

## Distribution of DNA repair-related ESTs in sugarcane

W.C. Lima, R. Medina-Silva, R.S. Galhardo and C.F.M. Menck\*

### Abstract

DNA repair pathways are necessary to maintain the proper genomic stability and ensure the survival of the organism, protecting it against the damaging effects of endogenous and exogenous agents. In this work, we made an analysis of the expression patterns of DNA repair-related genes in sugarcane, by determining the EST (expressed sequence tags) distribution in the different cDNA libraries of the SUCEST transcriptome project. Three different pathways - photoreactivation, base excision repair and nucleotide excision repair - were investigated by employing known DNA repair proteins as probes to identify homologous ESTs in sugarcane, by means of computer similarity search. The results showed that DNA repair genes may have differential expressions in tissues, depending on the pathway studied. These *in silico* data provide important clues on the potential variation of gene expression, to be confirmed by direct biochemical analysis.

### INTRODUCTION

The genome of all living beings is constantly subject to damage generated by exogenous and endogenous factors, reducing DNA stability and leading to an increase of mutagenesis, cancer, cell death, senescence and other deleterious effects to organisms (de Laat *et al.*, 1999). Repair mechanisms for such lesions are required, in order to maintain the necessary genomic integrity and organism survival. DNA repair pathways may thus affect mutation rates, chromosome segregation, recombination and genetic exchange within and among populations (Vonarx *et al.*, 1998)

The environmental exposure of plants to light [including ultraviolet (UV)], heat, desiccation and chemicals makes these organisms particularly susceptible to genotoxic effects. The inability of plants to minimize environmental exposure increases the importance of DNA repair mechanisms (Britt, 1996; Tuteja *et al.*, 2001). DNA damage results in various physiological effects, such as reduced protein synthesis, cell membrane destruction, damage to photosynthetic proteins, affecting growth and development of the whole organism (Britt, 1999). Particularly UV-B radiation (290 to 320 nm), the main damaging light that reaches the surface of Earth, has many direct and indirect effects on plants, including damage to DNA, proteins and membranes; alterations in transcription and photosynthesis; changes in growth, development and morphology; and genetic instability (Teramura and Sullivan, 1994). UV-B can also produce decreased levels of photosynthetic pigments, altered thylakoid integrity, increased stomatal diffusion and reduced photosynthesis activity (Strid *et al.*, 1994).

The recent efforts to sequence the genome of several plants, particularly the genetic model *Arabidopsis thaliana*

(The *Arabidopsis* Genome Initiative, 2000), have provided huge amounts of data that still need to be processed, in order to enable us to understand the physiological mechanisms of these organisms. This is the case of the DNA repair pathways. Although repair and damage tolerance mechanisms have been well described in bacteria, yeast, humans and rodents, remarkably little is known about these processes in plants.

The knowledge of the regulation of DNA repair gene expression is essential to the understanding of the biological relevance of this process regarding the resistance to cytotoxic and mutagenic effects of environmental and endogenous DNA-damaging agents (Shi *et al.*, 1997). The expression patterns of specific repair genes in plants have been analyzed by several authors (Shi *et al.*, 1997; Desprez *et al.*, 1998; Liu *et al.*, 2000; Deveaux *et al.*, 2000); however, further studies are required to disclose how such genes and proteins play their roles in different plant tissues.

EST (expressed sequence tags) sequencing projects for plants may provide hints on the expression patterns of specific genes, or even of whole pathways, by the analysis of the number of reads in different tissues. The sugarcane EST project (SUCEST) has identified close to 300,000 reads of cDNA sequences from the transcriptome of this monocotyledon, by analyzing several tissues and plant treatments, with a relatively low read redundancy. This database was used to identify genes related to DNA repair mechanisms, as described in this issue (Costa *et al.*, 2001).

In this work, we employed the SUCEST database to test the possibility of obtaining differential expression data by using bioinformatic similarity tools, the so-called electronic northern. Known DNA repair proteins from three different pathways - direct repair, base excision repair and

nucleotide excision repair - were employed as probes. Although the number of reads obtained was too small to reach any significant results for specific genes, we found that, if general pathways are taken into account, interesting data is revealed. The results indicate that most of these genes are similarly expressed throughout sugarcane lifespan, but some variation in gene expression among the different libraries provides important clues for further studies.

## METHODS

Known DNA repair protein sequences from *Arabidopsis thaliana* were used as probes for the search of homologous sugarcane ESTs. The protein probes were compared to the SUCEST Database using tBlastN (Altschul *et al.*, 1997) to find homolog clusters, at the level of translated cDNA sequences. Hits were considered significant whenever e-values were smaller than  $e^{-10}$ . These clusters were then separated into reads, allowing to identify the original libraries. Each cluster was only counted once, even when matched with different probes from paralogous genes.

Table I shows the correspondence of each library with the plant tissue (or treatment). The number of reads obtained for each probe was added to others of the same DNA repair pathway, and then the values were normalized in relation to the total number of reads for each library. The statistical significance of the differences between the values obtained and the average of the reads for each type of library was assessed by using Student's T-test.

## RESULTS AND DISCUSSION

In this report, we present an analysis of the EST expression of DNA repair-related genes in sugarcane. The identification of these ESTs was made by *in silico* observations, through similarity search against known DNA repair

**Table I** - Identification of SUCEST libraries.

Name	Library	Number of reads
AD	<i>Gluconacetobacter diazotrophicus</i> infected plantlets	18,137
AM	Apical Meristem	28,128
CL	Callus	11,872
FL	Sugarcane flower	64,095
HR	<i>Herbaspirillum rubrisubalbicans</i> infected plantlets	12,000
LB	Lateral Bud	18,047
LR	Leaf roll	18,141
LV	Leaves (etiolated)	6,432
RT	Root	31,487
RZ	Leaf-root transition zone	24,096
SB	Stem Bark	16,318
SD	Seed	21,406
ST	Stem	20,762

proteins. It is important to point out that, although the similarity limit used in this work (e-value  $< e^{-10}$  for EST clusters) was conservative, part of the ESTs obtained were not necessarily related directly to the probes. In fact, some included ESTs are components of gene families, not necessarily involved in any DNA repair pathway. However, since these paralogous genes are part of the central core of DNA metabolism (such as cell cycle control and transcriptional activators), they were accepted for this work.

A list of the protein probes tested in this analysis is presented in Table II. Two genes, encoding plant homologues of the yeast RAD23 and of the human CSB proteins, are highly represented (over 100 reads) among the ESTs found in sugarcane. One possibility to explain such high expression levels is that gene families are identified, with many representative genes in plants. RAD23, for example,

**Table II** - Number of clusters and reads identified in the SUCEST Data Bank.

Repair pathway	Probe gene	Number of clusters (e-value $< e^{-10}$ )	Number of Reads
Direct repair		9	40
	(6-4) PHR	1	1
	PHR1	1	6
	PHR2	1	13
	CRY1 E 2	6	20
Base excision repair		18	80
<i>Glycosilase</i>	3MG1	6	42
	MUTM 1 E 2	4	17
	NTH	1	7
	OGG	1	2
	UNG	2	2
	<i>Endonuclease</i>	ARP	4
Nucleotide excision repair		71	543
<i>TFIIH</i>	P44	1	8
	XPB	2	10
	XPD	3	5
<i>Recognition</i>	RPA1	6	25
	XAB1	5	37
	XAB2	2	8
<i>Incision</i>	ERCC1	1	6
	ERCC5	2	8
	XPF	2	11
<i>TCR</i>	XPG	1	1
	CSA	2	8
<i>GGR</i>	CSB	28	200
	RAD16	7	44
	RAD23	3	109
	XPC	1	7
	XPE	5	42

has ubiquitin and ubiquitin-associated domains, which are necessary for the proteolysis involved in the turnover of proteins required for the control of the cell cycle progression (Bachmair *et al.*, 2001). CSB is a helicase part of the SNF2/RAD54 family of proteins, involved in a variety of processes including transcription regulation (Eisen *et al.*, 1995).

A second limitation of this approach is the low number of reads from some genes that interferes with the significance of the data. Therefore, we decided to group the genes, in order to collect data concerning the number of ESTs related to a specific pathway in each library. Three pathways that resulted in a significant number of matches were chosen for the detection of ESTs, as discussed below.

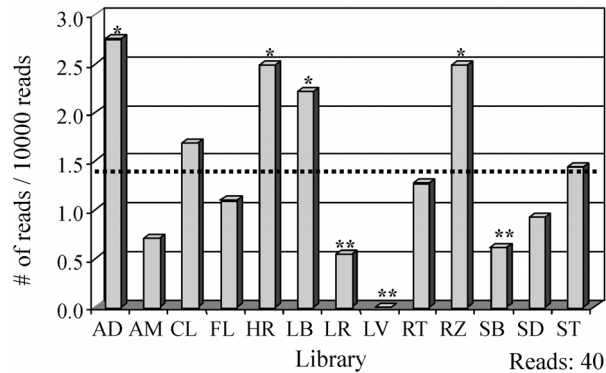
### Direct repair

Reversal of the lesion is the simplest mechanism by which damaged DNA can be repaired. The biochemical mechanism is based on a one-step reaction, where a specific enzyme recognizes and reverts the lesion to its normal configuration in an error-free manner. In the SUCEST database, only ESTs related to photoreactivation were identified; no EST related to alkyltransferase was found.

Photolyases are enzymes which revert the main lesions induced by UV, which are cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts, in a light-dependent reaction called photoreactivation (PHR) (Todo, 1999). CPD-photolyases were reported in all groups of organisms, including bacteria, fungi, plants, invertebrates and vertebrates (Eisen and Hanawalt, 1999). There are two classes of CPD-photolyases: the class I family is formed by microbial photolyases, whereas class II comprises eukaryotic photolyases (Sancar, 1994). (6-4) Photolyases were described in some invertebrates and vertebrates. CRY genes, which code for a class of plant blue-light photoreceptors, have no repair activity, although they are homologous to photolyases (Todo, 1999).

In this work, we used as probes the sequences of the photolyase family proteins, including the plant blue-light photoreceptors (CRY 1 and CRY 2). A noteworthy finding was the small number of reads detected. The identification of clusters with high similarity to the photolyase family led to an interesting distribution of the reads per library (Figure 1). There was a strong expression in the lateral bud (LB), root-leaf transition zone (RZ) and bacteria-infected plant libraries (AD and HR), but few reads in the apical meristem (AM) and stem bark (SB); no reads at all were detected in the etiolated leaves (LV).

The requirement of photolyases for photorepair implies a major expression in tissues exposed to light (Deisenhofer, 2000). The high number of ESTs related to this protein family in plants that were infected by bacteria (libraries AD and HR) is thus unexpected. It may be related to a general stress response in these plants. However, the



**Figure 1** - Distribution of reads of the photolyase gene family. The dashed line represents the average of reads observed in all tissues and those indicated by (\*) or (\*\*) are significantly different from the average ( $P < 0.01$  as determined by the *T*-test).

absence of such ESTs in etiolated leaves grown in the dark favors the idea that these ESTs are regulated by light exposure of the plant. In similar conditions, Taylor *et al.* (1996) reported that no transcript levels of photolyases were detected, which is in agreement with our observations.

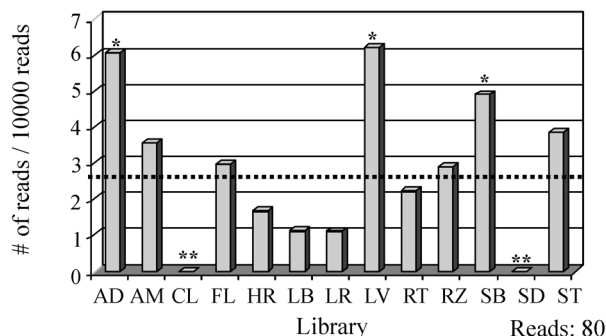
### Excision repair

Although efficient, the direct reversal of lesions is limited, due to its high substrate/enzyme specificity. The living organisms developed a more general DNA repair mechanism, by which damaged bases are removed and replaced by normal nucleotide sequences. This type of cellular response to DNA damage is generally called excision repair and includes the two different pathways described below.

#### Base excision repair (BER)

BER proteins play an important role in the excision of damaged bases generated by endogenous (reactive oxygen species, hydrolytic events, and cellular metabolites) and exogenous (various forms of radiation and environmental alkylating agents) factors (Cunningham, 1997). The first step of BER involves the removal of a single damaged base through the action of DNA glycosylases, which are able to recognize and excise the damaged base from the sugar phosphate backbone (Krokan *et al.*, 1997). This reaction results in an abasic (AP: apurinic/aprimidinic) site, which is recognized by another group of enzymes, the AP-endonucleases, which make an incision at the 5' or 3' phosphodiester bound at the AP site, generating a nucleotide gap. This gap is filled through polymerization and ligation of a new nucleotide to the DNA sequence (Seeberg *et al.*, 1995; Cadet *et al.*, 2000).

In this work, only DNA glycosylases and endonucleases were investigated (Table II), as they are the main representatives of BER. The results are shown in Figure 2. BER genes were preferentially expressed in leaves (LV), stem tissues (SB), as well as in plants infected with



**Figure 2** - Distribution of reads of the BER pathway. The dashed line represents the average of reads observed in all tissues and those indicated by (\*) or (\*\*) are significantly different from the average ( $P < 0.01$  as determined by the *T*-test).

*Gluconacetobacter* (AD). No expression was detected in seeds (SD) and callus (CL). The high expression of genes in differentiated tissues like stem and leaves, and their low expression in undifferentiated tissues (callus) correlate well with the protective role of BER genes, maintaining the integrity of the genome against lesions generated by endogenous agents, in organs with high metabolic rates.

In seeds, DNA lesions occur not only during dehydration associated with dormancy, but also as a consequence of genotoxic environmental exposure. Thus, during their dormancy and quiescence periods, there is little opportunity for DNA repair, in accordance with the low expression of BER genes (Shi *et al.*, 1997). However, before germination, plant embryos undergo a period of intense DNA repair prior to the initiation of replication, as a guarantee for the successful growth of seedlings (Osborne, 1993). Therefore, these DNA repair enzymes appear to be important for germination, and the absence of ESTs in seeds would be explained if stable BER related-proteins were stored in the embryo.

#### Nucleotide excision repair (NER)

NER is one of the most flexible and general DNA repair pathways, removing a large spectrum of structurally unrelated DNA lesions that generate considerable helical distortion in DNA (Prakash and Prakash, 2000; de Boer and Hoeijmakers, 2000). A single-stranded segment containing the lesion is removed by dual incision of the damaged strand with subsequent gap-filling (Wood, 1997; de Laat *et al.*, 1999). The basic NER mechanism is conserved from bacteria to humans. However, the prokaryotic and eukaryotic NER machinery has diverged significantly, since the involved proteins do not display sequence homology. Nevertheless, NER genes and proteins isolated from humans, yeast, and plants show significant sequence similarity (Eisen and Hanawalt, 1999).

In NER, two sub-pathways for damage detection were identified. The Global Genome Repair (GGR) system operates genome-wide (in non-transcribed DNA) and recognizes different lesions with different levels of efficiency.

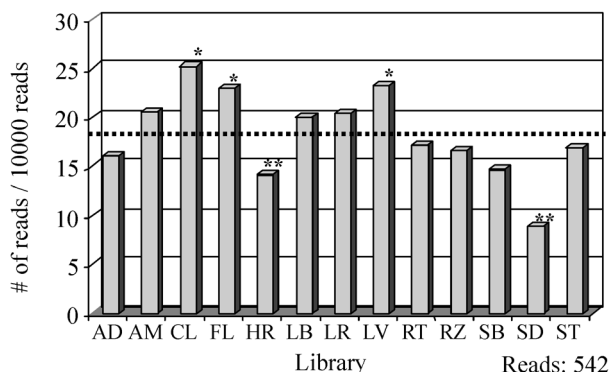
A second NER sub-pathway, named Transcription-Coupled Repair (TCR), was developed, which specifically removes lesions from the transcribed strand of active genes. TCR focuses on lesions that block ongoing transcription, and is important for the timely resumption of RNA synthesis after genomic insults (Selby and Sancar, 1990).

Considered as a whole, NER genes have a high number of reads with a uniform distribution of the ESTs in the different sugarcane libraries (Figure 3). However, a few differences were still observed: the reads were more frequent in callus (CL), flower (FL) and leaf (LV) libraries, and particularly less frequent in seeds. A similar expression pattern was observed when GGR and TCR were considered independently (data not shown).

The uniform expression of NER genes corroborates their important role in the maintenance of vital functions during development (including division, elongation and differentiation). The preferential expression of NER genes in callus may be related to the rapidly dividing cells, requiring high repair levels (Hanawalt, 2001; Balajee and Bohr, 2000), or may simply be due to the undifferentiated cell state in this tissue. On the other hand, the low expression level in seeds is also interesting, as the embryo may need these proteins in its early development. As for BER-related ESTs, this observation needs further investigation, with *in vitro* experiments, as it suggests that seeds may contain stable DNA repair proteins.

#### CONCLUDING REMARKS

The elevated number of reads in several libraries from different tissues, as well as the low redundancy of these libraries in the SUCEST transcriptome project, allow an *in silico* approach for the evaluation of gene expression in different tissues and treatment conditions in sugarcane. Computer analysis of the ESTs in three DNA repair pathways revealed some interesting expression patterns, which may be related to their roles in DNA repair during plant development and differentiation. However, this genomic analysis



**Figure 3** - Distribution of reads of the NER pathway. The dashed line represents the average of reads observed in all tissues and those indicated by (\*) or (\*\*) are significantly different from the average ( $P < 0.01$  as determined by the *T*-test).

should be regarded with caution, as it represents just a first screening that provided important hints and the basis for future biochemical investigation.

### RESUMO

As vias de reparo de DNA são requeridas para manter a necessária estabilidade genômica e garantir a sobrevivência do organismo, frente aos efeitos deletérios causados por fatores endógenos e exógenos. Neste trabalho, realizamos a análise dos padrões de expressão dos genes de reparo de DNA encontrados na cana-de-açúcar, pela determinação da distribuição de ESTs nas diferentes bibliotecas de cDNA no projeto de transcriptoma SUCEST. Três vias de reparo - fotorreativação, reparo por excisão de bases e reparo por excisão de nucleotídeos - foram estudadas através do uso de proteínas de reparo como sondas para identificação de ESTs homólogos em cana-de-açúcar, com base na procura computacional de similaridade. Os resultados indicam que os genes de reparo de DNA possuem uma expressão diferencial nos tecidos, dependendo da via de reparo analisada. Esses dados *in silico* fornecem importantes indícios da expressão diferencial, a qual deve ser confirmada por análises bioquímicas diretas.

### ACKNOWLEDGMENTS

This work was supported by FAPESP (São Paulo, SP, Brazil) and CNPq/PADCT (Brasília, DF, Brazil). C.F.M.M. is a CNPq Research Fellow; W.C.L., R.M.S. and R.S.G. have scholarships from FAPESP.

### REFERENCES

- Altschul, S.F., Madden T.L., Schaffer A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25 (17): 3389-402.
- Bachmair, A., Novatchkova, M., Potuschak, T., and Eisenhaber, F. (2001). Ubiquitylation in plants: a post-genomic look at a post-translational modification. *Trends Plant Sci.* 6 (10): 463-70.
- Balajee, A.S. and Bohr, V.A. (2000). Genomic heterogeneity of nucleotide excision repair. *Gene* 250 (1-2): 15-30.
- Britt, A.B. (1996). DNA damage and repair in plants. *Annu Rev Plant Physiol Plant Mol Biol.* 47: 75-100.
- Britt, A.B. (1999). Molecular genetics of DNA repair in higher plants. *Trends Plant Sci.* 4 (1): 20-25.
- Cadet, J., Bourdat, A.G., D'ham, C., Duarte, V., Gasparutto, D., Romieu, A. and Ravanat, J.L. (2000). Oxidative base damage to DNA: specificity of base excision repair enzymes. *Mut Res.* 462 (2-3): 121-128.
- Costa, R.M.A., Lima, W.C., Vogel, C.G.I., Berra, C.M., Luche, D.D., Medina-Silva, R., Galhardo, R.S., Menck, C.F.M. and Oliveira, V.R. (2001). DNA repair-related genes in sugarcane ESTs. *Gen Mol Biol.* 24 (1-4): 131-140.
- Cunningham, R.P. (1997). DNA glycosylases. *Mut Res* 383 (3): 189-196.
- De Boer, J. and Hoeijmakers, J.H. (2000). Nucleotide excision repair and human syndromes. *Carcinogenesis* 21 (3): 453-60.
- De Laat, W.L., Jaspers, N.G. and Hoeijmakers, J.H. (1999). Molecular mechanisms of nucleotide excision repair. *Genes Dev.* 13 (7): 768-85.
- Deisenhofer, J. (2000). DNA photolyases and cryptochromes. *Mutat Res* 460 (3-4): 143-9.
- Desprez, T., Amselem, J., Caboche, M. and Hofte, H. (1998). Differential gene expression in *Arabidopsis* monitored using cDNA arrays. *Plant J* 14 (5): 643-52.
- Deveaux, Y., Alonso, B., Pierrugues, O., Godon, C. and Kazmaier, M. (2000). Molecular cloning and developmental expression of AtGR1, a new growth-related *Arabidopsis* gene strongly induced by ionizing radiation. *Rad Res* 154 (4): 355-64.
- Eisen, J.A. and Hanawalt, P.C. (1999). A phylogenomic study of DNA repair genes, proteins and processes. *Mut Res.* 435: 171-213.
- Eisen, J.A., Sweder, K.S. and Hanawalt, P.C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* 23 (14): 2715-23.
- Hanawalt, P.C. (2001). Controlling the efficiency of excision repair. *Mut Res* 485: 3-13.
- Krokan, H.E., Standal, R. and Slupphang, G. (1997). DNA glycosylases in the base excision repair of DNA. *Biochem J.* 325: 1-16.
- Liu, Z., Hossain, G.S., Islas-Osuna, M.A., Mitchell, D.L. and Mount, D.W. (2000). Repair of UV damage in plants by nucleotide excision repair: *Arabidopsis* UVH1 DNA repair gene is a homolog of *Saccharomyces cerevisiae* Rad1. *Plant J* 21 (6): 519-28.
- Osborne, D.J. (1993). Function of DNA synthesis and DNA repair in the survival of embryos during early germination and in dormancy. *Seed Sci Res* 3: 43-53.
- Prakash, S. and Prakash, L. (2000). Nucleotide excision repair in yeast. *Mut Res* 451: 13-24.
- Sancar, A. (1994). Structure and function of DNA photolyase. *Biochemistry.* 33 (1): 2-9.
- Seeberg, E., Eide, L. and Bjoras, M. (1995). The base excision repair pathway. *Trends Biochem Sci.* 20: 391-97.
- Selby, C.P. and Sancar, A. (1990). Transcription preferentially inhibits nucleotide excision repair of the template DNA strand *in vitro*. *J Biol. Chem.* 265 (34): 21330-6.
- Shi, L., Kent, R., Bence, N. and Britt, A.B. (1997). Developmental expression of a DNA repair gene in *Arabidopsis*. *Mut Res.* 384 (3): 145-56.
- Strid, A., Chow, W.S. and Anderson, J.M. (1994). UV-B damage and protection at the molecular level in plants. *Photosynth Res.* 39: 475-89.
- Taylor, R., Tobin, A. and Bray, C. (1996) Nucleotide sequence of an *Arabidopsis* cDNA At-phrII encoding a protein with high homology to the Class II CPD Photolyases present in higher eukaryotes. *Plant Physiol.* 112: 862-3.
- Teramura, A.H. and Sullivan, J.H. (1994). Effects of UV-B radiation on photosynthesis and growth of terrestrial plants. *Photosynth Res.* 39: 463-73.

- The Arabidopsis Genome Initiative.** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.
- Todo, T.** (1999). Functional diversity of the DNA photolyase/blue light receptor family. *Mut Res.* 434 (2): 89-97.
- Tuteja, N., Singh, M.B., Misra, M.K., Bhalla, P.L. and Tuteja, R.** (2001). Molecular mechanisms of DNA damage and repair: progress in plants. *Crit Rev Biochem Mol Biol.* 36 (4): 337-97.
- Vonarx, E.J., Mitchell, H.L., Karthikeyan, R., Chatterjee, I. and Kunz, B.A.** (1998). DNA repair in higher plants. *Mutat Res.* 400 (1-2): 187-200.
- Wood, R.D.** (1997). Nucleotide excision repair in mammalian cells. *J Biol Chem.* 272 (38): 23465-8.