Nucleotide excision repair I: from *E. coli* to yeast

JAN H.J. HOEIJMAKERS

Genetic information is constantly deteriorating, mainly as a consequence of the action of numerous genotoxic agents. In order to cope with this fundamental problem, all living organisms have acquired a complex network of DNA repair systems to safeguard their genetic integrity. Nucleotide excision repair (NER), one of the most important of these, is a complex multi-enzyme reaction that removes a remarkably wide range of lesions. This is the first of a series of two reviews on this repair process. Part I focuses on the main characteristics of the NER pathway in E. coli and yeast. Part II, to appear in the next issue of TIG, deals with NER in mammals and compares it with the process in yeast.

DNA is a remarkably versatile and stable carrier of genetic information in living cells, even though it is continually subject to undesired chemical alterations. Various types of radiation, such as ultraviolet (UV) light and X-rays, and numerous chemical agents induce a wide range of lesions that interfere with the proper functioning of DNA. As well as its immediate hampering effect on vital processes, notably transcription and replication, DNA damage may also give rise to mutations and chromosomal aberrations that lead to inborn defects, carcinogenesis and cell death. The time-dependent accumulation of DNA damage and mutations may even contribute to (cellular) ageing. To minimize such problems, evolution has equipped all living organisms with an intricate network of repair pathways. Together, these systems act as a kind of intranuclear 'immune system' that is able to recognize and eliminate many types of lesions.

One of the most important, and general, repair processes is the nucleotide excision repair (NER) pathway. This system deals with a strikingly diverse array of structurally unrelated lesions, including various UV-induced photoproducts [cyclobutane pyrimidine dimers (CPD) and 6–4 photoproducts], chemical adducts and certain types of crosslinks. In all organisms, NER consists of five steps: damage recognition, incision of the damaged strand on both sides of the lesion and at some distance from it, excision of the lesion-containing oligonucleotide, synthesis of new DNA using the undamaged strand as template, and ligation. The reaction is in principle error-free. In this review I shall describe the main features of the NER pathway in *Escherichia coli* – the organism in which the mechanism is understood in the most detail – and compare it with the equivalent process in yeast. For a comprehensive review on DNA repair in general, see Ref. 1.

NER in *E. coli*

Figure 1 outlines the core components involved in NER in *E. coli*, and their role in the mechanism. The main properties of the key proteins UvrA–D and of two auxiliary factors, Phr and Mfd, are summarized in Table 1.

The NER reaction is thought to start when two molecules of UvrA dimerize in the presence of ATP and complex with one molecule of UvrB (Ref. 2). The heterotrimer binds DNA; its weak 5'→3' DNA helicase activity may then allow it to translocate along one of the strands of the helix, scanning it for distorting lesions[3]. If such an injury is encountered, translocation stops and UvrB is hooked onto the DNA, inducing a specific DNA conformation that includes a kink and a locally denatured region. The UvrA molecules dissociate, allowing UvrC to bind[2]. The UvrBC complex then makes two incisions in the damaged strand: one, located eight phosphodiester bonds 5' of the injury, is probably catalysed by UvrC. The other incision, five bonds 3' of the lesion, is presumably made by UvrB (Refs 2, 4). The concerted action of UvrD (which is also a helicase) and DNA polymerase I accomplishes the release of the 12–13-mer containing the damaged site, turnover of bound UvrB and UvrC, and synthesis of an undamaged copy of the sequence that has been removed[2]. Finally, DNA ligase seals the nicks in the repaired strand.

The crucial step, the detection of a lesion by discrimination between abnormal and normal DNA structures, is carried out by the cooperative action of just

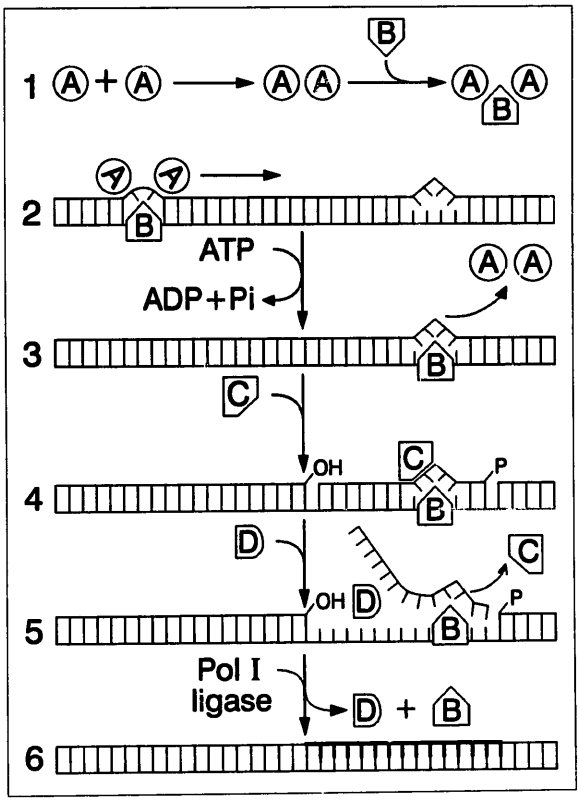

FIG 1

Model for the reaction mechanism of *E. coli* nucleotide excision repair (see text for a full explanation). A, B, C and D, molecules of UvrA, B, C and D respectively; Pol I, DNA polymerase I; ligase, DNA ligase. DNA strands are represented by notched lines. Sequential stages in NER are numbered.

TABLE 1. Properties of the major NER proteins of *E. coli*[a]

Protein	No. of amino acids	No. of molecules per cell[b]	Properties of protein
UvrA	940	25 (250)	Binds UV-irradiated dsDNA Can dimerize and complex with UvrB Two NTP binding sites Two zinc finger domains
UvrB	673	500 (2000)	Binds $UvrA_2$ 5'→3' DNA helicase activity as a $UvrA_2B$ complex One NTP binding site, probably makes 3' incision
UvrC	610	10	Binds ss and dsDNA Endonuclease activity in complex with UvrB, probably makes 5' incision
UvrD (helicase II)	720	3000 (4500)	3'→5' DNA or DNA–RNA helicase Binds NTP and DNA
Phr	471	10–20	Photoreactivates CPD Stimulates repair by UvrABC Binds $FADH_2$, pterin and UV-irradiated dsDNA
Mfd	1148	?	Confers specificity for transcribed strand on *in vitro* repair by UvrABC Displaces RNA polymerase Interacts with UvrA

[a]DNA polymerase I and DNA ligase are not included.
[b]The number of molecules per cell after SOS induction is shown in parentheses.

two proteins: UvrA and B. The A_2B complex is able to recognize an impressive spectrum of damaged structures ranging from thymine glycols to bulky chemical adducts and inter- and intrastrand crosslinks. Moreover, these lesions are usually present in trace amounts (concentrations of 1 in 10^6 nucleotides or less). It should be noted, however, that not all lesions are identified with the same efficiency, and some are substrates of more than one repair system.

The scheme shown in Fig. 1 is the core reaction of *E. coli* NER, which can be entirely mimicked *in vitro* using naked, damaged DNA as substrate and the purified proteins mentioned above. However, in the cell, additional factors are involved that modulate the system or permit it to act on chromatin. An interesting example is the strand-selective repair of active genes, a sub-pathway of NER that was first discovered in mammals by the pioneering work of Hanawalt and co-workers (reviewed in Refs 5, 6). It has been found *in vivo*[7], and for *E. coli* also *in vitro*[8], that the NER system removes transcription-blocking (CPD) lesions from the transcribed strand of an active gene faster than from the nontranscribed strand or from the genome overall. The system can apparently differentiate between the two strands of a gene and assigns a higher priority to eliminating DNA damage in more sensitive sites in the genome than to dealing with the same injury in less critical locations. In *E. coli* this sophisticated process requires at least one extra factor: the transcription–repair coupling factor, identified by Sancar and co-workers as the product of the *mfd* gene[8] (Table 1). It has recently been shown that *in vitro* the Mfd protein dissociates an RNA polymerase molecule stalled at a lesion and binds the UvrA subunit of the repair complex (C.P. Selby and A. Sancar, submitted). Some auxiliary components in this process have been identified, notably the photoreactivating enzyme photolyase. This enzyme itself monomerizes CPDs by direct photoreversal but, in addition, appears to facilitate the repair of these lesions by the NER pathway[9]. It is likely that more modulating factors will be discovered. Finally, NER in *E. coli* is under the control of the SOS response. This ingenious and versatile regulatory system, mediated by the *lexA–recA* regulon, permits a low constitutive level of expression of a number of NER genes under normal conditions and a strong induction when the bacterium is suddenly confronted with severe damage to its DNA (Table 1).

The *E. coli* NER system has recently been the subject of several excellent reviews[2,3,10,11], in which references to many of the original publications can be found. Although the basic molecular mechanism of NER in *E. coli* is considered a valid model for eukaryotes, the eukaryotic system is probably more complex.

NER in *Saccharomyces cerevisiae*

Although far less is known about excision repair in eukaryotes than in *E. coli*, it is clear that *S. cerevisiae* is a particularly relevant paradigm for NER in eukaryotes in general. Information about the yeast system has come chiefly from detailed analysis of excision-deficient mutants and cloned genes.

Yeast NER mutants

The collection of yeast NER mutants (Table 2) currently includes 11–12 members that comprise the *rad3* epistasis group. The list is certainly incomplete; it is likely that a significant number of NER genes have not been revealed by any mutant phenotype and remain hidden in the yeast genome. In fact, one of the yeast NER genes, *RAD25* (also designated *SSL2*), was not identified in the hunt for repair-deficient mutants. It was cloned on the basis of homology with a human NER gene, *ERCC3*, and a corresponding mutant was subsequently generated by gene disruption. The members of the *rad3* group display a varying degree of UV sensitivity, which is characteristic of NER deficiency (reviewed in Ref. 12). Some show, in addition, a pronounced sensitivity to crosslinking agents. As is observed for *E. coli*, the most sensitive mutants are totally deficient in NER. This applies to certain alleles of *rad1, 2, 3, 4, 10, 14* and probably *rad25*. The remaining mutants are partially deficient, presumably indicating that the corresponding genes are only implicated

in specific sub-pathways of excision repair or have an accessory function. It has been shown that *rad7* and *16* mutants, and to a lesser extent *rad24* mutants, are deficient in the repair of CPD lesions in the inactive *HMLα* locus of the yeast mating-type system; however, repair of the active *MATα* locus appears to be quite proficient[13]. It is not clear whether the repair process that is defective in these mutants is equivalent to the relatively slow and incomplete sub-pathway of NER that is involved in 'overall genome repair' in mammalian cells (see Part II). Mutants specifically impaired in the gene- or strand-specific repair of active genes (cf. the *mfd* mutant of *E. coli* discussed above) have not (yet) been identified in the *rad3* collection.

Some mutants in the *rad3* group also have defects in other aspects of DNA or RNA metabolism. For example, *rad1* and *rad10* mutants show deficiencies in a mitotic recombination pathway[14]. The *RAD3* and *RAD25* genes appear to be involved in a vital cellular process, since mutants carrying certain alleles of these genes are inviable[15]. As discussed below, recent findings suggest that, at least for *RAD25*, this essential process is concerned with one of the steps of gene expression. The involvement of several proteins in a number of distinct pathways of nucleic acid metabolism may indicate either that the individual polypeptides are multifunctional or that there is functional overlap between these pathways in the nucleus. RAD1 and RAD10 could, for example, participate in a multi-enzyme complex that has indispensable functions both in NER and in one of the mitotic recombination pathways. It is even conceivable that some of the subunits of such a complex might be directly required only for recombination. These proteins might appear to be needed for NER because they have to be physically present in the complex in order for it to be functional. Mutations in such proteins could still give rise to a repair-deficient phenotype, even though the protein did not have a direct catalytic role in NER. An explanation such as this may account, at least in part, for the large number of NER mutants found in yeast (and in mammals) and for the fact that there are subclasses of NER mutants that have the same pattern of additional impairments in other nucleic acid transactions.

Yeast NER genes and proteins

Almost all of the genes corresponding to the mutants in the current *rad3* epistasis group have been cloned, mostly by genetic complementation. The main properties of these genes and the encoded polypeptides are summarized in Table 2 (see Ref. 12 for an extensive review). A few general aspects deserve specific attention.

Gene regulation. So far there is no indication that the yeast genes are regulated by a common pathway analogous to the SOS response of *E. coli*. Most of the repair genes are very weakly expressed. Some (such

TABLE 2. Major properties of NER genes and mutants of *S. cerevisiae*

Gene	UV-sensitivity of mutant[a]	NER of mutant[b] Active *MATα*	NER of mutant[b] Inactive *HMLα*	Predicted no. of amino acids in protein	Homologous genes Human	Homologous genes *S. pombe*	Properties of protein[d]
RAD1	++	−	−	1100		*rad16*$^+$	Involved in recombination Chromatin binding? endonuclease?
RAD2	++	−	−	1031	*ERCC5*	*rad13*$^+$	Transcription inducible by UV
RAD3	++	−	−	778	*ERCC2*	*rad15*$^+$	Essential function 5'→3' DNA helicase chromatin binding?
RAD4	++	−	−	754	*XPCC*?		Chromatin binding? DNA binding?
RAD7	+	±	−	565			Transcription inducible by UV Membrane associated?
RAD10	++	−/±[c]	−/±[c]	210	*ERCC1*	*swi10*$^+$	Involved in recombination Binds ssDNA, catalyses DNA renaturation
RAD14	++	−/±[c]	−	247	*XPAC*		DNA binding zinc finger
RAD16	+	±	−	790			DNA binding zinc finger? DNA helicase? (Protein has partial homology to RAD5, RAD54 and ERCC6)
RAD23	+						
RAD24	+	+	−/±				
RAD25	+/++			843	*ERCC3*	*ERCC3*^{Sp}	Essential function DNA helicase?
MMS19	+						

[a]The greater the number of plus signs, the more sensitive the mutant.
[b]Minus indicates no NER, plus indicates NER activity.
[c]The strain tested may not be a completely null mutant.
[d]Question marks indicate characteristics inferred on the basis of the predicted amino acid sequences of the proteins.

as *RAD2*, 7 and *23*) appear to be UV-inducible, at least at the level of transcription, but the majority are not. The functional significance of this induction is not yet clear.

Homology. The (postulated) functional properties of most of the gene products (Table 2) have been inferred mainly from homology with known functional domains of other polypeptides. Only the RAD1, 3, 10 and 14 proteins have actually been purified and to some extent biochemically characterized. As indicated in Table 2 and discussed more extensively in Part II, there is striking overall homology with the products of human NER genes. The corresponding proteins are so similar that they almost certainly have the same functions; indeed it seems beyond any doubt that the entire NER system of eukaryotes has a common evolutionary origin. Remarkably, there is no significant resemblance to any of the *E. coli* Uvr proteins, suggesting that prokaryotic and eukaryotic NER systems have diverged substantially.

Involvement in other cellular processes. As mentioned above, RAD1 and RAD10 also have a role in one of the mitotic recombination pathways. These proteins form a strong complex both *in vivo* and *in vitro*[16,17]. Site-directed mutagenesis of *RAD1* has identified domains in the protein that are specifically implicated either in repair or in recombination, indicating that *RAD1* has distinct roles in each of these processes (S. Prakash, cited in Ref. 18). Recent studies have suggested that the function of the protein in recombination is to remove nonhomologous sequences from the 3' end of recombining DNA molecules[19]. Purified RAD10 protein has been shown to bind single-stranded DNA and to accelerate DNA renaturation[20]. The latter function can easily be envisaged as part of the recombination process in which RAD10 is known to participate, but it is less obvious how this enzymatic activity fits into the molecular mechanism of excision repair.

RAD3 and RAD25 have an additional role in a process that is essential for cell viability. RAD3 is a 5'→3' helicase that is able to unwind both double-stranded DNA and DNA–RNA hybrids[21]. The helicase activity is indispensable for NER but not for the vital function of the protein. RAD25 is also suspected to be a helicase, because its sequence features seven consecutive domains that are shared between two superfamilies of DNA and RNA helicases[22]; RAD3 also belongs to one of these families. Evidence of helicase activity was recently obtained for the human counterpart of RAD25, ERCC3 (see Part II). In contrast to RAD3, the proposed helicase activity of RAD25 is indispensable for the vital function of the protein[22]. A clue to the nature of the essential RAD25 function came unexpectedly from the recent work of Gulyas and Donahue[23]. These investigators searched for suppressors of the translational block that is imposed on a *HIS4* mRNA by a strong artificial hairpin in the 5' UTR. One of these *ssl* (suppressor of stem–loop) mutations, *ssl2*, appeared to be in the *RAD25* gene. This suggests that the RAD25 (SSL2) gene product is somehow involved in gene expression at the level of RNA. In view of the striking parallels between RAD25 and RAD3 it is possible that RAD3 also participates in the same aspect of RNA metabolism. Intriguingly, the *SSL1* gene was also recently found to encode an essential function associated with gene expression; some mutant alleles make yeast hypersensitive to UV radiation, suggesting that *SSL1* may be a previously undiscovered member of the *rad3* group[24]. The predicted 52 kDa SSL1 protein contains several putative zinc fingers. This protein may interact with RAD25 and perhaps also with RAD3 during both NER and some stage of gene expression.

Damage recognition. The deduced amino acid sequence of the RAD14 protein contains a putative zinc finger region that is postulated to have a DNA binding function[25]. The yeast protein was recently found to bind both single- and double-stranded DNA, with a particularly strong affinity for UV-irradiated DNA (S. Prakash, pers. commun.). RAD14 may therefore be (one of) the damage recognition protein(s) of eukaryotic NER. Further characterization of its lesion specificity is required to test this proposition.

NER of active and inactive genes. As discussed above, both *rad7* and *rad16* mutants show specific defects in the removal of CPD lesions from the inactive *HMLα* locus. While the primary amino acid sequence of the RAD7 protein does not reveal any clue to its function, the predicted amino acid sequence encoded by the recently cloned *RAD16* gene features a special type of DNA-binding zinc finger as well as seven helicase motifs (Ref. 26 and references therein). This strongly suggests that RAD16 is also a DNA helicase. The significance of this presumption is underscored by the finding that the 'helicase' part of RAD16 displays extraordinary homology to a similar domain in proteins that are implicated in various repair pathways and in transcriptional regulation. These include the yeast RAD54 protein (implicated in recombination repair), RAD5 (involved in postreplication repair) and the human ERCC6 protein (associated with preferential repair of active genes; see Part II). Thus every major multi-enzyme repair pathway in yeast is equipped with at least one member of this subfamily. Although the sequence homology is strikingly limited to the 'helicase' segment, the overall level of identity in this portion is much higher than among other members of the helicase superfamily. Thus these observations define a new subfamily of proteins that may share a specific type of helicase function. Recent evidence suggests that the transcription activators in this family exert their function via effects on chromatin[27]. In this light, RAD16 could be involved in opening the closed chromatin structure of inactive regions in the genome, such as *HMLα*, to allow repair to take place.

NER in *Schizosaccharomyces pombe*

In the last few years another, distantly related, yeast species, *Schizosaccharomyces pombe*, has enjoyed increasing popularity in the exploration of eukaryotic NER. Although many mutants have been generated in this organism, they have not been characterized as thoroughly as those in *S. cerevisiae*. Several cloned repair genes appear to be homologous to members of the *S. cerevisiae RAD3* series and in some cases also to human NER genes (see Table 2; Ref. 28). Sequence comparison of such distantly related versions of the same genes may identify important functional domains and may yield valuable information for

the isolation of homologs from mammals. In addition, new NER genes may even be discovered.

Concluding remarks

In the light of all the available information, how closely related are the *E. coli* and yeast NER systems, and is the prokaryotic process a valid model for the equivalent pathway in yeast and other eukaryotes? Several considerations support the notion that the fundamental features of the reaction mechanism of *E. coli* and eukaryotes are comparable.

(1) The systems share the basic steps of NER: damage recognition, dual incision (recently also demonstrated in mammals[29]), excision, repair synthesis and ligation.

(2) The substrate repertoire is essentially similar, suggesting a similar mechanism of damage recognition.

(3) Both systems entail preferential repair of the transcribed strand.

Despite these similarities, two remarkable differences can be noted: the number of NER genes is significantly higher in eukaryotes than in *E. coli*, and so far there appears to be no significant overall sequence homology between the NER proteins of prokaryotes and eukaryotes. It is clear that the main components of the bacterial NER pathway have been identified, since the entire reaction can be reconstructed *in vitro* using the purified proteins. In contrast, it is likely that a substantial fraction of the factors involved in the yeast system is still unidentified. It is possible, therefore, that the true yeast homologs of UvrA–D have not yet been isolated. In that case, we would have to account for the function of the many eukaryotic NER genes identified so far that appear to have no *E. coli* counterpart. This would imply an even greater level of complexity for the eukaryotic process. An alternative or additional explanation might be that the NER proteins have diverged to such an extent that no homology can be detected across the eukaryote–prokaryote divide. This could mean in turn that the underlying molecular mechanisms may have diverged substantially. An observation consistent with this idea is that the excised lesion-containing fragment in mammals is much longer than that in *E. coli*. It also seems likely that the fundamental differences in chromatin structure between prokaryotes and eukaryotes may dictate differences in their excision repair machinery.

Acknowledgements

I apologize to those whose important primary papers were not cited due to space restrictions. I thank all my colleagues in the DNA repair group of the Medical Genetics Centre for valuable discussions, and other colleagues in the field for sharing unpublished results, and R. Boucke for typing the manuscript. Research in our group is supported by the Netherlands Organization for the Advancement of Pure Science, through the Foundation of Medical Scientific Research, the Dutch Cancer Society and Euratom.

References

1 Friedberg, E.C. (1985) *DNA Repair*, Freeman
2 Lin, J-J. and Sancar, A. (1992) *Mol. Microbiol.* 6, 2219–2224
3 Grossman, L. and Yeung, A.T. (1990) *Photochem. Photobiol.* 51, 749–755
4 Sancar, A. and Rupp, W.D. (1983) *Cell* 33, 249–260
5 Hanawalt, P.C. (1989) *Genome* 31, 605–611
6 Bohr, V.A. (1991) *Carcinogenesis* 12, 1983–1992
7 Mellon, I. and Hanawalt, P.C. (1989) *Nature* 342, 95–98
8 Selby, C.P., Witkin, E.M. and Sancar, A. (1991) *Proc. Natl Acad. Sci. USA* 88, 11574–11578
9 Sancar, G.B. and Smith, F.W. (1989) *Mol. Cell. Biol.* 9, 4767–4776
10 van Houten, B. (1990) *Microbiol. Rev.* 54, 18–51
11 Sancar, A. and Sancar, G. (1988) *Annu. Rev. Biochem.* 57, 29–67
12 Friedberg, E.C. (1988) *Microbiol. Rev.* 52, 70–102
13 Terleth, C. *et al.* (1990) *Mol. Cell. Biol.* 10, 4678–4684
14 Schiestl, R.H. and Prakash, S. (1990) *Mol. Cell. Biol.* 10, 2485–2491
15 Naumovski, L. and Friedberg, E.C. (1983) *Proc. Natl Acad. Sci. USA* 80, 4818–4821
16 Bardwell, L., Cooper, A.J. and Friedberg, E.C. (1992) *Mol. Cell. Biol.* 12, 3041–3049
17 Bailly, V. *et al.* (1992) *Proc. Natl Acad. Sci. USA* 89, 8273–8277
18 Lehmann, A. *et al.* (1992) *Mutat. Res.* 273, 1–28
19 Fishman-Lobell, J. and Haber, J.E. (1992) *Science* 258, 480–484
20 Sung, P., Prakash, L. and Prakash, S. (1992) *Nature* 355, 743–745
21 Bailly, V., Sung, P., Prakash, L. and Prakash, S. (1991) *Proc. Natl Acad. Sci. USA* 88, 9712–9716
22 Park, E. *et al.* (1992) *Proc. Natl Acad. Sci. USA* 89, 11416–11420
23 Gulyas, K.D. and Donahue, T.F. (1992) *Cell* 69, 1031–1042
24 Yoon, H., Miller, S.P., Pabich, E.K. and Donahue, T.F. (1992) *Genes Dev.* 6, 2463–2477
25 Bankmann, M., Prakash, L. and Prakash, S. (1992) *Nature* 355, 555–558
26 Bang, D.D. *et al.* (1992) *Nucleic Acids Res.* 20, 3925–3931
27 Winston, F. and Carlson, M. (1992) *Trends Genet.* 8, 387–391
28 Lehmann, A.R., Carr, A.M., Watts, F.Z. and Murray, J.M. (1991) *Mutat. Res.* 250, 205–210
29 Huang, J-C., Svoboda, D.L., Reardon, J.T. and Sancar, A. (1992) *Proc. Natl Acad. Sci. USA* 89, 3664–3668

J.H.J. HOEIJMAKERS IS IN THE DEPARTMENT OF CELL BIOLOGY AND GENETICS, MEDICAL GENETICS CENTRE, ERASMUS UNIVERSITY, PO BOX 1738, 3000DR ROTTERDAM, THE NETHERLANDS.