Nucleic Acids Research, Vol. 20, No. 21 5541–5548

Cloning and characterization of the *Drosophila* homolog of the xeroderma pigmentosum complementation-group B correcting gene, *ERCC3*

Marcel H.M.Koken[+], Cees Vreeken[1,2], Sandra A.M.Bol[1], Ngan Ching Cheng[§], Iris Jaspers-Dekker, Jan H.J.Hoeijmakers, Jan C.J.Eeken[1,2], Geert Weeda and Albert Pastink[1,*]
Department of Cell Biology and Genetics, Medical Genetics Centre, Erasmus University, PO Box 1738, 3000DR Rotterdam, [1]Department of Radiation Genetics and Chemical Mutagenesis, Medical Genetics Centre, State University Leiden, PO Box 9503, 2300 RA Leiden, [2]J.A.Cohen Institute, Interuniversity Research Institute for Radiopathology and Radiation Protection, Wassenaarseweg 62, 2333 AL Leiden, The Netherlands

Received September 11, 1992; Revised and Accepted October 8, 1992 EMBL accession no. X68309

ABSTRACT

Previously the human nucleotide excision repair gene *ERCC3* was shown to be responsible for a rare combination of the autosomal recessive DNA repair disorders xeroderma pigmentosum (complementation group B) and Cockayne's syndrome (complementation group C). The human and mouse ERCC3 proteins contain several sequence motifs suggesting that it is a nucleic acid or chromatin binding helicase. To study the significance of these domains and the overall evolutionary conservation of the gene, the homolog from *Drosophila melanogaster* was isolated by low stringency hybridizations using two flanking probes of the human *ERCC3* cDNA. The flanking probe strategy selects for long stretches of nucleotide sequence homology, and avoids isolation of small regions with fortuitous homology. *In situ* hybridization localized the gene onto chromosome III 67E3/4, a region devoid of known *D.melanogaster* mutagen sensitive mutants. Northern blot analysis showed that the gene is continuously expressed in all stages of fly development. A slight increase (2–3 times) of *ERCC3*^{Dm} transcript was observed in the later stages. Two almost full length cDNAs were isolated, which have different 5′ untranslated regions (UTR). The SD4 cDNA harbours only one long open reading frame (ORF) coding for *ERCC3*^{Dm}. Another clone (SD2), however, has the potential to encode two proteins: a 170 amino acids polypeptide starting at the optimal first ATG has no detectable homology with any other proteins currently in the data bases, and another ORF beginning at the suboptimal second startcodon which is identical to that of SD4. Comparison of the encoded ERCC3^{Dm} protein with the homologous proteins of mouse and man shows a strong amino acid conservation (71% identity), especially in the postulated DNA binding region and seven 'helicase' domains. The *ERCC3*^{Dm} sequence is fully consistent with the presumed functions and the high conservation of these regions strengthens their functional significance. Microinjection and DNA transfection of *ERCC3*^{Dm} into human xeroderma pigmentosum (c.g. B) fibroblasts and group 3 rodent mutants did not yield detectable correction. One of the possibilities to explain these negative findings is that the *D.melanogaster* protein may be unable to function in a mammalian repair context.

INTRODUCTION

For all organisms it is of vital importance to secure reliability of genetic information. Because the genome of the cell is constantly under attack by a plethora of DNA damaging agents, all organisms had to develop efficient systems to recognize and remove DNA injury. Lesions in DNA—if unrepaired—have immediate deleterious effects on transcription and replication, or, after fixation into permanent mutations, they can change the coding properties of genes. In *Escherichia coli* several of the biochemical pathways leading to the removal of DNA damage are rather well understood. Among these is the nucleotide excision repair (NER) process in which a minimum of 6 proteins participates in the elimination of a considerable number of different DNA lesions (for review see (1)). A complex consisting of two molecules of UvrA and one molecule of UvrB, forming a DNA helicase, scans the DNA for local distortions, caused by

* To whom correspondence should be addressed

Present addresses: [+]Hôpital St Louis, Laboratoire des Rétrovirus et Rétrotransposons des Vertèbres, Batiment INSERM, 1610 Rue de la Grange aux Belles, 75010 Paris, France and [§]Department of Anthropogenetics, Academical Medical Centre, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

damage. After tracing DNA injury the two UvrA proteins leave the complex, whereas UvrB is tightly attached to the DNA, and serves as a tag for the third polypeptide involved, UvrC. The UvrBC complex makes incisions in the damaged strand at both sides of the lesion, after which a second DNA helicase, UvrD, removes the damaged part from the DNA backbone. Finally, DNA polymerase and ligase fill in and close the gap. The *E.coli* system is probably relatively simple when compared to the process in eukaryotes, where already more than eleven NER genes in yeast and man have been identified and in part cloned. Many of the cloned human NER genes have been isolated utilizing DNA transfection of normal human genomic DNA or cDNA into repair-deficient mutant cells, either laboratory-derived rodent cell lines or cells from patients with the rare autosomal recessive disease xeroderma pigmentosum (XP). Individuals suffering from this repair disorder are characterized by hypersensitivity to sunlight (UV), pigmentation abnormalities and a high incidence of skin tumors in sun-exposed areas (for review see (2)). Cell fusion experiments showed that the XP (and rodent mutant) cells can be divided into at least eight (respectively ten) complementation groups (c.g.). DNA transfections have been used for the cloning of six human excision repair genes (3, 4, 5, 6, 7, 8), a number of which are shown to be well-conserved during evolution. The ERCC1 geneproduct corrects the UV-sensitive, DNA excision repair deficient, Chinese hamster mutant cell lines of complementation group 1. The protein is homologous to the yeast NER protein RAD10 and shares additional regions of similarity with parts of the prokaryotic *E.coli* UvrA and UvrC polypeptides (9). The *XPAC* gene isolated after transfection of mouse genomic DNA into XP-A cells (7, 10), encodes a Zn^{2+}-finger protein conserved during evolution (11), and is homologous to the yeast RAD14 DNA excision repair geneproduct (12). *ERCC2* (4, in ref.13) and *ERCC3* (14), isolated by correction of hamster mutants of rodent complementation group 2 and 3, respectively, appear to be involved in xeroderma pigmentosum, complementation groups D (*ERCC2*) and B (*ERCC3*). Mutations in the *ERCC2* gene, the human homolog of the yeast DNA excision repair helicase, RAD3, are also underlying two other hereditary diseases, Trichothiodystrophy and Cockayne's syndrome (CS) that occur in some of the XP-D patients (in ref.13). Cockayne syndrome patients are characterized by a small stature, wizened appearance, sun-sensitivity, and often mental and physical retardation, but no elevated risk for cancers (15). The only three XP-B patients that have been identified, suffer also from CS ((14) and unpublished data). Both ERCC2 and ERCC3 are members of a recently defined group of putative DNA helicases, because they share 7 consecutive amino acid domains found in this family (16). When the *ERCC3* sequence was compared with databases and known NER genes no corresponding yeast equivalent was found. However, low stringency hybridization experiments with DNA of various eukaryotic organisms strongly suggested that the *ERCC3* gene is rather well conserved in evolution (17). To study this evolutionary conservation and functional and/or structural importance of putative domains in the encoded protein we isolated the *Drosophila melanogaster* (and both the *Saccharomyces cerevisiae* (in ref.13, 18) and *Schizosaccharomyces pombe* (in ref.13, 19)) homolog of the *ERCC3* excision repair gene, by applying a hybridization strategy which eliminates the isolation of many false positives and increases the chance to isolate homologous genes with low nucleotide sequence conservation. Here we present the isolation and characterization of the *D.melanogaster ERCC3* homolog (designated here as *ERCC3^{Dm}*), and provide further evidence that the eukaryotic DNA repair system as a whole is very well conserved during evolution.

MATERIALS AND METHODS

DNA/RNA isolation, restriction analysis, blot/library hybridizations, and cloning procedures

Standard DNA and RNA manipulations were done as described (20). Conditions for (restriction) enzyme incubations were as recommended by the manufacturer.

PCR primers for the preparation of two flanking probes of the human *ERCC3* gene were designed as follows:

		Met
Couple 1:	5′ sense:	5′ATGGGCAAAAGAGACCGAGCG 3′
	middle anti sense:	5′CAGCTCCATCCAGTTGGCTTG 3′
Couple 2:	middle sense:	5′CAGAATAATGGCTACATCGC 3′
		Stop
	3′ anti-sense:	5′TCATTTCCTAAAGCGCTTGAA 3′

the human *ERCC3* cDNA (position 1581) around which the probes were designed. Met and Stop (underlined) indicate the start and stop codon of human *ERCC3*. PCR was performed exactly as described (14) using the plasmid pCD1 (14) as template.

Total RNA from flies at different stages of development was isolated (21), after which poly(A)$^+$ RNA was prepared via oligo(dT) cellulose chromatography (20). RNA samples were fractionated on 1% agarose formaldehyde gels and transferred onto nitrocellulose membranes and hybridized overnight at 42°C (20).

DNA samples separated on 0.7−1% agarose gels were blotted onto nylon membranes (Hybond^R, Zetaprobe^R). Probe labelling and (low stringency (56°C)) hybridization were done as delineated before (20, 22).

The oligonucleotides used for the characterization of the SD2 and SD4 cDNAs were constructed as follows:

SD2 specific oligo:	5′ATACGTGATTTACCAACTATTTGCATT 3′
SD4 specific oligo:	5′GATATCTTTGTTTAATTTCTTGAGCTCGG 3′

Sequence analysis

*Eco*RI fragments of *ERCC3^{Dm}* lambda cDNA clones, SD2 and SD4, were subcloned into pUC119 and pUC120 phagemid vectors (20). Single-stranded DNA was prepared (20), and used for fluorescent DNA sequencing. Reactions were carried out according to Sanger *et al.*(23), using an AutoRead sequencing kit (Pharmacia/LKB) and analyzed on an ALF automatic sequencer (Pharmacia/LKB).

In situ hybridization

Salivary glands from third instar larvae of wild-type *Berlin K* and *Amherst M56i* flies were dissected and squashed. The 8.5kb genomic *SalI* fragment was nick-translated with biotinylated dUTP, and used for *in situ* hybridization as described (24). Signals were visualized using streptavidin-alkaline phosphatase (BlueGENE detection system, Gibco/BRL). No counterstaining was used, instead the preparations were analyzed by combining phase contrast with conventional light microscopy.

Complementation studies

DNA transfections, microinjection, UV survival and unscheduled DNA synthesis assays. Complementation experiments with the *ERCC3*^{Dm} made use of a mammalian expression vector, pSLM, which harbours a SV40 late promoter. In this modified pSVL vector (Pharmacia/LKB), the original polylinker (*Xho*I, *Xba*I, *Sma*I, *Sac*I, *Bam*HI) has been replaced by the more suitable multiple cloning site of pTZ19R (*Eco*RI, *Sac*I, *Kpn*I, *Sma*I, *Bam*HI, *Xba*I, *Sal*I/*Hinc*II, *Pst*I, *Sph*I, *Hind*III) (Pharmacia/LKB). Polylinker sites that were not unique any more after this interchange (*Eco*RI, *Kpn*I, *Sal*I, *Hind*III) have subsequently been removed from the plasmid backbone by standard procedures. With the exception of *Hinc*II, *Pst*I and *Sph*I all other sites are now unique.

For the subcloning of *ERCC3*^{Dm} into pSLM, oligonucleotide primers located in both arms of λgt11 flanking the insert were used to amplify the insert of lambda clone SD2:

5'oligo: 5'CAGCCCGGGTCGACGTAGCGACCGGCGCTCAG3'
3'anti-sense oligo: 5'CGTCCCGGGAGTCGACTGGAGCCCGTCAGTATCG3'

The *Sma*I cloning site is underlined, whereas the *Sal*I cloning site is indicated by shading.

The PCR product was blunt ended with the Klenow fragment of DNA polymerase, phosphorylated with T4 polynucleotide kinase, and introduced into the *Sma*I site of the plasmid pSLM. Two independent PCR derived clones were isolated, designated pSVD2 and pSVD3. To remove the first ATG the *Dra*I site, in front of the second ATG, specifying the *ERCC3*^{Dm} open reading frame was subsequently used to isolate from both clones a *Dra*I-*Sal*I (located in the 3' anti-sense oligo) fragment, harbouring *ERCC3*^{Dm}. This portion was subsequently ligated into the SmaI-SalI sites of plasmid pSLM, giving rise to the deletion constructs, pSVD2 and pSVD3. 6 μg of the *ERCC3*^{Dm} constructs together with 2 μg of pRSVneo were transfected into Chinese hamster 27-1 cells (belonging to rodent complementation group 3) using Lipofectin^{R} reagent (Gibco/BRL) as described (25). After 48 hours cells were selected on 800 μg/ml G418 for DNA uptake. UV-survival was determined on the exponentially growing neomycin resistant mass-population (26).

The expression constructs were also microinjected into the nuclei of XP homopolykaryons of the fibroblast cell line XPCS1BA (belonging to XP complementation group B) using procedures previously described (27). Cells were incubated for 24−48 hr allowing the injected DNA to be expressed. After this incubation time cells were assayed for UV-induced unscheduled DNA synthesis (UDS) as described (28). In all experiments plasmid pCD1, containing the human *ERCC3* driven by an SV40 early promoter, was used as a positive control.

RESULTS

Strategy for cloning the *Drosophila* homolog of *ERCC3*

Previously, we reported that Southern blots containing genomic DNAs from organisms, ranging from *Drosophila* to man, incubated under moderately stringent conditions with a ^{32}P-labelled total cDNA probe from the human *ERCC3* gene, showed many hybridizing fragments (17). This suggested the presence of a homologous gene among eukaryotes, and strong nucleotide sequence conservation. To avoid isolation and sequencing of many clones with fortuitous sequence homology we used two flanking cDNA probes. In this 'junction probe' strategy it is assumed that only homology spread over a long area of DNA is of interest. When different restriction enzyme digests of genomic DNA are used, and an extended area of nucleotide homology exists in the organism under investigation, there is a reasonable chance that the 5' as well as the 3' cDNA probe hybridize to only one genomic fragment; the junction fragment. Although small domainal homology can be very illustrative, and even can be the only conserved part of a gene, it more often does not represent the homologous gene looked for. Moreover, it appears that accidentally cloned plasmid contaminations in lambda cDNA or genomic libraries often show up due to vector DNA contamination in the geneprobes used. This problem was circumvented by the use of PCR-generated probes.

Cloning and nucleotide sequence of *D.melanogaster ERCC3*

From the human *ERCC3* cDNA two flanking PCR probes (see Materials and Methods) were prepared and hybridized at 56°C to blots with several digests of genomic DNA of *D.melanogaster*, a related species, *D.hydei*, and of the yeasts *S.cerevisiae* and *Ss.pombe*. As expected when an extended region of homology exists, junction fragments can be discerned in some of the DNA digests of all these species (Fig. 1). Subsequently, an EMBL3 lambda library made from genomic *D.melanogaster* DNA was screened (29). One replica filter was hybridized with the 5' PCR fragment (primer couple 1), the duplicate filter with the 3' PCR fragment (primer couple 2). Four double positive clones could

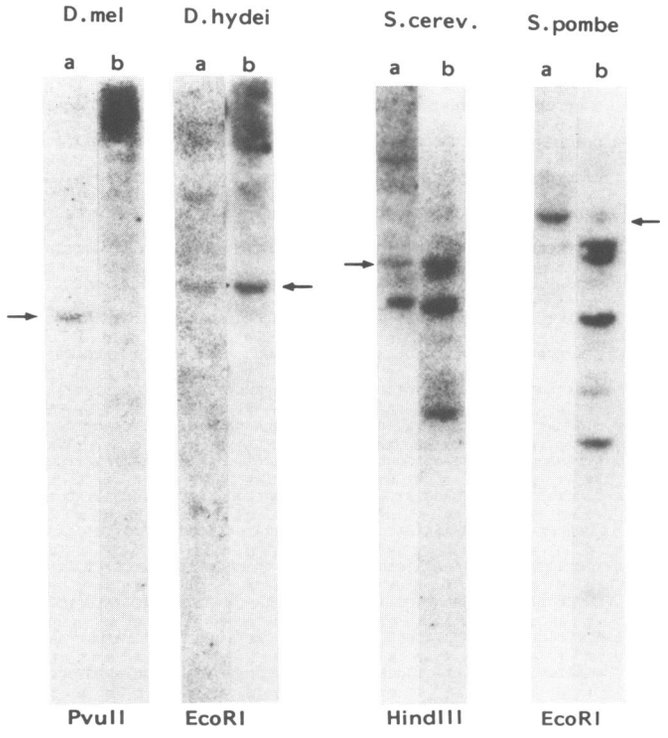
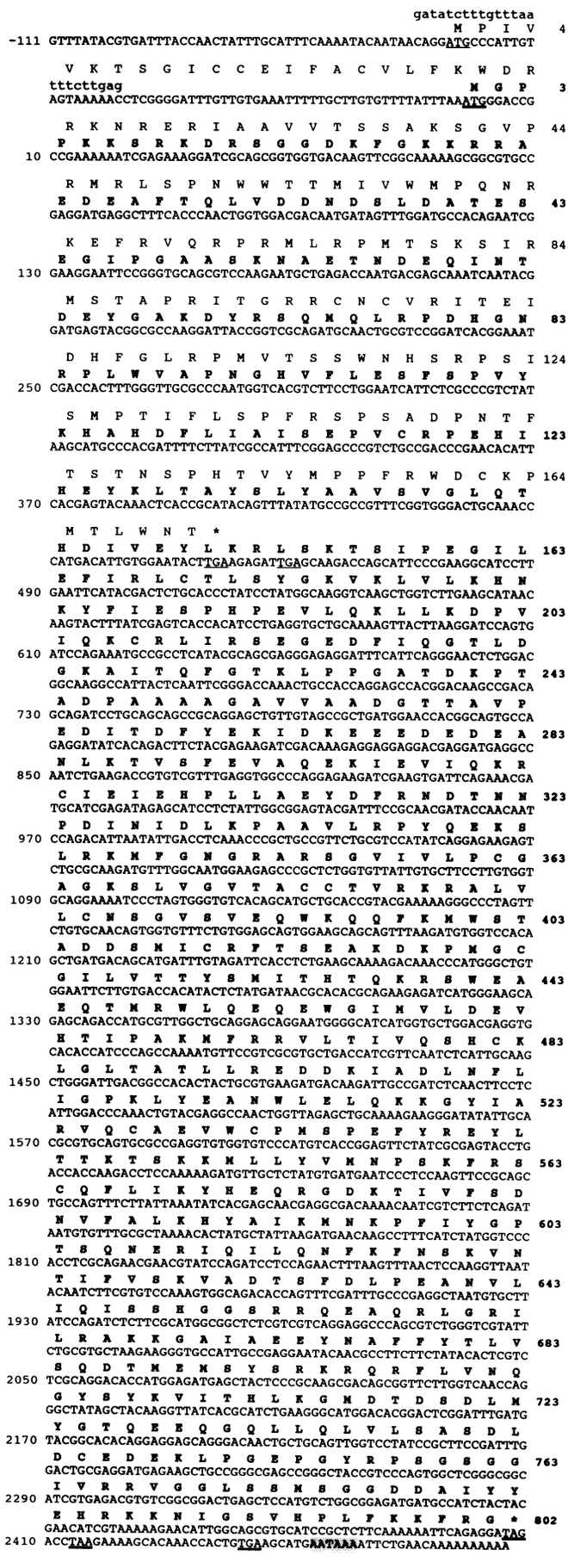


Figure 1. Southern hybridization of 5' and 3' human *ERCC3* probes to genomic DNA of *D.melanogaster* and *D.hydei*, and the yeasts *S.cerevisiae* and *S.pombe*. DNA digested with *Pvu*II, *Hind*III, *Eco*RI, *Bam*HI, and *Pst*I, was hybridized under low stringency conditions (56°C) with the two PCR probes of human *ERCC3* (a: 5'human *ERCC3* probe; b: 3'human *ERCC3* probe). Only the digests with clear junction bands are shown. In all other digests no common bands could be discerned. Arrows indicate the junction bands.



be isolated among about 100 phage plaques hybridizing to only one of the two probes. All double positive clones appeared to be derived from the same genomic region and hybridization studies showed that the complete gene is located on an 8.5 kb SalI fragment common to these clones. A λgt11 cDNA library, made from *Drosophila* head RNA (30) was screened with the genomic 2.4 kb BamHI-SalI subfragment, harbouring the 5′ region of the gene. Six positive clones were identified. The inserts of the two longest clones, SD2 (2579 bp) and SD4 (2534 bp), were subcloned in pUC119 and pUC120 vectors and sequenced. The complete nucleotide sequence is shown in Fig. 2. The two cDNAs were exactly the same with the exception of the utmost 5′ region. From position −42 further 5′ the nucleotide sequence deviates. In SD4, the shortest cDNA, this part of the sequence matches relatively well to the splice acceptor site consensus: $(y)_{\geq 11}$nyAG↓g (31) (it contains the conserved AG and a pyrimidine track). Therefore, it is possible that SD4 is partially processed and still contains part of a 5′ intron. The longest cDNA, SD2, which cannot readily be explained as containing an intron, harbours an ATG at position −62 (Fig. 2), which matches well to the *Drosophila* translation initiation consensus C/AAAC/AATG (32). This codon determines the start of a 170 amino acid open reading frame (ORF) without any homology to mouse or human *ERCC3*, nor could obvious homology be detected with any other sequence in various databases. The second ATG at position +1, although only weakly matching the ATG consensus, specifies the 802 amino acid ORF of *Drosophila ERCC3*. Both cDNAs harbour at the same position a poly(A)tail, preceded by a polyadenylation signal which completely fits the consensus: AATAAA (33, 34).

Expression of *ERCC3^{Dm}* during fly development

To study mRNA expression of *ERCC3^{Dm}* and to see whether the two cDNAs, SD2 (2579 bp) and SD4 (2534 bp) were nearby full length, Northern blots were prepared with poly(A)$^+$ RNA isolated from an established cell line, Dm2 (data not shown), and different developmental stages of the fly (Fig. 3). Hybridization with the *ERCC3^{Dm}* probe revealed a relatively low abundant transcript which migrates at about 2.7 kb (Fig. 3); the same length as the mouse and human transcripts (5, 17). Because the cDNAs are marginally shorter it is possible that still some 5′ sequences are lacking, although one has to take into account the *in vivo* length of the poly(A) tail. To correct for differences in the amount of RNA loaded on the gel, blots were rehybridized with a *D.melanogaster ras* probe. This gene has been shown to be expressed at a constant level throughout development (35). Scanning of the blots suggests small differences in *ERCC3^{Dm}* mRNA levels in the later stages of development (Fig. 3). About 2−3 times more *ERCC3^{Dm}* transcripts were detected in the 2nd instar larvae, pupae and adults compared to 0−16 hours embryos and first instar larvae.

Figure 2. Nucleotide sequence of the *D.melanogaster ERCC3* gene. Start and in frame stopcodons of the *ERCC3^{Dm}* open reading frame (bold) are indicated by double underlining. The (first) ATG and two in frame stopcodons which specify the 170 amino acid open reading frame are underlined. A putative polyadenylation signal is boxed. The 5′ sequence of the SD4 cDNA are indicated in small print, whereas the 5′ SD2 sequence is given in capitals.

Structural conservation between ERCC3^{Dm} and its human, mouse and yeast counterpart

The predicted *Drosophila* protein displays 71% overall amino acid sequence identity (78% strong similarity) with its human counterpart (Fig. 4). Comparison of the human, mouse and yeast (in ref.13, 18) gene products revealed that the majority of amino acid substitutions is concentrated in three regions. This general picture is corroborated when the *Drosophila* sequence is included (Fig. 4). The N- and C-terminus together with the region in front of the second acidic region show most of the amino acid changes. Seven short amino acid motifs identified in a recently defined family of presumed DNA/RNA helicases (16) were previously also found in the ERCC3 coding sequence (Fig. 4) (14). All essential amino acids of these domains are conserved in *D.melanogaster*, which underlines the functional or structural importance of these parts. The regions which link these seven domains are significantly less conserved: 83.5% similarity in between the domains compared with 91% in the motifs. All important residues in the putative helix-turn-helix DNA binding region (for review see (36)) are conserved; it is only remarkable that a rather bulky, weakly basic histidine replaces the small uncharged polar serine residue in the second postulated α-helix. In contrast to the human and mouse protein only the second of the two acidic regions, thought to be important for interactions with chromatin, has been conserved. The position of the previously identified potential nuclear location signal (NLS) in the human/mouse protein is not strictly conserved. The signal has moved more C-terminal and matches exactly the consensus for a bipartite NLS (37).

Localization of *ERCC3^{Dm}* to polytene salivary gland chromosomes by *in situ* hybridization using biotinylated probes

To identify a possible involvement of *ERCC3^{Dm}* in one of the many isolated *D.melanogaster* mutagen-sensitive mutants, *in situ* hybridization to polytene salivary gland chromosomes of two wild type strains, *Berlin K* and *Amherst M56i*, was performed with the genomic 8.5kb *SalI* fragment. In both strains a clear strong signal appears on chromosome III band 67E3/4 (Fig. 5A/B). No known mutagen-sensitive mutant is localized in this region. An additional weaker hybridization signal could be detected on chromosome II band 51A (Fig. 5C/D), but only in *Berlin K* flies. It is possible that this hybridization is due to a pseudogene of *ERCC3^{Dm}* present in these flies or to a gene with some sequence homology to the *ERCC3^{Dm}* probe, although in Dm2 genomic DNA no evidence for such a gene exists from Southern blots hybridized under low stringency salt conditions.

Functional crosscomplementation of *ERCC3^{Dm}* in xeroderma pigmentosum (c.g.B) cells

To examine whether the SD2 cDNA is able to correct the UV survival of the repair deficient Chinese hamster cell line, 27-1, the insert of this clone was amplified via PCR and placed behind the strong SV40 late promoter in a eukaryotic expression plasmid, pSLM (see Materials and Methods). To minimize the risk of PCR-induced mutations, two independant clones were used for the experiments: plasmids pSVD2 and pSVD3. The optimal first ATG, specifying the 170 amino acids ORF, which is present in these constructs, may prevent *ERCC3^{Dm}* to be efficiently translated from the second suboptimal startcodon. Therefore, this ATG was removed from pSVD2 and pSVD3, yielding plasmids: pSVD2Δ and pSVD3Δ (see Materials and Methods). All four plasmids were transfected together with a selection marker (Neo^R) for DNA uptake. UV survival was performed on the neomycin resistant mass culture. No correction could be detected (results not shown). The four cDNA constructs were also microinjected into the nucleus of the xeroderma pigmentosum cell line, XPCS1BA (complementation group B), and Unscheduled DNA Synthesis (UDS), another parameter for DNA repair ability, was monitored. But, also in this test no correction of the UDS negative phenotype of XPCS1BA could be found, whereas the human *ERCC3* gene did completely correct the repair defect. These negative findings may indicate that *ERCC3^{Dm}* is unable to significantly correct the XP-B defect.

DISCUSSION

Here we present the cloning and characterization of the *D.melanogaster* homolog of a human DNA excision repair gene. The strategy for evolutionary walking which we used is widely applicable for the isolation of homologous genes from other organisms. It selects for long stretches of nucleotide sequence homology and in that way avoids the isolation of nucleotide sequences with only marginal local identity to the probes used. This is clearly demonstrated by the fact that only four double positive clones were identified amidst about 100 single positive phage plaques.

From a *D.melanogaster* head cDNA library two different almost full length cDNAs were isolated. It is likely that one of them represents a partially spliced mRNA because of the presence of a putative splice acceptor site in the sequence. However, it is not excluded that the 5′ ends of both cDNAs are naturally occurring parts of mature spliced mRNAs resulting from alternative splicing. This could mean that the gene has two transcriptional start sites and/or two promoters, a situation found in some other *D.melanogaster* genes (38). Hybridization of two synthetic oligonucleotides, specific for each of the 5′cDNA

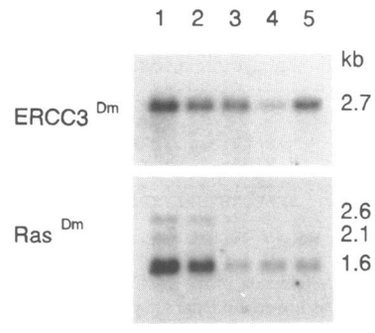


Figure 3. mRNA expression of *ERCC3^{Dm}* during fly development. 5μg poly(A)$^+$ RNA isolated from different developmental stages was fractionated on a 1% agarose-formaldehyde gel, blotted and hybridized with the insert of λ clone SD4 (upper panel). As a control for the amount of RNA loaded, the filter was rehybridized with a *D.melanogaster ras* probe (lower panel), found by others to be present at a constant level throughout development (35). This shows that lower amounts of RNA were loaded in lanes 3,4 and 5. Lane 1: 0−16 hr embryos; lane 2: first instar larvae; lane 3: second instar larvae; lane 4: pupae; lane 5: adults. The amount of *ERCC3* mRNA in the different stages was quantified by densitometrical scanning of the autoradiogram, and expressed relative to the amount of ras mRNA. Arbitrarily the value for the 0−16 hr embryos was fixed at 1. For the other stages the values are: 1 (1st instar larvae), 3 (second instar larvae), 2 (pupae), and 3 (adults).

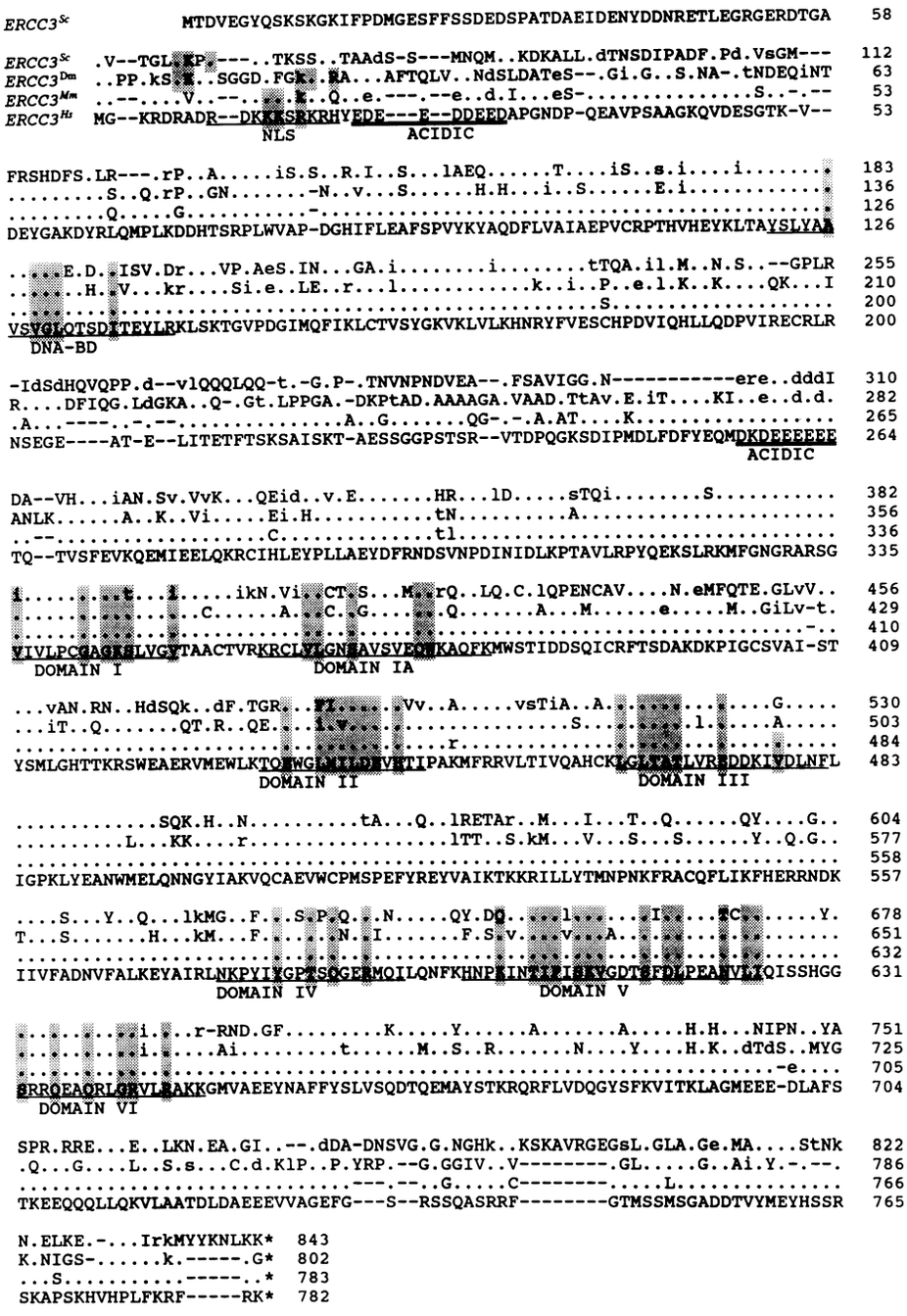


Figure 4. Amino acid comparison of the *ERCC3* homologs of man, mouse, *Drosophila* and yeast. Amino acids which are identical with the human protein are indicated by dots, whereas lowercase letters signify strongly conserved residues. Conserved amino acid residues: R and K; E and D; I, V, and L; T and S. Putative Nuclear Localization Signals, DNA-binding and helicase regions are underlined. Acidic regions are indicated by double underlining. Boldfaced, boxed residues are the consensus residues of the putative domains. Consensus sequences used: 1) NLS; KK/RxK/R 2) Helicase domain I; $+x_5$GxGKS/Tx_3+ 3) domain IA; x_4++xxo/Txx+xxo+x_5 4) domain II; xxoxx+++++DExH/Dxx 5) domain III; +x+T/SA/GT/Sx_3ox_4+xoxx 6) domain IV; x_5F/YxxS/Txoxxox$_3$ 7) domain V; oxxoxx+++xS/To+x_3G/Sxo+$x_3$ox++ 8) domain VI; oxxoxxQ/HxxGRxxRx_3. x = irrelevant amino acid residue. o = charged/polar amino acids (S, T, D, E, N, Q, K, R). + = hydrophobic residue (I, L, V, M, F, Y, W). '/' means 'or', and '-' indicates an introduced gap. In the putative helix-turn-helix DNA binding region the VGL sequence forms the turn region in-between the two α-helices.

regions, to the isolated genomic clones of *D.melanogaster ERCC3*, shows that both probes recognised the gene, and hence are part of it. Therefore, artefacts in the library, by which unrelated sequences are ligated to the *ERCC3* cDNA, can be excluded. It is remarkable that the SD2 cDNA has the potential to encode two unrelated proteins in two different reading frames. The first (optimal) ATG specifies a protein without any homology to other polypeptides. The second, suboptimal ATG defines the start of the ERCC3^{Dm} protein. It may be that translation of the first ORF does not intervene with translation of ERCC3^{Dm} by ribosomes that skipped the first ATG. Alternatively, the first ORF could in some way be involved in regulation of the quantity of

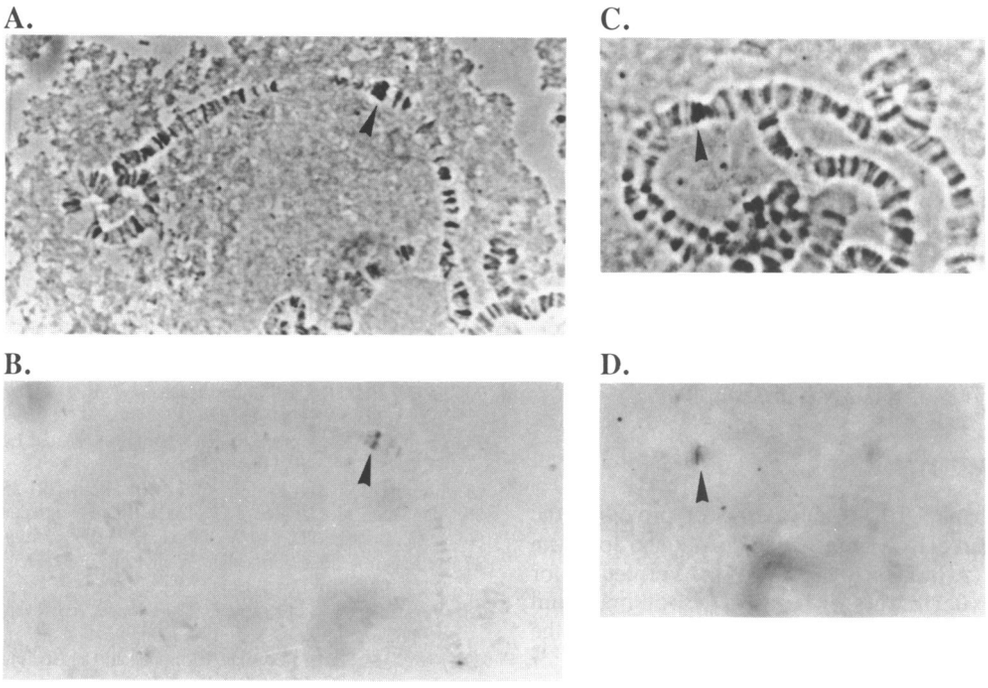


Figure 5. *In situ* hybridization of *ERCC3*^{Dm} to polytene salivary gland chromosomes. Salivary gland chromosomes from the wild-type strain *Berlin K* were used for *in situ* hybridisation with biotinylated probes. Panel A: phase contrast of chromosome III. Panel B:normal illumination of chromosome III. Panel C: phase contrast of chromosome II. Panel D: normal illumination of chromosome II. The arrowheads indicates the chromosomal position 67E3/4 (Panel A/B) and position 51A (panel C/D).

ERCC3^{Dm} protein, by interfering with translation. Also in the human (but not in the mouse) *ERCC3* mRNA a 5′ ORF is found (14, 17). However, this precedes the *ERCC3* ORF completely and does not overlap with the region coding for ERCC3 like in *D.melanogaster*.

Northern blot analysis indicates that *ERCC3*^{Dm} is continuously expressed throughout development. This is consistent with the notion that the protein is supposed to function in NER, a basic process operational in all cells and stages of the cell cycle. Similar constitutive RNA expression was also observed for *ERCC1* and *3* in mammals (17, 39). Furthermore, these findings are in line with an additional vital function of the gene at least in yeast and presumably also in other organisms (in ref.13, 18, 39). The significance and physiological consequences of sligthly increased *ERCC3*^{Dm} RNA levels in second instar, pupae and adult stages of *D.melanogaster* has to be established.

The finding that the N- and C-terminal areas of the protein are less conserved is in complete agreement with earlier data from comparisons between mouse and human ERCC3 (17). This is further underlined by the very recent cloning of the *S.cerevisiae SSL2* gene, reported while this manuscript was in preparation. This gene was isolated as a suppressor mutant that overcomes a block of translation initiation of the *his* 4 mRNA due to an artificial stem-loop structure in the 5′UTR (40). Unexpectedly, the SSL2 geneproduct appeared to be identical to the *S.cerevisiae* homolog of human ERCC3 independently isolated by us applying the junction probe strategy (in ref.13, 18). These findings suggest a dual function of the ERCC3 gene product: one in NER, the other in processing of (a subset of) mRNPs for translation initiation. The latter may define the essential function of the gene, demonstrated for the yeast homolog (in ref.13, 18) and postulated for the human gene. The association of ERCC3 with a bypass of a hairpin structure in mRNA supports a helicase function for this protein. The preservation of the helicase domains, of the nucleic acid binding region, and of the second acidic region during evolution, strengthen the functional significance of these protein parts. The identification of a nuclear location signal consensus in the N-terminus of the mammalian protein suggests that the ERCC3 geneproduct resides in the nucleus. The presence of a sequence exactly matching a bipartite NLS in the *D.melanogaster* protein supports this notion. Also the human and mouse NLS may be bipartite, although lysine residues are located nine instead of ten amino acids N-terminally of the important KKxK motif.

In situ hybridization using biotinylated geneprobes localized the *ERCC3* in region 67E3/4 of chromosome III. A number of lethal mutations are listed in this area as well as two genes involved in heterochromatinization processes. In addition, also the 'Enhancer of zeste' and *haywire* genes (41). Although none of these genes or mutants are known to be affected in DNA repair it could be that in view of the essential role of the gene in yeast and by inference also in *D.melanogaster* and man, only a limited set of mutations in *ERCC3* may be tolerated. Alternatively, mutations affecting the postulated vital function in a subtle fashion may generate unexpected phenotypes (cf.SSL2).

Functional crosscomplementation was attempted by cloning the *ERCC3*^{Dm} cDNA SD2 behind the strong SV40 late promoter. Two independent PCR clones were tested, to minimize the chance for PCR-derived inactivating mutations. Both, DNA transfection and microinjection studies did not provide evidence for significant correction of the repair-deficient phenotype of the 27-1 hamster cell line and XPCS1BA primary fibroblasts. These findings could suggest that the ERCC3^{Dm} protein can not function in a human or hamster context e.g. because it has to interact with members

of a mammalian protein complex, and has diverged too much to be able to do so. This in contrast to the *D.melanogaster* XP-A protein which has a lower degree of amino acid sequence conservation, yet is able to correct the human XP-A phenotype (11). Since the outcome of the interspecies complementation experiments with *ERCC3^{Dm}* is negative, many other explanations can be put forward. The only conclusive answer whether the absence of correction is due to an intrinsic inability of the *Drosophila* protein to function in a mammalian context would be to demonstrate that sufficient amounts of active ERCC3^{Dm} geneproduct are made. Unfortunately, at present a direct assay for measuring ERCC3^{Dm} activity is lacking. Nevertheless, cloning and characterization *ERCC3^{Dm}* enables studies on the functions of this gene in *Drosophila*.

ACKNOWLEDGEMENTS

We would like to thank Dr.P.M.Salvaterra for providing the λgt11 *Drosophila* head cDNA library. Jan de Wit and Joke van den Berg for the DNA transfections, and Wim Vermeulen for microinjection experiments. Drs. D.Bootsma and P.H.M.Lohman are acknowledged for critical reading of the manuscript. Mirko Kuit and Tom de Vries Lentsch for photography, and Sjozef van Baal for computer assistance. This work was supported by the Dutch Cancer Society (project IKR 88-2 and 90-20), and the European Community (contract B16-141-NL).

Note added in proof

Recent research of Drs L.Mounkes and M.T.Fuller identifies ERCC3 as the *Drosophila* haywire gene. Their results will appear in *Cell* shortly.

REFERENCES

1. Van Houten, B. (1990) *Microbiol. Rev.*, **54**, 18–51.
2. Cleaver, J.E. and Kraemer, K.H. (1989) *In* C.R.Scriver, A.L.B., W.S. Sly, D. Valle (ed.), Xeroderma pigmentosum. McGraw Hill Inc., 2949–2971.
3. Westerveld, A., Hoeijmakers, J.H.J., Van Duin, M., De Wit, J., Odijk, H., Pastink, A., Wood, R.D. and Bootsma, D. (1984) *Nature*, **310**, 425–429.
4. Weber, C.A., Salazar, E.P., Stewart, S.A. and Thompson, L.H. (1990) *EMBO. J.*, **9**, 1437–1447.
5. Weeda, G., van Ham, R.C.A., Masurel, R., Westerveld, A., Odijk, H., de Wit, J., Bootsma, D., van der Eb, A.J. and Hoeijmakers, J.H.J. (1990) *Mol. Cell. Biol*, **10**, 2570–2581.
6. Mudgett, J.S. and MacInnes, M.A. (1990) *Genomics*, **8**, 623–633.
7. Tanaka, K., Miura, N., Satokata, I., Miyamoto, I., Yoshida, M.C., Satoh, S., Kondo, A., Yasui, A., Okayama, H. and Okada, Y. (1990) *Nature*, **348**, 73–76.
8. Troelstra, C., Odijk, H., de Wit, J., Westerveld, A., Thompson, L.H., Bootsma, D. and Hoeijmakers, J.H.J. (1990) *Mol. Cell. Biol.*, **10**, 5806–5813.
9. Van Duin, M., De Wit, J., Odijk, H., Westerveld, A., Yasui, A., Koken, M.H.M., Hoeijmakers, J.H.J. and Bootsma, D. (1986) *Cell*, **44**, 913–923.
10. Tanaka, K., Satokata, I., Ogita, Z., Uchida, T. and Okada, Y. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5512–5516.
11. Shimamoto, T., Kohno, K., Tanaka, K. and Okada, Y. (1991) *Biochem. Biophys. Res. Comm.*, **181**, 1231–1237.
12. Bankmann, M., Prakash, L. and Prakash, S. (1992) *Nature*, **355**, 555–558.
13. Lehmann, A., Hoeijmakers, J.H.J., Van Zeeland, A.A., Backendorf, C.M.P., Bridges, B.A., Collins, A., Fuchs, R.P.D., Margison, G.P., Montesano, R., Moustacchi, E., Natarajan, A.T., Radman, M., Sarasin, A., Seeberg, E., Smith, C.A., Stefanini, M., Thompson, L.H., Van der Schans, G.P., Weber, C.A. and Zdzienicka, M.Z. (1992) *Mut. Res.*, **273**, 1–28.
14. Weeda, G., van Ham, R.C.A., Vermeulen, W., Bootsma, D., van der Eb, A.J. and Hoeijmakers, J.H.J. (1990) *Cell*, **62**, 777–791.
15. Nance, M.A. and Berry, S.A. (1992) *Am. J. Med. Gen.*, **42**, 68–84.
16. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1989) *Nucl. Acids Res.*, **17**, 4713–4730.
17. Weeda, G., Ma, L., Van Ham, R.C.A., Van der Eb, A.J. and Hoeijmakers, J.H.J. (1991) *Carcinogenesis*, **12**, 2361–2368.
18. Park, E., Guzder, S., Koken, M.H.M., Jaspers-Dekker, I., Weeda, G., Hoeijmakers, J.H.J., Prakash, S. and Prakash, L. (1992) *Submitted for publication*
19. Broughton, B.C., Barbet, N., Murray, J., Watts, F.Z., Koken, M.H.M., Lehmann, A.R. and Carr, A.M. (1991) *Mol. Gen. Genet.*, **228**, 470–472.
20. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a laboratory manual. Cold Spring Harbor, NY.
21. Pastink, A., Vreeken, C. and Vogel, E.W. (1988) *Mutat.Res.*, **199**, 47–53.
22. Koken, M.H.M., Reynolds, P., Jaspers-Dekker, I., Prakash, L., Prakash, S., Bootsma, D. and Hoeijmakers, J.H.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8865–8869.
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA.*, **74**, 5463–5467.
24. Engels, W.R., Preston, C.R., Thompson, P. and Eggleston, W.B. (1986) *FOCUS (Gibco-BRL)*, **8**, 6–8.
25. Karger, B.D. and Komro, C. (1990) *Focus*, **12**, 25–27.
26. Van Duin, M., Janssen, J.H., De Wit, J., Hoeijmakers, J.H.J., Thompson, L.H., Bootsma, D. and Westerveld, A. (1988) *Mutat.Res.*, **193**, 123–130.
27. de Jonge, A.J.R., Vermeulen, W., Klein, B. and Hoeijmakers, J.H.J. (1983) *EMBO J.*, **2**, 637–641.
28. Vermeulen, W., Osseweyer, P., de Jonge, A.J.R. and Hoeijmakers, J.H.J. (1986) *Mutat. Res.*, **165**, 199–206.
29. Koken, M.H.M., Reynolds, P., Bootsma, D., Hoeijmakers, J.H.J., Prakash, S. and Prakash, L. (1991) *Proc. Natl. Acad. Sci. USA.*, **88**, 3832–3836.
30. Itoh, N., Slemmon, J.R., Hawke, D.H., Williamson, R., Morita, E., Itakura, K., Roberts, E., Shively, J.E., Crawford, G.D. and Salvaterra, P.M. (1986) *Proc. Natl. Acad. Sci. USA.*, **83**, 4081–4085.
31. Mount, S.M. (1982) *Nucl. Acids Res.*, **10**, 459–472.
32. Cavener, D.R. (1987) *Nucl. Acids. Res.*, **15**, 1353–1361.
33. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) *Cell*, **41**, 349–359.
34. Wickens, M. (1990) *Trends Biochem. Sci.*, **15**, 277–281.
35. Mozer, B., Marlor, R., Parkhurst, S. and Corces, V. (1985) *Mol. Cell. Biol.*, **5**, 885–889.
36. Brennan, R.G. and Matthews, B.W. (1989) *J. Biol. Chem.*, **264**, 1903–1906.
37. Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991) *Cell*, **64**, 615–623.
38. Searles, L.L., Ruth, R.S., Pret, A.-M., Fridell, R.A. and Ali, A.J. (1990) *Mol. Cell. Biol.*, **10**, 1423–1431.
39. Van Duin, M., Van den Tol, J., Warmerdam, P., Odijk, H., Meijer, D., Westerveld, A., Bootsma, D. and Hoeijmakers, J.H.J. (1988) *Nucl. Acids Res.*, **16**, 5305–5322.
40. Gulyas, K.D. and Donahue, T.F. (1992) *Cell*, **69**, 1031–1042.
41. Lindsley, D.L. and Zimm, G.G. (1992) The genome of *Drosophila melanogaster*. Academic Press.