CLONING AND CHARACTERIZATION OF EXCISION REPAIR GENES

CLONING AND CHARACTERIZATION OF EXCISION REPAIR GENES

KLONERING EN KARAKTERISERING VAN EXCISIE HERSTEL GENEN

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR
AAN DE ERASMUS UNIVERSITEIT ROTTERDAM
OP GEZAG VAN DE RECTOR MAGNIFICUS
PROF. DR. P.W.C. AKKERMANS M.A.
EN VOLGENS BESLUIT VAN HET COLLEGE VAN DEKANEN.

DE OPENBARE VERDEDIGING ZAL PLAATSVINDEN OP
WOENSDAG 27 MAART 1996 OM 13:45 UUR

DOOR
PETRUS JOHANNES VAN DER SPEK
GEBOREN TE DELFT

PROMOTIECOMMISSIE

Promotoren:	Prof. Dr. D. Bootsma
	Prof. Dr. J.H.J. Hoeijmakers
Overige leden:	Prof. Dr. J.A. Grootegoed
	Prof. Dr. D. Lindhout
	Prof. Dr. Ir. A.A. van Zeeland

The studies described in this thesis were carried out in the Medical Genetics Centre South-West Netherlands at the department of Cell Biology and Genetics Erasmus University Rotterdam. This project was financially supported by the Medical Genetics Centre and the Dutch Cancer Society. The printing of this thesis was financially supported by: Ames B.V., Autron B.V., Bio-Rad Laboratories B.V., Biozym B.V., Eurogentec N.V., Het Kasteel van Rhoon, Pharmacia B.V., Schleicher & Schuell Nederland B.V. and Thieme's Echte Thee.

Front cover	:	Three dimensional representation of the protein structure of ubiquitin. In blue (identical) and in orange (similar) residues shared by the NER enzyme RAD23 The similar spacefilling model indicates the homologous residues of the conserved core. Molecular modeling and image processing was performed at the National Institutes of Health's division of computer research and technology, Bethesda, USA.
Illustrations	:	Mirko Kuit
Printing	:	Drukkerij Haveka B.V., Alblasserdam

The known is finite, the unknown infinite; intellectually we stand on an island in the midst of an illimitable ocean of inexplicability. Our business in every generation is to reclaim a little more land.

Het bekende is eindig het onbekende oneindig; verstandelijk staan we op een eilandje midden in een onmetelijke oceaan van onverklaarbaarheid. Het is onze plicht in elke generatie een beetje meer land droog te leggen.

T.H. Huxley (1887)

CONTENTS

ABBREVIATIONS

Ab	antibody
AP	apurinic or apyrimidinic (site in DNA)
AT	ataxia telangiectasia
BER	base excision repair
BLAST	basic local alignment search tool
bp	base pairs
BS	Bloom syndrome
BSA	bovine serum albumin
CDK	cyclin dependent kinase
cDNA	complementary DNA
CHO	chinese hamster ovary
CPD	cyclobutane pyrimidine dimer
CS	Cockayne syndrome
DAPI	4'6-diamino-2-phenylindole
DDB	damaged DNA binding protein
DDBJ	DNA database Japan
DNA	deoxyribonucleic acid
DNA-PK	DNA dependent protein kinase
dNTP	2'-deoxynucleoside 5'-triphosphate
DSBs	double strand breaks
DTT	dithiothreitol
EDTA	ethylene dinitrilo tetraacetic acid
EMBL	European molecular biology laboratory
ERCC	excision repair cross complementing
EST	expressed sequence tag
FA	Fanconi anemia
$FADH_2$	1,5-dihydroflavin adenine dinucleotide
FCS	fetal calf serum
FISH	fluorescence *in situ* hybridization
FTP	file transfer protocol
GDB	Genome Data Base
GGR	global genome repair
HHR23	human homolog of RAD23
HLA	histocompatibility antigens
HNPCC	hereditary nonpolyposis colorectal cancer
Ig	immunoglobulin
J	Joule
kb	kilobase
kD	kilo Dalton
MEC	mitotic entry control
MTHF	5,10-methenyltetrahydrofolate
MW	molecular weight
NER	nucleotide excision repair
nm	nanometer
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
PHR	photoreactivation DNA photolyase
RAD	radiation sensitive
REV	defective mutation reversion
RPA	replication protein A
RF-C	replication factor C
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RPA	replication protein A
SAD	S phase arrest deficient
scid	severe combined immunodeficiency
SDS	sodium dodecyl sulphate
SNM	sensitive nitrogen mustard
SSBs	single strand breaks
SSL	suppressor stem loop
STS	sequence tagged site
SV40	Simian virus 40
TCR	transcription coupled repair
TFIIH	transcription initiation factor IIH (BTF2)
TRIS	tris(hydroxymethyl)aminomethane
TTD	trichothiodystrophy
UBC	ubiquitin conjugating
UDS	unscheduled DNA synthesis
UV	ultraviolet
UVA	ultraviolet light (315-400 nm)
UVB	ultraviolet light (280-315 nm)
UVC	ultraviolet light (280-320 nm)
uvr	ultraviolet resistance
V(D)J	variable (diversity) joining
www	world wide web
XP	xeroderma pigmentosum
XP-C	XP complementation group C
XRCC	X-ray repair cross-complementing
YAC	yeast artificial chromosome
6-4PP	pyrimidine (6-4) pyrimidone photoproduct

CHAPTER I

General introduction

GENERAL INTRODUCTION

For all living organisms, it is of vital importance to maintain intact the genetic information stored in the nucleotide sequence of DNA. Numerous environmental and genotoxic agents can affect the DNA and lead to, for example, mutagenesis or carcinogenesis. Study of the mechanism of UV-carcinogenesis has become even more pressing given recent concerns about atmospheric ozone depletion (Mimms, 1994; Watson, 1995), because such atmospheric alterations would result in increased UVB at the earth's surface.

Carcinogenesis appears to be a multistep process through which normal cells progress from benign, through transitional stages to the fully malignant forms of cancer by the gradual accumulation of genetic errors (Bishop, 1995). Therefore, normal cells have an intricate quality control mechanism that recognizes and mends damage to the DNA helix.

A number of distinct DNA repair pathways have been identified. Such mechanisms counteract the process of carcinogenesis, and include nucleotide excision repair (NER), recombination repair, post-replication repair, mismatch repair, base excision repair and photoreactivation (reviewed in TIBS, DNA repair; Special issue vol 20, 1995). Impressive advances in our understanding of the biologic significance of these DNA repair pathways have come from studying human conditions in which the relevant genes are involved.

The NER pathway is one of the most important pathways, since this recognizes and removes a wide range of DNA lesions. The fundamental importance of this mechanism is illustrated by three rare, autosomal, recessive, clinical conditions: xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD). Of the three disorders, XP has been studied in the greatest detail. The discovery that this cancer-prone syndrome is caused by defective NER was made in 1968 by Cleaver. NER-deficiency patients show a marked sensitivity to sun exposure. UV-exposed skin of XP patients shows pigmentation abnormalities and a greater than 1000-fold increased risk of cancer, caused by defects in one of at least 7 different genes (*XPA* to *XPG*) (Cleaver and Kraemer, 1994). This clinical heterogeneity is also observed in CS, in which two complementation groups, CS-A and CS-B, are distinguishable. The phenotype of CS patients includes sun sensitivity, but less severe than in XP, and stunted growth, neural dysmyelination and disturbed sexual development. Unlike XP, neither CS nor TTD patients show an elevated risk of (skin) tumor formation (Lehmann, 1987). The rare inborn disorder TTD manifests the CS symptoms in addition to other hallmarks such as brittle hair/nails and ichthyosis (Itin and Pittelkow, 1990).

In contrast to the UV-sensitive xeroderma cells, ataxia telangiectasia (AT) cells are

abnormally sensitive to killing by ionizing radiation. AT is also an autosomal recessive human disorder, whose clinical features include progressive cerebellar ataxia, immunodeficiency, a greatly elevated incidence of cancer, hypersensitivity to ionizing radiation and radiomimetic chemicals, and an abnormal resistance to inhibition of DNA synthesis by ionizing radiation (Gatti *et al.*, 1991). Recently, the gene defective in AT, which encodes a 1-phosphatidylinositol 3-kinase was identified (Savitsky *et al.*, 1995). AT patients are likely to be defective in cellular checkpoint controls which normally function to allow repair in cells with damaged DNA, presumably contributing to the genetic and chromosomal instability characteristic of the disorder. A variant of AT, the Nijmegen Breakage Syndrome, manifests similar clinical and biochemical features as AT, with the major difference being a lack of the neurocutaneous features characteristic of the former syndrome (Weemaes *et al.*, 1994). Although cells from AT patients are ionizing radiation sensitive, they have not been shown to be defective in recombinational repair. A possible explanation is the high rate of intrachromosomal recombination demonstrated in some transformed AT cell lines (Meyn, 1993).

Recombinational repair mechanisms are major pathways for the repair of X-ray induced double strand breaks (DSBs). The physiological relevance of DSBs is exemplified by their role in the process of V(D)J recombination in B- and T- lymphocytes, and is inherent to the process of meiosis (Weaver, 1995a/b). Much evidence points to the marked genotoxicity of DSBs; DSBs are recombinogenic, and strong circumstantial evidence implicates DSBs in the formation of chromosomal aberrations. The latter might induce cancer.

Postreplication repair has been suggested to be defective in the variant form of xeroderma pigmentosum (Lehmann *et al.*, 1977). Another clinical condition that probably is associated with replication is Bloom syndrome (Ellis *et al.*, 1995). The post replication repair mechanism fixes the gaps in the daughter strand by strand exchange rather than by repairing the actual lesion itself. However, no definite clinical conditions corresponding to defects in postreplication repair genes have been identified as of yet.

Mismatch repair has been found to replace nucleotides misincorporated into DNA, thereby increasing the overall fidelity of DNA replication. Hereditary nonpolyposis colorectal cancer (HNPCC) is the most common form of genetic predisposition to colon cancer; defects in any of the several mismatch repair genes can cause HNPCC (Kolodner, 1995, for a review).

Base excision repair (BER) involves repair of alkylated bases such as *N*7-alkylguanine and *N*3-alkyladenine by DNA glycosylases and apurinic/apyrimidinic endonucleases (Seeberg

et al., 1995, for a review). Alkyltransferases can remove alkyl groups from the O^6-position of guanine and the O^4-position of thymine.

Photoreactivation (PHR) is a light-dependent enzymatic process in which DNA photolyases catalyze the repair of *cis,syn*-cyclobutane-dipyrimidines. The PHR enzyme has been shown to bind dimers in both prokaryotes and eukaryotes. Recent data suggest that a structural PHR homolog exists in placental mammals. However, unlike their lower eukaryotic counterparts, the existence of photoreactivating activity in placental mammals is a matter of controversy. The PHR protein has been shown to also play a role in NER in lower species. Thus the dimer binding by the mammalian enzyme may participate in an alternative pathway for dimer recognition. Hence, it is possible that this activity participates in NER (Sancar, 1990).

Isolation of genes involved in different DNA repair pathways has provided major insights into the relationship between DNA damage, mutagenesis, and carcinogenesis. Of the available approaches, four major strategies for isolating new human repair genes can be discerned: (1) gene transfer via transfection into repair-deficient mutants, (2) protein purification and gene cloning facilitated by amino acid sequencing, (3) heterologous hybridization if a homolog of another, evolutionarily related, species is available, or (4) computer-assisted screening of DNA and protein databases using sequence comparison algorithms. The rapidly-expanding list of cloned and characterized genes indicates that some DNA repair proteins are not restricted only to damage control, but also to basal transcription, cell cycle regulation, and other cellular processes. Clues for (dual) functions could come from approaches such as purification and association studies, from sequence analysis, or from transgenic animals.

Overview and scope of this thesis

The aim of the work described in this thesis was to isolate and characterize the human equivalents of the *S. cerevisiae RAD23* NER gene. Presumably because of evolutionary divergence, it was not possible to clone the RAD23 human equivalent via intermediate species such as *Schizosaccharomyces pombe* and *Drosophila melanogaster*. As a direct result of the human genome project, using computer algorithms two human partial cDNAs were identified based on homology to the baker's yeast *S. cerevisiae*. Since NER genes have proved to be highly conserved during eukaryotic evolution (Hoeijmakers, 1993a/b), *S. cerevisiae* was a valuable starting point from which to attempt to clone higher eukaryotic homologs of a gene known to be involved in NER. Because NER is the major mechanism

for the repair of UV-induced DNA damage, a brief review of DNA damage and repair is given in Chapter II.

Computer-assisted analysis was an important tool for the identification of the two human RAD23 homologs, and at a later stage, for the analysis of functional domains and motifs. Evolutionary duplication occurred in the cases of *RAD23* and *RAD6* DNA repair genes. The use of computer algorithms as a tool for cloning and functional sequence analysis is described in detail in Chapter III. Both human homologs of the *RAD23* gene, designated *HHR23A* and *HHR23B*, were identified by computer search - also described in Chapter III. A number of other homologous genes were identified, cloned and partially characterized, further emphasizing the value of these comparisons. Via computer analysis of the EST database, another xeroderma pigmentosum group E (XPE) equivalent gene was also identified. Cloning of this cDNA is also described in Chapter III. Other human genome project clones showing homology to lower (model) species were cloned, mapped, and analysed as described in Chapter III.

Complementation group C of XP represents one of the most common forms of this cancer-prone NER syndrome. Purification to homogeneity of the XPC-correcting activity revealed a tightly associated complex of XPC and the HHR23B protein, as described in Chapter IV. No mammalian mutants could be identified with a defect in either of the *HHR23* genes. The chromosomal localization of both *HHR23* genes described in Chapter V revealed no clues, which could have indicated their dysfunction in clinical defects or syndromes. Intriguingly, the *HHR23B* and *XPC* genes whose products form a tight complex, are located close to each other on chromosome 3p25.1. The fact that the *RAD23* gene was duplicated in evolution leading to two genes in human and in mouse, might explain the lack of mutants due to functional redundancy of the gene products. Cloning and characterization of the mouse RAD23 equivalents and identification of a repeated element in the RAD23 cognates with homology to a C-terminal domain in one of the ubiquitin-conjugating (E2) enzymes, are described in Chapter VI. This represents - in addition to the N-terminal ubiquitin like domain - a second link with the ubiquitin system. Chapter VII shows the necessity of the HHR23B protein for NER *in vitro*. Chapter VIII describes the biochemical behaviour and subcellular localization of both HHR23 gene products and the XPC protein, whilst the last Chapter IX deals with future directions and conclusions.

CHAPTER

II

DNA repair mechanisms

DNA REPAIR MECHANISMS

Introductory comments

The biological basis of UV-radiation-induced cancer has been under investigation for a number of years. UV carcinogenesis is a prominent feature of the class of DNA repair-deficient xeroderma pigmentosum (XP) patients (Cleaver, 1968; McKusick, 1992; Pearson *et al.*, 1994). Study of individuals such as those with the rare disorder XP attains a wider importance, as their genetic deficiencies may reflect the defects involved in the increasingly widespread, UV-related, non-hereditary skin cancer. A detailed understanding of the genetic ramifications of DNA damage from various sources is vital to the continued health of all living organisms; gene mutations in human somatic cells probably underlie a variety of pathological processes such as cancer and possibly ageing. These considerations provide relevance and motivation for the study of mechanisms which prevent DNA damage. The most important DNA lesions are the non-repairable lesions, since these are most likely to be mutagenic and/or lethal. Although repairable lesions can also be mutagenic, for example, when repair time is limiting (Brendel and Ruhland, 1984).

In an attempt to further the understanding of the fundamental mechanistic aspects of DNA repair, especially after cellular UV exposure, a number of genes involved in these pathways have been cloned. The present chapter reviews the current state of knowledge of the DNA lesions processed by the different repair pathways, with an emphasis on UV-induced DNA damage. The NER pathway responsible for removal of the major UV lesions can be subdivided into the transcription coupled repair (TCR) and global genome repair (GGR) subpathways.

The use of yeast as a relevant model system for NER by which to isolate higher eukaryotic homologs is discussed. The cloned yeast DNA repair genes can be classified into three major epistasis groups, covering nucleotide excision repair (RAD3), postreplication repair (RAD6), and the recombination repair (RAD52) pathways. This chapter also addresses the clinical conditions (XP, CS, TTD) resulting from a defect in the human homologs of the RAD3 epistasis group NER genes. Next, the actual and postulated roles of genes involved in DNA metabolism in the process of multistep carcinogenesis are considered. The chapter concludes with a section on GGR, which involves the XPC protein, a protein shown to interact with one of the human gene products (designated HHR23B) isolated as part of the work described in this thesis. The HHR23B protein, together with the HHR23A product are two human homologs of the

yeast DNA repair gene *RAD23*. These gene products have an unexpected relationship with ubiquitin, which is discussed in detail in the last part of this chapter. The information covered in the present chapter provides a background to the experimental work subsequently presented in Chapters III to IX.

DNA lesions

Lesions: causes and consequences

Several protective mechanisms are known for maintenance of genetic stability. DNA damage on parental template strands can lead to miscoding mutations or deletions in case lesions are not repaired before DNA replication occurs. When DNA damage arises without environmental pressure, it is regarded as spontaneous damage. DNA is the main biological target for (UV-and ionizing-) radiation-induced damage. The effects of UV-irradiation are largely confined to external epithelial surfaces, although it may give rise to distant effects (e.g., immunomodulation). In the skin, low dose of UVB can induce selective alterations that include photoimmunosuppression of the normal immune response (Vermeer and Hurks, 1994). In contrast, ionizing radiation exposure can directly contribute to internal malignancies, given its penetrating nature. Most mutagens induce multiple types of lesions, each with different mutagenic properties resulting in a variety of mutations. For instance, some lesions might directly miscode or distort the DNA helix, whereas strand breaks in the DNA probably give rise to deletions or to chromosomal rearrangements.

Mutagenesis studies have been highly useful in identifying those genes in which the mutation causes a phenotypic change, such that mutants can be isolated through selective killing of cells or organisms carrying the wild-type gene, or where phenotypic change is recognizable. The distribution of CPD lesions over the genome appears to be random. However, (promoter) areas of transcribed genes that have a more "open" structure, possibly are more prone to DNA damage by certain genotoxic agents (Tardiff *et al.*, 1994). Moreover, the surrounding sequence context can also affect the distribution of particular lesions.

Mutagenic and carcinogenic properties of many carcinogens depend upon their conversion to electrophilic derivatives that are capable of reacting with DNA. The covalent binding of carcinogens to DNA is causally related to tumorigenesis (Friedberg *et al.*, 1995; Tardiff *et al.*, 1994). The process of malignant transformation is currently envisaged to incorporate the sequential acquisition of genetic lesions at loci intimately

involved with the control of cellular proliferation and differentiation. This highlights the importance of removal of adducts from DNA. The biological impact of DNA damage is also exemplified by the existence of human genetic diseases where deficiencies in DNA repair are associated with a high risk of cancer, e.g., xeroderma pigmentosum. Figure 1 and Table 1 summarize the most well characterized DNA lesions, together with examples of DNA-damaging agents causing these lesions.

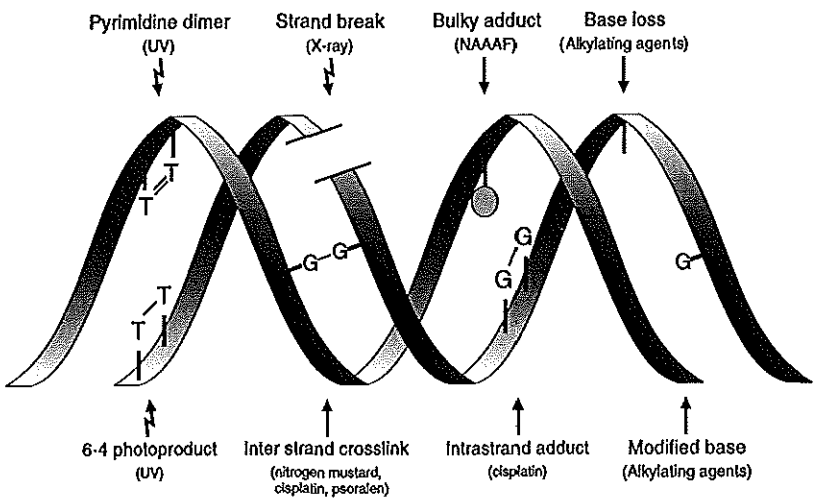


Figure 1 Schematic representation of different lesions on a DNA helix. Modified and adapted from Bohr, 1991 and Friedberg *et al.* (1995).

Major type of lesions	DNA damaging agents
Single and double strand breaks	Ionizing radiation, bleomycin
Pyrimidine dimers and 6-4 photoproducts	UV light (254 nm)
Alkylated bases, alkyl-phosphotriesters	Alkylating agents
Inter- and intra-strand DNA crosslinks	Psoralen + 340 nm UV, mitomycin C, *cis*-platinum
Bulky adducts	Aromatic amines

Table 1 Summary of the most common DNA lesions with the DNA damaging agents causing them. Modified and adapted from Friedberg *et al.* (1995).

X-ray induced lesions

Ionizing radiation is one of the most common carcinogenic agents to which humans are exposed because of background radiation and artificial sources such as medical diagnostic and therapeutic exposures (Goodhead, 1994; Peto and Darby, 1994). During the organogenesis period (weeks 8 to 15) of pregnancy, ionizing radiation is particularly harmful to the developing fetus. Ionizing radiation induces a broad spectrum of lesions in cellular DNA. The classes of lesions occurring are DNA strand breaks (both double- and single-stranded), base-free (AP) sites, and modified DNA bases which vary widely in their relative frequency and biological effects.

DNA double strand breaks (DSBs) are highly biologically damaging lesions. Chromosomal aberrations can, with a high degree of likelihood, be shown to arise directly from DNA DSBs in the cell (Obe *et al.*, 1992). Single strand breaks (SSBs) can be introduced by direct ionization of the DNA and by radicals formed in the surrounding aqueous medium. Furthermore, SSBs can occur as intermediates in recombination repair or nucleotide and base excision repair reactions (see Figure 1 and Table 1).

UV-induced lesions

Sun exposure is required for vitamin D production, but can also contribute to photoageing and non-melanoma skin cancer. Sunlight can be divided into visible light, ranging from 400 nm (violet) to 700 nm (red), infrared radiation, >700 nm, and ultraviolet radiation (UVR), < 400 nm. UVR can be subdivided into three different categories, all having different biological effects. The UVC (200-280 nm) rays do not reach the earth, since these rays are absorbed by the ozone layer. UVB (280-315 nm) represents less than 0.5% of the terrestrial sunlight. However, this part of the spectrum is primarily responsible for almost all biological effects following exposure to sunlight: sunburn, suntan, and, after many years, premature ageing of the skin and skin cancer. UVA (315-400 nm) rays are the least harmful; nevertheless, they can produce sunburn, tanning, skin ageing and skin cancer (Hawk, 1991).

The depth of UVR penetration into the skin is directly related to the wavelength. The more energetic photons of shorter wavelength are absorbed more superficially. Erythema is the most clinically apparent event of the sunburn reaction that introduces damage to the cell membranes and transient disturbances in DNA, RNA and protein synthesis. The primary cause of the erythema is the UVB portion of the UV spectrum. Photosensitivity is a general term to indicate an abnormal sensitivity of the human skin to sun exposure. The effect of non-ionizing radiation is not only limited to the skin. It can

also affect the eyes by inducing cataract, and it may have an immunomodulatory effect.

The major UV-induced lesion is the cyclobutane pyrimidine dimer, formed between two adjacent pyrimidines (Figure 1). Dimers have been shown to play a significant role in mutagenesis and cell death. Another UV-induced lesion that plays an important role in the mutagenic effects of UV light is the pyrimidine-pyrimidone (6-4) photoproduct, also formed at dipyrimidine sites. The ratio of dimers to (6-4) photoproducts is about 3:1 (Mitchell and Nairn, 1989).

Alkylating agents

Alkylating agents react with N- and O-atoms in DNA. These compounds include a wide variety of chemicals which are proved or suspected carcinogens. The most prominent types of adducts are phosphotriesters and alkylated bases. Potential reaction sites for alkylation have been identified in all four bases; however, not all of them have equal reactivity.

Crosslinking agents are those alkylating agents which are capable of reacting with two different nucleophilic centers of the DNA. Inter- and intra-strand DNA crosslinks represent an important class of chemical damage to DNA, since they prevent DNA strand separation and can form complete blocks to DNA replication and transcription. For this reason, a number of agents such as nitrous acid, mitomycin C, nitrogen mustard and sulfur mustard, various platinum derivatives (such as *cis*-platinum (II) diaminodichloride), and certain photoactivated psoralens, have been used extensively in cancer chemotherapy. In addition, UV radiation (254 nm) and ionizing radiation can contribute to the formation of inter-molecular DNA crosslinks. However, crosslinks represent a small fraction of DNA damage caused by radiation.

The anti-tumor drug cisplatin causes inter-strand crosslinks between guanines or adenines in opposing strands (Fig.1). Reaction of cisplatin with nucleophilic sites of the DNA not only results in the formation of intrastrand crosslinks, but also monoadducts and interstrand crosslinks.

In the presumed repair deficiency syndrome Fanconi anemia (FA), patients show a pronounced sensitivity to crosslinking agents (Auerbach, 1995). The FA group C gene has been cloned by complementation of the intrinsic sensitivity of these cells to DNA crosslinking agents (Wevrick *et al.*, 1993), and, recently, the genes for FA group A (Pronk *et al.*, 1995) and D (Whitney *et al.*, 1995) were mapped, by linkage analysis and microcell mediated gene transfer respectively.

Bulky adducts and base analogs as a source of DNA damage

Aromatic amines such as N-acetoxy-2-acetylaminofluorene, N-2-acetyl-aminofluorene, and N-hydroxy-aminofluorene are associated with an increased incidence of cancer in humans. AAAF predominantly reacts at the C^8-position of deoxyguanosine residues, causing (predominantly) dG-AF and dG-AAF adducts. DNA damage induced by 4-nitroquinoline (4-NQO) and AAF is often referred to as "UV radiation-like" because these chemicals produce bulky base damage of the type that, like cyclobutane pyrimidine dimers and (6-4) photoproducts, is repaired principally by the NER pathway.

Analogs of the four naturally occurring bases in DNA can be incorporated from the appropriate triphosphate substrates during DNA replication. These base analogs can cause mutations when present in template DNA undergoing DNA replication.

Lesions: removal and evolution

The evolutionarily long-established repair mechanisms are capable of detecting and removing a wide variety of DNA lesions. Valid interspecies extrapolations can only be performed when the mechanism of action of carcinogens and the protective DNA repair pathways are roughly similar. That DNA repair pathways are conserved throughout evolution, has been established in a number of ways, including sequence conservation and genetic confirmation by mutants from different species that resemble each other in phenotype with respect to sensitivities to various DNA damaging agents, induced mutagenesis, and lesion removal (reviewed by Friedberg *et al.*, 1995). The next section deals in detail with the genes involved in these different repair pathways and the mechanisms.

DNA REPAIR: GENES AND MECHANISMS

Pathways and evolution

Evidence that genetic factors can influence cellular radiation sensitivity comes from UV- and ionizing radiation-sensitive yeast and mammalian mutants. Identification of genes contributing to radiation resistance and cancer prevention is a significant goal of this thesis. Many genes that are involved in prevention of cancer play a role in DNA repair processes (Cleaver and Kraemer, 1994). The yeast *S. cerevisiae* is a convenient model organism for the study of DNA repair in eukaryotes. A large number of DNA damage-sensitive mutants in *S. cerevisiae* have been identified (Haynes and Kunz, 1981). These mutants show many repair defects similar to the mammalian repair-defective cells (Friedberg, 1991; 1994). Yeast mutants abnormally sensitive to killing by radiation are designated as *rad* with identifying locus and allele numbers. The locus members 1 to 49 refer to genes which primarily affect sensitivity to UV radiation or to both UV- and ionizing radiation. Locus numbers greater than 50 designate genes which primarily affect sensitivity to ionizing radiation. As one of the most simple eukaryotes, *S. cerevisiae* has a number of important, basic characteristics that are at least partially conserved throughout eukaryotic evolution. These include chromatin structure, cell cycle regulation, transcription, and DNA replication. This conservation has been proved to extend to the principal features of DNA repair, exemplified by the high level of sequence conservation between yeast and human genes, and the involvement of these genes in other conserved processes such as transcription. On the other hand, major differences exist between yeast and mammals, for example, in some aspects of recombination, mitosis, certain types of modification of DNA (e.g. methylation), and of chromatin (e.g. poly ADP-ribosylation), as well as in the importance and mechanisms of RNA splicing and cell/nuclear division.

The defective proteins in most of these UV and X-ray sensitive mutants are now being elucidated, and seem to be evolutionarily preserved based on sequence similarity. Moreover, cloned human genes are able to correct the defects in rodent mutants, and sometimes even yeast mutants, confirming the functional significance of structurally conserved genes.

Functional relationships of the *rad* genes emerged from analysis of single and double mutants to killing by DNA-damaging agents. In a case where the double mutant is no more sensitive than the most sensitive, single mutant parent, the interaction is termed epistatic. Epistasis has been taken to imply that the gene products mediate steps in the

same repair pathway. The three postulated epistasis groups are named after one of their most prominent member genes: *RAD3*, *RAD6*, and *RAD52*, involving nucleotide excision repair, postreplication repair and recombination repair, respectively. These pathways are subsequently discussed in this chapter.

Nucleotide excision repair

The RAD3 epistasis group includes genes that are involved in nucleotide excision repair (NER). Mutants of this group are sensitive to UV but not to X-rays and show enhanced UV-induced mutagenesis. These loci are known to control excision of pyrimidine dimers and (6-4) photoproducts, which are recognized by the NER pathway.

Besides the *RAD* mutants of *S. cerevisiae* and the naturally occurring XP/CS/TTD patients, laboratory-induced mutant rodent cell lines have been generated. By introducing human DNA (genes) into these mutant rodent cells, the NER defect could be complemented and the responsible gene cloned. For more detailed discussion of genetic complementation, and specifically with respect to NER, see Friedberg *et al.* (1995). Via this procedure, at least five Excision Repair Cross Complementing (ERCC) genes have been identified. The number of identified NER-related genes is still increasing. Many of these NER genes have been found to be related to rare genetic diseases such as XP and CS, characterized by skin hypersensitivity to sunlight. Table 2 summarizes most of the yeast, rodent and human NER counterparts involved in NER syndromes identified to date.

Yeast gene	UV-sensitivity	Human gene	Rodent gene	Features/function
RAD1[a]	very high	XPF	ERCC4[b]	incision 5' side
RAD2[c]	very high	XPG[d]	ERCC5	incision 3' side
RAD3[e]	very high	XPD[f]	ERCC2	5'->3' helicase
RAD4[g]	very high	XPC[h]	unknown	strong ssDNA binding
RAD7[i]	moderate	unknown	unknown	repair of inactive chromatin?
RAD10[j]	very high	ERCC1[k]	ERCC1	incision 5' side, XPA-interaction
RAD14[l]	very high	XPA[m]	unknown	Binds UV-damaged DNA
RAD16[n]	moderate	unknown	unknown	Putative helicase
RAD23[o]	moderate	HHR23A[p], HHR23B[p]	unknown unknown	ubiquitin fusion proteins
RAD25[q]	very high	XPB[r]	ERCC3	3'->5' helicase
RAD26[s]	marginal	CSB[t]	ERCC6	Putative helicase
RAD28[u]	not UV[s]	CSA[v]	ERCC8	WD-repeat protein

Table 2.

Table 2. Yeast DNA excision repair genes with their human and rodent cognates.
(a) Reynolds *et al.*, 1987, (b) Thompson *et al.*, 1994, (c) Madura and Prakash, 1986, (d) O'Donavan *et al.*, 1994; MacInnes *et al.*, 1993 (e) Naumovski *et al.*, 1983, 1985, (f) Weber *et al.*, 1990, (g) Gietz and Prakash, 1988, (h) Legerski and Peterson, 1992 and Masutani *et al.*, 1994, (i) Perozzi and Prakash, 1986, (j) Reynolds *et al.*, 1985, (k) van Duin *et al.*, 1986, (l) (m) Bankmann *et al.*, 1992, (n) Bang *et al.*, 1992, (o) Perozzi and Prakash, 1986, (p) Masutani *et al.*, 1994, (q) Gulyas and Donahue, 1992; Park *et al.*, 1992, (r) Weeda *et al.*, 1990, (s) van Gool *et al.*, 1994, (t) Troelstra *et al.*, 1992, (u) not officially assigned yet, (v) Henning *et al.*, 1995.

The names of the *XP* genes have superseded those of the corresponding *ERCC* genes (Lehmann *et al.*, 1994). Two subpathways can be discerned in NER, the fast transcription-coupled repair mechanism (TCR) and the slower global genome repair mechanism (GGR). The NER mechanism and the functions of some of the gene products involved in the different steps will be discussed below. A major finding in recent years has been that many of the genes shown to be involved in UV-sensitive, laboratory-induced mutant rodent cell lines have also been the genes implicated in a number of human recessive diseases (eg. XP,CS). This is addressed in more detail in the section on the relation to cancer and reflected in the columns of Table 2.

NER: mechanism and machinery

The molecular mechanism of NER has been studied in considerable detail in *E.coli* (Grossman and Yeung, 1990; Van Houten, 1990). The same basic steps have been demonstrated to occur in yeast and human, including incision of damaged DNA, repair synthesis, ligation, and preferential repair of pyrimidine dimers in actively transcribed genes, as depicted in Figure 2. A complex network of many different NER enzymes is involved in the excision and replacement of damaged sites on the DNA helix. Structural and functional similarity has been observed for eukaryotic gene products (yeast to human) involved in NER, which is less pronounced for prokaryotes.

In the initial stage of NER, damage recognition probably involves proteins such as XPA and XPE. The yeast *RAD14* gene encodes a highly hydrophilic protein similar to the human XPA protein, that is absolutely required for the incision process (Bankmann *et al.*, 1992). One of the three subunits of RPA (RPA2) was found to interact with XPA (Matsuda *et al.*, 1995). This single strand DNA binding heterotrimer has been found to cover a region of approximately 30 nucleotides, similar to the estimated patch removed after incisions made by specific DNA endonucleases. XPA is a 31 kD protein, containing a nuclear localization signal and a zinc-finger motif, that was found to interact directly

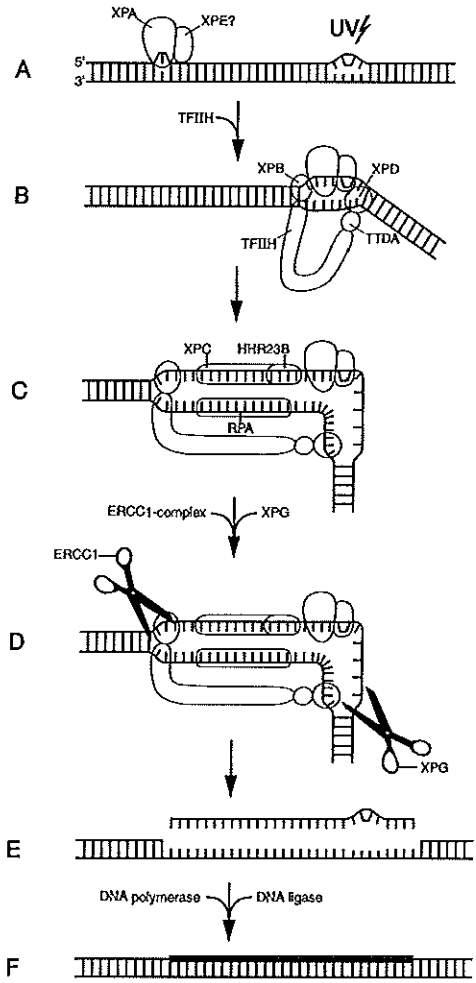


Figure 2. Schematic representation of a model representing the basic steps and the enzymes involved in the NER pathway: (A) Damage recognition, (B) Lesion marcation, (C) Local unwinding, (D) Dual incision, (E) Release of damaged patch, (F) Gapfilling by DNA synthesis and ligation.

with DNA; it has a strong affinity for UV- or cisplatin-damaged DNA (Jones and Wood, 1993). XP-A patients occur, like XP-C patients, at a relatively high frequency compared to the other complementation groups. In addition to the cutaneous lesions, pathological

changes of the central and peripheral nervous system have been described (Cleaver and Kraemer, 1994). There appear to be two types of xeroderma pigmentosum, group E: a DDB-positive form and a DDB-negative form. The damage-specific DNA-binding protein (DDB) or XPE has been purified to near homogeneity from human placenta and HeLa cells (Hwang and Chu, 1993; Keeney *et al.*, 1993) and from primate cells (Abramic *et al.*, 1991; Takao *et al.*, 1993). The DDB activity copurified with a heterodimeric complex consisting of 124 and 41 kD proteins. Microneedle injection of DDB into XP-E cells stimulated DNA repair to normal levels in those XP-E strains that lacked DDB activity (Keeney *et al.*, 1994). However, no stimulation was observed in other xeroderma complementation groups, and in XPE cells that did contain DDB activity. To date, no mutations have been described to occur in any of the subunits. Recently, a second human DDB(XP-E)-like large subunit was identified, as described in Chapter III. Identification of a putative XPE large subunit equivalent is also covered in Chapter III (Van der Spek, unpublished results). The yeast gene was recently found to be essential for viability, suggesting an additional function besides a possible repair function (Lombaerts, unpublished results). This is in contrast to the mild repair defect of XP-E cells. However, the latter can be explained by our finding of a second human *XPE-like* gene, that might be functionally redundant to the original human *XPE* gene.

The DNA helix at the site of the lesion is thought to be unwound by two xeroderma-associated proteins, XPB and XPD. Both XPB (in yeast RAD25) and XPD (equivalent to yeast RAD3) exhibit DNA helicase activities, in the 3'→5', and 5'→3' directions, respectively (Schaeffer *et al.*, 1994). Moreover, XPB and XPD are components of the TFIIH complex, required for basal transcription initiation of RNA polymerase II transcribed genes (Schaeffer *et al.*, 1993). The link with basal transcription points to an essential role for these proteins and for the entire complex. *S. cerevisiae RAD25* was shown to be an essential gene. RAD25 (Park *et al.*, 1992), also known as SSL2 in yeast, shares 55% identical and 72% similar amino acids with the human XPB protein. Patients belonging to the XPB complementation group have conditional mutations that might slightly affect the transcription function of the protein, since the gene is essential for viability. Notably, however, the NER function of the XPB protein can be severely reduced in these patients. The *XPB/ERCC3* gene complements the repair defect of rodent group 3, whereas the *XPD/ERCC2* gene corrects the UV sensitivity of CHO complementation group 2 (Table 1). TTD patients with mutations in the XPD gene, affecting DNA repair and probably to some extent the transcription activity as well have been described by Broughton *et al.* (1994). The yeast RAD3 protein is expected to be the

functional homolog of XPD (Weber *et al.*, 1990). RAD3 appeared essential for RNA polII transcription and yeast cell viability (reviewed by Friedberg *et al.*, 1995).

The exact incision event around the lesion requires the ERCC1/XPF complex, XPG, RPA (hSSB) and additionally, the multiprotein complex TFIIH (for a recent summary see Jaspers and Hoeijmakers, 1995). The 43 kDa ERCC1 protein, homologous to the yeast RAD10 protein (van Duin *et al.*, 1986), has not to date been correlated with a human disorder (van Duin *et al.*, 1988). Specific association of ERCC1 and XPA has been reported by Li *et al.*, 1994a). From *in vitro* repair studies, it became apparent that the ERCC1 protein is associated with the proteins correcting ERCC4, XPF and ERCC11 mutants (van Vuuren *et al.*, 1993; Biggerstaff *et al.*, 1993). The gene defective in UV complementation group 11 (ERCC11) is however, not yet cloned. From yeast, it is known that the RAD1 and RAD10 proteins form a stable complex involved in making the incision 5' of the lesion (Bailly *et al.*, 1992; Bardwell *et al.*, 1994). Very likely, the corresponding human cognates ERCC4 and ERCC1 will exhibit a similar enzymatic activity.

The human ERCC5 protein, homologous to the yeast RAD2 protein appeared to be defective in human XPG patients (O'Donavan and Wood, 1993). The XPG gene product was shown to encode an 133 kDa DNA endonuclease, homologous to FEN-1, that cleaves the damaged DNA strand 3' to the lesion (O'Donovan *et al.*, 1994; Harrington and Lieber, 1994). Additionally, XPB, XPD, and XPG patients with combined xeroderma and Cockayne's syndrome exist. Other cases have been described in which a repair defect in the *XPB* and *XPD* genes causes a TTD phenotype (Stefanini *et al.*, 1993a, 1993b). Highly purified TFIIH fractions specifically correct the NER defect of the XPB, XPD, TTDA NER complementation groups in an *in vitro* assay as well as with microneedle injection. Evidence for dual function of at least part of this complex came from activity studies that proved co-elution of transcription and repair activity in fractionation studies (van Vuuren *et al.*, 1994; Vermeulen *et al.*, 1994). Some features such as dwarfism and brittle hair can be explained by a defective transcription mechanism and not by a defective repair apparatus (Vermeulen *et al.*, 1994).

The completion of the NER reaction involves patch-displacement, gap-filling and religation. Enzymes required for these last steps are DNA polymerases ϵ and/or possibly δ, and ligase I, in combination with the accessory replication factors PCNA, RPA and RF-C (Coverley *et al.*, 1991; Li *et al.*, 1994b).

In vitro reconstitution of the NER process can be expected to further dissect the individual steps and the protein-protein interactions involved in this pathway. Classical

biochemical separation and purification of individual DNA repair enzymes has formed the basis for the reconstitution of the NER pathway. An *in vitro* system to mimic the NER reaction using damaged naked DNA as a substrate was designed by Wood *et al.* (1988) and independently by Sibghat-Ullah *et al.* (1989). In this system, cell extracts were tested for repair capacity of damaged plasmid DNA. Another cell-free DNA repair system with UV-irradiated SV40 minichromosomes as a substrate was described by Sugasawa *et al.*, 1993. Reconstitution of the NER process has been reported recently by Aboussekra *et al.* (1995). From the above studies, it appeared that most of the factors affected in the known mammalian NER mutants are essential for the NER reaction. The *in vitro* repair assay has provided mechanistic insight into the NER reaction with respect to the separation of the incision and excision/gap filling steps and the identification of additional factors such as PCNA (Shivji *et al.*, 1992) and RPA (Coverly *et al.*, 1992). Additionally, the identification of a protein complex containing ERCC1 and XPF (Biggerstaff *et al.*, 1993; Van Vuuren *et al.*, 1993) is a consequence of the *in vitro* repair analysis.

In yeast and mammalian cells, NER-mediated repair of (6-4) photoproducts occurs more rapidly than repair of CPDs (Friedberg *et al.*, 1995). Repair of the transcribed strand of active genes is a much faster process for some lesions, but not for 6-4PPs, than that of the remainder of the genome. TCR involves repair of DNA lesions on the transcribed strands of expressed genes. This is in contrast to lesions that occur on the non-transcribed strands of active genes, or in unexpressed genomic domains. The latter two are processed by the global genome repair pathway (Bohr, 1991). A CPD on the transcribed strand arrests RNA polymerase II, whereas a CPD on the nontranscribed strand has no effect on polymerase elongation (Hanawalt, 1994).

Defects in transcription coupled repair are present in the human disease, CS. Both *CS-A* (*ERCC8*) and *CS-B* (*ERCC6*) genes are cloned, and their products seem to act in a complex. Indications for an interaction between CSA and CSB proteins were obtained by the use of the yeast two-hybrid system (Henning *et al.*, 1995). The *CSA* gene maps to human chromosome 5 and the *CSB* gene was previously shown to be localized on 10q11.2 (Troelstra *et al.*, 1992). The CSA gene encodes a WD-repeat protein which may interact with the 44kDa subunit of TFIIH (Henning *et al.*, 1995), which is provocative with regard to its role in RNA polymerase II-mediated transcription. WD-repeat proteins are associated with different cellular metabolic events such as RNA processing, gene regulation, cell cycle regulation, regulation of cytoskeletal assembly and signal transduction (reviewed by Neer *et al.*, 1994).

Some xeroderma pigmentosum group G patients manifest typical clinical features of CS. Based on the relation of CSA and CSB with TCR, this suggests also some relation

for XPG with RNA polII transcription. The initiation of transcription from eukaryotic protein coding genes is a complex process requiring RNA polymerase II and an array of transcription factors (reviewed by Koleske and Young, 1995). Mutations in XPG might perturb conversion of the form of TFIIH required for transcription to that required for repair, and *vice versa* (Henning *et al.*, 1995).

Both NER and RNA polII mediated basal transcription involve multiprotein complexes consisting of many different subunits in yeast and human (Svejstrup *et al.*, 1995; Drapkin *et al.*, 1994). Because of the dual function of XPB and -D in NER and transcription, other proteins that are part of this transcription initiation complex might possess a NER function, or might be involved in transcription syndromes as postulated by Vermeulen *et al.*, 1994. Table 3 summarizes the human and yeast TFIIH subunits known to date, together with the properties of these gene products. Studies on the involvement of these factors in NER deficient mutants and transcription syndromes are ongoing.

Human gene	Yeast gene	Features/function
ERCC3/XPB	RAD25/SSL2	3'->5' helicase
ERCC2/XPD	RAD3	5'->3' helicase
p62	TFB1	unknown
p52	TFB2	unknown
p44	SSL1	2 Zn^{++}-fingers
MO15/CDK7	Kin28	CDK-like kinase
p34	Scp34[a]	SSL1-like Zn^{++}-finger
cyclinH	ccl1	homology to cyclins
MAT1	TFB3	ring Zn^{++}-finger

Table 3. Human and yeast TFIIH components with their features. Functional and structural equivalence of both human and yeast TFIIH subunits summarized above are reviewed by Roy *et al.* (1994b) and Feaver *et al.* (1993), respectively. (a) *S. cerevisiae* homologue present in Genebank database.

Postreplication repair

Members of the RAD6 group are sensitive to UV and X-rays and control postreplication repair (Lawrence, 1994). Moreover, defects in damage-induced mutagenesis have been reported for several mutants of this epistasis group (Lawrence, 1994). Some of the UV-sensitive yeast mutants are not epistatic with the mutants of the RAD3 epistasis group. A number of these mutants constitute genes required for DNA damage-dependent mutagenesis such as *RAD6*, *RADH (SRS2)*, *RAD18*, *REV1*, and *REV3*. Postreplication repair contributes to the overall fidelity of the DNA replication process. A number of distinct gene products have been associated with postreplication repair, including RAD6 (error-prone repair & error-free repair), RAD18 and RAD5 (exclusively error-free repair), and the RAD9, RAD15 and REV gene products (reviewed in Friedberg *et al.*, 1995).

Studies on *E. coli* have provided a paradigm for postreplication repair (also known as daughter strand gap repair): little is known of this repair mechanism in eukaryotes. The gaps in the daughter strand, rather than the lesion itself, are repaired by this process. This mechanism indirectly leads to a form of tolerance to DNA damage, by leaving a gap opposite the lesion in the template strand, and DNA synthesis restarts downstream of this lesion. The *E.coli* RecA protein is involved in homologous pairing and strand exchange with the undamaged sister molecule. Strand transfer past the lesion is followed by endonuclease cleavage at the Holliday junction, and finally, polymerases fill the gap of the parental strand.

The RAD6 protein has been extensively biochemically characterized and is one of twelve or more ubiquitin-conjugating (E2) enzymes (Ciechanover, 1994). RAD6 specifies a ubiquitin-conjugating enzyme that is shown to be able to ubiquitinate histones 2A and 2B *in vitro* (Jentsch *et al.*, 1987; Sung *et al.*, 1988). Histone ubiquitination is thought to be implicated in the modulation of chromatin conformation, required for various DNA-metabolizing processes; this could explain the extremely pleiotropic phenotype of *rad6* mutants, including defects in postreplication repair, induced mutagenesis, meiotic recombination, and sporulation (Siede, 1988). Structural and functional conservation of two human homologs of the yeast RAD6 was shown by Koken *et al.* (1991).

RAD18, a member of the RAD6 epistasis group, forms a distinct protein complex with RAD6. Unlike RAD6, RAD18 is a DNA-binding protein containing a cysteine-rich sequence (Zn^{2+}-finger) motif common to a number of diverse proteins that are thought to interact with DNA (Freemont *et al.*, 1991). The RAD6-RAD18 ubiquitin-conjugating-DNA binding complex might target either chromatin-modulating proteins that block DNA

repair enzymes or components of the DNA repair machinery for degradation (Bailly *et al.*, 1994). Genes such as *RAD6* and *RAD18* and possibly *RAD5* (*REV2*) are associated with mutagenic effects. The *REV1* and *REV3* genes are essential for UV mutagenesis, however, their mutants only show a modest level of UV sensitivity.

Single strand gaps occur opposite lesions that are generated by the postreplication repair pathway. Homologous recombination, mediated by the *RAD52* group, is thought to be necessary to repair these single strand gaps. From this it can be concluded that the classification of genes into different epistasis groups does not necessarily exclude their involvement in different pathways.

A number of *S. cerevisiae* yeast radiation checkpoint loci have been reported, including *RAD9*, *RAD17*, *RAD24*, *MEC1*, *MEC2*, and *MEC3* genes. Checkpoint control proteins recognize DNA damage and initiate DNA damage processing. The most extensively studied mutant of this class is the *rad9* null mutant, which has been assigned to the RAD6 epistasis group, based, however, on limited phenotypic characterization (Friedberg *et al.*, 1995). Cells treated with DNA-damaging agents such as ionizing radiation, arrest cell cycle progression, with G2 arrest being a classical response (Liu *et al.*, 1995). The G2 phase of the cell cycle is important for determining whether a cell will survive or die. Analysis of *S. cerevisiae* mutants has revealed that the *rad9* (and additionally *rad17* and *rad24*) mutants are defective in this G2 arrest following radiation (Weinert and Hartwell, 1988). The *rad9* gene has been cloned and is shown to encode a protein with a predicted molecular mass of 148 kDa (Schiestl *et al.*, 1989b). No apparent homolog of other species has been published to date, although a human homolog of this gene may now have been identified (see unpublished work presented in Chapter III). *RAD9* is not essential for growth of unirradiated cells, and deletion mutants appear to be viable (Weinert and Hartwell, 1990). An unusual characteristic of this mutant is the increased rate of chromosome loss in *rad9* cells growing under normal conditions. Mutations such as *rad9* inactivate feedback controls that detect DNA damage and allow time for lesion processing before cell division, and thereby abrogating lethal chromosome damage (Weinert and Hartwell, 1988). Recently, RAD17, RAD24 and MEC3 proteins are suggested to activate endonuclease activity that degrades the AC nucleotide stretch of DNA near telomeres (Lydall and Weinert, 1995). The RAD9 protein is suggested to act in another subpathway as an inhibitor of the exonuclease complex (Lydall and Weinert, 1995). Through mechanisms such as genomic instability, lesions in these controls are known to play an important role in the genesis and progression of cancer (Hartwell and Weinert, 1989), indicating the relevance of identification of equivalents from higher species.

Recombination repair

The third (RAD52) epistasis group of *S. cerevisiae* contains mutants principally involved in recombinational repair. Recombination is one of the most important repair processes involved in repair of DNA-double strand breaks (DSBs). DSBs can be induced by ionizing radiation, and by radiomimetic chemical agents, including endogenously-produced radicals (as discussed in the section on DNA lesions - Chapter II).

Homologous recombination and V(D)J recombination pathways both play a role in processing DSBs in vertebrate cells. V(D)J rearrangement occurs physiologically in the progenitors of B- and T-lymphocytes. During B cell differentiation, the functional genes encoding Ig heavy (H) and light (L) chains are generated by two types of joining. One type generates the exons encoding the variable (V) regions of the immunoglobulin chains; this is V(D)J recombination (reviewed by Lewis, 1994). The other type reconstructs a rearranged gene encoding an IgH chain by replacing the segment encoding a given constant (C) region (which determines the class of an immunoglobulin) with one encoding another, different C region; this is the immunoglobulin class switch (Weaver, 1995a/b). V(D)J recombination is initiated by specific DNA cleavages at recombinational signal sequences and it has been suggested that components of the RAD52 epistasis group for double-strand break repair could participate in this process (Jeggo *et al.*, 1995).

In cells undergoing meiosis, DSBs are involved in the formation of synaptonemal complexes, which cannot be resolved without recombination. The distribution of DSBs formed at recombination hotspots parallels that of meiotic crossovers. Presumably, these hotspots are more accessible for enzymes involving these modifications.

Most of the RAD52 epistasis group mutants were isolated on the basis of their sensitivity to ionizing radiation, exemplified by defective repair of DSBs (Friedberg *et al.*, 1995). At least eight genes: *RAD50*, *-51*, *-52*, *-54*, *-55*, *-57*, *MRE11* and *XRS2* belong to this epistasis group, as discussed in Table 4. However, the defect in this group of mutants is not restricted to the repair of X-ray induced DNA damage, but also involves meiotic recombination and/or mating type switching in yeast.

The mechanism is subdivided into three steps; the presynaptic, synaptic and post synaptic phase. During the first stage, DSBs are introduced and the DNA ends are processed. The synaptic phase covers repair synthesis after the search for homologous DNA, and the alignment of homologous DNA templates. Finally, during the last stage of the recombination process Holliday structures are formed. These recombination intermediates are resolved after branch migration of the junction (West, 1992; Heyer, 1994).

Yeast gene	X-ray sensitivity	Human gene	Features/function
RAD50[a]	high	unknown	NTP-binding, coiled-coil protein
RAD51[b]	very high	HHR51	Strand exchange?, RecA-like
RAD52[c]	very high	HHR52	Binds RAD51, not vital
RAD54[d]	very high	HHR54	Putative helicase
RAD55[e]	high	XRCC3?	Strand exchange?, RecA-like
RAD56[f]	moderate	unknown	not cloned
RAD57[g]	high	unknown	Strand exchange?, RecA-like
MRE11[h]	high	hMRE11	Binds RAD50
XRS2[i]	high	unknown	No homology apparent

Table 4. Yeast RAD52 epistasis group recombination repair proteins with their human cognates known to date. (a) Raymond and Kleckner, 1993 (b) Shinohara *et al.*, 1992, (c) Adzuma *et al.*, 1984, (d) Emery *et al.*, 1991, (e) Lovett, 1994, (f) Friedberg, 1988, (g) Kans and Mortimer, 1991, (h) Johzuka and Ogawa, 1995, (i) Ivanov *et al.*, 1994.

The *RAD50* gene is required for two interrelated events of the meiotic prophase: chromosome synapsis and recombination. Both processes are involved in proper segregation of chromosomes at the first meiotic division (Alani *et al.*, 1990). The RAD50 protein harbours a nucleotide-binding domain and a 'coiled-coil' region. Coiled-coil proteins are suspected to play a crucial role in chromosome condensation and/or maintenance of a condensed state (Gasser, 1995). The *RAD50*, *MRE11*, and *XRS2* mutants are deficient in repair of damaged DNA and in meiotic recombination. The MRE11 protein was found to interact with RAD50 and XRS2 proteins (Johzuka and Ogawa, 1995). The *MRE11*, *RAD50*, and *XRS2* mutants are proficient in mating-type switching and mitotic recombination; it has been suggested, that the corresponding enzymes are required for the introduction of DSBs, pointing to a role in the incision event during meiotic recombination.

The *RAD51*, *RAD52*, and *RAD54* genes seem to have a role in general recombination and in DNA repair that involves recombinational activity (Game, 1993). Mutants in these loci confer the most extreme X-ray sensitivity. Defects in *RAD51*, *RAD52*, or *RAD54* genes block both the repair of double-strand breaks and the radiation-induced mitotic recombination process. Human genes structurally homologous to the RAD51, (Shinohara *et al.*, 1993) RAD52 (Muris *et al.*, 1994), RAD54 (C. Troelstra &

R. Kanaar, unpubl. observation), and MRE11 (Petrini *et al.*, 1995) have recently been identified.

The RecA protein of *Escherichia coli* plays an essential role in genetic recombination in prokaryotic organisms and promotes synapsis and strand-transfer between homologous DNA molecules (West, 1992). RAD51, RAD55, RAD57, and DMC1 proteins share significantly homologous regions with the RecA protein (West, 1992). These sequence similarities may reflect functional biochemical similarities. The X-ray sensitive *rad55* mutant is deficient in X-ray induced mitotic recombination and this mutation appeared meiotically lethal (Lovett, 1994). Because of the cold-sensitive phenotype of both RAD55 and RAD57, it has been suggested that these gene products form a protein complex. This was confirmed by interaction shown with the two-hybrid system (Hays *et al.*, 1995). The *rad56* mutant is moderately sensitive to X-rays, but not much is known about this mutant. The *RAD53* gene (allelic to the *MEC2/SPK1/SAD1*) encodes a protein kinase, identified as being defective in the S-phase arrest (Zheng *et al.*, 1993; Allen *et al.*, 1994). To date this mutant is not regarded as a member of the RAD52 DSB repair pathway.

The other subset of recombinatorial enzymes specifically involves DNA end-binding. The DNA-dependent protein kinase (DNA-PK) is involved in double-strand break repair and lymphoid V(D)J recombination (Weaver, 1995b; Jackson and Jeggo, 1995). DNA-PK consists of three subunits: the Ku autoantigen heterodimer (p70, p80) which has DNA end-binding activity (Gottlieb and Jackson, 1993), and the kinase (p450) DNA-PK which is defective in *scid* mice (Fulop and Phillips, 1990; Biedermann *et al.*, 1991). The mechanistic overlap between rejoining DNA double strand breaks and V(D)J recombination is illustrated by rodent mutants defective in both processes: correction of some of the mammalian XRCC (X-ray Repair Cross Complementing) complementation groups and the *scid* mutant, with the Ku complex cDNAs, demonstrate this overlap (Jackson and Jeggo, 1995). The *scid* (severe combined immune deficiency) phenotype is characterized by an immune defect and radiosensitivity. The immune defect is caused by the inability to perform correct V(D)J recombination, whereas the inefficient rejoining of DSBs causes the radiosensitivity (Roth *et al.*, 1992). Several X-ray sensitive hamster mutant cell lines (*XRCC5*, *XRCC6*, and *XRCC7*) are now known to have genetic defects in these Ku complex-related genes. XRCC5-7 mutants arrest in the G2 phase of the cell cycle upon DNA damage, and appear to have wild-type G1-S phase cell cycle delays in response to DNA damage (Sipley *et al.*, 1995).

A number of high molecular weight kinases have been identified that participate in

meiotic and V(D)J recombination, chromosome maintenance and repair, cell cycle progression, and cell cycle checkpoints (Hartley *et al.*, 1995; Hunter, 1995; Keith and Schreiber, 1995). Dysfunction of these enzymes (e.g. DNA-PK, *ATM*: AT mutated) can result in medical disorders ranging from a loss of immunological function to cancer (Savitsky *et al.*, 1995). In humans, many chromosomal aberrations, which are often involved in carcinogenesis, originate from erroneous recombination. The processing of double-strand breaks thus represents an important cellular property with a relevance that exceeds DNA repair alone.

Other repair pathways

Apart from the three epistasis groups described above, other repair mechanisms exist, including: base excision repair (BER), mismatch repair, and photoreactivation. Most species possess more than one pathway (such as NER, BER and PHR) to ensure the repair of the major toxic and mutagenic UV photoproducts.

The BER pathway mends oxidative DNA damage and damage caused by alkylation (Seeberg *et al.*, 1995) and other types of damage, such as UV-induced thymine glycols. These lesions are recognized by DNA glycosylases (e.g. DNA uracil glycosylase), which excise the lesion, leaving an apurinic/apyrimidinic (AP) site. Next, AP endonuclease action leaves a one-nucleotide gap in the DNA. DNA polymerase ß is the gap-filling BER polymerase in mammalian cells whereas DNA ligase III is thought to be the major ligase involved in this pathway. Another mammalian factor that might be involved in this pathway is XRCC1, which forms a complex with DNA ligase III (Caldecott *et al.*, 1994). The *XRCC1* gene corrects the defective DNA strand break repair and the sister chromatid exchange in the rodent mutant *XRCC1* (Thompson *et al.*, 1990).

The mutagenic potential of lesions such as O^6-alkylguanine and O^4-alkylthymine is obvious since DNA polymerases read these damaged bases respectively as adenine and cytosine. The repair of these lesions involves DNA-alkyltransferases (reviewed by Friedberg *et al.*, 1995).

The *E.coli PHR* gene is required for light-dependent photoreactivation, which directly reverses CPDs back to normal pyrimidines (Sancar, 1990). This enzyme enhances also light-independent nucleotide excision repair of UV-induced DNA damage. The evolutionarily preserved homologs of this photolyase enzyme involved in repair of thymidine dimers are discussed extensively in Chapter III.

DNA mismatch repair mends mispaired bases that are formed during DNA replication, genetic recombination and as a result of damage to DNA (Modrich, 1994).

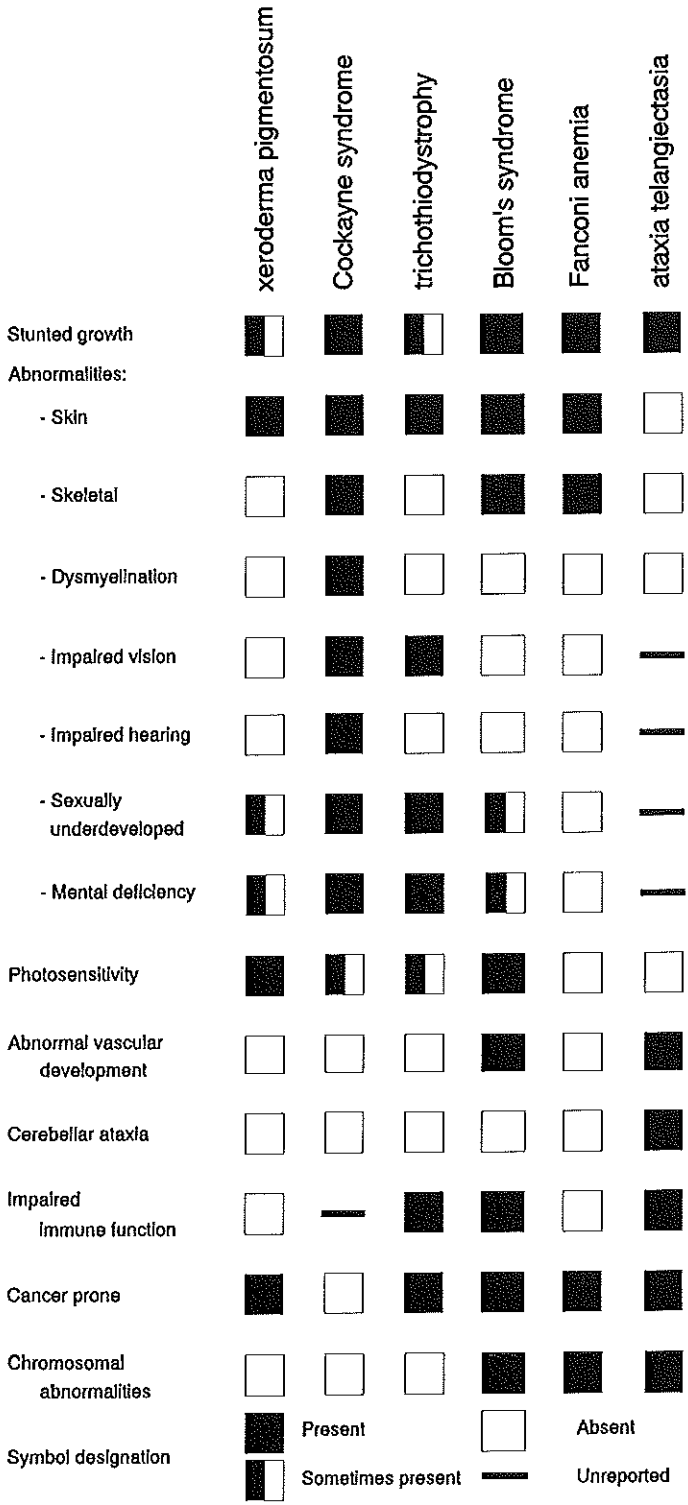
Eukaryote species contain a mismatch repair system related to the bacterial MutHLS system. Considerable evidence exists that this pathway plays an important role in maintaining replication fidelity, and processing recombination intermediates. Genetic defects in mismatch repair genes are known to play an important role in cancer-susceptibility syndromes and may play a role in sporadic cancers (Fishel and Kolodner, 1995). For example hereditary nonpolyposis colon cancer, characterized by microsatellite instability.

Repair syndromes and cancer

Genome instability, an early phenomenon in the tumor progression process, can be caused by defects in fundamental cellular processes such as DNA replication or repair, cell cycle arrest and apoptosis. Such genomic instability appears to pave the way for two other types of genetic events associated with tumor progression in humans. One of these types of genetic events is the altered expression/mutation of proto-oncogenes. The second type of event is the loss or inactivation of tumor suppressor genes which normally confer cellular growth control or growth restraint. The loss of function of such tumor suppressor genes can be caused by events such as deletions, point mutations or methylation (reviewed in Counts and Goodman, 1995). Areas of widespread hypomethylation, regional hypermethylation and increased cellular capacity for methylation have been associated with neoplasia (Baylin *et al.*, 1991), and may affect the transcription, replication and mutational rate of genes. However, methylation changes may not be major contributors to carcinogenesis. Loss of genetic information occurs in both sporadic cancers and human familial cancer predisposition syndromes. Studies with somatic cell hybrids have clearly shown that when malignant cells are fused with normal cells, the resulting hybrid cells are non-tumorigenic (Harris *et al.*, 1969; Stanbridge, 1976). This phenomenon of tumor suppression indicated that a gene (or genes) from a normal cell might replace a defective function in the cancer cell, thereby rendering it responsive to normal regulators of cell growth.

Cytogenetic analysis and molecular restriction fragment length polymorphism analysis indicate that specific chromosomal deletions are often associated with particular human malignancies (Stanbridge, 1990). In certain cancers, only a single tumor suppressor locus has been identified, whereas in others multiple loci are proved to be involved. Identification of cancer predisposition genes is relevant for early cancer detection and intervention strategies.

Carcinogenesis appears to be a multistep process in which an initiation event is followed by promotional events leading to tumor progression (Nowell, 1994). When DNA damage is not repaired before replication takes place, somatic mutation or chromosomal rearrangements or gene amplifications might occur and give rise to tumorigenesis. Incomplete, inefficient, and inaccurate repair in hereditary diseases such as xeroderma pigmentosum are clearly associated with increased carcinogenesis. This suggests that NER is a highly effective tumor prevention mechanism (Weeda *et al.*, 1990). Table 5 presents a provisional listing of syndromes and their features, directly or indirectly associated with radiation sensitivity due to a DNA repair defect.


xeroderma pigmentosum
Cockayne syndrome
trichothiodystrophy
Bloom's syndrome
Fanconi anemia
ataxia telangiectasia
Stunted growth
Abnormalities:
- Skin
- Skeletal
- Dysmyelination
- Impaired vision
- Impaired hearing
- Sexually underdeveloped
- Mental deficiency
Photosensitivity
Abnormal vascular development
Cerebellar ataxia
Impaired immune function
Cancer prone
Chromosomal abnormalities
Symbol designation
Present
Absent
Sometimes present
Unreported

Many of these DNA metabolism-affected conditions are associated with degenerative and neoplastic disorders. Carcinogenesis induced by radiation or chemicals in normal individuals might therefore be explained, for example, by inaccuracy of the repair mechanism.

The process of NER is especially important in the skin. Skin cancers seen in normal individuals are the same types seen at a high frequency in xeroderma pigmentosum. XP patients develop premalignant actinic keratoses and benign and malignant neoplasms at a median age of onset of eight years. Tumors in XP are predominantly basal- and squamous- cell carcinomas. The vast majority (97%) of these tumors occur on the face, head, and neck, the sites of greatest UV exposure. The spectrum of internal tumors found in XP patients is, however, different from that in the normal population. XP internal cancers are predominantly brain and central nervous system tumors, not clearly related to an environmental cause. Cleaver and Kraemer described a 2,000-fold increase in ocular neoplasms in XP patients under 20 years and a 10,000-fold increase in squamous cell carcinoma of the tip of the tongue (Cleaver and Kraemer, 1994). Ocular neoplasms occur predominantly in the anterior portion of the eye (lids, cornea, conjunctiva), the part that shields the posterior eye from UV radiation.

Xeroderma pigmentosum patients show a 1000-fold increased frequency of UV-induced skin cancer. CS patients exhibit no pigmentation abnormalities and increased (skin) tumors (Lehmann, 1987) and there is a lack of transcription coupled repair, in contrast to the increased levels of cancer observed in the global genome repair-deficient XP-C patients (as discussed in the next paragraph). CS patients are reported to exhibit growth failure, neurodevelopmental and neurological dysfunction, cutaneous photosensitivity, sensorineural hearing loss, dental caries, and cachetic dwarfism (reviewed by Nance and Berry, 1992). In order to gain insight into the function of different DNA repair genes and the resulting complex clinical symptoms, mouse models have been generated. Besides null-alleles, subtle mutations similar to those found in patients have been introduced via gene targeting into mouse embryonic stem cells in an attempt to mimic the clinical situation. If a double knock-out (that is, of both alleles) is created, the problem of lethality can arise, depending on the importance of the particular factor for the process of DNA repair and possible involvement in other essential processes. The molecular defect in the genetically heterogenous diseases xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy resides in the nucleotide excision repair pathway. For some of the NER genes (ERCC1, XPA, XPC and CSB) gene targeting of mouse embryonic stem cells yielded knock-out mice (respectively:

McWhir *et al.*, 1993; Weeda, unpubl. results; Nakane *et al.*, 1995; de Vries *et al.*, 1995; Sands *et al.*, 1995; van der Horst, unpubl. results). Both *XPA*- and *XPC*- deficient mice were defective in NER and were highly susceptible to UV-B induced skin carcinogenesis (de Vries *et al.*, 1995; Sands *et al.*, 1995). These NER-deficient mice thus provide a relevant *in vivo* model to study the high incidence of skin cancer in XP complementation group A and C patients.

Combined XP/CS phenotypes are observed in XP complementation groups -B, -D, and -G (Vermeulen *et al.*, 1993). TTD patients are characterized by sulphur-deficient brittle hair and nails, neurodysmyelination and some other symptoms overlapping with the characteristics of CS (Itin and Pittelkow, 1990). The TTD patients identified to date have a NER defect assigned to XP group B, -D and TTD-A. Notably, the CS patients as well as TTD patients having a NER defect are not cancer-prone. Transgenic mice should help enhance our understanding of mutagenesis and carcinogenesis in the absence of excision repair. Mouse models for HHR23A and HHR23B, presently in progress, should provide insight into the function of these different gene products in cellular events such as in DNA repair, meiosis and carcinogenesis.

Hereditary nonpolyposis colon cancer (HNPCC) is one of the most common inherited cancer susceptibility syndromes (see Kolodner, 1995 for a review). This type of cancer is characterized by an early age of onset and is inherited in an autosomal dominant fashion with high penetrance. The most common types of tumors observed in these families are colon tumors, while other types such as endometrial and ovarian tumors are found frequently in the group of patients that does not develop colon cancer. The mismatch repair pathway is known to regulate recombination events, repair of physical/chemical DNA damage and can trigger cell cycle arrest as part of the checkpoint control system (Modrich, 1994).

Many tumors are associated with chromosomal aberrations that may be involved in the initiation or promotion stages of carcinogenesis. Chromosomal alterations have been used as biological endpoints to study the mutagenic effects of ionizing radiation and chemicals. Analysis of the effect of DNA damage on normal and transformed cell lines reveals that mutations that affect cell cycle checkpoints play an important role in the development of cancer (Murray, 1992; Hartwell and Kastan, 1994; Carr and Hoekstra, 1995; Strauss *et al.*, 1995).

Irradiation of mammalian cells has different effects, depending on where the cells are in the cell cycle at the time of irradiation. Cells that are irradiated with X-rays in G1 phase may not pass the restriction point (checkpoint) and initiate DNA replication until

they have repaired the DNA damage. Cells that are irradiated in G2 delay mitosis entry until the damage has been repaired. In AT cells, the delay at the checkpoints does not occur, which presumably contributes to their genetic and chromosomal instability (Hartwell, 1992; Meyn, 1993). Further biochemical elucidation of the pathways involving the different XP and other gene products (for example *ATM*, defective in Ataxia telangiectasia (Savitsky *et al.*, 1995)) should provide a better understanding of the basic paths leading to carcinogenesis and tumorigenesis. AT cells are abnormally sensitive to killing by ionizing radiation, and abnormally resistant to inhibition of DNA synthesis by ionizing radiation. Specific chromosomal translocations or inversions are associated with several types of human tumors, especially of hematopoietic lineage. Lymphomas, usually of B-cell origin, and chronic lymphocytic leukemia of the T-cell type are frequently observed. There is a high frequency of chromosome breaks, not involving the locus itself, leading to translocations and inversions (Boehm and Rabbitts, 1989).

The autosomal recessive disorders Fanconi anemia (FA) and Bloom syndrome (BS) are, like AT, characterized by an increased susceptibility to the development of malignancy. At least five FA complementation groups can be distinguished (Buchwald, 1995; Pronk *et al.*, 1995). FA patients manifest sensitivity to DNA crosslinking agents, and predisposition to malignancy. Moreover, bone marrow elements are all prone to be affected and cardiac, renal, and limb malformations as well as dermal pigmentary changes are characteristic for this disease. The primary dermatologic manifestations are pigmentation abnormalities, such as hyperpigmentation, hypopigmentation and *cafe-au-lait* spots in addition to cutaneous malignancies (Fanconi, G., 1967). Because of the crosslinking agent sensitivity phenotype, newly identified human genes which are related to yeast mutants sensitive to crosslinking agents are prime candidate genes for these diseases. An example of such a gene (*SNM1*) is given in Chapter III.

The clinical features of Bloom syndrome patients include short stature, sun-sensitive facial erythema, hyper- and hypopigmented patches on the skin, immuno-deficiency, male infertility, and predisposition to neoplasia (Ellis *et al.*, 1995). The gene defective in the chromosomal instability disorder, *BLM* localized on chromosome 15 was found to encode a protein with homology to the *E.coli* RecQ helicase, involved in the recombination repair pathway (Ellis *et al.*, 1995).

NER: XPC, HHR23 proteins and ubiquitin

This chapter describes in detail some of the components specifically involved in the GGR pathway. The schematic representation of the NER pathway shown in Figure 3 contrasts the situation in normal individuals with the defective situation occurring in XP-C and CS patients. In patients with CS, TCR is impaired, but the global NER facility is unaffected (Venema *et al.*, 1990). In XP-C the NER defect is limited to GGR, whereas repair of the transcribed strand of expressed genes is not affected. Other XP complementation groups have been shown to be defective in both TCR as well as GGR. XP-C cells, in contrast to other XP and CS complementation groups, show no defect in RNA synthesis recovery after UV irradiation (Mayne and Lehmann, 1982).

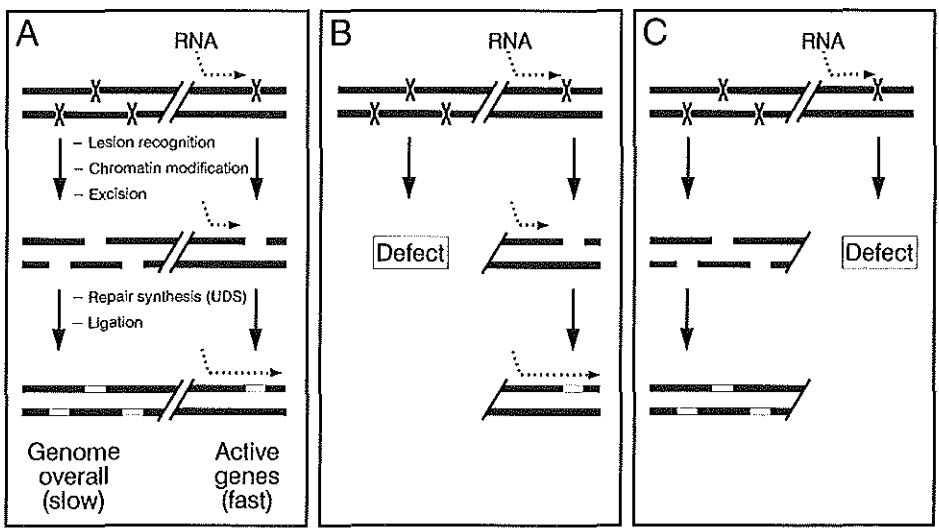


Figure 3 Stylized NER in health and disease. (A) Representation of the two distinct NER subpathways. On the right is shown transcription-coupled repair, characterized by the rapid and efficient removal of lesions in the transcribed strand of active genes, whereas the left part represents the slower global genome repair pathway. (B) This panel shows the situation for XP-C patients. The defective pathway is in the slower and less efficient repair of bulk DNA, including the non-transcribed strand of active genes (global genome repair). (C) CS patients show the converse opposite defect to that of XP-C patients. In this case, the defect is restricted to the faster TCR pathway.

The residual repair in XP-C could be completely abolished by a specific inhibitor of RNA pol II chain elongation (Carreau and Hunting, 1992). This finding is consistent with the fact that XP-C is deficient in global genome repair but not in repair of expressed genes. Therefore, residual repair in XP-C cells specifically represents the repair of the

transcribed strand of active genes. The defect in XP-C cells cannot simply be a failure in opening up inactive chromatin, as such differences in chromatin structure are not likely to exist between the two strands of active genes. Naturally-occurring deletions in the promoter region of the adenosine deaminase gene provided convincing evidence that more rapid repair could occur in potentially active genes (compared to the genome overall), in the absence of transcription (Venema *et al.*, 1992). *In vitro* NER experiments selectively reflect the global repair subpathway, since the UV-damaged plasmid used as a substrate is not transcribed under these conditions and XP-C extracts are very deficient in NER in these assays. Therefore, repair activity cannot be measured in this way in CS cells.

Two clinical forms of XP exist. One form involves progressive degenerative changes of the skin and eyes, whilst the other form also includes progressive neurologic degeneration. Within these two clinical forms of XP, complementation analysis allows further biochemical classification (Table 6).

Complementation group	DNA repair capacity UDS % of normal	Neurological abnormalities	Cancer prone
XP-A	< 2	severe	+
XP-B	3-7	variable	+
XP-C	10-20	none	+
XP-D	25-50	moderate	+
XP-E	40-50	none	+
XP-F	10-20	none	+
XP-G	2-25	severe	+
XP variant	100	none	+/-
CS-A	100	severe	-
CS-B	100	severe	-
TTD-A	10	variable	-

Table 6 Summary of repair characteristics and neurological abnormalities observed in XP patients (adapted from Cleaver and Kraemer, 1994).

The non-neurologic forms of XP involve most patients in groups C, E, and F, plus a "variant" form with normal excision repair (Cleaver and Kraemer, 1994). The neurologic forms consist of most patients in groups A, B, D, and G in which accellerated

neurodegeneration occurs. However, neurodysmyelination possibly caused by a general transcription defect is observed in CS, TTD, XPB, XPD and XPG patients. In addition to their repair function, the gene products involved have an additional role in transcription initiation.

RAD4/XPC

Experimental work described in Chapter IV of this thesis deals with the association of XPC and HHR23B proteins, which form a tight complex (Masutani *et al.*, 1994). This complex has also recently been described for *S. cerevisiae* RAD4 and RAD23 proteins (Guzder *et al.*, 1995a).

The XPC protein was biochemically identified in two ways, first as a deduced amino acid sequence from a correcting cDNA (Legerski and Peterson, 1992), and secondly by microsequencing of purified polypeptides and subsequent cloning of the corresponding cDNA (Masutani *et al.*, 1994). An Epstein-Barr virus-based cDNA expression vector library, that replicates extrachromosomally in human cells, and contains a selection marker, was transfected into XP-C cells. Transformant cells were selected by hygromycin resistance and UV survival. A 3.5 kb functional cDNA encoding a XPC correcting protein was recovered from transformants (Legerski and Peterson, 1992). Microsequencing of the large subunit (125 kD) homogenously purified, XP-C correcting protein complex, revealed striking homology of the to the deduced amino acid sequence of the cDNA identified by Legerski and Peterson (1992). However, at the N-terminus an extension of 117 amino acids was found besides a tightly associated small 58 kD subunit encoding one of the two human RAD23 homologs as described in detail in Chapter IV.

As indicated in the previous section on NER, the XPC/HHR23B protein complex is specifically associated with the process of GGR. Intriguingly, the similarity in amino-acid sequence between the RAD4, *Drosophila* XPC (Henning *et al.*, 1994) and the human XPC proteins are not reflected in the DNA-repair deficiency of the yeast *rad4* mutant and the XP-C patients. The *rad4* mutant cells are completely deficient in NER (McCready, 1994) while XP-C cells retain the ability to repair the transcribed strand of expressed class-II genes, the "fast" repair of transcriptionally active matrix bound DNA (Venema et al 1991). Phenotypically, the yeast *rad7* and *rad16* mutants resemble human XP-C cells. Both mutants exhibit wild-type levels of CPD repair of the transcribed strand, whereas repressed genes are not repaired as are the non-transcribed strands of active genes (reviewed by Hanawalt and Mellon, 1993; Hanawalt, 1994; see also R. Verhage *et al.*, 1994). The strand-specific repair in *rad7* and *rad16* explains the moderate sensitivity to UV in these mutants.

Mutations of the XPC gene from different XP-C cell lines have been reported to be point mutations; however, several frameshift mutations have also been described (Li *et al.*, 1993). The mouse homolog of XPC has extensive sequence homology to the human XPC protein (79% identity). Homozygous knock-out mice are viable and showed no apparent increased susceptibility to spontaneous tumor formation (Sands *et al.*, 1995). However, these mutant mice are highly susceptible to ultraviolet-induced carcinogenesis and display pathological skin and eye changes consistent with the human disease (Sands *et al.*, 1995).

RAD23/HHR23

The experimental work described in this thesis (Chapters IV, V, VI and VII) focusses on the isolation and characterization of the human equivalents of the *S. cerevisiae RAD23* gene. The yeast *RAD23* gene has been isolated by functional complementation of the mutant with a cosmid that was localized to chromosome 5L. The *rad1*, *rad2*, *rad3* and *rad4* mutants are completely unable to repair either CPDs or 6-4PPs and all exhibit a very high sensitivity to UV irradiation. This is in contrast to the *S. cerevisiae rad23* mutant, which is moderately sensitive to UV and psoralen crosslinks (Miller *et al.*, 1982). However, no repair of either (6-4)PPs or dimers is observed in the *rad23* mutant after 50J/m^2 (McCready, 1994; Verhage *et al.*, 1995). Therefore, in the *rad23* mutant, the intermediate UV-sensitivity is not due to a similar defect in GGR repair as described for *rad7* and *rad16* by Verhage *et al.* (1995). Gene disruption studies have disclosed that *RAD23* does not have a vital function in yeast.

The induction of *RAD23* mRNA after UV-irradiation observed in *S. cerevisiae* by Madura and Prakash, (1990) is not essential for the removal of CPDs from the global genome or for transcription-coupled repair of transcribed strands in expressed genes (Sweder and Hanawalt, 1992).

Some of the yeast DNA damage-inducible genes share a similar 5' upstream element (Siede and Friedberg, 1992). This suggests involvement of this sequence in transcriptional induction of these genes in *S. cerevisiae*. In this regard *RAD23* resembles the other genes, *RAD2*, *RAD6*, *RAD7*, *RAD18*, *PHR*, *SNM1* and *RAD54*, of which some also exhibit increased transcription during meiosis. In contrast to *RAD23*, the *RAD4* gene is not induced by UV-light damage in wild type yeast cells (Choi *et al.*, 1993). None of the mammalian excision repair genes (except for DNA polß only on RNA level) is found to be damage-inducible as observed in yeast to date. This inducible feature is likely to be more relevant for uni-cellular organisms, which have to adapt immediately to their environment, in contrast to the mammalian cells that are part of a whole organism, with a

relatively constant microenvironment.

Intrinsic radiation sensitivity can be modified by other signalling genes which may not be the primary gene defective in radiation response. Potential candidate genes influencing radiosensitivity include proto-oncogenes, tumor suppressor genes, genes involved in DNA repair, cell cycle regulatory genes, and genes encoding chromatin-associated proteins.

Analysis of specific single and double mutants of *rad7* and *rad23* suggest, however, a functional relationship between these genes, namely that the RAD23 protein can substitute for the N-terminal part of the RAD7 protein (Perozzi and Prakash 1986; Schiestl and Prakash, 1989).

Ubiquitin and repair

The RAD23 protein is found to encode a ubiquitin-like fusion protein as depicted in Chapter III, Figure 1. Mutation analysis has demonstrated that the ubiquitin-like region could be replaced by the ubiquitin sequence without affecting RAD23 activity (Watkins *et al.*, 1993). However, removal of this conserved domain affected the UV-survival.

Two distinct proteolytic pathways can be discerned in mammals: the lysosomal degradation pathway and the regulatory ubiquitin-conjugation pathway. The bulk of *in vivo* degradation of abnormal and naturally short-lived proteins is mediated by the ubiquitin-dependent pathway. A great deal of mechanistic information exists about ubiquitin and ubiquitin-related proteins in diverse species (Ciechanover, 1994). Along with selective protein degradation, this pathway covers strikingly diverse functions, including DNA repair, cell cycle control, and the stress response. Ubiquitin is an extremely strongly conserved and highly abundant protein of 76 amino acid residues that is present in all eukaryotic cells. The extreme evolutionary sequence conservation suggests that abnormal ubiquitin may be deleterious to cells. This abundant protein is distributed in the cytosol, nucleus, and on the cell surface (Ciechanover, 1994).

Ubiquitin encoding genes are present in two classes within the genome: a class of polyubiquitin genes, and a class of ubiquitin-fusion genes. The polyubiquitin genes encode from 2 to more than 50 head-to-tail repeats of ubiquitin. In contrast, the carboxy-terminal extension proteins encode ubiquitin followed by a downstream C-terminal extension of 76-80 amino acid residues, the latter encoding essential ribosomal proteins (Finley, *et al.*, 1989).

Besides these two classes of ubiquitin, evolutionarily diverged ubiquitin-like proteins exist. The class of ubiquitin-like proteins is involved in a variety of processes,

and the function of the diverged ubiquitin moiety is not known. Although the ubiquitin-like domain is expected to have a similar structure, the antigenic determinants differ substantially. Examples of ubiquitin-like fusion proteins are RAD23, GdX, fau, AN1A, AN1B, BAT3, NEDD8 (Watkins *et al.*, 1993; Toniolo *et al.*, 1988; Michiels *et al.*, 1993; Linnen *et al.*, 1993; Banerji *et al.*, 1990; Kumar *et al.*, 1993).

Proteinases which precisely cleave ubiquitin from proteins and peptides are necessary at various steps of the ubiquitin pathway. In order to recycle ubiquitin after protein degradation, isopeptidases are required to cleave ubiquitin from small peptides and conjugates that are reversibly ubiquitinated. C-terminal hydrolases recognize the RGG residues of ubiquitin. Similar to GdX, AN1a, AN1b, and BAT3, the RAD23 homologs bear no intact cleavage site and do not undergo proteolytic processing by a hydrolase.

The first enzyme, part of the ubiquitin-conjugation cascade, is the ubiquitin-activating enzyme E1. After the initial formation of an E1-ubiquitin thiolester, ubiquitin is transferred to one of a number of ubiquitin-conjugating enzymes known as E2 proteins. All known E2 proteins have a 16 kDa conserved motif in common, to which ubiquitin is covalently attached. The last enzyme of the cascade, E3 (ubiquitin-protein ligase), binds to cellular proteins and facilitates the transfer of ubiquitin from the E2 to the substrate.
The lysine 48 residue of ubiquitin can act as an acceptor for covalent ubiquitin linkage. The Lys residue is the site of attachment of a multi-ubiquitin chain, whose formation is required for the degradation of at least some N-end rule substrates (Chau *et al.*, 1989). Addition of multiple ubiquitin adducts to a substrate appears to act as the signal for degradation of the tagged protein. Polyubiquitinated proteins are targeted for proteolysis by an ATP-dependent 26S multisubunit protease containing the 20S multicatalytic protease (Jentsch and Schlenker, 1995). Other proteins are stably monoubiquitinated, for example histones H2A and H2B.

It is of interest that the mammalian $E2_{25K}$ (Chen *et al.*, 1991) is involved in catalyzing the formation of Lys-48-linked multiubiquitin chains from ubiquitin alone, and shares a domain similar to a repeated region present in the RAD23 protein. Recently, RAD23 has been found to be a suppressor of a ubiquitination mutant (Madura, personal communication).

Many proteolytic substrates of the ubiquitin system have been identified. Among them are mitotic cyclins, G1 cyclins, the tumor suppressor p53, transcriptional regulators NF kappa B and its inhibitor, I kappa B. The programmed degradation of cyclins at specific stages of the cell cycle is a dramatic example of regulation by degradation (Glotzer *et al.*, 1991). This is illustrated by cyclin ubiquitination which determines the end of mitosis due to proteolysis of the substrate (Murray, 1995). Many proto-

oncoproteins have rapid turnover rates, attributable to the ubiquitin degradation pathway, e.g., N- and *c-myc*, *c-fos*, and E1A (Jentsch, 1992).

The so called "N-end rule" was believed to regulate the *in vivo* half-life of a protein by the identity of its N-terminal residue. According to this rule, the type and eventual post-translational modification of the penultimate amino acid residue at the N-terminus of the substrate determines the metabolic stability of this protein (Varshavsky, 1992). However, the stabilizing or destabilizing capacity of the N-terminus is an important structural determinant for recognition by the ubiquitin system but is not the only recognition signal for degradation. Proteins with either free or blocked N-termini can be recognized via structural domains that are downstream and distinct from this residue (Ciechanover and Schwartz, 1994). However, recent data has suggested that the N-end rule may be less relevant than previously thought (Ciechanover, 1994), since limited subsets of protein substrates obey this rule.

Alterations in ubiquitin expression have been associated with diseases of the immune system (e.g. systemic lupus erythematosus -SLE-) and nervous system (e.g. Alzheimer's disease, and Parkinson's disease)(Mayer *et al.*, 1991). The exact role of protein ubiquitination in the course of neurodegenerative diseases or viral diseases is still unknown. It is not clear whether changes in ubiquitin expression and regulation contribute to these pathological changes, or are a consequence of these disorders.

The association of ubiquitin with DNA repair became apparent from the yeast enzyme UBC2, which proved to be identical to the RAD6 DNA repair protein (Jentsch *et al.*, 1987). As discussed in the previous section, the phenotype of *rad6* mutants includes slow growth, sensitivity to a variety of DNA damaging agents, and defects in induced mutagenesis, postreplication repair, and meiotic recombination. The yeast UBC2 enzyme has been isolated as one of the major ubiquitin-conjugating enzymes *in vitro* capable of conjugating ubiquitin to histones. A second link with repair was found with the RAD23 ubiquitin-like fusion protein involved in NER. Detailed analysis of the ubiquitin-like part of RAD23 is described in Chapter 3. The ubiquitination of chromosomal proteins may either alter the chromatin structure directly, or may direct 'patched' degradation of chromosomal proteins and allow access of repair enzymes to the sites of lesion.

In conclusion, the surprising variety of functions mediated by the ubiquitin pathway indicates important directions for future research. Insight into the mechanisms of the regulation of degradation of specific proteins such as proto-oncoproteins is of importance for understanding of the regulation of cellular growth, differentiation, cell cycle progression and malignant transformation (Hunter and Pines, 1991, 1994).

References:

Aboussekhra, A., Biggerstaff, M., Shivji, M.K.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protic, M., Hübscher, U., Egly, J.-M. and Wood, R.D. (1995). Mammalian DNA Nucleotide excision repair reconstituted with purified protein components. *Cell* 80: 859-868.

Abramic, M., Levine, A.S. and Protic, M. (1991). Purification of an ultraviolet-inducible, damage-specific DNA-binding protein from primate cells. *J. Biol. Chem.* 266: 22493-22500.

Adzuma, K., Ogawa, T. and Ogawa, H. (1984). Primary structure of the RAD52 gene in Saccharomyces cerevisiae. *Mol. Cel. Biol.* 4: 2735-2744.

Alani, E., Padmore, R. and Kleckner, N. (1990). Analysis of wild-type and *rad50* mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. *Cell* 61: 419-436.

Allen, J.B., Zhou, Z., Siede W., Friedberg, E.C. and Elledge, S.J. (1994). The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes and Development* 8: 2416-2428.

Auerbach, A.D. (1995). Fanconi anemia. *Dermatol. clin.* 13: 41-49.

Badame, A.J. (1989). Progeria. *Arch Dermatol.* 125: 540-544.

Bailly, V., Sommers, C.H., Sung, P., Prakash, L. and Prakash, S. (1992). Specific complex formation between proteins encoded by the yeast DNA repair and recombination genes RAD1 and RAD10. *Proc. Natl. Acad. Sci. USA* 89: 8273-8277.

Bailly, V., Lamb, J., Sung, P., Prakash, S. and Prakash, L. (1994). Specific complex formation between yeast RAD6 and RAD18 proteins: a potential mechanism for targeting RAD6 ubiquitin-conjugating activity to DNA damage sites. *Genes Dev.* 8: 811-820.

Banerji, J., Sands, J., Strominger, J.L. and Spies, T. (1990). A gene pair from human major histocompatibility complex encodes large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain. *Proc. Natl. Acad. Sci. USA* 87: 2374-2378.

Bang, D.D., Verhage, R., Goosen, N., Brouwer, J. and Putte, P. v.d. (1992). Molecular cloning of *RAD16*, a gene involved in differential repair in *Saccharomyces cerevisiae*. *Nucl. Acids Res.* 20: 3925-3931.

Bankmann, M., Prakash, L. and Prakash, S. (1992). Yeast *RAD14* and human xeroderma pigmentosum group A DNA repair genes encode homologous proteins. *Nature* 355: 555-558.

Bardwell, A.J., Bardwell, L., Iyer, N., Svejstrup, J.Q., Feaver, W.J., Kornberg, R.D. and Friedberg E.C. (1994). Yeast nucleotide excision repair proteins rad2 and rad4 interact with RNA polymerase II basal transcription factor b (TFIIH). *Mol. Cell. Biol.* 14: 3569-3576.

Baylin, S.B., Makos, M. and Wu, J. (1991). Abnormal patterns of DNA methylation in human neoplasia: potential consequences for tumor progression. *Cancer Cells* 3: 383-390.

Beamish, H. and Lavin, M.F. (1994). Radiosensitivity in ataxia telangiectasia:anomalies in radiation-induced cell cycle delay. *Int. J. Radiat. Biol.* 65: 175-184.

Biedermann, K.A., Sun, J.R., Giaccia, A.J., Tosto, L. and Brown, J.M. (1991). scid mutation in mice confers hypersensitivity to ionizing radiation and a deficiency in DNA double-strand break repair. *Proc. Natl. Acad. Sci. USA* 88: 1394-1397.

Biggerstaff, M., Szymkowski, D.E. and Wood, R.D. (1993). Co-correction of ERCC1, ERCC4 and

xeroderma pigmentosum group F DNA repair defects *in vitro*. *EMBO J.* 12: 3685-3692.

Bishop, J.M. (1995). Cancer: the rise of the genetic paradigm. *Genes and Development* 9: 1309-1315.

Boehm, T. and Rabbitts, T.H. (1989). A chromosomal basis of lymphoid malignancy in man. *Eur. J. Biochem.* 185: 1-17.

Bohr, V.A. (1991). Gene specific repair. *Carcinogenesis* 12: 1983-1992.

Brendel, M. and Ruhland, A. (1984). Relationships between functionality and genetic toxicology of selected DNA-damaging agents. *Mut. Res.* 133: 51-85.

Broughton, B.C., Steingrimsdottir, H., Weber, C.A. and Lehmann, A.R. (1994). Mutations in the xeroderma pigmentosum group D DNA repair/transcription gene in patients with trichothiodystrophy. *Nature Genetics* 7: 189-194.

Buchwald, M. (1995). Complementation groups: one or more per gene? *Nature Genetics* 11: 228-230.

Busch, D., Greiner, C., Lewis, K., Ford, R., Adair, G. and Thompson, L. (1989). Summary of complementation groups of UV-sensitive CHO mutants isolated by large-scale screening. *Mutagenesis* 4: 349-354.

Caldecott, K.W., McKeown, C.K., Tucker, J.D., Ljungquist, S. and Thompson, L.H. (1994). An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol. Cell. Biol.* 14: 68-76.

Carr, A.M. and Hoekstra, M.F. (1995). The cellular responses to DNA damage. *Trends in Cell Biology* 5: 32-40.

Carreau, M. and Hunting, D. (1992). Transcription-dependent and independent DNA excision repair pathways in human cells. *Mutat. Res.* 274: 57-64.

Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K. and Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* 243: 1576-1583.

Chen, Z., Niles, E.G. and Pickart, C.M. (1991). Isolation of a cDNA encoding a mammalian multiubiquitinating enzyme (E225k) and overexpression of the functional enzyme in *Escherichia coli*. *J. Biol. Chem.* 266: 15698-15704.

Choi, I.S., Kim, J.B., Jeon, S.H. and Park, S.D. (1993). Expression of *RAD4* gene of *Saccharomyces cerevisiae* that can be propagated in *Escherichia coli* without inactivation. *Biochem. Biophys. Res. Comm.* 193: 191-197.

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell* 79: 13-21.

Ciechanover, A. and Schwartz, A.L. (1994). The ubiquitin-mediated proteolytic pathway: Mechanisms of recognition of the proteolytic substrate and involvement in the degradation of native cellular proteins. *FASEB J.* 8: 182-191.

Cleaver, J.E. (1968). Defective repair replication in xeroderma pigmentosum. *Nature* 218: 652-656.

Cleaver, J.E. and Kraemer, K.H. (1994). Xeroderma pigmentosum and Cockayne syndrome. In: The metabolic basis for inherited disease. Seventh edition, Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., eds. (New York: McGraw-Hill Book Co.).

Collins, A.R. (1993). Mutant rodent cell lines sensitive to ultraviolet light, ionizing radiation and cross-linking agents: a comprehensive survey of genetic and biochemical characteristics. *Mutat. Res.* 293: 99-118.

Conaway, R.C. and Conaway, J.W. (1993). General initiation factors for RNA polymerase II. *Annu. Rev.*

Biochem. 62: 161-190.

Counts, J.L. and Goodman, J.I. (1995). Alterations in DNA methylation may play a variety of roles in carcinogenesis. *Cell* 83: 13-15.

Coverley D., Kenny M.K., Lane D.P. and Wood, R.D. (1992). A role for the human single-stranded DNA binding protein HSSB/RPA in an early stage of nucleotide excision repair. *Nucl. Acids Res.* 20: 3873-3880.

Coverley, D., Kenny, M.K., Munn, M., Rupp, W.D., Lane, D.P. and Wood, R.D. (1991). Requirement for the replication protein SSB in human DNA excision repair. *Nature* 349: 538-541.

De Vries, A., van Oostrom, C.Th.M., Hofhuis, F.M.A., Dornant, P.M., Berg, R.J.W., de Gruijl, F.R., Wester, P.W., van Kreijl, C.F., Capel, P.J.A., van Steeg, H. and Verbeek, S.J. (1995). Increased susceptibility to ultraviolet-B and carcinogens of mice lacking the DNA excision repair gene *XPA*. *Nature* 377: 169-173.

Ellis, N.A., Groden, J., Ye, T.-Z., Straughen, J., Lennon, D.J., Ciocci, S., Proytcheva, M. and German, J. (1995). The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* 83: 655-666.

Emery, H.S., Schild, D., Kellogg, D.E. and Mortimer, R.K. (1991). Sequence of RAD54, a Saccharomyces cerevisiae gene involved in recombination and repair. *Gene* 104: 103-106.

Fanconi, G. (1967). Familial constitutional panmyelocytopathy, Fanconi's anemia (FA) I. Clinical aspects. *Semin. Hematol.* 4: 233-240.

Feaver, W.J., Svejstrup, J.Q., Henry, N.L. and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell* 79: 1103-1109.

Finley, D., Bartel, B. and Varshavsky, A. (1989). The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* 338: 394-401.

Fishel, R. and Kolodner, R.D. (1995). Identification of mismatch repair genes and their role in the development of cancer. *Curr. Biol.* 5: 382-395.

Freemont, P.S., Hanson, I.M. and Trowsdale, J. (1991). A novel cysteine-rich sequence motif. *Cell* 64: 483-484.

Friedberg, E.C. (1988). Deoxyribonucleic acid repair in the yeast *S. cerevisiae Microbiol. Rev.* 52: 70-102.

Friedberg, E.C. (1991). Yeast genes involved in DNA-repair processes: new looks on old faces. *Mol. Microbiol.* 5: 2303-2310.

Friedberg, E.C. (1994). DNA repair: Looking back and peering forward. *BioEssays* 16: 645-649.

Friedberg, E.C., Walker, G.C. and Siede, W. (1995). DNA repair and mutagenesis. ASM Press Washington, D.C.

Fulop, G.M. and Phillips, R.A. (1990). The scid mutation in mice causes a general defect in DNA repair. *Nature* 347: 479-482.

Game, J. (1993). DNA double-strand breaks and the *RAD50-RAD57* genes in *Saccharomyces*. *Sem. Cancer Biol.* 4: 73-83.

Gasser, S.M. (1995). Coiling up chromosomes. *Curr. Biol.* 5: 357-360.

Gatti, R.A., Boder, E., Vinters, H.V., Sparkes, R.S., Norman, A. and Lange, K. (1991). Ataxia-telangiectasia: an interdisciplinary approach to pathogenesis. *Medicine* 70: 99-117.

Gietz, R.D. and Prakash, S. (1988). Cloning and nucleotide sequence analysis of the *Saccharomyces cerevisiae RAD4* gene required for excision repair of UV-damaged DNA. *Gene* 74: 535-541.

Glotzer, M.W., Murray, A.W. and Kirschner, M.W. (1991). Cyclin is degraded by the ubiquitin pathway. *Nature* 349: 132-138.

Goodhead, D. (1994). Initial events in the cellular effects of ionizing radiations: clustered damage in DNA. *Int. J. Radiat. Biol.* 65: 7-17.

Gottlieb, T.M. and Jackson S.P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* 72: 131-142.

Grossman, L. and Yeung, A.T. (1990). The UvrABC endonuclease of *Escherichia coli*. *Photochem. Photobiol.* 51: 749-755.

Gulyas, K.D. and Donahue, T.F. (1992). *SSL2*, a suppressor of a stem-loop mutation in the *HIS4* leader encodes the yeast homologue of human ERCC3. *Cell* 69: 1031-1042.

Guzder, S.N., Bailly, V., Sung, P., Prakash, L. and Prakash, S. (1995a). Yeast DNA repair protein RAD23 promotes complex formation between transcription factor TFIIH and DNA damage recognition factor RAD14. *J. Biol. Chem.* 270: 8385-8388.

Hanawalt, P. and Mellon, I. (1993). Stranded in an active gene. *Curr. Biol.* 3: 67-69.

Hanawalt, P.C. (1994). Transcription-coupled repair and human disease. *Science* 266: 1957-1958.

Hartley, K.O., Gell, D., Smith, G.C.M., Zhang, H., Divecha, N., Connelly, M.A., Admon, A., Lees-Miller, S.P., Anderson, C.W. and Jackson, S.P. (1995). DNA-dependent protein kinase catalytic subunit: A relative of phosphatidylinositol 3-kinase and the Ataxia telangiectasia gene product. *Cell* 82: 849-856.

Hartwell, L. and Weinert, T. (1989). Checkpoints: controls that ensure the order of cell cycle events. *Science* 246: 629-634.

Hartwell, L. (1992). Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell* 71: 543-546.

Hartwell, L. and Kastan, M.B. (1994). Cell cycle control and cancer. *Science* 266: 1821-1828.

Harrington, J.J. and Lieber, M.R. (1994). Functional domains within FEN-1 and RAD2 define a family of structure-specific endonucleases: implications for nucleotide excision repair. *Genes and Development* 8: 1344-1355.

Harris, H., Miller, O.J., Klein, G., Worst, P. and Tachibana, T. (1969). Suppression of malignancy by cell fusion. *Nature* 223, 363-368.

Hawk, J.L.M. (1991). Ultraviolet A radiation: staying within the pale. *B. M. J.* 302: 1036-1037.

Haynes, R.H. and Kunz, B.A. (1981). DNA repair and mutagenesis in yeast in: The molecular biology of the yeast *Saccharomyces*: Life cycle and inheritance (Stratern, J., Jones, E. and Broach, J. (eds.) Cold Spring Harbor Laboratories, Cold Spring Harbor, New York pp. 371-414.

Hays, S.L., Firmenich, A.A. and Berg, P. (1995). Complex formation in yeast double-strand break repair: participation of RAD51, RAD52, RAD55, and RAD57 proteins. *Proc. Natl Acad. Sci. USA* 92: 6925-6929.

Henning, K.A., Li, L., Iyer, N., McDaniel, L.D., Reagan, M.S., Legerski, R., Schultz, R.A., Stefanini, M., Lehmann, A.R., Mayne, L.V. and Friedberg, E.C. (1995). The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. *Cell* 82: 555-564.

Henning, K.A., Peterson, C., Legerski, R. and Friedberg, E.C. (1994). Cloning the *Drosophila* homolog of the xeroderma pigmentosum complementation group C gene reveals homology between the predicted human and *Drosophila* polypeptides and that encoded by the yeast *RAD4* gene. *NAR* 22: 257-261.

Heyer, W.D. (1994). The search for the right partner: homologous pairing and DNA strand exchange proteins in eukaryotes. *Experientia* 50: 223-233.

Hochstrasser, M. (1995). Ubiquitin, proteasomes, and the regulation of intracellular protein degradation. *Curr. Opin. Cell Biol.* 7: 215-223.

Hoeijmakers, J.H.J. (1993a). Nucleotide excision repair I: from E.coli to yeast. *TIG* 9: 173-177.

Hoeijmakers, J.H.J. (1993b). Nucleotide excision repair II: from yeast to mammals. *TIG* 9: 211-217.

Hunter, T. and Pines, J. (1991). Cyclins and cancer. *Cell* 66: 1071-1074.

Hunter, T. and Pines, J. (1994). Cyclins and cancer II: Cyclin D and CDK inhibitors come of age. *Cell* 79: 573-582.

Hughes, E.J., Eichbaum, Q. and Beatty, D.W. (1994). Human monoclonal antibodies: analysis of two antibodies derived from lymphocytes of a patient with acute rheumatic fever. *Med. Lab. Sci.* 48: 244-255.

Humbert, S., van Vuuren, H., Lutz, Y., Hoeijmakers, J.H.J. Egly, J.-M. and Moncollin. (1994). p44 and p34 subunits of the BTFII/TFIIH transcription factor have homologies with SSL1, a yeast protein involved in DNA repair. EMBO J. 13: 2393-2398.

Hunter, T. (1995). When is a lipid kinase not a lipid kinase? when it is a protein kinase. *Cell* 83: 1-4.

Itin, P.H. and Pittelkow, M.R. (1990). Trichothiodystrophy: Review of sulfur-deficient brittle hair syndromes and association with the ectodermal dysplasias. *J. Americ. Acad. Dermatol.* 22: 705-717.

Ivanov, E.L., Sugawara, N., White, C.I., Fabre, F., Haber, J.E. (1994). Mutations in XRS2 and RAD50 delay but not prevent mating-type switching in Saccharomyces cerevisiae. *Mol. Cell. Biol.* 14: 3414-3425.

Hwang, B.J. and Chu, G. (1993). Purification and characterization of a human protein that binds to damaged DNA. *Biochemistry.* 32: 1657-1666.

Jackson, S.P. and Jeggo, P.A. (1995). DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK. *Trends in Biochem. Sciences* 20: 412-415.

Jaspers, N.G.J. and Hoeijmakers, J.H.J. (1995). Nucleotide excision-repair in the test tube. *Curr. Biol.* 5: 700-702.

Jaspers, N.G.J., Gatti, R.A., Baan, C., Linssen, P.C.M.L. and Bootsma, D. (1988). Genetic complementation analysis of ataxia telangiectasia and Nijmegen breakage syndrome: a survey of 50 patients. *Cytogenet. Cell. Genet.* 49: 259-263.

Jeggo, P.A., Taccioli, G.E. and Jackson, S.T. (1995). Menage à trois: double strand break repair, V(D)J recombination and DNA-PK. *BioEssays* 17: 949-957.

Jentsch, S. (1992). The ubiquitin-conjugation system. *Annu. Rev. Genet.* 26: 179-207.

Jentsch, S. and Schlenker, S. (1995). Selective protein degradation: A journey's end within the proteasome. *Cell* 82: 881-884.

Jentsch, S., McGrath, J.P. and Varshavsky, A. (1987). The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* 329: 131-134.

Johnson, R.T. and Squires, S. (1992). The XPD complementation group. Insights into xeroderma pigmentosum, Cockayne's syndrome and trichothiodystrophy. *Mutat. Res.* 273: 97-118.

Johzuka, K. and Ogawa, H. (1995). Interaction of MRE11 and RAD50: two proteins required for DNA repair and meiosis-specific double-strand break formation in *Saccharomyces cerevisiae*. *Genetics* 139: 1521-1532.

Jones, C.J. and Wood, R.D. (1993). Preferential binding of the xeroderma pigmentosum group A complementing protein to damaged DNA. *Biochem.* 32: 12096-12104.

Keeney, S., Chang, G.J. and Linn, S. (1993). Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E. *J. Biol. Chem.* 268: 21293-21300.

Keeney, S., Eker, A.P.M., Brody, T., Vermeulen, W., Bootsma, D.,Hoeijmakers, J.H.J. and Linn, S. (1994). Correction of the DNA repair defect in xeroderma pigmentosum group E by injection of a DNA damage-binding protein. *Proc. Natl. Acad. Sci. USA* 91: 4053-4056.

Keith, C.T. and Schreiber, S.L. (1995). PIK-related kinases: DNA repair, recombination and cell cycle checkpoints. *Science* 270: 50-51.

Koken, M., Reynolds, P., Jaspers-Dekker, I., Prakash, L. and Prakash, S. Bootsma, D. and Hoeijmakers, J.H.J. (1991). Structural and functional conservation of two human homologs of the yeast DNA repair gene RAD6. *Proc. Natl. Acad. Sci. USA* 88: 8865-8869.

Koleske, A.J. and Young, R.A. (1995). The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends in Biochem. Sci.* 20: 113-116.

Kolodner, R.D. (1995). Mismatch repair: mechanisms and relationship to cancer susceptibility. *TIBS* 20: 397-402. *Trends in Biochem. Sci* 20: 397-401.

Kumar, S., Yoshida, Y. and Noda, M. (1993). Cloning of a cDNA which encodes a novel ubiquitin-like protein. *Biochem. Biophys. Res. Comm.* 195: 393-399.

Lawrence, C.W. (1994). The *RAD6* DNA repair pathway in *S.cerevisiae*: What does it do, and how does it do it? BioEssays 16: 253-258.

Legerski, R. and Peterson, C. (1992). Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C. *Nature* 359: 70-73.

Lehmann, A.R., Kirk-Bell, S., Arlett, C.F., Harcourt, S.A., de Weerd-Kastelein, E.A., Keijzer, W. and Hall-Smith, P. (1977). Repair of ultraviolet light damage in a variety of human fibroblast cell strains. *Cancer Res.* 37 904-910.

Lehmann, A.R. (1987). Cockayne's syndrome and trichothiodystrophy: Defective repair without cancer. *Cancer Rev.* 7: 82-103.

Lehmann, A.R., Bootsma, D., Clarkson, S.G., Cleaver J.E., McAlpine, P.J., Tanaka, K., Thompson, L.H. and Wood, R.D. (1994). Nomenclature of human DNA repair genes. *Mutat. Res.* 315: 41-42.

Lewis, S.M. (1994). The mechanism of V(D)J joining: Lessons from molecular, immunological, and comparative analyses. *Adv. in Immunology* 56: 27-150.

Li, L., Bales, E.S., Peterson, C.A. and Legerski, R.J. (1993). Characterization of molecular defects in xeroderma pigmentosum group C. *Nature Genet.* 5: 413-417.

Li, L., Elledge, S.J., Peterson, C.A., Bales, E.S. and Legerski, R.J. (1994a). Specific association between the human DNA repair proteins XPA and ERCC1. *Proc. Natl. Acad. Sci. USA* 91: 5012-5016.

Li, R., Waga, S., Hannon, G.J., Beach, D. and Stillman, B. (1994b). Differential effects by the p21 CDK inhibitor on PCNA-dependent DNA replication and repair. *Nature* 371: 534-537.

Linnen, J.M., Bailey, C.P. and Weeks, D.L. (1993). Two related localized mRNAs from Xenopus laevis

encode ubiquitin-like fusion proteins. *Gene* 128: 181-188.

Liu, V.F., Boubnov, N.V. and Weaver, D.T. (1995). Cell cycle checkpoints and repair of ionizing radiation damage. *Stem cells* 13: 117-128.

Lovett, S.T. (1994). Sequence of the *RAD55* gene of *Saccharomyces cerevisiae*: similarity of RAD55 to prokaryotic RecA and other RecA-like proteins. *Gene* 142: 103-106.

Lydall, D. and Weinert, T. (1995). Yeast checkpoint genes in DNA damage processing: implications for repair and arrest. *Science* 270: 1488-1491.

MacInnes, M.A., Dickson, J.A., Hernandez, R.R., Learmonth, D., Lin, G.Y., Mudgett, J.S., Park, M.S., Schauer, S., Reynolds, R.J., Strniste, G.F. and Yu, J.Y. (1993). Human ERCC5 cDNA-cosmid complementation for excision repair and bipartite amino acid domains conserved with RAD proteins of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* 13: 6393-6402.

Madura, K. and Prakash, S. (1986). Nucleotide sequence, transcript mapping, and regulation of the *RAD2* gene of *Saccharomyces cerevisiae*. *J. Bacteriol.* 166: 914-923.

Madura, K. and Prakash, S. (1990). Transcript levels of the *Saccharomyces cerevisiae* DNA repair gene *RAD23* increase in response to UV light and in meiosis but remain constant in the mitotic cell cycle. *Nucl. Acids Res.* 18: 4737-4742.

Mäkelä, T.P., Tassan, J.-P., Nigg, E.A., Frutiger, S., Hughes, G. and Weinberg, R.A. (1994). A cyclin associated with the CDK-activating kinase MO15. *Nature* 371: 254-257.

Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio K., Tanaka, K., van der Spek, P.J., Bootsma D., Hoeijmakers, J.H.J. and Hanaoka F. (1994). Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homolog of yeast RAD23. EMBO J. 8: 1831-1843.

Matsuda, T., Saijo, M., Kuraoka, I., Kobayashi, T., Nakatsu, Y., Nagai, A., Enjoji, T., Masutani, C., Sugasawa, K., Hanaoka, F., Yasui, A. and Tanaka, K. (1995). DNA repair protein XPA binds replication protein A (RPA). *J. Biol. Chem.* 270: 4152-4157.

Mayer, R.J., Arnold, J., Laszlo, L., Landon, M. and Lowe, J. (1991). Ubiquitin in health and disease. *Biochim. Biophys. Acta* 1089: 141-157.

Mayne, L.V. and Lehmann, A.R. (1982). Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Research* 42: 1473-1478.

McCready, S. (1994). Repair of 6-4 photoproducts and cyclobutane pyrimidine dimers in *rad* mutants of *Saccharomyces cerevisiae*. *Mut. Res.* 315: 261-273.

McWhir, J., Selfridge, J., Harrison, D.J., Squires, S. and Melton, D.W. (1993). Mice with DNA repair gene (ERCC1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. *Nature genetics* 5: 217-223.

McKusick, V.A. (1992). Mendelian inheritance in man. Catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes. (Baltimore: The Johns Hopkins university press).

Meyn, M.S. (1993). High spontaneous intrachromosomal recombination rates in Ataxia telangiectasia. *Science* 260: 1327-1330.

Michiels, L., Van der Rauwelaert, E., Van Hasselt, F., Kas, K. and Merregaert, J. (1993). fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an antisense in the Finkel-Biskis-Reilly

murine sarcoma virus. *Oncogene* 8: 2537-2546.

Miller, R.D., Prakash, L. and Prakash, S. (1982). Defective excision of pyrimidine dimers and interstrand crosslinks in *rad7* and *rad23* mutants of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 188: 235-239.

Mimms, F. (1994). UV-B and ozone observations. *Science* 265: 722.

Mitchell, D.L. and Nairn, R.S. (1989). The biology of the (6-4) photoproduct. *Photochem. Photobiol.* 49: 805-819.

Modrich, P. (1994). Mismatch repair, genetic stability, and cancer. *Science* 266: 1959-1960.

Muris, D.F., Bezzubova, O., Buerstedde, J.M., Vreeken, K., Balajee, A.S., Osgood, C.J., Troelstra, C., Hoeijmakers, J.H.J., Ostermann, K., Schmidt, H., Natarajan, A., Eeken, J., Lohman, P. and Pastink, A. (1994). Cloning of human and mouse genes homologous to RAD52, a yeast gene involved in DNA repair and recombination. *Mutat. Res.* 315: 295-305.

Murray, A.W. (1992). Creative blocks: cell-cycle checkpoints and feedback controls. *Nature* 359: 599-604.

Murray, A.W. (1995). Cyclin ubiquitination: The destructive end of mitosis. *Cell* 81: 149-152.

Nakane, H. Takeuchi, S., Yuba, S., Saijo, M., Nakatsu, Y., Murai, H., Nakatsuru, Y., Ishikawa, T., Hirota, S., Kitamura, Y., Kato, Y., Tsunoda, Y., Miyauchi, H., Horio, T., Tokunaga, T., Matsunaga, T., Nikaido, O., Nishimune, Y., Okada, Y. and Tanaka, K. (1995). High incidence of ultraviolet-B or chemical-carcinogen-induced skin tumours in mice lacking the xeroderma pigmentosum group A gene. *Nature* 377: 165-168.

Nance, M.A. and Berry, S.A. (1992). Cockayne syndrome: Review of 140 cases. *Am. J. Med. Genet.* 42: 68-84.

Naumovski, L. and Friedberg, E.C. (1983). A DNA repair gene required for the incision of damaged DNA is essential for viability in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA* 80: 4818-4821.

Naumovski, L., Chu, G., Berg, P. and Friedberg, E.C. (1985). *RAD3* gene of *Saccharomyces cerevisiae*: nucleotide sequence of wild-type and mutant alleles, transcript mapping, and aspects of gene regulation. *Mol. Cell. Biol.* 5: 17-26.

Neer, E.J., Schmidt, C.J., Nambudripad, R. and Smith, T.F. (1994). The ancient regulatory-protein family of WD-repeat proteins *Nature* 371: 297-300.

Obe, G., Johannes, C. and Schulte-Frohlinde, D. (1992). DNA double-strand breaks induced by sparsely ionizing radiation and endonucleases ascritical lesions for cell death, chromosomal aberrations, mutations and oncogenic transformation. *Mutagenesis* 7: 3-12.

Nowell, P.C. (1994). Cytogenetic approaches to human cancer genes. *FASEB J.* 8: 408-413.

O'Donavan, A., Davies, A.A., Moggs, J.G., West, S.C. and Wood, R.D. (1994). XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* 371: 432-435.

O'Donavan and Wood, R.D. (1993). Identical defects in DNA repair in xeroderma pigmentosum group G and rodent ERCC group 5. *Nature* 363: 185-188.

Park, E., Guzder, S., Koken, M.H.M., Jaspers-Dekker, I., Weeda, G., Hoeijmakers, J.H.J., Pearson, P., Francomano, C., Foster, P., Bocchini, C., Li, P. and McKusick, V. (1994). The Status of Online Mendelian Inheritance in man (OMIM) medio 1994. *Nucl. Acids Res.* 22: 3470-3473.

Perozzi, G. and Prakash, S. (1986). *RAD7* gene of *Saccharomyces cerevisiae*: Transcripts, nucleotide sequence analysis, and functional relationship between the RAD7 and RAD23 gene products. *Mol. Cell.*

Biol. 6: 1497-1507.

Peto, J. and Darby, S. (1994). Lung cancer. Radon risk reassesed. *Nature* 368: 97-98.

Petrini, J.H.J., Walsh, M.E., DiMare, C., Chen, X.N., Korenberg, J.R. and Weaver, D.T. (1995). isolation and characterization of the human MRE11 homologue. *Genomics* 29: 80-86.

Pronk, J.C., Gibson R.A., Savoia, A., Wijker, M., Morgan, N.V., Melchionda, S., Ford, D., Temtamy, S., Ortega, J.J., Jansen, S., Havenga, C., Cohn, R.J., de Ravel, T.J., Roberts, I., Westerveld, A., Easton, D.F., Joenje H., Mathew, C.G. and Arwert, F. (1995). Localisation of the Fanconi anemia complementation group A gene to chromosome 16q24.3. *Nature genetics* 11: 338-340.

Raymond, W.E. and Kleckner, N. (1993). RAD50 protein of *S. cerevisiae* exhibits ATP-dependent DNA binding. *Nucl. Acids Res.* 21: 3851-3856.

Reynolds, P., Prakash, L., Dumais, D., Perozzi, G. and Prakash, S. (1985). Nucleotide sequence of the *RAD10* gene of *Saccharomyces cerevisiae EMBO J.* 4: 3549-3552.

Reynolds, P., Koken, M.H.M., Hoeijmakers, J.H.J., Prakash, S. and Prakash, L. (1990). The $rhp6^{+}$ gene of *Schizosaccharomyces pombe*: a structural and functional homolog of the RAD6 gene from the distantly related yeast *Saccharomyces cerevisiae. EMBO J.* 9: 1423-1430.

Riboni, R., Botta, E., Stefanini, M., Numata, M. and Yasui, A. (1992). Identification of the eleventh complementation group of UV-sensitive excision repair-defective rodent mutants. *Cancer Res.* 52: 6690-6691.

Roth, D.B., Menetski, J.P., Nakajima, P.B. Bosma, M.J. and Gellert, M. (1992). V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in *scid* mouse thymocytes. *Cell* 70: 983-991.

Roy, R., Adamczewski, J.P., Seroz, T., Vermeulen, W., Tassan, J.-P., Schaeffer, L., Nigg, E.A., Hoeijmakers, J.H.J. and Egly, J.-M. (1994a). The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell* 79: 1093-1101.

Roy, R., Schaeffer, L., Humbert, S., Vermeulen, W., Weeda, G. and Egly,, J.-M. (1994b). The DNA-dependent ATPase activity associated with the class II basic transcription factor BTF2/TFIIH. *J. Biol. Chem.* 269: 9826-9832.

Sancar, G.B. (1990). DNA photolyases: Physical properties, action mechanism, and roles in dark repair. *Mut. Res.* 236: 147-160.

Sands, A.T., Abuin, A., Sanchez, A., Conti, C.J. and Bradley, A. (1995). High suseptibility to ultraviolet-induced carcinogenesis in mice lacking *XPC. Nature* 377: 162-168.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S., Ashkenazi, M., Pecker, I., Frydman, M., Harnik, R., Patanjali, S.R., Simmons, A., Clines, G.A., Sartiel, Gatti, R.A., Chessa, L., Sanal, O., Lavin, M.F., Jaspers, N.G.J., Taylor, M.R., Arlett, C.F., Miki, T., Weissman, S.M., Lovett, M., Collins, F.S. and Shiloh, Y. (1995). A single Ataxia telagiectasia gene with a product similar to PI-3 kinase. *Science* 268: 1749-1753.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H.J., Chambon, P. and Egly, J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) transcription factor. *Science* 260: 58-63.

Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J.H.J. and Egly, J.M. (1994). The ERCC2/DNA repair protein is associated with the class

II BTF2/TFIIH transcription factor. *EMBO J.* 13: 2388-2392.

Schiestl, R.H. and Prakash, S. (1989a). Interactions of the *RAD7* and *RAD23* excision repair genes of *Saccharomyces cerevisiae* with DNA repair genes in different epistasis groups. *Curr. Genet.* 16: 219-223.

Schiestl, R.H., Reynolds, P., Prakash, S. and Prakash, L. (1989b). Cloning and sequence analysis of the *S. cerevisiae RAD9* gene and further evidence that its product is required for cell cycle arrest induced by DNA damage. *Mol. Cell. Biol.* 9: 1882-1896.

Sedgwick, R.P. and Boder, E. (1991). in Hereditary neuropathies and spinocerebellar atrophies (Vianney de Jong, J.M.B. ed.), pp. 347-423, Elsevier.

Seeberg, E., Eide, L. and Bjørås, M. (1995). The base excision repair pathway. *Trends in Biochem. Sciences* 20: 391-397.

Sibghat-Ullah, Husain, I, Carlton, W, Sancar, A. (1989). Human nucleotide excision repair *in vitro*: repair of pyrimidine dimers, psoralen and cisplatin adducts by HeLa cell-free extract. *Nucl. Acids Res.* 17: 4471-4484.

Shinohara, A., Ogawa, H. and Ogawa, T. (1992). RAD51 protein involved in repair and recombination in *S.cerevisiae* is a RecA-like protein. *Cell* 69: 457-470.

Shinohara, A., Ogawa, H., Matsuda, Y., Ushido, N., Ikeo, K. and Ogawa, T. (1993). Cloning of human, mouse and fission yeast recombination genes homologous to RAD51 and RecA. *Nature Genetics* 4: 239-243.

Shivji, M.K.K., Kenny, M.K. and Wood, R.D. (1992). Proliferating cell nuclear antigen is required for DNA excision repair. *Cell* 69: 367-374.

Shivji, M.K.K., Eker, A.P.M. and Wood, R.D. (1994). DNA repair defect in xeroderma pigmentosum group C and complementing factor from HeLa cells. *J. Biol. Chem.* 269: 22749-22757.

Siede, W. (1988). The *RAD6* gene of yeast: a link between DNA repair, chromosome structure and protein degradation? *Radiat. Environ. Biophys.* 27: 277-286.

Siede, W., Friedberg, A.S. and Friedberg, E.C. (1993). *RAD9*-dependent G1 arrest defines a second checkpoint for damaged DNA in the cell cycle of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 90: 7985-7989.

Siede, W. and Friedberg, E.C. (1992). Regulation of the yeast *RAD2* gene: DNA damage-dependent induction correlates with protein binding to regulatory sequences and their deletion influences survival. *Mol. Gen. Genet.* 232: 247-256.

Sipley, J.D., Menninger, J.C., Hartley, K.O., Ward, D.C., Jackson, S.P. and Anderson, C.W. (1995). Gene for the catalytic subunit of the human DNA-activated protein kinase maps to the site of the *XRCC7* gene on chromosome 8. *Proc. Natl. Acad. Sci. USA* 92: 7515-7519.

Stanbridge, E.J. (1976). Suppression of malignancy in human cells. *Nature* 260: 17-20.

Stanbridge, E.J. (1990). Human tumor suppressor genes. *Annu. Rev. Genet.* 24: 615-657.

Stefanini, M., Collins, A.R., Riboni, R., Klaude, M., Botta, E., Mitchell, D.L. and Nuzzo, F. (1991). Novel Chinese hamster ultraviolet-sensitive mutants for excision repair form complementation groups 9 and 10. *Cancer Res.* 51: 3965-3971.

Stefanini, M., Lagomarsini, P., Giliani, S., Nardo, T., Botta, E., Peserico, A., Kleijer, W.J., Lehmann, A.R. and Sarasin, A. (1993a). Genetic heterogeneity of the excision repair defect associated with trichothiodystrophy. *Carcinogenesis* 14: 1101-1105.

Stefanini, M., Vermeulen, W., Weeda, G., Giliani, S., Nardo, T., Mezzina, M., Sarazin, A., Harper, J.L., Arlett, C.F., Hoeijmakers, J.H.J. and Lehmann, A.R. (1993b). A new nucleotide-excision-repair gene associated with the disorder trichothiodystrophy. *Am. J. Hum. Genet.* 53: 817-821.

Stewart, E., Kobayashi, H., Harrison, D. and Hunt, T. (1994). Destruction of Xenopus cyclins A and B2, but not B1, requires binding to $p34^{cdc2}$. *EMBO J.* 13: 584-594.

Strauss, M. Lukas, J. and Bartek, J. (1995). Unrestricted cell cycling and cancer. *Nature Medicine* 1: 1245-1246.

Stumm, M., Gatti, R.A., Reis, A., Udar, N., Chrzanowska, K., Seemanova, E., Sperling, K. and Wegner, R.D. (1995). The Ataxia-telangiectasia-variant genes 1 and 2 are distinct from the Ataxia telangiectasia gene on chromosome 11q23.1. *Am. J. Hum. Genet.* 57: 960-962.

Sugasawa, K., Masutani, C. and Hanaoka, F. (1993). Cell-free repair of UV-damaged simian virus 40 chromosomes in human cell extracts. I. Development of a cell-free system detecting excision repair of UV-irradiated SV40 chromosomes. *J. Biol. Chem.* 268: 9098-9104.

Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* 265: 1241-1243.

Sung, P., Prakash, S. and Prakash, L. (1988). The RAD6 protein of Sacharomyces cerevisiae polyubiquitinates histones, and its acidic domain mediates this activity. *Genes and Development* 2: 1476-1485.

Svejstrup, J.Q., Wang, Z., Feaver, W.J., Wu, X., Bushnell, D.A., Donahue, T.F., Friedberg, E.C. and Kornberg, R.D. (1995). Different forms of TFIIH for transcription and DNA repair: Holo-TFIIH and a nucleotide excision repairosome. *Cell* 80: 21-28.

Sweder, K.S. and Hanawalt, P.C. (1992). Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. *Proc. Natl. Acad. Sci. USA* 89: 10696-10700.

Swift, M., Reitnauer, P.J., Morrell, D., Chase, C.L. (1987). Breast and other cancers in families with ataxia-telangiectasia. New Engl. J. Med. 316: 1289-1294.

Takao, M., Abramic, M., Moos, M., Otrin, V.R., Wootton, J.C., Mclenigan, M., Levine, A.S. and Protic, M. (1993). A 127 kDa component of a UV-damaged DNA-binding complex which is defective in some xeroderma pigmentosum group E patients is homologous to a slime mold protein. *Nucl. acids Res.* 21: 4111-4118.

Tardiff, R.G., Lohman, P.H.M. and Wogan, G.N. (1994). Methods to assess DNA damage and repair: Interspecies Comparisons. By John Wiley & sons Ltd, Chichester, England.

Tatsumi, K., Toyoda, M., Hashimoto, T., Furuyama, J.I., Kurihara, T., Inoue, M. and Takebe, H. (1987). Differential hypersensitivity of xeroderma pigmentosum lymphoblastoid cell lines to ultraviolet light mutagenesis. *Carcinogenesis* 8: 53-57.

Thompson, L.H., Brookman, K.W., Jones N.J., Allen, S.A. and Carrano, A.V. (1990). Molecular cloning of the human *XRCC1* gene, which corrects defective DNA strand break repair and sister chromatid exchange. *Mol. Cell. Biol.* 10: 6160-6171.

Thompson, L.H., Brookman, K.W., Weber, C.A., Salazar, E.P., Reardon, J.T., Sancar, A., Deng, Z. and Siciliano, M.J. (1994). Molecular cloning of the human nucleotide-excision-repair gene *ERCC4*. *Proc. Natl. Acad. Sci. USA* 91: 6855-6859.

Toniolo, D., Persico, M. and Alcalay, M. (1988). A "housekeeping" gene on the X chromosome encodes a protein similar to ubiquitin. *Proc. Natl. Acad. Sci. USA* 85: 851-855.

Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D. and Hoeijmakers, J.H.J. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* 71: 939-953.

Varshavsky, A. (1992). The N-end rule. *Cell* 69: 725-735.

Van Duin, M., De Wit, J., Odijk, H., Westerveld, A., Yasui, A., Koken, M.H.M., Hoeijmakers, J.H.J. and Bootsma, D. (1986). Molecular characterization of the human excision repair gene *ERCC-1*: cDNA cloning and amino acid homology with the yeast DNA repair gene *RAD10*. *Cell* 44: 913-923.

Van Duin, M., van den Tol, J., Warmerdam, P., Odijk, H., Meijer, D., Westerveld, A., Bootsma, D. and Hoeijmakers, J.H.J. (1988). Evolution and mutagenesis of the mammalian excision repair gene *ERCC1*. *Nucl Acids Res.* 16: 5305-5322.

Van Gool, A.J. Verhage, R., Swagemakers, S.M.A., van de Putte, P., Brouwer, J. Troelstra, C., Bootsma, D. and Hoeijmakers, J.H.J. (1994). *RAD26* , the functional *S.cerevisiae* homolog of the Cockayne syndrome B gene *ERCC6*. *EMBO J.* 13: 5361-5369.

Van Hoffen, A., Venema, J., Meschini, R., van Zeeland, A.A. and Mullenders, L.H.F. (1995). Transcription-coupled repair removes both cyclobutane pyrimidine dimers and 6-4 photoproducts with equal efficiency and in a sequential way from transcribed DNA in xeroderma pigmentosum group C fibroblasts. *EMBO J.* 14: 360-367.

Van Houten, B. (1990). Nucleotide excision repair in Escherichia coli. *Microbiol. Rev.* 54: 18-51.

Van Vuuren, A.J., Appeldoorn, E., Jaspers N.G.J., Odijk, H., Yasui, A., Bootsma, D. and Hoeijmakers, J.H.J. (1993). Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. *EMBO J.* 12: 3693-3701.

Van Vuuren, A.J., Vermeulen, W., Ma, L., Weeda G., Appeldoorn, E., Jaspers N.G.J., van der Eb, A.J., Bootsma, D., Hoeijmakers, J.H.J., Humbert, S., Schaeffer, L. and Egly, J.M. (1994). Correction of xeroderma pigmentosum repair defect by basal transcription factor BTF2 (TFIIH). *EMBO J.* 13: 1645-1653.

Varshavsky, A. (1992). The N-end rule *Cell* 69: 725-735.

Venema, J., Bartosova, Z., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H.F. (1992). Transcription affects the rate but not the extend of repair of cyclobutane pyrimidine dimers in the human adenosine deaminase gene. *J. Biol. Chem.* 267: 8852-8856.

Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H.F. (1991). Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. *Mol. Cell. Biol.* 11: 4128-4134.

Venema, J., Van Hoffen, A., Natarajan, A.T., Van Zeeland, A.A. and Mullenders, L.H.F. (1990). The residual repair capacity of xeroderma pigmentosum complementation group C fibroblasts is highly specific for transcriptionally active DNA. *Nucl. Acids Res.* 18: 443-448.

Vennos, E.M. and James, W.D. (1995). Rothmund-Thomson syndrome. *Derm. Clin.* 13: 143-150.

Verhage, R., Zeeman, A.-M., de Groot, N., Gleig, F., Bang, D.D., van de Putte, P. and Brouwer, J. (1994). The *RAD7* and *RAD16* genes, which are essential for pyrimidine dimer removal from the silent mating type loci, are also required for repair of the nontranscribed strand of an active gene in

Saccharomyces cerevisiae. Mol. Cell. Biol. 14: 6135-6142.

Verhage, R., Zeeman, A.-M., Lombaerts, M., van de Putte, P. and Brouwer, J. (1995). Analysis of gene- and strand-specific repair in the intermediately UV-sensitive *S.cerevisiae rad23* mutant. Submitted.

Vermeer, B.J. and Hurks, M. (1994). The clinical relevance of immunosuppression by UV irradiation. *J. Photochem. Photobiol.* 24: 149-154

Vermeulen, W., van Vuuren, A.J., Chipoulet, J.M., Schaeffer, L.,Appeldoorn, E., Weeda, G., Jaspers, N.G.J., Priestly, A., Arlett, C.F., Lehmann, A.R., Stefanini, M., Mezzina, M., Sarasin, D., Bootsma, D., Egly, J-M. and Hoeijmakers, J.H.J. (1994). Three unusual repair deficiencies associated with transcription factor BTF2 (TFIIH): Evidence for the existence of a transcription syndrome. Cold Spring Harbor Symp. Quant. Biol. 317-329.

Vermeulen, W., Jaeken, J., Jaspers, N.G.J., Bootsma, D. and Hoeijmakers, J.H.J. (1993). Xeroderma pigmentosum complementation group G associated with Cockayne's syndrome. *Am. J. Human Genet.* 53: 185-192.

Watkins, J.F., Sung, P., Prakash, L. and Prakash, S. (1993a). The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* 13: 7757-7765.

Watson, R. (1995). European union tackles the greenhouse effect. *B. M. J.* 310: 758.

Weaver, D.T. (1995a). V(D)J recombination and double-strand break repair. *Adv. in Immunology* 58: 29-85.

Weaver, D.T. (1995b). What to do at an end: DNA double-strand-break repair. *Trends in Genetics* 11: 388-392.

Weber, C.A., Salazar, E.P., Stewart, S.A. and Thompson, L.H. (1990). *ERCC2*: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast *RAD3*. *EMBO J.* 9: 1437-1447.

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). Molecular cloning and biological characterization of the human excision repair gene *ERCC3*. *Mol.Cell. Biol.* 10: 2570-2581.

Weemaes, C.M.R., Smeets, D.F.C.M. and van der Burgt, C.J.A.M. (1994). Nijmegen breakage syndrome: a progress report. *Int. J. Radiat. Biol.* 66: S185-S188.

Weinert, T.A. and Hartwell, L.H. (1988). The *RAD9* gene controls the cell cycle response to DNA damage in *S. cerevisiae*. *Science* 241: 317-322.

Weinert, T.A. and Hartwell, L.H. (1990). Characterization of the *RAD9* gene of *S. cerevisiae* and evidence that it acts posttranslationally in cell cycle arrest after DNA damage. *Mol. Cell. Biol.* 10: 6554-6564.

West, S.C. (1992). Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* 61: 603-640.

Wevrick, R., Clarke, C.A. and Buchwald, M. (1993). Cloning and analysis of the murine Fanconi anemia group C cDNA. *Hum. Mol. Genet.* 2: 655-662.

Whitney, M., Thayer, M., Reifsteck, C., Olson, S., Smith, L., Jakobs, P.M., Leach, R., Naylor, S., Joenje, H. and Grompe, M. (1995). Microcell mediated chromosome transfer maps the Fanconi anamia group D gene to chromosome 3p. *Nature genetics* 11: 341-343.

Wood, R.D., Robins, P. and Lindahl, T. (1988). Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. *Cell* 53: 97-106.
Zheng, P., Fay D.S., Burton, J., Xiao, H., Pinkham, J.L. and Stern, D.F. (1993). SPK1 is an essential S-phase-specific gene of *Saccharomyces cerevisiae* that encodes a nuclear serine/threonine/tyrosine kinase. *Mol. Cell. Biol.* 13: 5829-5842.

DNA repair, transcription, cell cycle progression, and ubiquitin conjugation are extensively studied processes, and there are many important papers on these subjects. However, a number of recent reviews contain extensive lists of references. Therefore, I am only referencing the most recently published articles along with some selected older literature. I refer the reader to the previously published reviews as a source of most relevant references, and apologize to my colleagues for not being able to cite all of their work.

CHAPTER

III

Cloning of repair genes by sequence homology

CLONING OF REPAIR GENES BY SEQUENCE HOMOLOGY

Introduction to sequence comparison

The human genome is estimated to consist of 50,000 to 100,000 genes. Sequence similarity programs are important tools that can be used to identify possible protein coding regions and to provide clues to gene and protein structure and function. Sequence data can sometimes provide unexpected links among diverse biological fields due to sequence similarity. One approach to identify homologs of genes in other species and to discover new connections is database homology searching.

Global similarity algorithms optimize the overall alignment of two single sequences, which may include large stretches of low similarity. This is in contrast to the local similarity algorithms that seek only relatively conserved subsequences. Local similarity measures are generally preferred for database searches.

Genes expressed in human tissues are rapidly being identified by random partial sequencing of cDNA clones. Several large genome sequencing centers are analyzing the cDNA sequence data, and are using PCR and somatic cell hybrid mapping panels to localize these sequences to individual human chromosomes (Adams *et al.*, 1991, Wilcox *et al.*, 1991, Gieser and Swaroop, 1992, Khan *et al.*, 1992, Polymeropoulos *et al.*, 1992, Boguski, 1995, Adams *et al.*, 1995). Partial sequencing of cDNAs to generate expressed sequence tags (ESTs) is a rapid and efficient way to establish a broadly-based profile of genes expressed in a particular tissue or cell type. Comparison of ESTs from different laboratories with cDNAs of other laboratories has begun to provide a 'computerized northern blot' of high sensitivity (Boguski and Schuler, 1995). Strong preference has been observed for highly expressed cDNAs in the EST database. Therefore, genes expressed in a tissue-specific way, or encoding low-abundance gene products will be difficult to detect. At the end of 1995, 391,777 ESTs had been deposited in the public domain dbEST database, of which 307,214 are of human origin. These ESTs represent approximately 10,000 previously identified genes and 20,000 new unknown genes from a total amount of approximately 80,000 genes. Roughly estimated, 40% of the genes known to date are involved in basal cellular processes such as energy housekeeping, cell structure, and cell division; 22% is thought to be involved in protein synthesis.

Normalization of cDNA libraries opens the possibility of gridding libraries on small filters due to a large reduction in the total number of clones that have to be arrayed. Subtractive hybridization is now being used to reduce the population of highly represented sequences in cDNA libraries. In this way, an increase in the relative frequency of

occurrence of rare cDNA clones has been achieved (Soares *et al.*, 1994). However, the use of subtracted libraries has the drawback that a representation of the transcription level, a feature of primary libraries, is not obtained (Adams *et al.*, 1995).

Single-run DNA sequencing, such as has been employed so effectively in the EST projects (Adams *et al.*, 1995), has proved to be an efficient method of obtaining sequence data on cDNA clones. The overall accuracy of these sequences is estimated to be about 97%, based on matches to known genes (Boguski, 1995). The aim of the human genome project is to generate an STS/EST-based physical genome map, which facilitates the identification of cDNAs localized within 100 kb of a mapped marker (Wilcox *et al.*, 1991). Suitable primers are designed for PCR that are used to amplify the corresponding sequence from genomic DNA. In this way, the EST is converted to a sequence tagged site (STS) that can be mapped to a genomic location using radiation hybrids or genomic clones such as YACs (Venter and Merril, 1992; Durkin *et al.*, 1994; Berry *et al.*, 1995; Soares and Sikela, 1995). Mapping data and finished maps are being deposited in a variety of public databases, including GDB or EMBL databases. ESTs that map near a locus for a given disease or phenotype can be used as probes for candidate genes for the locus of interest.

At present there are genome sequencing projects for many organisms including *Bacillus subtilis*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Escherichia coli*, *Saccharomyces cerevisiae*, and *Homo sapiens*. The *Haemophilus influenzae Rd* genome was the first complete genome to be sequenced from a free living organism (Fleischmann *et al.*, 1995). The complete nucleotide sequence (580,070 base pairs) of the *Mycoplasma genitalium* genome has also been determined recently (Fraser *et al.*, 1995). This species contains the smallest known genome of any free-living organism, and therefore may be a paradigm for the minimal set of genes (482) necessary for independent life (Goffeau, 1995). The complete genomic sequence of *S.cerevisiae* is estimated to be available in early 1996, and 90% of the open reading frames (ORF) in *C. elegans* are expected to be in the public databases by the end of 1997. A recently exploited model organism is the *Fugu rubripes* (Pufferfish) (Brenner *et al.*, 1993). This organism has the advantages that its genome, although estimated to contain an equivalent number of genes to mammals, is ten times smaller. Such features greatly simplify comparative mapping.

Sequencing of homologous regions of human and mouse genomes can provide insights into gene evolution, function and regulation. Comparison of homologous sequences from different species greatly increases the amount of information that can be extracted from analysis of any particular sequence of a single species.

The characterization of variation between DNA sequences elucidates the role of mutations, rearrangements and repeat sequences in the evolution of genomes and species. Prediction of transcribed regions is currently possible by using computer programs such as the GRAIL algorithm (Uberbacher and Mural, 1991). Comparing the genomes of virulent and harmless strains of bacteria should further enable identification of disease-causing genes. Analysis of human genome loci will dramatically expand in the coming years, and is relevant for identification of those regions most likely involved in genetic diseases. As for other areas, these possibilities are also relevant for the analysis of the molecular mechanism of DNA repair processes and their involvement in genetic disorders and cancer predisposition. This chapter presents the present status of the human genome sequencing programme and its use for the field of DNA repair.

Sequence comparison

The computer algorithm to perform sequence comparison, Basic Local Alignment Search Tool (BLAST), was used to identify those sequences reported in this thesis. The *S. cerevisiae* DNA repair protein sequences of interest were used to query the EST database by TBLASTN (Altschul *et al.*, 1990), to determine whether matches occur in undocumented open reading frames of higher order species. Sequences of interest were periodically re-searched against the databases to identify new matches based on similarity to newly deposited sequences. There are a variety of databases available via Email and the Internet to query with BLAST (Altschul *et al.*, 1994). For cloning purposes, the EMBL, Genbank, DDBJ, and the EST databases including the daily updates were routinely checked for new deposited genes of interest. These different databases show considerable overlap (Benson *et al.*, 1994). Therefore, to reduce search time, "nonredundant" sequence databases have been developed. The latest version of BLAST scans a protein database at approximately 500,000 residues per second, requiring powerful computer systems.

The best approach for identification of new clones is to initially compare the complete (protein) sequence of interest with the EMBL/Genebank (DNA) databases. Depending on the occurrence of non-specific hits that might overwhelm and bury significant homologies in a large output list, a little editing can improve the results significantly. Low complexity regions are quite common in proteins and usually lead to spurious homology results. These low complexity regions consisting of acidic regions, basic regions, repeated strings of a single amino acid, certain types of zinc-fingers, and other common motifs which appear frequently, are best removed or masked from the

sequence of interest to increase the specificity. A sequence can be masked by replacing these subsequences with "X", a character that the BLAST program ignores.

Additionally, sequence comparison searches using DNA enquiries into DNA databases might provide additional overlapping clones not sharing enough homology to be detected by a protein to DNA search. It should be noted that DNA sequences are highly nonrandom, with locally biased base composition. AT-rich regions and repeated sequence elements (e.g., Alu sequences) will produce an enormous output of matches with little interest.

Results from sequence comparison with considerable biological interest were selected for cloning by RT-PCR. This paragraph describes some examples of genes identified by TBLASTN searches, all having important implications for understanding the DNA damage repair processes. A number of genes have now been isolated in this way, illustrating the relevance of this procedure. Furthermore, it should be noted that most of the findings described here cannot be obtained by low stringency hybridization due to large divergence of the DNA sequences of these genes. Evidence for this comes from the *RAD23* gene that does not cross hybridize with the highly homologous mammalian *HHR23* genes.

Examples of other genes that were identified in this way are: *Saccharomyces cerevisiae XPE*, a second human XPE homologous protein, the human equivalent of the yeast Sensitive Nitrogen Mustard gene 1 (*SNM1*), the yeast *RAD26* gene (which is the counterpart of the human Cockayne syndrome CSB-protein; van Gool *et al.*, 1994), a human gene with homology to the photoreactivating enzyme and two *S. cerevisiae*, mouse and human equivalents of the *Schizosaccharomyces pombe* DNA double strand break repair enzyme Rad21 (McKay *et al.*, unpublished results). Additionally, a human protein sharing two homologous domains with the *S. cerevisiae* cell cycle arrest protein RAD9 was identified.

Some examples of significant scores obtained with the BLAST algorithm are discussed in the next paragraph.

Identification of RAD23, SNM1, PHR, XPE, and RAD9 equivalents.

The overall aim at the onset of this project was cloning of a mammalian equivalent of the yeast *RAD23* gene.

Towards this end, an initial attempt was made to clone RAD23 homologs of the distantly related yeast species *Schizosaccharomyces pombe* and the insect *Drosophila melanogaster* via interspecies hybridization. The *S. cerevisiae RAD23* gene was subdivided into two non-overlapping parts, giving a 5' and a 3' probe of the *RAD23* gene. Southern blots containing *Eco*RI, *Bam*HI, *Hind*III and *Pst*I digested *Schizosaccharomyces pombe* and *Drosophila melanogaster* genomic DNA were cross-hybridized with both non-overlapping *S. cerevisiae* probes. The rationale of this approach is to search for similar sized fragments called junction bands, recognized by two non-overlapping probes because this points to a long region of sequence homology. No such junction bands were observed after low stringency hybridization followed by autoradiography indicating the absence of long stretches of conserved DNA sequences. As a positive control for these probes, the DNA of yeast species closely related to *S. cerevisiae* were analysed. These showed clear junction bands, indicating that the probes used were recognizing the *RAD23* equivalents of less diverged species (Figure 3.1). From these experiments, it was concluded that the cloning of higher order species equivalents of RAD23 was impossible via interspecies hybridization, due to too high a degree of evolutionary diversification.

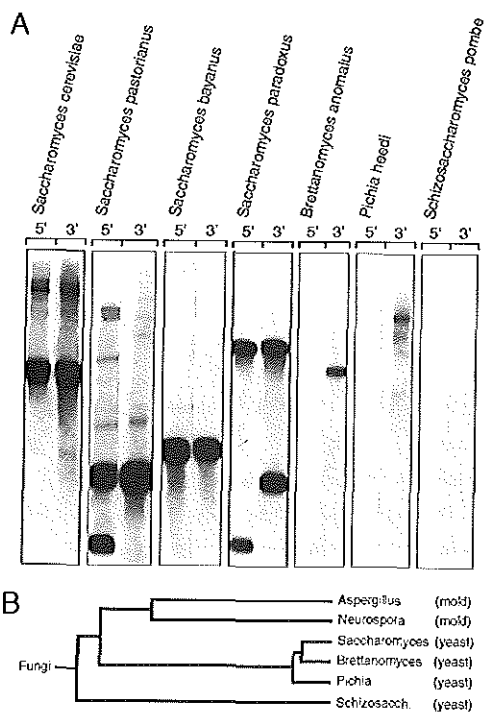


Figure 3.1

Evolutionary walking using yeast *S. cerevisiae* as a model organism.

(A) Southern blot analysis, demonstrating presence or absence of junction bands in yeast species. The blots were probed with both 5' and a 3' *S. cerevisiae RAD23* PCR-derived fragments of about 700 basepairs (see text).

(B) Phylogenetic tree of different yeast species indicating the relative evolutionary distances between these yeast strains.

Performing a sequence comparison search with the *S. cerevisiae* RAD23 protein sequence revealed a low but potentially significant homology of the C-terminus of RAD23 with a short part of a partial human cDNA sequence, isolated from a heart cDNA library by cross reactivity of acute rheumatic fever sera (Figure 3.2). Because of the correct sequential order of the four homologous regions, the percentage of homology for such distantly related species, their relative position compared to the yeast sequence and the correct position of the stop codon (not shown in Figure 3.2), it was judged that the human sequence was a structural homolog of *S. cerevisiae* RAD23. It was decided on this basis to clone the full length human gene. The overall homology extended to the N-terminal region that was missing on the partial cDNA obtained with the computer search. The full length cDNA isolated was a likely candidate for the human cognate of yeast *RAD23*, because of the additional high sequence homology and the presence of the start codon at the N-terminus.

```
YEAST:   264 SLRQVVSGNPEALRPLLENISARYPQLREHIMANPEVFVSMLLEAVG 310
             ++RQV+  NP  L +LL+ ++    PQL ++I  + E F+ ML E+ G
HUMAN:   174 NIRQVIQQNPALLPALLQQLGQENPQLLQQISRHQEQFIQMLNEPPG 314


YEAST:   172 AFNNPDRAVEYLLMGIPENLRQPEPQQQTAAAAEQPST 209
             ++NNP RAVEYLL GIP +         Q +   +EQP+T
HUMAN:     7 GYNNPHRAVEYLLTGIPGSPEPEHGSVQESQVSEQPAT 120


YEAST:   379 FACDKNEEAAANILFSDHAD 398
             + C+KNE  AAN L S++ D
HUMAN:   465 YLCEKNEXXAANFLLSQNFD 524


YEAST:   368 GFERDLVIQVY 378
             GF   LVIQ Y
HUMAN:   435 GFPESLVIQAY 467
```

Figure 3.2 Identification of the first human RAD23 equivalent, HHR23A by means of a TBLASTN search. Four homologous regions detected by using the *S. cerevisiae* RAD23 protein sequence as a query sequence. The numbers indicate the amino acid residues of the fragments identified to be homologous. The amino acid sequence is given in the one letter code, X indicates an uncertain amino acid, whereas + indicates similar amino acids. This cDNA identified with cardiac autoantigens was cloned from a heart cDNA library and deposited into the genbank database with accession number M77024 by Eichbaum *et al.*, (1991).

While cloning of the full length *HHR23A* cDNA was in progress, a second human homolog (*HHR23B*) recently deposited in the database was identified using the TBLASTN algorithm. This second candidate cDNA was present in the EST database and had diverged at a similar rate as the other human gene from the *S. cerevisiae* equivalent (Figure 3.3). It was concluded that like *RAD6* (Koken *et al.*, 1991), the *RAD23* gene has duplicated during evolution. Detailed description of the cloning of the *HHR23* genes is described in Chapter IV.

A

Score = 131 (58.6 bits), Expect = 1.3e-09, P = 1.3e-09

Identities = 27/66 (40%), Positives = 35/66 (53%), Frame = +3

```
RAD23:   333 GEAAAAGLGQGEGEGGSFQVDYTPEDDQAISRLCELGFERDLVIQVYFACDKNEEA 387
             G +++ G+++  +      ++ TP++ +AI RL  LGF   LVIQ YFAC KNE+
HHR23B:   72 GGGGSGGIAEAGSGHMXYIQVTPQEKEAIERLKALGFPEGLVIQAYFACXKNENL 246

RAD23:   388 AANILFSDHAD 398
             AAN L  ++ D
HHR23B:  247 AANXLLQQNFD 269
```

B

Score = 236 (105.8 bits), Expect = 3.2e-30, Sum P(2) = 3.2e-30

Identities = 48/65 (73%), Positives = 51/65 (78%), Frame = +3

```
HHR23A:  300 GEVGAIGEEAPQMNYIQVTPQEKEAIERLKALGFPESLVIQAYFACEKNENLAAN 354
             G  G      ++ M YIQVTPQEKEAIERLKALGFPE+LVIQAYFAC KNENLAAN
HHR23B:   81 GSGGIAEAGSGHMXYIQVTPQEKEAIERLKALGFPEGLVIQAYFACXKNENLAAN 260

HHR23A:  355 FLLSQNFDDE 364
              LL QNFD++
HHR23B:  261 XLLQQNFDED 275
```

Figure 3.3 Identification of the second human RAD23 equivalent, HHR23B by means of a TBLASTN search. The (A) panel shows the result with the yeast equivalent whereas the (B) panel shows the outcome for the first human gene product to be identified. The EST02187 Homo sapiens cDNA clone HFBCL77 with accession number M85669 is named HHR23B and was identified by sequencing of 2375 human brain genes (Adams *et al.*, 1992).

Sequence and structural homology of the ubiquitin-like domain of RAD23 with the ubiquitin protein.

At the N-terminal region, the RAD23 protein was found by computer assisted sequence analysis to contain a ubiquitin-like domain. The biological role of this domain is not known until now. Clustering of conserved residues in the close vicinity of the helix of the folded ubiquitin protein does suggest a similar function due to a similar structure. Deletion of the ubiquitin-like domain impairs the DNA repair function of the yeast RAD23 protein (Watkins *et al.*, 1993). Replacement of this domain by the authentic ubiquitin sequence restored the repair function (Watkins *et al.*, 1993). Figure 3.4 shows the homology between ubiquitin and the class of ubiquitin-like fusion proteins.

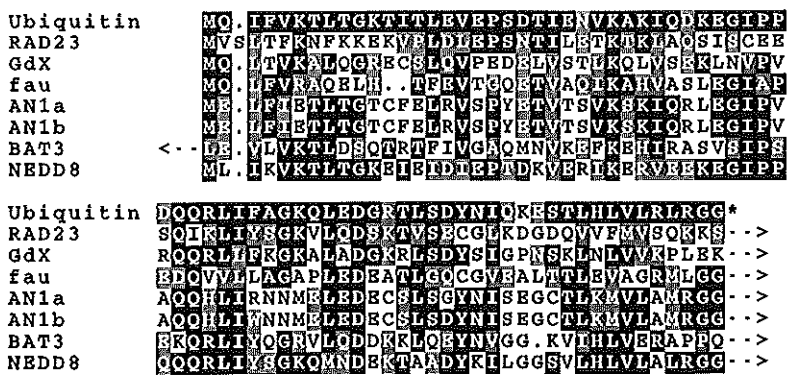


Figure 3.4 Alignment of different ubiquitin-like fusion proteins. Arrows indicate further sequence is present.

Computer algorithms (such as Predict protein) can be used to predict the secondary structure of a given protein sequence (Rost and Sander, 1993). Proteins are built up from combinations of secondary structure elements such as α helices and ß strands, which are connected by loop regions of various lengths and irregular shape. The interiors of protein molecules contain mainly hydrophobic side chains. The main chain in the interior is arranged in secondary structures to neutralize its polar atoms through hydrogen bonds. The formation of secondary structure in a local region of a polypeptide chain is to some extent determined by the primary structure. The similarity of the predicted secondary structure of the N-terminus of HHR23 proteins and the homologous ubiquitin protein is apparent from Figure 3.5.

```
Ubiquitin
         ....,....1....,....2....,....3....,....4....,....5....,....6....,....7....,....8....
AA       MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
PHD      EEEEEEE   EEEEEE     HHHHHHHH         HHHHEEEE  E                  EEEEEEEE
Rel      858888635717998742676157888752265699814542112231203675431125674542677664189

prH-     000000000000000000012467888765422200146654443321222212233331112211111211000
prE-     168888752147888763210000000000000000000012344442343210111111111112477776664l0
prL-     821111136741000125777521111123576788842211111254335676554446676653110001248 9
SUB      LEEEEEE.LL.EEEEE..LLL.HHHHHHH..LLLLLL..H..........LLL.....LLL.L..EEEEEE..LL

HHR23A
         ....,....1....,....2....,....3....,....4....,....5....,....6....,....7....,....8....
AA       MAVTITLKTLQQQTFKIRMEPDETVKVLKEKIEAEKGRDAFPVAGQKLIYAGKILSDDVPIRDYRIDEKNFVVVMVTKTKA->
PHD        EEEEEEEE   EEEEEE     HHHHHHHH           EEEEEE  EE              EEEEEEEE
Rel      9447788873663689872376516898875101169889972131455424131578753221534355377764237 9

prH-     011111111000000000000224788887644442111001345322222222221112233323232111111111100
prE-     026777877522378887531000000000000000000000011245665423541101122111111157777764310
prL-     862111011377621001357764110011244447888887532211113542367776444465556621001123589
SUB      L..EEEEEE.LL.EEEEE..LLL.HHHHHHH....LLLLLLL.....EE......LLLLL....L...LL.EEEEE...LL

HHR23B
         ....,....1....,....2....,....3....,....4....,....5....,....6....,....7....,....8....
AA       MQVTLKTLQQQTFKIDIDPEETVKALKEKIESEKGKDAFPVAGQKLIYAGKILNDDTALKEYKIDEKNFVVVMVTKPKA->
PHD       EEEEEEE   EEEEEE     HHHHHHHHHHHH         HHHEEEEEE EEE              EEEEEEE
Rel      837887735526898615854278888651112698899612413443221314666421114454646489986416 9

prH-     000000000000000000012357888876554521110014553222222222121234432222110000000000
prE-     167888752237888742000000000000000000000001113555423441101122221111126788876420 0
prL-     831111136752100147876410011124443788888743222111243235677543335566662100013578 8
SUB      L.EEEEE.LL.EEEEE.LLL..HHHHHHH....LLLLLLL.............LLL.......L.L..EEEEE..LLL
```

Figure 3.5 Prediction of the secondary structure of the ubiquitin protein and both HHR23 equivalents obtained with the EMBL PredictProtein Email server. This prediction method, designed for water-soluble globular proteins, indicates helix (H), strand (E) and loop (L) regions (Rost and Sander, (1993). The region indicated in bold represents the helix present in ubiquitin and its putative structural homologs.

When comparing homologous amino acid sequences from different species, it has been found that insertions and deletions of a few residues occur almost exclusively in the loop regions. During evolution, cores are much more stable than loops. The front cover of this thesis shows identical amino acids (in blue) and similar amino acids (red) between the N-terminal part of RAD23, its homologs and the ubiquitin protein. The space-filling representation shows the non-conserved areas looping out of the conserved core. In general, cores are much more stable than loops during evolution. Moreover, intron positions are often found at sites in structural genes that correspond to loop regions in the protein structure.

The function of every protein molecule depends on its three-dimensional structure, in turn determined by its amino acid sequence, in turn determined by the nucleotide sequence of the relevant gene. Knowledge of the three-dimensional structure facilitates the rational design of site-directed mutations in a protein. X-ray crystallography and NMR spectroscopy are the only ways to obtain such detailed structural information. Unfortunately, these techniques involve elaborate technical procedures, and many proteins fail to crystallize at all and/or cannot be obtained or dissolved in large enough quantities for these experiments (Eisenberg and Hill, 1989). The size of the protein is also a limiting factor for NMR (Wright, 1989).

In the absence of such experimental data, model-building on the basis of the known three dimensional structure of a homologous protein is at present the only reliable alternative method to obtain structural information. Protein domains with homologous amino acid sequences have similar three-dimensional structures, indicating the feasibility of model-building by sequence homology (Branden and Tooze, 1990). Comparisons of the tertiary structures of homologous proteins have shown that three-dimensional structures have been better conserved in evolution than protein primary sequences. Common tertiary structures were shown to have arisen independently several times during evolution (Kraulis, 1991). In such cases, the different primary sequences are deceptive: similarity of structure is not necessarily evidence of common evolutionary ancestry. Sequence alignments are often the limiting step for modelling, and an evaluation of the required homology between the sequence and the structure that will serve as the modelling template is very important (Branden and Tooze, 1990).

Three dimensional structure data can provide clues about candidate residues to mutate, knowing the position and context of residues in close vicinity. Subtle changes in a protein not resulting in drastic conformational changes should in many cases not affect protein complex formation. Mutation of conserved residues predicted to reside relatively closely positioned to each other and exposed on the surface of the protein are the best candidate positions for site-directed mutagenesis. Mutation of internal positioned residues will probably affect the structure of the protein, and clustered conserved amino acids might reflect an enzymatic active site present on the surface of the protein.

Differences between three-dimensional structures increase with decreasing sequence identity, and, accordingly, the accuracy of models built by homology decreases. The errors in a model built on the basis of a structure with 90% sequence identity may be as low as the errors in crystallographically-determined structures, except for a few individual side chains. A test case is illustrative: if a known protein structure is built from

another known structure, and 50% identity exists between the sequences, the error in the modelled coordinates can be as large as 1.5 Ångström (Branden and Tooze, 1990). However, local errors may be considerably larger. If the sequence identity is only around 25%, the alignment is the main bottleneck for model building by homology, and large errors are often observed. With less than 25% sequence identity the homology often remains undetected. A notable example of this is the structural similarity of protein G and ubiquitin (Kraulis, 1991). The deviations between a model built by homology and the "real" X-ray structure vary throughout the molecule, the largest deviations occurring in loops at the protein surface.

SNM1

The *S. cerevisiae SNM1* mutant is sensitive to the DNA crosslinking agent nitrogen mustard (HN2) (Ruhland *et al.*, 1981). Crosslink repair is of medical interest since the cytostatic potency of several anti-cancer drugs is based on the formation of such DNA lesions. The *SNM1* gene proved to be allelic to *PSO2*, and has been assigned by epistasis analysis to the RAD3 epistasis group, the latter known to be involved in DNA excision repair (Cassier-Chauvat and Moustacchi 1988; Henriques and Brendel, 1990; Richter *et al.*, 1992). *SNM1* is not an essential gene (Haase *et al.*, 1989). Many radiation-sensitive mutants are also cross-sensitive to chemicals of the mono- or bi-functional alkylating type.

A TBLASTN search with the *S. cerevisiae* SNM1 protein sequence revealed a human cDNA and an *Aspergillus niger* equivalent for this gene. An example of the output file of the computer search representing the list with hits for SNM1 is given in Figure 3.6. The *S. pombe* equivalent has been identified and partially sequenced by Morimyo (personal communication, 1995). The human gene has been mapped to the region 10q24 (Sijbers, unpubl. results). No mutants sensitive to crosslinking agents are known to be defective in this region of chromosome 10. Moreover, transfection of the gene to candidate rodent mutants did not correct their crosslink sensitivity (Sijbers and Zdzienicka, unpublished results).

Sequences producing High-scoring Segment Pairs:		Reading Frame	High Score	Smallest Sum Probability P(N)	N
emb|Z47071|SC9375	S.cerevisiae chromosome XIII cosm...	-2	3471	0.0	3
emb|X64004|SCSNM1	S.cerevisiae SNM1 (PSO2) gene	+1	3471	0.0	1
emb|X76917|SCSNM12TS	S.cerevisiae snm 1-2 ts gene, tem...	+2	1133	1.9e-150	1
dbj|D42045|HUMKIAAI	Human mRNA for ORF. >emb|D42045|H...	+1	106	3.2e-45	11
emb|A22974|A22974	A.niger bphA gene	+2	126	1.3e-30	5
gb|R13785|R13785	yf61b07.r1 Homo sapiens cDNA clon...	+2	106	0.00011	1
gb|U15801|CAU15801	Candida albicans chitinase (cht3)...	+2	84	0.25	1

Figure 3.6 P-values (Poisson) smaller than 10^{-1} represent significant homology (Altschul *et al.*, 1990). As can be concluded from the results table, the first six sequences represent significant homology, of which the first three are identical to the *S. cerevisiae* query sequence. The left columns of the table give the database containing the listed sequence file; emb (EMBL database), dbj (DNA Database of Japan) and gb (Genebank). The next column specifies the unique sequence identifier, the accession number. In addition to the name, a brief description, the reading frame and statistical values (P) and (N) are given. Sequences are sorted from most statistically significant (lowest P-value) to the least statistically significant (highest P-value).

Identification of a human DNA sequence with homology to the photoreactivating enzyme.

Most species are thought to possess more than one pathway to ensure the repair of toxic and mutagenic UV photoproducts. Prokaryotes and eukaryotes have evolved a number of pathways to either remove or cope with CPDs and (6-4) PPs. These pathways include NER, BER, and PHR (reviewed in Friedberg *et al.*, 1995).

Photoreactivation is a defence mechanism against ultraviolet light in living cells, in which photoreactivating enzyme (photolyase abbreviated as *PHR*) monomerizes UV-induced pyrimidine dimers. This section describes the identification of a human gene product highly homologous to the *E.coli* PHR enzyme. *PHR* genes are required for light-dependent photoreactivation and enhance light-independent excision repair of ultraviolet light (UV)-induced DNA damage (Sancar, 1990). This DNA repair mechanism was already present in the early period of evolution, as deduced from sequence similarities of the cloned photolyases of various organisms.

Two classes of photolyases can be distinguished based on sequence homology (reviewed by Yasui *et al.*, 1994). Class I members that have been identified are: *Halobacterium halobium*, *Streptomyces griseus*, *Escherichia coli*, *Salmonella typhimurium*, *Saccharomyces cerevisiae*, and the *Neurospora crassa* photolyase. Examples of class II photolyases are identified from *Drosophila melanogaster*, *Oryzias latipes* (Killifish), *Carassius auratus* (goldfish), *Potorous tridactylis* (rat kangaroo), and the archae-bacterium *Methanobacterium thermoautotrophicum.*

The *A.thaliana*, *S.alba*, and *Chlamydomonas reinhardtii* putative blue-light receptors have been shown to encode proteins with significant homology to classI photolyases (Ahmad and Cashmore, 1993; Malhotra *et al.*, 1995; Small *et al.*, 1995). This finding did shed a new light on the DNA photolyase/blue-light photoreceptor family, and suggests a new mechanism for signal transduction and opens the possibility that photolyases in addition to their role in repair might function as blue-light receptor.

The photolyase enzyme consists of an apoenzyme and two chromophores. All characterized photolyases contain (reduced) FAD, which is the photochemically active chromophore, and a second chromophore. This can either be of the MTHF (5,10-methenyltetrahydrofolate) type or 8-HDF (8-hydroxy-5-deazaflavin) type. The 1,5-dihydroflavin adenine dinucleotide ($FADH_2$) chromophore plays a central role in the photolysis reaction. The enzyme binds to dimer-containing DNA independently of light and upon absorbing a 300-500 nm photon, breaks the carbon-carbon bonds linking adjacent C5 atoms and C6 atoms in the cyclobutane ring of the dimer and thus restores

dipyrimidines in the DNA (Sancar, 1990). Both *A.thaliana* and *S.alba* blue-light receptors were shown to contain both $FADH_2$ and MTHF cofactors, however, no DNA repair activity in *E.coli* could be observed (Malhotra *et al.*, 1995).

Using the class I *Anacystis nidulans* photolyase protein sequence in a TBLASTN search, a human cDNA was identified, homologous to the class I enzymes.

Figure 3.7 depicts a partial alignment of the deduced amino acid sequence of the human cDNA equivalent to all cloned class I photolyases.

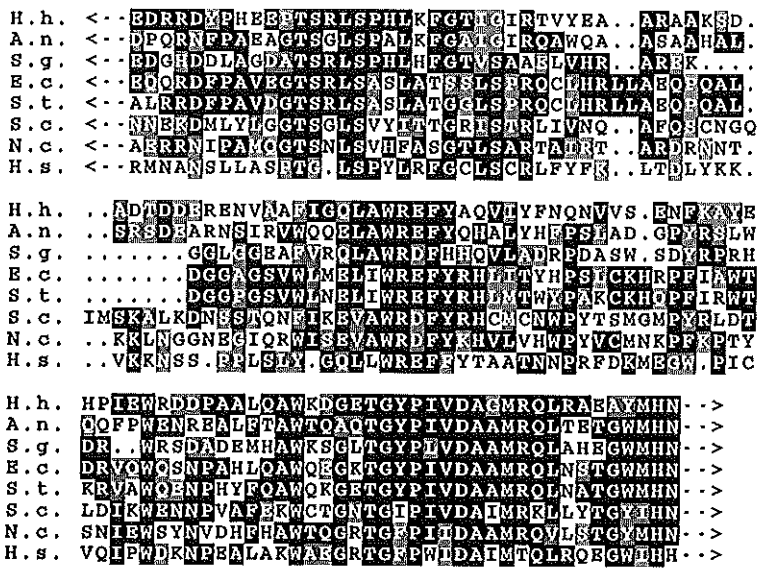


Figure 3.7 Partial alignment of Class I photolyases from different species. The amino acid sequences are given in the one letter code. Identical amino acids are presented by black boxes, whereas similar residues (A, S, T, and P; D, E, N, and Q; R, and K; I, L, M, and V; F, Y, and W) are given in gray boxes. Species are abbreviated as follows *Halobacterium halobium* (H.h.), *Anacystis nidulans* (A.n.), *Streptomyces griseus* (S.g.), *Escherichia coli* (E.c.), *Salmonella typhimurium* (S.t.), *Saccharomyces cerevisiae* (S.c.), *Neurospora crassa* (N.c.), and the *Homo sapiens* (H.s.) photolyase.

Alignment of different species photolyases enables the identification of common characteristics of photolyases in evolution and the determination of the conserved amino acid sequences essential for its enzymatic function. The high degree of sequence identity apparent from Figure 3.7 reflects an important common functional and/or structural domain and strongly suggests that the genes evolved from a common ancestral gene. The protein contains highly conserved tryptophane residues found in prokaryote and eukaryote PHR equivalents, present throughout the complete protein. However, these tryptophane residues are not part of highly conserved repeated units, known as WD-repeats as described by Neer *et al.* (1994) and Gutjahr *et al.* (1995). Recently, the three dimensional structure of the *E.coli* PHR enzyme has been resolved by X-ray-crystallography (Park *et al.*, 1995). Photolyases from prokaryotes, eukaryotes and archae-bacteria make essentially identical contacts with the DNA surrounding the dimer lesion (Park *et al.*, 1995). DNA photolyase acts specifically on the major UV photoproduct, *cis-syn* cyclobutyl pyrimidine dimers.

The existence of a photolyase restricted to the aplacental mammals was previously demonstrated by cloning of the gene from *Potorous tridactylis* (Yasui *et al.*, 1994) and by functional studies done before. However, the presence of photoreactivating enzyme in placental mammals is highly controversial (Li *et al.*, 1993). Sutherland and Bennett (1995) identified photolyase activity in human white blood cells, as follow up studies of very early claims for existance of PHR activity in a variety of human cells. However, others were unable to reproduce these findings (Sancar, 1993) and provided evidence for absence of photoreactivation in placental mammals. However, the BLAST search mentioned above identified a human cDNA clone with clear overall homology to class I photolyases (Figure 3.7). Enzymatic studies should reveal whether this is a bonafide photolyase. Since cells in UV-unexposed internal organs, such as testis and brain, of higher organisms express this structural homolog of photolyase, this light-dependent enzyme may bear a second function.

XPE

The third example of computer assisted sequence identification deals with the protein associated with xeroderma pigmentosum group E. Patients affected in this gene exhibit mild symptoms compared to other xeroderma pigmentosum complementation groups. Neurological abnormalities have not been observed in this subgroup of patients. Chu and Chang (1988) identified a nuclear factor that binds to DNA damaged by UV or cisplatin, that is absent in cells from complementation group E. The purification and characterization of this protein was reported by Hwang and Chu (1993) and Hwang *et al.* (1996). Moreover, the human 127 kD protein was identified and reported to show homology to a *Dictyostelium discoideum* (slime mold) gene product purified by Takao *et al.* (1993). Also Keeney *et al.* (1993) purified the human DNA damage-binding protein (DDB1), implicated in XPE, to near homogeneity from HeLa cells. DNA damage binding activity copurified with polypeptides of 124 and 48kD, two subunits respectively designated DDB1 and DDB2 of a heterodimeric protein. Keeney *et al.* (1994) microneedle-injected the purified human DDB protein complex into XPE cells and showed that DNA repair was stimulated to normal levels in those XPE strains that lacked the DDB activity but not in cells from other xeroderma pigmentosum groups. These data strongly indicate that defective DDB1 causes the repair defect in a subset of XPE patients, which in turn establishes a role for this activity in NER *in vivo*. However, mutations in DDB1 are not yet described that form the definite proof for this hypothesis. The gene encoding the large subunit (DDB1) is localized on human chromosome 11q12-q13 whereas the small subunit was assigned to the 11p11-p12 locus (Dualan *et al.*, 1995). The protein recognizes various forms of DNA damage, including cyclobutane pyrimidine dimers, 6-4 photoproducts, cisplatin adducts, and damage induced by nitrogen mustard (Payne and Chu, 1994; Reardon *et al.*, 1993).

The large subunit (DDB1) was used for a sequence comparison search and revealed both a yeast homolog and a second human gene. The TBLASTN result obtained with the human DDB1 gene product is given in Figure 3.8. A few observations make this a very significant match. First, the occurrence of six sequence segments in the correct strand and in the correct order. Second, conserved residues in this yeast sequence are also conserved in the two *C. elegans* equivalents that were identified in the same way. The yeast equivalent resides on yeast chromosome XIII (13L), and is being disrupted in order to see whether this causes a UV-sensitive phenotype (Lombaerts *et al.*, unpublished results). Surprisingly, gene disruption revealed a lethal phenotype, indicating the gene is essential in *S. cerevisiae*.

```
                Score = 119 (54.7 bits), Expect = 3.1e-06, Sum P(2) = 3.1e-06
                Identities = 31/120 (25%), Positives = 57/120 (47%), Frame = +2

HUMAN:  857 KLQTVAEKEVKGAVYSMVEFNGKLLSAINSTVRLYEWTTEKELRTECNHYNNIMA 911
            KL+ + +  V+     ++  F+G++L ++    +R+Y+ + +K LR   N +
HUMAN:  488 KLEFLHKTPVEEVPAAIAPFQGRVLIGVGKLLRVYDLGKKKLLRKCENKHIANYI 662

HUMAN:  912 LYLKTKGDFILVGDLMRSVLLLAYKPMEGNFEEIARDFNPNWMSAVEILDDDNFL 966
              ++T G  ++V+D+  S + + YK  E +    A D  P W+++  +LD D
HUMAN:  663 SGIQTIGHRVIVSDVQESFIWVRYKRNENQLIIFADDTYPRWVTTASLLDYDTVA 837

HUMAN:  967 GAENAFNLFV 976
            GA+   N+ V
HUMAN:  838 GADKFGNICV 847

                Score = 46 (21.1 bits), Expect = 3.1e-06, Sum P(2) = 3.1e-06
                Identities = 11/29 (37%), Positives = 13/29 (44%), Frame = +2

HUMAN: 1084 PATGFIDGDLIESFLDISRPKMQEVVANL 1112
            P    IDGDL E F  +   K ++V  +L
HUMAN: 1187 PVKNVIDGDLCEQFNSMEPNKQKNVSEEL 1273
```

Figure 3.8 The top line indicates the query sequence human DDB1 protein whereas the bottom line shows the human cDNA encoding a second human homolog of the large subunit.

The newly identified human homolog tentatively designated DDB1-like was a complete randomly sequenced cDNA clone identified by Nomura and coworkers for genome sequencing purpose. Cytogenetic analysis has shown the second human homolog of XPE to reside on chromosome 16q22-16q23 as shown in figure 3.9. The duplication of the DDB1 subunit might explain the relative mild symptoms of the XPE patients. It is very well possible that both human structural DDB1 homologs are functionally redundant (van der Spek, unpublished results).

To identify whether the *DDB2* gene has also duplicated during evolution, another TBLASTN search was performed. Although no DDB2-like gene was found it was discovered that the DDB2 protein shares significant homology to the CSA protein. The recently cloned Cockayne syndrome group A gene encodes a WD-repeat protein and is claimed to interact with CSB protein and a subunit of RNA polymerase II TFIIH (Henning *et al.*, 1995). Both DDB2 and CSA proteins are made up of highly conserved repeating units ending with Trp-Asp (WD). Thusfar, all WD-proteins described are regulatory proteins involved in different processes such as cell division, cell-fate determination, gene transcription, transmembrane signalling, and mRNA modification. Figure 3.10 shows the overall homology for the human DDB2 and CSA proteins.

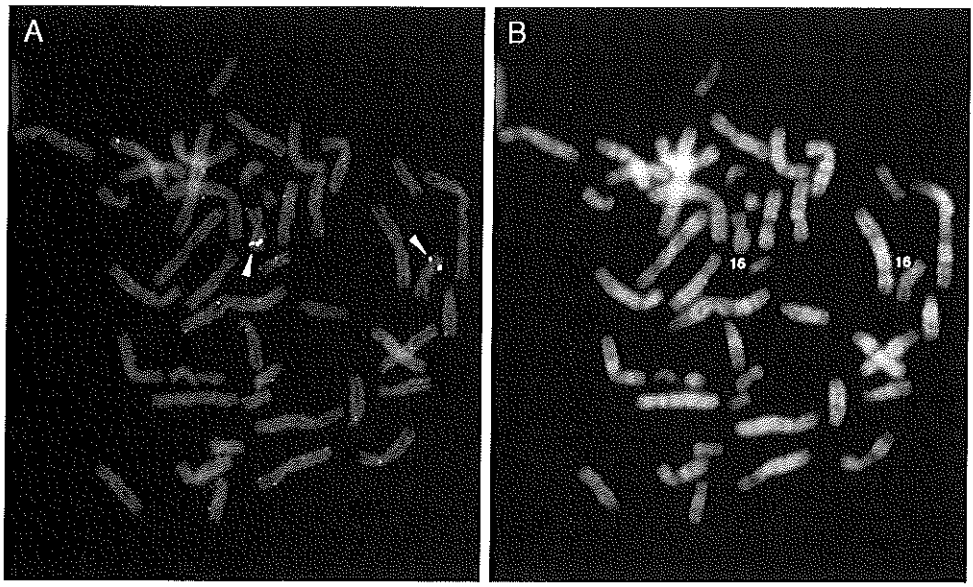


Figure 3.9 *In situ* hybridization of metaphase chromosomes with biotinylated *DDB1-like* cDNA probe. Based on morphology and banding pattern specific signal was observed on 16q22-16q23, panel (A). The Hoechst banding of the same metaphase is given on the right (B) panel.

```
DDB2 MASKKRP.ETQKTSETVLRPRNKRER..SPLELRPRAKKLCAKGEGPSRRCDSDCLWVGT
CSA  MLGFLSRRQTGLEDPLRLR.RAESTRRVLGLELKKD.RDVERIHGG...........GT

DDB2 AGPQILPPCRSIVRTLHQHKLGRASWPSVQQGLQQSFLHTLDSYRILQKAAPFDRRATSL
CSA  NTLDIEP.....VEGRYM..LSGGS.................DQVIVLYDLENSSRQ....

DDB2 AWHPTHPSTVAVGSKGGDIMLWNFGIKDKPTFIKGIGAGGSITGLKFNPLNTNQFYASSM
CSA  ....SYYTCKAVCSIG..........RDHPDVHR.....YSVETVQWYPHDTGMFTSSSF

DDB2 EGTTRLQDFKG.NILRVFASSDTINIWFCSLDVSASSRMVVTGDNVGNVILLNM.DGKEL
CSA  DKTLKVWDTNTLQTADVFNFEETVYSHHMS.PVSTKHCIVAVGTRGPKVQLCDLKSGSCS

DDB2 WNLRMHKKKVTHVALNPCCDWFLATASVDQTVKLWDIR.........QVRGKASFLYSL
CSA  HILQGHRQEILAVSWSPRYDYILATASADSRVKLWDVRRASGCLITLDQHNGKKSQAVES

DDB2 ...PHRHPVNAACFSPDGARLLT..TDQKSEIRVYSASQWDCPF...RTIPHR....HFQ
CSA  ANTAHNGKVNGLCFTSDGLHLLTVGTDNRMRLWNSSNGE.NTLVNYGKVCNNSKKGLKFT

DDB2 ...HLTPIKAAWHPRYNLIVV.GRYPDPNFKSCTPYELRTID..VFDGNSGKMMCQLYDP
CSA  VSCGCGS.EFVRVPYGSTIAVYTVYSGEQI.LMLKGHYKTVDCCVFQSNFQELYSGSRDC

DDB2 P.SSGISSLNEFNPMGD..TLASAV.GYHILIWS..QEERTRK*
CSA  NILAWVPSLYEPVPDDDETTTKSQLNPAFEDAWSSSDEEG*
```

Figure 3.10 Sequence alignment of the DDB2 subunit and the Cockayne syndrome A protein. Identical amino acid residues are presented in black boxes, whereas similar residues (A, S, T, P, and G; D, E, N, and Q; R, K, and H; I, L, V, and M; F, Y, and W) are given in gray boxes.

RAD9

As indicated in Chapter II, cells of the *S. cerevisiae rad9* mutant progress through mitosis without arresting the cell cycle to repair clastogenic insults, and therefore die as a result of lethal chromosome damage. Detailed genetic analysis has revealed that the *rad9* mutant is a checkpoint control mutant (Hartwell and Weinert, 1989). A human gene product, that shares limited homology to two parts of the yeast RAD9 protein was identified as shown in Figure 3.10. Until now, no RAD9 homologs of other species have been identified.

```
                Score = 63 (28.4 bits), Expect = 1.4, Sum P(2) = 0.74
                Identities = 17/60 (28%), Positives = 26/60 (43%), Frame = +2

YEAST:   116 PVMPTSLRMTRSATQAAKQVPRTVSSTTARKPVTRAANENEPEGKVPSKGRPAKN 175
             PV  ++        TQ           + T   PVT +  + EP   VP  + P+KN
HUMAN:   444 PVFHSTGQTEEIKTQIINSPEQNALNATFETPVTLSRINFEPILEVPETSSPSKN 498

YEAST:   211 VETKP 175
               +KP
HUMAN:   499 TMSKP 503

                Score = 59 (26.6 bits), Expect = 1.4, Sum P(2) = 0.74
                Identities = 14/60 (23%), Positives = 32/60 (53%), Frame = +2

YEAST:   654 DDINTNKKEGISDVVEGMELNSSITSQDVLMSSPEKNTASQNSILEEGETKISQS 708
             ++++T K  +I   VE  +  + +  Q+ L +SP K    +++++E   +K +
HUMAN:   586 NELDTKKESTIMSEVELTQELPEVEEQQDLQTSPKKLVVEEETLMEIKKSKGNSL 640

YEAST:   709 ELFDN 713
             +L D+
HUMAN:   641 QLHDD 645
```

Figure 3.10 Alignment of the two homologous domains shared by the yeast RAD9 protein and a human protein identified by sequence comparison. The randomly sampled full-length human cDNA clone, derived from a human immature myeloid cell line KG-1, was sequenced and deposited in the DNA database of Japan with accession number D13633 (Nomura *et al.*, 1994).

The human cDNA was cloned and mapped to human chromosome 14q21-14q22 shown in Figure 3.11. However, no candidate mutants are known to be affected in this region except for the X-ray sensitive *irs1SF* mutant. The gene defective in this mutant has recently been cloned and is distinct from this gene (Thompson, personal communication; 1995).

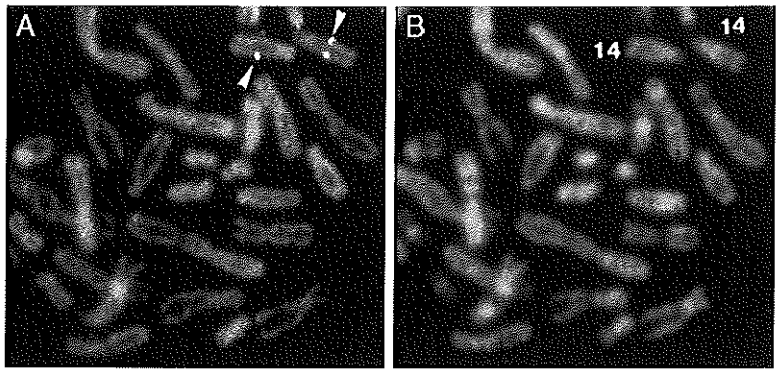


Figure 3.11 *In situ* hybridization of metaphase chromosomes with the biotinylated human cDNA probe. The arrowheads point to the region on chromosome 14q22-14q23 with a specific signal. (A) The *in situ* hybridization result. (B) The Hoechst banding of the same metaphase.

X-ray inducibility was checked and found to be not a conserved feature between *S. cerevisiae rad9* and the human sequence (van der Spek, unpublished results). The same holds for a number of other genes that are damage-induced in *S. cerevisiae* but not in mammals such as RAD6, RAD23 and RAD54 as discussed in Chapter 2.

The human cDNA is encoded by a 2.4 kb messenger RNA expressed specifically in testis, thymus and colon tissue (Nomura *et al.*, 1994). Interestingly, elevated RNA expression was observed during S phase of the mitotic cell cycle (van der Spek, unpublished results). As with some other cell cycle regulatory genes, evolutionary diversification could be a likely explanation for the fact that RAD9 equivalents in other species have not to date been identifiable using degenerated primers or low stringency hybridization. Additionally, these two regions of homology were not found to be present in other proteins present in the database, excluding this is a general structural element present in many other proteins. Recently, the human gene product was found to have a second human homolog, that was identified as a partially sequenced EST with accession number T07912. Whether the sequence homology extends in the not yet sequenced remainder of this gene remains to be clarified. Due to the very limited sequence homology to the *S. cerevisiae* RAD9 protein, it will be difficult to prove that these human gene products are functional equivalents of the yeast *RAD9* gene. However, the stage-specific expression during the mitotic cell cycle in combination with the presence of the two conserved domains suggests a function in cell cycle control for the regions shown in Figure 3.10.

Detection of loosely defined functional domains

In contrast to the approach of sequence comparison against a complete database to identify homologous sequences, particular DNA or protein sequences can be checked for the presence of precisely- or loosely-defined functional motifs or patterns. An example of a routine suitable for identification of specific well-defined domains is Prosite (Bairoch, 1992). The cloning and sequencing of a gene permits detailed analysis of the function. The protein sequences derived from the different species have to be aligned optimally to determine conserved residues. From these alignments, it is sometimes possible to recognize or deduce patterns or motifs representing a particular function. This approach can provide crucial clues to the (putative) function of the gene product under investigation. Newly deduced domains can be used to compare to complete databases, to see whether they represent a common fold (domain) or a rare motif. The Prosite program, although powerful, has some limitations. Notably, the number of motifs in the dataset cannot be increased by the user. In other words, Prosite cannot be used to scan for amino acid sequences of interest to the operator. Likewise, rigid definition of motifs and domains in the Prosite routine restricts detection sensitivity.

These disadvantages have been overcome by us with a program described here. This program allows the identification of more diverged and therefore loosely-defined motifs that are clustered in a particular order with variable inter-motif intervals. Figure 3.12 shows the way in which motifs or domains can be defined, together with an example of the identification of the 'helicase' domains present in one of the class of SNF2-type helicases, the ERCC6 protein (Troelstra *et al.*, 1992).

```
Consensus Domain I                                  Search percentage: 80%
+          G   G K T       +
                   S
1 . . . 5 . . . .10 . . . .15 . . . .20 . . . .25 . . . .30

Consensus Domain Ia                                 Search percentage: 80%
+ +      0    +     0 +
1 . . . 5 . . . .10 . . . .15 . . . .20 . . . .25 . . . .30

Consensus Domain II                                 Search percentage: 80%
0      + + + + D E  H
                    D
1 . . . 5 . . . .10 . . . .15 . . . .20 . . . .25 . . . .30

Consensus Domain III                                Search percentage: 80%
+    + T A T        0          +    0
       S G S
1 . . . 5 . . . .10 . . . .15 . . . .20 . . . .25 . . . .30

Consensus Domain IV                                 Search percentage: 80%
F      S    0      0
Y      T
1 . . . 5 . . . .10 . . . .15 . . . .20 . . . .25 . . . .30

Consensus Domain V                                  Search percentage: 80%
0      0      + + +    T 0 +         G   0 +         0    + +
                       S             S
1 . . . 5 . . . .10 . . . .15 . . . .20 . . . .25 . . . .30

Consensus Domain VI                                 Search percentage: 80%
0      0      Q      G R      R
A      H      H
1 . . . 5 . . . .10 . . . .15 . . . .20 . . . .25 . . . .30
```

Figure 3.12A

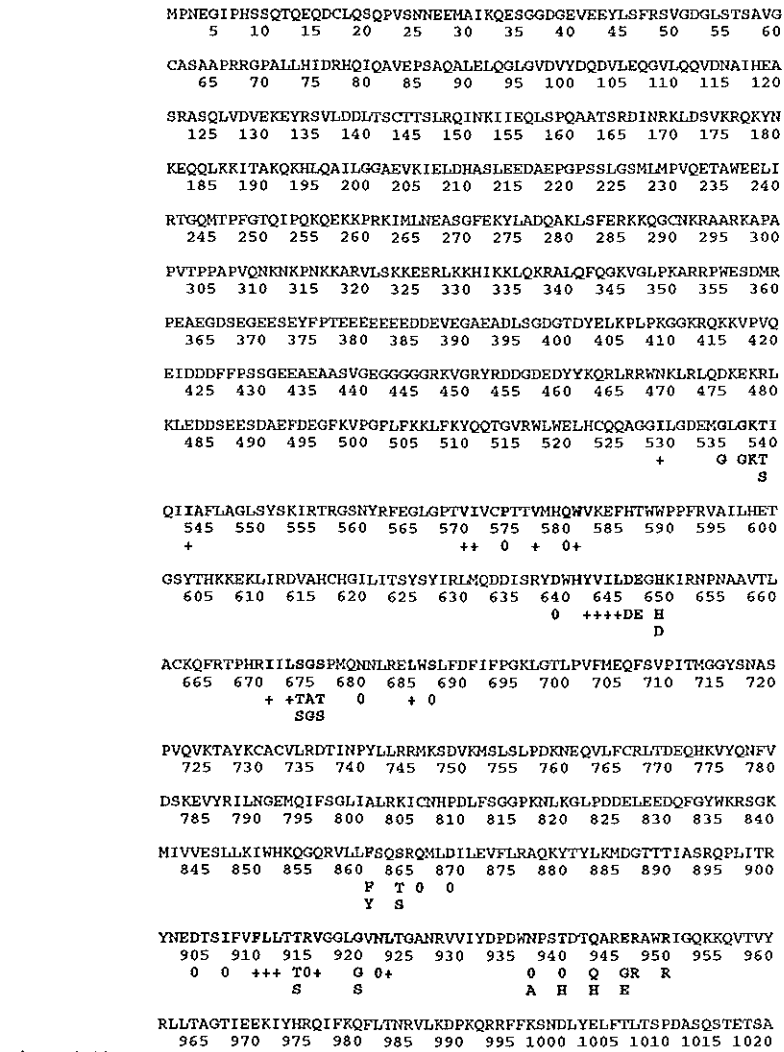


Figure 3.12B

Figure 3.12 (A) Definition of the consensus sequences of the DNA and RNA helicase domains postulated by Gorbalenya *et al.* (1989); + represents hydrophobic (V, L, I, M, F, Y, W) amino acid residues, whereas 0 indicates charged or polar residues (S, T, D, E, N, Q, K, R).

(B) Example of program output indicating the ERCC6 helicase region that contains the seven consensus motifs identified among two superfamilies of DNA and RNA helicases. The consensus sequences are given below the corresponding domains.

The routine has been designed using the presumed and putative helicase family defined by Gorbalenya *et al.* (1989) as a model. In the ERCC3 protein, an array of seven different domains occurs with variable spacing. Thus the program was made highly flexible and allows for example variations in the number of spacer amino acids between the domains. Functionality of the routine was shown by the identification of the helicase domains in the CSB/ERCC6 protein (described by Troelstra *et al.*, 1992).

The routine has an option to check any sequence of interest for numerous published consensus sequences of loosely defined motifs and domains in a single run. The advantages of this program can therefore be summarized as follows:

1. The possibility to search for domains with loosely defined consensus sequences in which for each position within a domain one or more (or even a class of amino acids) can be defined (Figure 3.12A).

2. The possibility to identify an array of different domains that belong together but which have variable inter-domain spacing in the sequence of interest. For example, linked domains occur in RNA polymerases, kinases and DNA polymerases (Figure 3.12B).

Information concerning (frequently) occurring patterns, motifs, or common structures can be obtained by dividing the sequence into parts of about 50 amino acids. These short 'query strings' can be compared again to the Genbank and EMBL databases. Isolated regions of similarity, corresponding for instance to structural motifs or active sites, can be detected in this manner. Alternatively, checking the Prosite database can provide evidence for the presence of putative functional domains. Finally, a self comparison can be performed to detect the presence of internal homology within the sequence of interest.

After performing BLAST searches and detection of significant homology, detailed analysis of homologous DNA and protein sequences using programs such as the NCBI Macaw algorithm can be highly informative. This multiple sequence alignment program compares limited amounts of homologous sequences (DNA or protein) and is meant for performing self comparisons (Schuler *et al.*, 1991). This allows the user to locate conserved areas of aligned sequence segments of a variety of species in a single run. To indicate the significance of the matches found, an internal statistics option is available (Karlin and Altschul, 1990). Mathematical analysis can indicate which similarities are unlikely to have arisen by chance and therefore merit special attention. In this way, self comparison of *Drosophila* XPC protein using the NCBI Macaw algorithm revealed that an internal duplication of approximately 50 amino acids starting on residue 514 occurred three times as shown in Figure 3.13. The duplication, unique for *Drosophila*, does not

show any homology to particular domains or motifs previously identified (van der Spek, unpublished results).

```
DmXPC 603   SLSSKLVEKSKHQKAHTSSKSDTSFDEKPSTSSSSKCLKEEYSELGLSK  651
            +LSSKLV KSK Q + +S KSDTSF+E PSTSSSSK LKEE + L+ SK
DmXPC 657   TLSSKLVLKSKNQSSFSSNKSDTSFEENPSTSSSSKSLKEETAKLSSSK  705
```

Figure 3.13 Alignment of the duplicated region unique for the *Drosophila* XPC protein. The amino acid sequence is given in the one letter code. Identical amino acids are presented by black boxes, whereas similar residues (A, S, T, P; D, E, N, Q; R, K; I, L, V, M; F, Y,W) are given in gray boxes. The *Drosophila melanogaster* XPC sequence with accession number Z28622 was described previously by Henning *et al.*, (1994).

Significance and implications

Most of the database sequence comparison programs produce an ordered list of imperfectly matching database similarities, but none of them need to have any biological significance. Therefore, correctly interpreting data obtained by sequence alignment is important for identifying sequences of interest for experimental study. A region of high similarity shared by two or more sequences might be evidence of evolutionary homology or of common function. Statistical methods for evaluating sequence patterns can be based on theoretical models or on permutation reconstructions of the observed data [Doolittle, 1981; Karlin *et al.*, 1983; Fitch *et al.*, 1983; Altschul and Erickson, 1985; for a review on patterns in DNA and amino acid sequences and their statistical significance, see Karlin and Altschul (1990). For a general discussion of the meaning of scoring matrices, see Altschul (1991)].

Many of the known human proteins may be members of families that include several genes with similar sequences, and presumably similar functions (Adams *et al.*, 1993). Striking examples of this are the classes of RNA and DNA helicases (Gorbalenya *et al.*, 1989; Koonin, 1992; Thommes and Hubscher, 1992), and ubiquitin-conjugating enzymes (Qin *et al.*, 1991).

Whenever "significant" sequence homology appears in comparisons between two species, the respective amino acid residues should be conserved in corresponding protein sequences of other species in order to be meaningful. All homologous segments should appear unidirectionally and in the correct order. The less frameshifts which occur, the more likely the result is of relevance. This is dependent on the quality of the EST sequences. Reciprocal comparison to the database should give similar results. Complementary DNA libraries may contain multiple truncated clones of the same gene, as well as chimeric clones in an artificial manner derived from different genes.

The statistical values reported by the BLAST programs are dependent on numerous factors, including the scoring scheme employed, the residue composition of the query sequence, an assumed residue composition for a typical database sequence, and the query and database lengths. The scoring scheme is a matrix for similarity measures defining all possible pairs of residues. Identities and conservative replacements have positive scores, while unlikely replacements have negative scores. Local alignments are given a score, and computed as the sum of the scores for aligned pairs of residues, making allowance for gaps. More powerful routines are neccessary to detect weak similarities between strongly diverged sequences. The price of increased detection sensitivity is greatly increased

computation time (Voght and Argos 1992). The BLAST default settings are therefore a compromise between specificity and computation time. The Poisson P-value (probability in the range 0-1) is a function of its expected frequency of occurrence and the number of high scoring segment pairs (N) observed against the same database sequence with scores at least as high, as described by Karlin and Altschul (1990). Sequence similarities identified by the BLAST programs are considered statistically significant when the poisson p-value is < 0.01 see (Figure 3.6). The p-value is the probability of as high a score occurring by chance amongst the residues in the query sequence and in the database. However, statistical significance does not necessarily mean functional similarity. Matches reported by the BLAST algorithm can also indicate the presence of a conserved domain or motif, or simply a common protein structure pattern. Furthermore, diverged sequences may not be detected using BLAST or FASTA sequence comparison software.

In conclusion

Improvements in DNA sequencing technologies have facilitated rapid progress in sequencing expressed genes in diverse organisms. The human EST sequence and mapping approach will provide a new resource for the analysis of the genome and for human gene discovery. The EST sequence data have been deposited in Genbank and the Genome DataBase linked to the OMIM database, all accessible via world wide web (Harper, 1995). All clones are available from the American Type Culture Collection and via FTP (file transfer protocol) using the accession number of the particular clone.

Within a few years, detailed information about gene expression, such as its timing and location, will be available for the estimated 50,000 to 100,000 human genes. However, it is likely that the last part of the human genome project will become a difficult task because of the barrier of identifying lowly expressed genes. Libraries from more specific cell and tissue types are necessary to track down rare transcripts expressed only in specialized tissues. Sequence comparisons, like the ones described in this chapter, can be of significant help in identifying the less conserved homologous genes in related species. Insight into coordinated gene expression during development and fundamental problems such as cancer will rapidly expand (Velculescu *et al.*, 1995). Access to these data should have enormous impact on diagnostics, gene therapy and eventually, in general clinical medicine. To greatly expediate gene cloning by linkage analysis, it will be possible to more rapidly and precisely identify the location of the gene(s) responsible for an inherited disease or trait.

References:

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, O., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991). Complementary DNA sequencing:expressed sequence tags and human genome project. *Science* 252: 1651-1656.

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992). 3,400 New expressed sequence tags identify diversity of transcripts in human brain. *Nature* 355: 632-634.

Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C. (1993). 3,400 New expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.* 4: 256-267.

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., FitzHugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle Jr., P.S., Kelley, J.M., Klimek, K.M., Kelley, J.C., Liu, L., Marmaros, S.M., Merrick, J.M., Moreno-Palanques, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Barthlow, R., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.-J., Dimke, D., Feng, P. Ferrie, A., Fisher, C., Hastings, G.A., He, W.W., Hu, J.-S., Huddleston, K.A., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Ji, H., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377: 3-17.

Ahmad, M. and Cashmore, A.R. (1993). *Hy4* gene of *A.thaliana* encodes a protein with characteristics of a blue-light photoreceptor. *Nature* 366: 162-166.

Altschul, S.F. (1991). Amino acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.* 219: 555-565.

Altschul, S.F. and Erickson, B.W. (1985). Significance of nucleotide sequence alignments: a method for random sequence permutation that preserves dinucleotide and codon usage. *Mol. Biol. Evol.* 2: 526-538.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Molec. Biol.* 215: 403-410.

Altschul, S.F., Boguski, M.S., Gish, W. and Wooton, J.C. (1994). Issues in searching molecular sequence databases. *Nature Genetics* 6: 119-129.

Bairoch, A. (1992). Prosite: a dictionary of sites and patterns in proteins. *Nucl. Acids Res.* 20: 2013-2018.

Benson, D.A., Boguski, M., Lipman, D.J. and Ostell, J. (1994). Genbank. *Nucl. Acids Res.* 17: 3441-3444.

Berry, R., Stevens, T.J., Walter, N.A.R., Wilcox, A.S., Rubano, T., Hopkins, J.A., Weber, J., Goold, R., Soares, M.B. and Sikela, J.M. (1995). Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nature Genetics* 10: 415-423.

Boguski, M.S. (1995). The turning point in genome research. *Trends in Biochem. Sci* 20: 295-296.

Boguski, M.S. and Schuler G.D. (1995). ESTablishing a human transcript map. *Nature Genetics* 10: 369-371.

Branden, C. and Tooze J. (1991). Introduction in protein structure. Garland publishing, Inc. New York and London.

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., Aparicio, S. (1993). Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature* 366: 265-268.

Cassier-Chauvat, C. and Moustacchi, E. (1988). Allelism between *pso1-1* and *rev3-1* mutant and between *pso2-1* and *snm1* mutants *Saccharomyces cerevisiae. Curr. Genet.* 13: 37-40.

Chu, G. and Chang, E. (1988). Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science* 242: 564-567.

Chu, G. and Chang, E. (1990). Cisplatin-resistant cells express increased levels of a factor that recognizes damaged DNA. *Proc. Natl. Acad. Sci. USA* 87: 3324-3327.

Doolittle, R.F. (1981). Similar amino acid sequences: chance or common ancestry? *Science* 214: 149-159.

Dualan, R., Brody, T., Keeney, S., Nichols, A.F., Admon, A. and Linn, S. (1995). Chromosomal localization and cDNA cloning of the genes (DDB1 and DDB2) for the p127 and p48 subunits of a human damage-specific DNA binding protein. *Genomics* 29: 62-69.

Durkin, A.S., Nierman, W.C., Zoghbi, H., Jones, C., Kozak, C.A. and Maglott, D.R. (1994). Chromosome assignment of human brain expressed sequence tags (ESTs) by analyzing fluorescently labeled PCR products from hybrid cell panels. *Cytogenet. Cell Genet.* 65: 68-91.

Eichbaum, Q.G., Beatty, D.W. and Parker, M.I. (1991). H.sapiens sequence isolated by cross-reactivity of ARF sera with heart cDNA library. *Unpublished results.*

Eisenberg, D. and Hill, C.P. (1989). Protein crystallography: more surprises ahead. *Trends Biochem. Sci.* 14: 260-264.

Fleischmann, R.D., Adams M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenny, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.l., McDonald, L.A., Small, K.V., Fraser, C.M., O'Smith, H.O. and Venter, J.C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496-512.

Fitch, W.M. (1983). Random sequences. *J. Mol. Biol.* 163: 171-176.

Fraser, C.M., Gocayne, J.D., Owen, W., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchinson III, C.A., Venter, J.C. (1995). The minimal gene complement of *Mycoplasma genitalium. Science* 270: 397-403.

Friedberg, E.C., Walker, G.C. and Siede, W. (1995). DNA repair and mutagenesis. ASM Press

Washington, D.C.

Gieser, L. and Swaroop, A. (1992). Expressed sequence tags and chromosomal localization of cDNA clones from a subtracted retinal pigment epithelium library. *Genomics* 13: 873-876.

Goffeau, A. (1995). Life with 482 genes. *Science* 270: 445-446.

Van Gool, A.J. Verhage, R., Swagemakers, S.M.A., van de Putte, P., Brouwer, J. Troelstra, C., Bootsma, D. and Hoeijmakers, J.H.J. (1994). *RAD26* , the functional *S.cerevisiae* homolog of the Cockayne syndrome B gene *ERCC6. EMBO J.* 13: 5361-5369.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucl. Acids Res.* 17: 4713-4730.

Gutjahr, T., Frei, E., Spicer, C., Baumgartner, S., White, R.A.H. and Noll, M. (1995). The polycomb-group gene, extra sex combs, encodes a nuclear member of the WD-40 repeat family. *EMBO J.* 14: 4296-4306.

Haase, E., Riehl, D., Mack, M. and Brendel, M. (1989). Molecular cloning of SNM1, a yeast gene responsible for a specific step in the repair of crosslinked DNA. *Mol. Gen. Genet.* 218: 64-71.

Harper, R. (1995). World wide web resources for the biologist. *Trend in Genetics* 11: 223-228.

Hartwell, L. and Weinert, T. (1989). Checkpoints: controls that ensure the order of cell cycle events. *Science* 246: 629-634.

Hwang, B.J. and Chu, G. (1993). Purification and characterization of a human protein that binds to damaged DNA. *Biochem.* 32: 1657-1666.

Hwang, B.J., Liao, J.C. and Chu, G. (1996). Isolation of a cDNA encoding a UV-damaged DNA binding factor defective in xeroderma pigmentosum group E cells. *Mut. Res.* 362: 105-117.

Henning, K.A., Li, L., Iyer, N., McDaniel, L.D., Reagan, M.S., Legerski, R., Schultz, R.A., Stefanini, M., Lehmann, A.R., Mayne, L.V. and Friedberg, E.C. (1995). The Cockayne syndrome group A gene encodes a WD-repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. *Cell* 82: 555-564.

Henriques, J.A.P. and Brendel. (1990). The role of *PSO* and *SNM* genes in DNA repair of the yeast *Saccharomyces cerevisiae. Curr. Genet.* 18: 387-393.

Karlin, S., Ghandour, G., Ost, F., Tavaré, S. and Korn, L.J. (1983). New approaches for computer analysis of nucleic acid sequences. *Proc. Natl. Acad. Sci. USA* 80: 5660-5664.

Karlin, S. and Altschul, S.F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* 87: 2264-2268.

Keeney, S., Chang, G.J. and Linn, S. (1993). Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E. *J. Biol. Chem.* 268: 21293-21300.

Keeney, S., Eker, A.P.M., Brody, T., Vermeulen, W., Bootsma, D., Hoeijmakers, J.H.J. and Linn, S. (1994). Correction of the DNA repair defect in xeroderma pigmentosum group E by injection of a DNA damage-binding protein. *Proc. Nat. Acad. Sci. USA* 91: 4053-4056.

Khan, A.S., Wilcox, A.S., Polymeropoulos, M.H., Hopkins, J.A., Stevens, T.J., Robinson, M., Orpana, A.K. and Sikela, J.M. (1992). Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature Genet.* 2: 180-195.

Koken, M., Reynolds, P., Jaspers-Dekker, I., Prakash, L., Prakash, S., Bootsma, D. and

Hoeijmakers, J. (1991). Structural and functional conservation of two human homologs of the yeast DNA repair gene *RAD6*. *Proc. Natl. Acad. Sci. USA* 88: 8865-8869.

Koonin, E.V. (1992). A new group of putative RNA helicases. *Trends in Biochem Sci.* 17: 495-497.

Kraulis, P.J. (1991). Similarity of protein G and ubiquitin. *Science* 254: 581-582.

Li, Y.F., Kim, S.T., Sancar, A. (1993). Evidence for lack of DNA photoreactivating enzyme in humans. *Proc. Natl. Acad. Sci. USA* 90: 4389-4393.

Malhotra, K., Kim, S, Batschauer, A., Dawut, L. and Sancar, A. (1995). Putative Blue-light photoreceptors from *Arabidopsis thaliana* and *Sinapsis alba* with a high degree of sequence homology to DNA photolyase contain the two photolyase cofactors but lack DNA repair activity. *Biochemistry* 34: 6892-6899.

Neer, E.J., Schmidt, C.J., Nambudripad, R. and Smith, T.F. (1994). The ancient regulatory-protein family of WD-repeat proteins *Nature* 371: 297-300.

Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K. and Tabata, S. (1994). Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA40) Deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA research* 1: 27-35.

Park, H.W., Kim, S.-T., Sancar, A., Deisenhofer, J. (1995). Crystal structure of DNA photolyase from *Escherichia coli*. *Science* 268: 1866-1872.

Payne, A. and Chu, G. (1994). Xeroderma pigmentosum group E binding factor recognizes a broad spectrum of DNA damage. *Mut. Res.* 310: 89-102.

Polymeropoulos, M.H., Xiao, H., Glodek, A., Gorski, M., Adams, M.D., Moreno, R.F., Fitzgerald M.G., Venter, J.C. and Merril, C.R. (1992). Chromosomal assignment of 46 brain cDNA's. *Genomics* 12: 492-496.

Qin, S., Nakajima, B., Nomura, M. and Arfin, S.M. (1991). Cloning and characterization of a Saccharomyces cerevisiae gene encoding a new member of the ubiquitin-conjugating protein family. *J. Biol. Chem.* 266: 15549-15554.

Reardon, J.T., Nichols, A.F., Keeney, S., Smith, C.A., Taylor, J.S., Linn, S. and Sancar, A. (1993). Comparative analysis of binding of human damaged DNA-binding protein (XPE) and Escherichia coli damage recognition protein (UvrA) to the major ultraviolet photoproducts: T[*c,s*]T, T[*t,s*]T, T[6-4]T, and T[Dewar]T' *J. Biol. Chem.* 28: 21301-21308.

Richter, D., Niegemann, E. and Brendel, M. (1992). Molecular structure of the DNA crosslink repair gene *SNM1*(*PSO2*) of the yeast *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 231: 194-200.

Rost, B. and Sander, C. (1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl. Acad. Sci. USA* 90 : 7558-7562.

Ruhland, A., Haase, E., Siede W. and Brendel, M. (1981). Isolation of yeast mutants sensitive to the bifunctional alkylating agent nitrogen mustard. *Mol. Gen. Genet.* 181: 346-351.

Sancar, G.B. (1990). DNA photolyases: Physical properties, action mechanism, and roles in dark repair. *Mut. Res.* 236: 147-160.

Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991). A workbench for multiple alignment construction and analysis. *Proteins* 9: 180-190.

Small, G.D., Min, B. and Lefebvre, P.A. (1995). Characterization of a *Chlamydomonas reinhardtii* gene

encoding a protein of the DNA photolyase/blue-light photoreceptor family. *Plant Mol. Biol.* 28: 443-454.

Soares, M.B., de Fatima Bonaldo, M., Jelene, P., Su, L., Lawton, L. and Erfstratiadis, A. (1994). Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. SCI. USA* 91: 9228-9232.

Sutherland, B.M. and Bennett, P.V. (1995). Human white blood cells contain cyclobutyl pyrimidine dimer photolyase. *Proc. Natl. Acad. Sci. USA* 92: 9732-9736.

Takao, M., Abramic, M., Moos, M., Otrin, V.R., Wootton, J.C., McLenigan, M., Levine, A.S. and Protic, M. (1993). A 127 kDa component of a UV-damaged DNA-binding complex, which is defective in some xeroderma pigmentosum group E patients, is homologous to a slime-mold protein. *Nucl. Acids Res.* 21: 4111-4118.

Thommes, P. and Hubscher, U. (1992). Eukaryotic DNA helicases: essential enzymes for DNA transactions. *Chromosoma* 101: 467-473.

Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D. and Hoeijmakers, J.H.J. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* 71: 939-953.

Uberbacher, E. and Mural, R. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* 88: 11261-11265.

van Gool, A.J., Verhage, R., Swagemakers, S.M.A., van de Putte, P., Brouwer, J., Troelstra, C., Bootsma, D. and Hoeijmakers, J.H.J. (1994). RAD26, the functional *S. cerevisiae* homolog of the Cockayne syndrome B gene ERCC6. *EMBO J.* 13: 5361-5369.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* 270: 484-487.

Voght, G. and Argos, P. (1992). Searching for distantly related protein sequences in large databases by parrallel processing on a transputer machine. *CABIOS* 8: 49-55.

Watkins, J.F., Sung, P., Prakash, L. and Prakash, S. (1993a). The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* 13: 7757-7765.

Wilcox, A.S., Khan, A.S., Hopkins, J.A. and Sikela, J.M. (1991). Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome: *Nucl. Acids. Res.* 19: 1837-1843.

Wright, P. (1989). What can two-dimensional NMR tell us about proteins? *Trends Biochemical Sci.* 14: 255-260.

Yasui, A., Eker, A.P.M., Yasuhira, S., Yajima, H., Kobayashi, T., Takao, M. and Oikawa, A. (1994). A new class of DNA photolyases present in various organisms including aplacental mammals. *EMBO J.* 13: 6143-6151.

CHAPTER

IV

Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23

PURIFICATION AND CLONING OF A NUCLEOTIDE EXCISION REPAIR COMPLEX INVOLVING THE XERODERMA PIGMENTOSUM GROUP C PROTEIN AND A HUMAN HOMOLOGUE OF YEAST RAD23

Chikahide Masutani[1], Kaoru Sugasawa[1,2], Junn Yanagisawa[3,4], Tadao Sonoyama[3], Michio Ui[3,5], Takemi Enemoto[3], Koji Takio[6], Kiyoji Tanaka[7], Peter J. van der Spek[8], Dirk Bootsma[8], Jan H.J. Hoeijmakers[8], and Fumio Hanaoka[9].

[1] Cellular Physiology Laboratory, [2] Biodesign Research Group and [6]Department of Research Fundamentals Technology, The Institute of Physical and Chemical Research (RIKEN), Wako-shi, Saitama 351-01, [3]Department of Physiological Chemistry, Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113, [7]Institute for Molecular and Cellular Biology, Osaka University, 1-3 Yamada-oka, Suita, Osaka 565, Japan and [8]Department of Cell Biology and Genetics, Medical Genetic Centre, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

[4]Present address: La Jolla Cancer Reseach Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

[5]Present address: Ui Special Laboratory, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama 351-01, Japan.

EMBO Journal 13: 1831-1843, 1994.

ABSTRACT

Complementation group C of xeroderma pigmentosum (XP) represents one of the most common forms of this cancer-prone DNA repair syndrome. The primary defect is located in the subpathway of the nucleotide excision repair system, dealing with the removal of lesions from the non-transcribing sequences ('genome-overall' repair). Here we report the purification to homogeneity and subsequent cDNA cloning of a repair complex by *in vitro* complementation of the XP-C defect in a cell-free repair system containing UV-damaged SV40 minichromosomes. The complex has a high affinity for ssDNA and consists of two tightly associated proteins of 125 and 58 kDa. The 125kDa subunit is an N-terminally extended version of previously reported XPCC gene product which is thought to represent the human homologue of the *Saccharomyces cerevisiae* repair gene *RAD4*. The 58 kDa species turned out to be a human homologue of yeast RAD23. Unexpectedly, a second human counterpart of RAD23 was identified. All RAD23 derivates share a ubiquitin-like N-terminus. The nature of the XP-C defect implies that the complex exerts a unique function in the genome-overall repair pathway which is important for prevention of skin cancer.

INTRODUCTION

DNA repair plays a key role in the prevention of carcinogenesis and mutagenesis. Nucleotide excision repair (NER) is the principal pathway for eliminating a broad spectrum of structurally unrelated lesions such as ultraviolet (UV)-induced cyclobutane pyrimidine dimers and [6-4] photoproducts, as well as bulky chemical adducts and certain cross-links (for review see Friedberg, 1985). At least five steps can be discerned in the reaction mechanism of NER: damage recognition, incision of the damaged strand on both sides of the lesion (Huang *et al.*, 1992), excision of the lesion-containing oligonucleotide, synthesis of new DNA using the undamaged strand as a template, and ligation. Although the molecular mechanism underlying NER is now well understood in the bacterium *Escherichia coli* (Van Houten, 1990; Hoeijmakers, 1993a; Sancar and Hearst, 1993), the mechanism of NER in mammals has not yet been clarified. The high level of sophistication of the NER system is illustrated by the existence of distinct subpathways. One of these deals with the preferential elimination of lesions that thwart ongoing transcription (transcription-coupled repair), a second subpathway effects the slower repair

of the rest of the genome ('genome-overall' repair) (Hanawalt and Mellon, 1993).

The association of a DNA repair defect with a human cancer-prone syndrome, xeroderma pigmentosum (XP) was first reported by Cleaver (1968). XP is a rare, autosomal recessive disease associated with a high incidence of sunlight-induced skin abnormalities including cancers (Cleaver and Kraemer, 1989). Coplementation tests by cell fusion have provided evidence for the existence of at least seven NER-deficient complementation groups: XP-A to XP-G.

Another important category of mammalian mutants is the class of laboratory-induced, UV-sensitive rodent cell lines. At least 11 NER complementation groups have been identified (Riboni *et al.*, 1992; Collins, 1993). DNA-mediated gene transfer has led to the cloning of human genes that correct the mutations in rodent complementation groups. These human genes are named 'excision repair cross complementing rodent repair deficiency' (*ERCC*) genes, followed by a number referring to the corrected complementation group. With the exception of ERCC1 (van Duin *et al.* 1989), all other cloned ERCC genes appeared to be also responsible for one of the XP defects or for one of the forms of another NER disorder, Cockayne's syndrome (CS). Thus *ERCC2*, *ERCC3*, *ERCC5* and *ERCC6* were found to be identical to the genes causing XP-D, XP-B, XP-G and CS-B, respectively (Weeda *et al.*, 1990; Fletjer *et al.*, 1992; Troelstra *et al.*, 1992; O'Donavan and Wood, 1993; for a recent review see Hoeijmakers, 1993b). Hence, a considerable overlap exists between the rodent mutants and the human disorders. In addition, phenotypic correction of XP-cells by genomic or cDNA transfection has resulted in the cloning of the genes implicated in XP-A (the *XPAC* gene, for XP-A correcting; Tanaka *et al.*, 1990) and XP-C (the *XPCC* gene; Legerski and Peterson, 1992).

Sequence analysis has revealed a striking evolutionary conservation. For all mammalian NER genes cloned to date, (presumed) yeast counterparts have been found (Hoeijmakers, 1993b; A. van Gool, C. Troelstra and J.H.J. Hoeijmakers, unpublished data). In *Saccharomyces cerevisiae* a minimum of 11 distinct NER mutants has been identified, collectively designated the *RAD3* epistasis group. The degree of similarity between the human and yeast genes strongly suggests that the NER pathways in both extremes of the eukaryotic spectrum are largely superimposable and mechanistically very related. However, for several yeast NER genes a mammalian equivalent is still lacking.

Another powerful tool for unravelling the molecular mechanism of excision repair is an *in vitro* system based on cell-free extracts capable of performing NER on a damaged naked DNA template. We have recently adapted this system originally developed by

Wood *et al.* (1988) and Sibghat-Ullah *et al.* (1989) to the use of SV40 minichromosomes (Sugasawa *et al.*, 1993); Masuitani *et al.*, 1993). Using this system as an assay, we report here the purification to homogeneity of a 125 kDa XP-C correcting protein from HeLa cells, the cloning of the corresponding cDNA, as well as the co-purification and cDNA cloning of a tighly associated protein of 58 kDa. The latter turned out to be homologous to the yeast RAD23 NER protein, thus filling in one of the remaining gaps in the parallels between yeast and man. Interestingly, a second human homologue of RAD23 was identified as well. Both human homologues of RAD23 (designated HHR23A and HHR23B) harbour a ubiquitin-like N-terminal domain. The XPCC-HHR23B complex is suspected to play a selective role in the genome-overall NER subpathway, since the repair defect in XP-C is limited to the genome-overall system (Kantor *et al.*, 1990; Venema *et al.*, 1990, 1991).

RESULTS

Purification of the XP-C correcting protein from HeLa cells

A cell-free system for DNA repair was constructed in which UV-damaged SV40 minichromosomes can be repaired during an incubation with extracts from human cells (Sugasawa *et al.*, 1993). The system contains UV-irradiated or unirradiated SV40 minichromosomes as well as unirradiated pUC19 supercoiled DNA. The following evidence indicates that DNA synthesis with UV-irradiated chromosomes is due to excision repair of UV-induced damage: (i) it is defective in extracts from all excision- deficient XP complementation groups, (ii) it is stimulated by the addition of T4 endonuclease V to XP extracts, (iii) it is complemented by mixing XP cell extracts of different complementation groups, (iv) it is complemented by the addition of purified XP-A complementing (*XPAC*) gene product to XP-A cell extracts, and (v) it is inhibited by the addition of antiserum raised against XPAC protein to repair-proficient cell extracts (Masutani *et al.*, 1993).

One important use of cell-free systems is for fractionation and biochemical identification of factors involved in the reactions. We used our cell-free DNA repair system for purification of a protein that corrects DNA repair defects of XP-C cell extracts. Activity that complements the repair defect of XP4PASV (group C) cell extracts was assayed in the cell-free system. XP-C complementing activity was detected in nuclear extracts from HeLa cells and purified by successive column chromatographies on phosphocellulose, single-stranded DNA - cellulose, FPLC CM cosmogel and FPLC Mono

Q (HR5/5) (for details see Materials and methods). The XP-C correcting activity bound strongly to a single-stranded DNA - cellulose column (being eluted between 0.6 and 1.5 M KCl), suggesting that the protein associates with DNA in cells. The purification procedure yielded a good recovery of the activity and ≈2000 fold increase in its activity over that of the starting material (Table I).

Table I Purification of XP-C correcting protein from HeLa cells.

	Protein (mg)	Activity (units)	Specific activity (U/mg)	Purification (fold)
Nuclear extract	1390	38 160	27.5	1
Phosphocellulose	634	22 360	35.3	1.28
ssDNA cellulose	1.86	15 680	8430	306
CM cosmogel	0.40	14 250	35 625	1295
Mono Q	0.23	12 780	55 565	2020

After FPLC Mono Q column chromatography, two polypeptides with apparent molecular masses of 125 and 58 kDa (p125 and p58) were detected by SDS-PAGE (Figure 1).

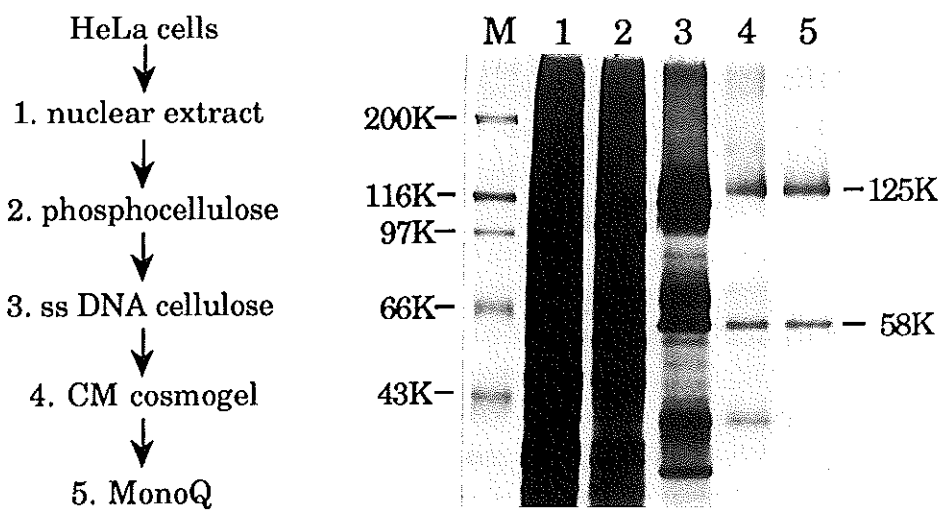


Fig.1. Purification of the XP-C correcting protein from HeLa cells.
The purification scheme is shown on the left. A sample at each purification step (indicated by numbers) was subjected to SDS-PAGE (8% polyacrylamide) and stained with silver. The marker proteins used were myosin, ß-galactosidase, phosphorylase B, bovine serum albumin and ovalbumin (lane M).

As shown in figure 2A, the XP-C correcting activity was eluted from a Sephacryl S-300 column at a position corresponding to a molecular weight of 500-550 kDa as estimated by a linear extrapolation. The two polypeptides were co-eluted with the activity (Figure 2B), indicating that these polypeptides form a physical complex and are associated with the XP-C correcting activity. Although the estimated molecular weight is much bigger than the sum of 125 and 58 kDa, it is unlikely that it is due to the protein aggregation, because we employed the gel filtration in the presence of 0.3 M KCl, 10% glycerol and 0.01% Triton X-100. In fact, the activity sedimented at 6.2S (Figure 2C) on glycerol density gradient centrifugation under the same solution condition as in the gel filtration except for the various concentrations of glycerol. Again the p125 and p58 polypeptides co-migrated with the activity (Figure 2D). The molecular weight of the p125 - p58 protein complex was estimated to be 110 kDa from the sedimentation position in the glycerol gradient, much smaller than that predicted from the results of gel filtration analysis and even smaller than the sum of 125 and 58 kDa, suggesting that the XP-C correcting protein is laminar in shape. We note that neither of the two proteins has the 93 kDa molecular weight predicted for the XPCC gene product, encoded by the cDNA cloned recently by Legerski and Peterson (1992). The purified XP-C correcting protein was tested for various enzymatic activities. No detectable DNA polymerase, DNA helicase, DNA ligase, DNA exonuclease or DNA endonuclease activity with UV-irradiated or unirradiated DNA was found under the conditions described in Materials and methods.

Specificity of complementation by the XP-C correcting protein fraction

The specificity of the activity of the p125 - p58 protein preparation to complement defects of XP-C cell extracts was examined. Addition of 10 ng of the purified XP-C cells (XP4PASV and XP3KA) induced a correction of the UV-specific repair synthesis to a level comparable with that of a repair-proficient cell extract (Figure 3A and B). In contrast, no significant increase in UV-dependent incorporation in the SV40 minichromosomes was observed in cell-free extracts from any of the six remaining excision-deficient XP complementation groups (Figure 3B).

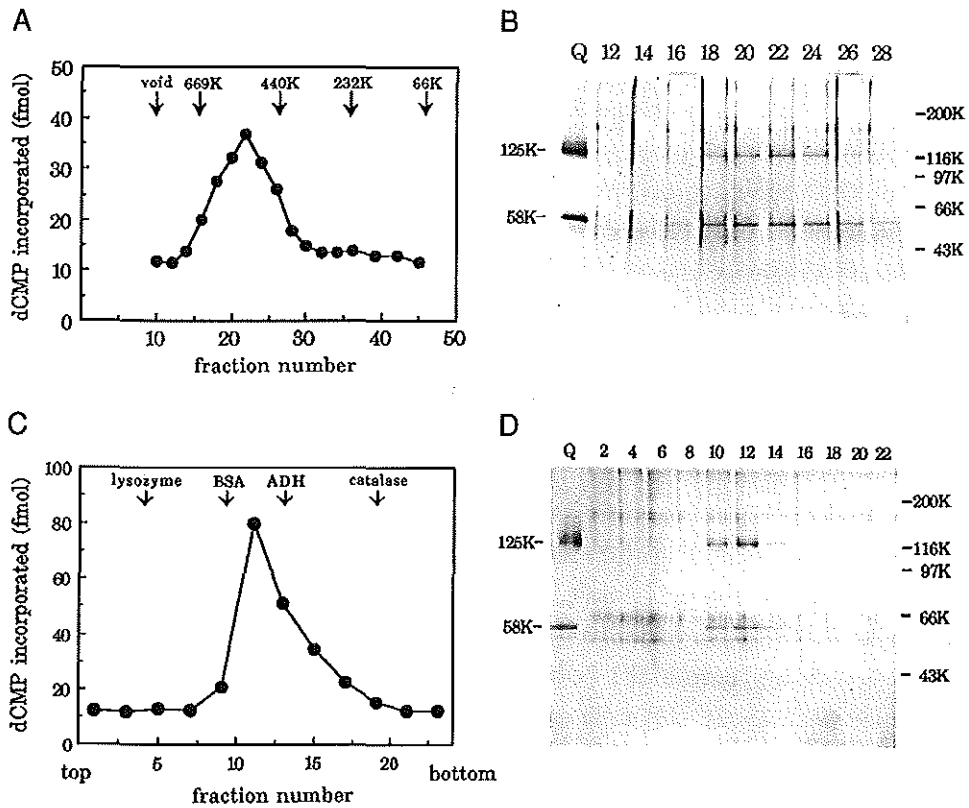


Fig.2. Physical properties of the XP-C correcting protein. (A) Purified XP-C correcting protein was subjected to Sephacryl S-300 column chromatography and the XP-C correcting activity in eluted fractions was assayed as described in Materials and methods. The incorporation of radioactive materials into UV-irradiated mini-chromosomes was quantified. The positions of elution of marker proteins (thyroglobulin, ferritin, catalase and bovine serum albumin) fractionated under identical conditions are indicated by their molecular weights. (B) Samples (20 μl) of fractions around the peak of activity in panel A were subjected to SDS-PAGE (8% polyacrylamide) and stained with silver. (C) Purified XP-C correcting protein was subjected to glycerol density gradient centrifugation and assayed for XP-C correcting activity. The sedimentation positions of marker proteins in a parallel gradient are indicated. (D) Samples (20 μl of fractions in panel C were subjected to SDS-PAGE (8% polyacrylamide) and stained with silver.

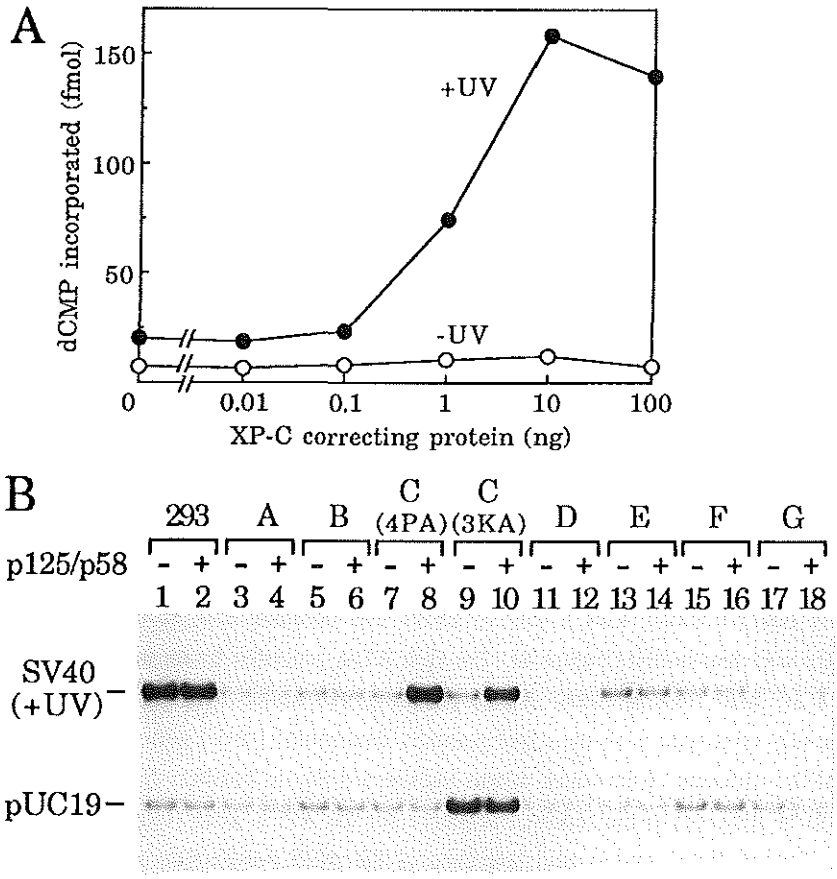


Fig.3. Specificity of complementation by p125-p58 XP-C correcting protein. (A) Dose-response of XP-C correction. UV-irradiated (closed circles) or unirradiated (open circles) SV40 minichromosomes were incubated in standard reaction mixtures with XP4PASV cell extracts with increasing amounts of the p125-p58 protein complex. The incorporation into minichromosomes was quantified. (B) Complementation group specificity. UV-irradiated SV40 mini-chromosomes were incubated in the standard reaction mixture in the presence (even numbered lanes) or absence (odd numbered lanes) of the p125-p58 protein complex (10 ng) with 80 μg of protein of 293 (lanes 1 and 2), XP2OSSV (XP-A) (lanes 3 and 4), CRL199 (XP-B) (lanes 5 and 6), XP4PASV (XP-C) (lanes 7 and 8), XP3KA (XP-C) (lanes 9 and 10), XP6BESV (XP-D) (lanes 11 and 12), XP2RO (XP-E) (lanes 13 and 14), XP2YOSV (XP-F) (lanes 15 and 16) or XP3BRSV (XP-G) (lanes 17 and 18) cell extracts. Purified DNA products were linearized with *Eco*RI and then subjected to 1% agarose gel electrophoresis as described in Materials and methods. An autoradiogram of the gel is shown. Although a higher level of DNA synthesis was observed with pUC19 plasmid DNA and the XP3KA cell extract than with other extracts, it must be an independent phenomenon from the repair event because the synthesis did not change on addition of p125-p58 protein complex in spite of the increase on addition of UV-irradiated chromosomes (lanes 9 and 10).

Isolation of the cDNA encoding the p125 subunit

To clone the cDNA for the p125 - p58 protein complex, the two polypeptides were separated from each other by gel filtration in the presence of guanidine-HCl (separation under physiological conditions failed. CNBr cleavage yielded completely different peptide profiles for the two proteins. Thus it is unlikely that p58 is a proteolytic product of the p125 subunit (data not shown). One partial amino acid sequence of p125 and two of p58 were determinied, none of which matched with the predicted amino acid sequence of the previously cloned *XPCC* gene. The sequence of p125 was > 50 amino acids in lenght. Since the same sequence was obtained for the undigested p125 polypeptide, it represents the N-terminal sequence of p125.

To prepare a DNA probe for screening cDNA libraries, two sets of oligonucleotide mixtures were synthesized according to the determined amino acid sequence of p125 (see Materials and methods) and used for the RT-PCR with poly(A)$^+$ RNA from HeLa cells. A PCR product of the expected length (132 bp) and nucleotide sequence was obtained and used for screening a λgt10 cDNA library prepared from HeLa cells. A positive clone with a 3.6 kb insert was obtained and its complete nucleotide sequence was determined (Figure 4). The first ATG, preceded by an in-frame stop codon, initiates an open reading frame (ORF) encoding 940 amino acids. The N-terminal part was entirely consistent with the experimentally determined partial amino acid sequence. In view of the different N-terminus, it was unexpected that at position 266 the sequence was found to be identical to the reported sequence for the XPCC gene (Legerski and Peterson, 1992). The predicted amino acid sequence of the p125 polypeptide was joined in-frame with the deduced ORF of the XPCC protein. As a result, there were 117 additional amino acids at the N-terminus of the XPCC protein and the calculated molecular mass increased from 93 to 106 kDa. Therefore, the deduced product of the reported *XPCC* gene is probably part of the p125 protein truncated in the N-terminal region. We infer that the p125 polypeptide represents the full-length XPCC gene product.

Legerski and Peterson (1992) reported that the putative XPCC protein shares limited homology with the *RAD4* gene product of *S. cerevisiae*. We could not find any significant additional homology with RAD4 or other proteins or any functional motifs in the newly identified N-terminal region of the p125.

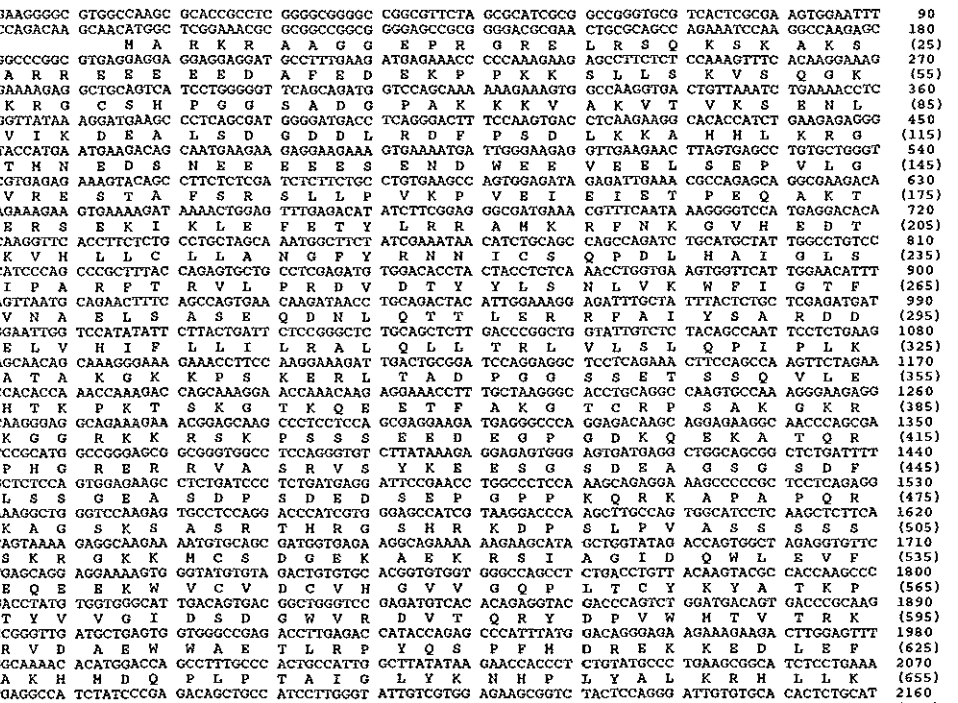


Fig.4. Nucleotide and predicted amino acid sequence of the p125/XPCC. Top numbers on the right are those of nucleotide residues and lower ones (in parentheses) are those of amino acids. A termination codon, TAG, in the 5' untranslated region is boxed. Two circled methionines are putative initiation codons for the p125 of the XP-C correcting protein and the previously reported XPCC protein (Legerski and Peterson, 1992), respectively. The arrow indicates the start position of the reported sequence. The asterisk indicates the termination codon, TGA, for this ORF. Doubly underlined amino acids represent a peptide sequence derived from the purified p125 polypeptide. Boxed nucleotides (nucleotide positions:286, 1601, 2166 and 3024) are different from those in the sequence reported by Legerski and Peterson (1992). The GenBank accession number for human XPCC (p125) is D21089.

Cloning and sequence analysis of the cDNA encoding the p58 subunit

To obtain a cDNA clone for the p58, an oligonucleotide mixture was synthesized according to one of the two determined amino acid sequences of p58 (see Materials and methods) and was used for screening a λgt10 cDNA library prepared from HeLa cells. A positive clone with a 2.9 kb insert was obtained and its complete nucleotide sequence was determined (Figure 5).

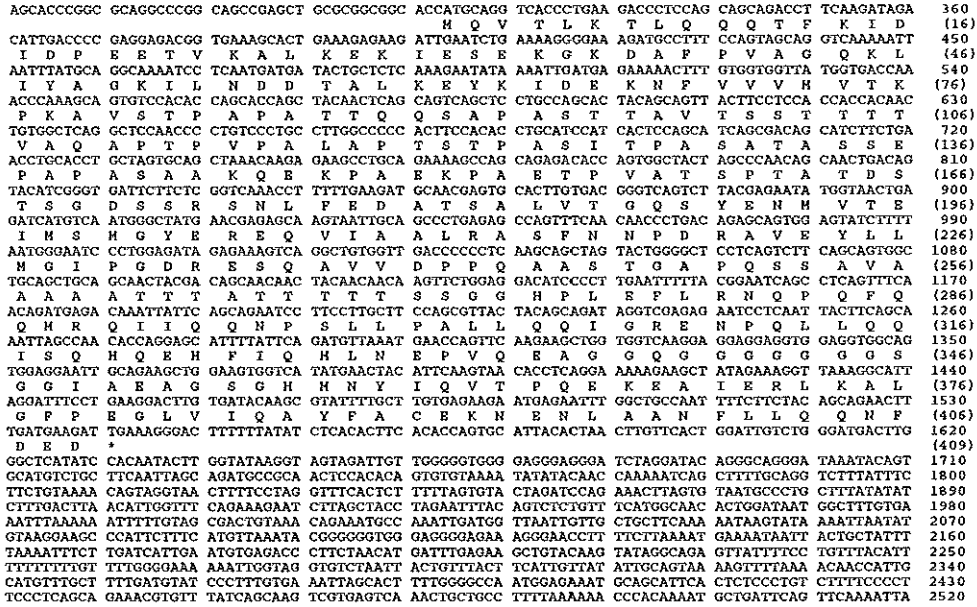


Fig.5. Nucleotide and predicted amino acid sequence of the p58/HHR23B. Top numbers on the right are those of nucleotide residues and lower ones (in parentheses) are those of amino acids. An in frame termination codon, TGA, in the 5' untranslated region is boxed. The asterisk indicates the termination codon, TGA, for this ORF. Doubly underlined amino acids represent peptide sequences derived from the purified p58 polypeptide. Putative polyadenylation signals (ATTAAA) in the 3' untranslated region are shown by bold boxes. Three ATTTA sequence motifs (mRNA degradation signals) are underlined. The GenBank accession number for human XPCC (p58/HHR23B) is D21090.

An ORF encoding 409 amino acids including both determined amino acid sequences was found. Although the calculated molecular mass of the protein was only 43 kDa, we concluded that the clone includes the full length of the coding region of the p58 polypeptide because a termination codon (TGA) was found in frame in an upstream region of the putative initiation codon. Consistent with this notion is our finding that the protein overproduced in *E.coli* by the cloned cDNA migrates at the same position as the p58 protein (unpublished results).

Searches in various databases for sequence homology to the p58 ORF revealed several interesting features:

(i) At the nucleotide sequence level, two expressed sequence tags (ESTs) with unknown function representing partial human cDNA clones of brain and a liver cell line [accession numbers M85669 (Adams *et al.*, 1992) and D12303 (Okubo *et al.*,1992)] were - with the exception of a few sequence uncertainties - identical to the corresponding part of the p58 cDNA sequence. These cDNAs are therefore expected to be derived from the p58 gene.

(i) Amino acid sequence comparison uncovered significant resemblance between the N-terminal 79 amino acids of p58 and ubiquitin and a similar domain in various ubiquitin fusion proteins (see below).

(iii) Interestingly, the p58 amino acid sequence appeared to share extensive overall sequence homology with the *S.cerevisiae RAD23* gene (Melnick and Sherman, 1993; sequence prior to publication kindly provided by S. Prakash, Galveston), a member of the *RAD3* NER epistasis group for which no human homologue was identified yet. The RAD23 gene is identical to the sygg-orf29 sequence, identified on chromosome 5 as part of the yeast genome sequencing project [accession number L10830 (Mulligan *et al.*, unpublished)].

(iv) Finally, using the BLAST algorithm (Altschul *et al.*, 1990), which is able to detect amino acid sequence homologies translated from all six possible reading frames, we identified several human partial cDNAs which exhibited some homology to the amino acid sequence of p58, when uncertainties in the sequence are taken into account. These cDNAs were derived from heart [accession number M77024 Eichbaum *et al.*, unpublished)] and a T lymphoblastoid cell-line (accession numbers Z15569, Z12748 and Z15568). Because of the presence of some sequence ambiguities and to find out whether this cDNA shared additional sequence similarity to p58, we decided to isolate the corresponding full-length cDNA by RT-PCR using total HeLa RNA combined with library screening.

The nucleotide and deduced amino acid sequence of the cDNA encoded by this p58-related gene, that we termed tentatively *HHR23A* for human homologue of RAD23 A is presented in Figure 6. The ORF, starting from the first ATG encodes an acidic protein (pI 4.4) of 363 amino acids, with a calculated molecular mass of 40 kDa. Also this protein synthesized in *E.coli* migrates well above its predicted molecular weight (P.J. van der Spek, unpublished results). The 3' UTR harbours a canonical AATAAA polyadenylation signal 12 bp before the start of the poly(A) tail.

```
GGGATCCCGG GGCCGCCGCG TCGCTCGGGC CCCGCCATGG CCGTCACCAT CACGCTCAAA ACGCTGCAGC AGCAGACCTT CAAGATCCGC    90
                                  M  A  V  T  I  T  L  K  T  L  Q  Q  Q  T  F  K  I  R              (18)
ATGGAGCCTG ACGAGACGGT GAAGGTGCTA AAGGAGAAGA TAGAAGCTGA GAAGGGTCGT GATGCCTTCC CCGTGGCTGG ACAGAAACTC   180
M  E  P  D  E  T  V  K  V  L  K  E  K  I  E  A  E  K  G  R  D  A  F  P  V  A  G  Q  K  L          (48)
ATCTATGCCG GCAAGATCTT GAGTGACGAT GTCCCTATCA GGGACTATCG CATCGATGAG AAGAACTTTG TGGTCGTCAT GGTGACCAAG   270
I  Y  A  G  K  I  L  S  D  D  V  P  I  R  D  Y  R  I  D  E  K  N  F  V  V  V  M  V  T  K          (78)
ACCAAAGCCG GCCAGGGTAC CTCAGCACCC CCAGAGGCCT CACCCACAGC TGCCCCAGAG TCCTCTACAT CCTTCCCGCC TGCCCCCACC   360
T  K  A  G  Q  G  T  S  A  P  P  E  A  S  P  T  A  A  P  E  S  S  T  S  F  P  P  A  P  T         (108)
TCAGGCATGT CCCATCCCCC ACCTGCCGCC AGAGAGGACA AGAGCCCATC AGAGGAATCC GCCCCCACGA CGTCCCCAGA GTCTGTGTCA   450
S  G  M  S  H  P  P  P  A  A  R  E  D  K  S  P  S  E  E  S  A  P  T  T  S  P  E  S  V  S          (138)
GGCTCTGTTC CCTCTTCAGG TAGCAGCGGG CGAGAGGAAG ACGCGGCCTC CACGCTAGTG ACGGGCTCTG AGTATGAGAC GATGCTGACG   540
G  S  V  P  S  S  G  S  S  G  R  E  E  D  A  A  S  T  L  V  T  G  S  E  Y  E  T  M  L  T          (168)
GAGATCATGT CCATGGGCTA TGAGCGAGAG CGGGTCGTGG CCGCCCTGAG AGCCAGCTAC AACAACCCCC ACCGAGCCGT GGAGTATCTG   630
E  I  M  S  M  G  Y  E  R  E  R  V  V  A  A  L  R  A  S  Y  N  N  P  H  R  A  V  E  Y  L          (198)
CTCACGGGAA TTCCTGGGAG CCCCGAGCCG GAACACGGTT CTGTCCAGGA GAGCCAGGTA TCGGAGCAGC CGGCCACGGA AGCAGCAGGA   720
L  T  G  I  P  G  S  P  E  P  E  H  G  S  V  Q  E  S  Q  V  S  E  Q  P  A  T  E  A  A  G          (228)
GAGAACCCCC TGGAGTTCCT GCGGGACCAG CCCCAGTTCC AGAACATGCG GCAGGTGATT CAGCAGAACC CTGCGCTGCT GCCCGCCCTG   810
E  N  P  L  E  F  L  R  D  Q  P  Q  F  Q  N  M  R  Q  V  I  Q  Q  N  P  A  L  L  P  A  L          (258)
CTCCAGCAGC TGGGCCAGGA GAACCCTCAG CTTTTACAGC AAATCAGCCG GCACCAGGAG CAGTTCATCC AGATGCTGAA CGAGCCCCCT   900
L  Q  Q  L  G  Q  E  N  P  Q  L  L  Q  Q  I  S  R  M  Q  E  Q  F  I  Q  M  L  N  E  P  P          (288)
GGGGAGCTGG CGGACATCTC AGATGTGGAG GGGGAGGTGG GCGCCATAGG AGAGGAGGCC CCGCAGATGA ACTACATCCA GGTGACGCCG   990
G  E  L  A  D  I  S  D  V  E  G  E  V  G  A  I  G  E  E  A  P  Q  M  N  Y  I  Q  V  T  P          (318)
CAGGAGAAAG AAGCTATAGA GAGGTTGAAG GCCCTGGGCT TCCCAGAGAG CCTGGTCATC CAGGCCTATT TCGCGTGTGA AAAAAATGAG  1080
Q  E  K  E  A  I  E  R  L  K  A  L  G  F  P  E  S  L  V  I  Q  A  Y  F  A  C  E  K  N  E          (348)
AACTTGGCTG CCAACTTCCT CCTGAGTCAG AACTTTGATG ACGAGTGATG CCAGGAAGCC AGGCCACCGA AGCCCCCACC CTACCCTTAT  1170
N  L  A  A  N  F  L  L  S  Q  N  F  D  D  E  *                                                    (363)
TCCATGAAAG TTTTATAAAA GAAAAAATAT ATATATATTC ATGTTTATTT AAGAAATGGA AAAAAAAATC AAAAATCTTA AAAAAACAAG  1260
CAAACAGTCC AGCTTCCTGT CCTCCTAAAG TGGCCCCTGT TCCCATCTCC CGGGCCAGAC AGCTGTCCCC CCGTCCTCCT CCCCAGCCCA  1350
GCCTGCTCAG AGAAGCTGGC AGGACTGGGA GGCGACAGAT GGGCCCCTCT TGGCCTCTGT CCCAGCTCTC TGCAGCCAGA CGGAAAGGCG  1440
GCTGCTTGCC TCTCCATCCT CCGAAAAACC CCTGAGGACC CCCCCCCATC CTCTTCTAGG ATGAGGGGAA GCTGGAGCCC CAACTTTGAT  1530
CCTCCATTGG AGTGGCCCAA ATCTTTCCAT CTAGGGCAAG TCCTGAAAGG CCCAAGGCCC CCTCCCAGTC TGGCCTTGGC CTCCAGCCTG  1620
GAGAAGGGCT AACATCAGCT CATTGTCAAG GCCACCCCCA CCCCAGAACA GAACCGTGTC TCTGATAAAG GTTTTGAAGT GAATAAAGTT  1710
TTAAAAACTA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA                                                        1750
```

Fig.6. Nucleotide and predicted amino acid sequence of the HHR23A. Top numbers on the right indicates the numbering of nucleotides; the numbers in parentheses correspond to those of the amino acids. The sequence 5' proximal to the ATG matches perfect with the optimal translation initiation sequence (Kozak, 1991). The asterisk indicates the termination of the ORF. The polyadenylation site present in the 3' untranslated region before the poly(A) tail is boxed. The GenBank accession number for HHR23A is D21235.

As shown by the amino acid sequence alignment in Figure 7A, the two human proteins exhibit a high overall homology to each other (57% identity, 76% similarity) and to the yeast *RAD23* gene product (30-34% identity, 41% homology). Furthermore, it is worth noting that regions rich in S, T, P and A amino acids are found at two locations. The first starts immediately following the ubiquitin-like domain: residues 79-144; 84% of

which is S, T, P, or A (figures for p58). The second runs from residues 241 to 272, 87% of which is made up of these residues (figures for p58). Finally, a glycine-rich stretch is present in p58 between residues 336-348.

The alignment of all three RAD23 homologues with (human) ubiquitin and with similar domains in other ubiquitin-like fusion proteins is presented in Figure 7B. The level of homology to ubiquitin is very similar for all three polypeptides (25%-31% identity, 55%-59% similarity) and is in the same range as that of other ubiquitin hybrid polypeptides. We conclude that both human proteins belong to the family of ubiquitin-fusion proteins and represent homologues of RAD23. Consistent with the designation HHR23A, we term the p58 HHR23B. Apart from the ubiquitin motif, no other functional domains could be identified in the HHR23/RAD23 sequence using the PROSITE software package or comparison to other proteins.

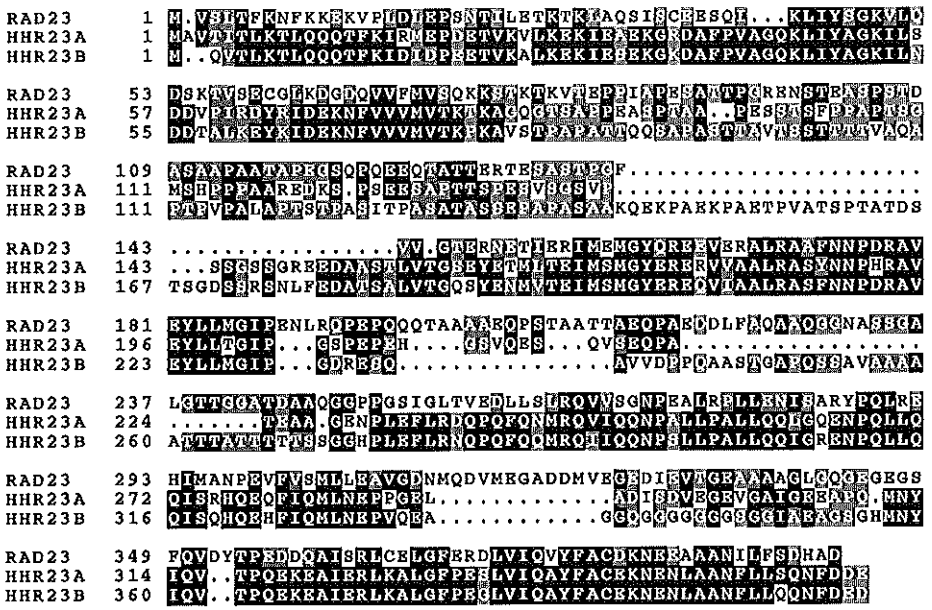


Figure 7A

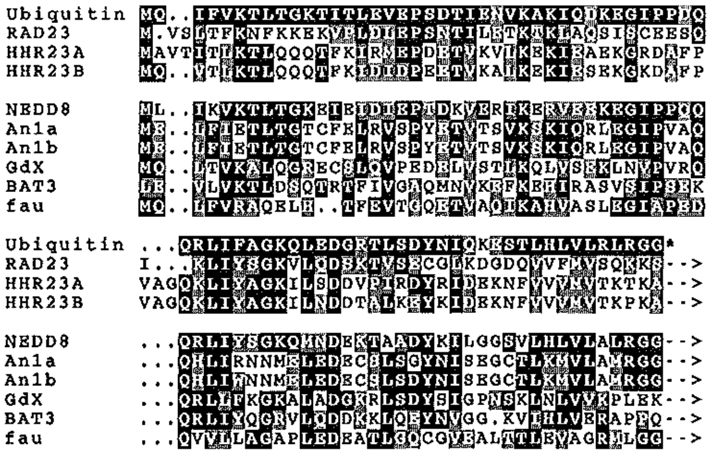


Figure 7B

Fig.7. Sequence alignment of the yeast and human homologues of RAD23 with each other and with ubiquitin. (A) Conserved sequences between yeast RAD23, HHR23A and p58/HHR23B. The amino acid sequence of the human HHR23A and p58/HHR23B proteins are compared with yeast RAD23. (B) Alignment of ubiquitin, RAD23, HHR23A, p58/HHR23B and ubiquitin-like sequences. The N-terminal conserved regions of the RAD23, HHR23A, p58/HHR23B and the ubiquitin-like domain in the NEDD8, AN1A, AN1B, GdX, BAT3 and fau proteins are compared with ubiquitin. Sequences used in this figure are NEDD8 (Kumar *et al.*, 1992), AN1A, AN1B (Linnen *et al.*, 1993), GdX (Toniolo *et al.*, 1988), BAT3 (Banerji et al., 1990) and fau (Kas et al., 1992). The amino acid sequence is given in the one letter code. Identical amino acids are presented by black boxes, whereas similar residues (A, S, T, P; D, E, N, Q; R, K; I, L, M, V; F, Y, W) are given in grey boxes.

DISCUSSION

XP-C correcting protein

XP-C is one of the most common forms of XP (Kraemer *et al.*, 1987). Group C patients display the (for XP obligate) features of hypersensitivity to sunlight (UV) and other cutaneous manifestations, including predisposition to skin cancer, but a second hallmark, accelerated neurodegeneration, is absent. Recently, the NER defect in XP-C was pinpointed to the genome-overall subpathway; 'transcription-coupled repair' functions normally in these cells (Kantor *et al.*, 1990; Venema *et al.*, 1990, 1991). This provides a plausible explanation for the relatively high cellular resistance to UV. Furthermore, transcription-coupled NER may be important for cunteracting neurodegeneration. However, since this repair process is limited only to the transcribed strand of active genes, it has no effect on mutagenesis in the non-transcribed strand nor in the rest of the genome. Presumably, this explains why XP-C patients cannot effectively avert sunlight-induced skin cancer. Here we have purified a protein complex that based on the nature of the XP-C mutation is expected to operate specificially in the 'genome-overall' repair pathway. More recently, six distinct mutations including point mutations, deletions and insertions were detected in the *XPCC* gene of five XP-C cell lines (Li *et al.*, 1993). Thus a defect in the p125 subunit gives rise to cancer proneness. The complex consists of two tightly associated polypeptides: a 125 kDa species representing the XP-C gene product and a 58 kDa protein, which turned out to be a human homologue of *S.cerevisiae RAD23*, one of the remaining yeast NER genes for which no human counterpart was known. Unexpectedly, a second human equivalent of *RAD23* appeared to exist. All *RAD23* homologues share an N-terminal ubiquitin-like domain.

A DNA-dependent ATPase, designated ATPase Q1, was previously found to be altered in XP-C cells in terms of its elution position from a FPLC Mono Q column (Yanagisawa et al., 1992). However, the XP-C correcting protein described here differs from the ATPase Q1 for the following reasons. First, we could not detect any DNA helicase activity in the XP-C correcting protein while the ATPase Q1 has relatively weak but detectable helicase activity. Second, the molecular weights of the two polypeptides purified in this work are different from that of purified ATPase Q1 (73 kDa on SDS-PAGE). Third, a cDNA clone for ATPase Q1 is different from the cDNA clones of p125 and p58. Fourth, purified or partially purified ATPase Q1 cannot complement the repair defects of XP-C cell extracts in our cell-free system (C. Masutani, unpublished observations). Despite the above facts, the alteration of elution of ATPase Q1 from Mono

Q column was observed with all five independent XP-C cell lines examined. At present, we do not know why two apparently different proteins, the XP-C correcting protein and the ATPase Q1, are altered in XP-C cells. A possible explanation is a direct or indirect effect of the XPCC protein (complex) on the post-translational protein modification. We are now examining this or other possibilities.

Parallels with yeast

Since the *S.cerevisiae RAD4* gene is likely to be the yeast equivalent of XP-C (Legerski and Peterson, 1992), one inference from our observations is that the yeast RAD23 and RAD4 proteins are likely to interact with each other. Intriguing discrepancies emerge when these parallels are extrapolated to the corresponding mutants and genes. *Rad4* and *rad23Δ* mutants are very different. *RAD4* is one of the seven *RAD* genes that appear to be absolutely required for NER, since *rad4* mutants do not show detectable incisions during incubation after UV exposure (Friedberg, 1988). In contrast, *rad23Δ* mutants exhibit only a partial NER defect, supporting the idea that this gene does not play an essential role in the NER process (Perozzi and Prakash, 1986). Furthermore, both genes differ in their transcriptional response to UV. Transcription of the *RAD23* gene is enhanced upon UV irradiation and during meiosis (Madura and Prakash, 1990) but that of *RAD4* is not (Fleer *et al.*, 1987). Although this damage-induced expression may be similar to the SOS response in Bacteria, its functional significance in yeast still needs to be established. Therefore, it will be of interest to examine whether the *RAD23* response is evolutionary conserved. In view of the likely participation of both proteins in the same complex, it is surprising that the mutant phenotypes are so different. One would assume that absence of one component would render the entire complex non-functional. Indeed we cannot separate the two human partners without inactivating the XP-C correcting activity. One possibility is that -like in man- a second *RAD23*-like gene is hidden in the yeast genome and that this related gene takes over part of the functions of *RAD23*. An alternative, although perhaps not so likely option is that *RAD4* is not the real yeast *XPCC* equivalent. One argument in favour of this idea is the prediction that a true yeast *XPCC* mutant should be specifically defective in the 'genome-overall' NER subpathway. When the relative contribution of this NER subpathway to survival is similar in yeast and man, one would expect a milder phenotype for an XP-C-like yeast mutant than actually revealed by *RAD4*. Unfortunatelly, the degree of homology between the XPCC gene product and the RAD4 protein is not conclusive.

Dual genes for RAD23 in man

Why do two homologues of RAD23 exist in man? All NER genes analysed to date appear to be unique. The only precedent of a repair gene duplication are the human homologues of *RAD6*, *HHR6A* and *HHR6B*, which are implicated in post-replication repair (Koken *et al.*, 1991). Concerning *HHR23A* and *HHR23B*, we have found that both genes are expressed in the same cells. In the XPCC purification scheme, however, only the HHR23B protein is found in a complex with p125/XPCC. It is possible that a second form of this complex involving HHR23A exists that has been missed. Alternatively, the HHR23A component may havedissociated from the complex during purification, or HHR23A is engaged in another complex with the human homologue of RAD4, when this gene is not the XP-C counterpart. Unfortunately, no human mutant defective in HHR23A has been identified so far. Transfection and microinjection experiments of this gene into any of the NER-deficient complementation groups for which no gene has been identified yet failed to induce correction, indicating that a *HHR23A* mutant is not existing in the class of known NER syndromes (P.J. van der Spek, unpublished observations).

Possible function of the XPCC-HHR23B complex

The function of the XPCC complex must be accommodated in a step unique to the genome-overall NER subpathway. The purification procedure indicates that the complex has a high affinity for ssDNA. At present we do not know which of the components (or both) is responsible for this property. Previously putative DNA binding motifs have been postulated for the RAD4 protein (Gietz and Prakash, 1988), however comparison with XPCC amino acid sequence reveals that these are not conserved. No obvious DNA binding domains are apparent from the sequence. Also no enzymatic activity was detected for the purified complex (see Results). The only striking domain recognizable using sequence comparison is the ubiquitin-like N-terminus of the RAD23 homologues. Ubiquitin itself is a highly conserved 76 amino acid polypeptide found in all eukaryotes. One or multiple ubiquitin moieties are covalently attached post-translationally to acceptor proteins. This reversible conjugation reaction appears to play an important role in a surprisingly diverse set of regulatory processes, such as selective protein degradation, DNA repair, protein translocation and cell cycle control (reviewed by Jentsch, 1992). Ubiquitin conjugation may also serve as a molecular chaperone.

A number of naturally occuring ubiquitin fusion proteins has been identified. From the alignment shown in Figure 7B, it is apparent that within this functionally diverse family, specific amino acid residues are conserved. Figure 8 shows the position of the

conserved amino acids of the ubiquitin-like family, when projected into the known tertiary structure of ubiquitin itself (Vijay-Kumar *et al.*, 1987). It is clear that most residues are clustered in the inner part of the molecule, whereas the periphery appears more prone to divergence. Particularly, the inner half of the α-helix displays a striking conservation. These observations suggest that the core of the molecule is important for the function of this domain. An additional notable feature is the strict conservation of lysine residue K^{48} in all RAD23 derivatives (Figure 8, arrow). This amino acid is involved in multi-ubiquitination since it can serveas point for attachment for ubiquitin conjugation (Jentsch, 1992).

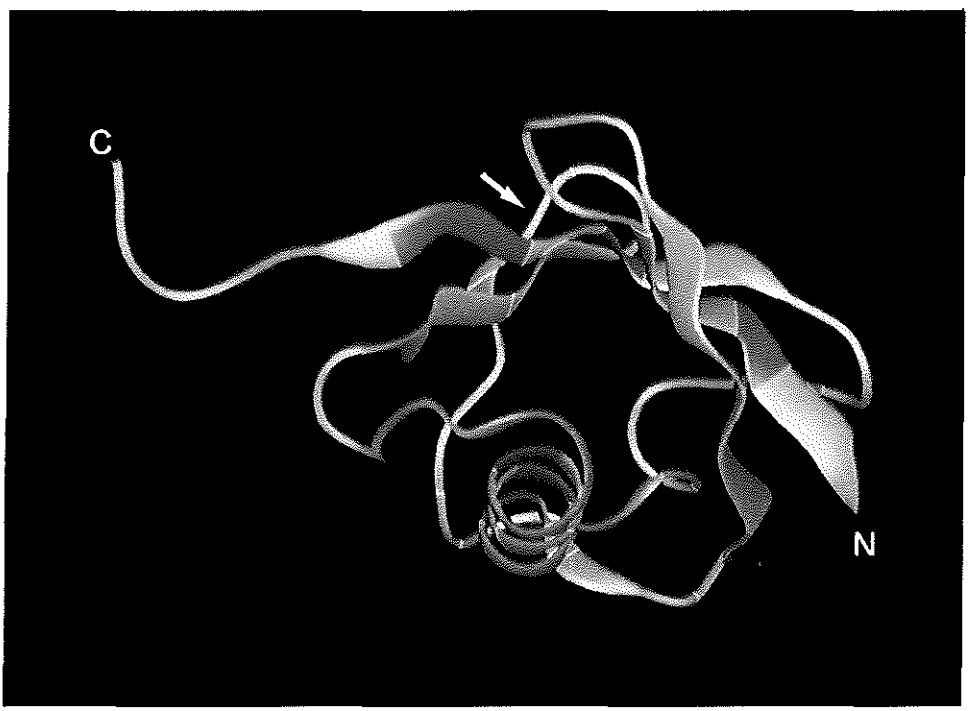


Fig.8. Selective conservation of HHR23 residues in the 'core' of ubiquitin. The computer drawing shows a model for the tertiary structure of ubiquitin including the presence of one α-helix and four ß-sheets. Secondary structure prediction revealed a similar pattern for the N-terminus of RAD23, HHR23A, p58/HHR23B as for ubiquitin (data not shown). The diagram shows in purple the residues of ubiquitin which are identical with those of RAD23, HHR23A and p58/HHR23B as well as with those of many other ubiquitin-like domains: K^{6}, P^{19}, T^{22}, K^{27}, K^{29}, L^{43}, I^{44}, G^{47}, K^{48}, L^{50} and D^{52}. Similar residues (I^{3}, V^{17}, I^{23}, I^{26}, I^{30}, L^{56}, L^{61}, L^{67}, L^{69}, V^{70}) are indicated in orange. It is apparent that intrapolation of these conserved residues into the structure of ubiquitin reveals selective conservation of the core of the protein. Particularly, the inside of the helix seems strongly conserved. The invariant K^{48} is indicated by an arrow.

The alignment in Figure 7B shows also that the C-terminal glycine doublet is absent in all RAD23 derivatives, suggesting that the ubiquitin moiety can not be cleaved off from the remainder. The function of the ubiquitin(-like) domain in different hybrid proteins is not known. Genetic studies in yeast indicated that the ubiquitin moiety of a ribosomal fusion protein might function as a chaperone, facilitating ribosome assembly (Finley *et al.*, 1989). In analogy with this idea, the ubiquitin-like motif in RAD23 may perform a similar role in assembly of the XPCC-HHR23B complex. If so the intrapolation of Figure 8 suggests that the core rather than the outside of the molecule is important for this function. During the preparation of this manuscript, it was demonstrated that the ubiquitin-like domain is required for RAD23 function in *S. cerevisiae* (Watkins *et al.*, 1993). No other functional clues are yielded up by the primary amino acid sequence of either XPCC or HHR23B.

The identification of the XPCC-HHR23B complex adds to the recent discovery of several multi-protein complexes in mammalian NER. The recently described ERCC1 complex consists of a minimum of three proteins: ERCC1, ERCC4, ERCC11 and XPFC (when this protein is not identical to ERCC4 or ERCC11) (Biggerstaff *et al.*, 1993; van Vuuren *et al.*, 1993). In analogy with the yeast RAD1/RAD10 counterpart, this complex may simultaneously be implicated in a mitotic recombination pathway (Schiestl and Prakash, 1990; Bailly *et al.*, 1992; Bardwell *et al.*, 1992). In addition, the *ERCC3* gene product, responsible for the rare XP complementation group B, was recently uncovered as one of the components of the multisubunit transcription initiation factor BTF2 (TFIIH) (Schaeffer *et al.*, 1993). This finding disclosed an unexpected functional overlap between basal transcription and NER. It is possible that the entire BTF2 transcription complex is involved in NER. The ERCC1 and ERCC3 complexes play a role in both transcription-coupled as well as genome-overall repair and are thus implicated in the core of the NER reaction mechanism (Hoeijmakers, 1993b). The XPCC-HHR23B complex is the first to be described which appears to be specific for the genome-overall subpathway. In view of the tight link between transcription and NER, the function of this complex could be to uncouple the NER machinery from the basal transcription proces, enabling it to scan the non-transcribed bulk of the genome for the presence of lesions. The availability of the protein complex and in vitro NER systems provide the necessary tools to investigate the function(s) of this NER component.

MATERIALS AND METHODS

Cells and cell culture

Five SV40-transformed fibroblast lines XP2OSSV (group A), XP4PASV (group C), XP6BESV (group D), XP2YOSV (group F) and XP3BRSV (group G), three non-transformed fibroblast lines CRL1199 (group B), XP3KA (group C) and XP2RO (group E), and repair-proficient lines 293 cells were cultured at 37°C in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. HeLa cells were grown in spinner flasks at 37°C in RPMI 1640 medium supplemented with 5% calf serum and harvested at a density of 10^5 cells/ml.

Preparation of whole cell extracts

The 293 cell line was grown at 37°C in 150 mm tissue culture plates (Falcon), treated with phosphate-buffered saline containing 0.05% Na_3EDTA and collected by gentle pipetting. XP cells were grown in 850 cm^2 roller bottles (Corning) and collected by scraping. The harvested cells were washed with phosphate-buffered saline, and whole cell extracts were prepared as described previously (Manley *et al.*, 1983; Wood *et al.*, 1988). Protein concentration was determined by the method of Bradford (1976) with bovine serum albumin as standard. Extracts contained 10-20 mg of protein/ml.

Preparation of SV40 minichromosomes and plasmid DNA

SV40 virions were prepared as described previously (Sugasawa *et al.*, 1993). Minichromosomes were obtained by alkali disruption of the SV40 virions as described (Christiansen *et al.*, 1977) and irradiated with 200 J/m^2 of UV light (254 nm) as described previously (Sugasawa *et al.*, 1993).

Plasmid pUC19 DNA was propagated in *E.coli* strain HB101. Closed circular DNA was prepared by the alkali lysis method and CsCl-ethidiumbromide equilibrium density gradient centrifugation as described (Sambrook et al., 1989). In our previous studies, we used a plasmid DNA sample prepared with a single CsCl centrifugation step (Masutani et al., 1993; Sugasawa et al., 1993). In these previous studies we observed a significant level of DNA synthesis with unirradiated pUC19 DNA. We found that the DNA preparations contained detectable amounts of nicked molecules. These molecules were likely to be used as a template for DNA synthesis, because on repeating the CsCl centrifugation one or two more times, the UV-independent DNA synthesis on pUC19 DNA decreased in proportion to reduction in the amount of nicked molecules. Therefore, in the present study we repeated CsCl centrifugation several times.

Cell-free DNA repair assay

The standard reaction mixture (20 μl) contained 40 mM creatine phosphate-Tris (pH 7.7), 1 mM dithiothreitol, 10 mM $MgCl_2$, 2 mM ATP. 50 μM each of dATP, dGTP and dTTP, 10 μM [α-^{32}P]dCTP (37-74 kBq), phosphocreatine kinase (Sigma, Type I; 0.5 μg), bovine serumalbumin (6.4 μg), whole cell extracts (80 μg of protein), unirradiated pUC19 RFI DNA (0.3 μg) and UV-irradiated (200 J/m^2) or unirradiated SV40 mini-chromosomes (0.3 μg). The reaction was performed at 30°C for 3 h. The products were purified from the reaction mixtures, linearized with *Eco*RI and electrophoresed in a 1% agarose gel as described previously (Sugasawa et al., 1993). Autoradiography was performed at -80°C with Fuji New RX X-ray film. The incorporation of radioactive materials into UV-irradiated or unirradiated SV40 minichromosomes was quantified with a FujixBAS2000 Bio-Imaging Analyzer.

Purification of XP-C correcting protein from HeLa cells

All procedures were carried out at 0-4°C. The purification is summarized in Figure 1 and Table I. A frozen stock of $5x10^{10}$ HeLa cells 176 ml of packed cell volume) was thawed, washed once with hypotonic buffer [10mm Tris-HCl (pH7.5), 1 mM Na_3EDTA, 2 mM $MgCl_2$, 5 mM dithiothreitol, 0.25 mM PMSF, 0.2 μg/ml aprotinin, 0.2 μg/ml leupetin, 0.1 μg/ml antipain, and 50 μM EGTA], suspended in 700 ml of hypotonic buffer and homogenized in an all-glass Dounce homogenizer by 15 strokes with a pestle A. The nuclei were obtainedby low speed centrifugation, washed twice with nuclei wash buffer [10 mM potassium phosphate (pH7.5), 1 mM Na_3EDTA, 2 mM dithiothreitol, 0.25 mM PMSF, 0.2 μg/ml aprotinin, 0.2 μg/ml leupeptin, 0.1 μg/ml antipain and 50 μg EGTA] and then suspended in 380 ml of buffer 1 [20 mM potassium phosphate (pH7.5), 1 mM Na_3EDTA, 5 mM dithiotreitol, 0.25 mM PMSF, , 0.2 μg/ml aprotinin, 0.2 μg/ml leupeptin, 0.1 μg/ml antipain and 50 μM EGTA]. A suspension was made in 0.3 M KCl by the addition of 0.1 vol of buffer 1 containing 3.3 M KCl. An extract was obtained by gentle stirring for 30 min followed by centrifugation for 1 h at 100.000 *g*. The supernatant was dialysed against buffer 2 [20 mM potassium phosphate (pH 7.5), 1 mM Na_3EDTA, 10% glycerol, 1 mM dithiotreitol , 0.01% Triton X-100, 0.25 mM PMSF, 0.2 μg/ml aprotinin, 0.2 μg/ml leupeptin, 0.1 μg/ml antipain and 50 mM EGTA] containing 0.15 M KCl and centrifuged for 1 h at 100.000 *g*. The supernatant (nuclear extract) was loaded onto a phosphocellulose column (Whatman P11; 90 ml) equilibrated with buffer 2 containing 0.15 M KCl. The column was washed with three column volumes of the same buffer and the adsorbed proteins were eluted with buffer 2 containing 1 M KCl. The eluate was loaded onto a single-stranded DNA-cellulose column (Sigma; 4.3 mg DNA/g cellulose; 6 ml) equilibrated with buffer 2 containing 0.6 M KCl. The column was washed with three column volumes of the same buffer and the adsorbed proteins were eluted with buffer 2 containing 1.5 M KCl. The eluate was dialysed against buffer 2 containing 0.3 M KCl and adjusted to 0.3 M KCl by dilution with buffer 2. The following two steps were performed with an FPLC system. The dialysate was loaded onto a column of CM cosmogel (Nakalai tesque; 8 mm ID x 75 mm) equilibrated with buffer 2 containing 0.3 M KCl. The column was washed with 10 ml of the same buffer and then proteins were eluted with buffer 2 containing 0.6 M KCl. The eluate was adjusted to 0.15 M KCl by diluting with buffer 2 and promptly loaded onto a column of MonoQ HR5/5 (Pharmacia) equilibrated with buffer 2 containing0.15 M KCl. The column was washed with 10 ml of the same buffer and then proteins were eluted with 25 ml of a linear gradient of 0.15 to 0.45 M KCl. in buffer2. XP-C correcting activity was eluted with ≈ 0.29 M KCl. The active fractions were pooled and stored at -80°C. A portion of the active fraction was dialysed against buffer 1 containing 0.2 M KCl. and 50% glycerol, and stored at -20°C. In both pools the XP-C correcting activity was stable for at least 3 months. The XP-C protein could be obtained by another purification procedure in which Tris-HCl (pH 7.5) and NaCl were used instead of potassium phosphate (pH 7.5) and KCl, respectively (data not shown).

XP-C correcting activity was assayed with XP4PASV cell extract in standard conditions. One unit of XP-C correcting activity was defined as the amount of protein required to increase the XP4PASV cell extract-mediated incorporation of 1 pmol of dCMP into UV-irradiated SV40 minichromosomes. As the incorporation of dCMP reached a maximum at 100-150 fmol in standard conditions, units of activity were determined at the order of 10^{-2} by titration.

Gel filtration of XP-C correcting protein

A portion (80 μl) of the MonoQ fraction was loaded onto a Sephacryl S-300 column (6 mm x 82 cm) equilibrated with buffer 2 containing 0.3 M KCl and run at 3 ml/h. Fractions (250 μl) were collected and used for assay of XP-C correcting activity and SDS-PAGE. Marker proteins were loaded in identical conditions and detected by SDS-PAGE followed by staining with Coomassie brilliant blue.

Glycerol density gradient centrifugation of XP-C correcting protein

A portion (60 μl) of the Mono Q fraction was layered on 4.8 ml of a 15-35% (v/v) glycerol gradient in buffer 1 containing 0.3 M KCl and centrifuged in a Hitachi RPS65T rotor at 260.000 *g* for 22 h at 2°C. Fractions (200 μl) were collected from the top of the gradient and assayed for XP-C correcting activity. An identical gradient containing marker proteins was run at the same time. The markers were detected by SDS-PAGE followed by staining with Coomassie brilliant blue.

Assays of enzyme activities

DNA polymerase activity was assayed with activated DNA as template as described previously (Suzuki *et al.*, 1989). The MonoQ fraction of the XP-C correcting factor (60 ng) was incubated at 37°C for 2 h in 30 μl ofa solution of 40 mM Tris-HCl (pH 8.0), 1 mM dithiotreitol, 10 mM $MgCl_2$, 2 mM ATP, 50 μM each of dATP, dGTP and dTTP, 10 μM [α-^{32}PdCTP (74 kBq), 0.32 mg/ml bovine serum albumin and 0.5 mg/ml of activated DNA. The reaction was terminated by chilling on ice and the radioactivity incorporated into acid-insolluble materials was measured.

DNA helicase activity was assayed as oligomer displacing activity. The MonoQ fraction (60 ng) was incubated at 37°C for 1 h in 20 μl of a solution of 50 mM Tris-HCl (pH 7.5), 20 mM 2-mercaptoethanol, 5 mM $MgCl_2$, 5 mM ATP, 0.5 mg/ml bovine serum albumin and 0.017 pmol of 5'^{32}P-labelled 21mer annealed to M13 DNA. After termination of the reaction, products were analysed by polyacrylamide (12%) gel electrophoresis followed by autoradiography as described previously (Yanagisawa *et al.*, 1992).

Exonuclease activities were detected in the DNA helicase assay by monitoring the amounts of labelled oligomers and their sizes.

DNA ligase activity was assayed indirectly with bacterial alkaline phoshatase. For this, 60 ng of the Mono Q fraction were incubated at 37°C for 2 h in 30 μl of a solution of 40 mM Tris-HCl (pH 7.5), 1 mM dithiotreitol, 10 mM $MgCl_2$, 2 mM ATP, 0.32 mg/ml bovine serum albumin and 50 ng of 5'[^{32}P]oligo(dT)$_{12\text{-}18}$-poly(dA)$_{400}$(1:5). Then 0.4 unit of bacterial alkaline phosphatase (Takara) was added and after incubation at 65°C for 1 h, the radioactivity remaining in the acid insoluble material was measured.

Endonuclease activities were measured as nicking activities with UV-irradiated or unirradiated closed circular form I pUC19. The Mono Q fraction (60 ng) was incubated at 37°C for 2 h in 20 μ l of solution containing 40 mM Tris-HCl (pH 8.0), 1 mM dithiotreitol, 10 mM $MgCl_2$, 2 mM ATP, 0.32 mg/ml bovine serum albumin and 0.1 μg of UV irradiated (500 J/m^2) or unirradiated closed circular form I pUC19. After the reaction, the plasmids were subjected to 1% agarose gel electrophoresis and detected by ethidiumbromide staining.

SDS-PAGE

SDS-PAGE was performed by the method of Laemmli (1970).

Determination of partial amino acid sequences

The Mono Q fractions of the purified XP-C correcting protein were adjusted at 6 M guanidine-HCl and 10 mM sodium phosphate (pH 6.0) and subjected to gel filtration using tandemly joined TSK G3000SW_{XL} and TSK G4000SW_{XL} columns (Tosoh; 7.8 x 300 mm ea.) and a Gilson HPLC system at a flow rate of 0.5 ml/min. Protein peaks corresponding to the 125 and 58 nkDa polypeptides were collected separately and digested with CNBr after removal of salts. The digests were applied to an Aquapore RP300 column (Applied Biosystems; 2.1 x 100 mm) and eluted with a linear gradient of 0.09% TFA to 80% acetonitrile-0.075% TFA in 40 min at a flow rate of 0.2ml/min. Materials in clearly isolated peptide peaks were collected and applied to a protein sequencer (Applied Biosystems; model 477A/120A).

Screening of cDNA libraries

For isolation of cDNA clones encoding p125, a cDNA library with relatively long inserts was constructed. Complementary DNAs were synthesized from 5 μg of HeLa cell poly(A)$^+$ RNA using a cDNA synthesis kit (Pharmacia). After addition of *Eco*RI-*Not*I adaptors and size-fractionation by agarose gel electrophoresis, double-stranded cDNAs of >2.5 kb were eluted from the gel and ligated to an *Eco*RI-digested λgt10 vector. Some of the recombinant DNAs were packaged in vitro into bacteriphage particles, then amplified in *E.coli* strain, C600 hflA. The resulting cDNA library contained 8.8 x 10^5 independent clones.

To obtain a probe for screening the cDNA library, RT-PCR was carried out using synthetic oligonucleotide mixtures and first-strand cDNA synthesized from HeLa cell poly(A)$^+$ RNA. The sequences of the oligonucleotides used were 5'-GCI(C/A)GIAA(A/G)- (C/A)GIGCIGCIGGIGGIGA-3' and 5'-(T/C)TT(T/C)TTIGGIGG(T/C)TT(T/C)TC(A/G)-TC(T/C)TC(A/G)AA-3', where I indicates inosine. PAGE revealed amplification of 132 bp DNA fragments, which were then purified from the gel and cloned into pUC19 DNA for sequencing. Since the sequence of the 132 fragment was consistent with the determined amino acid sequence, this fragment was reamplified from the plasmid, gel-purified and used for screening the cDNA library.

About one million recombinant bacteriophage plaques were transferred to Hybond-N membranes (Amersham) in duplicate. Prehybridization was carried out at 68°C for 4 h in 6x SSC (1x SSC: 0.15 M NaCl, 15 mM sodium citrate), 4x Denhardt's solution (1x Denhardt's solution: 0.02% Ficoll 400, 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone) and 50 mg/ml heat-denatured salmon sperm DNA. Hybridization was performed at 42°C overnight in 30% formamide, 4x SSC, 4x Denhardt's solution, 50 μg/ml heat-denatured salmon sperm DNA and the DNA probe radiolabelled with [α-^{32}P]dCTP and a multiprime DNA labelling system (Amersham). The membranes were successively washed at room temperature for 10 min and at 55°C for 10 min with 2x blot wash buffer (1x blot wash buffer: 1x SSC, 10 mM sodium phosphate, 0.025% SDS), at 55°C for 10 min with 1x blot wash buffer, at 55°C for 30 min with 0.5x blot wash buffer, at 55°C for 30 min with 0.2x blot wash buffer and twice at 65°C for 30 min 0.1x blot wash buffer. Then the membranes were air-dried and exposed at -80°C to Kodak X-OMAT film with intensifying screens. A positive plaque was picked up and purified by another round of plaque hybridization.

The 3.6 kb insert of the positive clone was obtained by *Not*I digestion and subcloned into the *Not*I site of pBluescript II KS$^+$. Deletion mutants were constructed by use of exonuclease III and mung bean nuclease (a deletion kit for kilo-sequencing; Takara Shuzo), and sequenced with a Taq Dye Deoxy Terminator cycle sequencing kit and an automated DNA sequencer (Applied Biosystems, model 373A).

For isolation of cDNA clones encoding the p58, an oligonucleotide, 5'-CCICCICC-ICC(C/T)TGICCICCIGC(C/T)TC(C/T)TGIACIGG(C/T)TC(A/G)TT-3', was used for screening a λgt10 cDNA library from HeLa cells. Screening was performed as described above. The 2.9 kb insert of the positive clone was obtained by *Eco*RI digestion and subcloned into the *Eco*RI site of pUC19. Deletion mutants were constructed and sequenced as described above.

Cloning and nucleotide sequence analysis of HHR23A

Total RNA (10 μg) was used for preparing cDNA with *HHR23A*-specific primers (see below). RNA was dissolved in 9 μl of annealing buffer [250 mM KCl, 10 mM Tris-HCl (pH 8.3), 1 mM EDTA]. Following the addition of 1 μl (100 pmol/μl) of primers, the samples were fist heated for 3 min at 80°C and transferred to a 37°C water bath for 1 h. Fifteen microlitres of cDNA buffer (24 mM Tris-HCl [pH 8.3], 16 mM $MgCl_2$, 8 mM DTT, 0.4 mM of GTP, dATP, dTTP, dCTP) and 5U of Moloney leukemia virus reverse transcriptase (Promega) were added and the tube was incubated at 37°C for 1 h. To 5 μl cDNA, 10 μl of Taq buffer [100 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 500 mM KCl, 2 mg/ml bovine serum albumin], 4 μl dNTPs (2.5 mM), 75 μl water, 1 μl of each primer (100 pmol/μl and 2U of Taq polymerase (Cetus) were added.

Oligonucleotide primers for cDNA, DNA amplification and DNA sequencing were synthesized in an Applied Biosystems DNA synthesizer. The PCR primers used for this purpose are: 5'-ATCCAGATGCTGAACGAGCC-3' and 5'-CGGCAGGTGATTCAGCA
GAAC-3'.

A PCR probe was used to screen a pre-B cell library and clones hybridizing with the PCR probe were picked up and examined by restriction enzyme analysis. Hybridization of human probes to human DNA was at 65°C in a hybridization mixture containing 10x Denhardt's solution, 10% dextransulfate, 0.1% SDS, 3xSSC, 50 mg of sonicated salmon sperm DNA per litre. Washings were performed twice for 20 min each in 0.3xSSC at 65°C. Hybridization was detected by autoradiography on Fuji medical X-ray film RX with intensifying screens at -80°C.

Lambda zap phages (Short *et al.*, 1988) were after two rounds of rescreens converted into Bluescript vectors and transformed to competent DH5αF' cells. Sequence analysis on double-stranded DNA was done by the T7 DNA polymerase modification (Pharmacia) of the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) using sequence- derived oligonucleotides prepared for sequencing both strands. For separation of the fragments, Hydrolink (AT Biochem, Malvern, PA) sequencing gels were used.

ACKNOWLEDGEMENTS

We are grateful to Drs M. Sekiguchi, A. Sarai, Christine Troelstra and Marcel Koken for their helpful suggestions, Sigrid Swagemakers for help with some of the experiments and to Satya Prakash (Galveston) who kindly provided us with the yeast *RAD23* gene sequence. Furthermore, we thank Dr. Jack Leunissen from the Nijmegen University COAS/CAMM center for his assistance with the molecular modelling packages. This work was supported by grants from the Ministry of Education, Science and Culture of Japan, the Biodesign Research Program of the Institute of Physical and Chemical Research (RIKEN), and the Cosmetology Research Foundation of Japan. The work at the Department of Cell Biology and Genetics of the Erasmus University was financially supported by the Medical Genetics Centre South-West Netherlands.

REFERENCES

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelly, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992) *Nature*, **355**, 632-634.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*,**215**, 403-410.

Bailly, V., Sommers, C.H., Sung, P., Prakash, L. and Prakash, S. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 8273-8277.

Banerji, J., Sands, J., Strominger, J.L. and Spies, T. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2374-2378.

Bardwell, L., Cooper, A.J. and Friedberg, E.C. (1992) *Mol. Cell Biol.*, **12**, 3041-3049.

Biggerstaff, M., Szymkowski, D.E. and Wood, R.D. (1993) *EMBO J.*, **12**, 3685-3692.

Bradford, M.M. (1976) *Anal. Biochem.*, **72** 248-254.

Christiansen, G., Landers, T., Griffith, J. and Berg, P. (1977) *J.Virol.*, **21**, 1079-1084.

Cleaver, J.E. (1968) *Nature*, **218**, 652-656.

Cleaver, J.E. and Kraemer, K.H. (1989) In Scriver,C.R., Beaudet,A.L., Sly,W.A. and Valle, D. (eds), *Xeroderma Pigmentosum. The Metabolic Basis of Inherited Disease.* Vol.II, McGraw-Hill BookCo., New York, pp.29449-2971.

Collins, A.R. (1993) *Mutat. Res.*, **293**, 99-118.

Finley, D., Bartel, B. and Varshavsky, A. (1989) *Nature*, **338**, 394-401.

Fleer, R., Nicolet, C.M., Pure, G.A. and Friedberg, E.C (1987) *Mol. Cell. Biol.*, **7**, 1180-1192.

Flejter, W.L., McDaniel, L.D. Johns, D., Friedberg, E.C. and Schultz, R.A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 261-265.

Friedberg, E.C. (1985) *DNA Repair.* W.H.Freeman and Co., New York.

Friedberg, E.C. (1988) *Microbiol. Rev.*, **52**, 70-102.

Gietz, R.D. and Prakash, S. (1988) *Gene*, **74**, 535-541.

Hanawalt, P. and Mellon, I. (1993) *Curr Biol.*, **3**, 67-69.

Hoeijmakers, J.H.J. (1993a) *Trends Genet.*, **9**, 173-177.

Hoeijmakers, J.H.J. (1993b) *Trends Genet.*, **9**, 211-217.

Huang, J.-C., Svoboda, D.L., Reardon, J.T. and Sancar, A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3664-3668.

Jentsch, S. (1992) *Annu Rev. Genet.*, **26**, 179-207.

Kantor, G.J., Barsalou,L.S. and Hanawalt,P.C. (1990) *Mutat. Res.*, **235**, 171-180.

Kas, K., Michiels, L. and Merregaert, J. (1992) *Biochem. Biophys. Res. Commun.*, **187**, 927-933.

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.

Kozak, M. (1991) *J. Biol. Chem.*, **266**, 19867-19870.

Kraemer, K.H., Lee, M.M. and Scotto, J. (1987) *Arch. Dermatol.*, **123**, 241-250.

Kumar, S., Tomooka, Y. and Noda, M. (1992) *Biochem. Biophys. Res. Commun.*, **185**, 1155-1161.

Laemmli, U.K. (1970) *Nature*, **227**, 680-685.

Legerski, R. and Peterson, C. (1992) *Nature*, **359**, 70-73.

Li, L., Bales, E.S., Peterson, C.A. and Legerski,R.J. (1993) *Nature Genet.*, **5**, 413-417.

Linnen, J.M., Bailey, C.P. and Weeks, D.L. (1993) *Gene*, **128**, 181-188.

Madura, K. and Prakash, S. (1990) *Nucleic Acids Res.*, **18**, 4737-4742.

Manley, J.L., Fire, A., Samuels, M. and Sharp, P.A. (1983) *Methods Enzymol.*, **101**, 568-582.

Masutani, C., Sugasawa, K., Asahina, H., Tanaka, K. and Hanaoka, F. (1993) *J. Biol. Chem.*, 268, 9105-9109.

Melnick, L. and Sherman, F. (1993) *J. Mol. Biol.*, **233**, 372-388.

O'Donovan, A. and Wood, R.D. (1993) *Nature*, **363**, 185-188.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y. and Matsubara, K. (1992) *Nature Genet.*, **2**, 173-179.

Perozzi, G. and Prakash, S. (1986) *Mol. Cell. Biol.*, **6**, 1497-1507.

Riboni, R., Botta, E., Stefanini, M., Numata, M. and Yasui, A. (1992) *Cancer Res.*, **52**, 6690-6691.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sancar, A. and Hearst, J.E. (1993) *Science*, **259**, 1415-1420.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H.J., Chambon, P. and Egly, J.-M. (1993) *Science*, **260**, 58-63.

Schiestl, R.H. and Prakash, S. (1990) *Mol.Cell. Biol.*, **10**, 2485-2491.

Short, J.M., Fernandez, M., Sorge, J.A. and Huse, W.D. (1988) *Nucleic Acids Res.*, **16**, 7583-7600.

Sibghat-Ullah, Husain, I., Carlton, W. and Sancar, A. (1989) *Nucleic Acids Res.*, **17**, 4471-4484.

Sugasawa, K., Masutani, C and Hanaoka, F. (1993) *J. Biol. Chem.*, **268**, 9098-9104.

Suzuki, M., Enemoto, T., Masutani, C., Hanaoka, F., Yamada, M. and Ui, M. (1989) *J. Biol. Chem.*, **264**, 10065-10071.

Tanaka, K., Miura, N., Satokata, I., Miyamoto, I., Yoshida, M.C., Satoh, Y., Kondo, S., Yasui, A., Okayama, H. and Okada, Y. (1990) *Nature*, **348**, 73-76.

Toniolo, D., Persico, M. and Alcalay, M. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 851-855.

Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D. and Hoeijmakers, J.H.J. (1992) *Cell*, **71**, 939-953.

van Duin, M., Vredeveldt, G., Mayne, L.V., Odijk, H., Vermeulen, W., Klein, B., Weeda, G., Hoeijmakers, J.H.J., Bootsma, D. and Westerveld, A. (1989) *Mutat. Res.*, **217**, 83-92.

Van Houten, B. (1990) *Microbiol. Rev.*, **54**, 18-51.

van Vuuren, A.J., Appeldoorn, E., Odijk, H., Yasui, A., Jaspers, N.G.J., Bootsma, D. and Hoeijmakers, J.H.J. (1993) *EMBO J.*, **12**, 3693-3701.

Venema, J., van Hoffen, A., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H.F. (1990) *Nucleic Acids Res.*, **18**, 443-448.

Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H.F. (1991) *Mol. Cell. Biol.*, **11**, 4128-4134.

Vijay-Kumar, S., Bugg, C.E., Wilkinson, K.D., Vierstra, R.D., Hatfield, P.M. and Cook, W.J. (1987) *J. Biol. Chem.*, **262**, 6396-6399.

Watkins, J.F., Sung, P., Prakash, L. and Prakash, S. (1993) *Mol. Cell. Biol.*, **13**, 7757-7765.

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.

Wood, R.D., Robins, P. and Lindahl, T. (1988) *Cell*, **53**, 97-106.

Yanagisawa, J. Seki, M., Ui, M. and Enomoto, T. (1992) *J. Biol. Chem.*, **267**, 3585-3588.

CHAPTER

V

Chromosomal localization of three repair genes: The xeroderma pigmentosum group C gene and two human homologs of yeast RAD23

CHROMOSOMAL LOCALIZATION OF THREE REPAIR GENES: THE XERODERMA PIGMENTOSUM GROUP C GENE AND TWO HUMAN HOMOLOGS OF YEAST RAD23

P.J. van der Spek[1], E.M.E Smit[1], H.B. Beverloo[1], K. Sugasawa[2], C. Masuitani[2], F. Hanaoka[2], J.H.J. Hoeijmakers[1] and A. Hagemeijer[1].

[1] Department of Cell Biology and Genetics, Medical Genetic Centre, Erasmus University, Rotterdam, The Netherlands; and: [2] Cellular Physiology Laboratory, The Institute of Physical and Chemical Research (RIKEN), Wako-shi, Saitama 351-01, Japan.

GENOMICS 23: 651-658, 1994.

ABSTRACT

The nucleotide excision repair (NER) disorder xeroderma pigmentosum (XP) is characterized by sun (UV) sensitivity, predisposition to skin cancer, and extensive genetic heterogeneity. Recently, we reported the cloning and analysis of three human NER genes, *XPC*, *HHR23A*, and *HHR23B*. The previously cloned *XPC* gene is involved in the common XP complementation group C, which is defective in excision repair of non-transcribed sequences in the genome. The XPC protein was found to be complexed with the product of *HHR23B*, one of the two human homologs of the *Saccharomyces cerevisiae* NER gene *RAD23*. Here we present the chromosomal localization by *in situ* hybridization using haptenized probes of all three genes. The *HHR23A* gene was assigned to chromosome 19p13.2. Interestingly, the *HHR23B* and *XPC* genes, the product of which forms a tight complex, were found to colocalize on band 3p25.1. Pulsed-field gel electrophoresis revealed that the *HHR23B* and *XPC* genes possibly share a *Mlu*I restriction fragment of about 625 kb. Potential involvement of the *HHR23* genes in human genetic disorders is discussed.

INTRODUCTION

The integrity of the DNA is under constant assault by genotoxic agents, such as ultraviolet light, X rays, and numerous chemical compounds that can damage the genetic material. A network of repair systems has evolved to minimize the deleterious effects of DNA injury. One of these pathways, the nucleotide excision repair (NER) process, removes a broad range of DNA lesions, such as UV-induced cyclobutane pyrimidine dimers and (6-4) photoproducts, bulky chemical adducts, and certain DNA crosslinks in a multienzyme reaction (Hoeijmakers, 1993a,b). Two NER subpathways can be discerned: a rapid and efficient repair of the transcribed strand of active genes (transcription-coupled repair) and a more slow and less efficient repair of the bulk DNA, designated herein genome overall repair (Bohr, 1991; Hanawalt and Mellon, 1993). The enzymes involved in NER play a role in inherited diseases such as xeroderma pigmentosum (XP), Cockayne syndrome (CS) and PIBIDS, in which the excision repair mechanism is defective.

The autosomal recessive disorder XP is clinically characterized by extreme sensitivity of the skin to sunlight (UV), sunlight-induced pigmentation abnormalities, and predisposition to skin cancer. Frequently, neurological complications are seen due to progressive neurodegeneration (for a review see Cleaver and Kraemer, 1994). Cell fusion

experiments have identified at least seven excision-deficient XP complementation groups (designated XP-A to XP-G) (Vermeulen *et al.*, 1991) in addition to a form of XP, called XP-variant, that is defective in postreplication repair (Lehmann *et al.*, 1975). This indicates involvement of a minimum of seven distinct NER genes in XP. Most of the XP complementation groups are defective in both NER subpathways, the overall genome repair and the transcription-coupled repair. However, patients belonging to XP group C, one of the most common complementation groups, are only defective in genome overall repair, and proficient in the removal of lesions from the transcribed strand of active genes (Venema *et al.*, 1991), indicating that specific factors are implicated in this subpathway of excision repair.

CS patients exhibit sun sensitivity, dwarfism, microcephaly, wizened appearance, deafness, and severe mental retardation. The neurological symptoms in this disorder are related to neurodysmyelination. CS is, unlike XP, not associated with an elevated risk for skin tumor formation (Lehmann, 1987). Two complementation groups have been identified within the classical form of the disease: CS-A and CS-B (Tanaka *et al.*, 1981; Lehmann, 1982). CS cells were found to be selectively disturbed in the transcription-coupled repair subpathway of NER (Venema *et al.*, 1990).

A third recently discovered NER disorder is PIBIDS, an acronym for: Photosensitivity, Ichthyosis, Brittle hair and nails, Impaired intelligence, Decreased fertility, Short stature (Stefanini *et al.*, 1993). Brittle hair and nails are hallmarks of trichothiodystrophy (TTD), a much broader genetic disease that includes PIBIDS. In many clinical aspects CS and PIBIDS resemble each other (Bootsma and Hoeijmakers, 1993, Hoeijmakers 1993b). At least two complementation groups have been described, one of which overlaps with XP-D (Stefanini *et al.*, 1993). Finally, rare cases have been identified displaying simultaneously the clinical hallmarks of XP and CS. These patients are assigned to XP groups B, D, and G (Hoeijmakers, 1993b). Recently, many of the clinical features of CS and TTD have been ascribed to subtle defects in the vital process of basal transcription, as the proteins affected appear to be involved in NER as well as in transcription initiation (Bootsma and Hoeijmakers, 1993, Hoeijmakers 1993b, Vermeulen *et al.* submitted for publication).

A second class of mammalian excision repair-deficient mutants is represented by laboratory-induced, UV-sensitive, rodent cell lines. Eleven complementation groups have been identified (Busch *et al.*, 1994; Riboni *et al.*, 1992). Human genes correcting these rodent mutants are designated *e*xcision *r*epair *c*ross *c*omplementing (*ERCC*) genes. Biochemical and genetic analyses of mammalian and yeast NER genes and proteins have

revealed that the entire NER pathway is strongly conserved in eukaryotic evolution (reviewed in Hoeijmakers 1993a,b).

Recently, we described the purification of a NER protein complex consisting of the 125 kD XPC gene product and a 58 kD protein with overall homology to the product of the *Saccharomyces cerevisiae RAD23* NER gene (Masutani et al., 1994). Simultaneously, we cloned a closely related second homolog of the yeast gene designated *HHR23A* (for *H*uman *H*omolog of *RAD23*), whereas the former was called *HHR23B*. This represents the first example of a NER gene duplication during eukaryotic evolution. The *RAD6* gene encoding an ubiquitin-activating enzyme involved in postreplication repair is also duplicated in mammals (Koken, *et al.*, 1992). RAD23 mutants show a partial defect in excision repair and the encoded protein begins with a strongly conserved ubiquitin-like domain that is essential for its repair function (Watkins, *et al.*, 1993). The XPC/HHR23B complex displays a strong affinity for ssDNA and appears to be selectively involved in the genome-overall NER subpathway. Here we report the chromosomal localization of these genes.

MATERIALS AND METHODS

In situ hybridization. Normal human lymphocytes were used for the preparation of metaphase spreads prior to *in situ* hybridization.

In situ hybridization experiments using the *XPC* cDNA in a pBluescript vector, the *HHR23A* genomic phage IV, the *HHR23B* genomic phage II, the *HHR23B* cDNA in a pBluescript vector , and the chromosome 19 centromer-specific marker, pG-A16 (Chérif *et al.*, 1990), as biotin-labeled or digoxigenin-labeled probes were performed as described elsewhere (Pinkel *et al.*, 1986).

After incubation with avidin D-FITC (Vector, USA), the biotin-labeled probes were visualized. The fluorescent signal was amplified with biotinylated goat anti-avidin D. The digoxigenin-labeled probes were visualized by incubation with sheep anti-digoxigenin TRITC followed by amplification with Donkey anti-Sheep Texas-Red conjugates (Fab fragments, Boehringer Mannheim). After immunochemical staining, the slides were dehydrated with ethanol, and air-dried. The slides were counterstained with propidium iodide and 4',6'-diamidino-2-phenylindole (DAPI) in antifade medium.

In case of hybridization with *HHR23A*, slides were banded with bisbenzimide H33258 (Hoechst), UV-irradiated and heat-denatured before amplification (Chérif *et al.*, 1990).

General procedures. Isolation, digestion and gel electrophoresis of the genomic λ clones hybridizing to the *HHR23A* and *HHR23B* cDNAs were performed according to established procedures (Sambrook *et al.*, 1989). Labeling of DNA probes, and hybridizations of Southern blots were carried out using routine protocols. Southern blotting to Zeta probe membranes was performed by alkaline transfer, as described by the manufacturer (Bio-Rad, Richmond, CA). Membranes were exposed at -80°C to Fuji RX

film with intensifying screens. After exposure, blots were stripped in 10 mM Tris, 1 mM EDTA, 1% SDS at 90°C for 5 min and rehybridized.

The *HHR23* genomic phages were derived from a λ EMBL-3 library prepared from genomic DNA of the CML-0 cell line (generously provided by Dr. G. Grosveld). Phages were used to infect *E.coli* LE392 cells. Analysis and identification of *HHR23* genomic fragments were carried out by restriction enzyme site mapping and hybridization using *HHR23* cDNA probes.

DNA for restriction fragment length polymorphism (RFLP) analysis was isolated from peripheral blood leucocytes. For pulsed field gel electrophoresis (PFGE), agarose-embedded leucocytes of a normal individual were lysed and digested with the appropriate restriction endonucleases according to the manufacturers' instructions. PFGE was carried out as detailed elsewhere (van Ommen and Verkerk, 1986).

RESULTS

In Situ Hybridization

HHR23A. Hybridization with the *HHR23A* genomic probe yielded a clear hybridizing signal with a chromosome that on the basis of Hoechst banding can be identified as human chromosome 19, in the area close to the p13.3-p13.2 border (Fig.1A). To verify the chromosome identification, the chromosome 19-specific centromere probe pG-A16 was used in combination with the *HHR23A* genomic probe. Figure 2 shows that the *HHR23A* signals reside on the same chromosome as that of the centromere probe. As an independent confirmation of the assignment of *HHR23A* to the p-arm of chromosome 19, we performed also simultaneous hybridization with a genomic probe of another known chromosome 19 gene, *ERCC1*, located on 19q13.2 (Mohrenweiser *et al.,* 1989). The results obtained (not shown) were in complete agreement with the localization of the *HHR23A* gene to 19p13.2. in every metaphase analysed.

HHR23B. For mapping of the *HHR23B* locus, *in situ* hybridization experiments were performed on metaphase spreads using the biotinylated *HHR23B* cDNA in a pBluescript vector as well as the genomic λ phage. The *HHR23B* genomic probe II contains a 7 kb *Eco*RI fragment present in the 3' end of the *HHR23B* gene. In agreement with the cDNA, the genomic probe gave specific hybridization on 3p25.1. (See Fig. 1B). Since unequivocal identification of chromosome 3 is possible on the basis of morphology and banding pattern, no double hybridizations with a control probe were performed. A representative example of chromosome 3 showing hybridization and the clear morphology in combination with the banding pattern is depicted in Fig.2.

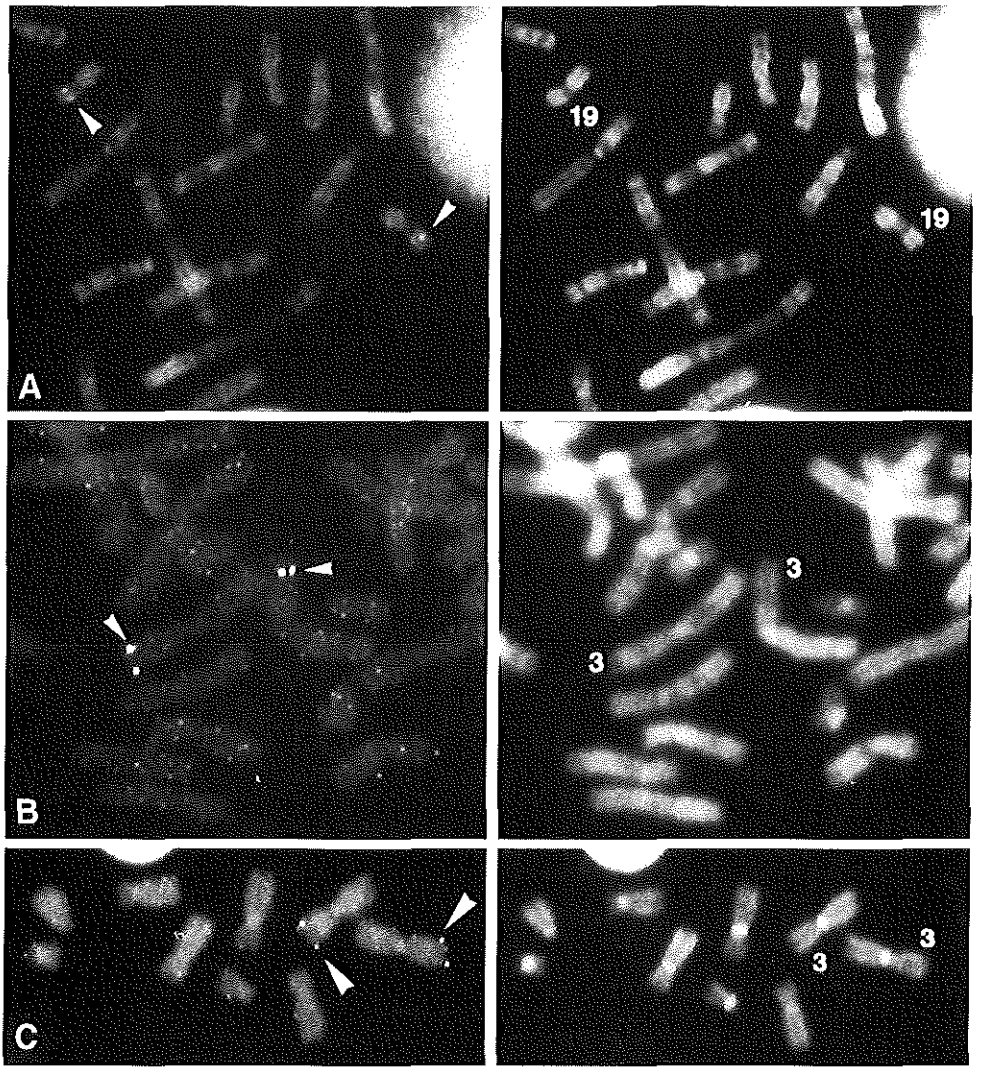


FIG.1 (A) *In situ* hybridization of metaphase chromosomes with biotinylated genomic *HHR23A* probe. Hybridization on a metaphase spread (partly shown) with the genomic *HHR23A* probe. The arrowheads point to the region with a specific signal on chromosome 19p13.2. (Left) The *in situ* hybridization results. (Right) The Hoechst banding of the same metaphases. (B) *In situ* hybridization of metaphase chromosomes with biotinylated genomic *HHR23B* probe. The arrowheads indicate the hybridization signal on chromosome 3p25.1. (Left) *In situ* hybridization results. (Right) The DAPI banding of the same metaphase. (C) *In situ* hybridization of metaphase chromosomes with biotinylated *XPC* cDNA probe. The arrowheads indicate the hybridization signal on chromosome 3p25.1. (Left) *In situ* hybridization results. (Right) The DAPI banding of the same metaphase.

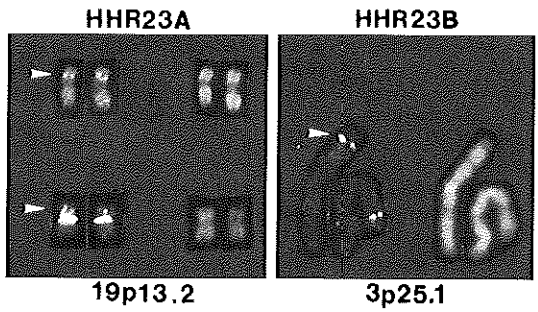


FIG.2 *In situ* hybridization of metaphase chromosomes with biotinylated genomic *HHR23* probes. (Left) *In situ* hybridization signal on chromosome 19 with the biotinylated genomic *HHR23A* probe and a cocktail of a centromeric probe specific for chromosome 19 in combination with *HHR23A*. (Right) The signal of the biotinylated *HHR23B* probe on chromosome 3.

XPC. For localization of the xeroderma pigmentosum group C correcting gene (*XPC*), the 3.6-kb cDNA was biotinylated and used for *in situ* hybridization. A representative *in situ* hybridization for *XPC*, of the more than 50 metaphases analyzed, is depicted in Fig.1C. Interestingly, like *HHR23B*, the *XPC* gene was also assigned to chromosome 3p25.1. Since both genes are located on 3p25.1, a double hybridization with both *XPC* and *HHR23B* was performed to see whether they hybridize to discernable locations. To this aim, the *XPC* cDNA probe was haptenized with biotin and visualized by FITC-labeled antibodies, whereas the *HHR23B* probe was provided with a digoxigenin label and visualized by TRITC/Texas red-labeled antibodies. The results (Fig.3) indicate a similar cytogenetic position on the chromosome. The cytogenetic colocalization of both probes is apparent from the fact that only one spot is present with a mixed colour composed of the green of *XPC* and red of the *HHR23B* probe. Detectable hybridization of both probes is also demonstrated by the finding of distinct spots in interphase nuclei (data not shown). Occasionally, in more elongated chromosomes distinct green (*XPC*) and orange/red (*HHR23B*) spots were observed, with the XPC hybridization in each case located more to the telomere. These data suggest the following order of these genes on the chromosome: centromere // *HHR23B* - *XPC* // telomere.

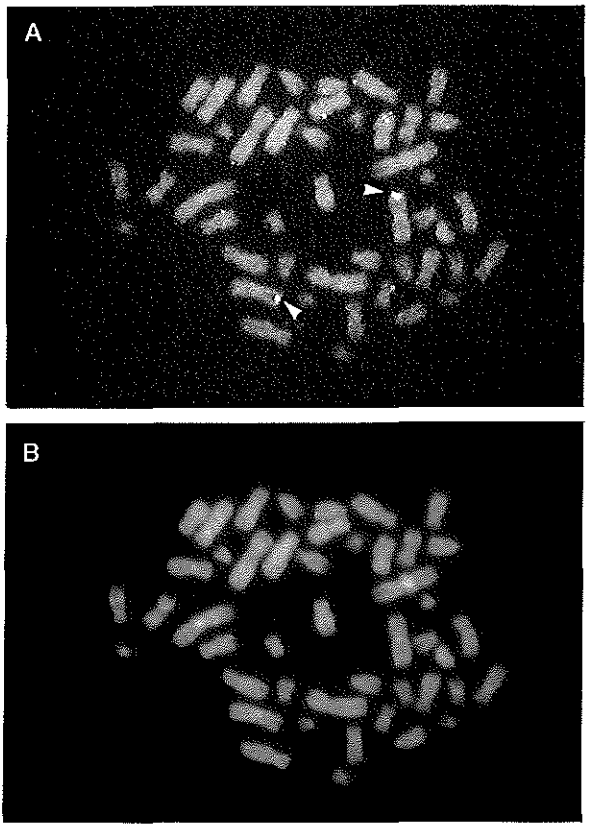


FIG.3. *In situ* hybridization of metaphase chromosomes with biotinylated *HHR23B* and *XPC* probes. The arrowheads indicate the hybridization signals on the short arm of chromosome 3. Digoxigenin-labeled *HHR23B* (visualized with TRITC/ Texas red) was hybridized in combination with biotin-labeled *XPC* (visualized with FITC). (A) *In situ* hybridization results. (B) The DAPI banding of the same metaphase spread.

Characterization of the Genomic Context of the XPC and HHR23 Genes by Pulsed-Field Gel Electrophoresis.

Since the cytogenetic data do not permit conclusions about the physical distance by which *XPC* and *HHR23B* are separated and in fact do not exclude the possibility that they are very close together, pulsed-field gel electrophoresis experiments were conducted. The following restriction enzymes were utilized: *Eag*I, *Mlu*I, *Sfi*I, *BssH*II, *Not*I, *Nru*I, and *Sal*I. The results summarized in Table 1a indicate that all but one enzyme generate different fragments for each gene. However, interestingly, *XPC* and part of the *HHR23B* gene hybridize to a *Mlu*I fragment of about 625 kb, raising the possibility that both genes reside on this fragment. The fact that the *HHR23B* cDNA probe visualizes two *Mlu*I bands implies that this site is situated within the *HHR23B* gene. In view of the large size of the *Mlu*I fragments ($\approx$625 and $>$1000 kb), it was not feasible to perform partial digestions to rule out the possibility that the hybridization of both genes to a similar-size

fragment is due to a coincidental correspondence in size. From the results obtained with *Eag*I and *Bss*HII, one can conclude that *XPC* and *HHR23B* must be at least 250 kb apart, assuming that both genes are not larger than 50 kb. The *Mlu*I digest sets an upper limit to the distance between both loci of 625 kb if they indeed are located on the same fragment. Finally, the findings of more than one *HHR23B* hybridizing fragment for four rare-cutting enzymes (*Eag*I, *Mlu*I, *Bss*HII, and *Sal*I) strongly suggests the presence of a CpG island within this gene. The *XPC* gene appears to contain sites for *Eag*I and *Sal*I.

We have also characterized the genomic context of the *HHR23A* gene by PFGE. The results are presented in Table 1b. From these data we can conlude that there are at least two *Nru*I sites and one *Cla*I site present in the *HHR23A* gene. Furthermore, the *Nru*I digests indicate that the gene has a minimal size of 16 kb.

(A)

Enzyme	*HHR23B* fragment (kb)	*XPC* fragment (kb)
*Eag*I	≈700	≈310
	≈290	<40
*Mlu*I	>1000?	≈625
	≈625	
*Bss*HII	≈725	<100
	≈310	
*Not*I	≈800	≈750
*Sfi*I	≈300	≈240
*Sal*I	>1000	≈450
	≈100	<50

(B)

Enzyme	*HHR23A* fragment (kb)
*Cla*I	≈300
	60-70
*Nru*I	≈240
	≈18
	≈16
*Not*I	≈190
*Mlu*I	<80
*Bss*HII	<80

Table 1 Pulsed-Field Gel electrophorese Analysis of XPC, HHR23A, and HHR23B.

RFLPs in the Areas of HHR23A and HHR23B.

To facilitate linkage analysis we have searched for RFLPs in the genomic areas of both *HHR23* genes. The *HHR23A* gene, on the short arm of chromosome 19, does not appear to be highly polymorphic for *Eco*RI, *Pst*I, *Rsa*I, *Msp*I and *Taq*I. Similarly, the *HHR23B* 3p25.1 locus does not appear to be highly polymorphic for *Pst*I, *Rsa*I, *Msp*I, *Pvu*II and *Taq*I. However, the *HHR23B* locus seems to contain a *Bgl*II-polymorphic site. Figure 4 shows part of the blot in which the RFLP is presented. The allelic frequency of the less common b allele is estimated from the small sample to be in the order of 25% in the Caucasian population.

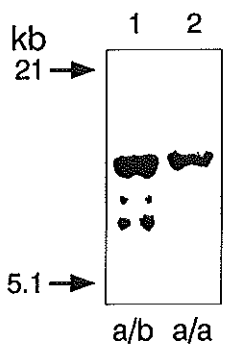


FIG.4. A *Bgl*II polymorphism in the *HHR23B* gene. Autoradiogram of *Bgl*II-digested DNA of two unrelated Caucasian individuals hybridized with the full-length *HHR23B* cDNA probe. The polymorphic band which was detected in DNA of four out of eight unrelated individuals. Lane 1, DNA with the a/b haplotype; lane 2, DNA with the a/a haplotype.

DISCUSSION

The consequences of inefficient or deficient repair are illustrated by genetic repair diseases that in general predispose individuals to cancer due to the fact that mutations arising from unrepaired lesions accumulate at a high rate. The genetics of NER in mammalian (including human) cells is very complex since many genes are involved in this pathway. Rodent mutant cell lines defective in NER fall into at least 11 distinct genetic complementation groups. The human NER-defective hereditary disease XP is characterized by at least 7 genetic complementation groups. CS and PIBIDS account for at least 3 other complementation groups. Furthermore, one of the PIBIDS complementation groups shows overlap with XP (XP group D). The parallelism between yeast repair genes and these NER syndromes indicates the importance of this evolutionarily conserved pathway (Hoeijmakers 1993a,b). Here we report the chromosomal localization of 3 recently isolated human repair genes. The *XPC* gene is localized on the short arm of chromosome 3, like the *HHR23B* gene. The *HHR23A* gene maps on the short arm of chromosome 19.

In the process of reviewing of our manuscript, we became aware of earlier work by Legerski *et al.* (1994) providing unequivocal evidence in favor of the localization of *XPC* on 3p25 based on southern blotting of hybrids and *in situ* hybridization studies and contrasting with preliminary results by Kauer and Atwahl (1993) suggesting chromosome 5, based on microcell-mediated correction studies. Our independent observation strongly corroborates Legerski's finding and finally settles the question of the *XPC* assignment in favor of chromosome 3p25.

As shown in Table 2, the human NER genes mapped to date are distributed over the genome. This resembles the situation in *S. cerevisiae*. Since yeast represents the other end of the eukaryotic spectrum, it is likely that random distribution of NER genes occurs in all eukaryotes.

Human gene	Yeast homolog	Corrected human NER syndrome	Chromosomal localization	References
XPA	*RAD14*	XPA	9q34	Kauer and Athwal, 1989 Ishizaki, 1990
XPB(ERCC3)	*RAD25*	XPB	2q21	Weeda *et al.*, 1989
XPC	*RAD4*	XPC	3p25.1	This report
XPD(ERCC2)	*RAD3*	XPD	19q13.2	Siciliano, 1986
XPG(ERCC5)	*RAD2*	XPG	13q32-33	Mudgett and MacInnes, 1990
ERCC1	*RAD10*	Unknown[a]	19q13.2	Mohrenweiser *et al.*, 1989
ERCC4	Unknown	Unknown	16p13.13-p13.2	Liu, 1993
HHR23A	*RAD23*	Unknown	19p13.2	This report
HHR23B	*RAD23*	Unknown	3p25.1	This report
CSB(ERCC6)	Unknown	CSB	10q11-21	Troelstra *et al.*, 1992
DNA Ligase I	Ligase	46BR	19q13.2	Barnes *et al.*, 1992

[a] Not any of the known NER-deficient complementation groups.

Table 2 Chromosomal Localization of Human NER Genes.

Many repair genes are found on chromosome 19: *ERCC1*, *2*, the gene for DNA ligase I, and the X-ray repair gene *XRCC1* all reside on 19q13.2. The *HHR23A* gene can be added to this list, but in contrast to the others this gene resides on the short arm. The location of a considerable fraction of repair genes on chromosome 19 is in line with the known high density of genes on this chromosome (Human Gene Mapping 11, 1991).

In humans, 2 pairs of NER genes are localized close together. Previously we reported that the *ERCC1* and *ERCC2* genes are situated 250-300 kb apart on 19q13.2 (Smeets *et al.*, 1990). Here we find close proximity of the *XPC* and *HHR23B* genes, on band 3p25.1 at a distance of 250-625 kb. This colocalization is a remarkable finding, since the gene products form a tight complex, which is not the case for *ERCC1* and *ERCC2*. The question remains whether this is pure coincidence or whether the colocalization has a functional significance, for instance, coregulation at the transcription level. At present it is not known whether other genes are located in between *XPC* and *HHR23B* as with *ERCC1* and *ERCC2*.

The Chromosomal Context of HHR23A.

Loss of heterozygosity studies indicate an underrepresentation of the involvement of chromosome 19 in human cancers, which is in contrast to chromosome 3p (Seizinger *et al.*, 1991). Many expressed genes have been mapped to the 19p13 region (McKusick, 1990), as have breakpoints for several translocations (Kamps *et al.*, 1990). The t(1;19)(q23;p13) chromosomal translocation is observed in 25% of children with pre-B-cell acute lymphoblastic leukemia (ALL) (Hunger *et al.*, 1991). However, the gene(s) disrupted by the translocation have not yet been cloned, and the possibility that the *HHR23A* gene is involved cannot be excluded at present. Furthermore, it should be noted that many other breakpoints involved in ALL have been described (Ahuja and Cline, 1988).

A rare heritable folate-sensitive fragile site at 19p13 was detected in four brothers (Chodirker *et al.*, 1987). The clinical significance of this fragile site and possible involvement of *HHR23A* are unknown. Other well-characterized loci on 19p are the low density lipoprotein receptor (Francke *et al.*, 1984), insulin receptor (Yang-Feng *et al.*, 1985), and the human Ro ribonucleoprotein (52kDa) autoantigen (McCauliffe *et al.*, 1990). The MHC class II regulatory factors RFX1 and RFX2, defective in hereditary HLA II deficiency (Bare lymphocyte syndrome), respectively mapped on 19p13.1 and 19p13.2-p13.3 (Pugliatti *et al.*, 1992). These data in combination with the *HHR23A* locus should help guide molecular studies to characterize further 19p13 breakpoints and mapping of genes in this chromosomal region.

The Chromosomal Context of HHR23B and XPC.

Like *HHR23B*, the *XPC* gene maps on 3p25.1. A number of studies suggested the presence of important genetic loci on the short arm of chromosome 3. Chromosomal abnormalities of 3p have been observed in breast cancer, lung cancer, renal cell carcinoma, ovarian carcinoma, various hematological malignancies and cervical cancer (Naylor and Carritt, 1991). Loss of 3p in some of these malignancies suggests the presence of one or multiple tumor suppressor genes on the short arm of chromosome 3. One of these is the gene responsible for the genetic disorder von Hippel-Lindau, which was recently cloned (Latif *et al.*, 1993). Furthermore, the developmental disorder Greig craniopolysyndactyly syndrome which has been associated with a t(3;17)(p21;p13) balanced translocation. The 3p25 region has been characterized quite well by several groups working on the von Hippel-Lindau disease gene. Chromosome 3p allele loss has been described in four tumor types: renal cell carcinoma, haemangioblastoma, phaeochromocytoma and pancreatic tumor, suggesting a common mechanism of tumorigenesis in all types of tumor in von Hippel-Lindau disease. Our data obtained by pulsed-field analysis were not easy to implement in the map generated for the von Hippel-Lindau region (Szymanski *et al.*, 1993).

The close vicinity of the *XPC* and *HHR23B* genes opens the possibility of a common deletion inactivating both genes. Since impairment of the two functions may yield a phenotype more severe than regular XP-C, we searched for XP-C patients with additional features. An XPC patient (XP1MI) was described having a unique combination of symptoms that correspond to two sun-sensitive conditions: xeroderma pigmentosum (XPC) conferring sensitivity to UV-B, and systemic lupus erythematosus (SLE), with an exaggerated response to UV-A (Hananian and Cleaver, 1980). This XPC patient shows no detectable mRNA on Northern blot analysis (Legerski and Peterson, 1992) and a point mutation in one of the alleles of XP1MI was reported using RT-PCR (Li, *et al.*, 1993). No functional studies demonstrating inactivation of the gene by this mutation have been performed, and it is not known whether this sequence alteration is present on one or both alleles, nor whether the patient has lost the second allele. However, *in situ* hybridization on lymphocytes and fibroblasts of patient XP1MI (our own unpublished data) show that both alleles of *XPC* and *HHR23B* are present. This was also found in another XPC patient (Halley, *et al.*, 1979) showing no detectable mRNA in Northern blot analysis is XP4PA (Legerski and Peterson, 1992). Recently, a mutation was reported in the DNA of this patient, involving a deletion of two nucleotides at the mRNA level and a frameshift in the central part of the protein (Li, *et al.*, 1993). No information on the other allele was provided.

Relationship of the HHR23A and B Genes with Human NER Syndromes.

Virtually all NER-deficient XP, CS, TTD, and rodent complementation groups for which no repair gene is isolated have been tested for the ability of the *HHR23A* gene to correct their defect. However, no correction was found (P.J. van der Spek and W. Vermeulen unpublished results). Similar studies using the *HHR23B* gene are in progress. Given the high amino acid sequence homology between both gene products (57% identity and 76% similarity) it is possible that HHR23A and HHR23B proteins have largely overlapping functions. When functional redundancy exists, it would require the unlikely event of simultaneous inactivation of both *HHR23* genes for clinical symptoms to become manifest, which may explain the absence of a known repair disorder for *HHR23A* and possibly also for *HHR23B*. Targeted gene replacement in mouse embryonal stem cells opens the possibility of generating HHR23-defective cell lines and mice in the laboratory. Via cloning of the mouse homologs, valuable insight can be gained in translation of a molecular defect in the *HHR23* function into clinical features, particularly the predisposition to cancer and other clinical hallmarks of human NER disorders.

ACKNOWLEDGMENTS

We are grateful to Arjenne Hesseling-Janssen, D.J.J. Halley, Peter Heutink, Guido Breedveld, and Cécile Visser for their help with the PFGE and RFLP experiments. Furthermore, we thank Christine Troelstra and Prof. D. Bootsma for the helpful suggestions and discussions. We also thank Dr. M. Siciliano for sharing his independent results on the XPC assignment. Mirko Kuit, Ruud Koppenol, and Tom de Vries Lentsch are acknowledged for photography. This work was financially supported by a subsidy of the Dutch government to the Research School Medical Genetics Centre (MGC).

REFERENCES

Ahuja, H.G., and Cline, M.J. (1988). Genetic and cytogenetic changes in acute lymphoblastic leukemia. *Med. Oncol. Tumor. Pharmacother.* **5:** 211-222.

Barnes, D.E., Tomkinson, A.E., Lehmann, A.R., Webster, A.D., and Lindahl, T. (1992). Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. *Cell* **69:** 495-503.

Bohr, V.A. (1991) Gene specific DNA repair. *Carcinogenesis* **12:** 1983-1992.

Bootsma, D., and Hoeijmakers, J.H.J. (1993). DNA repair: Engagement with transcription. *Nature* **363:** 114-115.

Busch, D., Greiner, C., Rosenfeld, K.L., Ford, R., de Wit, J., Hoeijmakers, J.H.J., and Thompson L.H. (1994). Complementation group assignments of moderately UV-sensitive CHO mutants isolated by large scale screening (FAECB). *Mutagenesis*, in press.

Chérif, D., Julier, C., Delattre, O., Derré, J., Lathrop, G.M., and Berger R. (1990). Simultaneous localization of cosmids and chromosome R-banding by fluorescence microscopy: application to regional mapping of human chromosome 11. *Proc. Natl. Acad. Sci. USA* **87**: 6639-6643.

Chodirker, B.N., Chudley, A.E., Ray, M., Wickstrom, D.E., and Riordan, D.L. (1987). Fragile 19p13 in a family with mental illness. *Clin. Genet.* **31**: 1-6.

Cleaver, J.E., and Kraemer, K.H. (1994). Xeroderma pigmentosum. In: "The Metabolic Basis of Inherited Disease," 7th ed. (C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle Eds.) Vol. 2, pp. 2949-2971, McGraw-Hill, New York.

Francke, U. Brown, M.S., and Goldstein, J.L. (1984). Assignment of the human gene for the low density lipoprotein receptor to chromosome 19: synteny of a receptor, a ligand, and a genetic disease. *Proc. Natl. Acad. Sci. USA* **81**: 2826-2830.

Halley, D.J.J., Keijzer, W., Jaspers, N.G.J., Niermeijer, M.F., Kleijer, W.J., Boué, J, Boué, A., and Bootsma, D. (1979). Prenatal diagnosis of xeroderma pigmentosum (group C) using assays of unscheduled DNA synthesis and postreplication repair. *Clinical Genetics* **16**: 137-146.

Hananian, J., and Cleaver, J.E. (1980). Xeroderma pigmentosum exhibiting neurological disorders and systemic lupus erythematosus. *Clinical Genetics* **17**: 39-45.

Hanawalt, P., and Mellon, I. (1993). Stranded in an active gene. *Current Biology* **3**: 67-69.

Hoeijmakers, J.H.J. (1993a). Nucleotide excision repair I: from E.coli to yeast. *Trends Genet.* **9**: 173-177.

Hoeijmakers, J.H.J. (1993b). Nucleotide excision repair II: from yeast to mammals. *Trends Genet.* **9**: 211-217.

Human Gene Mapping 11 (1991). *Cytogenet. Cell Genet.* **58**: 1-2200.

Hunger, S.P., Galili, N., Carroll, A.J., Crist, W.M., Link, M.P., and Cleary, M.L. (1991). The t(1;19)(q23;p13) results in consistent fusion of E2E and PBX1 coding sequences in acute lymphoblastic leukemias. *Blood* **77**: 687-693.

Ishizaki, K., Oshimura, M., Sasaki, M.S., Nakamura, Y., and Ikenaga, M. (1990). Human chromosome 9 complements UV sensitivity of xeroderma pigmentosum group A cells. *Mutat. Res.* **235**: 209-215.

Kamps, M.P., Murre, C., Sun, X.H., and Baltimore, D. (1990). A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. *Cell* **60**: 547-555.

Kauer, G.P., and Athwal, R.S. (1989). Complementation of a DNA repair defect in xeroderma pigmentosum cells by transfer of human chromosome 9. *Proc. Natl. Acad. Sci. USA* **86**: 8872-8876.

Kauer, G.P., and Athwal, R.S. (1993). Complementation of a DNA repair defect in xeroderma pigmentosum cells of group C by transfer of human chromosome 5.
Som. Cell. Mol. Genet. **19**: 83-93.

Koken, M.H.M., Smit, E.M.E., Jaspers-Dekker, I., Oostra, B.A., Hagemeijer, A., Bootsma, D., and Hoeijmakers, J.H.J. (1992). Localization of two human homologs, HHR6A and HHR6B, of the yeast DNA repair gene RAD6 to chromosomes Xq24-q25 and 5q23-q31. *Genomics* **12**, 447-453.

Latif, F., Tory, K., Gnarra J., Yao, M., Duh, F.M., Orcutt, M.L., Stackhouse, T., Kuzmin, I.,

Modi, W., Geil, L., Schmidt, L., Zhou, F. Li, H., Wei, M.H., Chen F., Glenn, G., Choyke, P., Walther M.M., Weng Y., Duan, D.S.R., Dean, M. Glavac, D., Richards, F.M., Crossey P.A., Ferguson-Smith, M.A., Le Paslier, D, Chumakov, I., Cen, D., Chinault, A.C., Maher, E.R., Linehan, W.M., Zbar B., and Lerman, M.I. (1993). Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**: 1317-1320.

Legerski, R.J., and Peterson, C.K. (1992). Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C. *Nature* **359**: 70-73.

Legerski, R.J., Liu, P., Li.,L, Peterson, C.A., Zhao, Y., Leach, R.J., Naylor, S.L., and Siciliano M.J. (1994). Assignment of xeroderma pigmentosum group C (*XPC*) gene to chromosome 3p25. *Genomics* **21**: 266-269.

Lehmann, A. R., Kirk-Bell, S., Arlett, C., Paterson, M.C., Lohman, P.H.M., de Weerd-Kastelein, E.A., and Bootsma D. (1975). Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation. *Proc. Natl. Acad. Sci.* **72**: 219-223.

Lehmann, A. R. (1982). Three complementation groups in Cockayne syndrome. *Mut. Res.* **106**: 347-356.

Lehmann, A. R. (1987). Cockayne's syndrome and trichothiodystrophy: Defective repair without cancer. *Cancer Rev.* **7**: 82-103.

Li, L., Bales, E.S., Peterson, C.A., and Legerski R.J. (1993). Characterization of molecular defects in xeroderma pigmentosum group C. *Nature genetics*. **5**: 413-417.

Liu, P., Siciliano, J., White, B., Legerski, R., Callen, D., Reeders, S., Siciliano, M.J., and Thompson, L.H. (1993). Regional mapping of human DNA excision repair gene *ERCC4* to chromosome 16p13.13-p13.2. *Mutagenesis* **8**: 199-205.

Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio K., Tanaka, K., van der Spek, P.J., Bootsma D., Hoeijmakers, J.H.J., and Hanaoka F. (1994). Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homolog of yeast RAD23. EMBO J. **8**: 1831-1843.

McCauliffe, D.P., Lux, F.A., Lieu, T.S., Sanz, I., Hanke, J., Newkirk, M.M., Bachinski, L.L., It, Y., Siciliano, M.J., and Reichlin, M. (1990). Molecular cloning, expression, and chromosome 19 localization of a human Ro/SS-A autoantigen. *J. Clin. Invest.* **85**: 1379-1391.

McKusick, V.A. (1990). "Mendelian inheritance in man," 9th ed., The Johns Hopkins Univ. Press, Baltimore, MD.

Mohrenweiser, H.W., Carrano, A. V., Feritta, A., Perry, B., Thompson, L.H., Tucker, J.D., and Weber, C.A. (1989). Refined mapping of the three DNA repair genes ERCC1, ERCC2, and XRCC1 on chromosome 19. *Cytogenet. Cell Genet.* **52**: 11-14.

Mudgett, J.S., and MacInnes, M.A. (1990). Isolation of the functional human excision repair gene *ERCC5* by intercosmid recombination. *Genomics* **8**: 623-633.

Naylor, S., and Carritt, B. (1991). Report of the committee on the genetic constitution of chromosome 3. *Cytogenet. Cell Genet.* **58**: 170-230.

Pinkel, D., Straume, T., and Gray, J. (1986). Cytogenetic analysis using quantitative highly sensitive, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **83**: 2934-2938.

Pugliatti, L., Derré, J., Berger R., Ucla, C., Reith, W., and Mach, B. (1992). The genes for MHC class II regulatory factors RFX1 and RFX2 are located on the short arm of chromosome 19. *Genomics* **13**:

1307-1310.

Riboni, R., Botta, E., Stefanini M., Numata, M., and Yasui, A. (1992). Identification of the eleventh complementation group of UV-sensitive excision repair-defective rodent mutants. *Cancer Res.* **52:** 6690-6691.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Seizinger, B.R., Klinger, H.P., Junien, C., Nakamura, Y., Le Beau, M., Cavenee, W., Emanuel, B., Ponder, B., Naylor, S., Mitelman, F., Louis, D., Menon, A., Newsham, I., Decker, J., Kaelbling, M., Henry, I., and v. Deimling, A. (1991). Report of the committee on chromosome and gene loss in human neoplasia. *Cytogenet. Cell. Genet.* **58:** 1080-1096.

Siciliano, M.J., Carrano, A.V., and Thompson, L.H. (1986). DNA repair gene on chromosome 16 and a third repair gene on chromosome 19. *Am. J. Hum. Genet.* **39:** A42.

Smeets, H., Bachinski, L., Coerwinkel, M., Schepers, J., Hoeijmakers, J.H.J., van Duin, M., Grzeschik, K.H., Weber, C.A., de Jong, P., Siciliano, M.J., and Wieringa, B. (1990). A long-range restriction map of the human chromosome 19q13 region: close physical linkage between CKMM and the *ERCC1* and *ERCC2* genes. *Am. J. Hum. Genet.* **46:** 492-501.

Stefanini, M., Vermeulen, W., Weeda, G, Giliani, S., Nardo, T., Mezzina, M., Sarasin, A., Harper, J.I., Arlett, C.F., Hoeijmakers, J.H.J., and Lehmann, A.R. (1993). A new nucleotide-excision-repair gene associated with the disorder trichothiodystrophy. *Am. J. Hum. Genet.* **53:** 001-005.

Szymanski, S.C., Hummerich, H., Latif, F., Lerman, M.I., Röhrborn, G., and Schröder, E. (1993). Long range restriction map of the von Hippel-Lindau gene region on human chromosome 3p. *Hum. Genet.* **92:** 282-288.

Tanaka, K., Kawai, K., Kumahara, Y., Ikenaga, M., and Okada, Y. (1981). Genetic complementation groups in Cockayne syndrome. *Somat. Cell. Genet.* **7:** 445-455.

Troelstra, C., Landsvater, R.M., Wiegant, J., van der Ploeg, M., Viel, G., Buys, C.H.C.M., and Hoeijmakers J.H.J. (1992). Localization of the Nucleotide Excision Repair Gene ERCC6 to human chromosome 10q11-q21. *Genomics* **12:** 745-749.

Van Ommen, G.J.B., and Verkerk, J.M.H. (1986). Restriction analysis of chromosomal DNA in a size range up to two million base pairs by pulsed field gradient electrophoresis.
In "Human genetic diseases, A Practical Approach" (K.E. Davies, Ed.), pp. 113-133, IRL Press, Oxford/Washington, DC.

Venema, J., Mullenders, L.H.F., Natarajan, A.T., van Zeeland, A.A., and Mayne, L. (1990). The genetic defect in Cockayne's syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc. Natl. Acad. Sci. USA* **87:** 4707-4711.

Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland, A.A., and Mullenders, L.H.F. (1991). Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. *Mol. Cell. Biol.* **11:** 4128-4134.

Vermeulen, W., Stefanini, M., Giliani, S., Hoeijmakers, J.H.J., and Bootsma, D. (1991). Xeroderma pigmentosum complementation group H falls into complementation group D. *Mutat. Res.* **255:** 201-208.

Watkins, J.F., Sung, P., Prakash, L., and Prakash, S. (1993). The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological

function. *Mol. and Cell. Biol.* **13:** 7757-7765.

Weeda, G., Wiegant, J., van der Ploeg, M., Geurts van Kessel, A. H. M., van der Eb, A.J., and Hoeijmakers J.H.J. (1991). Localization of the xeroderma pigmentosum group B-correcting gene ERCC3 to human chromosome 2q21. *Genomics* **10:** 1035-1040.

Yang-Feng, T.L., Ullrich, A., and Francke, U. (1985). Mapping of the human insulin receptor gene (INSR) to the distal short arm of chromosome 19 (19p13.2-p13.3). *Cytogenet. Cell Genet.* **40:** 785.

CHAPTER

VI

Cloning, Comparative Mapping and RNA Expression of the Mouse Homologues of the *S. cerevisiae* Nucleotide Excision Repair Gene *RAD23*

CLONING, COMPARATIVE MAPPING AND RNA EXPRESSION OF THE MOUSE HOMOLOGUES OF THE *S. CEREVISIAE* NUCLEOTIDE EXCISION REPAIR GENE *RAD23*.

Peter J. van der Spek[1], Cécile E. Visser[1], Fumio Hanaoka[2], Bep Smit[1], Anne Hagemeijer[1], Dirk Bootsma[1], and Jan H.J. Hoeijmakers[1].

[1] Department of Cell Biology and Genetics, Medical Genetic Centre, Erasmus University, Rotterdam, The Netherlands; and: [2] Cellular Physiology Laboratory, The Institute of Physical and Chemical Research (RIKEN), Wako-shi, Saitama 351-01, Japan.

GENOMICS 31: 20-27, 1996.

ABSTRACT

The *Saccharomyces cerevisiae RAD23* gene is involved in nucleotide excision repair (NER). Two human homologs of *RAD23*, *HHR23A* and *HHR23B* (HGMW-approved symbols RAD23A and RAD23B), were previously isolated. The HHR23B protein is complexed with the protein defective in the cancer-prone repair syndrome xeroderma pigmentosum, complementation group C, and is specifically involved in the global genome NER subpathway. The cloning of both mouse homologs (designated MHR23A and MHR23B) and detailed sequence comparison permitted the deduction of the following overall structure for all RAD23 homologs: an ubiquitin-like N-terminus followed by a strongly conserved 50-amino-acid domain that is repeated at the C-terminus. We also found this domain as a specific C-terminal extension of one of the ubiquitin-conjugating enzymes, providing a second link with the ubiquitin pathway. By means of *in situ* hybridization, *MHR23A* was assigned to mouse chromosome 8C3 and *MHR23B* gene to 4B3. Because of the close chromosomal proximity of human *XPC* and *HHR23B*, the mouse *XPC* chromosomal location was determined (6D). Physical disconnection of the genes in mouse argues against a functional significance of the colocalization of these genes in human. Northern blot analysis revealed constitutive expression of both *MHR23* genes in all tissues examined. Elevated RNA expression of both *MHR23* genes was observed in testis. Although the RAD23 equivalents are well conserved during evolution, the mammalian genes did not express the UV-inducible phenotype of their yeast counterpart. This may point to a fundamental difference between the UV response of yeast and human. No stage-specific mRNA expression during the cell cycle was observed for the mammalian *RAD23* homologs.

INTRODUCTION

The fundamental importance of DNA repair systems is illustrated by a number of cancer-prone human genetic disorders that are thought to be due to defects in DNA surveillance mechanisms like xeroderma pigmentosum (XP), Fanconi's anemia and Bloom syndrome (for a comprehensive review, see Friedberg *et al.*, 1995). XP and two other conditions, Cockayne's syndrome (CS) and trichothiodystrophy (TTD), are associated with defects in the nucleotide excision repair (NER) pathway. This system deals with the elimination of a diverse array of structurally unrelated lesions, including various UV-

induced photoproducts (cyclobutane pyrimidine dimers and 6-4 photoproducts), chemical adducts and crosslinks (Van Houten, 1990). A defect in one of at least seven genes XP-A to XP-G can cause the genetic recessive disease XP. XP patients show extreme sun sensitivity, pronounced pigmentation abnormalities in UV-exposed areas of the skin, and frequently, accelerated neurodegeneration. Importantly, the disease is associated with a > 1000-fold increased risk of skin cancer (see Cleaver and Kraemer, 1995, for a review). In contrast, CS patients are not cancer-prone and display a less severe sun sensitivity. Instead, this disorder is characterized by developmental impairment, including severe neurological abnormalities due to dysmyelination of the nervous system (Nance and Berry, 1992). Two CS complementation groups have been identified: CS-A and CS-B (Lehmann, 1987). TTD patients share many clinical features with CS, but have, in addition, brittle hair and nails (due to a reduced content of a class of ultra-high cysteine matrix proteins) and ichthyosis (Itin and Pittelkow, 1990). Three TTD genes are known, two of which are identical to XPB and XPD (Hoeijmakers, 1994). A striking selective correlation is found among all three TTD genes and mutations in subunits of basal transcription factor TFIIH, which has a dual involvement in transcription initiation and NER. Therefore, we have proposed that the 'non-XP' features of TTD are due to a partial impairment of the transcription function of the complex, in addition to a defect in the NER function. Thus, TTD can be regarded as a combined repair-transcription syndrome (Vermeulen *et al.*,1994; Bootsma and Hoeijmakers, 1993).

Recently, we reported the identification and cloning of two human homologs of the yeast NER gene *RAD23*: *HHR23A* and *HHR23B* (HGMW-approved symbols RAD23A and RAD23B) (Masutani *et al.*, 1994). Yeast *rad23* null mutants display an intermediate UV-sensitive phenotype, suggesting that the affected product is not indispensable for NER. The RAD23 protein contains an N-terminal ubiquitin-like domain (Watkins *et al.*, 1993). The HHR23B gene product was found in a tight complex with the XPC protein. This complex has a very high ssDNA binding activity. XPC cells harbour a specific defect in the repair of the nontranscribed sequences of the genome, including the nontranscribed strand of active genes, whereas the NER subpathway that accomplishes the preferential repair of the transcribed strand of active genes (transcription-coupled repair) is still operational (Venema *et al.*, 1989). This implies a selective role of the XPC complex in the global genome NER system (Van Hoffen *et al.*, 1995). Here we report the cloning and characterization of the mouse homologs of RAD23, the chromosomal localization, the expression profile, and the identification of novel domains in the primary amino acid sequence.

MATERIALS AND METHODS

Cloning and nucleotide sequence analysis

General molecular biological procedures for nucleic acid isolation, (sub)cloning, DNA sequencing, PCR, and hybridization were as detailed in Sambrook *et al.* (1989). PCR-derived probes of both *HHR23* cDNAs were used to screen a mouse testis cDNA library (Stratagene). Clones hybridizing with the human probes were picked, rescreened for purification, and examined by restriction enzyme analysis. Hybridization of human probes to mouse DNA was at 62°C in a hybridization mixture containing 10x Denhardt's solution, 10% dextransulfate, 0.1% SDS, 3xSSC, 50 mg of sonicated salmon sperm DNA per litre. Washing was performed twice for 20 min each in 3x SSC, twice for 20 min each in 1x SSC, and twice for 20 min each in 0.3x SSC at 62°C. Hybridization was detected by autoradiography on Fuji RX medical X-ray film with intensifying screens at -80°C. After two rounds of rescreens, EMBL-3 phages were converted into pBluescript II KS (Stratagene) or pTZ19R (Pharmacia) vectors and transformed to competent DH5αF' cells. Sequence analysis on double-stranded DNA was performed with the T7 DNA polymerase modification (Pharmacia) of the dideoxynucleotide chain termination method (Sanger *et al.*, 1977), using sequence-derived oligonucleotides, and exonuclease clones for sequencing both strands. For separation of the fragments, 6% acrylamide sequencing gels were used. The 5' end of *MHR23B* was derived from a mouse 17.5-day embryo 5' stretch cDNA library (ML1029a, Clontech). Genomic clones were derived by screening a mouse CCE library with PCR fragments covering the mouse cDNAs of both *MHR23* genes.

Chromosomal localization

In situ hybridization experiments using biotin-labeled or digoxigenin-labeled genomic fragments of *MHR23A* and *MHR23B* were performed as described elsewhere (Pinkel *et al.*, 1986). After incubation with avidin D-FITC (Vector, U.S.A), the biotin-labeled probes were visualized by FISH. The fluorescent signal was amplified with biotinylated goat anti-avidin D. Probes labeled with digoxigenin were visualized using Texas red-conjugated antidigoxigenin antibody.

Northern blot analysis

RNA samples were separated on an 1% agarose gel and transferred to Zeta probe membrane (Bio-Rad) as described by the manufacturer. Total RNA was isolated from adult (Balb/c) mice using the LiCl/urea method (Auffray and Rougeon, 1980). Filters were hybridized using mouse *MHR23A* and *MHR23B* ^{32}P-labeled cDNA probes. Labeling of DNA probes was carried out using the random priming protocol (Sambrook *et al.*, 1989).

Cell culture

HeLa cells were grown on F10/DMEM medium supplemented with 10% fetal calf serum, penicillin, and streptomycin. HeLa cells were synchronized using double thymidine block (Bootsma *et al.*, 1964, Galavazi *et al.*, 1966), to study stage-specific expression during the cell cycle. A primary culture of epidermal keratinocytes derived from human foreskin was established as described elsewere (Rheinwald and Green, 1975). Prior to RNA isolation, keratinocytes were UV-irradiated with 30 J/m^2, and HeLa cells were irradiated with 1 J/m^2 (UV-C).

Computer analysis

Sequences were analysed and compared using the BLAST algorithm (Altschul *et al.*, 1990) and the Genbank and EMBL databases.

RESULTS

Cloning and Sequence Analysis of the cDNAs Encoding the Mouse RAD23 equivalents.

To obtain cDNA's encoding the murine homologs of *HHR23A* and *HHR23B*, a mouse testis library was screened with both human cDNA probes. Analysis of several cDNA clones indicated the presence of two distinct *RAD23* genes, tentatively designated *MHR23A* and *MHR23B* (*M*ouse *H*omologs of *R*AD23). Additionally, for a *MHR23B*, a 17.5-day mouse embryo library was screened to obtain the missing 5' end of the cDNA. The *MHR23A* ORF encodes an acidic protein (pI 4.4) of 363 amino acids, with a calculated molecular mass of 39,772 Da. The *MHR23B* ORF encodes an acidic protein (pI 4.6) of 416 amino acids with a calculated molecular mass of 43,520 Da. Both the mouse A and the B polypeptides share, respectively, 96 and 88% sequence identity and 99 and 95% sequence similarity to their human counterparts. The human and mouse 23A proteins are equal in size. A notable difference between the human and mouse 23B proteins is the insertion of a stretch of seven glycine residues in the latter at position 339. The MHR23A and MHR23B gene products exhibit 58% amino acid sequence identity and 77% homology to each other. Both proteins share about 62% homology to their yeast counterpart (Fig. 1A).

Detailed comparison of the mouse, human and yeast RAD23 homologs permitted the identification of strongly conserved and more variable parts of the proteins that were not apparent from the human-yeast comparison and the identification of several domains in the RAD23 amino acid sequence that were not previously noted (Fig. 1). Together, these sequences account for the major part of the primary protein sequence. Figure 1B shows schematically the deduced overall structure of the RAD23 A and B proteins. At the N-terminus, a previously identified 80-amino-acid ubiquitin-like domain is present. It is followed by a region of approximately 65 amino acids predominantly (86%) composed of the physicochemically related amino acids proline, alanine, serine and threonine (termed here 'PAST' domain). This segment shows many sequence changes between human and mouse, suggesting that the precise primary sequence is probably not critical. Computer analysis revealed within the remainder of the protein two internal repetitive 50-amino-acid elements with significant homology to each other. The first domain is followed by a variable PAST-rich region, and the second is preceded by such a segment.

```
ScRAD23  M.VSHTFKNFKKEKVPFDVEPSNTLETKTKIAQSISCEESQI...KLIYSGKVLQDSKT
MHR23A   MAVTITLKTLQQQTFKIRMEPDETVKVLKEKIEAEKGRDAFPVAGQKLIYAGKILSDDVP
HHR23A   MAVTITLKTLQQQTFKIRMEPDETVKVLKEKIEAEKGRDAFPVAGQKLIYAGKILSDDVP
MHR23B   M..QVTLKTLQQQTFKIDIDPEETVKALKEKIESEKGKDAFPVAGQKLIYAGKILSDDTA
HHR23B   M..QVTLKTLQQQTFKIDIDPEETVKALKEKIESEKGKDAFPVAGQKLIYAGKILSDDTA

ScRAD23  VSECGLKDGDQVVFMVSQKKSTKTKVTEPPIAPESAATPGRENSTEASPSTDASAAPAAT
MHR23A   IRDYHIDEKNFVVVMVTKAKAGQGISAPPEASPTAV..PEPSTPFPPVLASGMSHPPPTS
HHR23A   IRDYKIDEKNFVVVMVTKAKAGQGTSAPPEASPTAA..PESSTSFPPAPTSGMSHPPPAA
MHR23B   LKEYKIDEKNFVVVMVTKPKAVTTAVPATTQPSSTPSPTTVSESPAVAAAQAPAPTPALA
HHR23B   LKEYKIDEKNFVVVMVTKPKAVSTPAPATTQQSAPASTTAVESSTTTTVAQAPTPVPALA

ScRAD23  APEGSQPQEEQTA.....................................TTERT
MHR23A   REDKS.PSEESTTTTSPESISGSVP.........................SSGSSGREE
HHR23A   REDKS.PSEESAPTTSPESVSGSVP.........................SSGSSGREE
MHR23B   PTSTPASTTPASTTASSEEAPAGATQPEKPAEKPAQTPVLTSPAPADSTPGDSSRSNLFE
HHR23B   PTSTPASITPASATASSEEAPASAAKQEKPAEKPAETPVATSPVTTDSTSGDSSRSNLFE

ScRAD23  ESASTPGFVVGTERNETIERIMEMGYQREEVERALRAAFNNPDRAVEYLLMGIPENLRQP
MHR23A   DAAST..LVTGSEYETMLTEIMSMGYERERVVAALRASYNNPHRAVEYLLTGIPGS...P
HHR23A   DAAST..LVTGSEYETMLTRIMSMGYERERVVAALRASYNNPHRAVEYLLTGIPGS...P
MHR23B   DATSA..LVTGQSYENMVTEIMSMGYEREQVIAALRASFNNPDRAVEYLLMGIPGD...R
HHR23B   DATSA..LVTGQSYENMVTEIMSMGYEREQVIAALRASFNNPDRAVEYLLMGIPGD...R

ScRAD23  EPQQQTALVAAEQPSTAATTAEQPAEDDLFAQAAQGGNASSQALGTTGQATDAAQGGPPGS
MHR23A   EPEH....GSVQESQR.AEQPA.....................TEAA.GENPLE
HHR23A   EPEH....GSVQESQV.SEQPA.....................TEAA.GENPLE
MHR23B   ESQAVVDPPEQAVSTGTPQSPAV.....................AAAAATTTATTTTTSGGHPLE
HHR23B   ESQAVVDPP.QAASTGAPQSSAVA....................AAAATTTATTTTTSSGGHPLE

ScRAD23  IGLTVEDLLSLRQVVSGNEEALRPLLENISARYPQLREHLMANPEVFVSMLLEAVGDNMQ
MHR23A   FLRDQPQFQNMRQVIQQNPALLPALLQQLGQENPQLLQQISRHQEQFIQMLNEPPGEL..
HHR23A   FLRDQPQFQNMRQVIQQNPALLPALLQQLGQENPQLLQQISRHQEQFIQMLNEPPGEL..
MHR23B   FLRNQPQFQQMRQIIQQNPSLLPALLQQIGRENPQLLQQISQHQEHFIQMLNEPVQEA..
HHR23B   FLRNQPQFQQMRQIIQQNPSLLPALLQQIGRENPQLLQQISQHQEHFIQMLNEPVQEA..

ScRAD23  DVMEGADDMVEGEDI.EVTGEAAAAGLGQGEGKGSFQVDYTPEDQAISRLCELGFERDL
MHR23A   ...........ADISDVEGEVGAIGEEAP...QMNYIQVTPQEKEAIERLKALGFPESL
HHR23A   ...........ADISDVEGEVGAIGEEAP...QMNYIQVTPQEKEAIERLKALGFPESL
MHR23B   ....GGQGGGGGGGGGGGGGGGGGIAEAGSG..HMNYIQVTPQEKEAIERLKALGFPEGL
HHR23B   ....GGQ.......GGGGGGGSGGIAEAQSG..HMNYIQVTPQEKEAIERLKALGFPEGL

ScRAD23  VIQVYFACDKNEEAAANILFSDHAD
MHR23A   VIQAYFACEKNENLAANFLLSQNFDDE
HHR23A   VIQAYFACEKNENLAANFLLSQNFDDE
MHR23B   VIQAYFACEKNENLAANFLLQQNFDED
HHR23B   VIQAYFACEKNENLAANFLLQQNFDED
```

Figure 1A

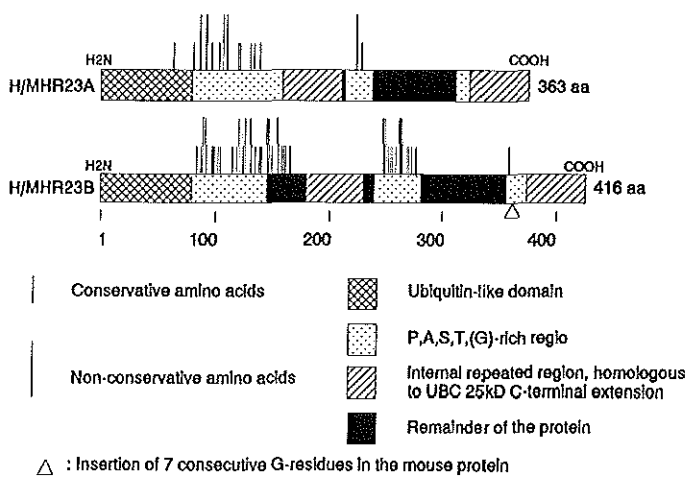


Figure 1B

FIG.1. (A) Sequence alignment of the yeast, mouse and human homologues of RAD23 with each other. Conserved sequences between yeast RAD23, MHR23A, HHR23A, MHR23B and HHR23B. The amino acid sequence of the mouse and human proteins are compared with that of yeast RAD23. Identical amino acids are presented in black boxes, whereas similar residues (A, S, T, P, and G; D, E, N, and Q; R, K, and H; I, L, V, and M; F, Y, and W) are given in gray boxes. (B) Schematic representation of the RAD23 protein homologs. Schematical presentation of conserved (short bar) and nonconserved (long bar) amino acid changes in mouse and human RAD23 proteins. The different domains discussed in the text are indicated.

The internal repeated sequence is fully conserved between mouse and human, presumably a reflection of its functional importance. The significance of this domain is further underlined by the finding of clear sequence similarity with a C-terminal extension present in a bovine ubiquitin-conjugating enzyme E2(25K) (Chen *et al.*, 1991), shown in Fig.2. Comparison with various databases identified in addition homology to two genome project clones: one of *Arabidopsis thalia* (accession number Z26691) and one of *Plasmodium falciparum* (accession number T09564). These partial clones represent the RAD23 homologs in the corresponding species. Based on the presence of multiple conserved residues including a C-terminal cysteine residue these short EST sequences are very likely the RAD23 homologs of these species.

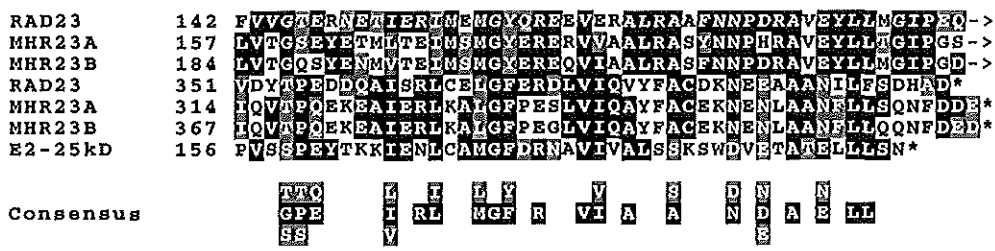


FIG. 2. Sequence alignment of the internally duplicated regions of RAD23 and homologues. Alignment of the internally duplicated regions of yeast, mouse RAD23, and the C-terminus of a bovine ubiquitin-conjugating enzyme E2(25K). The human sequences are not included in this figure since these are exactly identical to the mouse sequences. Identical amino acids are presented in black boxes, whereas similar residues (A, S, T, P, and G; D, E, N, and Q; R, K, and H; I, L, V, and M; F, Y, and W) are given in gray boxes.

Chromosomal Localization, Comparative Mapping

Both human *RAD23* homologs are located on different autosomal chromosomes. Interestingly, the *HHR23B* and *XPC* genes, the products of which form a tight complex, were found to colocalize to human chromosome 3p25.1 (Legerski *et al.*, 1994; van der Spek *et al.*, 1994). We have mapped the mouse equivalents to determine whether this colocalization is preserved during evolution. Genomic clones isolated from a mouse CCE genomic λ library were utilized for mapping purposes. Biotinylated probes hybridized with mouse metaphase spreads assigned the *MHR23A* gene to chromosome 8C3 (Fig. 3A) and the *MHR23B* to chromosome 4B3 (Fig. 3B). To confirm the *MHR23B* localization, a mouse leukemia cell line (Red8) that contains three copies of chromosome 4 was used. Furthermore, we determined that the mouse *XPC* gene maps to the 6D locus (shown in Fig. 3C), being in complete agreement with the localization of the human equivalent to the syntenic locus 3p25 (Legerski *et al.*, 1994; van der Spek *et al.*, 1994). Therefore, in contrast to the human situation, *XPC* and *MHR23B* do not colocalize in mouse.

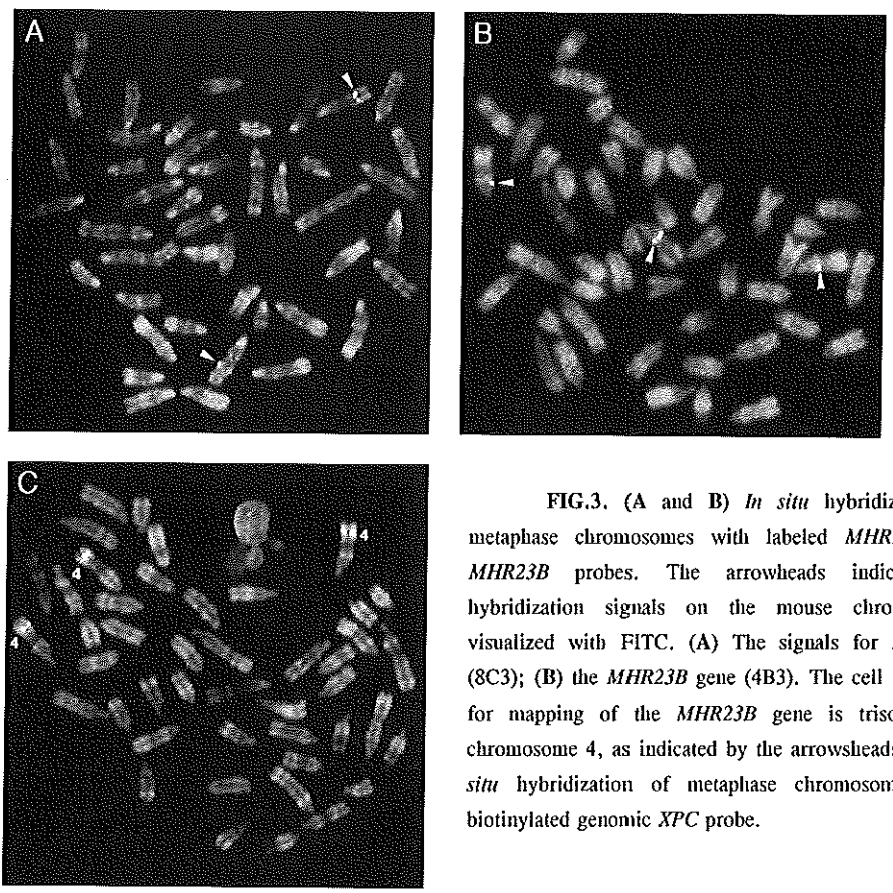


FIG.3. (A and B) *In situ* hybridization of metaphase chromosomes with labeled *MHR23A* and *MHR23B* probes. The arrowheads indicate the hybridization signals on the mouse chromosomes visualized with FITC. (A) The signals for *MHR23A* (8C3); (B) the *MHR23B* gene (4B3). The cell line used for mapping of the *MHR23B* gene is trisomic for chromosome 4, as indicated by the arrowsheads. (C) *In situ* hybridization of metaphase chromosomes with biotinylated genomic *XPC* probe.

Northern Blot Analysis of Mouse Tissues

To obtain clues about the potential function of the mammalian homologs of RAD23, we studied the expression properties of the genes. Northern blot analysis was performed using RNA from different mouse tissues. As shown in Fig. 4, both genes are expressed at the RNA level in all tissues and organs examined. When corrected for slight differences in the amount of RNA loaded in each lane (see 18S rRNA intensity from the ethidium bromide-stained gel), it is apparent that most tissues have roughly similar levels of *MHR23* transcripts. Clearly elevated expression was repeatedly observed for both *MHR23A* and *MHR23B* mRNAs in testis tissue compared to that in other tissues examined. The *MHR23A* probe visualized a transcript of 1.7 kb. The *MHR23B* gene was found to specify transcripts of 3.2 kb and 2.7 kb, both migrating just below the 28S ribosomal band. This size difference is most likely due to alternative polyadenylation of this gene, which was also observed in certain human tissues (data not shown). Additionally, the *MHR23A* probe detected enhanced expression in muscle tissue, whereas *MHR23B* shows enhanced RNA levels in brain, although in both cases, expression was lower than that in testis (Fig. 4).

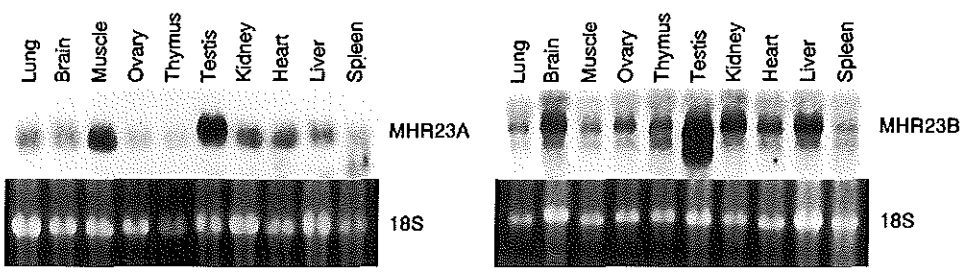


FIG.4. Northern blot analysis of different mouse tissues for mRNA expression of MHR23A and MHR23B. RNA (15 μg total cell RNA) was size-fractionated on a 0.8% agarose gel and, after blotting to Bio-Rad zetaprobe-GT filters, was hybridized with *MHR23A* and *MHR23B* probes. The intensity of the ethidium bromide-stained 18S rRNA band reflects the amount of RNA loaded in each lane for all the tissues examined.

We also examined whether expression of any of the *HHR23* transcripts changed during the mitotic cell cycle. RNA isolated from different stages (G1, S, G2 and mitosis). HeLa cells were synchronized by double thymidine block as described by Bootsma *et al.* (1964), and did not display significant differences in RNA levels of either *HHR23* gene (data not shown). In this regard, it is worth noting that no periodic fluctuations in *RAD23* mRNA levels were observed during the cell cycle in yeast (Madura and Prakash, 1990).

UV-Inducibility of *HHR23A* and *HHR23B* mRNA Expression

RAD23 belongs to a small subset of yeast repair genes, the expression of which is induced at the mRNA level upon UV exposure and during meiotic prophase (Madura and Prakash, 1990). Although the significance of this phenomenon is unknown, it is thought that it constitutes part of a mechanism similar to the well-studied SOS response in *Escherichia coli* (Walker, 1985). Whether such a system also exists in higher organisms is still an open question. To see whether the feature of UV inducibility is conserved, we analyzed the levels of *HHR23* mRNA in response to UV irradiation in a cell type that is most relevant in this respect: human cultured keratinocytes. RNA isolated at various timepoints after UV exposure (30 J/m^2) was analyzed by Northern blot hybridization. The *HHR23A* probe visualizes a mRNA of 1.7 kb, whereas the *HHR23B* probe detects a transcript of 2.8 kb. Figure 5 shows that the levels of *HHR23A* and *HHR23B* transcripts actually decline upon UV irradiation in contrast to expression of the *SPRR2* gene, a known UV-inducible gene included as a positive control (Gibbs *et al.*, 1993). Similarly, we did not find any evidence for UV induction in HeLa cells exposed to a lower UV dose (1 J/m^2) that were analyzed at time points much shorter after UV challenge (data not shown). We conclude that *HHR23A* and *HHR23B* do not express the UV-inducible phenotype of their yeast counterpart.

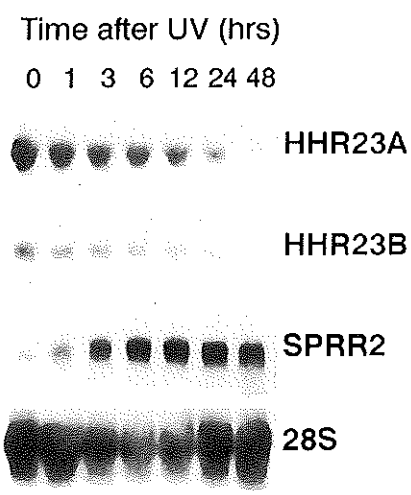


FIG.5. Effect of UV-irradiation on HHR23 transcription. RNA (15 μg total cell RNA) derived from UV-irradiated keratinocytes was size- fractionated on a 0.8% agarose gel. RNAs were transferred to Bio-Rad zetaprobe-GT filter and hybridized with *MHR23A* and *MHR23B* probes. Various timepoints after UV-irradiation (30J/m^2) were analysed. The 18S ribosomal band is visualized by autoradiography to indicate the amounts of RNA loaded.

DISCUSSION

Relationship Between RAD23 and Nucleotide Excision Repair.

The phenotype of the yeast *rad23* mutant suggests that the protein is not indispensable for yeast NER, as gene disruption only induces a partial UV sensitivity (Miller *et al.*, 1982). However, the *rad23* mutants were found to be almost totally defective in repair of 6-4 photoproducts and cyclobutane pyrimidine dimers after a UV dose of 50 J/m^2 (McCready, 1994; J. Brouwer, pers. communication, Leiden). One of the human homologs of RAD23, HHR23B, is complexed with the XPC protein. Cells from XP-C individuals harbor a specific defect in the global genome repair subpathway of NER, with transcription-coupled repair being unaffected (Venema *et al.*, 1991). Assuming that this property of RAD23 is conserved, one predicts that the protein in yeast is also selectively implicated in the global genome repair process and that it is complexed with RAD4, the presumed yeast XPC homolog. However, curiously, *rad4* mutants display a total NER deficiency (Verhage *et al.*, 1994).

Unfortunately, no mammalian mutants of any of the RAD23 homologs have been identified (Masutani *et al.*, 1994; van der Spek, unpublished observation), possibly due to functional redundancy of the two homologous gene products. The only biochemical activity assigned to the HHR23B-XPC complex is a very high affinity for ssDNA. The latter is likely to be due to the XPC component, as the purified HHR23B polypeptide does not exhibit strong DNA binding activity (van der Spek, manuscript in preparation).

Links with the Ubiquitin Pathway.

Concerning the function of RAD23 and its mammalian homologs, the findings reported here point to a specific connection with the ubiquitin pathway. The protein carries a strongly conserved ubiquitin-like N-terminus. First, the N-terminal 80 amino acids of RAD23 and its mammalian homologs contain a ubiquitin-like moiety, and deletion analysis has provided evidence that it is indispensable for the NER function of RAD23 in yeast (Watkins *et al.*, 1993). Consistent with this finding, the entire amino acid sequence of this region is strictly conserved between mouse and human, including lysine 48, which in ubiquitin is required for polyubiquitination (Chau *et al.*, 1989). Conservation of this residue suggests the possibility of covalent attachment of ubiquitin to RAD23.

A second independent link with the ubiquitin pathway was disclosed by the finding that the internally duplicated region (Fig. 2) displays sequence homology with the C-terminus of the bovine E2-(25K) member of the family of ubiquitin-conjugating (UBC)

enzymes. The significance of this homology is potentially multiple. First, both 50-amino-acid regions are strictly conserved between human and mouse and are the most strongly conserved segments of the protein from yeast to human. This strongly suggests that they have an important function. Second, the idea that they represent distinct domains is supported by the observation that both domains are preceded by regions with the same general characteristics: a high content of PAST(G) residues and a rapid evolution. Furthermore, the homology to the C-terminus of the E2(25K) ubiquitin-conjugating enzyme starts exactly behind the core domain shared with all UBC enzymes and constitutes the entire part of the protein that remains. Notably, in RAD23, the second element also contains the last 50 amino acids of the polypeptide. Third, the level of homology is highly significant. Using the sequence of one of the RAD23 elements in a computer search, the E2(25K) C-terminal extension stands out from all sequences in the database. The above considerations strongly support the idea that the RAD23 internal repeat constitutes a highly conserved, distinct domain with a specific function similar to that carried out by one of the ubiquitin-conjugating enzymes.

Ubiquitin, one of the most highly conserved and ubiquitous polypeptides known, marks proteins for non-lysosomal proteolytic degradation or for proper folding, and has a regulatory role in cellular homeostasis, the stress response, organelle biosynthesis, protein translocation across membranes and in DNA repair (see Ciechanover, 1994 for a recent review). Covalent attachment of one or multiple ubiquitin moieties to a target protein is the result of a number of activation and conjugation steps. Recently, Spence *et al.* (1995) described a ubiquitin mutant with specific defects in DNA repair and multiubiquitination.

The members of the UBC family (in yeast, at least 12 members) share a highly homologous 150- amino-acid domain containing the Cys residue (C^{88}) required for the ubiquitin-E2 thiol ester intermediate (Jentsch, 1992). A number of UBC members have in addition a unique C-terminal extension, presumably related to their specific function. Since each of the UBC enzymes is believed to target a selective set of proteins, it is possible that at least one of the functions of the unique C-terminal domain is to provide the specificity of interaction with target proteins or ubiquitin-protein ligase. For instance, the very acidic C-terminal domain of the yeast UBC repair enzyme RAD6 is thought to mediate interaction with the basic histones. Ubiquitination of histones may be part of the chromatin transactions required for the post replication repair, mutagenesis, and sporulation events in which this enzyme is implicated (Jentsch *et al.*, 1987).

Unfortunately, the specific function of the bovine E2(25K) UBC protein that could provide clues to the role of the equivalent duplicated domain in RAD23 is unknown, and

any yeast homologs have not yet been identified. Consequently, no mutants are available to assess the biological function of the protein. However, in light of the affiliation of the ubiquitin system with the postreplication repair pathway, one might speculate that RAD23 is implicated in modulation of chromatin structure in the context of the global genome repair subpathway. RAD23 thereby provides the first indication for a possible link between NER and the ubiquitin system.

Gene Duplication

The *in situ* hybridization results obtained for mouse *XPC* and *MHR23A* are in complete agreement with the localization of their human counterparts on chromosome 3p25 and 19p13 (Legerski *et al.*, 1994; van der Spek *et al.*, 1994). The *MHR23B* gene provides a new reference anchor locus on mouse chromosome 4B3, which could be useful for comparative gene mapping and linkage analysis in mammals (Copeland et al., 1993). The colocalization of *XPC* and *HHR23B* on human chromosome 3p25.1 is not observed in mouse. This argues against the possibility that the colocalization in human has an important function, e.g., in gene regulation or in complex formation.

Since the evolutionary distance of budding yeast to human is about 1200 x 10^6 years, one can roughly estimate from the degree of divergence of the human, mouse, and yeast genes that the *RAD23* gene duplicated approximately 700-800 x 10^6 years ago. This calculation assumes that the rate of evolution for the different homologs has remained constant. The reason for this duplication of RAD23 is not known. Because of the high level of sequence homology, both gene products might possess a redundant function. The finding that they are both expressed simultaneously in all tissues examined supports this presumption.

Expression Properties

Yeast *RAD23* mRNA is induced fivefold in meiotic prophase, coinciding with recombination (Madura and Prakash, 1990). Although *rad23* mutants undergo sporulation and produce viable spores, it is not known whether meiotic recombination is affected. Enhanced expression of both *MHR23* equivalents was found in testis tissue. Thus, the feature of meiosis-specific induction of RAD23 expression may be a preserved property of the gene, perhaps pointing to its involvement in meiotic recombination.

RAD23 represents the second example of evolutionary duplication of a repair gene, with the first being RAD6 (Koken *et al.*, 1991). A number of additional parallels between these genes can be noted. In addition to the links with the ubiquitin pathway

mentioned above, both genes display elevated levels of expression during mammalian spermatogenesis and yeast meiosis and sporulation. Both are -for repair genes- abundantly expressed at the protein level based on calculation of molecules per cell (van der Spek, manuscript in preparation). Finally, in yeast, both belong to the small subset of repair genes of which the expression is UV-inducible (Koken *et al.* (submitted for publication; this report). Notably, in both cases this feature does not appear to be conserved in mammals. This may point to a principle difference in the UV response between lower organisms that are fully subject to sudden changes in their environment and higher species in which the cellular milieu is kept relatively constant.

Cloning of the mouse homologs of repair genes can give valuable insight into the clinical consequences of molecular defects in the relevant gene. In particular, the predisposition to cancer and other clinical hallmarks of human NER disorders can be investigated by means of targeted gene replacement in MHR23-defective cell lines and mice. These experiments are in progress.

ACKNOWLEDGEMENTS

We are grateful to Dr. C. Troelstra and Dr. M. McKay for helpful suggestions and comments on the manuscript. We also thank Frans Lohman for providing us with RNA samples from UV-irradiated keratinocytes. Mirko Kuit is acknowledged for photography. The work at the Department of Cell Biology and Genetics of the Erasmus University was financially supported by the Medical Genetics Centre South-West Netherlands and in part by a Human Frontier Science Program Research grant.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215, 403-410.

Auffray, C., and Rougeon, F. (1980). Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* 107, 303-314.

Bootsma, D., and Hoeijmakers J.H.J. (1993). DNA repair: Engagement with transcription. *Nature* 363, 114-115.

Bootsma, D., Budke, L., and Vos, O. (1964). Studies on synchronous division of tissue culture cells initiated by excess thymidine. *Exp. Cell. Res.* 33, 301-309.

Chau, V., Tobias, J.W., Bachmair, A., Mariott, D., Ecker, D.J., Gonda, D.K., and Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* 243, 1576-1583.

Chen, Z., Niles, E.G., and Pickart, C.M. (1991). Isolation of a cDNA encoding a mammalian multiubiquitinating enzyme ($E2_{25KD}$) and overexpression of the functional enzyme in *Escherichia coli*. *J.*

Biol. Chem. 266(24), 15698-15704.

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell*, 79, 13-21.

Cleaver, J.E., and Kraemer, K.H. (1995). Xeroderma pigmentosum. In: The metabolic basis for inherited disease. Vol.II, Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., eds. (New York: McGraw-Hill Book Co.).

Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J., Miller, J.C., Dietrich, W.F., Weaver, A., Lincoln, S.E., Steen, R.G., Stein L.D., Nadeau, J.H., and Lander, E.S. (1993). A genetic linkage map of the mouse: current applications and future prospects. *Science* 262, 57-66.

Friedberg, E.C., Walker, G.C., and Siede W. (1995). DNA repair and mutagenesis American Society for Microbiology press, Washington, D.C.

Galavazi, G., Schenk, H., and Bootsma, D. (1966). Synchronization of mammalian cells *in vitro* by inhibition of the DNA synthesis. *Exp. Cell. Res.* 41, 428-437.

Gibbs, S., Fijneman, R., Wiegant, J., van Kessel, A.G., van de Putte P., and Backendorf, C. (1993). Molecular characterization and evolution of the SPRR family of keratinocyte differentiation markers encoding small proline-rich proteins. *Genomics* 16(3), 630-637.

Hoeijmakers, J.H.J. (1994). Human nucleotide excision repair syndromes:molecular clues to unexpected intricacies. *Eur. J. Cancer 30A*, 1912-1921.

Itin, P.H., and Pittelkow M.R. (1990). Trichothiodystrophy: review of sulfur-deficient brittle hair syndromes and association with the ectodermal dysplasias. *J. Americ. Acad. Dermatol.* 22, 705-717.

Jentsch, S. (1992). The ubiquitin conjugating system. *Annu. Rev. Genet.*, 26, 179-207.

Jentsch, S., McGrath, J.P., and Varshavsky, A. (1987). The yeast DNA repair gene *RAD6* encodes a ubiquitin-conjugating enzyme. *Nature*, 329, 131-134.

Koken, M.H.M., Reynolds, P., Jaspers-Dekker, I., Prakash, L., Prakash, S., Bootsma, D., and Hoeijmakers, J.H.J. (1991). Structural and functional conservation of two human homologs of the yeast DNA repair gene RAD6. *Proc. Natl. Acad. Sci. U.S.A.* 88, 8865-8869.

Legerski, R.J., Liu, P., Li, L., Peterson, C.A., Zhao, Y., Leach, R.J. Naylor., S.L. and Siciliano, M.J. (1994). Assignment of xeroderma pigmentosum group C (XPC) gene to chromosome 3p25. *Genomics* 21(1), 266-269.

Lehmann, A.R. (1987). Cockayne's syndrome and trichothiodystrophy: defective repair without cancer. *Cancer Rev.* 7, 82-103.

Madura, K. and Prakash S. (1990). Transcript levels of the *Saccharomyces cerevisiae* DNA repair gene *RAD23* increase in response to UV light and in meiosis but remain constant in the mitotic cell cycle. *Nucl. Acids Res.* 18(16), 4737-4742.

McCready, S. (1994). Repair of 6-4 photoproducts and cyclobutane pyrimidine dimers in *rad* mutants of *Saccharomyces cerevisiae*. *Mut. Res.* 315, 261-273.

Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enemoto, T., Takio K., Tanaka, K., van der Spek, P.J., Bootsma, D., Hoeijmakers, J.H.J., and Hanaoka, F. (1994). Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homolog of yeast RAD23. *EMBO J.*, 13, 1831-1843.

Miller, R.D., Prakash, L.,and Prakash, S. (1982). Defective excision of pyrimidine dimers and interstrand crosslinks in *rad7* and *rad23* mutants of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 188, 235-

239.

Nance, M.A. and Berry S.A. (1992). Cockayne syndrome: review of 140 cases. *Am. J. Hum. Genet.* 42, 68-84.

Pinkel, D., Straume, T., and Gray, J. (1986). Cytogenetic analysis using quantitative highly sensitive, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA*, 83, 2934-2938.

Rheinwald, J.G. and Green, H. (1975). Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell* 6, 331-344.

Sambrook, J., Fritsch, E.F. and Maniatis, T.E. (1989). *Molecular cloning, a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74, 5463-5467.

Spence, J., Sadis, S., Haas, A.L., Finley, D. (1995). A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol. Cell Biol.* 15, 1265-1273.

van der Spek, P.J., Smit, E.M.E., Beverloo, H.B., Sugasawa, K., Masutani, C., Hanaoka, F., Hoeijmakers, J.H.J., and Hagemeijer, A. (1994). Chromosomal localization of three repair genes: the xeroderma pigmentosum group C gene and two human homologs of yeast RAD23. *Genomics*, 23, 651-658.

Van Hoffen, A., Venema, J., Meschini, R., van Zeeland, A.A., and Mullenders, L.H.F. (1995). Transcription-coupled repair removes both cyclobutane pyrimidine dimers and 6-4 photoproducts with equal efficiency and in a sequential way from transcribed DNA in xeroderma pigmentosum group C fibroblasts. *EMBO*, 14, 360-367.

Van Houten, B. (1990). Nucleotide excision repair in Escherichia coli. *Microbiol Rev.* 54, 18-51.

Venema, J., van Hoffen, A., Natarajan, A.T., van Zeeland, A.A., and Mullenders, L.H.F. (1989). The residual repair capacity of xeroderma pigmentosum complementation group C fibroblasts is highly specific for transcriptionally active DNA. *Nucl. Acids Res.* 18, 443-448.

Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland A.A., and Mullenders L.H.F. (1991). Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. *Mol. Cell. Biol.* 11, 4128-4134.

Verhage, R., Zeeman, A.M., de Groot, N., Gleig, F., Bang, DD. van de Putte, P., and Brouwer, J. (1994). The *RAD7* and *RAD16* genes are essential for repair of non-transcribed DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 14, 6135- 6142.

Vermeulen, W., van Vuuren, A.J., Appeldoorn, E., Weeda, G., Jaspers, N.G.J., Priestly, A., Arlett, C.F., Lehmann, A.R., Stefanini, M., Mezzina, M., Sarasin, D., Bootsma, D., Egly, J-M., and Hoeijmakers J.H.J. (1995). Cold Spring Harbor Symp. Quant. Biol. (1994 in press).

Walker, G.C. (1985). Inducible DNA repair systems. *Annu. Rev. Biochem.*, 54, 425-457.

Watkins, J.F., Sung, P., Prakash, L. and Prakash, S. (1993). The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol. Cell. Biol.*, 13, 7757-7765.

CHAPTER

VII

Requirement of HHR23B, a human RAD23 homologue, for nucleotide excision repair *in vitro* as a stimulatory factor of the XPC protein

REQUIREMENT OF HHR23B, A HUMAN RAD23 HOMOLOGUE, FOR NUCLEOTIDE EXCISION REPAIR IN VITRO AS A STIMULATORY FACTOR OF THE XPC PROTEIN.

Kaoru Sugasawa[1,2], Chikahide Masutani[3], Akio Uchida[1,3], Takafumi Maekawa[3], Peter J. van der Spek[2], Dirk Bootsma[2], Jan H.J. Hoeijmakers[2], and Fumio Hanaoka[1,3].

[1] The Institute for Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama 351-01, Japan, [2] Department of Cell Biology and Genetics, Medical Genetic Centre, Erasmus University, Rotterdam, The Netherlands and [3] Institute for Molecular and Cellular Biology, Osaka University, 1-3 Yamada-oka Suita, Osaka 565, Japan.

SUBMITTED, 1995.

ABSTRACT

A protein complex which specifically complements defects of XP-C cell extracts *in vitro*, was previously purified to near homogeneity from HeLa cells. The complex consists of two tightly associated proteins: the XPC gene product and HHR23B, one of two human homologues of the *Saccharomyces cerevisiae* repair gene RAD23 (Masutani *et al.*, EMBO J. 13: 1831-1843, 1994). To elucidate the roles of these proteins in 'genome-overall' repair, we expressed the XPC protein in a baculovirus system and purified it to near homogeneity. The recombinant human XPC protein (rhXPC) exhibited a high affinity for single-stranded DNA and corrected the repair defect in XP-C cell extracts without extra addition of recombinant HHR23B protein (rHHR23B). However, Western blot experiments revealed that XP-C cell extracts contained excess endogenous HHR23B protein, which might be able to form a complex upon addition of the rhXPC protein. To investigate the role of HHR23B, we fractionated the XP-C cell extracts and constructed a reconstituted system in which neither endogenous XPC nor HHR23B proteins were present. In this assay system, rhXPC alone corrected the repair defect only very weakly, while significant enhancement of the correcting activity was observed upon co-addition of recombinant HHR23B protein, indicating that HHR23B stimulates the XPC function in the genome-overall subpathway.

INTRODUCTION

Various environmental agents such as radiation and chemicals cause DNA damage which may lead to alterations in the genetic information. DNA repair plays a crucial role in the prevention of such mutagenesis and consequent carcinogenesis and/or cell death. Nucleotide excision repair (NER) is one of the most important DNA repair pathways because it eliminates a wide variety of base lesions including ultraviolet light (UV)-induced cyclobutane pyrimidine dimers and [6-4]photoproducts, as well as certain chemical adducts (8). The molecular mechanism of NER in *Escherichia coli* is now understood in detail (9, 13, 21, 38). In this organism, only six proteins, *uvrA*, *uvrB*, *uvrC*, *uvrD*, DNA polymerase I and DNA ligase, are sufficient to complete the NER reactions *in vitro*, whereas a much greater number of gene products are now known to be involved in eukaryotic NER reactions (14). The eukaryotic NER system consists of at least two distinct subpathways.
One of these, transcription-coupled repair, preferentially eliminates DNA damage which

hinders the advance of RNA polymerases on transcribed strands (4, 12, 25). Damage to the rest of the genome is repaired more slowly, and less efficiently for some lesions, by the 'genome-overall' repair mechanism.

A number of NER gene products have been genetically identified. Xeroderma pigmentosum (XP) is a human autosomal recessive NER disease, which is associated with extreme sensitivity to sunlight exposure and high incidence of skin abnormalities including cancer. XP cells are hypersensitive to UV-irradiation and complementation analysis by cell fusion has identified at least seven complementation groups, XP-A to XP-G, all of which show defects in early steps of the NER reaction (36). Two genes implicated in XP, *XPA* and *XPC*, have been cloned by transfection of XP cells with mouse genomic DNA (35) or a human cDNA expression library (19). Another set of mammalian NER mutants has been established with UV-sensitive rodent culture cell lines, in which at least eleven complementation groups have been identified (6). Several human genes have been cloned, which phenotypically correct the UV-sensitivity of these mutants. Among these *ERCC* (excision repair cross-complementing rodent repair deficiency) genes, *ERCC2*, *ERCC3*, *ERCC5* genes appeared to be identical to *XPD*, *XPB* and *XPG* genes, respectively (7, 27, 45). Additionally, the *ERCC6* gene has been shown to be responsible for complementation group B of another known NER disorder, Cockayne's syndrome (37). The XPG protein was found to be an endonuclease, implicated in the incision on the 3' side of DNA adducts (26). Another cloned *ERCC* gene product, ERCC1, has recently been suggested to be complexed with protein factors which complement ERCC groups 4 and 11 and XP group F (3, 39), and to be possibly involved in incision on the 5' side of DNA damage (1). More recently, the XPB and XPD proteins, both DNA helicases, were reported to be subunits of TFIIH, one of the basal transcription factors required for transcriptional initiation by eukaryotic RNA polymerase II (29, 30), revealing a dual functioning of these proteins in basal transcription and NER.

XP group C is unique in that its defect is limited to the genome-overall NER subpathway, transcription-coupled repair being normal (40). The converse is found in Cockayne's (24, 41). By transfection with a human cDNA expression library, Legerski and Peterson isolated a partial cDNA clone, which corrected the UV-sensitivity of XP-C cells (19). This gene, named *XPC*, shares limited homology with a known *S. cerevisiae* NER protein, RAD4. The level of *XPC* mRNA was greatly reduced in most XP-C cell lines tested (19), and nonsense and missense mutations have been identified in several cell lines (20).

Development of a cell-free system which faithfully reproduces the in vivo reactions is a powerful strategy to investigate complicated biochemical processes, and several

laboratories have reported such cell-free systems for detection of NER (2, 32, 33, 43, 46). We have also developed a cell-free system which uses whole-cell extracts and UV-irradiated simian virus 40 (SV40) minichromosomes as templates (33). This system reproduced the NER defects of all XP complementation groups (22), and enabled us to identify biochemically a protein factor which corrects the defect in XP-C cell extracts (23). This factor, named XP-C correcting protein, was purified from HeLa cell nuclear extracts and was found to be a tight complex of two polypeptides with apparent molecular masses of 125 and 58 kDa. Cloning and sequence analysis of cDNAs encoding these proteins revealed that the 125-kDa polypeptide is an N-terminally extended version of the XPC gene product reported previously (23). The 58-kDa species was identified as one of two human homologues of the *S. cerevisiae* RAD23 NER protein, designated as HHR23B (human homologue of RAD23) (23). Both the RAD23 and HHR23B proteins, as well as HHR23A, another human RAD23 cognate, contain ubiquitin-like sequences on their N-termini. For yeast the importance of this domain in NER function has been suggested genetically (44). Unfortunately, for none of the HHR23 genes corresponding human or rodent NER mutants have been identified. It has been recently reported that yeast RAD23 protein may promote interaction with RAD14 and yTFIIH (10), but the biological significance of complex formation between the XPC and HHR23B proteins and the functions of HHR23B in NER, particularly in the genome-overall repair subpathway, have not yet been clarified. In the present study, recombinant XPC and HHR23B proteins were prepared to investigate the roles of these proteins in NER reactions *in vitro*.

MATERIALS AND METHODS

Cell culture and media. Human 293 cells, XP2OSSV (group A), XP4PASV and XP3KA (group C) cells were grown at 37°C in Dulbecco's modified Eagle's medium (Nissui) supplemented with 10% fetal bovine serum. An insect cell line, *Spodoptera frugiperda* Sf9, was cultured at 27°C in TNM-FH medium supplemented with 10% heat-inactivated (56°C, 30 min) fetal bovine serum. TNM-FH medium was prepared from Grace's insect cell culture medium (Gibco-BRL), TC yeastolate and TC lactalbumin hydrolysate (Difco) as described elsewhere (34).

Construction of recombinant baculoviruses. Plasmid pBS.XPC, which carries human *XPC* cDNA cloned into the *Not*I site of pBluescript II KS+ (23), was digested with *Not*I and the resulting 3.6-kb cDNA fragment was isolated. The fragment was inserted into the *Not*I site of a baculovirus transfer vector, pVL1393 (Invitrogen), to generate plasmid pVL.XPC. For simultaneous expression of XPC and HHR23B proteins, the *HHR23B* cDNA cloned into pUC19 (23) was digested with *Ban*I and *Bln*I to obtain the 1.5-kb cDNA fragment.

After treatment with Klenow fragment and subsequent addition of *Eco*RI linkers, the cDNA fragment was cloned into the *Eco*RI site of pAcUW31 (Clontech) generating plasmid pAcUW.HHR23B. The translational initiation site of XPC in the plasmid pBS.XPC was converted to a *Nde*I site using an oligonucleotide, 5'-GACAAGCAACATATGGCTCGGAAAC-3' and a site-directed mutagenesis system, Mutan-K (Takara Shuzo), essentially according to the method of Kunkel *et al.* (16). The resulting plasmid, pBS.XPC-*Nde*I, was digested with *Nde*I and *Bln*I to obtain the 3.3 kb XPC cDNA fragment. This cDNA fragment and *Bam*HI-digested pAcUW.HHR23B were blunt-ended with Klenow fragment and then ligated to each other to generate plasmid pAcUW(XPC-HHR23B). Sf9 cells were cotransfected with either of the constructed plasmids and the BaculoGold DNA (Pharmingen) to produce recombinant viruses, designated as vVL.XPC and vAcUW(XPC-HHR23B), respectively. Both viruses expressed the human *XPC* gene under control of the polyhedrin promoter, while transcription of HHR23B in vAcUW(XPC-HHR23B), was driven by the p10 promoter (Fig. 1A).

Baculovirus infection and extract preparation from the infected cells. Monolayers of Sf9 cells were infected with the recombinant baculoviruses at room temperature for 1 h at a multiplicity of infection of 5-10. At 3 days post-infection, cells were collected by low-speed centrifugation and washed twice with ice-cold phosphate-buffered saline. To examine the total cellular proteins, the cell pellets were lysed in 10 volumes, relative to the packed cell volume (10 x PCV), of 2 x SDS sample buffer [1 x concentration: 62.5 mM Tris-HCl (pH 6.8), 1% SDS, 5% glycerol, 2% 2-mercaptoethanol] and heated to 95°C for 10 min. For fractionation, the pellets of the infected cells were suspended in 8 x PCV of ice-cold NP lysis buffer [25 mM Tris-HCl (pH 8.0), 1 mM EDTA, 10% glycerol, 1% Nonidet P-40, 1 mM dithiothreitol (DTT), 0.25 mM phenylmethyl-sulfonyl fluoride (PMSF), 50 μM ethylene-glycol-bis(ß-aminoethyl ether)-*N, N, N', N'*-tetraacetic acid (EGTA), 0.2 μg/ml of aprotinin, 0.2 μg/ml of leupeptin, and 0.1 μg/ml of antipain]. All subsequent steps were carried out at 4°C. After incubation on ice for 30 min, the suspension was centrifuged at 800 x G for 10 min to obtain the supernatant fraction, designated as the fraction 'S1'. The remaining precipitate was resuspended by gentle pipetting in 8 x PCV of NP lysis buffer containing 0.3 M NaCl, and then incubated on ice for 30 min with occasional agitation. The suspension was centrifuged at 12,000 x g for 15 min to be divided into the supernatant (S2) and precipitate (P) fractions. To examine the proteins remaining in the fraction 'P', the pellets were homogenized in NP lysis buffer containing 0.3 M NaCl by sonication.

Purification of recombinant proteins from insect cells. For purification of recombinant proteins, ten 150-mm culture dishes of Sf9 cells were routinely infected. Behavior of the recombinant proteins was monitored by SDS-PAGE and Western blot analysis using antibodies raised against each protein (see below). For purification of recombinant XPC-HHR23B protein complex, the fraction 'S2' prepared from vAcUW(XPC-HHR23B)-infected Sf9 cells was loaded onto a phosphocellulose column (Whatman, P11; 6 ml) equilibrated with buffer A [25 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10% glycerol, 0.01% Triton X-100, 1 mM DTT, 0.25 mM PMSF] containing 0.3 M NaCl. The column was washed with the same buffer and the adsorbed proteins were eluted with buffer A containing 1 M NaCl. The eluate was adjusted at 0.6 M NaCl by dilution with buffer A and loaded onto a single-stranded DNA-cellulose column (Sigma, 4.3 mg DNA/g; 2 ml) equilibrated with buffer A containing 0.6 M NaCl. The column was washed with the same buffer and then proteins were eluted with buffer A containing 1.5 M NaCl. The eluate was dialyzed against buffer A containing 0.3 M NaCl and stored at -80°C. The recombinant XPC protein which was free of HHR23B protein was purified from Sf9 cells infected with vVL.XPC by the same procedures.

For purification of recombinant HHR23B protein (rHHR23B), the fraction 'S1' from vAcUW(XPC-HHR23B)-infected Sf9 cells was used as a starting material. The conductivity of the fraction 'S1' was adjusted to that of buffer A containing 0.05 M NaCl and loaded onto a phosphocellulose column equilibrated with the same buffer. The following two chromatography steps were performed by the use of an FPLC system (Pharmacia). Flow-through fractions from the phosphocellulose column were collected and then loaded at 0.1 ml/min onto a HiTrap-Q column (Pharmacia; 5 ml) equilibrated with buffer A containing 0.05 M NaCl. After the column was washed with 15 ml of the same buffer, proteins were eluted with 50 ml of a linear gradient of 0.05 to 0.4 M NaCl in buffer A. The recombinant HHR23B protein was eluted at around 0.25 M NaCl. The peak fractions were collected and then further loaded at 0.5 ml/min onto a Bio-scale CHT2-I hydroxy-apatite column (Bio-Rad; 2 ml) equilibrated with buffer A containing 0.2 M KCl. The column was washed with 10 ml of the same buffer and elution was carried out with 24 ml of a linear gradient of 0 to 0.5 M potassium phosphate (pH 7.5) in buffer A containing 0.2 M KCl. The recombinant HHR23B protein was eluted around 0.06 M potassium phosphate. The peak fractions were adjusted to 35% saturation of ammonium sulfate, kept on ice for 30 min and then centrifuged at 12,000 x g for 15 min. The precipitates were suspended in buffer A and dialyzed against buffer A containing 0.3 M NaCl. Insoluble materials were removed by centrifugation and the supernatant fraction was stored at -80°C.

Preparation and fractionation of whole-cell extracts. Whole-cell extracts for cell-free repair reactions were prepared as described previously. For fractionation, conductivity of the extracts was adjusted to buffer B [25 mM *N*-(2-hydroxyethyl)piperazine-*N'*-2-ethanesulfonic acid (HEPES)-KOH (pH 7.9), 1 mM EDTA, 10% glycerol, 0.01% Triton X-100, 1 mM DTT and 0.25 mM PMSF] containing 0.2 M KCl by the addition of 3 M KCl. Approximately 50 mg of protein from the extract was loaded onto a phosphocellulose column (5ml) equilibrated with buffer B containing 0.2 M KCl. The column was washed with the same buffer and the adsorbed proteins were eluted with buffer B containing 1 M KCl. The peak fractions from the flow-through (CFI) and the eluate (CFII) were concentrated by dialysis against buffer B containing 0.1 M KCl and 20% sucrose, and stored at -80°C. Shivji *et al.* (31) originally loaded whole-cell extracts onto a phosphocellulose column at 0.1 M KCl. We examined the effect of the salt concentration and found that properties of the reconstituted systems were essentially unaffected between 0.1 and 0.2 M KCl. However, a higher level of repair synthesis was obtained with CFII bound at 0.2 M KCl in our system.

Cell-free DNA repair assay. The standard reaction mixture (20 μl) contained 40 mM creatine phosphate-Tris (pH 7.7), 1 mM DTT, 5 mM $MgCl_2$, 2 mM ATP, 50 μM each of dATP, dGTP and dTTP, 10 μM [α-^{32}P]dCTP (37-74 kBq), phosphocreatine kinase (Sigma, Type I; 0.5 μg) bovine serum albumin (6.4 μg), whole-cell extracts (80 μg of protein), unirradiated pUC19 RFI DNA (0.3 μg), and UV-irradiated (400 J/m^2) or unirradiated SV40 minichromosomes (0.3 μg of DNA). Where indicated, fractionated extracts (CFI and/or CFII) and purified proteins [replication protein A (RPA), proliferating cell nuclear antigen (PCNA), XPC and HHR23B] were substituted for whole-cell extracts. The reactions were incubated at 30°C. for 3 h. DNA was purified from the reaction mixtures, linearized by *Eco*RI digestion and then electrophoresed in 1% agarose gels as described previously (33). Autoradiography was performed at -80°C with Fuji New RX X-ray film and Kodak intensifying screens. The incorporation of radioactive materials into viral DNA was quantified with a Fujix BAS2000 Bio-Imaging Analyzer.

Purification of RPA and PCNA. The RPA (15) and PCNA (18) were purified from HeLa and 293 cells as described previously.

Antibodies. For immunization, rhXPC was partially purified from the fraction 'P' prepared from vVL.XPC-infected Sf9 cells. The fraction 'P' was homogenized by sonication in buffer A containing 0.5 M NaCl. After the mixture was centrifuged at 12,000 x g for 15 min, the resultant supernatant was mixed with single-stranded DNA-cellulose equilibrated with the same buffer and rotated gently at 4°C overnight. The resin was collected by low-speed centrifugation, washed with buffer A containing 0.5 M NaCl and then packed into a column. After the column was further washed with the same buffer, bound proteins were eluted with buffer A containing 2 M NaCl and 50 % ethyleneglycol. The peak fractions determined by SDS-PAGE were dialyzed against buffer A containing 0.5 M NaCl. rhXPC was purified by excision from preparative SDS-PAGE and used for immunization of rabbits. Anti-XPC antibodies were affinity-purified using the partially purified rhXPC fraction.

To obtain anti-HHR23B antibodies, HHR23B was expressed in *E.coli* as a fusion protein with either glutathione S-transferase (GST) or maltose-binding protein (MBP), using plasmids pGEX-2T (Pharmacia) or pMAL-c2 (New England Biolabs), respectively. The GST-HHR23B fusion protein was purified with glutathione-Sepharose (Pharmacia) under standard conditions and used for immunization of rabbits. Anti-HHR23B antibodies were affinity-purified with the MBP-HHR23B fusion protein. As the MBP-HHR23B protein was not bound to amylose resins, it was partially purified by conventional column chromatography using HiTrap-Q and Mono Q.

Other methods. SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described by Laemmli (17). For Western blot analyses, electrophoresed proteins were transferred onto PVDF membranes (Immobilon-P; Millipore) at 8 V/cm for 12 h in ice-cold transfer buffer (50 mM Tris, 38.4 mM glycine, 0.01% SDS, and 15% methanol). The membranes were successively incubated in blocking buffer (5% skim milk in 25 mM Tris-HCl (pH 7.5), 0.15 M NaCl, 0.1% Tween 20), first antibodies (anti-XPC) or anti-HHR23B) in blocking buffer, and then anti-rabbit F(ab')2 antibodies conjugated with horseradish peroxidase (Amersham). Detection was carried out with the ECL system (Amersham) and Fuji New RX X-ray film. Protein concentration was measured according to the method of Bradford (5) using bovine serum albumin as a standard and reagents purchased from Bio-Rad Laboratories.

RESULTS

Expression and extractability of recombinant XPC and HHR23B proteins.

To obtain large amounts of human XPC and HHR23B proteins, two types of recombinant baculoviruses were coonstructed (Fig. 1A). One of the recombinant viruses, vVL.XPC, overexpressed XPC protein in insect cells, while the other, vAcUW(XPC-HHR23B), expressed both XPC and HHR23B proteins simultaneously. As shown in Fig. 1B, the recombinant human XPC protein (rhXPC) was easily detected by Coomassie Blue staining in total protein extracts from Sf9 cells infected with either of the recombinant viruses. Expression of the recombinant HHR23B protein (rHHR23B) was also detected in the lysate from vAcUW(XPC-HHR23B)-infected cells.

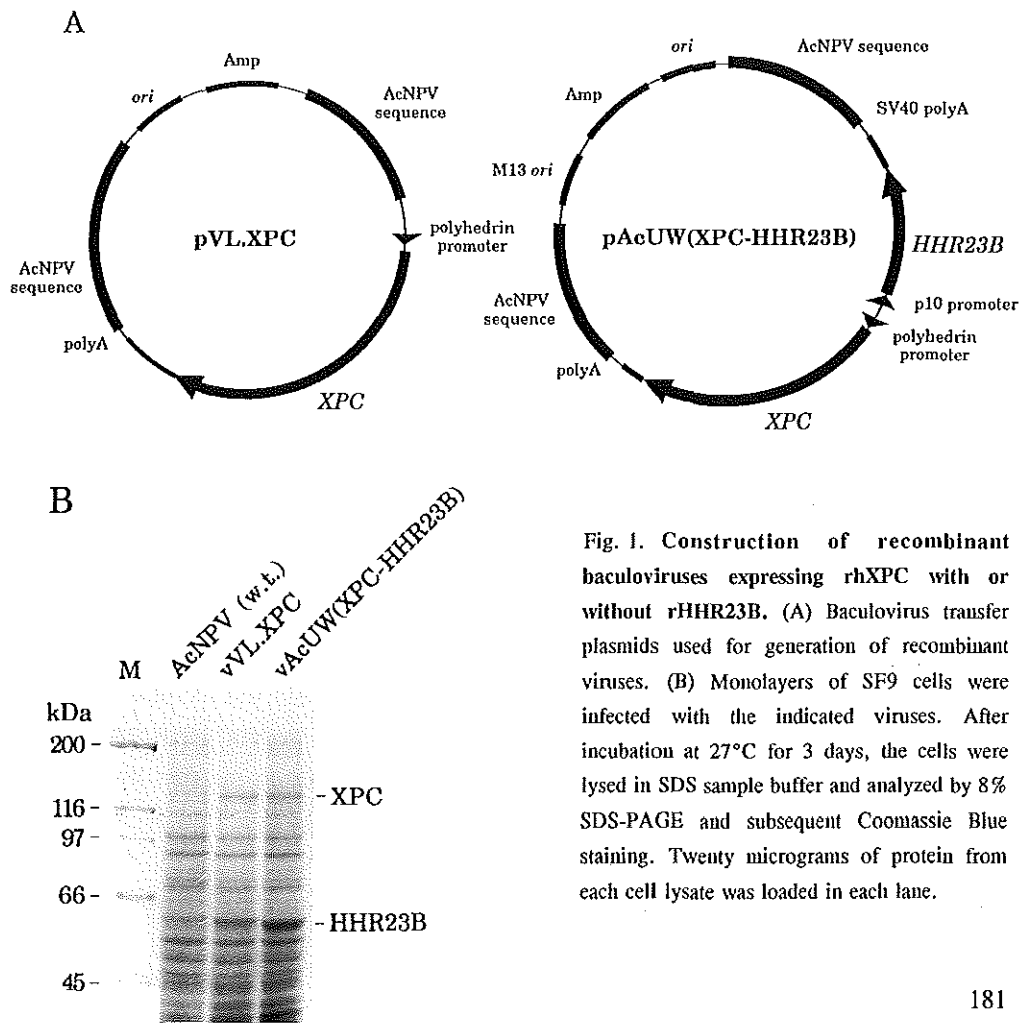


Fig. 1. Construction of recombinant baculoviruses expressing rhXPC with or without rHHR23B. (A) Baculovirus transfer plasmids used for generation of recombinant viruses. (B) Monolayers of SF9 cells were infected with the indicated viruses. After incubation at 27°C for 3 days, the cells were lysed in SDS sample buffer and analyzed by 8% SDS-PAGE and subsequent Coomassie Blue staining. Twenty micrograms of protein from each cell lysate was loaded in each lane.

To examine the extractability of these expressed proteins, the infected Sf9 cells were divided into three fractions. The infected cells were extracted with hypotonic buffer containing 1% Nonidet P-40 (NP lysis buffer; see MATERIALS AND METHODS) to obtain the supernatant (S1) fraction. The precipitate fraction was further extracted with buffer containing 0.3 M NaCl to obtain the supernatant (S2) and the precipitate (P) fractions. As shown in Fig. 2A, the vast majority of rhXPC remained unextracted. These extraction properties of rhXPC were essentially unaffected by coexpression of rHHR23B. In marked contrast, most of the rHHR23B expressed in vAcUW (XPC-HHR23B)-infected cells was recovered in the fraction 'S1', suggesting that most of rhXPC and rHHR23B were not complexed with each other when coexpressed in the insect cells.

The rhXPC remaining in the fraction 'P' was solubilized by sonication and then partially purified by single-stranded DNA-cellulose column chromatography. Using this protein fraction as an antigen, anti-XPC polyclonal antibodies were obtained. These antibodies cross-reacted with a 125-kDa protein species in the XPC protein complex purified from HeLa cells as well as in total lysates from repair-proficient human 293 cells and a XP-A cell line, XP2OSSV, but not in lysates from two XP-C cell lines, XP4PASV and XP3KA (Fig. 3A) confirming that the band corresponded to the XPC protein. Polyclonal antibodies were raised against GST-HHR23B fusion protein produced in *E.coli*, and affinity-purified with a MBP-HHR23B fusion protein. The antibodies reacted specifically with 58-kDa protein in the HeLa XPC complex and also in the 293 and XP-A cell lysates (Fig. 3B). In contrast to XPC, the 58-kDa bands were also present in the XP-C cell lysates, indicating that HHR23B protein was normally expressed in the XP-C cells. Furthermore, the intensity of XPC protein band in 20 μg of the 293 cell extract was much lower than the signal of 5 ng of the purified XPC/HHR23B complex (Fig. 3A), while the signal of HHR23B in the same amount of the extract was comparable to 30 ng of the purified complex (Fig. 3B). Thus HHR23B protein is present in large excess over XPC protein even in repair-proficient cells, suggesting that most of the HHR23B protein is not complexed with XPC protein.

Using these antibodies, expression and extractability of rhXPC and rHHR23B were re-examined. Although most of the rhXPC remained in the fraction 'P', Western blot analyses revealed that a small but significant portion of the protein was extracted in the fraction 'S2' from Sf9 cells infected with either vVL.XPC or vAcUW(XPC-HHR23B) (Fig. 2B). Similarly, a portion of rHHR23B was found to be resistant to the first hypotonic extraction, and was present in the fractions 'S2' and 'P' (Fig. 2C and 2D). Thus the fractions 'S2' and 'P' from coexpressing cells contained both rhXPC and rHHR23B. Although rHHR23B in the fraction 'S1' was detected as a single band, two HHR23B bands, with

apparent molecular masses of 61 and 58 kDa, were detected in the fractions 'S2' and 'P' when larger amounts of the fractions were subjected to Western blot analyses (Fig. 2D). Since the lower band of the doublet co-migrated with HHR23B in human cell extracts, a subpopulation of rHHR23B expressed in insect cells may undergo some post-translational modification, such as phosphorylation.

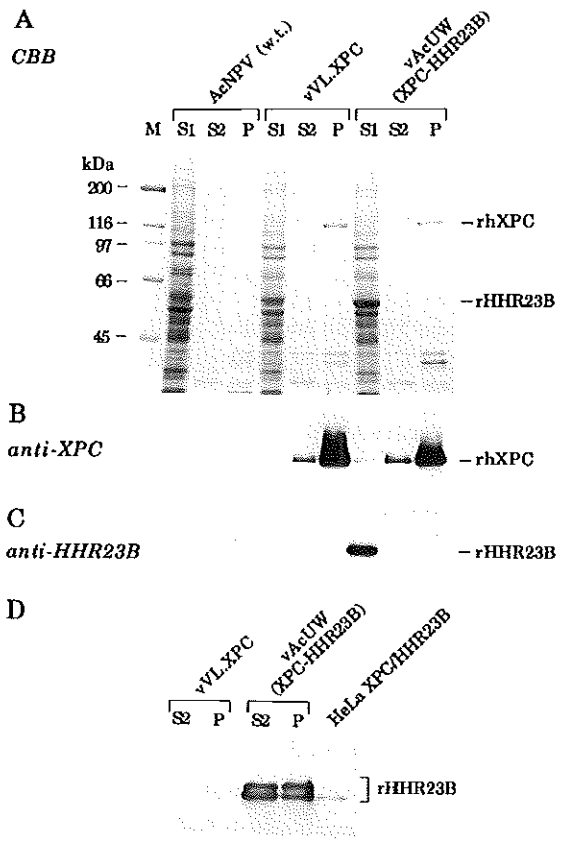


Fig. 2. **Extraction properties of rhXPC and rHHR23B expressed in insect cells.**
(A) Monolayers of Sf9 cells in 60-mm dishes were infected with the indicated viruses and incubated at 27°C for 3 days. The cells were collected and fractionated as described in MATERIALS AND METHODS. The volume of each of the resulting fractions, 'S1',' S2' and 'P', was adjusted to 500 μl, and 3.3 μl of each fraction was analyzed by 8% SDS-PAGE and subsequent Coomassie Blue staining. (B and C) The same samples as shown in A were transferred onto a PVDF membrane and subjected to Western analysis using anti-XPC (B) or anti-HHR23B (C) polyclonal antibodies. In this experiment, 0.33 μl of each fraction was loaded per lane. (D) The rHHR23B in 5 μl aliquots of the fractions 'S2' and 'P' from cells infected with the indicated viruses was visualized by Western blot analysis.

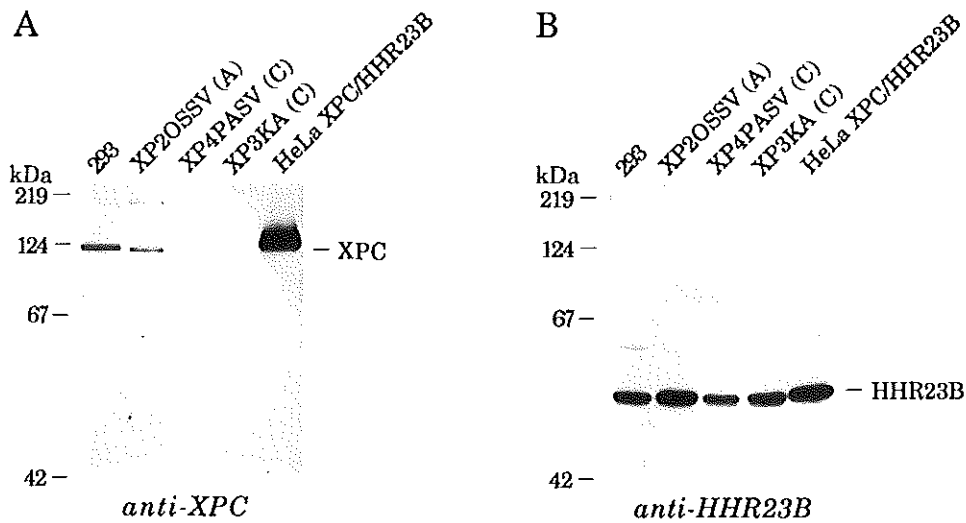


Fig. 3. **Specificity of anti-XPC and HHR23B antibodies.**
Total cell lysates (each 20 μg protein) from 293, XP2OSSV (group A), XP4PASV and XP3KA (group C) were analyzed by Western blotting with anti-XPC (A) or anti-HHR23B (B) polyclonal antibodies. Five (A) or thirty (B) nanograms of XPC/HHR23B complex purified from HeLa cells was loaded in parallel.

Purification of recombinant XPC and HHR23B proteins.

We used the fraction 'S2' from coexpressing cells as a starting material for purification of rhXPC/rHHR23B complex, because the results in Fig. 2A showed that, in other fractions, 'S1' and 'P', the molar ratio of the two recombinant proteins were quite different. Purification was carried out using procedures similar to those published previously for the purification of XPC protein complex from HeLa cell nuclear extracts (23). Two steps of column chromatography using phosphocellulose and single-stranded DNA-cellulose gave a protein fraction consisting of three bands with apparent molecular masses of 125, 61, and 58 kDa on SDS-PAGE (Fig. 4A). The largest protein reacted with anti-XPC antibodies, and the smaller two reacted with anti-HHR23B antibodies as expected (Fig. 4A). Moreover, when the same purification were applied to the fraction 'S2' from Sf9 cells expressing rhXPC alone, the 61- and 58-kDa bands were not copurified with rhXPC (Fig. 4B), confirming that the two bands corresponded to subforms of rHHR23B which were complexed with rhXPC. The HHR23B-free rhXPC bound to single-stranded DNA-cellulose in the presence of 0.6 M NaCl, as observed for the XPC/HHR23B complex. HHR23B itself did not bind to single-stranded DNA-cellulose (Data not shown), suggesting that the DNA binding activity can be attributed to XPC protein, not to HHR23B.

We also purified XPC-free rHHR23B from the fraction 'S1' prepared from insect cells infected with the coexpression virus. Purification was performed with three steps of column chromatography and ammonium sulfate precipitation, which gave a protein fraction containing a 58-kDa polypeptide cross-reactive with anti-HHR23B antibody (Fig. 4C).

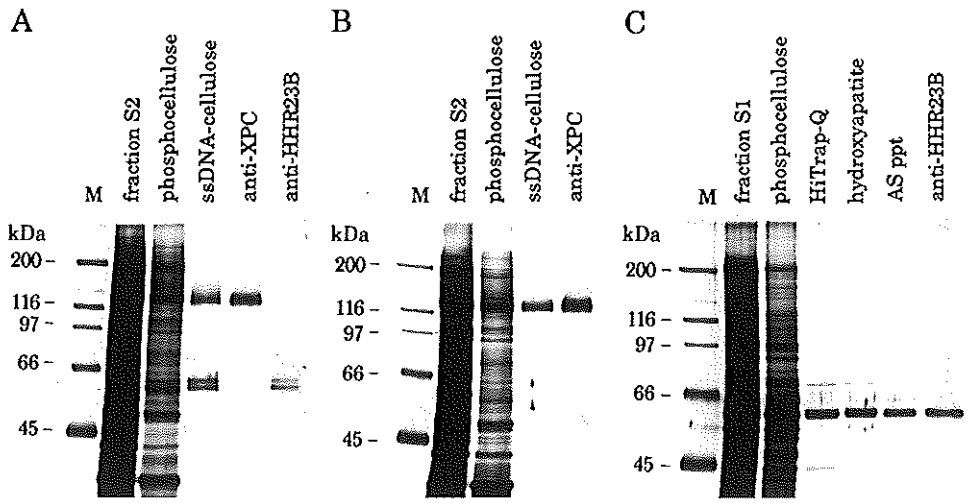


Fig. 4. **Purification of recombinant proteins.**
(A) Purification of rhXPC/rHHR23B complex. Aliquots of samples at each purification step indicated above the lanes were subjected to 8% SDS-PAGE. Protein bands were visualized by silver staining. The final purified sample was subjected to Western blot analysis using anti-XPC or anti-HHR23B antibodies as indicated. (B) Purification of free rhXPC. (C) Purification of free rHHR23B.

rhXPC alone can complement the repair defect of XP-C cell extracts.

Each recombinant protein fraction obtained above was assayed for XPC correcting activity in our cell-free DNA repair system using UV-irradiated SV40 minichromosomes. As shown in Fig. 5, the rhXPC/rHHR23B complex stimulated the cell-free NER reactions in whole-cell extracts from the XP-C cell line XP4PASV, although the correcting activity of the recombinant protein complex was in all experiments lower than that of similar amounts of the authentic XPC/HHR23B protein complex purified from HeLa cells. This could be due to the difference of the protein modification between human cells and the insect cells. The rhXPC, free of HHR23B, also complemented the repair defect in XP4PASV whole-cell extracts. The rhXPC, free of HHR23B, also complemented the repair defect in XP4PASV whole-cell extracts. This observation was not surprising because Western blot analyses (Fig. 3) showed HHR23B protein to be normally expressed in XP4PASV cells. It is likely that

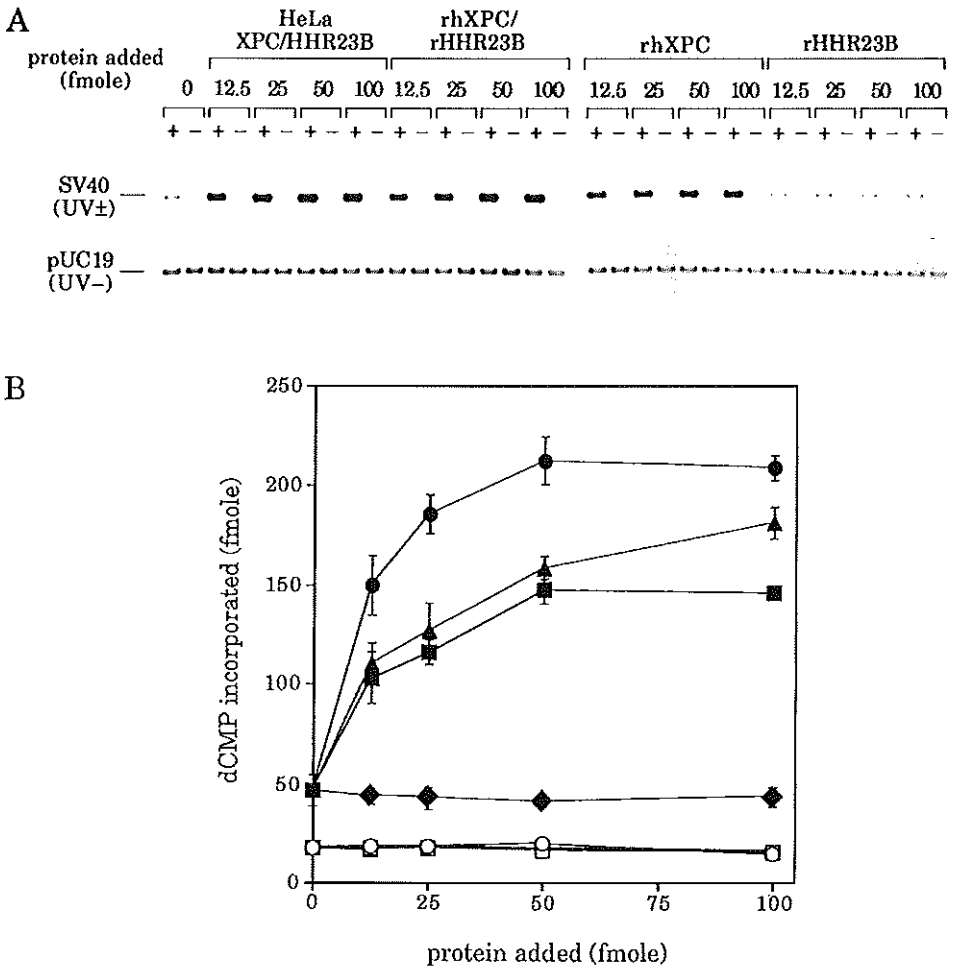


Fig. 5. **XP-C correcting activity of the purified recombinant proteins in the cell-free NER system using XP4PASV whole-cell extracts.** (A) UV-irradiated (UV+) or unirradiated (UV-) SV40 minichromosomes were incubated at 30°C for 3 h in standard reaction mixtures containing whole-cell extracts from XP4PASV cells and various amounts of purified proteins as indicated. DNA was purified, linearized by *Eco*RI, and then analyzed by 1% agarose gel electrophoresis and subsequent autoradiography. (B) Incorporation of ^{32}P-radioactivity into each viral DNA band shown in A was quantified. Closed symbols show the incorporation into UV-irradiated viral chromosomal DNA and open symbols show that into unirradiated viral DNA. Averages and experimental errors were calculated from thee experiments including the one shown in (A). (○ ●) XPC/HHR23B complex purified from HeLa cells. (▲ △) rhXPC/rHHR23B complex. (□ ■) free rhXPC. (♦ ◇) free rHHR23B. In the present work, the amount of XPC/HHR23B complex was calculated on the basis of the observation that the two proteins form a 1:1 complex. Therefore, the reactions labeled as "100 fmol of XPC/HHR23B" contained 100 fmol each of the two proteins.

exogenous rhXPC binds to endogenous HHR23B protein present in the XP4PASV cell extracts, resulting in reconstitution of functional rhXPC/HHR23B complex. We also tested the activity of rhXPC purified from the fraction 'P': the 'bulk' rhXPC hardly exhibited the XP-C correcting activity (data not shown). Therefore, only a small portion of rhXPC expressed in insect cells possesed the activity, which was preferentially extracted into the fraction 'S2'. As expected, rHHR23B alone could not stimulate the cell-free NER reactions by XP4PASV cell extract.

Fractionation and reconstitution of the cell-free DNA repair system with XP-C cell extract.

As whole-cell extract from XP-C cells contained endogenous HHR23B protein, it remained unclear whether the HHR23B protein was necessary for the cell-free NER reactions. To answer this question, we depleted HHR23B protein from the cell-free NER reactions by fractionation and reconstitution of whole-cell extracts. Phosphocellulose collumn chromatography of repair-proficient cell extracts yielded two protein fractions, a flow-through fraction (CFI) and a bound fraction (CFII). Both CFI and CFII are required for the cell-free NER reactions, but CFI can be replaced by two purified proteins, replication protein A (RPA) and proliferating cell nuclear antigen (PCNA) (Fig. 6A; see Ref. 31).

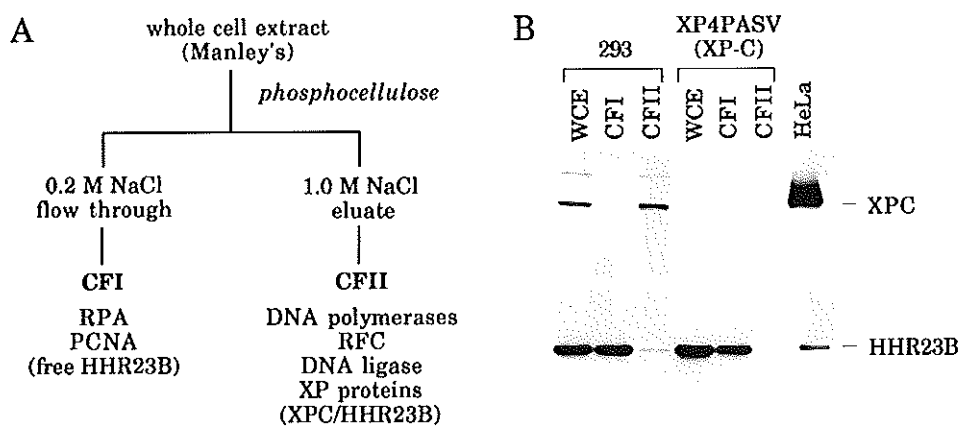


Fig. 6. **Fractionation of whole-cell extracts for the cell-free NER system.** (A) Scheme of fractionation of whole-cell extracts. (B) Whole-cell extracts from 293 or XP4PASV cells were fractionated as shown in A, and the presence of XPC (upper panel) or HHR23B (lower panel) protein in CFI and CFII was visualized by Western blotting. Forty micrograms of protein from whole-cell extracts (WCE), 15 μg protein of CFI and 10 μg protein of CFII were loaded per lane.

As shown in Fig. 7, the NER reactions using SV40 minichromosomes were completely dependent on both CFI and CFII derived from 293 whole-cell extracts. Purified RPA and PCNA could substitute for CFI, and omission of either of the two proteins resulted in reduction of the repair synthesis level. Western blot analysis revealed that most of the HHR23B protein in 293 cell extracts was recovered in CFI upon this fractionation procedure, whereas practically all XPC protein was present in CFII (Fig. 6B). Figure 6B also shows that a small portion of HHR23B protein was detected in the CFII from 293 cells. It is very likely that this subpopulation of HHR23B protein is complexed with XPC protein, because CFII prepared from XP4PASV cells under the same conditions contained no detectable level of HHR23B protein (Fig. 6B). Therefore, the replacement of CFI with RPA and PCNA resulted in omission of endogenous free HHR23B protein from the cell-free NER reactions.

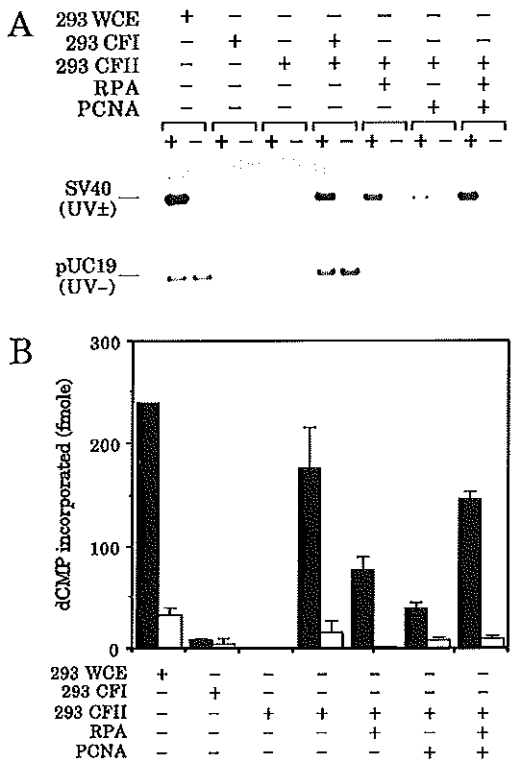


Fig. 7. **Reconstitution of the cell-free NER reactions.** (A) Cell-free NER reactions were carried out in which 293 WCE was replaced by the indicated (+) components. Amounts of the components used were 80 μg for WCE, 30 μg for CFI, 20 μg for CFII, 0.5 μg for RPA and 24 ng for PCNA. (B) Incorporation of ^{32}P-radioactivity into each viral band shown in A was measured. The solid bars indicate incorporation into UV-irradiated viral DNA and the open bars indicate that into unirradiated viral DNA. Averages and experimental errors were calculated from three experiments including the one shown in (A).

Stimulation of XP-C correcting activity by HHR23B protein in the reconstituted system.

The recombinant proteins were assayed for XP-C correcting activity in the reconstituted system using RPA, PCNA and CFII from XP4PASV cells. As shown in Fig. 8, the HeLa XPC/HHR23B complex stimulated repair synthesis in the reconstituted system to archieve about 3-fold stimulation over background synthesis. The rhXPC/rHHR23B

complex also stimulated repair synthesis in this system, but the maximum level of stimulation was approximately 40-50% of the HeLa XPC/HHR23B complex. Interestingly, free rhXPC alone showed only very weak stimulatory activity in the reconstituted system, in marked contrast to the results with whole-cell extracts (Fig. 5).

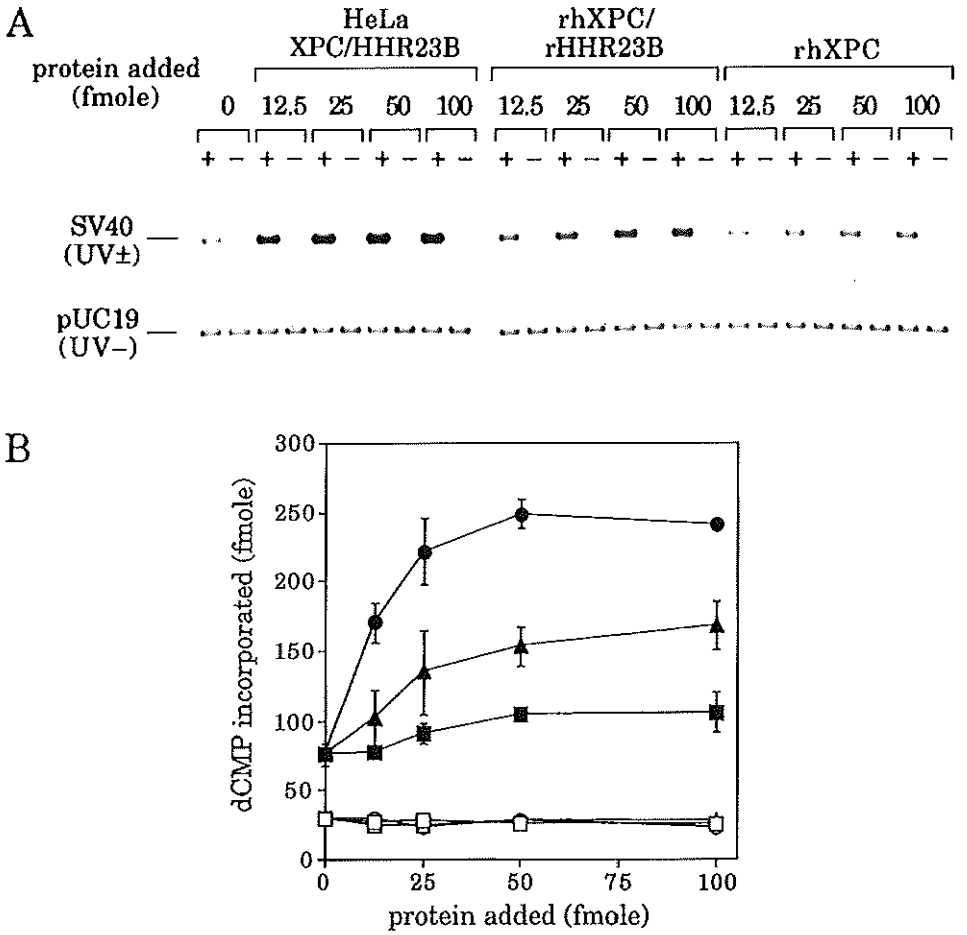


Fig. 8. **Free rhXPC could not stimulate repair synthesis in the reconstituted NER system using CFII from XP4PASV cells.** (A) Cell-free NER reactions were performed with 20 μg of XP4PASV CFII, 0.5 μg of RPA, 24 ng of PCNA, and various amounts of proteins purified from HeLa or insect cells as indicated. (B) The incorporation of radioactivity in each viral band shown in A was measured and plotted. The incorporation into UV-irradiated viral DNA is shown by closed symbols and that into unirradiated viral DNA is by open symbols. Averages and experimental errors were taken from three experiments including the one shown in (A). (○ ●) HeLa XPC/HHR23B complex, (▲ △) rXPC/rHHR23B complex, (□ ■) free rhXPC.

To examine the roles of HHR23B protein, rHHR23B was added to the reconstituted repair system in combination with rhXPC. In the presence of free rhXPC, rHHR23B stimulated repair synthesis in a dose-dependent manner up to a level nearly equivalent to that achieved by rXPC/rHHR23B complex (Fig. 9A). In contrast, rHHR23B showed little stimulatory activity by itself or in the presence of XPC protein prebound to HHR23B protein. Also when the amount of rhXPC was varied in the presence or absence of a fixed amount of rHHR23B, rhXPC again exhibited very weak stimulatory activity in the absence of rHHR23B but the co-addition of rHHR23B significantly enhanced the activity of rhXPC (Fig. 9B). These findings indicate that the HHR23B protein plays an important role in the mammalian NER system.

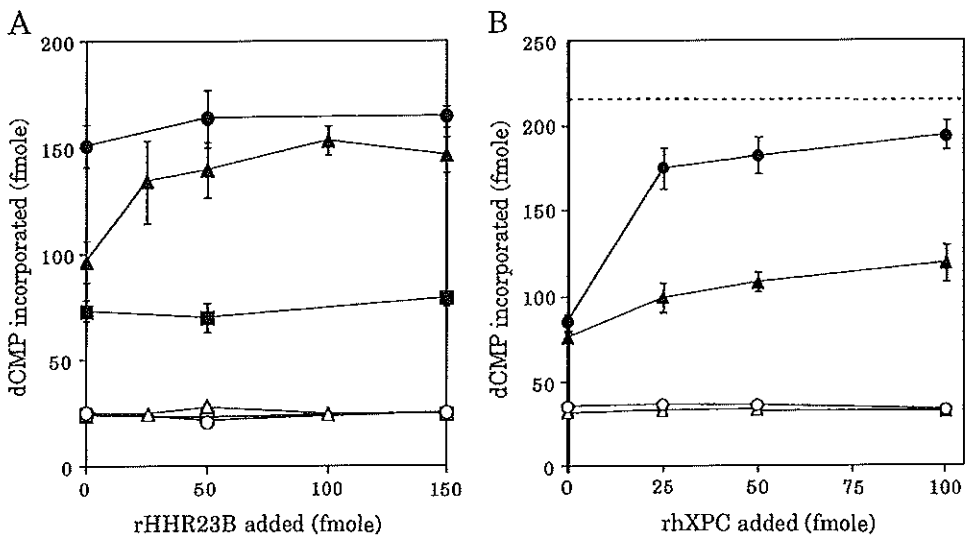


Fig. 9. **Requirement of HHR23B protein.**

(A) Various amounts of rHHR23B were added to the reconstituted cell-free NER reactions using CFII from XP4PASV cells in the presence or absence of 50 fmol of each indicated protein. (○ ●) rhXPC/rHHR23B complex, (▲ △) free rhXPC, (□ ■) none. Closed symbols represent incorporation into UV-irradiated viral DNA and open symbols represent that into unirradiated viral DNA. HeLa XPC/HHR23B complex (50 fmol) caused 260±15 fmol of dCMP incorporation. (B) Repair synthesis in the reactions containing various amounts of free rhXPC in the presence (○ ●) or absence (▲ △) of 100 fmol of rHHR23B was plotted. The dashed line shows the level of repair synthesis archieved by 100 fmol of rhXPC/rHHR23B complex. In both (A) and (B), averages and experimental errors were calculated from the results of three experiments.

DISCUSSION

In the present study, recombinant human XPC and HHR23B proteins were produced in insect cells using the baculovirus expression system. In *E.coli*, HHR23B protein was well expressed, but no detectable level of XPC protein expression was observed. This might be due to a number of AGA/AGG triplets encoding arginine in the *XPC* gene whose tRNAs are very rare in *E.coli*. Although high levels of XPC protein expression were obtained in insect cells, most of the rhXPC was not active in our *in vitro* NER system and also not sollubilized under the rather mild extraction conditions adopted in the present study. Sedimentation analyses using glycerol density gradients showed that the 'bulk' rhXPC was severely aggregated, while the 'soluble' rhXPC was sedimented as a single monomer peak around the expected molecular weight (data not shown). It is likely that the bulk rhXPC is incorrectly folded and/or modified post-translationally in insect cells.

We constructed a recombinant baculovirus expressing human XPC and HHR23B proteins simultaneously, based on the anticipation that the coexpressed proteins may form specific complexes in insect cells. Although a small portion of each protein extracted in the fraction 'S2' was purified in complex form, the results shown in Fig. 2 indicate that the majority of the coexpressed proteins was not complexed. This may be explained by the finding that most of the rhXPC expressed in insect cells was not synthesized in a functional form as mentioned above. Interstingly, Western blot analyses indicated that HHR23B protein exists in large excess over XPC protein even in human cells (see Figs. 3 and 6). Moreover, HHR23B protein is expressed normally in XP-C cell lines, suggesting that this protein may play additional role(s) in cells other than complex formation with the XPC protein. Further investigations are necessary to elucidate the putative multiple functions of HHR23B protein.

Using a cell-free NER system, we purified the XPC/HHR23B protein complex which complemented the repair defect in an XP-C cell extract. The recombinant XPC protein, free of HHR23B, could stimulate the repair synthesis by XP-C cell extracts (Fig. 5). These results support the idea that the disease XP group C is due to inactivation of the *XPC* gene and is not due to an indirect effect on the HHR23B protein. However, our results also indicate that HHR23B protein plays an important role in the mammalian NER reaction as well (Figs. 8 and 9). Disruption of the *S. cerevisiae RAD23*, *RAD7*, and *RAD16* genes were shown to cause only a moderate degree of UV-sensitivity (28). Recently, the *RAD7* and *RAD16* genes have been shown to be essential for NER on nontranscribed DNA strands as well as of the silent mating type loci, but not for NER on transcribed strands (42). This phenotype closely resembles that of XP-C cells. However, yeast *rad23* mutants may display a phenotype

different from *rad7* and *rad16* mutants, because RAD23 protein was recently shown to be complexed with RAD4 protein (11) which is the putative XPC counterpart but -in contrast to XPC- involved in both NER subpathways (42). The presence of free HHR23B and another RAD23 homologue, HHR23A also suggests multiple functions of this set of proteins.

In all mammalian *in vitro* NER reactions described thus far, XP-C cell extracts appear to be almost totally inactive. Since XP-C cells are still proficient in the transcription-coupled repair, this means that the *in vitro* repair reactions are driven, almost exclusively, by the mechanism of genome-overall repair. Therefore, it is difficult to determine whether HHR23B protein plays any role in transcription-coupled repair. Guzder *et al.* have recently described (10) that yeast RAD23 protein forms a higher-order complex with TFIIH and RAD14 protein (the yeast XPA counterpart). Although it is unclear whether a fraction of RAD23, even if not overexpressed, is also present as a free protein in yeast, this might be a function of free HHR23B. Alternatively, HHR23B might catalyze the assembly of repair complexes more efficiently, when complexed with XPC. In this context, it should be interesting to study the effect of RAD4 (XPC) protein on the complex formation among TFIIH, RAD14 (XPA) and RAD23 (HHR23B).

In the cell-free NER reactions using XP4PASV whole-cell extracts, the maximum level of repair stimulation archieved by the rhXPC/rHHR23B complex was 80 to 90% of that by the authentic XPC/HHR23B complex purified from HeLa cells (Fig. 5). On the other hand, in the reconstituted system repair synthesis stimulated by rhXPC/rHHR23B complex, as well as a combination of separately purified rhXPC and rHHR23B, reached only 30 to 40% of that by the HeLa protein complex (Figs. 8 and 9). One possible explanation is that the recombinant proteins may undergo some post-translational modification in insect cells different from those which occur in human cells. In fact, rHHR23B expressed in insect cells showed a hetergeneous mobility on SDS-PAGE, whereas human HHR23B was detected as a single band. To gain the maximum level of activity, the one or both recombinant proteins may need to be remodified by some enzymes present in whole-cell extracts but absent in CFII. Protein kinases and/or phosphatases may be considered as candidate 'activators' of the recombinant proteins. Future research using reconstituted systems such as that established recently by Aboussekhra *et al.* (2) should shed light on this possibility.

In our partial reconstitution system using UV-irradiated SV40 minichromosomes, the combination of CFII and RPA was capable of inducing a quite strong UV-dependent incorporation, with no detectable signal in the pUC19 internal control and the unirradiated SV40 minichromosomes (Fig. 7), indicating that PCNA dependency is not so strong in our cell-free system. In preliminary experiments, the weak dependency on PCNA was found to

be specific for the chromatin template. Clarification of this difference between a naked DNA substrate and the chromatin template might help elucidating the function(s) of PCNA in the *in vitro* and *in vivo* NER reaction.

There is another human homologue of the yeast RAD23 gene product, HHR23A. Since HHR23A and HHR23B proteins exhibit a high degree of overall homology (57% identity, 76% similarity), it will be interesting to determine whether these two proteins are functionally interchangable. This possibility and the characterization of HHR23A protein in human cells, is currently under investigation in our laboratory and will be subject of a forthcoming paper.

ACKNOWLEDGMENTS

We are grateful to Akio Kobayashi and Takahisa Hachiya (Medical & Biological Laboratories Co., Ltd.) for preparing anti-XPC and anti-HHR23B antibodies. We also thank Michael McKay for reading the manuscript.

This work was supported by grants from the Ministry of Education, Science and Culture of Japan, the Biodesign Research Program of the Institute of Physical and Chemical Research (RIKEN), and the Mitsubishi Foundation. The work at the Department of Cell Biology and Genetics of the Erasmus University was financially supported by a human frontier program research grant.

REFERENCES

1. **Aboussekhra, A. and R.D. Wood.** 1994. Repair of UV-damaged DNA by mammalian cells and *Saccharomyces cerevisiae*. Curr. Opinion in Genetics and Development. **4**: 212-220.
2. **Aboussekhra, A., M. Biggerstaff, M.K.K. Shivji, J.A. Vilpo, V. Moncollin, V.N. Podust, M. Protic, U. Hübscher, J.-M. Egly, and R.D. Wood.** 1995. Mammalian DNA Nucleotide excision repair reconstituted with purified protein components. Cell **80**: 859-868.
3. **Biggerstaff, M., D.E. Szymkowski, and R.D. Wood.** 1993. Co-correction of ERCC1, ERCC4 and xeroderma pigmentosum group F DNA repair defects *in vitro*. EMBO J. **12**: 3685-3692.
4. **Bohr, V.A., C.A. Smith, D.S. Okumoto, and P.C. Hanawalt.** 1985. DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. Cell **40**: 359-369.
5. **Bradford, M.M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. **72**: 248-254.
6. **Collins, A.R.** 1993. Mutant rodent cell lines sensitive to ultraviolet light, ionizing radiation and cross-linking agents: a comprehensive survey of genetic and biochemical characteristics. Mutat. Res. **293**:

99-118.

7. **Flejter, W.L., L.D. McDaniel, D. Johns, E.C. Friedberg, and R.A. Schultz.** 1992. Correction of xeroderma pigmentosum complementation group D mutant cell phenotypes by chromosome and gene transfer: Involvement of the human ERCC2 DNA repair gene. Proc. Natl. Acad. Sci. USA **89:** 261-265.
8. **Friedberg, E.C., G.C. Walker, and W. Siede.** 1995. DNA repair and Mutagenesis. ASM Press, Washington, D.C.
9. **Grossman, L. and S. Thiagalingam.** 1993. Nucleotide excision repair, a tracking mechanism in search of damage. J.Biol. Chem. **268:** 16871-16874.
10. **Guzder, S.N., V. Bailly, P. Sung, L. Prakash, and S. Prakash.** 1995. Yeast DNA repair protein RAD23 promotes complex formation between transcription factor TFIIH and DNA damage recognition factor RAD14. J.Biol. Chem. **270:** 8385-8388.
11. **Guzder, S.N., Y. Habraken, P. Sung, P., L. Prakash, and S. Prakash.** 1995. Reconstitution of yeast nucleotide excision repair with purified RAD proteins, replication protein A, and transcription factor TFIIH. J. Biol. Chem. **270:** 12973-12976.
12. **Hanawalt, P. and I. Mellon.** 1993. Stranded in an active gene. Curr. Biol. **3:** 67-69.
13. **Hoeijmakers, J.H.J.** 1993. Nucleotide excision repair I: from *E.coli* to yeast. Trends Genet. **9:** 173-177.
14. **Hoeijmakers, J.H.J.** 1994. Human nucleotide excision repair syndromes: molecular clues to unexpected intricacies. Eur. J. Cancer **30A:** 1912-1921.
15. **Kenny, M.K., U. Schlegel, H. Furneaux, and J. Hurwitz.** 1990. The role of human single-stranded DNA binding protein and its individual subunits in simian virus 40 DNA replication. J. Biol. Chem. **265:** 7693-7700.
16. **Kunkel, T.A., J.D. Roberts, and R.A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection, p.367-382. In R. Wu and L. Grossman (ed.), Methods in Enzymology, vol. 154. Academic Press, Inc., San Diego, CA.
17. **Laemmli, U.K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680-685.
18. **Lee, S.-H., Y. Ishimi, M.K. Kenny, P. Bullock, F.B. Dean, and J. Hurwitz.** 1988. An inhibitor of the *in vitro* elongation of simian virus 40 DNA replication is overcome by proliferating-cell nuclear antigen. Proc. Natl. Acad. Sci. USA **85:** 9469-9473.
19. **Legerski, R., and C. Peterson.** 1992. Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C. Nature **359:** 70-73.
20. **Li, L., E.S. Bales, C.A. Peterson, and R.J. Legerski.** 1993. Characterization of molecular defects in xeroderma pigmentosum group C. Nature Genet. **5:** 413-417.
21. **Liu, J.J. and A. Sancar.** 1992. (A)BC exinuclease: the *Escherichia coli* nucleotide excision repair enzyme. Mol. Microbiol. **6:** 2219-2224.
22. **Masutani, C., K. Sugasawa, H. Asahina, K. Tanaka, and F. Hanaoka.** 1993. Cell-free repair of UV-damaged simian virus 40 chromosomes in human cell extracts. II. Defective DNA repair synthesis by xeroderma pigmentosum cell extracts. J. Biol. Chem. **268:** 9105-9109.
23. **Masutani, C., K. Sugasawa, J. Yanagisawa, T. Sonoyama, M. Ui, T. Enemoto, K. Takio, K.**

Tanaka, P. van der Spek, D. Bootsma, J.H.J. Hoeijmakers, and F. Hanaoka. 1994. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. EMBO J. **13**: 1831-1843.

24. Mayne, L.V., and A.R. Lehmann. 1982. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. Cancer Research **42**: 1473-1478.
25. Mellon, I., G. Spivak, and P.C. Hanawalt. 1987. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. Cell **51**: 241-249.
26. O'Donavan, A., A.A. Davies, J.G. Moggs, S.C. West, and R.D. Wood. 1994. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. Nature **371**: 432-435.
27. O'Donavan, and R.D. Wood. 1993. Identical defects in DNA repair in xeroderma pigmentosum group G and rodent ERCC group 5. Nature **363**: 185-188.
28. Prakash, S., P. Sung, and L. Prakash. 1990. Structure and function of RAD3, RAD6 and other DNA repair genes of *Saccharomyces cerevisiae*. p.275-292. *In* P.R. Strauss and S.H. Wilson (e.d.), The eukaryotic nucleus, Telford Press, Inc., Caldwell, N.J.
29. Schaeffer, L., V. Moncollin, R. Roy, A. Staub, M. Mezzina, A. Sarasin, G. Weeda, J.H.J. Hoeijmakers, and J.-M. Egly. 1994. The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. EMBO J. **13**: 2388-2392.
30. Schaeffer, L., R. Roy, S. Humbert, V. Moncollin, W. Vermeulen, J.H.J. Hoeijmakers, P. Chambon, and J.-M. Egly. 1993. DNA repair helicase: a component of BTF2 (TFIIH) transcription factor. Science **260**: 58-63.
31. Shivji, M.K.K., M.K. Kenny, and R.D. Wood. 1992. Proliferating cell nuclear antigen is required for DNA excision repair. Cell **69**: 367-374.
32. Sibghat-Ullah, I. Husain, W. Carlton, and Sancar. 1989. Human nucleotide excision repair in vitro: repair of pyrimidine dimers, psoralen and cisplatin adducts by HeLa cell-free extract. Nucleic Acids Res. **17**: 4471-4484.
33. Sugasawa, K., C. Masutani, and F. Hanaoka. 1993. Cell-free repair of UV-damaged simian virus 40 chromosomes in human cell extracts. I. Development of a cell-free system detecting excision repair of UV-irradiated SV40 chromosomes. J. Biol. Chem. **268**: 9098-9104.
34. Summers, M.D. and G.E. Smith. 1987. A manual of methods for baculovirus vectors and insect cell culture procedures. Texas Agricultural Experiment Station Bulletin. vol. 1555.
35. Tanaka, K. N.Miura, I. Satokata, I. Miyamoto, M.C. Yoshida, S. Satoh, A. Kondo, A. Yasui, H. Okayama, and Y. Okada. 1990. Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. Nature **348**: 73-76.
36. Tanaka, K. and R.D. Wood. 1994. Xeroderma pigmentosum and nucleotide excision repair of DNA. Trends Biochem. Sci. **19**: 83-86.
37. Troelstra, C., A. van Gool, J. de Wit, W. Vermeulen, D. Bootsma, and J.H.J. Hoeijmakers. 1992. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell **71**: 939-953.
38. van Houten, B. 1990. Nucleotide excision repair in Escherichia coli. Microbiol. Rev. **54**: 18-51.
39. Van Vuuren, A.J., E. Appeldoorn, N.G.J. Jaspers, H. Odijk, A. Yasui, D. Bootsma, and J.H.J.

Hoeijmakers. 1993. Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. EMBO J. **12**: 3693-3701.

40. **Venema, J., A. Van Hoffen, A.T. Natarajan, A.A. Van Zeeland, and L.H.F. Mullenders.** 1990. The residual repair capacity of xeroderma pigmentosum complementation group C fibroblasts is highly specific for transcriptionally active DNA. Nucl. Acids Res. **18**: 443-448.

41. **Venema, J., L.H.F. Mullenders, A.T. Natarajan, A.A. van Zeeland, and L.V. Mayne.** 1990. The genetic defect in Cockayne's syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. Proc. Natl. Acad. Sci. USA **87**: 4707-4711.

42. **Verhage, R., A.-M. Zeeman, N. de Groot, F. Gleig, D.D. Bang, P. van de Putte, and J. Brouwer.** 1994. The *RAD7* and *RAD16* genes, which are essential for pyrimidine dimer removal from the silent mating type loci, are also required for repair of the nontranscribed strand of an active gene in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **14**: 6135-6142.

43. **Wang, Z., X. Wu, and E.C. Friedberg.** 1993. Nucleotide-excision repair of DNA in cell-free extracts of the yeast *Saccharomyces cerevisiae* Proc. Natl. Acad. Sci. USA **90**: 4907-4911.

44. **Watkins, J.F., P. Sung, L. Prakash, and S. Prakash.** 1993. The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological function. Mol. Cell. Biol. **13**: 7757-7765.

45. **Weeda, G., R.C.A. Van Ham, W. Vermeulen, D. Bootsma, A.J. Van der Eb, and J.H.J. Hoeijmakers.** 1990. A presumed DNA helicase encoded by *ERCC3* is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. Cell **62**: 777-791.

46. **Wood, R.D., P. Robins, and T. Lindahl.** 1988. Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. Cell **53**: 97-106.

CHAPTER VIII

XPC and human homologs of RAD23: intracellular localization and relationship with other nucleotide excision repair complexes

XPC AND HUMAN HOMOLOGS OF RAD23: INTRACELLULAR LOCALIZATION AND RELATIONSHIP TO OTHER NUCLEOTIDE EXCISION REPAIR COMPLEXES

Peter J. van der Spek[1], André Eker[1], Suzanne Rademakers,[1], Cécile Visser[1], Kaoru Sugasawa[1,2], Chikahide Masutani[3] Fumio Hanaoka[2,3], Dirk Bootsma[1], and Jan H.J. Hoeijmakers[1].

[1]Department of Cell Biology and Genetics, Medical Genetic Centre, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. [2] The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama 351-01, Japan. [3] Institute for Molecular and Cellular Biology Osaka University 1-3 Yamada-oka Suita, Osaka 565, Japan.

SUBMITTED, 1996

ABSTRACT

The syndrome xeroderma pigmentosum complementation group C (XP-C) is due to a defect in the global genome repair subpathway of nucleotide excision repair (NER). The XPC protein is complexed with HHR23B, one of the two human homologs of the yeast NER protein, RAD23 (Masutani *et al.*, EMBO J. 8: 1831-1843, 1994). Using heparin chromatography, gel fractionation and native gel electrophoresis we demonstrate that the majority of HHR23B is in a free, non-complexed form, and that a minor fraction is associated with XPC. In contrast, all detectable HHR23A seems to be free. Thus the HHR23 proteins may have an additional function independent of XPC. The fractionation behaviour suggests that the non-bound forms of the HHR23 proteins are not necessary for the core of the NER reaction. Although both HHR23 proteins share a high level of overall homology, they migrate very differently on native gels, pointing to a difference in conformation. Gel filtration suggests the XPC-HHR23B heterodimer resides in a high MW complex. However, immunodepletion studies starting from repair-competent Manley extracts fail to reveal a stable association of a significant fraction of the HHR23 proteins or the XPC-HHR23B complex with the basal transcription/repair factor TFIIH, or with the ERCC1 repair complex. Consistent with a function in repair or DNA/chromatin metabolism, immunofluorescence studies show all XPC, HHR23B and (the free) HHR23A to reside in the nucleus.

INTRODUCTION

A complex network of DNA repair mechanisms protects the genetic information from continuous genotoxic pressure caused by the DNA-damaging effect of exogenous and genotoxic agents. Such damage can lead to inborn defects, cell death or neoplasia. Nucleotide excision repair (NER) is one of the most important DNA damage repair pathways, since this process recognizes a wide variety of lesions. Impaired NER activity has been extensively investigated in cells from three human disorders: xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD) (Cleaver and Kraemer, 1994, Itin *et al.*, 1990). These genetic diseases are characterized by sun (UV) hypersensitivity, genetic instability and a marked clinical and genetic heterogeneity. Many genes involved in XP, CS and TTD complementation groups have been cloned

using cells from human patients or from NER-deficient, UV-sensitive chinese hamster ovary mutants (Bootsma *et al.*, 1995). Extensive sequence homology between the mammalian and yeast NER proteins has become apparent, indicating that the NER pathway is strongly conserved in eukaryotic evolution (Hoeijmakers, 1993).

At least two NER sub-pathways can be discerned: transcription-coupled repair (TCR) and global genome repair (GGR) (Hanawalt, 1994; Bohr, 1991). In contrast to all other xeroderma pigmentosum groups, only group C patients are defective in GGR (Venema *et al.*, 1991). Previously, we reported the identification and cloning of two human homologs of the yeast NER gene *RAD23*: *HHR23A* and *HHR23B* (Masutani *et al.*, 1994). The yeast *S. cerevisiae rad23* null mutants display an intermediate UV-sensitive phenotype (McCready, 1994), suggesting that the affected protein is not indispensable for NER. The HHR23B gene product, which forms a tight complex with the XPC protein, has a high affinity for ssDNA and both proteins were found to be indispensible for *in vitro* NER (Sugasawa *et al.*, submitted for publication 1995). Like the human XPC/HHR23B complex, the *S. cerevisiae* RAD4 and RAD23 protein homologs were also determined to be complexed with each other (Guzder *et al.*, 1995). XP-C cells harbour a specific defect in the repair of non-transcribed sequences of the genome, including the non-transcribed strand of active genes, whereas the NER subpathway that accomplishes the preferential repair of the transcribed strand of active genes (TCR) is still operational (Venema *et al.*, 1990, 1991). This implies a selective role for the XPC complex in the global genome NER system (Hanawalt, and Mellon, 1993). Additionally, XP-C cells are claimed to be defective in the repair of rDNA (Christians and Hanawalt, 1994).

RAD23 is a ubiquitin-like fusion protein displaying significant homology to ubiquitin at the N-terminus (Watkins *et al.*, 1993; Masutani et al., 1994). A second link exists between RAD23 and the ubiquitin pathway: a twice repeated element, homologous to a C-terminal extension of a Class II ubiquitin-conjugating enzyme (E2) has been identified (van der Spek *et al.*, 1996). The ubiquitin-conjugating pathway is involved in proteolytic degradation of proteins, and additional involvement in cellular processes such as DNA repair, chromosome condensation and decondensation, and cell cycle control has been reported (Ciechnover, 1994; Jentsch and Schlenker, 1995). The link to these other DNA-metabolizing processes presumably comes from ubiquitin-mediated proteolytic degradation of key proteins involved in these events.

Here we present data on the purification and stable association of HHR23A, HHR23B, and XPC proteins with other known NER factors. Additionally, the sub-cellular localization of HHR23A, HHR23B and XPC was determined by immunofluorescence.

MATERIALS AND METHODS

General procedures

Purification of nucleic acids, restriction enzyme analysis, gel electrophoresis of nucleic acids and proteins, transformation of *E.coli*, etc. were performed according to standard procedures (Sambrook *et al.*, 1989). RNA samples were separated on 1% agarose gels and transferred to Zeta probe membrane (Bio-Rad) as described (Reed and Mann, 1985). Labelling of DNA probes was carried out using the random priming protocol (Feinberg and Vogelstein, 1983). Immunoblotting was performed as described elsewhere (Harlow and Lane, 1988). HHR23A and HHR23B proteins were translated *in vitro* using a rabbit reticulocyte lysate system as recommended by the manufacturer (Promega) using 50 μCi of [^{35}S]methionine (1 mCi/mmol). After polyacrylamide gel electrophoresis (PAGE) and native gel electrophoresis, both labelled proteins were blotted and visualized by autoradiography.

For non-denaturing gel electrophoresis, a 4-15% gradient polyacrylamide gel in TBE buffer and 12% glycerol was prepared (TBE: 90 mM Tris, 80 mM boric acid and 2.5 mM EDTA). The gel was pre-run for 30 min at 70 V, loaded with samples and run for 2 h at 70 V followed by 16-20 h at 150V. Proteins included as molecular mass standards used for estimation of the native molecular weight of HHR23A and HHR23B were ferritin (440 kDa), catalase (240 kDa), lactate dehydrogenase (140 kDa), and albumin (67 kDa) (Boehringer Mannheim). The western blot was stained with Ponceau-S to visualize the molecular weight markers for determination of the apparent MW of the HHR23 proteins. HHR23A and HHR23B proteins were detected by polyclonal antibodies and autoradiography.

Production of recombinant proteins and antibodies.

The full-length *HHR23A* and *HHR23B* cDNAs were cloned into the pET11D vector (Novagen), transferred into *E.coli* strain BL21(DE3), and gene expression was induced over 4 hours by IPTG. Cells were homogenized in PBS, and after sonication, cleared by centrifugation. Approximately 20 g cells were disintegrated by sonication, followed by centrifugation to remove cell debris. Recombinant HHR23A and HHR23B proteins were purified by chromatography on a Q Sepharose Fast Flow column (1 x 12 cm, flow rate 18 ml/h). For HHR23A, the column was eluted with a linear gradient 0 -> 0.5 M NaCl in 0.1 M NaCl, 10 mM K-phosphate pH 7, while for HHR23B a gradient 0 -> 0.4M NaCl in 10mM NaCl, 10 mM K-phosphate pH 7 was used. HHR23A protein eluted at 0.3 M, HHR23B protein at 0.11 M NaCl. For antibody production these proteins were subjected to SDS-PAGE. Bands were cut from Coomassie stained gels, electroeluted and concentrated with Centricon 30 concentrators (Amicon). The identity of the eluted proteins was verified by amino acid sequencing. For large scale purification, Q Sepharose fractions were pooled, brought on to 20% ammoniumsulphate and loaded on a Butyl Sepharose Fast Flow column (1 x 12 cm, flow rate 18 ml/h). Columns were eluted with a linear gradient 20 -> 0% ammoniumsulphate in 0.1 M NaCl, 10 mM K-phosphate pH 7. Fractions containing HHR23A or HHR23B protein (eluted at 4.5% ammoniumsulphate) were dialysed and kept frozen after addition of 1/5 volume of glycerol. Polyclonal antibodies were raised in rabbits against the *E.coli*-overproduced human HHR23 proteins, as described (Harlow and Lane, 1988). Affinity-purified antibodies were derived from precise elution of antibodies specifically bound to recombinant antigen immobilized on nitrocellulose after transfer from SDS-polyacrylamide gels (Smith and Fisher, 1984).

The synthetic peptide (KTKREKKAAASHLFPFEKL), corresponding the C-terminus of XPC, was used to produce a polyclonal antibody in rabbits. Prior to injection, the peptide was cross-linked to KLH carrier protein. Affinity-purified antibodies were derived by eluting XPC-peptide from a column to which it had been coupled. As a second antibody, alkaline phosphatase-conjugated goat anti-rabbit was used, the latter visualised by 5-bromo-4-chloro-3-indolyl phosphate. Immunoblots were incubated with monoclonal antibodies (Mab3C9) against the p62 subunit of TFIIH (generously provided by Dr. J.-M. Egly, Strasbourg), as published earlier by Fischer *et al.* (1992). A polyclonal antiserum raised against p89, a GST-ERCC3 fusion protein containing an internal part (amino acids 82-480) was used to detect the p89/ERCC3/XPB component of TFIIH.

Cell lines and extracts

HeLa cells and Chinese hamster ovary (CHO9) cells were grown in F10/DMEM medium (1:1) supplemented with 10% fetal calf serum, penicillin 100 U/ml and streptomycin 0.1 mg/ml. Cells were harvested and extracts were prepared from 2-5 ml of packed cell pellets by the method of Manley, as modified by Wood (Manley *et al.*, 1983; Wood *et al.*, 1988) dialysed in buffer A and stored at -80°C until use. XPA and XPC patient cell lines used in these experiments were XP7CA (CW12, XPA) (Wood *et al.*, 1987) and XP4PA (XPC) (Halley *et al.*, 1979; Li *et al.*, 1993). COS-1 SV40-transformed African green monkey kidney fibroblasts were seeded semi-confluent in 6-well plates, and grown on F10/DMEM medium (1:1) supplemented with 5% fetal calf serum, penicillin and streptomycin. For DEAE-dextran/chloroquine transfection, SV40 promoter-driven constructs were used (pSVL derived pSLM vector; Pharmacia biotech) containing full length *HHR23A* and *HHR23B* cDNAs. The empty pSLM vector was used as a negative transfection control in parallel with *HHR23A* and *HHR23B* genes. A 10% DMSO shock for 1.5 minutes was given 4 hours after transfection. Transient expression of the corresponding HHR23A and HHR23B protein was analysed by immunofluoresence 48 hours after transfection.

Antibody depletion of NER-proficient extracts.

Protein A-Sepharose CL-4B beads (Pharmacia-Biotech) (70 μg) were washed three times with PBS, then incubated with 10 μl anti-XPC antibodies or pre-immune serum for 15 min at 4°C. The beads were then washed three times in buffer A (25 mM Hepes-KOH [pH 7.8], 0.1 M KCl, 12 mM $MgCl_2$ 1 mM EDTA, 2 mM DTT and 17% glycerol) and added to a repair-competent HeLa extract for 30 min at 4°C. The supernatant obtained after spinning down the beads was used as a depleted HeLa extract and tested on immunoblots for co-depletion. After boiling the protein A-Sepharose beads, the depleted "bound" fraction was analysed by immunoblot analysis.

Fractionation of whole-cell extracts.

A HeLa cell free extract (14 μg/μl; 750 μl) was applied to a heparin-sepharose column. Proteins were eluted with a linear 30 ml gradient (15 ml PBS containing 10 mM 2-mercaptoethanol, 15 ml 1M NaCl containing 10 mM 2-mercaptoethanol) from 0.15 to 1.15 M NaCl in PBS containing 10 mM 2-mercaptoethanol buffer.

Whole-cell extracts were prepared for fractionation on a phosphocellulose column. This 5 ml column equilibrated with buffer A [25 mM *N*-(2-hydroxyethyl)piperazine-*N'*-2-ethanesulfonic acid (HEPES)-

KOH (pH 7.9), 1 mM EDTA, 10% glycerol, 0.01% Triton X-100, 1 mM DTT and 0.25 mM PMSF] contained 0.2 M KCl. The column was washed with the same buffer and the adsorbed proteins were eluted with buffer A containing 1 M KCl. The peak fractions from the flow-through (CFI) and the eluate (CFII) were concentrated by dialysis against buffer A containing 0.1 M KCl and 20% sucrose, and stored at -80°C.

Size-fractionation of HeLa Manley extracts was performed on a Sephacryl S300-HR column. HeLa nuclear extracts were loaded on a 1 x 46.4 cm column with a flow rate of 0.92 ml/7.5 min and eluted with PBS (7.4 ml/hr). The resulting fractions were concentrated. Protein profiles of HHR23A, HHR23B and XPC were visualized on immunoblots using alkaline phosphatase-labeled secondary antibodies. Proteins included as molecular mass standards used for estimation of the native molecular weight were thyroglobulin (669), ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa), and albumin (67 kDa) (Boehringer Mannheim).

Immunofluoresence

HeLa, XP7CA, and XP4PA cells were grown on slides in F10/DMEM medium supplemented with 10% fetal calf serum, penicillin and streptomycin, and washed prior to fixation in PBS. Cells were fixed for 10 min in 2% paraformaldehyde-PBS, followed by incubation with methanol at room temperature for 20 minutes. After extensive washing (3x5 min.) with PBS supplemented with 0.15% glycine and 0.5% BSA (PBS^+) the slides were incubated with affinity-purified primary antibodies (1:100 dilution in PBS) for 1½ hour in a moist incubation chamber at room temperature.

Immunocytochemical controls were routinely included (omission of the primary antibody incubation step and incubation with pre-immune serum). Background was negligible.

Slides were washed in PBS^+ and incubated with goat anti-rabbit-FITC-conjugated antiserum (1:80 dilution) for 1½ hour. Slides were washed in PBS and preserved with vectashield™ mounting medium (Brunschwig). The DNA was stained with 4'-6 diamino-2-phenylindole (DAPI) whereas the fluorescein-labeled second antibody visualized the antigen of interest. Fluorescence microscopy was performed with an Aristoplan laser beam microscope. Image modification for figures was performed by using the Adobe Photoshop program on an IBM Compaq deskpro XE 560.

RESULTS

Characterization of HHR23A, HHR23B and XPC proteins

To characterize XPC, affinity-purified anti XPC polyclonal antibodies were generated and tested by immunoblotting, using *in vitro* translated XPC protein, XPC protein purified from HeLa cells, and XPC in total cell extracts. Figure 1 shows their specificity on HeLa, XP-A and XP-C protein extracts. A clear band of the expected molecular weight of 125 kD as determined by *in vitro* translation and purified XPC (Masutani *et al.*, 1994) was observed in total cell extracts of wildtype HeLa and XP-A cells. The XP-C extract from patient XP4PA is useful for testing the specificity of the antibody. Due to a homozygous frameshift mutation that is predicted to result in a premature termination of the protein (Li *et al.*, 1993), this patient lacks the C-terminal XPC region encoding the part used to raise the antibodies. Further evidence for the specificity of the anti-XPC antibodies is derived from the immunofluoresence data depicted in Figure 8.

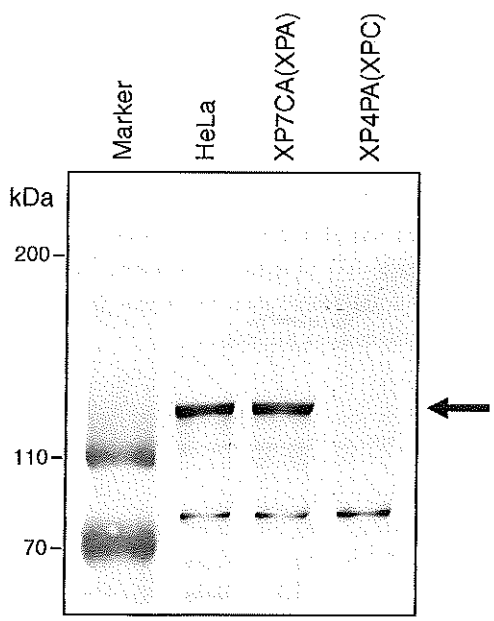


Figure 1 **Specificity of XPC antiserum by immunoblot analysis in total cell extracts.** Immunoblot analysis of wildtype cell extract (WCE) of HeLa, XPA (XP7CA) and XPC (XP4PA) cells incubated with the affinity-purified anti-XPC antibodies. 10 μg protein was loaded in each lane and separated on a 6% SDS-PAGE gel. The cross-reacting band present in all three lanes indicates equal loading.

Overproduction and purification of the two HHR23 proteins is described in the materials and methods section. Both recombinant HHR23 polypeptides behaved similarly during purification. *E. coli* overproduced proteins were present in the soluble fraction of total sonicated extracts (Fig. 2, first lanes, upper panels). Due to their low iso-electric point, Q-sepharose binds both HHR23 products efficiently. The presence of extensive hydrophobic stretches in the primary amino acid sequence suggested that a purification step based on hydrophobic interaction might be successful. Figure 2 indicates that butylsepharose yielded a very powerful purification (a full description of the entire HHR23 purification will be described elsewhere). The specificity of affinity-purified polyclonal rabbit antibodies raised against the HHR23A and HHR23B proteins was checked on recombinant *E.coli*-overproduced human protein, HeLa wildtype total cell extracts and *in vitro* translated proteins.

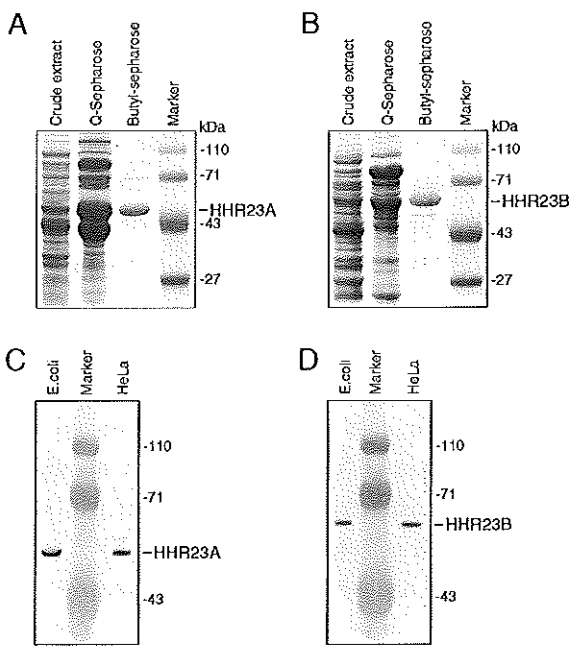


Figure 2 **Protein purification and antibody characterization of HHR23A and HHR23B.**
The top panels show respectively the purified CBB-stained 11% acrylamide gel containing crude, Q-sepharose and Butyl-sepharose purified fractions of the HHR23A and HHR23B proteins. The lower panels show affinity-purified HHR23A and HHR23B antibodies on 3 ng purified recombinant protein (butyl-sepharose) fraction and 10 μg HeLa WCE run at 8% acrylamide gels.

In immunoblotting experiments, anti-HHR23A antibodies visualized a 50 kDa band in repair-competent Manley extracts of HeLa cells (Fig. 2C, third lane). A similarly-sized protein was found for the *E.coli*-overproduced recombinant HHR23A protein (Fig. 2A and 2C, first lane) and the *in vitro* translated protein (not shown). The recombinant HHR23B protein migrated at 58 kDa, similar to the *in vitro* translated protein, the protein of a HeLa Manley extract (Fig. 2B and 2D, first and last lanes), and the protein from the purified XPC/HHR23B complex-containing fraction (Masutani *et al.*, 1994). Although both HHR23 gene products share 76% sequence similarity, polyclonal antibodies raised for each of the individual HHR23 proteins did not show any cross-reacting activity.

Besides human NER-deficient patients, a second class of excision repair-deficient mutants is represented by laboratory-induced, UV-sensitive, rodent cell lines. Eleven complementation groups have been identified (Riboni *et al.*, (1992) and Collins (1993)) which partially overlap with the genes defective in several of the human NER syndromes (Hoeijmakers, 1993). It is not known whether XPC or either of the HHR23 proteins is represented among these rodent mutants. Furthermore, defects in one product may lead to decreased stability of other proteins in the same complex. Previously, it was found that the amount of ERCC1 protein is significantly reduced in XPF cells (Biggerstaff *et al.*, 1993; van Vuuren *et al.*, 1994), although no primary defect in this gene is present in XPF. Since XPF and ERCC1 are known to form a protein complex (Biggerstaff *et al.*, 1993; van Vuuren *et al.*, 1993), a mutation in the XPF protein most likely results in breakdown of ERCC1 protein due to instability (van Vuuren *et al.*, 1993). Therefore, the above characterized antibodies were used to check Manley extracts of cells from all known repair-deficient mammalian complementation groups for abnormalities of the XPC, HHR23A and HHR23B proteins, that could point to a possible involvement in any of these mutants. With the exception of XPC in XP group C, no alterations were observed for any of the proteins in the extracts from NER-deficient individuals. As an example, Figure 3 shows the human NER complementation groups. XP complementation group E, not included in this panel, also showed a normal expression pattern. In addition to this, no abnormalities were observed in any of the CHO groups analysed (data not shown). This is consistent with the idea that none of the HHR23 proteins is implicated in the available NER mutants and that XPC is not affected in the rodent mutants. Moreover, the HHR23B protein is not destabilized as a consequence of an XPC defect.

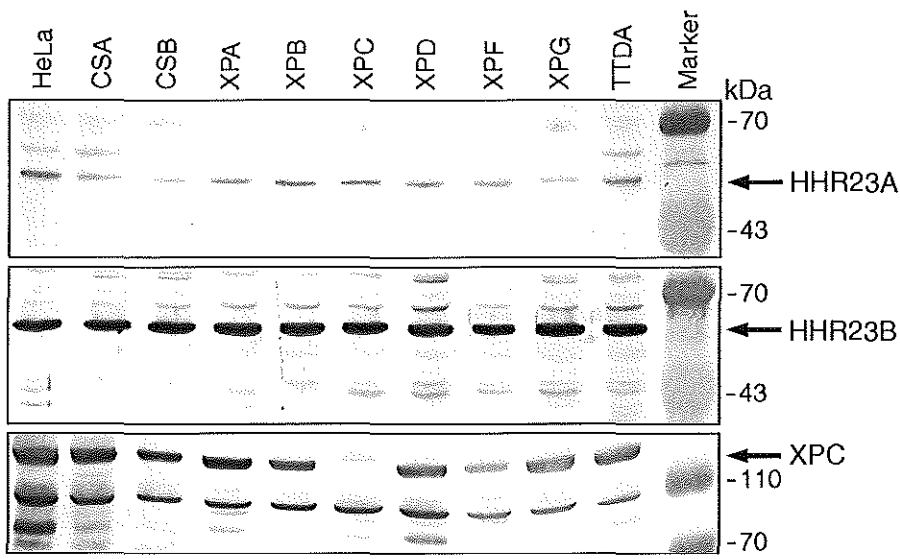


Figure 3 **Immunoblot analysis of human NER syndromes.**
Immunoblotting of cells representative of different NER syndrome complementation groups analysed with anti-HHR23A, anti-HHR23B, and anti-XPC antibodies. The crossreacting bands provide an internal control for protein loading in the different lanes. Proteins were run on 11% polyacrylamide gels and blotted as described in Materials and methods.

Behaviour of HHR23A, HHR23B and XPC in fractionation procedures

Previously, we determined XPC to be complexed with HHR23B protein (Masutani *et al.*, 1994). An association of XPC with TFIIH has been claimed (Drapkin *et al.*, 1994). In an attempt to identify stable associations with other repair components or factors involved in the basal transcription machinery, systematically-purified protein fractions were tested for the presence or absence of HHR23A, HHR23B and XPC proteins. Purification protocols were used which are known to leave large protein complexes such as TFIIH intact (Gerard *et al.*, 1991).

To separate the components of the different general transcription factors, a HeLa cell-free extract competent for *in vitro* repair and transcription was fractionated by heparin ultrogel column chromatography. Figure 4A shows the load and the elution fractions analysed with the affinity-purified HHR23A, HHR23B, XPC and TFIIH antibodies.

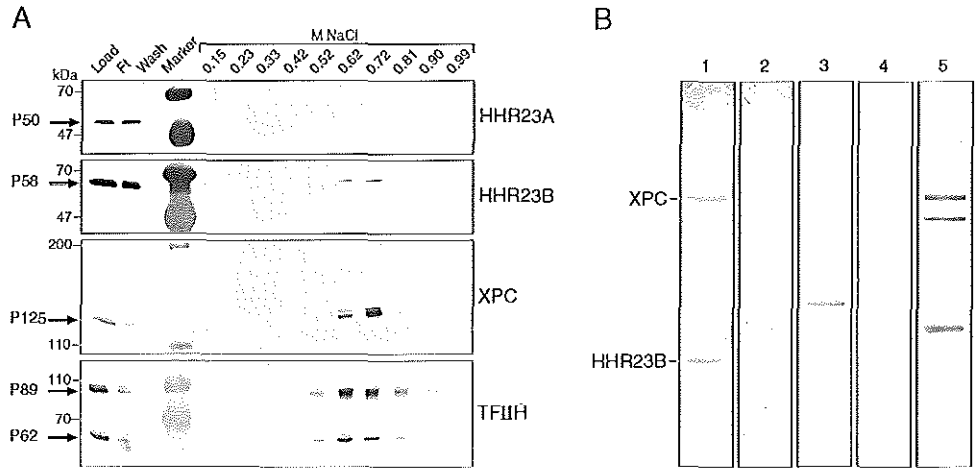


Figure 4 **Heparin fractionation and immunodepletion**

(A) Heparin-ultrogel fractions were assayed by immunoblotting with anti-HHR23A(p50), anti-HHR23B(p58), anti-XPC(p125), and anti-TFIIH (p89 and p62) antibodies. Affinity-purified polyclonal antibodies were visualized by alkaline phosphatase-labeled conjugates. The different bands of XPC probably represent different phosphorylation status. The minor amounts of XPC and TFIIH in the flow-through fraction are able to bind to the heparin upon a second application to the same column and thus do not represent a different species of these complexes. (B) HeLa extract (lane 1), HeLa extract depleted for XPC (lane 2), were both analyzed with anti-XPC and anti-HHR23B antiserum. HeLa extract depleted for XPC was analyzed with anti-p62 monoclonal antibody (lane 3). The protein fraction bound to the XPC-immunobeads was analysed by immunoblotting with anti-p62 antiserum (lane 4) and anti-XPC (lane 5), some of the extra bands are derived from the antibodies released from the beads.

All detectable HHR23A protein was found to reside in the flowthrough fraction. The vast majority of the HHR23B protein also resided in the flowthrough, whereas approximately 10-20% was found in the XPC-containing fractions (0.62-0.72 M NaCl), consistent with the existence of an XPC/HHR23B complex. The elution profile of TFIIH (represented by the p89 and p62 proteins) partly overlaps with that of XPC, although the TFIIH elution profile is slightly broader than that of XPC and the elution peaks are different. These findings support the existence of two subfractions of HHR23B and indicate that HHR23A is not detectably associated with XPC nor with TFIIH, but do not permit any conclusion concerning the existence of stable complex formation between XPC and TFIIH. Therefore, we performed immunodepletion experiments and other types of protein

fractionation. A wildtype repair-competent extract depleted for XPC still contained p62 protein besides free noncomplexed HHR23B protein, as determined by immunoblot analysis (Fig 4B). Moreover, the XPC-containing protein fraction bound to the protA beads showed no detectable co-depletion of the p62 subunit of the TFIIH complex. These data together with other fractionation studies including phosphocellulose (data not shown) and HA-His-tagged TFIIH complex (B. Winkler unpublished results), strongly suggest that the vast majority of XPC-HHR23B complex is not stably associated with TFIIH in repair-active Manley extracts. Furthermore, highly purified (Sulphopropyl 5-PW) protein fractions containing ERCC1-correcting activity as described by van Vuuren *et al.* (1995), showed no detectable amount of XPC or any of the HHR23 proteins, therefore excluding a stable association of significant quantities of XPC with the ERCC1 protein complex in the extracts used.

Non-denaturing gel electrophoresis

To detect possible complexes of HHR23 with other proteins separation of repair-proficient HeLa Manley extracts under non-denaturing gel electrophoresis conditions was performed using wild-type cell extracts and anti-HHR23A and anti-HHR23B antibodies. The results are shown in Figure 5. The HHR23A protein was detected as a single band migrating at about 70 kDa. For HHR23B, two forms with approximate sizes of 140 kDa were distinguished in HeLa cell extracts. To determine the specificity of the apparent molecular weights of both HHR23 proteins, *in vitro* translated protein and recombinant *E.coli* overproduced HHR23 proteins were run in parallel. After immunoblotting and autoradiography, all HHR23A protein samples were found to migrate at the same size, suggesting that the recombinant polypeptide has a similar conformation as the corresponding HeLa and *in vitro* translated proteins (Figure 5). The lower HHR23B band observed in the HeLa lane migrates at the same position as the recombinant protein and the *in vitro* translated form. Thus, the upper band may represent a modified form of the HHR23B protein. The notion that both bands are derived from HHR23B is supported by the observation that both signals disappeared when the antibodies were competed with excess recombinant HHR23B protein (data not shown).

Given the high sequence homology of both HHR23A and HHR23B proteins (57% identity, 76% similarity) a large apparent size difference was observed under native conditions (Figure 5). From these data, it was concluded that HHR23A does not form a protein complex with HHR23B. Furthermore, these data confirm that the majority of both HHR23 proteins is present in free non-complexed form in wildtype extracts, and that the HHR23A and part of the HHR23B proteins do not undergo a gross post-translational

modification. Analysis with anti-XPC antibodies revealed no detectable signal for XPC. Moreover, the XPC-HHR23B complex could also not be detected with the anti-HHR23B antibodies, suggesting a size or charge problem for migrating into the native gel.

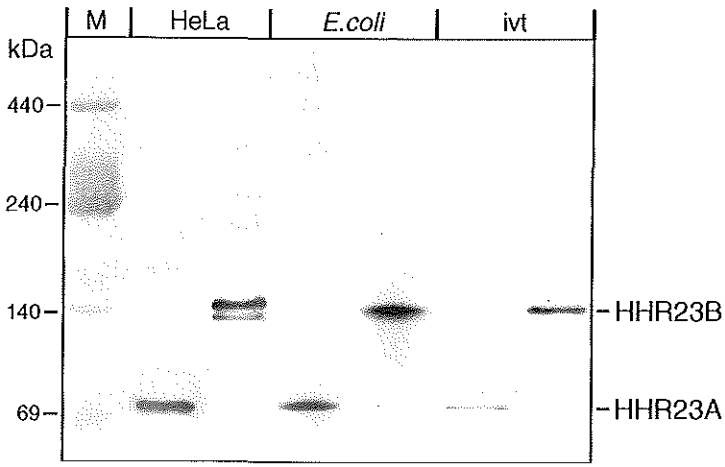


Figure 5 **Non-denaturing gel electrophoresis of HHR23A, and HHR23B proteins.** Migration pattern of HHR23A and HHR23B in HeLa Manley extracts as detected by immunoblotting, in vitro translated (ivt) HHR23 proteins and purified recombinant HHR23A and HHR23B *E.coli* proteins. Proteins were separated on a 4-15% polyacrylamide gradient gel.

To investigate the native size of XPC and the HHR23 proteins and possible association with other factors in a different manner, size-fractionation was performed using gel filtration. HeLa whole cell extracts, in which repair and transcription factors reside in an active configuration, were used as a starting point. The profiles of a S300 sephacryl fractionation are shown in Figure 6A. Both XPC and part of the HHR23B proteins resided in similar overlapping fractions. However, the size of the XPC-HHR23B complex in this configuration is much higher than the apparent size of the purified complex as determined by glycerol gradient and Sephacryl fractionation (Masutani *et al.*, 1994). The large size provides a possible explanation for the failure to detect this complex under native gel electrophoresis conditions (see above). The size determined for the free form of HHR23B is slightly bigger than that for HHR23A, which is in agreement with the data of the native gel electrophoresis (as shown in Figure 5).

The above findings indicate that the vast majority of the HHR23A and B molecules is in a free form. To determine how many A and B molecules reside in a cell,

immunoblot titration experiments were performed using HeLa total cell extracts and standard amounts of recombinant HHR23A and HHR23B (Fig. 6B). From the data we calculate that approximately 2-4 x 10^5 molecules of each are present in a single HeLa cell. From the blots (e.g. see Figs 2, 5) we estimate that maximally one fifth of HHR23B is complexed to XPC. In the absence of free XPC and on the assumption of a 1:1 stoichiometry in the complex, the XPC concentration is of the order of 4-8 x 10^4 copies per cell.

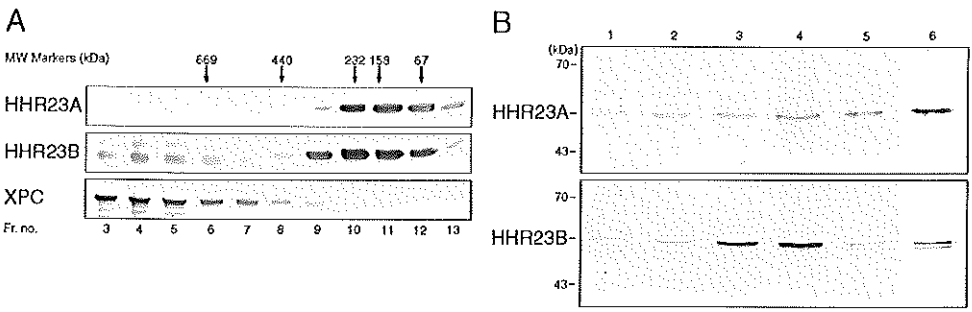


Figure 6 **Size fractionation and titration experiments**

(A) Size fractionation of HHR23 and XPC proteins was performed on a Sephacryl S300-HR column.(A) Immunoblot analysis of fractions for the presence of HHR23A, HHR23B and XPC respectively, using affinity-purified antibodies. The arrowheads indicate the size (kDa) of the marker proteins used as reference molecules. (Similar results were obtained using S500 Sephacryl fractionation). (B) Titration experiments to estimate the amount of HHR23 proteins per cell. 6 μg and 12 μg HeLa total cell extract (lanes 1, 2), 6 μg and 12 μg HeLa Manley extract, and 3 ng and 10 ng recombinant protein (lanes 5, 6) were analyzed by immunoblotting using anti-HHR23A and anti-HHR23B antibodies. Intensities were used to calculate the amounts of molecules per cell.

Intracellular localization of HHR23A, HHR23B and XPC.

Analysis of HHR23A, HHR23B and XPC protein sequences for the presence of a DNA binding domain or a nuclear localization signal (Dingwall and Laskey, 1991) revealed no clear matches conforming with the known consensus sequences. Moreover, we failed to detect DNA binding activity of the isolated recombinant HHR23A and HHR23B proteins (unpublished observations). To define the subcellular distribution of the

free HHR23A, HHR23B protein molecules and the XPC/HHR23B complex, we performed indirect immunofluoresence in HeLa cells, COS-1 transfected cells and fibroblasts derived from wildtype and XPC patients. No labeling was seen after treatment with secondary antibodies alone or after competition with excess of the recombinant HHR23 proteins or the XPC peptide used for immunization. The specificity of the primary antibody was confirmed by the use of pre-immune sera in all experiments, included as a negative control. For both human equivalents, HHR23A and HHR23B, a clear nuclear localization was observed, and the protein appeared to be absent from the nucleoli (Figure 7A and 7C).

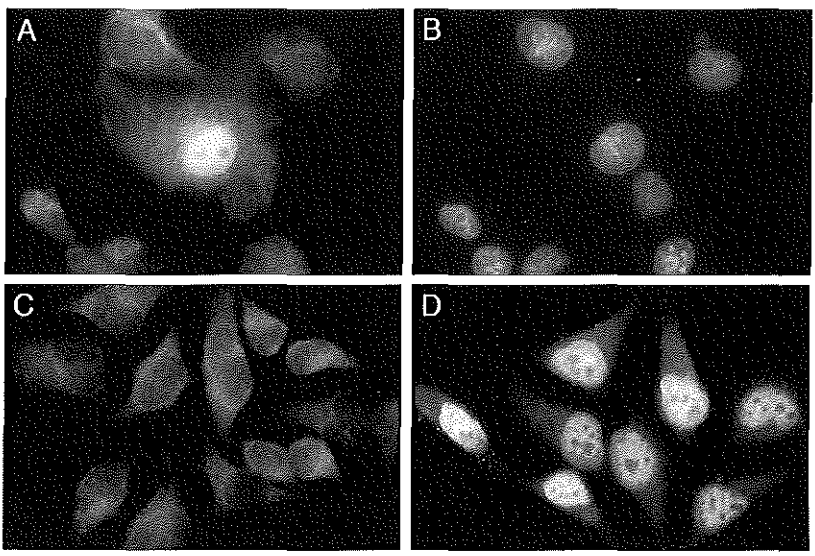


Figure 7 **Nuclear subcellular localization of the HHR23 proteins.**
Panel (A) the affinity-purified HHR23A antibody staining on COS-1 transfected cells, visualized by FITC-labelled secondary antibodies. Panel (B) the DAPI staining of the same cells. The bottom left panel (C) shows HeLa cells with the pre-immune serum of HHR23B, whereas the bottom right panel (D) shows the endogenous HHR23B protein in HeLa cells detected with affinity-purified antibodies.

The subcellular localization of the XPC protein was exclusively present in the nucleus of HeLa and xeroderma pigmentosum group A (XP7CA) fibroblasts during interphase, as shown in Figure 8. The XP4PA (XPC) cells gave no signal, confirming the absence of the intact XPC protein (as also shown by immunoblot analysis in Fig. 1 and Fig. 2).

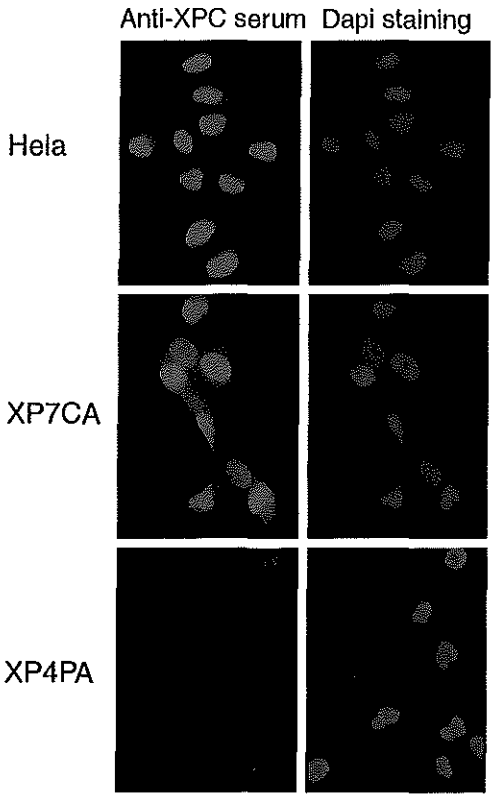


Figure 8

Figure 8
Immunofluorescent localization of XPC in HeLa cells, xeroderma group A (XP7CA) and group C (XP4PA) fibroblasts. The left panel visualizes the FITC-conjugated anti-XPC antibodies whereas the right panel depicts DAPI staining of the same cells.

Additionally, we investigated the subcellular localization of the XPC and HHR23 protein during mitosis. This was carried out in human XP7CA, XP4PA fibroblasts and in HeLa cells. When the cell entered mitosis, XPC (and both HHR23) proteins relocalized in the cytoplasm. During metaphase, they were present diffusely throughout the whole cell, without particular association with chromatin. Unexpectedly, in the later stages of mitosis, during anaphase and telophase, the XPC protein became specifically associated with chromatin (Figure 8 panels C and D). This anaphase/telophase specific association was not observed for either of the HHR23 proteins. However, due to excess of the free HHR23B protein in the cell (see above), we cannot exclude whether this also holds for the small amount of HHR23B complexed with XPC. Cells from the XP7CA (XPA) patient showed the same subcellular localization as HeLa. The HHR23 protein distribution in XP4PA (XPC) cells was indistinguishable from that in HeLa or XP7CA (XPA).

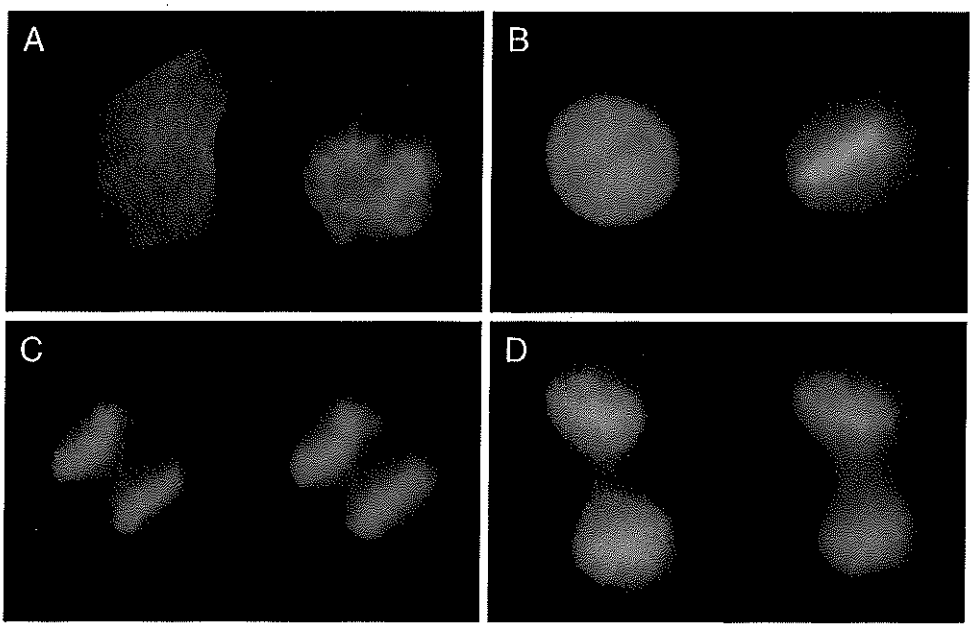


Figure 9 Different stages of mitosis analysed with affinity-purified anti-XPC antibodies. The panels represent (A) prophase, (B) metaphase, (C) late anaphase and (D) telophase. The green signals on the left represent XPC, visualized by FITC-conjugated secondary antibodies, whereas the blue signals on the right depict the DAPI staining of the same cells.

DISCUSSION

This article describes the partial characterization of HHR23A, HHR23B and XPC proteins, known to be specifically involved in global genome nucleotide excision repair. GGR, defective in XPC cells, deals with the repair of bulk DNA, including the non-transcribed strand of active genes (Kantor *et al.*, 1990; Venema *et al.*, 1990, 1991) and is important for preventing carcinogenesis. Evidence for this comes from the lack of enhanced cancer risk in patients with the transcription-coupled NER disorder, Cockayne syndrome (Lehmann, 1987) and the high cancer predisposition when the GGR subpathway is defective as in XP-A and XP-C. Purification of the XPC-correcting NER activity revealed a heterodimeric protein complex consisting of XPC and HHR23B (Masutani *et al.*, 1994). However, the functional significance of the association of HHR23B with XPC is not known, and could be, for instance, stabilization of the XPC protein. The yeast RAD23 protein also has a role in NER, and was recently found to form a protein complex with the yeast RAD4 protein (Guzder *et al.*, 1995), a structural homolog of XPC.

Except for potential phosphorylation sites, analysis of the primary amino acid sequence of XPC gave no clues about a particular function. The primary amino acid sequence of RAD23 protein and its mammalian homologues indicated that they are N-terminal ubiquitin-like fusion proteins (Watkins *et al.*, 1993, Masutani *et al.*, 1994). In addition, a second link with the ubiquitin pathway was observed. Two repeated domains in the RAD23 amino acid sequence shared homology to a C-terminal extension in a bovine ubiquitin-conjugating enzyme (E2-25kD) (van der Spek *et al.*, 1996 in press). This suggests that the RAD23 protein may have an involvement in the ubiquitin system, within the context of NER or in another process, implying a dual functionality. Other NER proteins have also been found to have dual functions. Examples include the XPB and XPD proteins in the multisubunit TFIIH transcription repair factor (Schaeffer *et al.*, 1993) and the RAD1-RAD10 complex, additionally involved in mitotic recombination (Schiestl *et al.*, 1993).

In the present studies, we tried to find evidence for a stable association of XPC and HHR23 proteins with each other and with previously identified protein complexes which have defined enzymatic activity, involving endonuclease-mediated incision (ERCC1/ERCC4) or transcription initiation activity (TFIIH). Heparin fractionation experiments revealed that HHR23A and a large fraction of HHR23B resided in the flowthrough fraction (Fig. 4A). Native gel electrophoresis indicated that the vast majority of both HHR23 proteins was present in the free, non-complexed form (Fig. 5) a finding

supported by the gelfiltration experiments. The heparin, phosphocellulose and size fractionation experiments as well as the immunodepletion studies all confirmed the complex formation of XPC protein with HHR23B protein (Figs. 4A, 4B and 6A). No HHR23A protein could be detected in these purified heparin or phosphocellulose fractions containing XPC. From these findings we conclude that for HHR23B two forms exist: the majority is in a free form, whilst a small fraction is complexed with XPC. For HHR23A, we can only detect a free form, although it is not excluded that a fraction below our detection level is complexed with XPC or another protein. The absence of detectable quantities of HHR23A in the XPC-HHR23B containing high-salt fractions from the heparin (Fig. 4A) and phosphocellulose chromatography suggests that HHR23A may not be functionally fully equivalent to HHR23B. The small fraction of HHR23B that is complexed with XPC is necessary for NER (Sugasawa *et al.*, submitted). This raises the question whether HHR23A and the free from of HHR23B are involved in NER at all and/or whether they have an additional function. These proteins resided in the flow-through of the phosphocellulose fractionation. Previously, Aboussekhra *et al.* (1995) showed that only the RP-A complex and the PCNA protein from this fraction are necessary for *in vitro* NER. These data therefore suggest that HHR23A does not play a role in the core NER reaction. However, the *in vitro* system might not reflect the step in which this protein plays a role *in vivo*. If only (the XPC-bound) HHR23B has a role in NER, one might wonder why no rodent or human mutants for HHR23B were found (Fig. 3). A possible explanation for the absence can be the dual function, that might give rise to an unexpected phenotype. Alternatively, HHR23A may bind to XPC when HHR23B is absent. A clear answer on what is the function of the free form of both mammalian RAD23 equivalents and whether they are functionally redundant should come from analysis of mutants generated by gene targeting and from *in vitro* reconstitution experiments (both experiments in progress).

Gel filtration studies suggested that the XPC protein can be part of protein complexes of large size (Fig. 6A). The purified XPC/HHR23B complex was previously determined to have a molecular weight of 500-550 kDa by gel filtration and a value of 110 kDa by glycerol gradients (Masutani *et al.*, 1994). Here we found in fractionated Manley type cell extracts a molecular weight bigger than these previously determined values. This suggests that the XPC/HHR23B proteins are part of a bigger complex, that can easily fall apart during purification in a stable XPC/HHR23B subcomplex and other proteins. However, it cannot theoretically be excluded that XPC protein selectively multimerizes or aggregates. Therefore, it was investigated whether the large molecular

weigth XPC-containing complex also includes TFIIH and ERCC1 components, conforming with the 'repairosome' model reported by Svejstrup *et al.* (1995) for yeast NER. Non-denaturing gel electrophoresis showed distinct bands for HHR23A (70 kDa) and HHR23B (150 kDa), for both the HeLa proteins as well as the *E. coli* produced recombinant polypeptides. In this context, it should be noted that native molecular weight estimations themselves should be taken with caution, since these may strongly depend on the conformation of the proteins or protein complexes. However, bands migrating at a similar position can provide evidence for complex formation. Therefore, it is evident that both native molecular weights are different from the 280 kDa previously described for the ERCC1 complex (van Vuuren *et al.*, 1994) and the minimal molecular weight calculated for TFIIH.

Fractions containing highly purified ERCC1 complex described by van Vuuren *et al.* (1995) were also checked for the presence of HHR23 proteins, and were found to be negative. This makes a tight association of these proteins with the ERCC1 complex highly unlikely.

From the data presented here one can also conclude that there is no stable complex of a significant fraction of XPC/HHR23B and TFIIH, under these conditions and in our (Manley) extracts. This is in conflict with the findings of Drapkin *et al.* (1994), who after six purification steps for TFIIH components, still detected XPC protein in the purified fractions. However, their TFIIH complex is not completely pure, and no data are provided with respect to the yield and the fraction (percentage) of XPC present in the TFIIH preparation. Therefore, it is hard to identify whether this is a significant amount of XPC and whether cross-contamination is excluded. From our unpublished results, it appeared that the XPC complex by coincidence behaved in a similar way during several purification steps as TFIIH. Moreover, no physical interaction was shown (e.g. immunodepletion) by Drapkin *et al.* (1994). On the other hand it should be stressed that, our data on XPC/TFIIH interaction do not exclude a transient association, as reported by Bardwell *et al.* (1994) for the *S. cerevisiae* system. The absence of any detectable interactions between these factors is of relevance in the context of the evidence for a "repairosome" in *S. cerevisiae* (Svejstrup *et al.* 1995), in which (almost) all NER components are represented in one super complex. A difference might exist between yeast and mammals. Alternatively, a 'NERosome' in mammals may be more delicate, and might disassociate sooner than its yeast counterpart. Therefore, the extract preparation procedure can be of crucial relevance.

The subcellular localization of a protein can provide possible clues about its

function. Immunofluoresence data displayed a clear nuclear localization of the XPC protein on interphase cells (Fig. 7). This finding supports a function in DNA metabolism and is in agreement with the previously described ssDNA binding activity (Masutani *et al.*, 1994). The observation that during anaphase and telophase, XPC specifically associates with chromatin, suggests a role for XPC after the metaphase/anaphase transition, and is consistent with the DNA binding activity of the XPC-HHR23B protein complex. However, such a function remains to be clarified.

Nuclear localization was also found for both HHR23 proteins, consistent with a role for these proteins in DNA or chromatin metabolism. These results are in accordance with previous *S. cerevisiae* data, in which the RAD23 protein was described to be nuclear (Watkins *et al.*, 1993). Both recombinant HHR23 proteins were found not to have specific affinity for ss or dsDNA, and therefore, a direct role for these proteins in DNA damage recognition can be regarded as unlikely.

Additional functions for both HHR23 proteins seem however likely, based on the excess of both HHR23 proteins in the cell compared to XPC, as deduced from the experiments presented in this paper.

ACKNOWLEDGEMENTS

We are grateful to Dr. Michael McKay for helpful suggestions and critically reading of the manuscript. Sigrid Swagemakers is acknowledged for her help with the COS-1 transfections and immunofluoresence experiments and Dr. J.-M. Egly (Strasbourg) is acknowledged for some of the antibodies used in this work. We thank Mirko Kuit for photography. The work at the Department of Cell Biology and Genetics of the Erasmus University was financially supported by the Medical Genetics Centre South-West Netherlands and in part by a Human Frontier Science Program Research grant. Dr Kaoru Sugasawa was supported by a grant of the Biodesign Research Program of the Institute of Physical and Chemical Research (RIKEN).

REFERENCES

Aboussekhra, A., Biggerstaff, M., Shivji, M.K.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protic, M., Hübscher, U., Egly, J.-M. and Wood, R.D. (1995). Mammalian DNA Nucleotide excision repair reconstituted with purified protein components. *Cell* 80: 859-868.

Bardwell, A.J., Bardwell, L., Iyer, N., Svejstrup, J.Q., Feaver, W.J., Kornberg, R.D., and Friedberg E.C. (1994). Yeast nucleotide excision repair proteins rad2 and rad4 interact with RNA

polymerase II basal transcription factor b (TFIIH). *Mol. Cell. Biol.* 14: 3569-3576.

Biggerstaff, M., Szymkowski, D.E., and Wood, R.D. (1993). Co-correction of ERCC1, ERCC4 and xeroderma pigmentosum group F DNA repair defects *in vitro*. *EMBO J.* 12: 3685-3692.

Bohr, V.A. (1991). Gene specific repair. *Carcinogenesis* 12: 1983-1992.

Bootsma, D., Weeda, G., Vermeulen, W., van Vuuren, H., Troelstra, C., van der Spek, P.J. and Hoeijmakers, J.H.J. (1995). Nucleotide excision repair syndromes: molecular basis and clinical symptoms. *Phil. Trans. R. Soc. Lond.* 347: 75-81.

Christians, F.C. and Hanawalt, P.C. (1994). Repair in ribosomal RNA genes is deficient in xeroderma pigmentosum group C and in Cockayne's syndrome cells. *Mutat. Res.* 323: 179-187.

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell* 79: 13-21.

Cleaver, J.E. and Kraemer, K.H. (1994). Xeroderma pigmentosum and Cockayne syndrome. In: The metabolic basis for inherited disease. Seventh edition, Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D., eds. (New York: McGraw-Hill Book Co.).

Collins, A.R. (1993). Mutant rodent cell lines sensitive to ultraviolet light, ionizing radiation and cross-linking agents: a comprehensive survey of genetic and biochemical characteristics. *Mutat. Res.* 293: 99-118.

Dingwall, C. and Laskey R.A. (1991). Nuclear targeting sequences -a consensus?- Trends Biochem. Sci. 16: 478-481.

Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.C., Zawel, L., Ahn, K., Sancar, A. and Reinberg, D. (1994). Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* 368: 769-772.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132: 6-13.

Fischer, L., Gerard, M., Chalut, C., Lutz Y., Humbert, S., Kanno, M., Chambon, P., and Egly, J.-M. (1992). Cloning of the 62-kilodalton component of basic transcription factor BTF2. *Science* 257: 1392-1395.

Gerard, M. Fischer, L., Moncollin, V., Chipoulet, J.M., Chambon, P. and Egly, J.M. (1991). Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J. Biol. Chem.* 266: 20940-20945.

Guzder, S.N., Bailly, V., Sung, P., Prakash, L. and Prakash, S. (1995). Yeast DNA repair protein RAD23 promotes complex formation between transcription factor TFIIH and DNA damage recognition factor RAD14. *J. Biol. Chem.* 270: 8385-8388.

Halley, D.J.J., Keijzer, W., Jaspers, N.G.J., Niermeijer, M.F., Kleijer, W.J., Boue, J., Boue, A. and Bootsma, D. (1979). Prenatal diagnosis of xeroderma pigmentosum (group C) using assays of unscheduled DNA synthesis and postreplication repair. *Clin. Genet.* 16: 137-146.

Hanawalt, P.C. (1994). Transcription-coupled repair and human disease. *Science* 266: 1957-1958.

Hanawalt, P. and Mellon, I. (1993). Stranded in an active gene. *Curr. Biol.* 3: 67-69.

Harlow, E., and Lane, D. (1988). Antibodies: A laboratory manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Hoeijmakers, J.H.J. (1993). Nucleotide excision repair II: from yeast to mammals. *TIG* 9: 211-217.

Itin, P.H. and Pittelkow, M.R. (1990). Trichothiodystrophy: Review of sulfur-deficient brittle hair syndromes and association with the ectodermal dysplasias. *J. Americ. Acad. Dermatol.* 22: 705-717.

Jentsch, S., and Schlenker, S. (1995). Selective protein degradation: A journey's end within the proteasome. *Cell* 82: 881-884.

Kantor, G.J. Barsalou, L.S., and Hanawalt, P.C. (1990). Selective repair of specific chromatin domains in UV-irradiated cells from xeroderma pigmentosum complementation group C. *Mut. Res.* 235: 171-180.

Lehmann, A.R. (1987). Cockayne's syndrome and trichothiodystrophy: Defective repair without cancer. *Cancer Rev.* 7: 82-103.

Li, L., Bales, E.S., Peterson, C.A. and Legerski, R.J. (1993). Characterization of molecular defects in xeroderma pigmentosum group C. *Nature Genet.* 5: 413-417.

Manley, J.L., Fire, A., Samuels, M. and Sharp, P.A. (1983). *In vitro* transciption: whole cell extract. *Methods Enzymol.* 101: 568-582.

Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio K., Tanaka, K., van der Spek, P.J., Bootsma D., Hoeijmakers, J.H.J., and Hanaoka F. (1994). Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homolog of yeast RAD23. EMBO J. **8:** 1831-1843.

McCready, S. (1994). Repair of 6-4 photoproducts and cyclobutane pyrimidine dimers in *rad* mutants of *Saccharomyces cerevisiae*. *Mut. Res.* 315: 261-273.

Reed, K.C., and Mann, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucl. Acids. Res.* 13 7207-7221.

Riboni, R., Botta, E., Stefanini, M., Numata, M. and Yasui, A. (1992). Identification of the eleventh complementation group of UV-sensitive excision repair-defective rodent mutants. *Cancer Res.* 52: 6690-6691.

Sambrook, J., Fritsch, E.F. and Maniatis, T.E. (1989). *Molecular cloning, a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H.J., Chambon, P. and Egly, J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) transcription factor. *Science* 260: 58-63.

Schiestl, R.H., and Prakash, S. (1990). *RAD10*, an excision repair gene of Saccharomyces cerevisiae, is involved in the *RAD1* pathway of mitotic recombination. *Mol. Cell. Biol.* 10: 2485-2491.

Smith, D.E., and Fisher, P.A. (1984). Identification, Developmental regulation, and response to heat shock of two antigenically related forms of a major nuclear envelope protein in Drosophila embryos: Aplication of an improved method for affinity purification of antibodies using polypeptides immobilized on nitrocellulose blots. *J. Cell Biol.* 99: 20-28.

Sugasawa, K., Masutani, C., Uchida, A., Maekawa, T., van der Spek, P.J., Bootsma, D., Hoeijmakers, J.H.J., and Hanaoka, F. (1995). HHR23B, a human RAD23 homologue and a component of xeroderma pigmentosum group C correcting protein, is required *in vitro* for nucleotide excision repair. (submitted for publication, 1995).

Svejstrup, J.Q., Wang, Z., Feaver, W.J., Wu, X., Bushnell, D.A., Donahue, T.F., Friedberg, E.C. and Kornberg, R.D. (1995). Different forms of TFIIH for transcription and DNA repair: Holo-TFIIH and a nucleotide excision repairosome. *Cell* 80: 21-28.

van der Spek, P.J., Visser, C.E., Hanaoka, F., Smit, B., Hagemeijer, A., Bootsma, D., and Hoeijmakers, J.H.J. (1996). Cloning, comparative mapping, and RNA expression of the mouse

homologues of the *Saccharomyces cerevisiae* nucleotide excision repair gene RAD23. *Genomics* 31, In press.

Van Vuuren, A.J., Appeldoorn, E., Jaspers N.G.J., Odijk, H., Yasui, A., Bootsma, D. and Hoeijmakers, J.H.J. (1993). Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. *EMBO J.* 12: 3693-3701.

van Vuuren, A.J., Appeldoorn, E., Odijk, H., Humbert, S., Moncollin, V., Eker, A.P.M., Jaspers, N.G.J., Egly, J.-M. and Hoeijmakers, J.H.J. (1995). Partial characterization of the DNA repair complex, containing ERCC1, ERCC4, ERCC11 and XPF correcting activities. *Mut. Res.* 337: 25-39.

Van Vuuren, A.J., Vermeulen, W., Ma, L., Weeda G., Appeldoorn, E., Jaspers N.G.J., van der Eb, A.J., Bootsma, D., Hoeijmakers, J.H.J., Humbert, S., Schaeffer, L. and Egly, J.M. (1994). Correction of xeroderma pigmentosum repair defect by basal transcription factor BTF2 (TFIIH). *EMBO J.* 13: 1645-1653.

Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H.F. (1991). Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. *Mol. Cell. Biol.* 11: 4128-4134.

Venema, J., Van Hoffen, A., Natarajan, A.T., Van Zeeland, A.A. and Mullenders, L.H.F. (1990). The residual repair capacity of xeroderma pigmentosum complementation group C fibroblasts is highly specific for transcriptionally active DNA. *Nucl. Acids Res.* 18: 443-448.

Watkins, J.F., Sung, P., Prakash, L. and Prakash, S. (1993). The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* 13: 7757-7765.

Wood, R.D., Robins, P. and Lindahl, T. (1988). Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. *Cell* 53: 97-106.

Wood, C.M., Timme, T.L., Hurt, M.N., Brinkley, B.R., Ledbetter, D.H. and Moses, R.E. (1987). Transcription of DNA repair-deficient human diploid fibroblasts with Simian virus 40 plasmid. *Exp. Cell Res.* 169: 543-553.

CHAPTER IX

Concluding remarks

CONCLUDING REMARKS

General considerations

A major goal of the human genome project is to analyse the genetic blueprint of our genome and to elucidate the mechanisms of control of gene expression in ordered time- and cell type-dependent fashion. As a consequence, enormous amounts of DNA sequence data are being generated. To cope with this large amount of information, powerful computers and sequence comparison algorithms are essential. This thesis illustrates examples of the use of the data resulting from the human genome project for the field of DNA repair. New genes are now being identified by computer-assisted sequence comparison. Such approaches are expected to have a major impact in biology (The nature genome directory).

Dysfunctional DNA repair has recently been shown to play a major role in the genesis of a diverse array of cancers. At present, the vast majority of approaches to the treatment of malignant disease are empirically-based. It follows that an enhanced understanding of the mechanism of DNA repair holds promise for the future development of targeted and highly specific molecular-based therapies for a variety of malignancies. UV-sensitive, NER defective xeroderma pigmentosum patients have a defect in NER which predisposes these individuals to skin cancer (Cleaver and Kraemer, 1994).

Non-melanoma skin cancer is one of the most frequently occurring neoplasms in Caucasian populations, and strong epidemiologic and other evidence implicates solar UV-irradiation in its genesis. According to the multi-step model of tumorigenesis (Nowell, 1976), the induction of DNA damage is considered an important initiating step for gene mutation and the subsequent initiation of cancer.

Other DNA repair disorders are also associated with increased cancer risk. For example, hereditary non-polyposis colon cancer is caused by microsatellite instability as a result of a defective mismatch repair system. Cancers at other body sites may arise by a similar molecular defect in this repair pathway (Fishel and Kolodner, 1995).

Defective cell cycle regulatory genes involved in cell cycle progression can also be associated with cancer (Hartwell and Kastan, 1994; Lydall and Weinert, 1995). Ample evidence points to disordered cell cycle regulation in neoplasia (Hunter and Pines, 1994). Given the fundamental role for ubiquitin in cyclin degradation, the ubiquitin-conjugating pathway is likely to have a central role in the development of those tumors in which cyclin dysfunction plays a role in carcinogenesis (Barinaga, 1995).

Relation of RAD23 to HHR23 proteins and xeroderma pigmentosum

As shown in the studies reported in this thesis, sequence homology is of significant relevance for the cloning of homologous genes in higher organisms. Many yeast *RAD* genes proved to have human homologs that are involved in genetic disorders such as xeroderma pigmentosum and Cockayne syndrome (Hoeijmakers, 1993a/b), suggesting that functional aspects of the DNA excision repair pathway are well conserved during evolution. Cloning and characterization of both mouse and human homologs of the yeast *RAD23* NER gene are described in detail in this thesis. The duplicated human cognates of RAD23 are putative candidate genes for the remaining NER complementation groups for which the defective gene has not yet been identified. However, for both these human gene products, no mammalian mutant has as yet been identified.

Purification of the XPC protein using an *in vitro* cell-free repair system containing UV-damaged SV40 minichromosomes as a substrate, uncovered a second protein of 58 kDa; the protein copurified with the 125 kDa XPC protein (see Chapter III). The 58 kDa protein appeared to be by one of the two human homologs of yeast RAD23, HHR23B. Strengthening the observation of an association between human XPC and HHR23B proteins, Guzder *et al.* described complex formation between the structurally homologous yeast RAD4 and RAD23 proteins (Guzder *et al.*, 1995). In XPC cells, only the transcription-coupled repair pathway is active. However, as discussed in Chapter 2, this is contradictory with the phenotype of a *rad4* mutant, a mutant defective in both transcription-coupled and genome overall repair.

It was shown by Mullenders *et al.* (1984) that repair synthesis, confined to active genes in XPC cells, was preferentially associated with the attachment sites of chromatin loops at the nuclear matrix. Indirect evidence for a separate repair pathway came from previous work on CS cells (Mayne and Lehmann, 1982). Transcription coupled repair is directed towards the transcribed strand (Mellon, 1987). This process was shown to be defective in CS patients. Lesions in the nontranscribed DNA strand and overall genome are thus repaired normally in these patients. In contrast to CS, XPC individuals have a GGR defect, whereas TCR is apparently normal. Notably, XP-C individuals are cancer-prone whereas CS individuals are not. These facts suggest that defects in the GGR NER subpathway, rather than TCR, are linked to the processes involved in carcinogenesis. Cancer-prone XPC patients do not manifest the accelerated neurodegeneration characteristic of some of the other XP complementation groups (Bootsma *et al.*, 1995). This is consistent with the idea that the transcription-coupled repair pathway, defects in which may be crucial to this type of neurodysfunction, is not affected in XPC.

Details concerning the interactions of the HHR23 and XPC repair proteins with each other and with DNA were assessed. No clear association with TFIIH has been shown for the human enzymes encoding HHR23A, HHR23B or XPC. Mammalian data do not corroborate *S. cerevisiae* data which claimed an association of RAD23 and the yeast general transcription factor b (Guzder *et al.*, 1995). It is possible that the reason for this discrepancy relates to a difference in the stability of yeast *versus* mammalian protein complexes: subtle differences in stability of the complexes may result in their dissociation during purification for mammals, but not for yeast. Alternatively, there may be a true difference between yeast and mammals in this regard. The biochemical behaviour of the HHR23 and XPC enzymes is distinct from the properties of the transcription repair complex TFIIH, containing the p62, XPB and XPD proteins (see Chapter VIII). Analysis of purified fractions containing ERCC1/ERCC4(XPF) complexes for the presence of XPC and HHR23 equivalents did not suggest any stable association of a significant proportion of these complexes with each other. Additionally, the relationship of HHR23 and XPC proteins to CSB was studied to further investigate whether the claimed yeast repairosome existed in mammals. No copurification was found for any of the HHR23, XPC factors with the CSB protein.

The yeast *RAD7* and *RAD16* genes are involved in the repair of lesions on the non-transcribed DNA strand, representing a similar phenotype to that observed for XPC in human. It is possible that their as of yet unidentified human counterparts also subserve the same function in the genome overall NER sub-pathway. RAD7, RAD16 and RAD23 gene products have been described as necessary for the removal of pyrimidine dimers from the nontranscribed strand of transcriptionally active genes (Verhage *et al.*, 1994). All three gene products may be involved in alterations of chromatin structure that provide greater accessibility for the catalytically active repair complex. However, in contrast to *rad7* and *rad16*, the *rad23* disruption mutant is defective in GGR and TCR. Therefore, the intermediate UV-sensitivity of this mutant cannot be explained by a defect in one or the other NER subpathway, as has been proposed to explain the intermediate UV-sensitivity of *RAD7* and *RAD16* mutants. The specific function of the XPC/HHR23B complex in genome-overall repair still remains to be elucidated.

In both normal and cancerous cells, control of proliferation and cell cycle progression is exerted through changes in the levels of transiently or constitutively short-lived regulatory proteins. Ubiquitin conjugation is known to be an important essential mechanism in cellular proteolysis (Ciechnover, 1994). Based on sequential homology of the primary amino acid sequence, RAD23 exhibits a dual link with the ubiquitin-conjugating pathway (see Chapter VI). The N-terminal region of the RAD23 homologs harbors a ubiquitin-like part that has

been demonstrated to be required for the repair function of the yeast protein (Watkins *et al.*, 1993). A second link exists between RAD23 and the ubiquitin pathway: a twice repeated element, homologous to a C-terminal extension of a Class II ubiquitin-conjugating enzyme (E2) has been identified (see Chapter VI). In addition to degradation of proteins tagged by ubiquitin, conjugation of ubiquitin is also involved in DNA repair, the stress response, chromosome (de)condensation, and cell cycle progression/control (Ciechanover, 1994).

As is the case for almost all NER genes. The nuclear ubiquitin-like fusion protein RAD23 has no prokaryotic homologs, and is more likely to have a regulatory role rather than a direct role in DNA repair. However, the relatively high abundance of the human homologs of RAD23 compared to other repair enzymes such as XPC, suggests a more general function, such as a chromatin modulating function analogous to the RAD6 protein that is involved in ubiquitin-conjugating pathway.

Future directions

Since NER is strongly conserved between yeast and humans, knock-out mice might be expected to mimic the clinical defects of human XP, CS and TTD patients. In particular the sensitivity to genotoxic agents and the predisposition to cancer of the xeroderma patients can be assessed with these models. Furthermore, the phenotype of the mice might provide clues for potential candidate syndromes which show a similar clinical picture. Therefore, experiments are in progress to generate MHR23A and MHR23B knockout mice. Although yeast *rad23* null mutants are viable and sporulation is not affected, it remains to be seen whether disruption of *MHR23* genes results in a vital phenotype. Particular aspects of interest in the resulting phenotypes will be the status of NER and spermatogenesis. The presumed function of RAD23 in meiosis as suggested by studies in yeast can be evaluated during oogenesis and spermatogenesis in these mice.

The vast majority of transgenic animals reported to date have been knock-out mutations (to generate null alleles). In these animals, the effects of complete absence of gene function can be assessed. However, the recent introduction of targeted subtle genetic alterations can facilitate more comprehensive study of the functional role of the gene of interest, to model human genetic syndromes.

The MHR23-deficient mice will be intercrossed with hairless mice. Using such models, the influence of the mammalian RAD23 homologs on UV-light-induced skin cancer will be investigated. Furthermore, the influence of various chemical mutagens (e.g., DMBA,

acetoxy-AAF and benzo(a)pyrene) on carcinogenesis will be analysed. Crossing of various repair-deficient and cancer gene-inactivated mice can indicate the contribution of the different repair systems to multistep carcinogenesis. The induced tumors can be analysed to identify which oncogenes are activated during the process of malignant transformation. Moreover, HHR23B and XPC-deficient mice can be crossed and the consequences of impaired genome overall repair for spontaneous and induced mutagenesis thereby investigated using the *hprt* gene as a marker. Furthermore, this can be achieved by crossing these mice with transgenic mice containing several copies of the LacZ reporter gene for scoring mutations.

Double knock-outs might provide clues about the functional redundancy of both *HHR23* gene products: if HHR23A and HHR23B knockouts are independently viable but the double mutant has a lethal phenotype, this would be strong evidence for functional redundancy of the gene products.

Both XPA and XPC-deficient mice mimic the patient phenotype with respect to carcinogenesis. In contrast to the XPC-deficient mice, CSB-deficient mice should provide clues to the contribution of transcription-coupled repair to the clinical features of NER patients. It will be also of interest to examine crossings of XPC homozygous null mutant animals with CS-B knockouts, to investigate the effect of knocking out both known NER subpathways; likewise, crossing MHR23B and CS-B knockout animals will be of interest given that XPC and HHR23B are complexed.

In order to further understand the biochemical function of both HHR23 proteins, a deduction or prediction of the three-dimensional structure of the proteins would be highly informative. The striking homology with ubiquitin at the N-terminus of the RAD23 protein and its various homologs pointed to similar structural features of these ubiquitin-like fusion proteins. Since the crystal structure of ubiquitin is known, it was determined via molecular modeling whether RAD23 could have a similar structure. Recombinant proteins obtained from cloned HHR23 genes in efficient expression vectors are being purified to homogeneity in large quantities, to facilitate attempts to crystallize both HHR23 proteins. The three-dimensional protein structure, together with biochemical and genetic data, should provide greater insight into the exact function of both mammalian RAD23 equivalents.

Understanding the molecular mechanisms of fundamental cellular processes such as DNA repair, transcription, recombination, and cell division, should continue to increase, in view of the rapidly expanding genome sequence data. The impact of this mega project will undoubtedly continue to be felt throughout all fields of biology.

References

Barinaga, M. (1995). A new twist to the cell cycle. *Science* 269: 631-632.

Bootsma, D., Weeda, G., Vermeulen, W., van Vuuren, H., Troelstra, C., van der Spek, P.J. and Hoeijmakers, J.H.J. (1995). Nucleotide excision repair syndromes: molecular basis and clinical symptoms. *Phil. Trans. R. Soc. Lond.* 347: 75-81.

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell* 79: 13-21.

Cleaver, J.E. and Kraemer, K.H. (1994). Xeroderma pigmentosum and Cockayne syndrome. In: The metabolic basis for inherited disease. Seventh edition, Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D., eds. (New York: McGraw-Hill Book Co.).

Fishel, R., and Kolodner, R.D. (1995). Identification of mismatch repair genes and their role in the development of cancer. *Curr. Biol.* 5: 382-395.

Friedberg, E.C., Walker, G.C. and Siede, W. (1995). DNA repair and mutagenesis. ASM Press Washington, D.C.

Guzder, S.N., Bailly, V., Sung, P., Prakash, L., and Prakash, S. (1995). Yeast DNA repair protein RAD23 promotes complex formation between transcription factor TFIIH and DNA damage recognition factor RAD14. *J. Biol. Chem.* 270: 8385-8388.

Hartwell, L. and Kastan, M.B. (1994). Cell cycle control and cancer. *Science* 266: 1821-1828.

Hoeijmakers, J.H.J. (1993a). Nucleotide excision repair I: from E.coli to yeast. *TIG* 9: 173-177.

Hoeijmakers, J.H.J. (1993b). Nucleotide excision repair II: from yeast to mammals. *TIG* 9: 211-217.

Hunter, T., and Pines, J. (1994). Cyclins and cancer II: Cyclin D and CDK inhibitors come of age. *Cell* 79: 573-582.

Mayne, L.V., and Lehmann, A.R. (1982). Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Research* 42: 1473-1478.

Mellon, I., Spivak, G., and Hanawalt, P.C. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 51: 241-249.

Mullenders, L.H., van Kesteren, A.C., Bussmann, C.J., van Zeeland, A.A. Natarajan, A.T. (1984).Preferential repair of nuclear matrix associated DNA in xeroderma pigmentosum complementation group C. *Mut. Res.* 141: 75-82.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* 194: 23-28.

The nature genome directory. *Nature* supplement to Vol. 377.

Verhage, R., Zeeman, A.-M., de Groot, N., Gleig, F., Bang, D.D., van de Putte, P. and Brouwer, J. (1994). The *RAD7* and *RAD16* genes, which are essential for pyrimidine dimer removal from the silent mating type loci, are also required for repair of the nontranscribed strand of an active gene in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 14: 6135-6142.

Watkins, J.F., Sung, P., Prakash, L., and Prakash, S. (1993). The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* 13: 7757-7765.

Summary

Samenvatting

SUMMARY

Genome instability is regarded to be caused by defects in processes such as DNA repair, cell cycle arrest or apoptosis, and is presumably an early step in carcinogenesis. Several pathways exist that are capable of removing chemical or radiation-induced DNA damage. One of the major DNA damage repair mechanisms is the nucleotide excision repair (NER) pathway. This system removes a broad spectrum of DNA lesions, for example (6-4) photoproducts, cyclobutane pyrimidine dimers (the main UV-induced injuries) and bulky chemical adducts. Three NER-deficient human syndromes, all characterized by marked photosensitivity are known: xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD). XP patients show pigmentation abnormalities and an over thousand-fold increased risk of skin cancer, caused by defects in one of at least seven genes (*XPA* to *XPG*). CS patients display overall developmental impairment. The two genes responsible for the CS phenotype have been cloned (*CSA* and *CSB*). TTD is characterized by brittle hair, ichthyosis and many CS symptoms. The NER defect in TTD is due to mutations in the *TTDA*, *XPB* or *XPD* genes. No increased risk of cancer is reported for CS and TTD patients. Another subclass of patients shows a combined XP-CS phenotype. These patients can be assigned to complementation groups XP-B, XP-D or XP-G.

Chapter I of this thesis is a general introduction. Genes and gene products involved in different repair pathways are discussed in Chapter II, with special attention to the NER pathway. The multi-step NER reaction entails damage recognition, local unwinding of both complementary strands, dual incision of the injured strand, removal of the damage-containing oligonucleotide, DNA repair synthesis to fill the single strand gap and ligation. Two NER subpathways exist: rapid transcription-coupled repair limited to the transcribed strand of active genes, and the less efficient global genome repair. TTD and most XP groups carry defects that affect both subpathways, CS is specifically deficient in transcription-coupled repair, whereas XP-C individuals are impaired in the global genome repair pathway.

The value of sequencing model organisms and the complete human genome have been shown to be of major relevance for geneticists in general, and in this project, for the identification of DNA repair-related genes that might be important to the study of the complex process of carcinogenesis. One of the first demonstrations of a powerful new technology called "computer-cloning" is the identification of two human homologs of the *S.cerevisiae RAD23* NER gene designated, *HHR23A* and *HHR23B* (described in Chapter

III). Such computer-assisted analysis is described in detail for the yeast *RAD23* gene, together with other examples of genes (SNM1, XPE/DDB, PHR) identified in this way.

The identification and characterization of a DNA repair protein complex, involved in the genome overall repair pathway, is described in Chapter IV of this thesis. This complex consists of the xeroderma pigmentosum group C correcting protein, stably associated with the HHR23B gene product. This protein heterodimer exhibits a high affinity for ssDNA and is specifically implicated in global genome NER. Computer comparison revealed the presence of an N-terminal ubiquitin-like domain in all RAD23 derivatives.

Two protein complexes involved in NER (TFIIH; ERCC1/XPF) were identified during the last few years: the TFIIH complex (which contains the NER genes, *XPB* and *XPD*, mutations of which occur in XP, XP-CS and TTD); and the XPF/ERCC1 complex which has a dual role, one involved in making the incision in the injured DNA strand, as well as a second function in the mitotic recombination pathway. The XPC-HHR23B complex is the third such complex, and is involved in global genome NER. Many but not all components of these repair complexes represent a particular NER syndrome or hamster complementation group. Involvement of NER genes in the basal transcription complex TFIIH has important clinical implications, since some symptoms in TTD, not explained by a NER defect, are likely to be caused by subtle insufficiencies in basal transcription.

In Chapter V, the chromosomal localization of the *XPC* and the *HHR23* genes is described. The *HHR23A* gene was assigned to 19p13.1. Both *XPC* and *HHR23B* reside on chromosome 3p25.1. The colocalization of these genes is remarkable since both gene products form also a tight complex at the protein level. The chromosomal locations of the human HHR23 genes did not match any known deletion or rearrangement sites in mammalian NER-defective mutants. Given the high amino sequence homology between both gene products it is possible the HHR23A and HHR23B proteins have largely overlapping functions. To unravel the molecular mechanism of different repair pathways in mammals and to assess the biological relevance of these processes, we have isolated, mapped and characterized the mouse homologs of RAD23, *MHR23A* and *MHR23B* genes, as reported in Chapter VI. Physical disconnection of *MHR23B* and *XPC* genes in mouse argues against an important functional significance of the colocalization of these genes observed in human. Moreover, a dual link with the ubiquitin-conjugating pathway was found. All RAD23 homologs contain a ubiquitin-like N-terminus followed by a strongly conserved 50-amino-acid domain that is repeated at the C-terminus. This domain is highly

homologous to a specific C-terminal extension of one of the ubiquitin-conjugating enzymes. These putative functional implications are reinforced by the recent finding in yeast that RAD23 has been found to be a suppressor of a ubiquitination mutant (Madura, personal communication). Chapter VII deals with the role of HHR23B in repair, tested in an *in vitro* reconstituted repair assay. It was shown that both HHR23B and XPC proteins are required for NER. Moreover, it was found that only a small portion of HHR23B exists in a bound form with XPC, while a large fraction is present in unbound form. In Chapter VIII, the protein characterization of the XPC and HHR23 homologs and their relation to TFIIH components is addressed. Purification and immuno-depletion experiments suggested that the XPC activity was not stably associated with the activity of basal transcription initiation. The HHR23A protein has been found to reside in the free non-bound form. The relatively high expression of HHR23 equivalents compared to the XPC protein points to an additional function for the HHR23 proteins. Immunofluoresence data have shown HHR23 and XPC proteins to be localized to the nucleus. Moreover, the nuclear XPC protein specifically associated with chromatin during anaphase and telophase.

These mouse HHR23 genes are being used to generate (NER-) deficient cell lines and mice through homologous recombination in embryonic stem cells. The impact of these gene products on processes such as DNA repair, recombination, meiosis and carcinogenesis can then be assessed *in vivo*.

SAMENVATTING

Genoom instabiliteit wordt gezien als een gevolg van een defect in processen zoals DNA herstel, cel cyclus regulatie of apoptose en is mogelijk een vroege stap in carcinogenese. Een van de belangrijkste DNA herstelmechanismen is het nucleotide excisie reparatie (NER) systeem. Dit mechanisme is in staat een breed spectrum van DNA schades zoals 6-4 fotoprodukten, cyclobutaan-dimeren (de meest voorkomende UV-geïnduceerde schades) en 'bulky' adducten te verwijderen. Drie NER-deficiënte humane syndromen gekarakteriseerd door fotosensitiviteit zijn: xeroderma pigmentosum (XP), Cockayne syndroom (CS) en trichothiodystrofie (TTD). XP patiënten vertonen pigmentatie abnormaliteiten naast een meer dan duizendmaal verhoogde kans op huidkanker, veroorzaakt door een defect in één van de op z'n minst zeven genen (*XPA* t/m *XPG*). CS patiënten vertonen zonlichtgevoeligheid, ernstige groeistoornissen en afwijkingen in het centraal zenuwstelsel. De twee genen verantwoordelijk voor het CS fenotype zijn gekloneerd, te weten *CSA* en *CSB*. TTD wordt gekarakteriseerd door breekbare haren, ichthyosis en een aantal CS symptomen. Het NER defect in TTD wordt veroorzaakt door mutaties in het *TTDA*, *XPB* of *XPD* gen. Voor CS en TTD patiënten is echter geen verhoogde kans op kanker gerapporteerd. Een andere subklasse van patiënten vertoont een gecombineerd XP-CS fenotype. Deze patiënten behoren tot de complementatie groepen XP-B, XP-D of XP-G.

Hoofdstuk I van dit proefschrift is een algemene introductie. Genen en genprodukten, betrokken in de verschillende reparatieprocessen zijn beschreven in hoofdstuk II, met de nadruk op het NER mechanisme. De meerstaps NER reactie omvat schadeherkenning, lokale ontwinding van de complementaire DNA strengen, dubbele incisie van de beschadigde streng, verwijdering van de beschadigde oligonucleotide, DNA synthese ter opvulling van het ontstane enkelstrengs gat en ligatie. Er bestaan twee NER subroutes, te weten: het snelle transcriptie gekoppelde herstel dat gelimiteerd is tot de getranscribeerde streng van de actieve genen, en het minder efficiënte globale genoom herstel. TTD en de meeste XP groepen dragen defecten die beide subroutes beïnvloeden, terwijl CS specifiek deficiënt is in transcriptie gekoppeld herstel, en in XPC-patiënten is het globale genoom herstel aangedaan.

De relevantie van sequencing projecten van model organismen en het humane genoom heeft aangetoond van wezenlijk belang te zijn voor genetici in z'n algemeenheid en in dit project voor de identificatie van DNA-reparatie gerelateerde genen. Deze genen zijn van belang voor het bestuderen van een complex proces als carcinogenese. Een voorbeeld van gen identificatie middels computervergelijking, bekend als computer-kloneren, is de identificatie van twee humane homologen van het *S. cerevisiae RAD23* NER gen, aangeduid

als *HHR23A* en *HHR23B* (beschreven in hoofdstuk III). De identificatie van RAD23 homologen middels computervergelijking is in detail beschreven, samen met andere voorbeelden van andere genen (*SNM1*, *XPE/DDB* en *PHR*), waarvan op gelijke wijze homologen geïdentificeerd zijn.

In hoofdstuk IV van dit proefschrift is de identificatie en karakterisering beschreven van een DNA herstel eiwit complex, dat specifiek betrokken is bij globaal genoom herstel. Dit complex bestaat uit het xeroderma pigmentosum group C eiwit, dat een stabiele associatie heeft met het HHR23B eiwit. Deze heterodimeer vertoont hoge affiniteit voor enkelstrengs DNA en is specifiek betrokken bij globaal genoom herstel. Computervergelijking heeft uitgewezen, dat alle RAD23 equivalenten een ubiquitine-verwant domein bevatten aan de N-terminus.

Twee eiwit complexen betrokken in NER (TFIIH en ERCC1/XPF) zijn in de afgelopen jaren reeds beschreven: het TFIIH complex (dat onder andere bestaat uit de NER genen *XPB* en *XPD*; waarin mutaties gevonden zijn in XP, XP-CS en TTD patiënten) en het XPF/ERCC1 complex, dat een dubbele rol heeft, ten eerste in het maken van de 5' incisie van de beschadigde DNA streng, en ten tweede in mitotische recombinatie. Het XPC-HHR23B complex is het derde NER complex. Dit complex speelt een specifieke rol in het globaal genoom herstel. Veel, maar niet alle componenten van deze reparatie complexen zijn betrokken bij een bepaald NER syndroom of een NER deficiënte hamster complementatie groep. De betrokkenheid van NER genen in basale transcriptie zoals bij TFIIH het geval is heeft belangrijke klinische implicaties, vanwege het feit dat sommige klinische verschijnselen in TTD niet verklaard kunnen worden door een defect in NER maar waarschijnlijk het gevolg zijn van een subtiel defect in basale transcriptie.

In hoofdstuk V is de chromosomale lokalisatie van het XPC en de HHR23 genen beschreven. Het *HHR23A* gen is gelokaliseerd op chromosoom 19p13.1. Zowel het *XPC* als het *HHR23B* gen liggen op chromosoom 3p25.1. Deze co-lokalisatie is opmerkelijk daar beide genprodukten tevens een stabiel eiwit complex vormen. De chromosomale lokalisatie van beide humane *HHR23* genen gaf geen aanwijzingen met betrekking tot bekende deleties of genherschikkingen in NER deficiënte zoogdier mutanten. Teneinde inzicht te verkrijgen in het moleculaire mechanisme van de verschillende herstel systemen in zoogdieren en om de biologische relevantie van deze processen te kunnen bestuderen, zijn de muize-homologen van RAD23 geïsoleerd, MHR23A en MHR23B, gelokaliseerd en gekarakteriseerd zoals beschreven in hoofdstuk VI. Co-lokalisatie van *MHR23B* en *XPC* genen is afwezig bij de muis, hetgeen een mogelijke functionele betekenis van de co-lokalisatie bij de mens minder waarschijnlijk maakt. Tevens is een tweeledig verband met het ubiquitine-systeem naar voren

gekomen. Alle RAD23 homologen bevatten naast een ubiquitine-verwante N-terminus, een sterk geconserveerd domein van 50 aminozuren dat gerepeteerd aanwezig is in de C-terminus. Dit domein is sterk homoloog aan een specifieke C-terminale extensie van één van de ubiquitine conjugerende enzymen. Deze aanwijzingen voor een ubiquitine gerelateerde functie worden versterkt door de recente bevinding in gist dat het RAD23 een suppressor is van een ubiquitine mutant (Dr. Madura). Hoofdstuk VII behandelt de rol van het HHR23B eiwit in DNA reparatie, getest in een *in vitro* nagebootste reparatie reactie. Aangetoond is dat zowel het HHR23B als het XPC eiwit nodig zijn voor NER. Daarnaast is gevonden dat een klein deel van HHR23B gecomplexeerd is met XPC eiwit en dat het grootste gedeelte van het HHR23B eiwit in vrije vorm voorkomt. In hoofdstuk VIII is de eiwit karakterisering van XPC en de HHR23 homologen en de relatie met TFIIH componenten belicht. Zuivering en immuno-depletie experimenten suggereren dat de XPC activiteit niet in een stabiel complex geassocieerd is met basale transcriptie initiatie reparatie factor TFIIH. Het HHR23A eiwit is enkel detecteerbaar in vrije vorm in de cel. De relatief hoge expressie van de HHR23 equivalenten in vergelijking met XPC eiwit duidt op een additionele functie voor deze HHR23 eiwitten. Immunofluorescentie experimenten hebben aangetoond dat de HHR23 en XPC eiwitten in de celkern gelokaliseerd zijn. Daarnaast is voor XPC een specifieke associatie met het chromatine gevonden tijdens de anafase en telofase.

De geïsoleerde muize-genen worden gebruikt om (NER)-deficiënte cellijnen en muizen, middels homologe recombinatie, in embryonale stamcellen te verkrijgen. Het effect van deze genprodukten op processen als DNA herstel, recombinatie, meiose en carcinogenese wordt hierdoor toegankelijk *in vivo*.

Curriculum vitae

Acknowledgements

CURRICULUM VITAE

Name:	Petrus Johannes van der Spek
Born:	2 november 1966, Delft, The Netherlands.
September 1986:	Graduated High School 'De Lage Waard', Papendrecht, The Netherlands.
August 1989:	Student at the Department of Cell Biology and Genetics, Erasmus University, Rotterdam, the Netherlands. Supervisor: Prof. Dr. J.H.J. Hoeijmakers
September 1990:	Graduated HLO-Biochemistry. 'Hogeschool West-Brabant', Etten-Leur, The Netherlands.
August 1990:	PhD. student, Medical faculty, Department of Cell Biology and Genetics, Erasmus University, Rotterdam, the Netherlands. Promotors: Prof. Dr. D. Bootsma Prof. Dr. J.H.J. Hoeijmakers
November 1991:	Oxford University Certificate of proficiency in English.
January 1992:	Course Pharmacogenesis; Toxicology and Pharmacology at the Dr. Paul Janssen Stichting, Janssen Research Council, Belgium.
May 1996:	Long-term post-doctoral position at the MRC Cell Mutation Unit, University of Sussex, Brighton, United Kingdom, and short-term position at the Institute for Molecular and Cellular Biology, University of Osaka, Japan.

List of Publications:

Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23.

Masutani, C., Sugasawa,. K., Yanagisawa, J., Sonoyama, T., Ui, M., Enemoto, T., Takio, K., Tanaka, K., Van der Spek, P.J., Bootsma, D., Hoeijmakers, J.H.J. and Hanaoka, F.

EMBO J. 13: 1831-1843, 1994

Chromosomal localization of three repair genes: the xeroderma pigmentosum group C gene and two human homologs of yeast RAD23.

Van der Spek, P.J., Smit, E.M.E., Beverloo, H.B., Sugasawa, K., Masutani, C., Hanaoka, F., Hoeijmakers, J.H.J. and Hagemeijer, A.

Genomics 23: 651-658, 1994

Nucleotide excision repair syndromes: molecular basis and clinical symptoms.

Dirk Bootsma, Geert Weeda, Wim Vermeulen, Hanneke van Vuuren, Christine Troelstra, Peter van der Spek, and Jan Hoeijmakers

Phil. Trans. Roy. Soc. Lond. 347: 75-81, 1995

Involvement of a spliced and defective human foamy virus in the establishment of chronic infection.

Ali Saïb, Marcel H.M. Koken, Peter J. van der Spek, Jorge Périès, and Hugues de Thé.

Journal of Virology 69: 5261-5268, 1995

Cloning, comparative mapping and RNA expression of the mouse homologs of the *S. cerevisiae* nucleotide excision repair gene *RAD23*.

Peter J. van der Spek, Cécile Visser, Fumio Hanaoka, Bep Smit, Anne Hagemeijer, Dirk Bootsma and Jan H.J. Hoeijmakers.

Genomics 31: 20-27, 1996

Conference report: 10th International Congress of Radiation Research Würzburg, Germany, August 27-September 1, 1995.

Michael McKay, Peter van der Spek, Henk Roest and Dirk Bootsma.

Mutation Research, in press, 1996

Requirement of HHR23B, a human RAD23 homologue, for nucleotide excision repair *in vitro* as a stimulatory factor of the XPC protein.

Kaoru Sugasawa, Chikahide Masutani, Akio Uchida, Takafumi Maekawa, Peter J. van der Spek, Dirk Bootsma, Jan H.J. Hoeijmakers and Fumio Hanaoka.

Submitted, 1995

XPC and human homologs of RAD23: intracellular localization and relationship to other nucleotide excision repair complexes.

Peter J. van der Spek, André Eker, Suzanne Rademakers, Cécile Visser, Kaoru Sugasawa, Chikahide Masutani, Fumio Hanaoka, Jan H.J. Hoeijmakers and Dirk Bootsma.

Submitted, 1996

Different levels of ERCC1 are required for repair of crosslinks and UV-damage.

Anneke M. Sijbers, Peter J. van der Spek, Hanny Odijk, Joke van den Berg, Marcel van Duin, Andries Westerveld, Nicolaas G.J. Jaspers, Dirk Bootsma, and Jan H.J. Hoeijmakers.

Submitted, 1996

Evolutionary conservation in eukaryotes of the *Schizosaccharomyces pombe* DNA double- strand break repair gene *rad21*.

Michael J. McKay, Christine Troelstra, Peter J. van der Spek, Roland Kanaar, Bep Smit, Anne Hagemeijer, Dirk Bootsma and Jan H.J. Hoeijmakers.

Submitted, 1996

A human photolyase homolog with similarity to blue light receptors.

P.J. van der Spek, K. Kobayashi, D. Bootsma, M. Takao, A.P.M. Eker and A. Yasui.

Manuscript in preparation, 1996

ACKNOWLEDGEMENTS

Iedereen die een bijdrage heeft geleverd aan de totstandkoming van dit proefschrift ben ik zeer dankbaar. Op de eerste plaats mijn ouders, die mij altijd gestimuleerd en gemotiveerd hebben, zonder hen was ik niet zo ver gekomen. Mijn broer Joost, die mij telkens weer probeerde uit te leggen dat er meer was dan het lab; bedankt, ééns tik ik je weer van de tennisbaan met 6-0, 6-0 !

Professor Bootsma, ik ben U zeer erkentelijk voor de gelegenheid die U mij heeft geboden en het vertrouwen dat in mij is gesteld. Prof. Hoeijmakers, beste Jan, jij als stimulator, promotor van dit onderzoek, jouw nimmer aflatende enthousiasme, ook bij de soms kleine stapjes vooruit, mag niet onvermeld blijven. Heel veel dank voor de leerzame jaren onder jouw vleugels.

Sigrid, en René, paranimfen, hartelijk bedankt, niet alleen voor jullie grote belangstelling voor mijn onderzoek, maar vooral ook voor de hechte vriendschap en de geweldige steun het afgelopen jaar.

Medewerkers van de Instituten Genetica, Celbiologie, en Klinische Genetica van de Erasmus Universiteit Rotterdam (EUR), bedankt voor de prettige samenwerking tijdens mijn onderzoek. Dit door het MGC geïnitieerde project kwam mede tot stand door de goede verstandhouding met verschillende mensen uit Leiden en Rotterdam.

In het bijzonder bedank ik mijn collega's van de DNA-herstel groep, niet alleen voor de fijne werksfeer en voor alle steun, die ik gedurende mijn studenten- en promotietijd van jullie kreeg, maar ook voor de gezelligheid buiten werktijd. Cécile Visser en Suzanne Rademakers wil ik bedanken voor het werk dat zij voor mij hebben gedaan als stagaires en later als analistes. Paula, bedankt voor de part-time assistentie en home-made boterkoek. Andere medewerkers van het "herstel-oord", Alain, André (zuiverings goeroe), Anja, Anneke (del *.*), Bas, Betty (pronto), Christine (screen queen), Coen, Esther, Geert, Hanny, Henk, Ingrid, Jan de Boer, Jan de Wit (kweek godfather), Jeroen, Wim (ouwe prikker) en Wouter, dank ik voor ieders bijdrage, op eigen wijze, aan dit boekje. De discussies, onmisbare literatuur en correspondentie macro's van Koos mogen hier ook niet onvermeld blijven. Marcel Koken, plezierige herinneringen heb ik aan de begin periode van mijn onderzoek, toen jij mij ("snel") hebt leren pipetteren. Nog vaak denk ik aan de dag dat er een nieuwe trend werd gezet in het kloneren. Het doet mij goed te kunnen zeggen dat computer cloning nu eindelijk een erkende procedure is geworden! Roland, er zijn vele wegen die naar Würzburg leiden!? Tijd voor 'n puntje. Bert, Kaoru en Harry, wens ik veel succes met het vervolg onderzoek aan de MHR23 knockout muizen. Dr. Levendag, de werkbesprekingen en

discussies op de afdeling klinische radiobiologie van de Dr. Daniel den Hoed Kliniek waren een bijzondere ervaring en een nieuwe dimensie.

Beste collega's van 24, dank voor jullie gastvrijheid en met name Frans voor de gezellige lease-bench bij jullie boven en lease-stoel beneden. Andre, Nicole en Marianne, dank voor de inzet en goede suggesties voor het eiwitwerk. Arnold Reuser de lage drempel en de altijd open staande kamerdeur heb ik zeer op prijs gesteld! Tevens moet het chromosomen-lab op de achtste vermeld worden. Met name Prof. A. Hagemeijer, Bep, Berna en Ellen voor de prettige samenwerking en het inplakken van de postzegels.

De heer Lansbergen, Ans, John en Ron van het Centraal Dierenlaboratorium van de EUR, heel veel dank voor jullie deskundigheid en assistentie bij het uitvoeren van de dierexperimenten. Deze lijst zou verre van compleet zijn zonder de volgende namen: Rita, en Marike van het secretariaat, Jopie, Elly, Joke en Robby van de keuken, handy man Piet en inkoop Rein, Melle en Mieke. Zonder de medewerking van vooral Mirko, Tom en Ruud had dit boekje geen foto's en figuren gehad. Mirko bedankt voor de vele "kleine" uurtjes die hierin zijn gaan zitten.

I am indebted to all persons who have contributed to the realization of this thesis. I especially would like to thank Fumio Hanaoka and Kaoru Sugasawa for the open collaboration, sharing of results and friendship during the past three years. Moreover, from Japan, Akira Yasui and Nobuo Nomura have to be mentioned for the collaboration on *PHR* and *XPE* genes respectively. With regard to my visits to NIH in Bethesda (USA), I am especially indebted to Eric Ackerman who was a great host during the different stays at Washington. I also feel privileged to have met Vilhelm Bohr, David Busch, Tony Carr, Carlo Crocé, Alan Lehmann, Jos van Renswoude, Larry Thompson, Christine Weber and many others who made my various stays both scientifically as socially memorable ones. Jean-Marc Egly, à votre santé.

Cecilia en René, naast alle andere vrienden die er begrip voor hadden als ik weer eens "even" moest werken verdienen jullie een speciaal plekje. Grazie per la fantastica atmosfera dei giorni passati ad Arezzo e per il sostegno e l'attenzione mostrati. Angelique en Rene, de gezellige zeil vakanties en vlieg activiteiten hebben mij op de been en letterlijk van de straat gehouden, hiervoor veel dank. Ruud, zoals je ziet, 't grapje met het computer programma is niet voor niets geweest. Bedankt voor de goede introductie in de binaire wereld. Paul, veel dank voor de goede discussies, industriële zakelijke visie, zaterdag acties, vele reizen, en de leerzame gastvrije periode op jouw kantoor. And last but not least Michael McKay, many thanks for the helpful discussions, clinical views, and critical reading of chapters and manuscripts. It was a pleasure to have you, Kim, and the boys around. Thesis finished! Faired dinkum pal.