fragment *F7* was subsequently used to isolate clones from a *TE98* library protruding into the distal border of *TE98* (98/2). 297 is a *copia*-like element that has inserted into the Canton-S chromosome. The same strategy was used for the proximal border. The fragment *F8* which flanks the proximal border was used to screen a random *Sau*3A pool of *TE98*. Only clones like 98/13 which stopped just outside the *TE* or unstable clones were obtained. However, an *Xho*I fragment defined by *F8* on a whole genome Southern analysis could be subcloned in Charon-30 and isolated (98/51). Both ends of *TE98* could be shown by cross hybridization to *FB-1* and by homoduplex analysis to contain an *FB* element. *NEB* is a *copia*-like element. Black bars indicate the cloned segments in λ, whereas the open bars indicate the corresponding chromosomal DNA. Abbreviations for restriction enzymes are: B = *Bam*HI, R = *Eco*RI, S = *Sal*I, X = *Xho*I. The size of single restriction enzyme fragments was resolved to 600 bp.

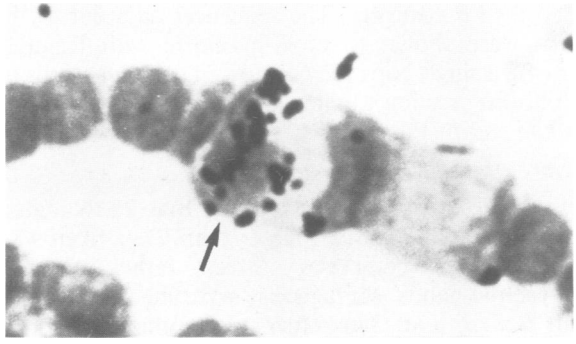

Fig. 3. An *in situ* hybridization of the target DNA (M100, Figure 2) to the *TE98/D*^{3} *CxF* chromosome. Two bands of hybridization flank the *TE* (arrow) whereas a single band is found on the balancer chromosome. Hybridization on both sides of *TE98* prove that the target site must be contained in the λ clone M100. The labeling has the same intensity on both sides, indicating that the insertion must be approximately in the middle of M100. This is confirmed by the location of the target site in fragment *F8*.

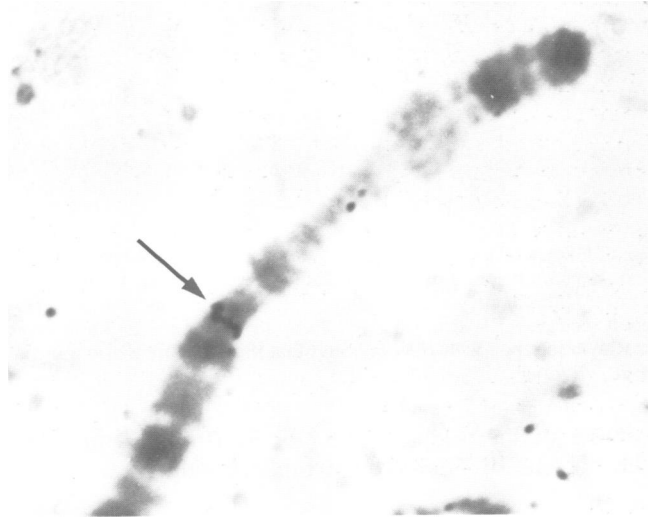

Fig. 4. *In situ* hybridization of DNA sequences from the distal end of *TE98* to the X chromosome. Fragment T2B.2 (clone 98.2, Figure 2) hybridizes to bands 3C 5-7 (arrow) the putative *roughest* locus on the X chromosome. The *white* locus maps further distal at 3C1/2. Therefore, the distal sequences represent the *roughest* end of the *TE*.

reported so far, the sequence at the ends of a transposon are essential for its mobility. Electron microscopic examinations of renatured homo- and heteroduplexes of 98/2 DNA revealed a large inverted repeat structure which coincides with the end of the *TE* (Figure 5). This element was found to be a

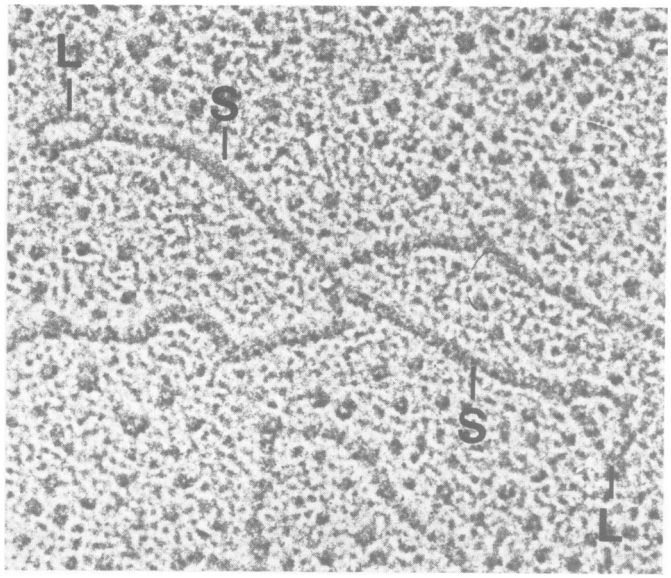

Fig. 5. Homoduplex of the clone 98/2 containing the *roughest* end of *TE98*. This clone (98/2, Figure 2) contains a large inverted repeat. The stem (S) size is ~700 bp and the loop (L) ~500 bp long. This structure was shown by cross-hybridization to *FB-1* to be a member of the *FB* family.

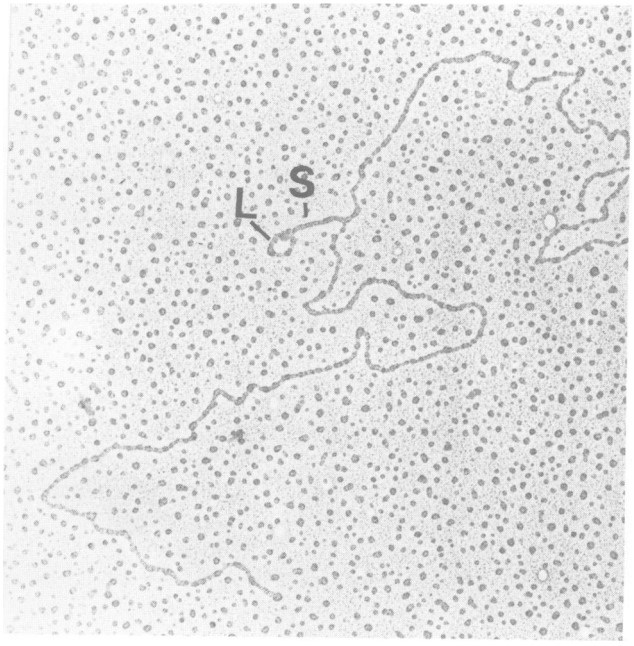

Fig. 6. Foldback DNA in clone 98/51 containing the *white* end of *TE98*. Upon denaturation and brief reannealing this clone (98/51, Figure 2) also forms a foldback structure indicating the presence of a large inverted repeat sequence. The stem (S) is ~1.4 kb, the loop (L) ~1.1 kb in length. Cross-hybridization with *FB-1* indicates that this structure is also a member of the *FB* family.

member of the foldback *(FB)* family (Potter *et al.*, 1980), since the end of *TE98* cross-hybridizes with *FB-1*.

Isolation of TE98 border sequences from the proximal end

The strategy to clone the *TE* border sequences by way of the target segment should have provided the probes to isolate both ends of *TE98*. Our attempt to screen the *TE98 Eco*RI partial library for the proximal end of *TE98* failed because

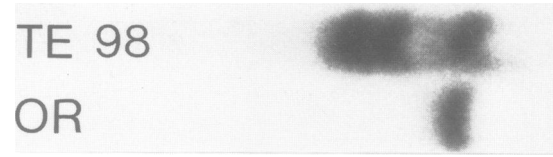

Fig. 7. Whole genome Southern analysis of the *TE98 white* region. Single copy sequences from the *white* locus region (fragment D, Figure 8) were hybridized to total *Xho*I digests of *TE98*-bearing strains and to Oregon-R wild-type strains. The largest fragment in the *TE98* digests co-migrates with the wild-type Oregon-R fragment, indicating that this fragment originates from the normal site at the X chromosome. The two additional fragments in the *TE98* digest give further evidence to the finding that *TE98* contains two copies of the *white* locus region and is a partially duplicated *TE* (for further detail see text).

the appropriate *Eco*RI fragment was too large to be cloned in a λ vector, as became evident from whole genome blotting experiments. A *Sau*3A partial library was constructed and screened unsuccessfully, either because most clones ended just before the *TE* border or because the respective clones were unstable. Finally, an *Xho*I fragment appropriate for cloning in λ was identified by whole genome Southern blotting. This fragment was cloned from a library obtained by complete *Xho*I digestion of *TE98/Df(3R)E-229* DNA (see Materials and methods) and insertion into the Charon-30 vector. In this case, the *TE98* chromosome was crossed over the large deletion *Df(3R)E-229* (Gausz *et al.*, 1981) which extends from 86F6-7 to 87B1-2, thereby eliminating all of the *TE98* target sequences from the balancer chromosome. Attempts to construct *TE98* homozygotes failed, probably because of a secondary lethal mutation on the chromosome carrying *TE98*. By using *F8* as a probe, λ-clone 98/51 (Figure 2) was isolated. The length of both proximal and distal restriction fragments of this clone correspond to those present in the *TE98* chromosome as determined by whole genome Southern analysis. Cross-hybridization of 98/51 to *FB-1* DNA also revealed an *FB* element at this end of the *TE*. The presence of an inverted repeat structure was confirmed by electron microscopy (Figure 6). This proves that both ends of *TE98* are flanked by *FB* elements. The sequences adjacent to the *FB* element were shown to cross-hybridize with fragment D (Figure 8), a single copy sequence from the *white* locus region (see following section), indicating that 98/51 originates from the *white* end of the *TE*.

Structure of TE98

Several pieces of evidence indicate that *TE98* underwent partial duplication. *TE98* is larger than *TE-1* from which it originated and comprises three, rather than two, chromosomal bands. Mutagenesis experiments performed by Awad, Bencze and Gausz (personal communication) have shown that *TE98* most probably contains two copies of the *white* gene. The frequency of ethyl methane sulfonate-induced mutations of the *white* marker in *TE98* was much lower than for any comparable recessive gene on the 3rd chromosome. Our own results confirm this finding. Whole genome Southern anaysis of *TE98* shows three bands of hybridization to single copy *white* sequences (see below): one from the original location on the X chromosome, and two from *TE98* (see Figure 7). Furthermore, *in situ* hybridization of single copy *white* sequences to chromosomes of males carrying *TE98* shows more intense labeling on the *TE* than at 3C. However, it is unlikely that the entire *TE* is duplicated since it

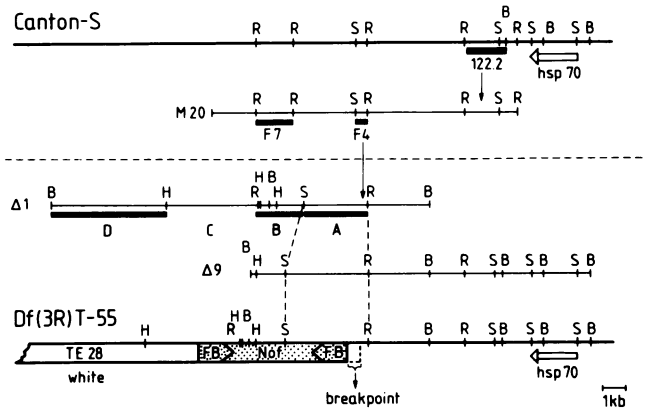

Fig. 8. Isolation of sequences from the *white* end of a *w^{a}rst$^+$-TE*. For this purpose a deficiency mutant *Df(3R)T-55* of *TE28* was used which brought the *TE28 white* end into the region which had previously been cloned (see Figure 2). Fragment *F4* from the Canton-S wild-type chromosome was used to screen a *Df(3R)T-55* library. Two overlapping clones were obtained which protruded into the *TE* (Δ1 and Δ9). The end of *TE28* was shown to be flanked by a rather large *FB* element, termed *FB-NOF*. Fragment A in Δ1 is smaller than in Δ9 which is probably a cloning artefact. Restriction sites: B = *Bam*HI, R = *Eco*RI, S = *Sal*I, X = *Xho*I, H = *Hind*III.

contains only three rather than four chromosomal bands as expected for a complete duplication. *In situ* hybridization of the element NEB, which can be taken as a marker for the *rst* end of the *TE*, only labels the distal border in *TE98*. Therefore, we conclude that *TE98* contains a duplication of the *white* locus region.

Cloning of the white locus region

Due to the initial difficulties in cloning the *white* end of *TE98*, we used an alternative way to clone the *white* locus region. For this purpose we used *TE28* located at 86EF. In the mutant *Df(3R)T-55* (Gausz *et al.*, 1981) *TE28* sequences are fused to the 87A7 heat-shock region by deletion of the intervening chromosome bands (see Figure 8). The clones isolated in the *TE98* 'walk' were used to cross the *Df(3R)T-55* breakpoint and to enter into the *TE28 white* sequences. A detailed description of the isolation of the *white* locus region using *TE28* is found in a previous publication (Goldberg *et al.*, 1982).

Comparison by whole genome Southern analysis of *TE28* DNA and *Df(3R)T-55* DNA has revealed that the deletion has left intact nearly the entire *TE* (data not shown). However, the repetitive element found at the breakpoint of *Df(3R)T-55* is ~2.5 kb smaller than the corresponding element in *TE28*. Electron microscopic analysis of homoduplexes of the clone Δ1 (Figure 8) and cross-hybridization with *FB-1* DNA demonstrates that *TE28* is flanked by a member of the *FB* family which has a rather large, characteristic internal segment of 4.5 kb (Goldberg *et al.*, 1982). This specific member of the *FB* family, located at the *white* end of *TE28*, was designated as *FB-NOF* (Nofretete) (Goldberg *et al.*, 1982).

Correlation of FB sequences with other large transposable elements of the white locus

The finding that *FB* elements are flanking the w^a*rst*$^+$*-TE* strongly suggests that they might be responsible for the mobility of this large segment of DNA. Therefore, it was of interest to see whether they are also associated with the independently isolated *Tp*s derived from *white-crimson* (Green, 1969a, 1969b) and the *Tp* w^{+IV} transposon (Rasmuson *et al.*, 1980).

For this purpose, fragments containing the internal segment of *FB-NOF* (fragment B; Figure 8) and the repeat (fragment A) were subcloned in pBR322. *In situ* hybridization to polytene chromosomes of w^c strains showed strong hybridization of both fragments at 3C, the expected location of the w^c allele (data not shown). Fragment B furthermore hybridizes to five other sites in the genome, whereas fragment A hybridizes to many other dispersed sites in the genome. These results are consistent with those obtained by Collins and Rubin (1982) and Levis *et al.* (1982b). They have cloned the element supposed to cause the w^c instability. Their published restriction map of the element corresponds exactly to that of the member of the *FB* family which we termed *FB-NOF* (Goldberg *et al.*, 1982).

We further analyzed Tp w^c-1, a transposition of the w^c allele to the 3rd chromosome. *In situ* hybridization of single copy *white* DNA to Tp w^c-1 places the transposon to 62E, which is in agreement with the genetic mapping data (Green, 1969a). This same site was labeled by the two fragments of *FB-NOF* indicating that the w^c transposon is also associated with *FB* sequences. A more detailed analysis of the w^c transposon will be the content of a further communication.

The same analysis was performed with the *Tp* w^{+IV} transposon which has moved from the X to the 4th chromosome. The *white* gene in this new position gave rise to a mottled phenotype, which suggested that it might have moved into a heterochromatic region (Rasmuson *et al.*, 1980). Hybridization with single copy *white* sequences placed the transposon close to the telomere of the 4th chromosome at 102F. Again both fragments of *FB-NOF* hybridize to the same new position (data not shown). Although the resolution of *in situ* hybridization to polytene chromosomes does not provide conclusive proof, it gives at least suggestive evidence that *Tp* w^{+IV} is also associated with *FB-NOF* sequences.

FB sequences at the ends of TEs are heterogeneous

The *FB* sequences found at the ends of *TEs* were shown to cross-hybridize with *FB-1*, a member of the *FB* family (Potter *et al.*, 1980). However, the homology is presumably confined to the inverted repeat structure. Analysis of the various elements cloned from *TE* borders reveal the same heterogeneity in size and insert content, as Potter and co-workers found for randomly isolated *FB* elements. Even within closely related *TEs* this heterogeneity is observed. The insert segment defining *FB-NOF* (fragment B; Figure 8) found in *TE28*, for example, was not found to hybridize to *TE98* or *TE77* in polytene chromosomes (see below). We presume that in these two cases this internal segment was missing, leaving only the inverted repeats. Since *TE98* has been shown to retain transposing properties (G.Ising, personal communication) this would argue that either only the inverted repeats are responsible for the transposing property or alternatively, that a putative transposition function can be provided *in trans* by complete *FB* elements or host genes.

Transposon sequences as a means to isolate genes into which the TE has integrated

Cloned single copy sequences from the *TE* are of substantial interest for gene cloning. Since *TEs* have been isolated and mapped to well over 100 different sites all over the genome, and new positions can be screened genetically, the single copy sequences at the *TE* borders can be used to isolate

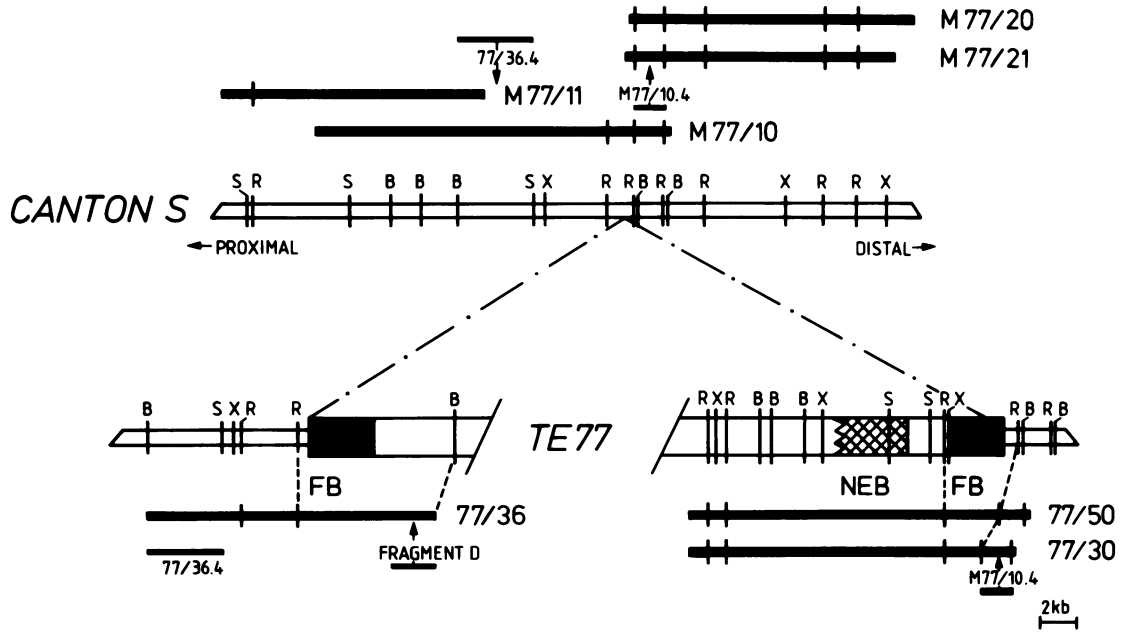

Fig. 9. Isolation of *TE77* border sequences and of adjacent target sequences. Fragment D (Figure 8) from the *white* end of the *TE* was used to screen a *Sau*3A partial pool of *TE77* DNA inserted into Charon-30. The clone 77/36 was shown by cross-hybridization with *FB* DNA to contain the *TE* border and to extend into the flanking sequences. The single copy fragment 77/36.4 was *in situ* hybridized to wild-type polytene chromosomes showing hybridization at 89DE, the target site of *TE77* (data not shown). This fragment was furthermore used to isolate DNA segments from a wild-type Canton-S library spanning the target site. The clone M77/10 was shown by whole genome Southern analysis and *in situ* hybridization to *TE77* chromosomes to contain the entire target site. The fragment M77/10.4 was used to enter again into the *TE77* and to isolate the *rst* end. This end looks similar to the *rst* end of *TE98*, although some degree of internal rearrangement has apparently occurred. The *NEB* element is partially deleted and flanked by some unknown repetitive DNA. The interrupted lines denote the difference between the cloned fragments containing *FB* sequences and the expected size of the fragment as found by whole genome Southern blotting. This discrepancy presumably results from the fact that *FB* sequences are very reluctant to be cloned in λ and often are deleted. The proximal-distal orientation of the DNA was deduced from *in situ* hybridization to the *TE77* chromosome of flanking fragments. The fragment 77/36.4 labeled the proximal end of *TE77*, whereas M77/10.4 labeled the distal end. For abbreviations of restriction enzymes see Figure 2.

λ-clones which extend into neighboring DNA sequences. This strategy allows the isolation of the sequence of interest into which the *TE* has integrated and provides a general strategy for gene cloning (Gehring and Paro, 1980).

TE77, which has integrated close to the *bithorax* region was used to test this method and characterize an additional *TE*. As starting point, single copy sequences from the *white* end of the *TE* were used (fragment D; Figure 8). A *Sau*3A partial digest of *TE77* cloned into Charon-30 was screened. Clones spanning the border were distinguished from those originating from the normal location of *white* at 3C, by back-screening with *FB* DNA. The clone 77/36 was shown to cross the end of *TE77* and extend into adjacent sequences (Figure 9). In wild-type chromosomes, the fragment 77/36.4 hybridizes to 89DE, the insertion site of *TE77*. Using this fragment, the sequences surrounding the target site were cloned from the Canton-S library (Figure 9). Thus, the unique sequences of the *TE* can be used for the isolation of DNA sequences at the target site, in this case the *bithorax* region. The target site has been mapped ~70 kb proximal to *bithorax*^{34e} by W.Bender (personal communication). As indicated in Figure 7, fragment M77/10.4 was subsequently used as a probe for the isolation of the distal end of *TE77*. As expected, we again found an *FB* element at the edge of the *TE*.

Discussion

Molecular cloning of *TE* borders has revealed the presence of a moderately repetitive element at each end. These elements were shown to be members of the previously characterized *FB* family (Potter *et al.*, 1980; Truett *et al.*, 1981; Goldberg *et al.*, 1982). Indirect evidence indicates that *FB* elements are mobile by themselves. Their association with the ends of the *TE* suggests that they are also involved in the mobilization of this large composite transposon. For all transposable elements described so far (for review see Kleckner, 1981; Shapiro and Cordell, 1982) the ends of the element are essential for mobility. They appear to be the substrate for the enzymes involved in transposition, and in some cases the terminal repeats or internal sequences of the element may code for functions that are required for transposition. Therefore, the presence of *FB* elements at the ends of the *TE* suggests that *FB* elements are not only mobile by themselves, but also that two *FB* elements can mobilize an intermediary DNA segment.

In a previous publication (Gehring and Paro, 1980), we proposed that this mobilization might be due to another dispersed repetitive element, *copia*, which was found to be associated with both the *TE* and the w^a mutation. Subsequent studies (Goldberg *et al.*, 1982; Bingham and Judd, 1981; Bingham *et al.*, 1981; Levis *et al.*, 1982a) have shown that w^a mutations arose by an insertion of *copia* into the *white* locus. Since the w^a allele is carried by the original *TE*, we found this strong link between *copia* and the *TE*. However, many new transpositions have been recovered from red variants of the *TE* (Ising and Block, 1981) which have lost the *copia* element. This evidently separates the transposing property of *copia* from the transposing property of the *TE*. Similarly, we can rule out *NEB* as a possible cause for mobility since it is present only in some *TE*s but not in others. On the other hand, the correlation between *FB* elements and the capacity for transposition is very strong. Not only have we found it at the ends of all the *TE*s analyzed but also on *Tp* w^c*-1* and

Tp w^{+IV}, two additional transposons isolated independently. An additional piece of evidence comes from the analysis of the w^{DZL} mutant (Bingham, 1980; Levis *et al.*, 1982b). This mutation is due to the insertion of a large DNA segment into the *white* locus region. The insert consists of a piece of 'foreign' DNA flanked by a pair of *FB* elements. Since in wild-type stocks, the 'foreign' DNA sequence is unique and located on the second chromosome, w^{DZL} can also be interpreted as the result of a transposition of a large DNA segment flanked by a pair of *FB* elements. Therefore, we propose that a pair of *FB* elements can mobilize intermediary DNA segments.

The individual *FB* elements are highly variable with respect to the size of the inverted repeats as well as the nature of the central insert, even when closely related *TEs* are compared. This raises the question of how to define an intact and functional *FB* element. Some of the elements may be partially defective since they lack the internal segment and retain only the inverted terminal repeats. On the basis of genetic evidence Green (1969a, 1969b) postulated the existence of a 'controlling element' separable from the wild-type *white* locus. This 'controlling element' was thought to have induced the *white-crimson* (w^c) mutation by inserting into the *white* locus and to be the cause for both the instabilities associated with this allele, and also for the transpositions of w^c to the third chromosome. Our data indicate that the member of the *FB* family which we designated as *FB-NOF* (Goldberg *et al.*, 1982) is the molecular entity corresponding to the genetically defined 'controlling element' that is present in both w^c and w^c-derived transpositions. *FB-NOF* is a rather large element characterized by a 4.5-kb internal segment. This is confirmed by Collins and Rubin (1982) and Levis *et al.* (1982b) who also found that w^c is caused by an *FB* element that has the same restriction pattern as *FB-NOF*. Furthermore, the *FB-NOF* element was found to be associated with the *white* end of *TE28*, but in this case the site of insertion is more distal than in w^c. *In situ* hybridization experiments also suggest that *FB-NOF* is associated with *Tp* w^{+IV} (Rasmuson *et al.*, 1980). Therefore, *FB-NOF* could represent the intact functional *FB* element. However, the *FB* elements flanking *TE98* and *TE77* lack the internal segment characteristic of *FB-NOF*. Since at least *TE98* was shown to transpose further although at a very low frequency (G.Ising, personal communication), the internal segment may not be essential. In this case, the inverted repeats alone may code for the functions required for transposition. The sequencing data of Potter (1982a) make this possibility rather unlikely, since the inverted repeats consist of short tandemly repeated sequences with frequent translational stop codons in all three possible reading frames and are unlikely to code for a protein. However, the inverted termini may only serve as a substrate for 'transposases' encoded by other transposable elements or cellular genes. There may be a correlation between the length of the inverted repeats and the frequency of transposition, but our sample size is too small to decide on this point. The function of the inverted terminal repeats may be tested by constructing artificial transposons and introducing them into the germline of different recipient stocks.

The combination of short tandem repeats within large inverted repeats makes the *FB* elements highly recombinogenic (Potter, 1982b). Such a structure allows for unequal crossing over within a chromosome and between different chromosomes, which may lead to the variability that we observe. Recombination between *FB* elements may also provide a mechanism for transposition of the *TE*. In this case one would predict that the *TE* inserts only into target sites where an *FB* element is located, and that the *TE* excises from its site of origin. Both of these predictions can be tested experimentally. The recombinogenic properties of the *FB* elements are also reflected by their instability in *Escherichia coli*. It proved to be difficult to clone these elements in λ vectors, resulting in the underrepresentation of *TE* borders in λ libraries and in clones having partially deleted inverted repeats as compared with the *in vivo* situation. The same difficulties were met by Collins and Rubin (1982) in cloning *FB* sequences. To some extent this problem can be circumvented by cloning into plasmids in a *recA*$^-$ host.

The occurrence of composite transposons mobilizing large segments of cellular DNA like the *TEs* provides an explanation for some important evolutionary phenomena. In particular, such composite transposons provide a mechanism for generating dispersed gene duplications. Two basically different mechanisms for transposition can be envisaged: in one case the transposon is excised from its original chromosomal site and inserted into a new site. Alternatively, the transposon may remain at the original site and a replica transposes to a new location. In higher organisms like *Drosophila*, both mechanisms can lead to a duplication, since following sexual reproduction the gamete carrying a new transposition will be combined with a gamete carrying the transposon in the original position. In the case of interchromosomal transposition the independent reassortment of the chromosomes during subsequent generations will give rise to a duplication. If the transposition is intrachromosomal, recombination between the two sites may generate a duplication-carrying chromosome. This mechanism may explain, for example, the dispersed arrangement of the multiple actin genes in *Drosophila* (Fyrberg *et al.*, 1980; Tobin *et al.*, 1980). Since such duplications may be followed by subsequent divergence of the different gene copies, this mechanism provides the raw material for the generation of new genes in evolution. Composite transposons might also mobilize parts of genes like promoters or exons and combine them with parts of other genes which can account for some of the 'tinkering' (Jacob, 1981) occurring during evolution.

Materials and methods

Drosophila strains

Canton-S: wild-type strain.

Oregon-R: wild-type strain.

TE98 (3-50.6): *w;ri st cu TE98/CxF, ru h D*^{3}. *TE98* is a transposition of the red form of *TE-1* to the third chromosome at 87A5-6 near the 70-kd heat-shock genes (Ising and Block, 1981). *w/w*$^+$*;th st cu TE98/th st Df(3R)E-229:* in this strain the *TE98*-bearing chromosome was crossed to a third chromosome carrying a large deletion. This deletion extends from 86F6-7 to 87B1-2 (Gausz *et al.*, 1981) thereby eliminating all of the target site of *TE98* on the homologous chromosome.

TE77 (3-58.7). *w;TE77*: a strain carrying *TE77* in homozygous condition at 89DE on the third chromosome. *TE77* is a transposition of the red form of *TE-1* (Ising and Block, 1981).

TE28 (3-50.3). *w;(ri)cu TE28: TE28* is a transposition of the *white-apricot* form of *TE-8* to the third chromosome at 86EF (Ising and Block, 1981).

Df(3R)T-55. ri th Df(3R)T-55 sr e^s*/TM3, Sb e*^{s}*Ser*: this strain has a deletion between *TE28* and the 70-kd heat-shock genes (Gausz *et al.*, 1981).

Tp w^c*-1/Ubx*^{130} (3–3.8). *w;Tp* w^c*-1/Ubx130*: this strain contains a transposition of the *white-crimson* allele to the third chromosome (62E) (Green, 1969a).

Tp w^{+IV} (4th chromosome). *C(1)DX w f/y;Tp* w^{+IV} *x sc z* w^{is}*;Tp* w^{+IV}: this strain carries a transposition of the w^{sp} w^{17G} duplication to the telomere of the fourth chromosome at 102F. The *white* gene in this position gives a mottl-

ed phenotype (Rasmuson *et al.*, 1980).

For a description of the marker mutants and balancer chromosomes see (Lindsley and Grell, 1968).

Drosophila recombinant libraries

The isolation of *Drosophila* DNA and the construction of recombinant libraries containing partial digests of *Drosophila* DNA were already described in Goldberg *et al.* (1982). Where λSep6 (constructed by E.Meyerowitz and D.Kemp; defined in Davis *et al.*, 1980) and Charon-30 (Rimm *et al.*, 1980) were used as vectors, the tails were isolated by digesting λ DNA with the appropriate restriction enzyme, separation on an agarose gel and recovery by electroelution. All procedures used with recombinant λ vectors and for 'walking along the chromosome' were described in Goldberg *et al.* (1982). The Canton-S library was kindly provided by J.Lauer (Maniatis *et al.*, 1978).

Hybridizations

Southern blot analysis and *in situ* hybridizations to polytene chromosomes have been described previously in Goldberg *et al.* (1982).

Orcein staining of polytene chromosomes

Where heat-shock puffs had to be induced in order to mark a certain region on the chromosome, the larvae were heat-shocked at 37°C for 30 min. Salivary glands were dissected out of the larvae in a drop of 45% acetic acid. Glands were fixed for 30 s in 1 N HCl and stained in a drop of Orcein solution (2% Orcein in equal parts of 85% lactic acid, glacial acetic acid and distilled water; the mixture was boiled in a steam bath for 2 h and filtered while hot). After 10–15 min the stain was diluted with several drops of lacto-acetic acid (equal parts of lactic-acid and 60% acetic acid). The glands were transferred to a drop of lacto-acetic acid on a microscope slide and sealed with a coverslip. The cells were broken and the chromosomes spread by tapping with the rubber end of a pencil to the coverslip.

Restriction and heteroduplex mapping

Restriction enzymes were obtained from New England Biolabs and Boehringer. Digestions were carried out according to the instructions given by the suppliers. Homo- and heteroduplex mapping was performed according to Davis *et al.* (1971).

Acknowledgements

We wish to thank G.Ising, M.M.Green, D.Ish-Horowicz and B.Rasmuson for their valuable strains; S.Potter for *FB-1* DNA and M.Goldschmidt-Clermont for λ122 DNA. We are also grateful to Monika Zulliger and Pam Beattie for electron microscopy; Käthi Paro-Peter and Urs Kloter for technical assistance; Erika Wenger-Marquardt for typing; and R.Klemenz, R.Garber and M.Levine for critical reading of the manuscript. This work was supported by the Kanton Basel-Stadt, the Swiss National Science Foundation to WJG, and a Roche Stipendium to MLG.

References

Bingham,P.M. (1980) *Genetics,* **95**, 341-353.
Bingham,P.M. and Judd,B.H. (1981) *Cell,* **25**, 705-711.
Bingham,P.M., Levis,R. and Rubin,G.M. (1981) *Cell,* **25**, 693-704.
Collins,M. and Rubin,G.M. (1982) *Cell,* **30**, 71-79.
Davis,R.W., Simon,M. and Davidson,M. (1971) *Methods Enzymol.,* **21**, 413-428.
Davis,R.W., Botstein,D. and Roth,J.R. (1980) *Advanced Bacterial Genetics,* published by Cold Spring Harbor Laboratory Press, NY, p. 241.
Flavell,A.J. and Ish-Horowicz,D. (1981) *Nature,* **292**, 591-595.
Fyrberg,E.A., Kindle,K.L. and Davidson,N. (1980) *Cell,* **19**, 365-378.
Gausz,J., Gyurkovics,H., Bencze,G., Awad,A.A., Ish-Horowicz,D. and Holden,J.J. (1981) *Genetics,* **98**, 775-789.
Gehring,W.J. and Paro,R. (1980) *Cell,* **19**, 897-904.
Goldberg,M.L., Paro,R. and Gehring,W.J. (1982) *EMBO J.,* **1**, 93-98.
Goldschmidt-Clermont,M. (1980) *Nucleic Acids Res.,* **8**, 235-252.
Green,M.M. (1967) *Genetics,* **56**, 467-482.
Green,M.M. (1969a) *Genetics,* **61**, 429-441.
Green,M.M. (1969b) *Genetics,* **61**, 423-428.
Ish-Horowicz,D. and Pinchin,S.M. (1980) *J. Mol. Biol.,* **142**, 231-245.
Ising,G. and Ramel,C. (1976) in Ashburner,M. and Novitski,E. (eds.), *The Genetics and Biology of Drosophila,* Vol. **1b**, Academic Press, NY, pp. 947-954.
Ising,G. and Block,K. (1981) *Cold Spring Harbor Symp. Quant. Biol.,* **45**, 527-551.
Jacob,F. (1981) *Le Jeu des Possibles,* published by Fayard, Paris.
Kleckner,N. (1981) *Annu. Rev. Genet.,* **15**, 341-404.
Levis,R., Bingham,P.M. and Rubin,G.M. (1982a) *Proc. Natl. Acad. Sci. USA,* **79**, 564-568.
Levis,R., Collins,M. and Rubin,G.M. (1982b) *Cell,* **30**, 551-565.
Lindsley,D.L. and Grell,E.H. (1968) *Carnegie Inst. Washington Publ. No. 627.*
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K., Efstratiadis,A. (1978) *Cell,* **15**, 687-701.
Potter,S.S., Truett,M., Phillips,M. and Maher,A. (1980) *Cell,* **20**, 639-647.
Potter,S.S. (1982a) *Nature,* **297**, 201-204.
Potter,S.S. (1982b) *Mol. Gen. Genet.,* **188**, 107-110.
Rasmuson,B. and Green,M.M. (1974) *Mol. Gen. Genet.,* **133**, 249-260.
Rasmuson,B., Montell,I., Rasmuson,A., Svahlin,H. and Westerberg,B.-M. (1980) *Mol. Gen. Genet.,* **177**, 567-570.
Rimm,D.L., Horness,D., Kucera,J. and Blattner,F.R. (1980) *Gene,* **12**, 301-309.
Rubin,G.M., Kidwell,M.G. and Bingham,P.M. (1982) *Cell,* **29**, 987-994.
Rubin,G.M., Spradling,A.C. (1982) *Science (Wash.),* **218**, 348-353.
Schedl,P., Artavanis-Tsakonas,S., Steward,R., Gehring,W.J., Mirault, M.E., Goldschmidt-Clermont,M., Moran,L. and Tissières,A. (1978) *Cell,* **14**, 921-929.
Shapiro,J.A. and Cordell,B. (1982) *Biol. Cell.,* **43**, 31-54.
Shiba,T. and Saigo,K. (1983) *Nature,* **302**, 119-124.
Spradling,A.C. and Rubin,G.M. (1981) *Annu. Rev. Genet.,* **15**, 219-264.
Spradling,A.C. and Rubin,G.M. (1982) *Science (Wash.),* **218**, 341-347.
Tobin,S.L., Zulauf,E., Sanchez,F., Craig,E.A. and McCarthy,B.J. (1980) *Cell,* **19**, 121-131.
Truett,M.A., Jones,R.S. and Potter,S.S. (1981) *Cell,* **24**, 753-763.