Characterization and analysis of Chinese hamster ovary cell ERCC1 mutant alleles

Rhonda L. Rolig, Megan P. Lowery, Gerald M. Adair and Rodney S. Nairn

Department of Carcinogenesis, The University of Texas M.D. Anderson Cancer Center, Science Park-Research Division, PO Box 389, Smithville, TX 78957, USA

We isolated and characterized the ERCC1 coding sequence from three Chinese hamster ovary (CHO) parental (CHO-AA8, CHO-AT3-2 and CHO-9) and 10 ERCC1 mutant cell lines. Two general classes of mutations were observed: two mutant cell lines exhibited nucleotide additions or deletions to produce frameshift mutations and seven mutant cell lines exhibited point mutations that resulted in transitions or transversions, including nonsense mutations and mutations that generated intron/exon splicing errors. One mutant (UV201) which had been provisionally assigned to ERCC1 complementation group 1 (CG1) had no detectable mutation in its coding sequence. Of the nine ERCC1 mutant alleles characterized two mutations were identified in the Xpa binding region of the Ercc1 protein; no mutations were found in the N-terminal portion of the Ercc1 protein. Results of Northern hybridization analysis showed that the relative levels of ERCC1 mRNA differed significantly both among the parental cell lines and among the mutant cell lines derived from each parental cell line. Western analysis with a CHO Ercc1-specific antibody detected Ercc1 protein in each of the parental cell lines and also in UV201. The marked reduction in Ercc1 protein levels observed in all the other mutants examined supports the hypothesis that ERCC1 mutations may destabilize this polypeptide.

Introduction

Nucleotide excision repair (NER) is a complex multistep process involving a large number of genes. In all cells NER functions to remove bulky, helix-distorting DNA lesions and other types of DNA damage that, if left unrepaired, can cause cell death or mutation or promote cellular transformation. Defects in NER have been directly linked to the human disease syndromes xeroderma pigmentosum (XP), Cockayne syndrome and trichothiodystrophia (Lindhahl et al., 1997). Studies conducted in bacteria, yeast and mammalian cells have shown that the processes of NER and transcription are intimately linked (Bohr et al., 1985; Mellon and Hanawalt, 1989; Sweder et al., 1988, 1990; Fishman-Lobell and Haber, 1992). Two general classes of mutations were observed: two mutant cell lines exhibited nucleotide additions or deletions to produce frameshift mutations and seven mutant cell lines exhibited point mutations that resulted in transitions or transversions, including nonsense mutations and mutations that generated intron/exon splicing errors. One mutant (UV201) which had been provisionally assigned to ERCC1 complementation group 1 (CG1) had no detectable mutation in its coding sequence. Of the nine ERCC1 mutant alleles characterized two mutations were identified in the Xpa binding region of the Ercc1 protein; no mutations were found in the N-terminal portion of the Ercc1 protein. Results of Northern hybridization analysis showed that the relative levels of ERCC1 mRNA differed significantly both among the parental cell lines and among the mutant cell lines derived from each parental cell line. Western analysis with a CHO Ercc1-specific antibody detected Ercc1 protein in each of the parental cell lines and also in UV201. The marked reduction in Ercc1 protein levels observed in all the other mutants examined supports the hypothesis that ERCC1 mutations may destabilize this polypeptide.

Investigation of NER in mammalian cells has been based largely on studies using cell lines established from patients with XP and other disease syndromes and the large collection of somatic cell mutants isolated from Chinese hamster and other rodent cell lines (reviewed in Collins, 1993). In rodent cells 11 genetic complementation groups representing mutants affected in NER have been reported (Weeda et al., 1993); many, but not all, of these complementation groups overlap with XP complementation classes. Many human genes involved in NER have been cloned based on their complementation of UV-sensitive phenotypes of CHO mutants and have thus been called excision repair cross-complementing (ERCC) genes (Thompson and Bootsma, 1988). While mutants from each of the CHO NER complementation groups characteristically exhibit a UV-hypersensitive phenotype, the CHO ERCC1 and ERCC4/XPF complementation groups are also uniquely defined by hypersensitivity to agents causing DNA interstrand cross-links (Hoy et al., 1985; Busch et al., 1989, 1997; Collins, 1993). Some mutants from these two groups have also been shown to be sensitive to X- or γ-irradiation under hypoxic (but not oxic) conditions (Murray and Rosenberg, 1995); although the mechanism underlying this sensitivity to hypoxic ionizing radiation is not well understood, it has been shown to be independent of NER (Murray and Rosenberg, 1996).

The Ercc1 polypeptide functions in a complex with the Ercc4/Xpf protein as a structure-specific 5′ double-strand–3′ single-strand endonuclease (Mu et al., 1996; Sijbers et al., 1996a; Blessho et al., 1997). This activity is critical for the dual incision step of NER, during which an oligonucleotide fragment containing the DNA damage is excised (Huang et al., 1992; Moggs et al., 1996); the ERCC5/XPG gene product catalyzes nicking on the 3′-side of DNA damage (O’Donovan et al., 1994). Biochemical evidence suggests that 3′ nicking precedes 5′ nicking in the dual incision step (Mu et al., 1996). Experiments with ERCC1 knockout mice suggest that ERCC1 may play a role in embryonic development and cellular proliferation in addition to its function in NER (McWhir et al., 1993; Weeda et al., 1997). A recent study in an ERCC1 knockout CHO cell line has provided evidence suggesting that ERCC1 plays a role in mitotic recombination and perhaps in maintaining genomic stability in mammalian cells (Sargent et al., 1997). The pronounced sensitivities to DNA crosslinking agents observed in the ERCC1 and ERCC4/XPF complementation groups are consistent with involvement of these two NER genes in genetic recombination, since DNA–DNA interstrand cross-links may be subject to repair by recombinational mechanisms (Cole, 1973). The pronounced sensitivities to DNA crosslinking agents observed in the ERCC1 and ERCC4/XPF complementation groups are consistent with involvement of these two NER genes in genetic recombination, since DNA–DNA interstrand cross-links may be subject to repair by recombinational mechanisms (Cole, 1973).
and recombination (Thompson et al., 1987). Most of these studies have used four CHO ERCC1 mutants derived from three different CHO parental cell lines: UVV20 (derived from CHO-AA8 cells; Thompson et al., 1980), 43-3B (from CHO-9 cells; Wood and Burki, 1982) and either UVL9 or UVL10 (derived from CHO-AT3-2 cells; Clarkson et al., 1983). We were interested in investigating: (i) the precise nature of the particular ERCC1 mutation in each of these mutant cell lines; (ii) whether there are any differences among the ERCC1 coding sequences in the three parental CHO cell lines. We were also interested in characterizing the ERCC1 mutant alleles from a larger panel of CHO ERCC1 mutants (listed in Table I) produced by conventional mutagenesis with a variety of different physical and chemical mutagens. Some of these mutants exhibit heterogeneous phenotypes with respect to NER proficiency or cytotoxic and mutational responses (Thompson et al., 1981; Busch et al., 1989, 1994, 1997; Collins, 1993). Location of the mutations in these cell lines could potentially identify regions of functional importance in the Erccl protein.

In human and mouse, the translational start site for the Erccl protein is located in exon 2 of the ERCC1 cDNA sequence (van Duin et al., 1986, 1988b). Therefore, in order to identify mutations in the ERCC1 coding sequence in CHO cell mutants we used reverse transcription–PCR (RT-PCR) with parental RNA to amplify sequence corresponding to exons 2–10 of the CHO ERCC1 cDNA. Using this approach we have identified mutations in the ERCC1 coding sequence in nine of the 10 CHO ERCC1 mutant cell lines examined and have shown that there are no differences in the ERCC1 coding sequence among the three parental CHO cell lines. We also observed a surprising degree of variability in ERCC1 gene expression among both parental and mutant cell lines as determined by steady-state mRNA levels. In every mutant in which we identified a coding sequence change, protein analysis by Western blotting showed very little or no protein expression, consistent with a previous suggestion (Biggerstaff et al., 1997) that mutations in either polypeptide may destabilize the Erccl–XpF complex and result in increased degradation of its components.

### Materials and methods

#### Cell lines and culture conditions

The three CHO parental and 10 ERCC1 mutant cell lines characterized in this study are listed in Table I.

<table>
<thead>
<tr>
<th>Parental cell line</th>
<th>Mutant cell line</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHO-AA8</td>
<td>UV4 (Hoy et al., 1985)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV20 (Thompson et al., 1980, 1981)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV60 (Busch et al., 1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV71 (D. B. Busch, unpublished)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV121 (D. B. Busch, unpublished)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV201 (Busch et al., 1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UV203 (Busch et al., 1997)</td>
<td></td>
</tr>
<tr>
<td>CHO-AT3-2</td>
<td>UVL9 (Clarkson et al., 1983; Sage et al., 1996)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UVL10 (Clarkson et al., 1983)</td>
<td></td>
</tr>
<tr>
<td>CHO-9</td>
<td>43-3B (Wood and Burki, 1982)</td>
<td></td>
</tr>
</tbody>
</table>

1980; Clarkson et al., 1993) consistent with a previous suggestion (Biggerstaff et al., 1997) by Western blotting showed very little or no protein expression, determined by steady-state mRNA levels. In every mutant in the three parental CHO cell lines, we also have shown that there are no differences in the coding sequence among the three parental CHO cell lines. We have identified mutations in the coding sequence in all the mutant cell lines examined and have identified mutations in the coding sequence in CHO parental cell lines to amplify sequence corresponding to exons 2–10 of the CHO ERCC1 cDNA. To identify mutations in the ERCC1 coding sequence in CHO mutants we used reverse transcription-PCR (RT-PCR) and subcloned according to the protocol recommended by the manufacturer of the TA Cloning Kit (In Vitrogen, San Diego, CA).

#### Northern analysis

Polyadenylated RNA was isolated from each parental and ERCC1 mutant cell line using a FastTrack 2.0 Kit (In Vitrogen) with the modifications described in Rolig et al. (1997). RNA samples were electrophoresed in 0.66 M formaldehyde-agarose gels, blotted and hybridized essentially as described by Maniatis et al. (1982), but with the addition of 50 μg/ml polyadenylic acid to the hybridization solution. To quantify mRNA a VISAGE 60 camera (Biolmage, Ann Arbor, MI) was used to capture autoradiograph image data; analysis was performed utilizing the Biolmage Whole Band Analyzer software (Biolmage).

#### Western analysis

The anti-Erccl antibody used in this study, RW-017, was a generous gift of Dr Rick Wood (Imperial Cancer Research Fund, Clare Hall Laboratories, UK). This antibody, described in Biggerstaff et al. (1993), specifically recognizes the conserved C-terminal region of the human and CHO Erccl protein. For each of the ERCC1 parental and mutant cell lines nuclear extracts (40 μg) were prepared essentially as described (Dignam et al., 1983), then electrophoresed in 10% SDS-polyacrylamide gels and transferred by electrophoresing to nitrocellulose membranes. The membranes were blocked with 5% non-fat dry milk in Tris-buffered saline containing 0.1% Tween 20 (blocking solution) and incubated with RW-017 primary antibody. The membrane was then incubated with horseradish peroxidase-conjugated goat anti-rabbit IgG antibody (Sigma, St Louis, MO) in blocking solution. Secondary antibody was detected by enhanced chemiluminescence following the manufacturer’s protocol (Amersham, Arlington Heights, IL).

#### Genomic PCR

DNA was isolated from cell monolayers as described previously (Rolig et al., 1997). To amplify ERCC1 coding exons, exons 2–10 were amplified individually from each cell line using primers designed using CHO ERCC1 genomic sequence (unpublished data). The primer sets, listed in Table I, were used for Touchdown PCR (Don et al., 1991), performed using 200 ng genomic DNA in a reaction mixture containing 2.5 U Taq polymerase (Promega, Madison, WI) associated with TouchStart Antibody (Clontech, Palo Alto, CA) according to the protocol supplied by Clontech. Thirty five total PCR cycles were performed with initial and final annealing temperatures of 65 and 55°C respectively.

#### DNA sequence analysis

Sequencing of cDNAs and genomic exons was performed with 0.1 μl each PCR reaction (total reaction volume 50 μl) using the ThermoSequenase Kit (Amersham, Arlington Heights, IL) essentially according to the manufacturer’s protocol for dideoxynucleosine sequencing. The primers used to sequence the CHO ERCC1 exons are listed in Table I. Nucleotide and amino acid sequence data were analyzed using the MacVector and AssemblyLIGN sequence analysis programs (IBI, New Haven, CT) and the Genbank database.
Table II. Genomic PCR and sequencing primers

<table>
<thead>
<tr>
<th>Exon Primer</th>
<th>PCR forward</th>
<th>PCR reverse</th>
<th>Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5'-ATATGACAGTGGGCGGGAATG-3'</td>
<td>5'-ACACTGACCCACCCCACATCT-3'</td>
<td>5'-TTTCCAAAACACTGCAAAC-3'</td>
</tr>
<tr>
<td>3</td>
<td>5'-GGATGAGTGGGAGGAATG-3'</td>
<td>5'-AGAAAGAAGTGGCAGAAGAAT-3'</td>
<td>5'-CCGCGGCTTCTTATCATGGTGTA-3'</td>
</tr>
<tr>
<td>4</td>
<td>5'-CGGCCTTCTTACATGTCGTT-3'</td>
<td>5'-ACGCGTCTTCTTATCATGGTGTA-3'</td>
<td>5'-CCGCGGCTTCTTATCATGGTGTA-3'</td>
</tr>
<tr>
<td>5</td>
<td>5'-GGATGAGTGGGAGGAATG-3'</td>
<td>5'-AGAAAGAAGTGGCAGAAGAAT-3'</td>
<td>5'-CCGCGGCTTCTTATCATGGTGTA-3'</td>
</tr>
<tr>
<td>6</td>
<td>5'-ACACTGACCCACCCCACATCT-3'</td>
<td>5'-TTTCCAAAACACTGCAAAC-3'</td>
<td>5'-TTTCCAAAACACTGCAAAC-3'</td>
</tr>
<tr>
<td>7</td>
<td>5'-GGATGAGTGGGAGGAATG-3'</td>
<td>5'-AGAAAGAAGTGGCAGAAGAAT-3'</td>
<td>5'-CCGCGGCTTCTTATCATGGTGTA-3'</td>
</tr>
<tr>
<td>8</td>
<td>5'-CGGCCTTCTTACATGTCGTT-3'</td>
<td>5'-ACGCGTCTTCTTATCATGGTGTA-3'</td>
<td>5'-CCGCGGCTTCTTATCATGGTGTA-3'</td>
</tr>
<tr>
<td>9</td>
<td>5'-GGATGAGTGGGAGGAATG-3'</td>
<td>5'-AGAAAGAAGTGGCAGAAGAAT-3'</td>
<td>5'-CCGCGGCTTCTTATCATGGTGTA-3'</td>
</tr>
<tr>
<td>10</td>
<td>5'-ACACTGACCCACCCCACATCT-3'</td>
<td>5'-TTTCCAAAACACTGCAAAC-3'</td>
<td>5'-TTTCCAAAACACTGCAAAC-3'</td>
</tr>
</tbody>
</table>

Results

Expression of ERCC1 in CHO mutants

A 344 bp fragment containing the C-terminal region of the CHO ERCC1 cDNA was generated by RT-PCR and used as a probe for Northern analysis of polyadenylated mRNA (Figure 1). Figure 1A shows a Northern blot of parental and mutant cell lines hybridized with the 344 bp ERCC1-specific probe. To control for RNA loading this membrane was stripped and re-hybridized with a glyceraldehyde 3-phosphate dehydrogenase (GAPDH) probe (Figure 1B). The CHO parental cell lines CHO-AT3-2 (lane 2), CHO-AA8 (lane 5) and CHO-9 (lane 13) express different levels of ERCC1 mRNA. Interestingly, the CHO-AT3-2 cell line exhibits a signal approximately one third that observed for cell lines CHO-AA8 and CHO-9. Therefore, expression in each CHO ERCC1 mutant cell line was normalized relative to expression in its parental cell line (Table III). The CHO-AT3-2-derived cell lines, UVL9 (lane 3) and UVL10 (lane 4), express ERCC1 mRNA levels of 85 and 67% respectively relative to CHO-AT3-2. The CHO-AA8-derived cell lines exhibit a wider range of relative ERCC1 mRNA expression: UV4, 73%; UV20, 15%; UV71, 157%; UV121, 11%; UV201, 37%; UV203, 30%; UV60, 71% (lanes 6–12). One mutant cell line, 43-3B, derived from CHO-9, exhibited an ERCC1 mRNA level equivalent to its parent (lane 14).

To analyze the effect of ERCC1 mutation on protein expression Western analysis was performed on nuclear extracts using an antibody which specifically recognizes the Erccl C-terminal region (Biggerstaff et al., 1993). The CHO Erccl protein migrates at ~38 kDa (Biggerstaff et al., 1993). The immunoblots (Figure 2) show that very little or no protein was detected in any of the 10 ERCC1 mutant cell lines examined (lanes 3, 5–7, 10–12, 14 and 15) or in an ERCC1 knock-out cell line derived from CHO-AT3-2 as described in Rolig et al. (1997), which was included as a negative control (lane 4). In contrast, Erccl protein was detected in each of the parental cell lines (lanes 2, 8 and 9) and also in one ERCC1 mutant cell line, UV201 (lane 13).

Structure of the CHO ERCC1 coding sequence

The CHO ERCC1 cDNA sequence was obtained by nucleotide sequencing of RT-PCR and genomic PCR products generated using primers designed based on CHO genomic ERCC1 sequence (our unpublished data). An identical sequence was obtained for each of the parental cell lines, CHO-AT3-2, CHO-AA8 and CHO-9. Analysis of the CHO ERCC1 cDNA sequence identified a 779 bp open reading frame extending from the ATG/Met in exon 2. Translation of this reading frame yielded a predicted 293 amino acid protein sequence, which is shown compared with the human and mouse ERCC1 homologs in Figure 3. As expected, the CHO Erccl amino acid sequence is highly conserved when compared with human and mouse ERCC1 homologs. In Figure 3 light gray shading denotes amino acid identity observed among the three homologs, while non-identical amino acids are highlighted in black. Overall, when compared with human and mouse CHO amino acid identity is 83 and 89% respectively. Allowing conservative substitutions (where I = L = M = V, A = G = P = S = T, F = W = Y, D = E = N = Q and H = K = R) the overall amino acid homology is 90.5 and 93% respectively. Like the mouse Erccl polypeptide, the amino acid sequence corresponding to the middle and C-terminal regions of the CHO Erccl protein (downstream from human Ala91) is the most highly conserved. In this region CHO Erccl amino acid identity is 92 and 97% respectively when compared with human and...
Table III. CHO ERCCI mutations

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Mutagen</th>
<th>Sequence change</th>
<th>Polypeptide change</th>
<th>mRNA levela</th>
<th>Protein expressionb</th>
</tr>
</thead>
<tbody>
<tr>
<td>UV40</td>
<td>ICR170</td>
<td>C addition</td>
<td>Frameshift</td>
<td>0.73</td>
<td>−</td>
</tr>
<tr>
<td>UV20</td>
<td>EMS</td>
<td>G→A</td>
<td>Ser→Asn, splicing error</td>
<td>0.15</td>
<td>−</td>
</tr>
<tr>
<td>UV60</td>
<td>EMS × 4</td>
<td>4 bp deletion</td>
<td>Frameshift</td>
<td>0.71</td>
<td>−</td>
</tr>
<tr>
<td>UV71</td>
<td>UV</td>
<td>T→C</td>
<td>Leu→Pro</td>
<td>1.57</td>
<td>−</td>
</tr>
<tr>
<td>UV121</td>
<td>MNG</td>
<td>C→T</td>
<td>Amber stop</td>
<td>0.11</td>
<td>−</td>
</tr>
<tr>
<td>UV201</td>
<td>EMS</td>
<td>None</td>
<td>None</td>
<td>0.37</td>
<td>−</td>
</tr>
<tr>
<td>UV203</td>
<td>EMS</td>
<td>C→T</td>
<td>Amber stop</td>
<td>0.30</td>
<td>−</td>
</tr>
<tr>
<td>UV9L9</td>
<td>EMS</td>
<td>G→A</td>
<td>Splicing error</td>
<td>0.85</td>
<td>−</td>
</tr>
<tr>
<td>UV1L0</td>
<td>UV</td>
<td>A→T</td>
<td>Amber stop, splicing error</td>
<td>0.67</td>
<td>−</td>
</tr>
<tr>
<td>43-3B</td>
<td>ENU</td>
<td>T→A</td>
<td>Val→Glu</td>
<td>1.00</td>
<td>−</td>
</tr>
</tbody>
</table>

*aRelative to parental cell line.

bWithin detectable limits; all parental cell lines showed expression.

Fig. 2. Western analysis of nuclear extracts obtained from CHO ERCCI parental and mutant cell lines. Protein samples from mutant and parental CHO cell lines as indicated, isolated as described in Materials and methods and probed with RW-017 antibody, reactive with CHO cell ERCC1 polypeptide

From Table II, mutations identified in genomic DNA were closely related to mutations identified in the pre-mRNA sequence as shown in Figure 2A. In all cases, overlapping 5' and 3' subclones were generated and sequenced. When RT-PCR products were analyzed by agarose gel electrophoresis prior to subcloning into M13 vectors, a V3-11 and a V3-21 subclone included a nucleotide (n) mutation in the exon 3-5' flanking region ERCCI allele upstream of the initiation codon. When the subcloned PCR products were sequenced, the n mutation was also seen at the initiation codon. All n mutations were confirmed by nucleotide sequence analysis. This provided independent confirmation of RT-PCR results and eliminated the possibility that any mutation identified by RT-PCR was the result of polymerase error. In cases where a mutation was identified by genomic PCR, a second, independent, PCR reaction was performed using genomic DNA and amplifiers were sequenced to provide confirmation. Only one ERCCI mutation was identified in each mutant cell line, except UV201. These mutations were localized to particular sites in the ERCCI coding sequence as shown in Figure 3; the mutations and the changes generated from each of the mutant cell lines and analyzed by nucleotide sequence analysis. This provided independent confirmation of RT-PCR results and eliminated the possibility that any mutation identified by RT-PCR was the result of polymerase error.
Fig. 3. *ERCCI* coding sequence. Amino acid alignment of human, mouse and CHO cell *ERCCI* homologs and CHO cell *ERCCI* cDNA sequence.

Coping errors. Two mutant cell lines exhibited an addition or deletion producing frameshift mutations: UV4 has a CC→CCC addition at the site of CHO amino acid Leu166 and UV60 has a 4 bp deletion (GCAG) at Arg101. Point mutations were identified in seven of the remaining CHO *ERCCI* mutant cell lines. Two of these mutants, UV71 and 43-3B, had non-conservative amino acid substitutions: in UV71 a T→C transition at Leu248 changes the amino acid from leucine to proline; in 43-3B a T→A transversion at Val98 replaces the valine with a glutamic acid residue. In three mutants (UV203, UV121 and UV10) point mutations generated amber stop codons: mutants UV203 and UV121 exhibit C→T transitions at Gin174 and Gln246 respectively; UV10 shows an A→T transversion at Lys157. UV10 also exhibited aberrant splicing at the exon 361.
4/5 junction. Sequence analysis of the the UV20 cell line RT-PCR products also revealed aberrant splicing at the exon 6/7 junction. Sequencing of two independent exon 7 genomic PCR amplimers from UV20 identified a G→A transition which altered the first amino acid of exon seven, Ser196, to an asparagine, suggesting that this mutation may affect this splice acceptor site. Sequence analysis of UVL9 RT-PCR products showed that aberrant splicing was occurring at the exon 4/5 junction, but failed to identify a mutation, however, genomic PCR of exon 4 and a portion of the adjacent intronic sequence identified a G→A transition 2 nt into the 4/5 intron sequence. No mutation was identified in the ERCC1 coding sequence of the putative ERCC1 mutant cell line UV201.

Discussion

In this study we determined the coding sequence mutations in a panel of CHO ERCC1 mutants isolated from several different parental CHO cell lines following mutagenesis with various mutagens. Among the ERCC1 mutants analyzed were four that have been widely used in investigations of the role of ERCC1 in repair and mutagenesis, UV20, 43-3B, UVL9 and UVL10, isolated from three different parental CHO cell lines. UV20 was the first ERCC1 mutant to be isolated and thoroughly characterized (Thompson et al., 1981, 1982, 1985); 43-3B was the recipient cell line used for cloning the first human DNA repair gene (Westerveld et al., 1984); both UVL9 and UVL10 have been used in numerous investigations of repair, mutagenesis and recombination (Clarkson et al., 1983; Adair, 1987; Mitchell et al., 1988; Nairn et al., 1988, 1991; MacLeod et al., 1991; Sage et al., 1996). In many studies these mutants have been used as recipient cells in transfection experiments in which genomic DNA sequences or cDNA minigenes were used to complement mutagen sensitivity phenotypes (Thompson et al., 1985; van Duin et al., 1986, 1988a,b; Zdziennicka et al., 1987; Darroudi et al., 1989; Nairn et al., 1989; Sijbers et al., 1996b). Interpretation of the results obtained in such transfection experiments requires some knowledge of the endogenous mutant allele, since co-expression of transfected and endogenous mutant alleles may result in complex phenotypes.

Cloning and sequencing of the ERCC1 coding sequence from the three parental CHO cell lines (CHO-AA8, CHO-9 and CHO-AT3-2) from which most ERCC1 mutants have been derived revealed that the wild-type sequences are identical; no polymorphisms have arisen during years of culture in different laboratories. Comparison of the human, mouse and CHO ERCC1 coding sequences (Figure 3) shows that these homologs are highly conserved, particularly from the beginning of the Xpa binding site (amino acids 92–119) in the middle region of Ercc1 to the C-terminus. Within this region (amino acids 92–297) only 19 amino acids are found to be identical in all three mammalian species. For 18 of these 19 amino acid positions the amino acids are identical in two of the three ERCC1 homologs: 6/19 are identical in human and mouse, 10/19 in mouse and CHO cells and 2/19 in human and CHO cells.

The N-terminal region of the ERCC1 coding sequence (amino acids 1–90) is less well conserved. Of the 90 amino acids in this region 40 are variable between species. It is interesting to note that five amino acids, four of which are identical in human and mouse, are not represented in the CHO ERCC1 coding sequence (see Figure 3). Of the remaining 35 amino acids seven are identical in human and mouse, 14 in mouse and CHO cells and five in CHO cells and human. If conservative substitutions are allowed 24 non-conservative amino acid changes are observed in the N-terminal region. In only two cases (human Gln54 and Cys76) are the amino acids found not to be conserved between any of the ERCC1 homologs.

One ERCC1 mutant in our panel (UV4) was isolated after mutagenesis with the frameshift-inducing chemical ICR-170. Intercalating acridine mustard derivatives such as ICR-170 typically yield frameshift mutations, predominantly through single base insertions (Taft et al., 1994). Mutant cell line UV4 was found to have a one base (C addition) frameshift mutation in exon 5 of the CHO ERCC1 coding sequence, consistent with this mechanism.

Of the six CHO ERCC1 mutant cell lines in our panel isolated after mutagenesis with alkylating agents, five exhibited single base pair substitution, transition or transversion mutations. In mutants UV121 and UV203 C→T transitions generated amber stop codons in exons 8 and 6 respectively. CHO mutant 43-3B, isolated after N-ethyl-N-nitrosourea (ENU) mutagenesis, exhibited a T→A transversion in exon 3, within the Xpa binding region. ENU has been shown to react with the O2 and O4 positions of thymidine, as well as with the O6 position of guanine (Beranek, 1990). Two other missense mutants, UV20 and UVL9, exhibited the predominant G→A transition mutation characteristic of ethylmethane sulphonate (EMS) mutagenesis. The UV20 mutation was located in exon 7 and the UVL9 mutation was in the intron between exons 4 and 5 of the ERCC1 coding sequence. Another EMS-induced mutant (UV60) exhibited a very unusual frameshift mutation, a 4 bp deletion in exon 3 of the CHO ERCC1 coding sequence, within the Xpa binding site. This mutation is interesting because frameshifts are rarely observed in EMS-induced mutation spectra. However, this particular mutant was isolated after four successive rounds of EMS mutagenesis and therefore is unlikely to be typical.

Neither of the two UV-induced CHO ERCC1 mutant cell lines analyzed in this study (UV71 and UVL10) exhibited the C→T or tandem CC→TT transitions that are considered to be the 'signature mutations' of UV mutagenesis. In UV71, a T→C transition was identified in a run of pyrimidines in exon 8. In UVL10 an A→T transversion mutation was found in exon 5 of the ERCC1 coding sequence. Both types of mutation have previously been reported in other UV mutational spectra (Miller, 1985; Sage et al., 1996).

Analysis of 5' and 3' RT-PCR products obtained from each ERCC1 mutant cell line revealed aberrant pre-mRNA splicing in several mutants, similar to that previously observed among mutants with splice donor or splice acceptor site mutations (Carothers et al., 1993; Valentine and Heffich, 1995, 1997). UVL9 exhibited aberrant splicing at the exon 4/5 junction. DNA sequencing of CHO ERCC1 genomic PCR products obtained from UVL9 located the G→A transition in position 2 of the donor splice site in the intron sequence adjacent to exon 4. The aberrant splicing observed for UV20 occurred at the exon 6/7 junction. Genomic sequence analysis identified a G→A transition at position 1 of the exon 7 acceptor splice site; this mutation altered the first amino acid of exon 7, Ser196, to an asparagine. The intronic missense mutation in UVL9 and the exonic mutation in UV20 were the only mutations identified in these cell lines, suggesting that these mutations are responsible for the defective pre-mRNA splicing.
observed. Aberrant splicing was also observed in the mutant cell line UVL10, which was found to have normal splice donor and acceptor sites but had a nonsense mutation in exon 5 of the ERCC1 coding sequence. Exon skipping and other types of aberrant pre-mRNA splicing have previously been observed in other mutants with nonsense mutations in internal exons (Carothers et al., 1993; Valentine and Heflich, 1995, 1997).

Only one ERCC1 mutant in our panel, UV201, failed to exhibit a mutation in either ERCC1 coding or splicing sequences. This cell line, which was previously assigned (provisionally) to rodent complementation group 1 (CG1) based on the results of somatic cell hybridization, is only slightly UV hypersensitive (1.5X relative to CHO-AA8) and exhibits near wild-type UV mutability (Busch et al., 1997). Our results indicate that UV201 expresses both ERCC1 mRNA (Figure 1) and Ercc1 protein (Figure 2). The reduced level of expression of ERCC1 mRNA in this mutant (33%) relative to its parental cell line, CHO-AA8, suggested that the UV phenotype exhibited by UV201 might result from decreased ERCC1 gene expression. However, comparison of ERCC1 mRNA expression levels among the three parental cell lines analyzed in this study revealed that wild-type CHO-AT3-2 also expresses ERCC1 mRNA at a level approximately one third that observed for the parental lines CHO-AA8 and CHO-9. Therefore, it appears unlikely that this level of ERCC1 mRNA expression is responsible for the phenotype exhibited by UV201. Our failure to identify any mutations in the ERCC1 coding sequence in this mutant raises the possibility that UV201 could be misclassified and that its slight UV sensitivity may be due to causes other than mutation in the ERCC1 gene.

The CG1 assignments of the other nine CHO ERCC1 mutant cell lines examined are supported by the results of this study. Western analysis, performed using an antibody that specifically recognizes the C-terminal region of Ercc1 (Biggerstaff et al., 1993), detected little or no Ercc1 protein in nuclear extracts from each of these cell lines. These results are consistent with the generation of a 3' truncated Ercc1 protein in the UV121, UV203, UVL9 and UVL10 mutants and with loss of reading frame in mutants UV4 and UV60. The point mutant 43-3B had previously been shown to express Ercc1 at a greatly reduced level (Biggerstaff et al., 1993), but the profound effects of the missense mutations in UV20 and UV71 on Ercc1 protein expression were unexpected. These results suggest that mutations in ERCC1 can either directly induce protein instability or, if complex formation (e.g. with the XpF protein) is required to stabilize Ercc1, may indirectly result in protein instability by affecting protein–protein interactions. Our results are consistent in this regard with those from a recent study (Sijbers et al., 1996b) in which site-directed mutagenesis was used to generate missense mutations in a human ERCC1 cDNA minigene construct, primarily within a highly conserved region of ERCC1 possibly involved in DNA binding (human Ala138—Pro150). Mutated ERCC1 minigene constructs were transfected into the ERCC1 mutant cell line 43-3B and, in cases where protein levels were determined, it was found that the level of Ercc1 protein was reduced (six cases) or not detectable (two cases).

The mutations we identified in nine CHO ERCC1 mutant alleles were located throughout the middle and C-terminal regions of the ERCC1 nucleotide sequence, from mutations in the XpA binding region in exon 3 to mutations in exon 8 (see Figure 3). No mutations were identified in the N-terminal third of the ERCC1 coding sequence, consistent with reports that this region is not required for complementation of UV and mitomycin C sensitivities in ERCC1-deficient cells (van Duin et al., 1988a,b; Sijbers et al., 1996b). In their study Sijbers et al. introduced missense mutations in a region of ERCC1 corresponding to what is considered to be the most evolutionarily conserved amino acid sequence when compared with the Swi10 and Rad10 proteins from Schizosaccharomyces pombe and Saccharomyces cerevisiae respectively. Based upon structural motifs, it has been suggested that this region may be important for DNA binding (van Duin et al., 1988b; see Figure 3). However, of the mutants in our panel (all of which were generated by conventional mutagenesis followed by screening for UV-sensitive phenotypes) only one (UVL9) was found to have a mutation in this region.

Of the two missense alleles in our mutant panel that did not exhibit splicing errors one is located in exon 3 in the XpA binding region (43-3B) and one is located in exon 8 (UV71). In the case of 43-3B it is certainly possible that the non-conservative Val—>Glu substitution could affect interaction between Ercc1 and XpA; the fact that we and others (Biggerstaff et al., 1993) have observed little or no Ercc1 protein in 43-3B would support the hypothesis that protein–protein interaction may stabilize Ercc1. In the case of UV71 the missense mutation, resulting in a Leu—>Pro substitution, is in exon 8, which has been shown to be alternatively spliced and required for UV complementation (van Duin et al., 1986). Since UV71 also exhibits no detectable Ercc1 protein (although steady-state mRNA levels are not reduced, see Table 3), it is possible that mutations in exon 8 may likewise affect protein stability, perhaps due to an inability of the ERCC1 allele in UV71 to form a stable protein complex. Sijbers et al. (1996b) suggested that the region of Ercc1 corresponding to human Thr236—Leu289 may form a specific α-helix—turn—α-helix conformation followed by a long loop and two helices similar to that observed for the crystal structure of Taq polymerase (Kim et al., 1995). The Leu—>Pro missense mutation in UV71 occurs within the second α-helix at the only position (human Leu253) that appears to be absolutely conserved in the consensus sequence presented by Sijbers et al. (1996b); a proline substitution at this site could certainly result in altered protein conformation.

At least three mouse models with knock-out or nonsense mutations in ERCC1 have been generated (McWhir et al., 1993; Weeda et al., 1997). These mice exhibit interesting phenotypes with respect to growth failure and premature senescence, in addition to repair defects. These animal models have excellent potential for correlating ERCC1 gene alterations with disease processes (Weeda et al., 1997). Recently we constructed several CHO cell ERCC1 knock-out mutants (Rolig et al., 1997; Sargent et al., 1997; unpublished results) for studies of the role of ERCC1 in recombination in mammalian cells and to generate true ERCC1 null CHO cell mutants with no detectable mRNA or protein expression for use as transfection recipients. Transfection of such CHO cell null mutants with ERCC1 alleles that have been specifically altered at sites at which missense mutations have been localized (e.g. 43-3B and UV71) or with site-directed mutations in critical regions, such as the XpA binding region, has potential for defining the phenotypes of specific ERCC1 mutant alleles relatively rapidly and inexpensively in cell culture experiments; identification of such alleles and their associated cellular phenotypes could subsequently be used to create specific ERCC1 alterations in mouse models. These experimental
approaches could lead to the molecular dissection of functionally important domains of ERCC1 and aid in defining its roles in growth control and development, as well as in DNA repair and recombination.

Acknowledgements

We wish to thank Drs David Busch, Larry Thompson and Rick Wood for providing cell lines, we are also grateful to Dr Wood for Ercti antisera. In addition, R.L.R. thanks Dr Pratap Kumar, Dr Riaa Ghosh and Amy Pavone for their expert advice. This research was supported by USPHS grant CA36361 from the National Cancer Institute; we also acknowledge support from NIEHS Center Grant ES07784. R.L.R. is the recipient of a Rosalie B Hise predoctoral fellowship.

References


Klein,H.L. (1988) Different types of recombination events are controlled by the RAD1 and RAD5 genes of Saccharomyces cerevisiae Genes, 120, 367–377.


CHO cell ERCCI mutant alleles


Received on October 10, 1997; accepted on January 26, 1998