Articles

Purification, Cloning, and Characterization of the CEL I Nuclease†

Bing Yang,‡ Xiao Wen,‡,§ Nagendra S. Kodali,‡ Catherine A. Oleykowski,‡ C. Glenn Miller,‡ Joanne Kulinski,‡ David Besack,‡ Jason A. Yeung,‡,|| David Kowalski,⊥ and Anthony T. Yeung*,‡

Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, and Roswell Park Cancer Institute, Buffalo, New York 14263

Received October 13, 1999; Revised Manuscript Received December 23, 1999

ABSTRACT: CEL I, isolated from celery, is the first eukaryotic nuclease known that cleaves DNA with high specificity at sites of base-substitution mismatch and DNA distortion. The enzyme requires Mg^{2+} and Zn^{2+} for activity, with a pH optimum at neutral pH. We have purified CEL I 33 000-fold to apparent homogeneity. A key improvement is the use of α-methyl-mannoside in the purification buffers to overcome the aggregation of glycoproteins with endogenous lectins. The SDS gel electrophoresis band for the homogeneous CEL I, with and without the removal of its carbohydrate moieties, was extracted, renatured, and shown to have mismatch cutting specificity. After determination of the amino acid sequence of 28% of the CEL I polypeptide, we cloned the CEL I cDNA. Potential orthologs are nucleases putatively encoded by the genes *BFN1* of *Arabidopsis*, *ZEN1* of *Zinnia*, and *DSA6* of daylily. Homologies of CEL I with S1 and P1 nucleases are much lower. We propose that CEL I exemplifies a new family of neutral pH optimum, magnesium-stimulated, mismatch duplex-recognizing nucleases, within the S1 superfamily.

Nucleases are important to many aspects of cellular functions in all organisms (*1*). Some are highly specialized for DNA recombination, replication, and repair while others are for general nucleic acid degradation. The latter includes mung bean nuclease (MBN)[1] (*2*−*4*), S1 (*5*), P1 (*6*), and the pancreatic DNase I (*7*). However, the biological functions of many nucleases have yet to be revealed.

We recently discovered a novel family of mismatch-specific endonucleases in plants (*8*). These nucleases are abundant in various tissues, including roots, stems, leaves, flowers, and fruits. The nucleases appear to be mannosyl-glycoproteins by their ability to bind to Concanavalin A (ConA). Zn^{2+} was necessary for activity, but Mg^{2+} is also required for efficient DNA nicking at the 3′ side of the mismatch nucleotide. The pH optimum is neutral. One such nuclease from celery, CEL I, was used to develop a fluorescence-based mutation detection assay that is highly effective for insertion/deletion and base-substitution mismatches (*8*). In this report, we describe the purification, cloning, sequence alignment, and characterization of CEL I.

EXPERIMENTAL PROCEDURES

Materials. Plasmid DNA pUC19 was isolated with the QIAGEN Maxi Kit from DH5α host cells, following the manufacturer's instructions. Calf thymus DNA was obtained from Sigma and purified by repeated cycles of proteinase K digestion and phenol extraction (*9*). Chromatography resins and columns were purchased from Pharmacia Biotech. Toluidine Blue O and Ponceau S were from Sigma. Endo H_f was from New England Biolabs. Phosphocellulose P11 was from Whatman.

Purification of CEL I. All steps were performed at 4 °C. The nuclease activity was monitored by using a RF-I (replicative form I) nicking assay (*10*).

Step 1: Preparation of the crude extract. A 105 kg sample of chilled celery stalks was homogenized with a juice extractor. The juice was collected (total 79.34 L) and adjusted to the composition of Buffer A (100 mM Tris-HCl, pH 7.7, 100 μM PMSF). Solid $(NH_4)_2SO_4$ was slowly added to the juice and gently stirred, to a final concentration of 25% saturation. After 30 min, the suspension was centrifuged at 27000*g* for 1.5 h. The supernatant (total 70.56 L) was pooled, and the concentration of $(NH_4)_2SO_4$ was adjusted to 80% saturation. After 30 min of stirring, the mixture was centrifuged at 27000*g* for 2 h. The pellets were resuspended in Buffer B (0.1 M Tris-HCl, pH 7.7, 0.5 M KCl, 100 μM

† This work was supported, in part, by NIH Grant CA71426 and U.S. Army Grant DMAD17-97-1-7286 to A.T.Y., by NIH Grant GM30614 to D.K., by institutional grants from the National Institutes of Health to the Fox Chase Cancer Center (CA06927, RR05539), and by an appropriation from the Commonwealth of Pennsylvania. The DNA sequence(s) reported in this paper has been submitted to GenBank with accession number AF237958.

* To whom correspondence should be addressed at the Fox Chase Cancer Center, 7701 Burholme Ave., Philadelphia, PA 19111. Tel: (215) 728-2488. FAX: (215) 728-3647. Email: AT_Yeung@fccc.edu.

‡ Fox Chase Cancer Center.

§ Current address: Du Pont Pharmaceuticals Co., Wilmington, DE 19880.

|| Current address: University of Pennsylvania, Philadelphia, PA 19104.

⊥ Roswell Park Cancer Institute.

[1] Abbreviations: bp, base pair(s); BSA, bovine serum albumin; CEL I, celery mismatch nuclease I; ConA, Concanavalin A lectin; MBN, mung bean nuclease; nt, nucleotide(s); PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; UTR, untranslated region.

10.1021/bi992376z CCC: $19.00
Published on Web 03/08/2000

PMSF) and thoroughly dialyzed against Buffer B.

Step 2: Concanavalin A−Sepharose 4B affinity chromatography. ConA resin (100 mL) (cross-linked with dimethyl suberimidate) was added to the 7.71 L sample in bottles that were gently rolled overnight. The resin was packed into a 2.5 cm diameter column. The flow-through fraction, containing no CEL I activity, was discarded. CEL I was eluted at 4 °C by 200 mL of Buffer B containing 0.3 M α-methyl-mannoside. The elution step was repeated 10 more times until no more nuclease activity could be eluted. The elutate was combined and dialyzed against Buffer C (50 mM Tris-HCl, pH 8.0, 5 mM α-methyl-mannoside, 0.01% Triton X-100, and 100 μM PMSF).

Step 3: DEAE-Sephacel chromatography. The dialyzed sample from step 2 (total 2.5 L) was applied to a 400 mL DEAE-Sephacel column of 5 cm diameter previously equilibrated with Buffer C. The subsequent steps were performed using FPLC. The column was washed with 400 mL of Buffer C. CEL I was eluted with a 1 L linear gradient of 10 mM to 1 M KCl in Buffer C containing 50 mM α-methyl-mannoside at a flow rate of 5 mL/min, followed by 400 mL of Buffer C containing 1 M KCl and 50 mM α-methyl-mannoside at a flow rate of 8 mL/min. The CEL I enzyme peak eluted between 0.1 and 0.2 M KCl concentration. The most active CEL I fractions were pooled and dialyzed against Buffer D (25 mM potassium phosphate, pH 7.0, 5 mM α-methyl-mannoside, 0.01% Triton X-100, and 100 μM PMSF).

Step 4: Phosphocellulose P-11 chromatography. The dialyzed CEL I pool from step 3 (120 mL) was applied to a 5 cm diameter column packed with 400 mL of P-11 resin. The column was previously equilibrated with Buffer D at a flow rate of 5 mL/min. After sample loading, the column was washed with 625 mL of Buffer D containing 50 mM α-methyl-mannoside at a flow rate of 5 mL/min. CEL I was eluted with a 800 mL linear gradient of 20 mM KCl to 1 M KCl in Buffer D containing 50 mM α-methyl-mannoside at a flow rate of 5 mL/min. The column was further washed with 400 mL of Buffer D containing 1 M KCl and 50 mM α-methyl-mannoside at a flow rate of 8 mL/min. The CEL I enzyme peak eluted as a broad peak between 0.3 and 0.8 M KCl concentration. The most active fractions were pooled and dialyzed against Buffer E (50 mM potassium phosphate, pH 7.0, 5 mM α-methyl-mannoside, 0.01% Triton X-100, and 100 μM PMSF) containing 1.5 M $(NH_4)_2SO_4$.

Step 5: Phenyl Sepharose CL-4B chromatography. The dialyzed CEL I pool from step 4 (480 mL) was applied to a 5 cm diameter column packed with 400 mL of Phenyl Sepharose CL-4B. The column was previously equilibrated with Buffer E containing 1.5 M $(NH_4)_2SO_4$ at a flow rate of 5 mL/min. After sample application, the column was washed with 400 mL of Buffer E containing 1.5 M $(NH_4)_2SO_4$ and 50 mM α-methyl-mannoside at a flow rate of 5 mL/min. CEL I was eluted from the column with a 500 mL linear reversed salt gradient from 1.5 to 0 M $(NH_4)_2SO_4$ in Buffer E containing 50 mM α-methyl-mannoside at a flow rate of 5 mL/min. The CEL I enzyme peak eluted at about 133 mM $(NH_4)_2SO_4$ concentration. The most active fractions were pooled and dialyzed against Buffer F (50 mM Tris-HCl, pH 8.0, 5 mM α-methyl-mannoside, 0.01% Triton X-100, and 100 μM PMSF).

Step 6: Mono Q anion-exchange chromatography. A Pharmacia prepacked Mono Q HR 16/10 column was thoroughly washed and equilibrated with Buffer F. The dialyzed CEL I pool from step 5 (336 mL) was applied at a flow rate of 5 mL/min followed by 100 mL of Buffer F containing 50 mM α-methyl-mannoside at a flow rate of 10 mL/min. CEL I was eluted with a 250 mL linear gradient of 0−1 M KCl in Buffer F containing 50 mM α-methyl-mannoside at 2 mL/min. CEL I enzyme activity was eluted between 0.1 and 0.3 M KCl concentration.

Step 7: Superdex 75 size-exclusion chromatography using the SMART system. The active fractions of step 6, fractions 11 and 12, were combined and concentrated by using Centricon 3 centrifugal concentrators. Aliquots of the concentrated enzyme were applied to a prepacked Superdex 75 PC 3.2/30 column equilibrated with Buffer G (50 mM Tris-HCl, pH 8.0, 100 mM KCl, 10 μM $ZnCl_2$, 0.01% Triton X-100, and 100 μM PMSF) containing 50 mM α-methyl-mannoside. Five milliliters of Buffer G containing 50 mM α-methyl-mannoside was used to elute CEL I at a flow rate of 0.05 mL/min. CEL I enzyme activity elutes at its monomer position on this column. The purity of the active fractions was checked by SDS−PAGE. When additional protein bands were present, the fractions were pooled, concentrated, and purified again using the same size-exclusion chromatography until CEL I reached apparent homogeneity.

SDS−Polyacrylamide Gel Electrophoresis (SDS−PAGE). Polyacrylamide gel electrophoresis in SDS was carried out as previously described (*11*). Protein bands were detected by using the Gelcode Blue Stain Reagent (Pierce). Molecular weights of the protein bands were determined by using the semilogarithmic plot of the molecular weights of protein standards versus their relative electrophoretic mobilities. Activity gel assay was performed essentially as described (*12, 13*).

Endo H_f Removal of N-Linked Oligosaccharides from CEL I. CEL I sample was denatured in 0.5% SDS at 100 °C for 10 min. An appropriate amount of Endo H_f was added, and the reaction was incubated in G5 buffer (50 mM sodium citrate, pH 5.5) at 37 °C overnight.

Renaturation of CEL I from SDS−PAGE. This method is a modification of a procedure previously described (*13*, *14*). The CEL I fractions were loaded onto the SDS−PAGE in two consecutive lanes. After electrophoresis, the gel was split between the two lanes. Half of the gel was stained with Gelcode Blue Stain Reagent (Pierce) and then aligned with the other half that was not stained. The gel slice corresponding to the CEL I band in the unstained gel was excised and eluted using an AMICON model 57005 electroeluter, for 2 h at 20 mA per sample, using the elution buffer (50 mM Tris-HCl, pH 7.5, 180 mM NaCl, 0.1% SDS, 0.1 mg/mL bovine serum albumin, BSA). After elution, the sample was concentrated by using a Centricon 3 unit. Centrifugation was overnight at 7000*g*. The volume of the sample was measured, and 4 volumes of distilled acetone (−20 °C) was added. The sample was incubated in a dry ice−ethanol bath for 30 min and then centrifuged at 14000*g* for 10 min. The precipitated proteins were washed with a buffer consisting of 20% Dilution and Renaturation Solution (50 mM Tris-HCl, pH 7.5, 10% glycerol, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 2 μM $ZnCl_2$, and 0.1 mg/mL BSA) and 80% acetone. The sample was precipitated again at 14000*g* for 10 min.

The supernatant was discarded. The residual acetone was decanted by inverting the tube for 10 min. The pellet was air-dried for at least 10 min. Twenty microliters of Renaturation Solution (6 M guanidine hydrochloride, 50 mM Tris-HCl, pH 7.5, 10% glycerol, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 2 μM $ZnCl_2$, and 0.1 mg/mL BSA) was then used to dissolve the pellet. After 20 min of incubation at room temperature, 1 mL of Dilution and Renaturation Solution was added, and the protein was further renatured at room temperature for 12 h.

Preparation of the CEL I Sample for Sequencing. The purified CEL I sample was subjected to 10% SDS−PAGE analysis. After electrophoresis, the protein in the gel was electrophoretically transferred to an Immobilon-PSQ PVDF membrane by using a Western transfer apparatus (Novex). The transfer buffer contained 12 mM Trizma base, 96 mM glycine, and 20% methanol. The transfer condition was 1 h at 25 V (constant voltage). The membrane was next washed extensively with water, and stained with Ponceau S. The CEL I band was excised, destained with water, and sent to the Protein/DNA Technology Center of Rockefeller University for N-terminal and internal peptide micro-sequencing by automated Edman degradation reaction. The N-terminal sequence was determined first (*15*). The remaining protein fractions were digested with either trypsin or GluC. The digested peptides were purified by HPLC, and sequenced with Edman degradation (*16*).

Cloning of the cDNA of the CEL I mRNA. Total RNA was prepared from fresh celery using the phenol SDS method for plant RNA preparation (*17*). First-strand cDNA was synthesized using Stratagene's ProStar First Strand RT-PCR kit. Degenerate PCR primers were chosen from the amino acid sequences determined by Edman degradation analysis of the pure CEL I protein, and used to amplify the CEL I cDNA in two segments, using the AmpliTaq DNA polymerase, and cloned in *E. coli* for DNA sequencing. The two fragments provided most of the reading frame of the CEL I protein. Using 5′ and 3′ RACE methods (Clonetech Marathon cDNA amplification kit), the 5′ and 3′ coding regions and untranslated regions (UTR) of CEL I cDNA were obtained. To confirm the authenticity of the cDNA, two PCR primers were designed: one in the 5′ UTR, and one in the 3′ UTR. These two primers were used to amplify the CEL I cDNA as one fragment from a fresh preparation of celery RNA, using the high-fidelity Pfu DNA polymerase for the amplification. The new sequence was cloned in *E. coli*. The DNA sequence confirmed the authenticity of the previous cDNA sequence except for one nucleotide difference that gives another codon for the same amino acid.

Sources of Mung Bean Nuclease. MBN was purchased from Pharmacia Biotech, #27-0912, herein called 'MBN-A', or purified as previously described (*18*), herein called 'MBN-B'. MBN assay conditions and the measurement of protein concentrations vary in different laboratories and may partially influence the quantitation in this study. MBN-A is FPLC-purified and homogeneous, with a specific activity of 1.64×10^6 units/mg in the manufacturer's assay conditions, but 1.42×10^6 units/mg in our assay conditions. The enzyme exhibits a single band in SDS−PAGE. MBN-B is an older preparation of the original MBN of Kowalski and has a specific activity of 4×10^5 units/mg in the assay conditions used in this report. The enzyme appeared as a single band of about 39 kDa on nonreducing SDS−PAGE (data not shown). One unit of MBN-A single-strand DNase activity equals 0.7 ng of enzyme in our assay.

RF-I Nicking Assay. A 1.1 μg aliquot of a 2.9 kbp pUC19 plasmid derivative was incubated with the designated amount of MBN or CEL I for 30 min at 37 °C in a volume of 30 μL of Buffer H (20 mM sodium acetate, pH 5.5, 10 mM KCl) or Buffer I (20 mM HEPES, pH 7.5, 10 mM KCl) in the presence or absence of 3 mM $MgCl_2$. To stop the reaction, 5 μL of stop solution (50 mM Tris-HCl, pH 6.8, 3% SDS, 4.5% β-mercaptoethanol, 30% glycerol, and 0.001% Bromophenol Blue) was added. Then 24 μL of the final mixture was loaded onto a 0.8% agarose gel. After electrophoresis and staining with ethidium bromide, a photograph of the gel was taken, and the negative was scanned using the IS-1000 Digital Imaging System (Alpha Innotech Corp.). The RF-I band was quantified using IS-1000 v2.02 software.

Single-Strand DNase Assay. The DNA solubilization assay was similar to that previously described (*19*). Fifty micrograms of heat-denatured calf thymus DNA (Calbiochem #2618, purified by repeated Pronase treatment, phenol extraction, and dialysis) was incubated with 0.7 ng of MBN-A, or 1.9 ng of MBN-B, or 16 ng of CEL I, in 100 μL of Buffer H or Buffer I, with or without 3 mM $MgCl_2$. At the designated times, 100 μL of cold 20 mM $LaCl_3$ in 0.2 N HCl was added to stop the reaction. After centrifugation (21000*g*, 40 min), the absorbance at 260 nm of the supernatant was measured using a spectrophotometer to determine the amount of DNA that had become acid-soluble.

Mismatch Endonuclease Assay. The mismatch endonuclease assay was performed as previously described (*8*). Briefly, PCR products were amplified using genomic DNA of individuals that are heterozygous for certain alterations in three different exons in the *BRCA1* gene. The forward primer was 5′-labeled with 6-FAM (blue), and the reverse primer was 5′-labeled with TET (green). The locations of the mismatches in the *BRCA1* gene are the 300, 4184, 4421, and 5382 nt positions. They correspond to a T → G base substitution in exon 5, a 4 base deletion in exon 11, a C → T polymorphism in exon 13, and a C insertion in exon 20, respectively. The four resulting heteroduplexes provide a 235 bp PCR product containing a T/C or a G/A base-substitution mismatch, a 387 bp PCR product containing a 4 base loop, a 323 bp product containing either a C/A or a T/G base-substitution mismatch, and a 402 bp product containing an extrahelical C or an extrahelical G. Fifty nanograms of the fluorescently labeled heteroduplex was incubated with 7 ng of MBN-A, or 11 ng of MBN-B, or 10 pg of CEL I (0.3 unit) for 30 min at 37 or 45 °C in a reaction volume of 20 μL in Buffer I in the presence or absence of 3 mM $MgCl_2$. The reactions were processed as described (*8*), loaded onto a denaturing 34 cm well-to-read 6% polyacrylamide gel on an ABI 377 Sequencer, and analyzed using GeneScan 3.1 software (Perkin-Elmer). The results can be displayed as a gel image or as an electropherogram (a display of the peak profile of each lane of the gel image).

RESULTS

Purification of CEL I. CEL I was purified to homogeneity, more than 33 000-fold over its specific activity in the buffered celery juice. Table 1 summarizes the purification

Table 1: Purification of CEL I from Celery[a]

purification step	volume (L)	total protein (mg)	total act. (CEL I units)	sp act. (units/mg)	protein (*x*-fold purification)
buffered juice	79.34	19399	1.9×10^7	9.7×10^2	
25% $(NH_4)_2SO_4$ supernatant	70.56	17005	1.6×10^7	9.2×10^2	1
80% $(NH_4)_2SO_4$ pellet	8	2072	9.0×10^6	4.4×10^3	4.5
ConA−Sepharose 4B	2.5	6.75	3.6×10^6	5.4×10^5	553.8
DEAE-Sephacel	0.12	2.69	2.4×10^6	8.8×10^5	907.6
phosphocellulose P-11	0.48	0.408	1.5×10^6	3.8×10^6	3854
Phenol Sepharose CL-4B	0.34	0.054	5.6×10^5	1.0×10^7	10676
Mono Q	0.03	0.03	3.6×10^5	1.2×10^7	12316
Superdex 75	0.0005	0.005	3.1×10^5	3.1×10^7	33000

[a] The fold purification is calculated with respect to the activity measured in the 25% ammonium sulfate supernatant fraction because assays of the crude extract are inaccurate. The fractions from the Mono Q step were purified individually by the SMART system Superose 75 size-exclusion chromatography, and the total yield of the Superose 75 column chromatography cycles is reported. Protein determination was performed using the Bicinchoninic acid protein assay (Pierce). One unit of CEL I is defined as one-thousandth of one single-strand nuclease unit. One unit of single-strand nuclease activity is defined as the amount of enzyme (32 ng of CEL I) that produces 1 μg of acid-soluble material at pH 5.5 in 1 min at 37 °C in the absence of Mg^{2+} when purified sheared single-stranded calf thymus DNA is used as substrate. Homogeneous CEL I from the Superose 75 size-exclusion chromatography has a specific activity of 3.1×10^7 CEL I units/mg, at 33 000-fold purification. 270 units of CEL I will nick 1 μg of supercoiled pUC19 plasmid RF-I at pH 7.5 in 30 min at 37 °C. The pUC19 RF-I nicking assay was used to quantify CEL I throughout purification. That this assay reflects the proper location of CEL I was confirmed by mismatch incision assays. In contrast, the mismatch specificity of CEL I is evident in that 0.3−0.6 unit (10−20 pg) of CEL I is optimal in a CEL I mutation detection Genescan assay.

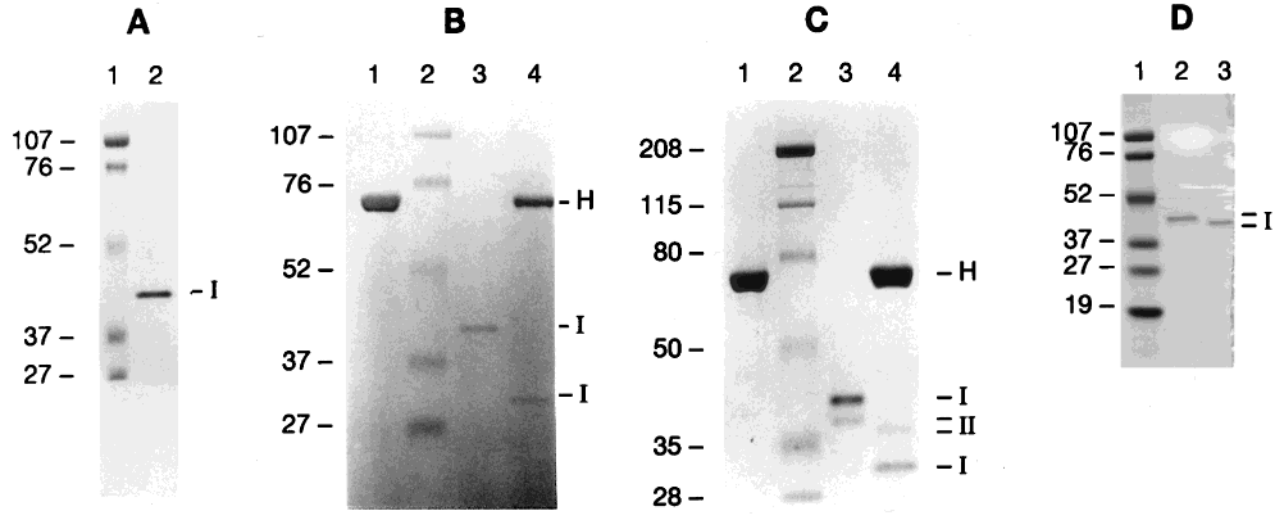

FIGURE 1: SDS−polyacrylamide gel analysis of purified CEL I and CEL II. The gels were stained with Gelcode Blue. (A) Lane 1, molecular mass standards shown in kilodaltons on the side. Lane 2, 1 μg of homogeneous CEL I enzyme. Panels B and C examine the mobility changes in the CEL I and CEL II protein bands due to Endo H_f treatment. Samples in panel B contain only CEL I. Samples in panel C contain a mixture of CEL I and CEL II. Panel D shows the mobility change of homogeneous CEL I after sulfhydryl reduction. (B) Lane 1, Endo H_f. Lane 2, molecular mass standards. Lane 3, homogeneous CEL I, about 30 ng. Lane 4, CEL I digested with Endo H_f. (C) Lane 1, Endo H_f. Lane 2, molecular mass standards. Lane 3, purified CEL I with a small amount of CEL II. Lane 4, CEL I and CEL II digested with Endo H_f. (D) Purified CEL I was boiled for 2 min in SDS sample buffer in the presence (lane 2) or absence (lane 3) of 1% β-mercaptoethanol. Lane 1, molecular mass standards. H = Endo H_f, I = CEL I, II = CEL II.

of CEL I from 105 kg of celery stalks. The active band of CEL I is of the same size throughout purification as judged by an activity gel assay (data shown as Figure 1 in the Supporting Information). There are two nuclease bands that copurify during all the purification steps. We show below that the minor band is not derived from the major band. The major nuclease activity, designated CEL I, migrates at 43 kDa on SDS−PAGE (Figure 1A). The minor activity at 39 kDa is a putative isozyme we named CEL II (Figure 1C, lane 3), also capable of cutting at mismatches.

Isoelectric Points of CEL I and CEL II. A sample of CEL I, containing a small amount of CEL II, was loaded onto an isoelectric focusing gel (pH 3−10, from Novex). After the gel was stained, the p*I* values of CEL I and CEL II were obtained by comparison with the standards (Bio-Rad). The p*I* of the CEL I band was between 6.0 and 6.5, and the p*I* of the CEL II band was between 6.5 and 6.8 (data not shown). After minimizing the N-linked oligosaccharides by Endo H_f, the 43 kDa major celery nuclease band shifted to the 29 kDa position (Figure 1B,C, lanes 4), and the 39 kDa minor celery nuclease band shifted to the 37 kDa position (Figure 1C, lane 4). If CEL II were a degradation product of CEL I, after Endo H_f treatment its polypeptide length should be equal or less than 29 kDa.

Effects of Reducing Agents on CEL I. When 1% β-mercaptoethanol was used in the sample buffer for SDS−PAGE analysis of the CEL I band, CEL I was shifted upward (Figure 1D, lane 2) but intact. DTT was also tested, and similar results were obtained (data not shown). The simplest interpretation is that the CEL I polypeptide does not contain any breakage in the backbone. Instead, disulfide bonds were broken that resulted in the enzyme becoming more extended

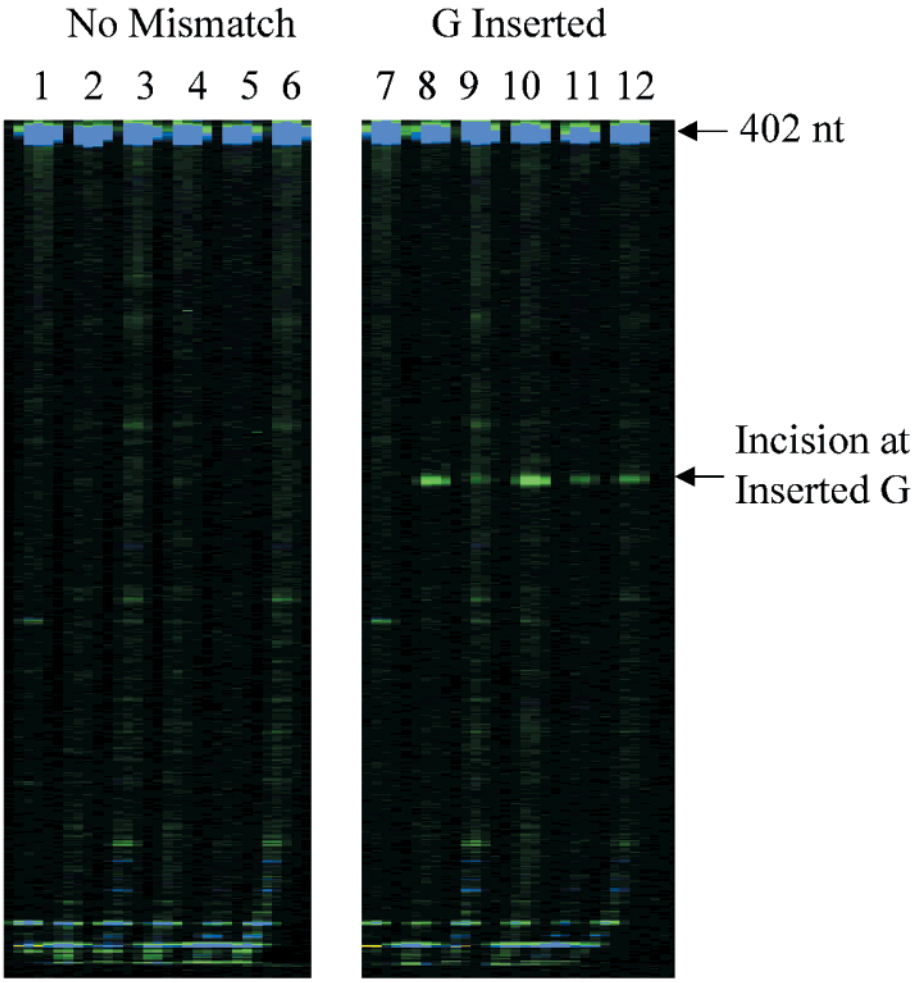

FIGURE 2: Incision at mismatch substrate by CEL I and CEL II proteins renatured from the SDS gel, before and after removal of the carbohydrate moieties. CEL I and CEL II protein bands were excised from a SDS gel and renatured as described under Experimental Procedures. The renatured enzyme was used to digest a 402 bp fluorescently labeled PCR product of exon 20 of the *BRCA1* gene. Lanes 1−6 are homoduplexes made from wild-type DNA samples containing no mismatch in exon 20. Lanes 7−12, because of the heterozygous nature of this sequence in the sample, the PCR product is a heteroduplex in which one strand contains a G residue insertion. CEL I incision at the 3′ side of this extrahelical G residue produces a green band as indicated in the figure. Lanes 1 and 7, substrate with no CEL I treatment. Lanes 2 and 8, incision of the substrate by purified native CEL I. Lanes 3 and 9, incision of the substrate by the renatured 29 kDa CEL I polypeptide band originated from Endo H_f digestion of the 43 kDa CEL I band. Lanes 4 and 10, incision of the substrate by the renatured 37 kDa CEL II polypeptide band originated from Endo H_f digestion of the 39 kDa CEL II band. Lanes 5, 6, 11, and 12, incision of the substrate by the renatured 43 kDa CEL I band.

in the reduced state, and hence slower in electrophoretic mobility.

Renaturation of Homogeneous CEL I and CEL II. Individual celery nuclease bands were excised from the 10% SDS−PAGE and eluted as described under Experimental Procedures. These bands included the 43 kDa band, the 39 kDa band, and their corresponding bands after Endo H_f digestion. The eluted enzyme fractions were concentrated and renatured. Plasmid nicking assays were carried out to show that the renatured samples were all active nucleases (data not shown). The renatured CEL I before or after Endo H_f digestion and CEL II after Endo H_f digestion were able to incise DNA at a mismatch substrate (Figure 2). In this experiment, the mismatch incised is a G residue insertion. This experiment is necessarily qualitative because of the uncertainties in the recovery of proteins and activity in the gel elution and renaturation steps. However, the data strengthen the conclusion that CEL I and CEL II are homogeneous and each able to incise at a DNA mismatch, and that most of the carbohydrates on CEL I and CEL II are not essential for activity.

The Cloning of CEL I cDNA. The amino acid sequences of the N-terminal and three other internal proteolytic peptides of CEL I, identified by Edman degradation performed by the Protein/DNA Technology Center of the Rockefeller University, are shown in Figure 3 in boldface letters. The 72 amino acids identified represent about 28% of the CEL I polypeptide and were completely accounted for in the cDNA sequence. CEL I without the leader sequence is a protein of 274 amino acid residues, with a calculated molecular weight of 31 440.2. Compared with the apparent molecular mass of 43 kDa determined by SDS−PAGE, CEL I is 27% carbohydrate by weight.

Alignment of the CEL I cDNA with homologues in GenBank by the PSI-Blast program at NCBI (*20*) revealed that CEL I has relatively low identity to the *Aspergillus* S1 nuclease (accession number P24021, 27% of 273 amino acids) and the P1 nuclease (accession number P24289, 30% of 277 amino acids) (data not shown). However, among all the homologues of CEL I in plants, three stand out to be of a very high degree of identity, namely, ZEN1 (accession number AB003131, 80% identity of 269 amino acids), DSA6

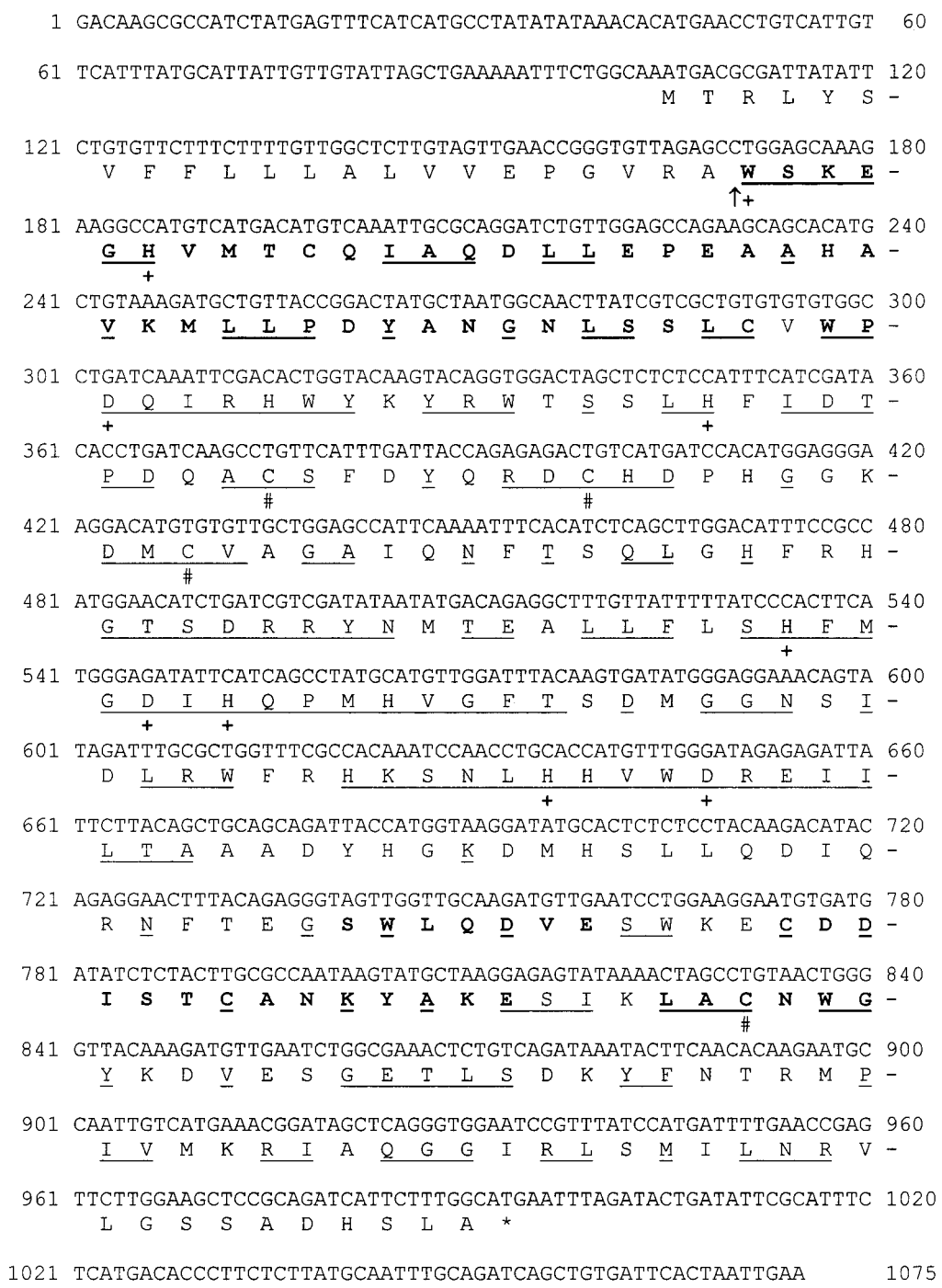

FIGURE 3: cDNA sequence of the CEL I mRNA. The amino acid sequences of CEL I determined by Edman degradation are shown in boldface. They consist of the N-terminal sequence: WSKEGHVMTCQIAQDLLEPEAAHAVKMLLPDYANGXLSSLXVWP; internal peptide from GluC digest: XSWLQDVE; internal peptides from tryptic digest: CDDISTCANKYAKE and LACNWGYK. The residues identical with DSA6, BFN1, and ZEN1 are underlined. The conserved Cys residues are shown with # underneath. The nine conserved residues shown to be ligands for the three Zn atoms in P1 nuclease are shown with + underneath. Nucleotides 1−37 are vector sequence.

(accession number AF082031, 73% identity of 271 amino acids), and BFN1 (accession number U90264, 72% identity of 274 amino acids). We propose that these three proteins are probably orthologs of CEL I because all other homologues are at a range of 45% identity or lower. Moreover, when one superimposes the sequences of CEL I and these three orthologs on the secondary structure of P1 nuclease, most of the sequence differences among these four putative orthologs are in the flexible loop regions that connect consecutive helices (Figure 2 of Supporting Information) and in the flexible COOH-terminal region. Thus, it is very likely that these orthologs share the enzymatic properties of CEL I and may not share the catalytic properties and substrate specificity of the S1 nuclease.

Mg^{2+} and pH Dependence of CEL I. A gel image of the automated DNA sequencer analysis of the CEL I incision at the mismatch of a T → G base substitution is shown in Figure 4. Lanes 1−4 are mock reactions without CEL I. The full-length 235 nt PCR product is seen on top of the image, and imperfect PCR products are seen as the bands dispersed below. In lane 5, in the presence of CEL I, Mg^{2+}, and pH 7.5, the blue incision band of 156 nt and the green incision

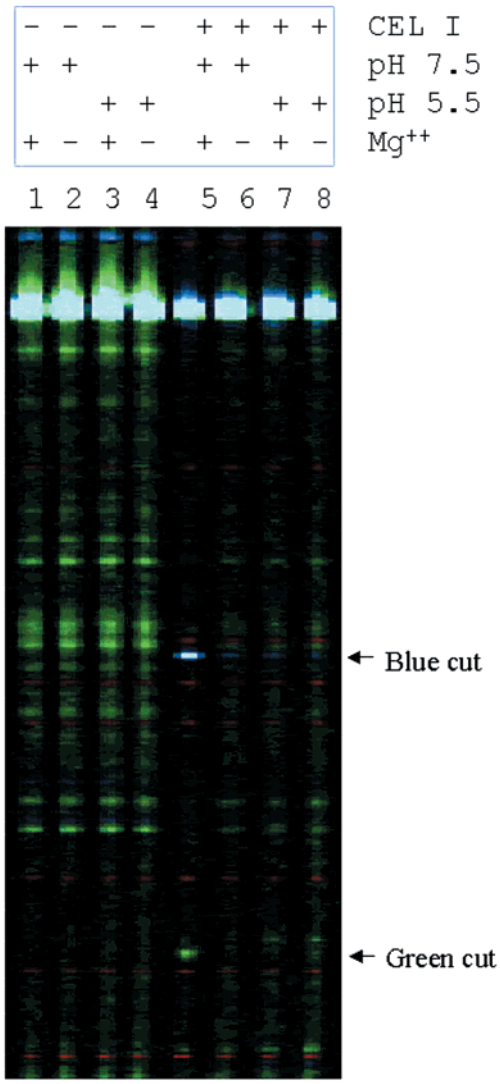

FIGURE 4: Effects of Mg^{2+} and pH on CEL I mutation detection. The picture is a gel image of mutation detection analyses on a Perkin-Elmer automated DNA sequencer running the GeneScan program. The substrate is a 235 bp PCR product of the *BRCA1* gene exon 5 containing a T → G polymorphism. It is labeled at the 5′ terminal with 6-FAM (Blue) in the top strand and with TET (Green) on the bottom strand. The substrates were incubated with 0.5 unit of CEL I for 30 min at 45 °C and then analyzed as described in Figure 5. In lane 5, the blue band at 156 nt corresponds to CEL I mismatch-specific cutting on the 6-FAM-labeled strand, and the green peak at 80 nt corresponds to the mismatch-specific cutting on the TET-labeled strand. The red bands in a gel image are the internal size standards in each lane.

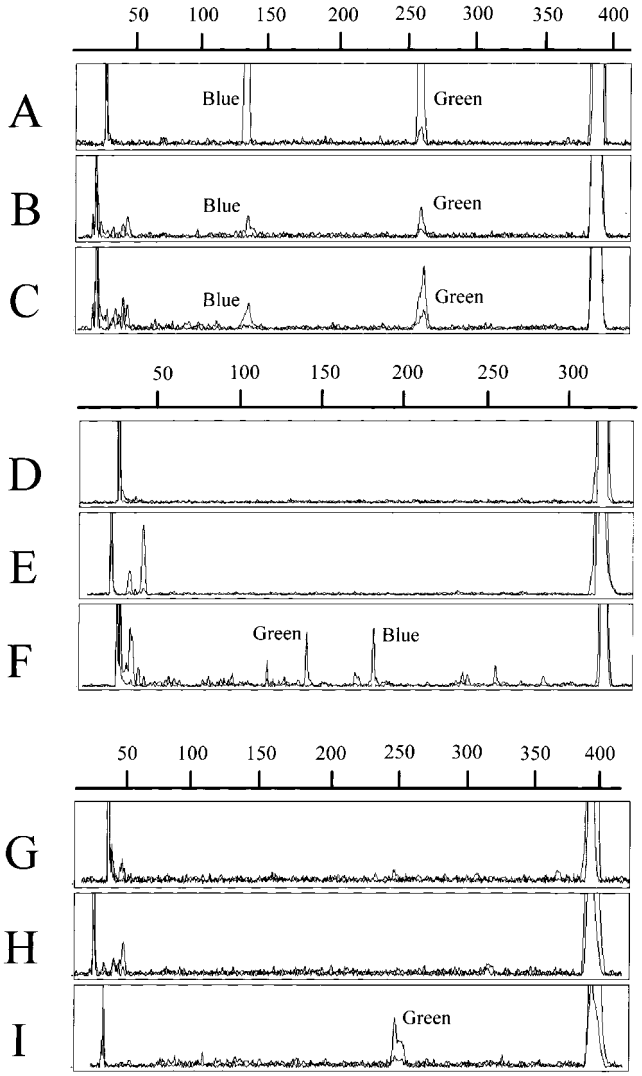

FIGURE 5: Mismatch detection by CEL I and MBN. Electropherograms of Genescan fragment analysis on a PE-Biosystems automated DNA sequencer are shown. Two color fluorescent heteroduplexes of PCR products of the *BRCA1* gene were prepared as described under Experimental Procedures. Vertical axis, relative fluorescence units; horizontal axis, DNA length in nucleotides. Panels A, D, and G, the DNA was incubated with 7 ng of MBN-A. Panels B, E, and H, the DNA was incubated with 11 ng of MBN-B. Panels C, F, and I, the DNA was incubated with 10 pg of CEL I. These reactions were performed in Buffer I with 3 mM $MgCl_2$ for 30 min at 37 °C. In panels A, B, and C, the substrate was a 387 bp heteroduplex containing a 4 nt deletion. In panels D, E, and F, the substrate was a 323 bp product containing a C → T base substitution mismatch. In panels G, H, and I, the substrate was a 402 bp heteroduplex containing a C insertion in one strand. In each of panels A, B, and C, the blue peak at 129 nt corresponds to cutting at the 4 base insertion on the 6-FAM-labeled strand; the green peak at 258 nt corresponds to cutting at the 4 base insertion on the TET-labeled strand. In panels D, E, G, and H, no mismatch-specific cutting is seen by the two MBNs. In panel F, the blue peak at 183 nt corresponds to CEL I-mismatch-specific cutting on the 6-FAM-labeled strand, and the green peak at 142 nt corresponds to the mismatch-specific cutting on the TET-labeled strand. In panel I, the green peak at 252 nt corresponds to the CEL I-specific cutting at the extrahelical G on the TET-labeled strand.

band of 80 nt are observed as indicated. In the absence of Mg^{2+} or in pH 5.5 (lanes 6−8), mismatch-specific incisions are not significant. This experiment also illustrates how the imperfect PCR byproducts seen in lanes 1−4 are eliminated by CEL I in lanes 5−8, especially under the conditions of lane 5.

The Mismatch Endonuclease Activity of CEL I and MBN. The nicking of DNA duplexes containing mismatches by MBN and CEL I is shown in Figure 5. The mismatch with a four base loop is nicked by CEL I and both preparations of MBN at pH 7.5 (A−C). Note the higher amounts of MBN needed in this reaction. However, even at 1000 times more enzyme than CEL I, MBN is unable to specifically nick at base-substitutions at a single base mismatch (D, E, G, and H). When the same amount of MBN protein is incubated with DNA substrates at pH 5.5 as at pH 7.5, the substrate is almost completely digested (data not shown). When a lesser, more appropriate amount of MBN is incubated with the DNA substrate at pH 5.5, no mismatch-specific nicking is seen (data not shown). CEL I nicks at the base-substitution

mismatch (panel F) and at the extrahelical nucleotide (panel I). In panel F, the blue peak at position 183 nt corresponds to the nick at the 3′ side of the mismatch on the 6-FAM-labeled strand of the heteroduplex, and the green peak at position 142 nt corresponds to the nick at the 3′ side of the mismatch on the TET-labeled strand. Some of the other blue peaks are nonspecific cutting by CEL I; it is important to note that if one incubates the reaction for a longer time, or with more CEL I enzyme, most of these non-mismatch-specific peaks will be removed while the mismatch-specific peaks will remain (Figure 4). The reason is that these background bands are often nonspecific heteroduplexes of PCR products in which the two DNA strands do not base-pair properly. These duplexes are nicked by CEL I at nonspecific positions, and their signal becomes diffused. In panel I, the green peak at 252 nt corresponds to the nick at the 3′ side of the extrahelical G on the TET-labeled strand of the PCR product. A blue peak corresponding to the nick at the extrahelical C on the 6-FAM-labeled strand is expected at position 151 nt, but is not seen. CEL I may have nicked the 6-FAM-labeled strand near its 5′ end removing the dye, making it unable to score the blue peak in the assay. Alternatively, the insert C substrate may have been out-competed by the insert G substrate.

DISCUSSION

Purification of Glycoproteins. We previously described a purification protocol that produced highly enriched CEL I, but was unable to provide the enzyme as a single band on a SDS−PAGE gel (*8*). To identify the source of contamination, we repeated the purification with *Arabidopsis* callus, and observed the same problem of aggregation. We made mouse antibodies to the purest fraction and used the antiserum to identified clones of two different genes from an *Arabidopsis* cDNA expression library (*21*) (accession number AC001645, genes PID: g2062157 and PID: g2062159) (unpublished data). These clones were found to be highly homologous to two jasmonate inducible proteins of *Brassica napus* that are known to function as ConA-like lectins (accession number CAA72271, 62% identity in 475 amino acids) (*22*). Such lectins are coded for by over 30 genes in *Arabidopsis* and can be a problem when the glycoprotein to be purified is less abundant than the lectins. The presence of mannose in the buffers in the present protocol has overcome this obstacle and has provided a homogeneous preparation of CEL I.

Alignment of CEL I cDNA. In an alignment of CEL I mRNA with all the S1 homologues in GenBank (data not shown), the universally conserved residues are the N-terminal tryptophan residue, five histidine residues, and three aspartate residues, located in different regions of the polypeptide (Figure 3). These nine residues are brought together to bind the three Zn^{2+} atoms, as revealed by the X-ray crystallography structure of the P1 nuclease (*23*−*24*). The conservation of the catalytic active site suggests that these nucleases share the same mechanism for the cleavage of the phosphodiester bonds, necessitating the conservation of the enzyme structure to form the catalytic domain. The differences in substrate preference may lie in the mechanism of substrate recognition, separate from catalysis, such that S1 family nucleases are specific for single-stranded nucleic acids whereas CEL I shows high specificity for mismatch heteroduplexes. The sequences that enable the recognition of different substrates may reside in amino acid sequences that are less conserved.

Properties of CEL I. The results of the renaturation experiments described in Figures 1 and 2 suggest that the carbohydrate moieties of CEL I are not essential for its mismatch incision activity or for its renaturation from a SDS gel. If CEL I and S1 should use the same catalytic mechanism for phosphodiester bond cleavage, their differences may lie in how the substrates are recognized. Mg^{2+} is required for optimal CEL I incision at mismatches in double-stranded DNA (Figure 4). The role of Mg^{2+} may lie in a structural role for substrate recognition and not in DNA hydrolysis (*25*). As judged by the high specific activity of CEL I on mismatch substrates in comparison to MBN (Figure 5), it is clear that CEL I is different from the S1 family of nucleases.

The high-resolution X-ray structure of the P1 nuclease showed that a double-stranded helix cannot fit into the P1 DNA binding grove (*21*, *22*). It will be interesting to see how CEL I is able to accommodate the double-stranded DNA structure and incise a mismatch in the helix without making single-stranded DNA a preferred substrate. Moreover, that the spinach ortholog is unable to recognize guanine residues in mismatches (*26*) may provide insights toward the mechanism of nucleotide recognition. The mRNA of the homologue *DSA6* in daylily is induced by 111-fold during senescence (*27*). Yet CEL I and its orthologs in many plants are constitutively present in all tissues tested. The roles and mechanisms of this unique enzyme family are yet to be revealed.

ACKNOWLEDGMENT

We thank Drs. Bellacosa and Matsumoto for critical reading of the manuscript, and Dr. Joseph Fernandez for lending us his expertise in micro-sequencing. Protein sequence determination was performed by the Protein/DNA Technology Center of the Rockefeller University.

SUPPORTING INFORMATION AVAILABLE

Figure 1 is polyacrylamide activity gel analysis of the CEL I purification fractions. Figure 2 is ClustalW alignment of the CEL I cDNA sequence with homologous sequences (4 pages). This material is available free of charge via the Internet at http://pubs.acs.org

REFERENCES

1. Linn, S. M., Lloyd, R. S., and Roberts, R. J., Eds. (1993) in *Nucleases*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Kowalski, D., Kroeker, W. D., and Laskowski, M., Sr. (1976) *Biochemistry 15*, 4457−4462.
3. Sung, S., and Laskowski, M., Sr. (1962) *J. Biol. Chem. 237*, 506−511
4. Kowalski, D., Natale, D. A., and Eddy, M. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9464−9468.
5. Shank, T. E., Rhodes, C., Rigby, P. W. J., and Berg, P. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 989−993.
6. Maekawa, K., Tsunasawa, S., Dibo, G., and Sakiyama, F. (1991) *Eur. J. Biochem. 200*, 651−661.
7. Lacks, S. A. (1981) *J. Biol. Chem. 256*, 2644−2648.
8. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K., and Yeung, A. T. (1998) *Nucleic Acids Res. 26*, 4597−4602.

9. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
10. Yeung, A. T., Mattes, W. B., Oh, E. Y., and Grossman, L. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6157−6161.
11. Laemmli, U. K. (1970) *Nature 227*, 680−685.
12. Blank, A., Silber, J. R., Thelen, M. P., and Dekker, C. A. (1983) *Anal. Biochem. 135*, 423−430.
13. Hager, D. A., and Burgess, R. R. (1980) *Anal. Biochem. 109*, 76−86.
14. Kennedy, J. F., and Robertson, E. R. (1996) *Bioseparation 6*, 1−15.
15. Fernandez, J., Gharahdaghi, F., and Mische, S. M. (1998) *Electrophoresis 19*, 1036−1045.
16. Fernandez, J., Andrews, L., and Mische, S. M. (1994) *Anal. Biochem. 218*, 112−117.
17. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidiman, J. G., Smith, J. A., and Struhl, K., Eds. (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.
18. Kowalski, D., Kroeker, W. D., and Laskowski, M., Sr. (1976) *Biochemistry 15*, 4457−4462.
19. Sung, S., and Laskowski, M., Sr. (1962) *J. Biol. Chem. 237*, 506−511.
20. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
21. Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., and Ecker, J. R. (1993) *Cell 72*, 427−441.
22. Geshi, N., and Brandt, A. (1998) *Planta 204*, 295−304.
23. Volbeda, A., Lahm, A., Sakiyama, F., and Suck, D. (1991) *EMBO J. 10*, 1607−1618.
24. Romier, C., Dominguez, R., Lahm, A., Dahl, O., and Suck, D. (1998) *Proteins: Struct., Funct., Genet. 32*, 414−424.
25. Katz, A. K., and Glusker, J. P. (1998) *Adv. Mol. Struct. Res. 4*, 227−279.
26. Oleykowski, C. A., Bronson Mullins, C. R., Chang, D. W., and Yeung, A. T. (1999) *Biochemistry 38*, 2200−2205.
27. Panavas, T., Pikula, A., Reid, P. D., Rubinstein, B., and Walker, E. L. (1999) *Plant Mol. Biol. 40*, 237−248.

BI992376Z