애기장대로부터 DNA 수선 단백질인 XRCC1 을 암호화하는 유전자의 클로닝과 분석

김정섭^{1*}, 크멕 애릭²

¹제주대학교 응용생물산업학과, 제주도 690-756, 한국 ²델라웨어생명공학연구소, 뉴왁시, 델라웨어주 19711, 미국

Cloning and Analysis of the Gene Encoding a DNA Repair Protein XRCC1 from *Arabidopsis thaliana*

Jung-Sup Kim^{1*}, Eric B. Kmiec²

¹Department of Applied Biotechnology, Cheju National University, Jeju 690-756, Korea ²Delaware Biotechnology Institute, Newark, DE 19711, USA

ABSTRACT: X-Ray Cross-Complementing group 1 (XRCC1) is a DNA repair protein that responds to DNA damage. A cDNA clone. XRCC1. from *Arabidopsis* thaliana isolated, sequenced, and, through combined analysis of the cDNA and genomic DNA, found to contain seven introns. It was determined that the XRCC1 coding sequence is 1062 bp in length and encodes a 353 amino acid, 40 kDa protein. The predicted amino acid sequence of XRCC1 contains a consensus BRCT domain and is similar to, but significantly shorter than, the XRCC1 counterparts from animals. XRCC1 was shown to increase resistance to methylmethane sulfonate (MMS) in wild type yeasts, whereas mutant yeast cells deficient in RAD51 or RAD52 were not rescued by overexpression of XRCC1.

INTRODUCTION

Damaged DNA created from exposure to radiation can lead to the disruption of

cellular processes, such as DNA replication and recombination. To counteract this damage. the cell has evolved tolerance mechanisms in which damaged DNA is rapidly repaired through the precise and concerted activities of a variety of proteins. DNA repair mechanisms also maintain genetic stability and suppress the formation of aberrant aging events that can cripple the cell. XRCC1 (X-Ray Cross-Complementing group 1) is a DNA repair protein needed to repair DNA damage caused by ionizing radiation or DNA methylating agents. XRCC1 has been primarily characterized in Homo sapiens (Thompson et al., 1990). Cricetulus griseus (Shen et al., 1998), and Drosophila melanogaster (Taylor et al., 2000). Human XRCC1 consists of three distinct domains, including the N-terminal domain (NTD), which can interact with DNA polymerase (Marintchev et al., 2000), and two other domains known generally as breast cancer carboxy-terminals (BRCT), which are thought to serve as modules for protein- protein interactions (Zhang et al., 1998). BRCT-I and BRCT-II motifs may interact with poly (ADP-Ribose) polymerase (PARP; Pleschke et al., 2000) and DNA ligase III, respectively.

As part of a long-term investigation on the process of DNA repair mechanisms in plants. we have initially identified genes involved in DNA repair processes from Arabidopsis This genome, which has been thaliana. completely sequenced (The Arabidopsis Genome Initiative, 2000), provides fertile ground for isolating important genes that can participate in or regulate repair events. The partial gene structure of XRCC1 was recently reported (Theologis et al., 2000) from A. thaliana and contains a coding sequence with significant homology to animal XRCC1 (Taylor et al., 2002). But, functional information about this gene was not described; thus, we characterized the gene encoding XRCC1 with the immediate goal of elucidating function of the XRCC1 protein. Here, we report the cloning of the Arabidopsis gene encoding a DNA repair protein, XRCC1, but also extend the study to an examination of its activity in yeast strains deficient in a variety of DNA repair activities.

MATERIALS and METHODS

Plant materials

Arabidopsis thaliana ecotype Columbia (Col) was raised in a 23C growth chamber with a light intensity of 100 uE.m⁻¹.s⁻¹ and a photo-period of 14 hours light followed by 10 hours darkness.

DNA sequence analysis

An EST clone (GenBank Acc. No. N37375, cDNA ID 210J9T7) was obtained from the

Arabidopsis Biological Resource Center (ABRC, Ohio State University). The clone was derived from A. thaliana (Col) cDNA library (D'Alessio et al. 1992) prepared in Zip-Lox (GibcoBRL, Inc.) containing plasmid pZL1. DNA sequencing was performed by the dideoxynucleotide chain termination method with an automatic sequencer (AlFexpress DNA sequencer, Pharmacia Biotech, Inc.). Nucleotide sequences and amino sequences were compared with sequences present in the GenBank and EMBL databases and analyzed using BLAST and PILEUP of the GCG Wisconsin Package Biology WorkBench 3.2 (http://workbench.sdsc.edu: San Diego Supercomputer Center: University of California San Diego). Comparison of sequences with published animal XRCC1 sequences was performed at the nucleotide and amino acid level.

Overexpression of XRCC1 in yeasts and MMS resistance test

To investigate the function of At XRCC1. the cDNA was subcloned into the yeast expression vector pYES2/CT (Invitrogen Corp., Carlsbad, CA) by KpnI and XbaI sites to generate plasmid pYES-XRCC1: expression of XRCC1 is under the control of the inducible GAL1 promoter. pYES-XRCC1 or pYES2/CT (as a negative control) were transformed into a wild-type yeast strain (LSY678: Liu et al. 2001) or strains lacking RAD51 (LSY402) or RAD52 (LSY386), respectively. Transformants were selected and maintained on agar containing SD supplemented media without Ura (SD-Ura, Invitrogen Corp.: Liu et al. 2002).

The resistance to DNA damage of yeast strains conferred by plasmid pYES-XRCC1 was

assessed using a traditional radiomimetic drug-sensitivity assay as follows: A single yeast colony was inoculated and grown in SD-Ura liquid media to $OD_{600} = 0.7$. XRCC1 was then induced by replacing the SD-Ura second SD-Ura media media with a containing 2% galactose, which induces the GAL1 promoter in the plasmid. The culture was grown to OD₆₀₀=0.5-0.9 over a period of 9 to 24 hours. The culture was diluted 1/10⁻⁵, and plated on SD-Ura+Gal media containing various concentrations mutagen methylmethane sulfonate (MMS) 0-0.05%. The highest ranging from concentration of MMS (0.05%) is sufficient to kill the entire wild-type population.

RESULTS

Identification of XRCC1 cDNA and sequence analysis

DNA repair genes typically exhibit sequence similarity, even among widely species. To identify genomic sequences from Arabidopsis plants based on sequence similarity to known DNA repair genes, we performed a database search with various gene fragments known to encode conserved regions of DNA repair proteins. Queries to the GenBank database using human XRCC1 protein sequences (GenBank Acc. No. M36089) by TBLASTN revealed an Arabidopsis gene with similar sequence (GenBank Acc. No. AC018849), and we identified it as a putative Queries to the Arabidopsis XRCC1 gene. EST database with the genomic sequence identified an EST 210J9T7 (GenBank Acc. No. N37375) as another putative XRCC1 coding region. The EST clone (210J9T7) was obtained directly from the Arabidopsis Biological Resource Center (ABRC) and was processed for DNA sequence analyses.

Sequence analysis revealed that the 210J9T7 clone contains a full length cDNA of the Arabidopsis XRCC1 gene, and that its sequence was identical to the genomic sequence (GenBank Acc. No. AC018849) derived from BAC clone T21F11 in Chromosome 1. The 1390bp cDNA contains an open reading frame (ORF) of 1062bp, and a 141bp 3'-untranslated region ending with a 30 nucleotide polyA-tail (Fig 1). But full, unimpeded extension from first start codon to the 5' end of the cDNA clone was viewed as likely, due to the presence of three stop codons existing in frame with the coding sequence.

The ORF, however, did encode a 353 amino acid polypeptide with a calculated molecular of 39.8kDa and a theoretical weight isoelectric point of 6.2. Inspection revealed 58-146 acids are highly that amino homologous to the BRCT-I motif (Fig 2A) present in XRCC1 homologs from H. sapiens, R. norvegicus, C. griseus, and D. melanogaster (Fig 2B), and typical for this subfamily of poly(ADP-ribose) polymerase (PARP)-interacting proteins (Masson et al., 1998). The deduced amino acid sequence of XRCC1 (353 amino acids) is similar to but shorter than the mammalian counterparts (Table 1). XRCC1 possesses a BRCT domain associated with PARP binding, but appears to lack the N-terminal domain for interacting with DNA polymerase. The C-terminal BRCT-II motif is not totally conserved in the plant gene, but some partial motifs are located in downstream from the BRCT-I domain.

GUTTCATAATAGTGTTTCTTTGTAGGACGAAAGCTGTGTTCAAGTAACTAGTGGCTTAAA GGACAAGATGTCTCAGAAGCGAAAGCTGTGTTTAGGATTCCACCTGTAAA GGACAAGATGTCTCAGAAGCGAAAGCTGTTTTGTGTATGGTATTCTCCACCTGTAAA M S Q K R M L P S M M S S R D P E I TACTCCCAGGTAAGTCTCATTGTAAGAAGCCTAAGGATGAAGGACCAAAGAGCCAAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGAGCCAAGTCCAACAAGCCAAGAGCCAAGTCCAAAGAACCACGAAGCCCAAGTCCAAAGAACCACGAAGCCCAAGTCCAAGAACCACGCAAGTCCAAAGAACCACGAAGCCCTAACACCTGAA CAGTAAGAAATGCCCCCTTCCCAACAAGTCAGAACACCGCCGAGGCCAAGTTCAAATACCACTGAA S R N A P S N K S E H A E P S S N T T E 58 ATTCTTCTAAACTTATG GAAAGGCTTGTTTTTTGTACTTTCGGGCTTTGTTTAATCCTGAAGG F S K L M E G V V F V L S G F V N P E R 78 GAGTACACCTGAGGCTTGACAATGGGAGACCCCCTATCAAACCTCTGACCTGAACACCTGAACCTGACCTGACCTGACCTGAACCCTGACCTGACCTGACAGCCTTGACAATGGAGACCCCCACTATCAACCTCCAAGATTCAAGACCTGAACACCTGACCTGACAGCCTGACCTGACATGGAACCCCGAAGAAGCACTGACTG	· v	
GGACAAGATGTTCTTTGTAGGACGAAAGCTGTGTTAGTATTCTCCCCCTGTAAA GGACAAGATGTCTCAGAAGCGAAATCTTCCTTCTTGGATGAGTTCTAGGGATCCCGAGAAT M S O K R M L P S M M S R D P E I TACTCCCAGGTAAGTCTCATTGTAAGAAGCCTAAGGATAGAACCGAAAGACCGAAAGACCGAAAGACCGACGA	GAAGCCACGGGGGAGCGGTAGCACTTAAAG ATTTGCACCGTTTTCTCTCTGGCTTTGTG	60
CAGACAAAATGCCCCATCAAAAAACCAAAAAAACCAAAAAAAA	TTGTGCTAGATGCGATGCAATAGCAATATTTTCCTGGAGTTTCAAGTAACTAGTGGCTTA	120
TACTCCCAGTAAOTCTCATTGTAAGAAGCCTAAAGAATGAAGCCAAACGCAAAAGAAGCCAAAAGCCAAAAGCCAAAGCCACAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCAAAGCCCCAAAGCCCCCTCCCAAAAGAAAG	GGTTCATAATAGTGTTTCTTTGTAGGACGAAAGCTGTGTTAGTATTCTCTCCACTGTAAA	180
TACTCCCGAGTAAGTCTCATTGTAAGAAGCCTAAGGATGAAGGACCAACCGAAGAGGCCAACCGATT TO S K S H C K K P K D B G P T E B H N 38 38 36 36 36 36 36 5 5 5 5 5 5 5 5 5 5 5 5		240
T P S K S H C K K P K D B G P T E B H N 38 CAGTAGAAATGCCCCTTCCAACAAGTCAGAACACGCCGGAGGCCAAGTTCAAATACCACTGA 360 S R N A P S N K S B H A B P S S N T T E 58 ATTCTCTAAACTTATG GAAGGCGTTGTTTTTGTACTTTCGGGCTTTGTTAATCCTGAGAG 420 F S K L M B G V V F V L S G F V N P B R 78 GAGTACACTGAGGTCACAAGCCTTGACAATGCGACCCCCAAAGTTCCAACTGAACCCGGAAGCCC S T L R S Q A L T N G A T Y Q P D N N A 98 GBCT domain) TGGCTCTACCTTGTTGTTTTTTCTTACACTCCAAAGTTCCGACACGC 480 GBC T L L I C A F P N T P K F R Q V E T 118 AAATGGTGGAACAATTATCTCAAAG GAGTGAGAACCACCTGAAAGTTCCGTCAAGTTTCAGAACCGACAAGAACCATGGAAAAAAACCAATTATCTCAAAG GAGTGAAAAAAACCAAGAAAAAAAACCAAAAGAAAAAAAA	MSQKRNLPSWMSSRDPEI	18
CAGTAGAAATGCCCCTTCCAACAGATCAGAGCCGAGCCAAGTTCAAATACCACTGA S R N A P S N K S E H A E P S S N T T E 58 ATTCTCTAAACTTATG GAAGGCGTTGTTTTTGTACTTCGGGCTTTGTTAATCCTGAAGG P S K L N B G V V F V L S G P V N P B R GAGTACACTGAGGCCACAAGCCTTGACAATGGGAGCCACCTATCAACCCTGACTGGAACGC S T L R S Q A L T N G A T Y Q P D N N A GRCT domain) TGGCTCTACCTTGTTGATTTGTGCTTTTCCTAACACTCCAAAGTTCCGCAAGATTGAAACGC G S T L L I C A F P N T P K F R Q V B T AAATGGTGGAACAATTATCTCAAAG GAGTGGATACAACACTCCAAAGTTCAACAGTAAGAAAAAAACAAATTATCTCAAAG GAGTGGAAAAACAATTATCTCAAAG GAGTGGAAAAAAAACCAGAAGAAACAATTATCTCAAAG GAGTGGAAAAAAAACCAGAAGAAACAAAGGGAAACACCCTGTCTAAAAAACCAGAGAAAACAAGGAAGAACACCTTGTCTAAAAAAACCAGAGAAACAAGGGCAAACACCCTGTCTAAAAAAACCAGAGAAACAAGGGCAAACACCCTGTCTAAAAAACCAGAGAAAGCAATGAGC Q D A N R B K R B H L S K K P E K Q V E 178 CCAAAGATGCTAATCCGA AAAAGAGAGAACACCATCTACCTCAACACAAGAAAAAACCAGAGAAACAAAGGGGAAACACCATCTACCTCATCCAAGAAAAAAAA	TACTCCGAGTAAGTCTCATTGTAAGAAGCCTAAGGATGAAGGACCAACCGAAGAGCACAA	
S R N A P S N K S E H A E P S S N T T E TO ATTOCTARACTTATO GAACGCOTTOTTTTTTTTTTTCTCTTTTTTTTTTTTTTTT	TPSKSHC <u>KRPK</u> DEGPTEEHN (MLS)	38
ATTICTCTAAACTTATG GAAGGCGTTGTTTTGTACTTGGGGCTTTGTTAATCCTGAGAGG $P \subseteq N \subseteq $	CAGTAGAAATGCCCCTTCCAACAAGTCAGAACACGCCGAGCCAAGTTCAAATACCACTGA	360
ATTOCTTARACTTATG GAAGGCGTTGTTTTTGTACTTTCGGGCTTTGTTATACTCTGAGGG 78 P		58
CAGTACACTGAOGCTCACAAGCCTTGACAATGGGAGCCACCTATCAACCTGACTGA	· · · · · · · · · · · · · · · · · · ·	420
S T L R S Q A L T M G A T Y Q P D N N A 98	F S K L M B G V V F V L S G F V N P B R	78
TOGGETCTACCTIGTTGATTTGTGCTTTTCCTTACACCTCCAAAGTTCCAAGTTGAAAC G S T L L I C A F P N T P K P R Q V B T AAATGGTGGAACAATTATCTCAAAG GAGTGGAAAACCATGGAGGAAAAAGCAGGTAGTCC AAATGGTGGAACAATTATCTCAAAG GAGTGGAAAACCATGGAGGAAAAAAGCAGGAAGAAAGCAAGC	GAGTACACTGAGGTCACAAGCCTTGACAATGGGAGCCACCTATCAACCTGACTGGAACGC	480
TOGCTCTACCTTGTTGATTTGTCTTTTCCTAACACTCCAAAGTTCCGTCAAGTTGAAAC G S T L L L I C A F P N T P K P R Q V B T AAATTGGTGGAACAATTATCTCAAAG V AAATTGGTGGAACAATTATCTCAAAG G G T I I S K B N I T B C Y A Q K K L 138 GGTAGATATCGAGCAATACCTTATGCATGGTGGAAAACCATGGAGGAAAAACAGTAGTCC V D I B Q Y L N H A G K P N R K S S S P 158 GCAAGATGCTAATCGAG AAAAGAGAAACCATGTCAAAAAAACCAGAGAAAGCAAGTAGA V CCAAGATGCTAATCGAG AAAAGAGAAACCCATCTCAAAAAAACCAGAGAAGCAAGTAGA K K T E T R G T P S T S S K N R S A C N 198 CGAAAAAAACAAGAGAGAACCCATCTTACCTCATCCAAG AATAGATCAGCTTGCAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAGAGCCATTTTCTGTTACAGAGGTGAAAACCAGGGGGAAATCAAGCG Q T I S N L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTACCTGCTTACAAGACCCATCTTGAACAAGAACCAGGCGAAATCAAGCG Q T I S N L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTACCTGCTTACAAGACCCCATCTTGTAACAAGAACCATCTCTTGAACAAGAACCATCTTACAAGACCATCTTACAAGACCCATCTTTTTTTT	STLRSQALTMGATYQPDNNA	98
TOAAACCATCTCATGGCTTGAAGAGGCAAGAGGAAGAAGCCATGGCGGAAAGAGGCCCATGAAGCAGGCGAAAGAGGCCCTTCATGCATG		
AAATOGTOGAACAATTATCTCAAAG GAGTOGATAACTGAGTGTTATGCGCAGAAGAAGTT 18	TGGCTCTACCTTGTTGATTTGTGCTTTTCCTAACACTCCAAAGTTCCGTCAAGTTGAAAC	
N	G S 1 B B 1 C X F P N 1 P K P R Q V B T	118
N	A NA TOUTOCA A CA A TITLA TOUTCA A DE CA CITATA DE CACITA	
GGTAGATATCGAGCAATACCTTATGCATGCTGGAAAACCATGGAGGAAAAGCAGTAGTCC V D I B Q Y L M H A G K P W R K S S S P 158 CCAAGATGCTAATCGAG AAAAGAGAGAACACCTGTCTAAAAAAACCAGAGAAGCAAGTAGA 720 Q D A N R B K R B H L S K K P E K Q V E 178 GAAAAAAACAGAAACAAGGGGAACACCATCTACCAAG AATAGATCAGGCTTOCAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAGAGCCATTTTCTGTTACAGAGGTGAAAAAACCAGGGCGAAATCAAGG 840 L V K E P F S V T E V K K W A R D D L S 218 TCAAACCATCTCATGGCTTGAGAGTCAGGAGGAAAAA CCAGAACCAGGCGAAATCAAGGG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGGTTAACCTGCTTACAAGACGCCATAGATTCCTTGAACAGAA 960 I A A E G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S F V P R V V K E L 278 TGGAAAGATGGAACCAGTCTCTTAAAAAGGAAAATTCAACAGCATCAAAGAAGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGGATCTTACGAAGCTAGAGCTAGAGAACCAGGGCGAAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGAGTAAACCTGGGCGAAGAA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGTATTAGACCTTTTAGATCAAACCTTGT 1260 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGAAGATCGAACCTTCTTAAAAAGGAACCTTTCTTT		
V D J B Q Y L M H A G K P W R K S S S P 158 V CCAAGATGCTAATCGAG AAAAGAGAGAACCCTGTCTAAAAAAACCAGAGAAGCAAGTAGA 720 Q D A N R E K R E H L S K K P E K Q V E 178 GAAAAAAACAGAAACAAGGGGAACACCATCTACCTCATCCAAG AATAGATCAGCTTGCAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAGAGCCATTTTCTGTTACAGAGGTGAAGAAATGGGCCAGGGAAATCAAGCCAGCAAAG L V K E P F S V T E V K K W A R D D L S 218 TCAAACCATCTCATGGCTTGAGAGTCAGGAGGAAAAAA CCAGGAACCAGGCGAAATCAAGCG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGAGCGCCATAGATTCTCTTGAACAGAA 960 I A A E G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAGAGGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAAG		
CCAAGATGCTAATCGAG AAAAGAGAGAACACCTGTCTAAAAAAACCAGAGAAAGCAAGTAGA 720 Q D A N R E K R E H L S K K P E K Q V E 178 V GC-AG SP11c1bg Site V GC-AG SP11c1bg Site QAAAAAAACAGAAACAAGGGGAACACCATCTACCTCATCCAAG AATAGATCAGCTTGCAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAAGAGCCATTTTCTGTTACAGAGGTGAAGAAATGAGCCAGGGAAATCAAGCG 900 L V K E P P S V T E V K K W A R D D L S 218 TCAAAACCATCTCATGGCTTGAGAGTCAGGAGGGAAAAA CCAGAACCAGGCGAAATCAAGCG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCATAGATTCTCTTGAACAGAA 960 I A A E G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGCATCAGTTACAGAGCTTAGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S P V P R V V K E L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGGAAAAAAGATCTACGAAAGTTAGACCAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGAATAGACCGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGTATAGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGAATCGAACTTGCTTATAAGGAACGTTTCTTTAGAGTGTCTTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATTAGGAACTTCCTAAGGAATTCCTAAGAATTCTTTGTTTTTTCTTTTTTTT		
CCAAGATGCTAATCGAG AAAAGAGAGAACACCTGTCTAAAAAACCCAGAGAAGCAAGTAGA 720 Q D A N R E K R E H L S K K P E K Q V E 178 V GC-AG SPIIcing site GAAAAAAACAGAAACAAGGGGAACACCATCTACCTCATCCAAG AATAGATCAGCTTGCAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAGAGGCCATTTTCTGTTACAGAGGTGAAAAAACCAGGATGACCTAAG 840 L V K E P F S V T E V K K W A R D D L S 218 TCAAACCATCTCATGGCTTGAGAGGTCAGGAGGAAACCAGGCGAAATCAAGCG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGAGGAGTCTTAACCTGCTTACAAGACGCCATAGATCTCTTGAACAGAA 960 I A A E G V L T C L Q D A I D S L E Q K 258 T A A E G V T E L W S F V P R V V K E L 276 TGGAAACATGTCTCTTCTAAAAAGGAAAATTCAACAGCATCTCGTTTACAAGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S F V P R V V K E L 276 TGGAAACAAGCAAACTGTCTTCTAAAAAGGAAAAATTCAACAGCATCAAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGCCGTGGGAAAAAAGAATCTACGAAGCTAAGAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGAGAAGCTGAGCTTAGAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATTATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGATCGAACTTGCTTAAAGGAACCGTTTTTTAAGGACCTTTTTTAAGGACTTTTTTTAAGAACTTGTTTTTTTT		
V GC-AG SPILLING SILE GAAAAAAACAGAAACAAGGGGAACACCATCTACCTCATCCAAG AATAGATCAGCTTGCAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAGAGCCATTTTCTGTTACAGAGGTGAAGAAATGGCCCAAGGAACCCAGGGTAAGAAATGGCCCAAGGAACCCAGGGTAAGAAATGGGCCAGGGAAATCAAGCG 218 TCAAACCATCTCATGGCTTGAGAGTCAGGAGGAAAAAA CCAGAAACCAGGCGAAAATCAAGCG 200 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTACAGAGCTATGGAGCCTAAGAACCAGGCGAAAATCAAGCG 238 AATAGCTGCGGAAGGAGTCTTACACGTGCTTACAAGACGCCCATAGATTCTCTTGAACAGAA 258 V GCAG GACATTGGATCAGTTACAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S P V P R V V K E L 278 TGGAAAGATGGATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGGAAAATTCAACAGCATCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGAATGACCGGTTGAGTGAACCTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGAAACTCGAACTTGCTTATAGGAACTTCTTTTAGACAGCTTTTTTAGACCAACCTTTTTTTT		720
GANAAAACAGAAACAAGGGGAACACCATCTACCTCATCCAAG AATAGATCĀGCTTGČAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAGAGCCATTTTCTGTTACAGAGGTGAAGAAATGGGCCAGGGATGACCTAAG L V K E P P S V T E V K K W A R D D L S 218 TCAAACCATCTCATGGCTTGAGAGGAGGAAAAA CCAGAACCAGGCGAAATCAAGCG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGGATCAACGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG		
GANAAAACAGAAACAAGGGGAACACCATCTACCTCATCCAAG AATAGATCĀGCTTGČAA 780 K K T E T R G T P S T S S K N R S A C N 198 TCTTGTGAAAGAGCCATTTTCTGTTACAGAGGTGAAGAAATGGGCCAGGGATGACCTAAG L V K E P P S V T E V K K W A R D D L S 218 TCAAACCATCTCATGGCTTGAGAGGAGGAAAAA CCAGAACCAGGCGAAATCAAGCG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGGATCAACGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG	V 00-1014-4	
TCTTGTGAAAGAGCCATTTTCTGTTACAGAGGTGAAGAAATGGGCCAGGGATGACCTAAG L V K E P P S V T E V K K W A R D D L S 218 V TCAAACCATCTCATGGCTTGAGAGTCAGGAGGAAAAAA CCAGAACCAGGCGAAATCAAGCG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCATAGATTCTCTTGAACAGAA 960 I A A E G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAGAGCTTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGTTTG 1020 Q D I G S V T E L W S P V P R V V K E L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGGATCTACGAAGCTGAGCTAGACCAGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATTATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGAACTAGAACTTGCTTATAGGAACGTTTCTTTTAGAGTGTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAAGGAATTCCAAGAATTCTTGTTTATACTCGGTTTGTTC 1320	GAAAAAACAGAAACAAGGGGAACACCATCTACCTCATCCAAG AATAGATCAGCTTGCAA	780
V TCAAACCATCTCATGGCTTGAGAGGTCAGGAGGAAAAAA CCAGGACCAGGCGAAATCAAGCG Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCATAGATTCTCTTGAACAGAA AA E G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S P V P R V V K E L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGAACTGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAACTTCCAAGAAATTCTTGTTTATACTCGGTTTGTTC 1320	KKTETRGTPSTSSKNRSACN	198
TCAAACCATCTCATGGCTTGAGAGTCAGGAGGAAAAA CCAGAACCAGGCGAAATCAAGCG 900 Q T I S W L E S Q E E K P E P G E I K R 238 AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCCATAGATTCTCTTGAACAGAA 960 I A A E G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S P V P R V V K E L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGAACTAGACTTATAGGAACGTTTCTTTAGAGTGTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * TTCATTTTGCATATGATCTCTTAAGGAATTCCAAGAAATTCTTGTTTATACTCGGTTTGTTC 1320		840
AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCATAGATTCTCTTGAACAGAA AA B G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T B L W S F V P R V V K E L 278 TGGAAGGATGGATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGGAAAAAGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGAACTGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCAAGAATTCTTGTTTATACTCGGTTTGTTC 1320	LVKEPFSVTBVKKWARDDLS	218
AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCATAGATTCTCTTGAACAGAA AA B G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T B L W S F V P R V V K E L 278 TGGAAGGATGGATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGGAAAAAGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGAACTGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCAAGAATTCTTGTTTATACTCGGTTTGTTC 1320	▽	
AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCATAGATTCTCTTGAACAGAA I A A E G V L T C L Q D A I D S L E Q K 258 V GCAG GACATTGGATCAGTTACAAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S P V P R V V K E L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGATCTACGAAGCTGAGGCTAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGAACTGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCAAGAATTCTTGTTTATACTCGGTTTGTTC 1320		
V GCAG GACATTGGATCAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S F V P R V V K E L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAATTCAAACGCATCAAAGGAAGATTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGGATCTACGAAGCTAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGCATATGATCTCTTAAGGAATTCCAAGAATTCTTGTTTATACTCGGTTTGTTC 1320		238
GCAG GACATTGGATCAGTTACAGAGCTATGGAGGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T B L W S P V P R V V K B L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M B S S S K K B N S T A S K B B V C 298 CAAACAAGCAAAGTCGTGGAAAAAGATCTACGAAGCTGAGCTGAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y B A B L A K P G B D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 B S T S R V A C G Y D S D M T V B M T B 338 AGAAGAAGATCGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTTAGATCAAACTTGT 1260 B B I B L A Y R N V S L B C L * 353 TTCCATTTGCATATGCATATGATCTCTTAAGGGAATTCCTAAGAATTCTTGTTTATACTCGGTTTGTTC 1320	AATAGCTGCGGAAGGAGTCTTAACCTGCTTACAAGACGCCATAGATTCTCTTGAACAGAA	
GCAG GACATTGGATCAGTTACAGAGCTATGGAGCTTTGTCCCTCGTGTAGTGAAGGAGCT 1020 Q D I G S V T E L W S F V P R V V K E L 278 TGGAAAGATGGAATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAGGAAGAAGTTTG 1080 G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA 1140 K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGATCGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCCAAGAATTCTTGTTATACTCGGTTTGTTC 1320		258
TOGARAGATOGAATCGTCTTCTAAAAAGGAAAATTCAACAGCATCAAAAGGAAGAAGTTTG G K M E S S S K K E N S T A S K E E V C 298 CAAACAAGCAAAGTCGTGGAAAAAGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGATCGAACTTGCTTATAGGAACGTTTCTTTAGATCTAAACTTGT 1260 E E I E L A Y R N V S L E C L * TTCATTTTGCATATGATCTCTTAAGGAATTCCCAAGAATTCTTGTTATACTCGGTTTGTTC 1320		
CAAACAAGCAAAGTCGTGGAAAAAGATCTTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAAGATCGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTGCATATGCATATGCATCTTAAGGAACTTCCAAGAATTCTTGTTATACTCGGTTTGTTC 1320		
CAAACAAGCAAAGTCGTGGAAAAAGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGATCGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCCAAGAATTCTTGTTATACTCGGTTTGTTC 1320		
K Q A K S W K K I Y E A E L A K P G E D 318 TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA 1200 E S T S R V A C G Y D S D M T V E M T E 338 AGAAGAGATCGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGCATATGATCTCTTAAGGAATTCCTAAGAATTCTTGTTATACTCGGTTTGTTC 1320	G K M E S S S K K E N S T A S K E E V C	298
TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA E S T S R V A C G Y D S D M T V E M T B 338 AGAAGAACTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGCATATGATCTCTTAAGGAATTCCTAAGAATTCTTGTTATACTCGGTTTGTTC 1320	CAAACAAGCAAAGTCGTGGAAAAAGATCTACGAAGCTGAGCTAGCAAAACCGGGCGAAGA	1140
B S T S R V A C G Y D S D M T V E M T B 338 AGAAGAGATCGAACTTGCTTATAGGAACGTTTCTTTAGATGTGTCTTTAGATCAAACTTGT 1260 B E I B L A Y R N V S L B C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCTAAGAATTCTTGTTATACTCGGTTTGTTC 1320	KQAKSWKKIYBABLAKPGBD	318
AGAAGAGATCGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTAGATCAAACTTGT 1260 E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGCATATGCATCTCTTAAGGAATTCCCAAGAATTCTTGTTATACTCGGTTTGTTC 1320	TGAATCGACCTCTCGGGTTGCTTGTGGCTATGACAGTGATATGACCGTTGAGATGACTGA	1200
E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCAAGAATTCTTGTTATACTCGGTTTGTTC 1320	ESTSRVACGYDSDMTVEMTE	338
E E I E L A Y R N V S L E C L * 353 TTCATTTTGCATATGATCTCTTAAGGAATTCCAAGAATTCTTGTTATACTCGGTTTGTTC 1320	AGAAGAGATCGAACTTGCTTATAGGAACGTTTCTTTAGAGTGTCTTTAGATCAAACTTGT	1260
	EBIBLAYRN VSLECL*	1400
	TTCATTTTGCATATGATCTCTTAAGGAATTCCCAAGAATTCTTGTTATACTCGGTTTGTTC	353

Fig 1. XRCC1 cDNA sequence.

Features of a XRCC1 cDNA sequence was presented. Mark represents the junction of intron and exon in splicing site and the GC-AG site is written followed by mark. Nucleus localization signal (NLS) and conserved BRCT domain were underlined with bold-type or italics, respectively. The stop codons and a poly(A) signal is indicated by an asterisk and a box, respectively.

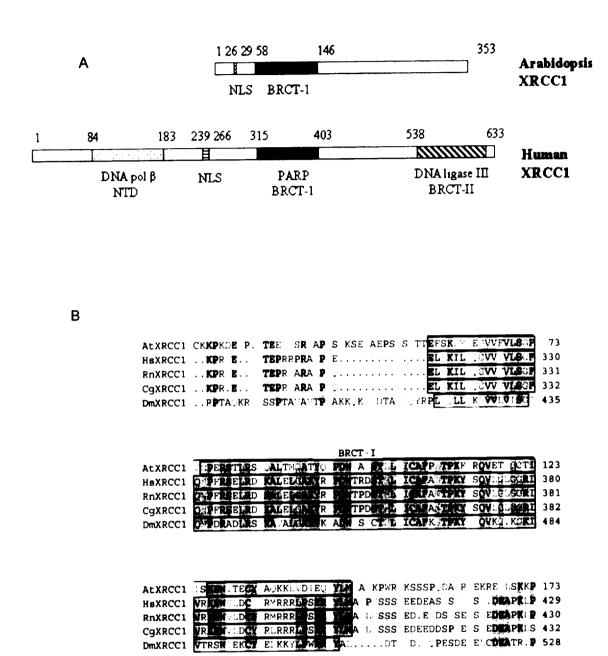


Fig 2. Amino acid sequence homology in XRCC1.

A: Schematic diagram of XRCC1 domains in Arabidopsis and human. B: The GCG program PILEUP (blosum62.cmp program) was used to generate the alignment. Dashes indicate gaps introduced to optimize the alignment. The identical amino acids among 4 out of 5 were shadowed and the conserved BRCT-I domain was boxed. The GenBank accession numbers for each sequence are as follows: Arabidopsis thaliana (AtXRCC1), AF324348, this study: Cricetulus griseus (CgXRCC1), AF034203: Rattus norvegicus (RnXRCC1), AF290895: Homo sapiens (HsXRCC1), M36089: Drosophila melanogaster (DmXRCC1), AF132142.

Table 1. Amino acid sequence homology among XRCC1 from Arabidopsis and other organisms.

Species	Numbers of Amino acids	Sequence identity to AtXRCC1	Sequence similarity to AtXRCC1	Conserved domains
Arabidopsis thaliana	353	_	_	NLS, BRCT-I
Drosophila melanogaster	614	28%	36%	NTD, NLS, BRCT-I
Homo sapiens	633	34%	43%	NTD, NLS, BRCT-I, BRCT-II
Rattus norvegicus	631	30%	39%	NTD, NLS, BRCT-I, BRCT-II
Cricetulus griseus	633	29%	38%	NTD, NLS, BRCT-I, BRCT-II

NTD: N-terminal domain

NLS: Nuclear localization signal

BRCT-I: Breast cancer carboxy-terminal domain I BRCT-II: Breast cancer carboxy-terminal domain II

Nuclear localization signal

The N-terminal sequence of XRCC1 contains a putative nuclear localization signal (NLS) consisting of KKPK in amino acids 26-29, as predicted by the PSORT program (http://psort.nibb.ac.jp/, mainly developed by K. Nakai, University of Tokyo). These residues were classified as a four-residue pattern (pat4) composed of four basic amino acids (K or R), or composed of three basic amino acids (K or R) and either H or P (Hicks et al., 1995).

XRCC1 genomic DNA

Cloning of *XRCC1* genomic DNA from *A. thaliana* (Col) revealed an identical sequence to the genomic sequence previously reported by Arabidopsis Genome Project (GenBank Acc. No. AC018849). Combined analysis by The Genetics Computer Group (GCG) software with the cDNA and genomic sequences revealed that the *XRCC1* gene consists of eight exons, which is clearly different from

the gene organization (four exons) suggested by using several softwares at The Institute for Genomic Research (TIGR). The softwares for gene prediction used by TIGR are Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN. html), GeneMarkHMM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (Mihaela Pertea, http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org).

Through careful combined analysis of the cDNA and genomic DNA in *A. thaliana*, a minor GC-AG splicing site was detected among seven introns (Fig 3). By using this splicing site, the predicted molecular weight of XRCC1 is increased to 39.8kDa (with the minor splicing site) from 26.2kDa (without the site).

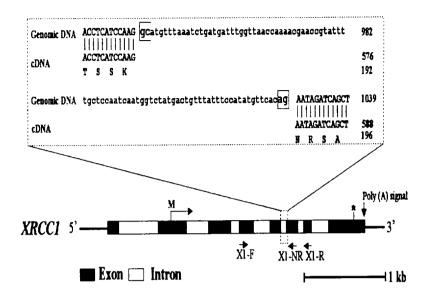
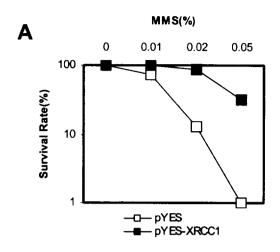


Fig 3. Genomic structure of *XRCC1*.

The start and stop codons are indicated by M or *, respectively. The numbers in Genomic and cDNA sequences are beginning from start codon.

MMS resistance

Methylmethane sulfonate (MMS) induces DNA damage through double strand breaks. Cells do not survive if the damage is not properly or efficiently repaired. We can, therefore, predict that overexpression of DNA repair proteins, such as XRCC1, would result in an increased resistance to DNA damaging MMS. The effect of agents. such as expressing the XRCC1 gene on the resistance to DNA damage was investigated in a wild-type yeast strain (LSY678) bearing a fully-active set of DNA repair genes. results, shown in Fig 4A, indicated that the wild-type yeast strains containing pYES-XRCC1 exhibit enhanced resistance to MMS relative to yeast cells containing the control plasmid pYES2/CT. In complementary experiments, overexpression of pYES-XRCC1 strains deficient in RAD51 or RAD52 function reveal that XRCC1 cannot provide protection against MMS toxicity.



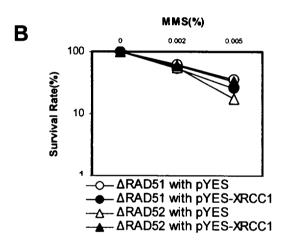


Fig 4. MMS resistance test after overexpression of XRCC1 in yeasts.

Each transformants with pYES-XRCC1 in wild-type (A) and mutant yeasts (B) were measured MMS resistance by survival tests in log scale of Y-axis.

DISCUSSION

We have sequenced and characterized and genomic clones encoding the XRCC1 gene from A. thaliana, a flowering plant. The XRCC1 sequence displays significant homology with other XRCC1 sequences from several animals, including *H. sapiens*, *R. norvegicus*, *C. griseus*, and *D. melanogaster* with a particularly high homology in the BRCT domain. The genomic sequence of *XRCC1* is identical to a previously reported sequence by Arabidopsis Genome Project (Theologis et al., 2000). The results reported here, therefore, confirm and extend the first identification of the *XRCC1* gene from *A. thaliana* (Theologis et al., 2000).

The XRCC1 gene encodes a protein with several important protein domains, including a nuclear localization signal (NLS) and The existence of NLS suggests a BRCT. potential role of XRCC1 in the nucleus or least. for nuclear transport (Schmidt-Zachmann and Nigg, 1993). Although it is possible that a protein without an NLS can enter the nucleus via cotransport with a protein that has NLS, most nuclear proteins such as ribosomal proteins or transcription factors have their own NLSs (Schmidt et al... 1995). Presently, NLSs are classified into two categories (reviewed in Hicks and Raikhel, 1995). The classical type of NLSs. e.g. SV40 large T antigen, is sub-divided into two rules. One rule in this type is a four-residue pattern (pat4) composed of four basic amino acids (K or R), or composed of three basic amino acids (K or R) and either H or P (Hicks et al., 1995). Another rule in this type is a pattern (pat7) starts with P. followed within three residues by a basic segment containing three K/R residues out of 4 residues. And the signature sequence (KKPK) for NLS of At XRCC1 falls into pat4 type in the classical category. Another type of NLS is the type of *Xenopus* nucleoplasmin proposed by Robbins et al. (Robbins et al., 1991). The pattern is: 2 basic residues, 10

residue spacer, and another basic region consisting of at least 3 basic residues out of 5 residues. Many nuclear proteins are generally rich in basic residues: If the sum of K and R compositions are higher than 20%, then the protein is considered to have higher possibility of being nuclear than cytoplasmic (Makkerh et al., 1996).

The highly-conserved BRCT domain is known as a site for protein-protein interaction with such proteins as poly(ADP-ribose) polymerase (PARP) (Caldecott et al., 1996; Masson et al., 1998; Pleschke et al., 2000; Taylor et al., 2002). The deduced amino acid sequence of XRCC1 (353 amino acids) is similar to, but shorter than, other XRCC1 proteins. But the absence of the N-terminal domain (NTD) and C-terminal BRCT-II motif in At XRCC1 suggest a limited role in DNA repair by interaction with DNA polymerase or DNA ligase III. Wild-type yeast strains containing pYES-XRCC1 exhibit enhanced resistance to MMS relative to wild-type cells bearing only the control plasmid pYES2/CT. Furthermore, expression of XRCC1 in mutant strains deficient in RAD51 or RAD52 genes indicates that XRCC1 fails to complement MMS sensitivity in those yeast strains. At present, no yeast homolog for XRCC1 has been found, and our homology-based searches did not reveal any candidate genes. assumed that XRCC1 functions in signaling to recruit DNA repair proteins to DNA such mismatch damage sites. as single-strand break regions (Caldecott et al., 1996: Pleschke et al., 2000)

XRCC1 functions on the pathway in single-strand break repair in animals. The damaged DNA sites mostly on sugar, detected by XRCC1, are repaired firstly by

PARP the damaged termini and are converted to 3'-OH or 5'-P residue by DNA polymerase or polynucleotide kinase. Lastly. the gap is filled by DNA ligase. The function recruiting the key enzymes such as PARP, DNA polymerase, or DNA ligase greatly enhances efficiency in single-strand break repair. AtXRCC1 has only one conserved domain of BRCT-I for interacting PARP which suggest a limited role on DNA repair and existence of additional pathways for singlestrand break repair to survive cells in plants.

Sequence date from this article have been deposited with the Gen Bank data libraries under accession number AF324348.

ACKNOWLEDGMENTS

We wish to thank Arabidopsis Biological Resource Center (ABRC) at the Ohio State University, Columbus for providing an EST clone (210J9T7). This work was supported by a grant from BioGreen 21 Program, Rural Development Administration, Republic of Korea (J.K.).

요 약

X선 상호보상군1 (XRCC1) 은 DNA 손상에 반응하는 DNA 수선 단백질이다. 애기장대로부터 XRCC1을 암호화하는 cDNA 클론이 분리되고 염기서열분석을한 후 cDNA 와 genomic DNA 의 종합 분석의 결과 7개의 인트론으로 구성되어 있음을 발견했다. 또한 XRCC1를 암호화하는 서열은 1062 염기쌍의 길이를 가지며 353 개의 아미노산과 40 kDa 단백질을 암호화한다는 것이 결정되었다. XRCC1 의 예상 아미노산 서

열은 잘 보존된 BRCT 도메인을 함유하고 있고 동물로부터 온 XRCC1 과 유사하나 동물에 비해 특이적으로 작았다. XRCC1 은 야생형 효모에서 MMS에 저항성을 중가시키는 것을 보였으나 RAD51 이나 RAD52가 결여된 돌연변이 효모들은 XRCC1 의 과발현에 의해 회복되지 않았다.

REFERENCES

- Britt AB (1996). DNA damage and repair in plants. Annu. Rev. Plant Physiol. and Plant Mol. Biol. 47: 75-100.
- Caldecott KW, Aoufouchi S, Johnson P, Shall S. (1996) XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. Nucleic Acids Res. 24:43874394.
- D'Alessio JM, Bebee R, Hartley JL, Noon MC, Polayes D (1992). Lambda ZipLox: Automatic subcloning of cDNA. Focus 14: 76-79.
- Hicks GR, Raikhel NV (1995). Protein import into the nucleus: an integrated view. Ann. Rev. Cell Dev. Biol., 11: 155-188.
- Liu L, Cheng S, van Brabant AJ, Kmiec EB. (2002) Rad51p and Rad54p, but not Rad52p, elevate gene repair in Saccharomyces cerevisiae directed by modified single-stranded oligonucleotide vectors. Nucleic Acids Res. 30:2742-2750.
- Liu L. Rice MC, Kmiec EB. (2001) In vivo gene repair of point and frameshift mutations directed by chimeric RNA/DNA oligonucleotides and modified single-stranded oligonucleotides. Nucleic Acids Res. 29: 4238-4250.
- Makkerh JP, Dingwall C, Laskey RA. (1996) Comparative mutagenesis of nuclear

- localization signals reveals the importance of neutral and acidic amino acids. Curr Biol. 6:1025-1027.
- Marintchev A. Robertson A. Dimitriadis EK. Prasad R. Wilson SH. Mullen GP (2000). Domain specific interaction in the XRCC1-DNA polymerase beta complex. Nucleic Acids Res. 28: 2049-2059.
- Masson M, Niedergang C, Schreiber V, Muller S, Menissier-de Murcia J, de Murcia G (1998). XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. Mol. Cell. Biol. 18: 3563-3571.
- Murashige T. Skoog F (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol Plant 15: 473-497.
- Pleschke JM, Kleczkowska HE, Strohm M, Althaus FR. (2000). Poly(ADP-ribose) binds to specific domains in DNA damage checkpoint proteins. J Biol Chem. 275:40974-40980.
- Robbins J. Dilworth SM, Laskey RA, Dingwall C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell. 64:615-623.
- Schmidt C, Lipsius E, Kruppa J. (1995) Nuclear and nucleolar targeting of human ribosomal protein S6. Mol Biol Cell. 6:1875-1885.
- Schmidt-Zachmann MS, Nigg EA. (1993)
 Protein localization to the nucleolus: a search for targeting domains in nucleolin. J Cell Sci. 105:799-806.
- Shen MR, Zdzienicka MZ, Mohrenweiser H, Thompson LH, Thelen MP (1998). Mutations in hamster single-strand break repair gene XRCC1 causing defective DNA repair. Nucleic Acids Res. 26:1032-1037.

- Taylor RM, Moore DJ, Whitehouse J, Johnson P, Caldecott KW (2000). A cell cycle-specific requirement for the XRCC1 BRCT II domain during mammalian DNA strand break repair. Mol. Cell. Biol. 20: 735-740.
- Taylor RM, Thistlethwaite A, Caldecott KW. (2002). Central role for the XRCC1 BRCT I domain in mammalian DNA single-strand break repair. Mol Cell Biol. 22:2556-2563.
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.
- Theologis.A., Ecker, J.R., Palm, C.J., Federspiel, N.A., Kaul, S., White, O., Alonso, J., Altaf, H., Araujo, R., Brooks, S.Y., Bowman, C.L., Buehler, E., Chan, A., Chao, Q., Chen, H., Cheuk, R.F., Chin, C.W., Chung, M.K., Conn, L., Conway, A.B., Conway, A.R., Creasy, T.H., Dewar, K., Dunn, P., Etgu, P., Feldblyum, T.V., Feng, J., Fong, B., Fujii, C.Y., Gill, J.E., Goldsmith, A.D., Haas, B., Hansen, N.F., Hughes, B., Huizar, L., Hunter, J.L., Jenkins.J.. Johnson-Hopson, C., Khan, S., Khaykin, E., Kim, C.J., Koo, H.L., Kremenetskaia.I., Kurtz.D.B., Kwan,A., Lam,B., Langin-Hooper, S., Lee, A., Lee, J.M., Lenz, C.A., Li,J.H., Li,Y., Lin,X., Liu,S.X., Liu,Z.A., Luros, J.S., Maiti, R., Marziali, A., Militscher, J., Miranda, M., Nguyen, M., Nierman, W.C., Osborne, B.I.. Pai, G., Peterson, J., Pham.P.K.. Rizzo, M., Rooney, T., Rowley, D., Sakano, H., Salzberg, S.L., Schwartz, J.R., Shinn.P., Southwick, A.M., Sun, H., Tallon, L.J.,

- Tambunga, G., Toriumi, M.J., Town, C.D., Utterback, T., van Aken, S., Vaysberg, M., Vysotskaia, V.S., Walker, M., Wu, D., Yu, G., Fraser, C.M., Venter, J.C. and Davis, R.W. (2000) Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. Nature 408:816-820.
- Thompson LH, Brookman KW, Jones NJ, Allen SA, Carrano AV (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.
- Whitehouse CJ, Taylor RM, Thistlethwaite A, Zhang H, Karimi-Busheri F, Lasko DD, Weinfeld M, Caldecott KW (2001). XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. Cell. 104:107-117.
- Zhang X, Morera S, Bates PA, Whitehead PC, Coffer AI, Hainbucher K, Nash RA, Sternberg MJ, Lindahl T, Freemont PS (1998). Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. EMBO J. 17:6404-6411.