The High Mobility Group Domain Protein Cmb1 of Schizosaccharomyces pombe Binds to Cytosines in Base Mismatches and Opposite Chemically Altered Guanines*

(Received for publication, May 5, 1998, and in revised form, July 29, 1998)

Oliver Fleck‡, Christophe Kunz, Claudia Rudolph, and Jürg Kohli

From the Institute of General Microbiology, University of Bern, Baltzer-Strasse 4, CH-3012 Bern, Switzerland

DNA can be damaged by a variety of physical and chemical agents. Most types of lesions are efficiently removed from DNA by repair mechanisms. The nucleotide excision repair (NER) system is able to correct a broad spectrum of adducts, including cyclobutane pyrimidine dimers and 6-4 photoproducts caused by ultraviolet light, intrastrand cross-links formed by cis-diamine-dichloroplatinum(II) (cisplatin), as well as O6-methylguanine (O6meG) produced by methylating agents such as N-methyl-N′-nitro-N′-nitrosoguanidine (MNNG) (1–3). The mutHLS-like mismatch-repair pathway efficiently corrects most base mismatches and small DNA loops, which frequently occur during replication by misincorporation of bases and strand slippage, respectively (4, 5). A defect in the human genes hMSH2 (a homologue of Escherichia coli MutS), hMLH1, hPMS1, or hPMS2 (all of which are MutL homologues) is responsible for the heritable cancer syndrome hereditary non-polyposis colon cancer. The clinical symptoms of patients with homozygously mutated hMSH2, hMLH1, or hPMS2 genes are accompanied by microsatellite instability and mismatch repair deficiency, whereas there is no indication that hPMS1 has a role in mismatch repair (4, 6, 7). The human mismatch-binding activity hMutSα, a heterodimer of the two MutS homologous proteins hMSH2 and GTBP (8, 9), also recognizes DNA lesions containing O6meG, O4-methylthymine, and cisplatin-induced 1,2 GpG intrastrand cross-links (10). Thus, mismatch-repair proteins and NER proteins might compete in repair of some types of DNA lesions. Whereas repair mediated by the NER system normally restores wild type information, the action of mismatch-repair proteins on lesions rather contribute to cytotoxicity (11, 12). Several high mobility group (HMG) proteins are known to bind to GpG intrastrand cross-links (13). Binding by HMG-1 shields the lesion from excision by the human ex-cinuclease protein complex (14). Consistently, a Saccharomyces cerevisiae mutant, defective in the HMG-box gene IXR1, is more resistant to cisplatin than wild type (15).

We have identified a mismatch-binding activity in Schizosaccharomyces pombe crude extracts showing high affinity to cytosine-containing mismatches (16). Here, we describe the purification of this activity, Cmb1, and cloning of the respective gene. DNA sequencing revealed that Cmb1 contains a HMG domain. Cmb1 was purified from an Escherichia coli overexpression strain and tested for binding to mismatches, small DNA loops, and substrates containing either GpG intrastrand cross-links or O6meG. Furthermore, we addressed the question of whether a cmb1 mutant responds differently to the chemicals cisplatin, transplatin, and MNNG.

EXPERIMENTAL PROCEDURES

DNA Binding Analyses—Binding of Cmb1 protein to DNA substrates was tested with a band shift assay. Substrates and competitor DNA used to test binding to mismatches, loops, and O6meG lesions are in the M13mp9 polynucleotide sequence context (16). The sequence of the platinated substrates and their unplatinated controls is shown in Fig. 6A. For platination, 1 nmol of plus strand was incubated for 24 h at 37 °C in 100 μl of 10 mM Tris-HCl (pH 7.5) containing either 0.1 mM cisplatin or 0.1 mM transplatin (Sigma). Unreacted cisplatin or transplatin was removed by chromatography on Sephadex G25 (Amersham Pharmacia Biotech). The extent of platinated oligonucleotides was checked on a 6% polyacrylamide gel and was estimated to be approximately 60–80%. Modified and unmodified strands were 5′-end-labeled by T4 polynucleotide kinase (Promega) in the presence of [γ-32P]ATP and annealed with their complementary strands as described (16). To test binding to the various substrates, 20-μl reactions were performed for 15 min at 4 °C in 25 mM Tris-HCl (pH 7.5), 0.5 mM dithiothreitol, 4 mM spermidine, 0.5 mM EDTA, 10% glycerol, 0.01 mM ZnCl2, 50–200 μM NaCl in the presence of 40 fmol radiolabeled substrate. The amount of unlabeled DNA sequencing was performed for 15 min at 4 °C in 25 mM Tris-HCl (pH 7.5), 0.5 mM dithiothreitol, 4 mM spermidine, 0.5 mM EDTA, 10% glycerol, 0.01 mM ZnCl2, 50–200 mM NaCl in the presence of 40 fmol radiolabeled substrate. The amount of unlabeled DNA sequencing was performed for 15 min at 4 °C in 25 mM Tris-HCl (pH 7.5), 0.5 mM dithiothreitol, 4 mM spermidine, 0.5 mM EDTA, 10% glycerol, 0.01 mM ZnCl2, 50–200 mM NaCl in the presence of 40 fmol radiolabeled substrate. The amount of unlabeled

Experimental Procedures

DNA Binding Analyses—Binding of Cmb1 protein to DNA substrates was tested with a band shift assay. Substrates and competitor DNA used to test binding to mismatches, loops, and O6meG lesions are in the M13mp9 polynucleotide sequence context (16). The sequence of the platinated substrates and their unplatinated controls is shown in Fig. 6A. For platination, 1 nmol of plus strand was incubated for 24 h at 37 °C in 100 μl of 10 mM Tris-HCl (pH 7.5) containing either 0.1 mM cisplatin or 0.1 mM transplatin (Sigma). Unreacted cisplatin or transplatin was removed by chromatography on Sephadex G25 (Amersham Pharmacia Biotech). The extent of platinated oligonucleotides was checked on a 6% polyacrylamide gel and was estimated to be approximately 60–80%. Modified and unmodified strands were 5′-end-labeled by T4 polynucleotide kinase (Promega) in the presence of [γ-32P]ATP and annealed with their complementary strands as described (16). To test binding to the various substrates, 20-μl reactions were performed for 15 min at 4 °C in 25 mM Tris-HCl (pH 7.5), 0.5 mM dithiothreitol, 4 mM spermidine, 0.5 mM EDTA, 10% glycerol, 0.01 mM ZnCl2, 50–200 mM NaCl in the presence of 40 fmol radiolabeled substrate. The amount of unlabeled

The mismatch-binding activity Cmb1 of Schizosaccharomyces pombe was enriched from wild type cells, and N-terminal sequencing enabled the cloning of the respective gene. The deduced amino acid sequence of cmb1* contains a high mobility group domain, a motif that is common to a heterogeneous family of DNA-binding proteins. In crude protein extracts of a cmb1 gene-disruption strain, specific binding to C/T, C/A, and C/A was abolished. Weak binding to C/C revealed the presence of a second mismatch-binding activity, Cmb2. Cmb1, enriched from S. pombe and purified from Escherichia coli, bound specifically to C/C, C/T, C/A, C/A, and C/A but showed little or no affinity to other mismatches and small loops. Cmb1 recognizes 1,2 GpG intrastrand cross-links, produced by the chemotherapeutic drug cisplatin, when two cytosines are opposite the cross-linked guanines but not when other bases are present. Consistently, O6-methylguanine:C but not O6-methylguanine/T lesions were bound. Thus, cytosines in mismatches and opposite chemically modified guanines are the preferred target of Cmb1 recognition..cmb1 mutant cells are more sensitive to cisplatin than wild type cells, indicating a role of Cmb1 in repair of cisplatin-induced DNA damage.

The Journal of Biological Chemistry
Vol. 273, No. 46, Issue of November 13, pp. 30398–30405, 1998
Printed in U.S.A.

© 1998 by The American Society for Biochemistry and Molecular Biology, Inc.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AJ002513.

† Supported by a Forschungstipendium of the Deutsche Forschungsgemeinschaft. To whom correspondence should be addressed. Tel.: 41-31-631-4656; Fax: 41-31-631-4684; E-mail: fleck@imb.unibe.ch.

‡ This work was supported by the Swiss National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AJ002513.
homoduplex DNA as competitor and the source of protein samples are indicated in the text. Reaction samples were separated by electrophoresis (110 V, 4 °C) in 6% nondenaturing polyacrylamide gels in 40 mM Tris-HCl (pH 7.5), 20 mM sodium acetate, 1 mM EDTA. Subsequently, gels were stained and scanned for autoradiography.

Partial Purification of Cmb1 from S. pombe and N-terminal Peptide Sequencing—The procedure for obtaining S. pombe cells for protein extracts was as described (16), except that we used Buffer A (25 mM Tris-HCl (pH 7.5, adjusted at room temperature), 150 mM NaCl, 1 mM EDTA, 0.5 mM spermidine, 0.1 mM spermine, 5 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol) for washing and suspending cells. To estimate the degree of purity of the samples from different purification steps were separated on 12–13% SDS-polyacrylamide gels by electrophoresis and analyzed after staining with silver nitrate or Coomassie Blue. For protein purification, all steps were done at 4 °C. 485 g of cells (720 ml), previously frozen in liquid nitrogen and stored at −70 °C, were thawed. A 400-ml chamber for a Bead beater (BioSpec Products) loaded with 350 g of acid-washed glass beads (0.5 mm in diameter) was used for four consecutive steps of cell disruption. Each step included ten 30-s intervals, with at least 2 min of cooling on ice between each interval. The suspensions were pooled and centrifuged for 1 h at 32,000 rpm in a Ti45 rotor (Beckman). The 510-ml supernatant was separated from cell debris and lipids (fraction I). As templates, either 300 ng of genomic DNA corresponding to the construction of the gene was amplified by PCR (annealing at 50 °C) using primer EC1 (5′-CTCTGCTACGTTGATGATATGATCATGCTTG-TAGGCCAAATGCTCC-3′) and primer EC2 (5′-CTGTTTCAAATGTGCTTCG-CATCTGTTATGTTACATGG-3′) derived from KNGLQK, primer EC1 (5′-CTCTGCTACGTTGATGATATGATCATGCTTG-TAGGCCAAATGCTCC-3′) (both derived from the amino acid sequence KNGLQK) or primer EC3 (5′-CTCTGCTACGTTGATGATATGATCATGCTTG-TAGGCCAAATGCTCC-3′) (derived from KNGLQK) and primer CECC-3′. The PCR product of the expected size was obtained with primer pair CECC-3′/CECC-3′-nts. Amplification was performed by three initial cycles with an annealing temperature of 40 °C, followed by 30 cycles with an annealing temperature of 65 °C. PCR products were separated on 1.5% agarose gels. A band of the expected size (102 bp) was eluted from the gel (17) and used for reamplification with primers CC-1b and CC-3 and for PCR with primers CC-1b and CC-2. This reaction (annealing at 47 °C) produced a 67-bp fragment, indicating that the 102-bp fragment contained the sequence of interest. The resulting 102-bp DNA fragment was digested with BamHI, cloned into pUC18 and sequenced. One clone (pOL50) was used for further screening.

Library Screening—The pOL50-insert was random-prime-labeled with [32P] (Ready to Go DNA labeling kit, Amersham Pharmacia Biotech) and used to screen a plasmid-based library of S. pombe (18) by colony hybridization. From this library several clones with a truncated open reading frame (ORF) of cmb1 were identified. A region containing the 5′ end of the ORF was amplified by PCR with primer CC-6 (5′-CTGTGTTGAGCCAAATGCTCC-3′) and primer CC-7 (5′-GATCTCCTACGCACTCAGG-3′) (annealing at 54 °C). The PCR product was radiolabeled with [32P] and hybridized to ordered cosmids and plasmid libraries (19, 20). Two cosmids and four P1 clones were obtained which all map on chromosome I between ras1 and edc3 (19). From one cosmid (ICRF690137), three fragments from the region of interest were subcloned in pUC18 (pOL4 to pOL5).

Sequence Analysis—Double-stranded plasmid–DNA (pOL54 to pOL59) was sequenced by the dyeodeoxy method using a sequencing kit (United States Biochemical Corp.). The cmb1 ORF was determined from both strands with primers derived from pUC18 and with primers designed on the basis of the sequence. The remainder of the inserts (not shown). Sequences were analyzed with the Wisconsin program package, version 9 (Genetics Computer Group, Madison, WI).

Gene Disruption of cmb1—The following strategy resulted in replacement of the complete ORF of cmb1 by his3′. A 342-bp fragment of the 5′-flanking region and a 512-bp fragment of the 3′-flanking region of cmb1 were amplified by PCR (annealing at 50 °C) with Pfu DNA polymerase (Stratagene). For the 5′-flanking region, primers DC-1 (5′-TCAAGTACTGAGCTATATC-3′) and DC-2 (5′-TAAATCTGACTGACTATATC-3′) were used, and for the 3′-flanking region, primers DC-3 (5′-ACACTTCTTGGCTCAGACATC-3′) and DC-4 (5′-TATT-GGGCTGTAATATTG-3′) were used. The primers contain restriction sites at their 5′-ends for cloning (not shown). PCR products were inserted to the left and right of the his3′ gene, previously digested in pBluescript KS. A 1.5-kb fragment containing this cmb1::his3′-con-struction was transformed into a S. pombe his+ strain (21). His+ transformants were checked by Southern hybridization for correct gene disruption (data not shown). Strain OLI42 (h+ cmb1::his3′-his3-D1) was stored for further experiments.

Overexpression of Cmb1 in E. coli and Purification—The full-length cmb1 gene was amplified by PCR (annealing at 54 °C) using Pfu DNA polymerase (Stratagene), primer EC1 (5′-CTCTGCTACGTTGATGATATGATCATGCTTG-TAGGCCAAATGCTCC-3′) and primer EC2 (5′-CTCTGCTACGTTGATGATATGATCATGCTTG-TAGGCCAAATGCTCC-3′) derived from the 5′- and 3′-ends of the coding region, respectively. A truncated cmb1 gene missing the 5′ nucleotide sequence coding for the amino acids 2–41 was amplified with primer EC1 (5′-CTCTGCTACGTTGATGATATGATCATGCTTG-TAGGCCAAATGCTCC-3′) and primer EC2 (annealing at 55 °C), purging gels. Primers EC1 and EC2 contain a NcoI restriction site, and primer EC2 contains a BamHI restriction site. The PCR products were digested with NcoI and BamHI and then ligated with digested pt7-7. The ligation mixtures were transformed into E. coli strain XL1Blue, and plasmids containing correct inserts were checked for mutations by sequencing. For expression of the protein, the plasmids were transformed into E. coli strain BL21/D3. Stationary phase
The reactions contained radiolabeled C/\(\Delta\)A an chemical Corp., 0.64 cm\(^2\) ml/h on a double-stranded DNA-cellulose column (United States Bio-

...fraction III (1.9 mg of protein in 7 ml). MonoQ fraction 29, containing partially purified Cmb1, was tested using a 120-fold excess of unlabeled homoduplex as competitor. The NaCl concentration in the reaction was 80 mM. An activity was detected that showed high affinity to all cytosine-containing mismatches and T/T, but no or only weak binding to other mismatches and homoduplex.

cultures of the expression strains were inoculated 1:40 in fresh LB medium containing 100 \(\mu\)g/ml ampicillin and incubated at 37 °C until an \(A_{600}\) of about 0.7 was reached. Isopropylthio-\(\beta\)-d-galactoside was added to a final concentration of 1 mM. After 5 h of induction, cells were harvested by 10 min of centrifugation at 8300 \(\times g\) and suspended in Buffer A. After brief freezing in liquid nitrogen, cells were stored at −70 °C until further use. To prepare crude protein extracts containing Cmb1 or Cmb1Δ41, Buffer B150 was added to the thawed cells (1/5 of the total volume). Cells were disrupted by sonication on ice (10 times for 30 s, with at least 1 min cooling on ice between each round). After 20 min of centrifugation at 10,000 \(\times g\), the supernatant was removed (crude protein extracts, fraction I).

The strain BL21/DE3 containing pT7–7 with the truncated cmb1 gene was used for purification of the respective protein, Cmb1Δ41. Fraction I (137 mg of protein in 28 ml) was obtained from a 750-ml E. coli culture and loaded at a flow rate of 19 ml/h on a Heparin-agarose column (type I, Sigma, 1.8 cm\(^2\) \(\times 11\) cm) previously equilibrated with Buffer B150. Bound proteins were eluted with 10 ml of Buffer B300 followed by a 100-ml gradient from 300 to 800 mM NaCl. Cmb1Δ41 peaked at 580 mM. The purest fractions were collected and dialyzed in Buffer B150 to give fraction II (2.8 mg of protein in 7 ml). Fraction II was loaded at a flow rate of 8 ml/h on a F11 phosphocellulose column (Whatman, 0.64 cm\(^2\) \(\times 4.7\) cm). Bound proteins were eluted with 10 ml of B350 (Buffer B containing 350 mM NaCl), followed by a 60-ml gradient from 350 to 800 mM NaCl. Cmb1Δ41 eluted between 400 and 550 mM and peaked at 480 mM. At this step, a few faint bands of other proteins were detected in early fractions (around 400 mM) but not in the other fractions containing Cmb1Δ41. However, an additional purification step was included with later fractions. These fractions were pooled and the NaCl concentration was adjusted to 300 mM by addition of Buffer B0 (same as Buffer B150 but without NaCl). The resulting fraction III (1.9 mg of protein in 7.3 ml) was loaded at a flow rate of 7 ml/h on a double-stranded DNA-cellulose column (United States Biochemical Corp., 0.64 cm\(^2\) \(\times 4.7\) cm), equilibrated with B300. After washing, a 40-ml gradient from 300 to 600 mM NaCl was applied. Cmb1Δ41 peaked at 400 mM. No other proteins could be detected in a silver-stained SDS-polyacrylamide gel. Cmb1Δ41 was then dialyzed against Buffer S and stored in aliquots at −20 °C (fraction IV, 1.8 mg of protein in 7 ml).

**Physiological Tests**—To test sensitivity of *S. pombe* cells to various chemicals, \(1 \times 10^7\) cells derived from stationary phase cultures were incubated with either cisplatin or transplatin for 90 min or with MNNG for 30 min. The cells were harvested, washed with 5 ml of 0.85% NaCl, pelleted again, and resuspended in 0.85% NaCl. Appropriate dilutions were plated on YE (0.5% yeast extract; 3% glucose; and 1.5% agar supplemented with adenine, histidine and uracil (each 0.01%)). Plates were incubated for 3 days at 30 °C before scoring colonies. The experiments were carried out six times in the case of cisplatin, twice with transplatin, and six times with MNNG.

**RESULTS**

**Partial Purification of Cmb1 from *S. pombe***—An activity binding to cytosine-containing single-base mismatches was previously identified in *S. pombe* crude extracts (16). We enriched this activity, Cmb1, from extracts of an *S. pombe* wild type strain (fraction I) by ammonium sulfate precipitation (fraction II), followed by chromatography through DEAE-cellulose (fraction III), double-stranded DNA-cellulose (fraction IV), Heparin-agarose (fraction V), and MonoQ columns (fraction VI) (see under “Experimental Procedures”). Analysis of the fractions from the last purification step by a band-shift assay revealed a strong enrichment of the mismatch-specific activity in fractions 30 and 31 (Fig. 1A). The intensity of the band shift correlated well with the intensity of a 22-kDa band visualized in a SDS-polyacrylamide gel by Coomassie Blue staining (Fig. 1B). The purest fractions also contained a second dominant
HMG Domain Protein Cmb1

protein of about 65 kDa and a few other proteins at low concentrations. These fractions were tested for their mismatch-binding ability. We found strong affinity to C/A, T/C, and C/G and some binding to C/C and T/T. No or weak binding was observed to homoduplex and to substrates with a T/G, A/A, G/G, or G/A mismatch (Fig. 1A).

Cloning of the cmb1+ Gene—400 pmol of the 22-kDa Cmb1 protein were used for N-terminal sequencing and resulted in the determination of 29 amino acids (Fig. 2A). Four degenerate primers derived from the peptide sequence of Cmb1 were used for PCR (Fig. 2A). A band of the expected size, obtained with primer pair CC-1b/CC-3, was eluted from the gel and subjected to PCR with primer pair CC-1b/CC-3. A single 67-bp band was produced, indicating that the 102-bp fragment contained the sequence of interest. Sequencing of the 102-bp fragment from five clones revealed inserts with the coding capability for the sequence previously determined from the PCR clones. A PCR product obtained with primers CC-1b/CC-3 was used to screen a plasmid-based library (18). Several clones containing a truncated ORF of cmb1+ were identified. The putative start codon (ATG) is indicated. The insert end, a Sau3A restriction site, is marked with GATC. The amplification product obtained with primers CC-6/CC-7 was used to identify cmb1+ on ordered cosmids/P1 filters (19, 20). C, DNA fragments were isolated from cosmid ICRP60H037 and cloned into pUC18 to give plasmids pOL54 to pOL59. Identical fragments cloned in different orientations are shown below the chromosomal map (e.g. pOL54/pOL55). The restriction map was constructed by Southern hybridization with chromosomal DNA, with cosmid DNA, and by restriction analysis of the plasmid clones pOL54–pOL59. The left end, containing a NsiI and SalI site, is not drawn to scale. Abbreviations for restriction sites: Bg, BglII; EV, EcoRV; N, NsiI; Sa, SalI; and Sc, ScaI. During sequencing of cmb1+, the arg3+ gene was identified about 600 bp downstream. Open reading frames are indicated by black arrows.

Fig. 2. Cloning of the cmb1+ gene. A, 29 amino acids were determined from purified Cmb1 by N-terminal peptide sequencing. Four primers derived from the amino acid sequence were used for PCR (see under “Experimental Procedures”). B, a PCR fragment obtained with primers CC-1b/CC-3 was used to screen a plasmid-based library (18). Several clones containing a truncated ORF of cmb1+ were identified. The putative start codon (ATG) is indicated. The insert end, a Sau3A restriction site, is marked with GATC. The amplification product obtained with primers CC-6/CC-7 was used to identify cmb1+ on ordered cosmids/P1 filters (19, 20). C, DNA fragments were isolated from cosmid ICRP60H037 and cloned into pUC18 to give plasmids pOL54 to pOL59. Identical fragments cloned in different orientations are shown below the chromosomal map (e.g. pOL54/pOL55). The restriction map was constructed by Southern hybridization with chromosomal DNA, with cosmid DNA, and by restriction analysis of the plasmid clones pOL54–pOL59. The left end, containing a NsiI and SalI site, is not drawn to scale. Abbreviations for restriction sites: Bg, BglII; EV, EcoRV; N, NsiI; Sa, SalI; and Sc, ScaI. During sequencing of cmb1+, the arg3+ gene was identified about 600 bp downstream. Open reading frames are indicated by black arrows.

Cmb1 Is a HMG Domain Protein—By data base searches with the deduced amino acid sequence of cmb1+, a HMG domain was identified in the C-terminal part (Fig. 3). This DNA-binding motif is common to a large family of HMG domain proteins characterized in various eukaryotes (13, 23). Originally, HMG-1 and HMG-2 were found to be abundant non-histone components of chromatin, migrating with high mobility in gels (28). A number of proteins were then identified that contain one to six regions of significant homology to regions of HMG-1 and HMG-2 (23, 29). These regions were named HMG domains. Some HMG domain proteins are transcription factors that recognize specific DNA sequences, whereas some are known to bind to specific DNA structures, such as four-way junctions (30, 31), intrastrand cross-links (13, 15) or intermediates of V(D)J recombination (26). Interestingly, our data base searches revealed that also hPMS1, a human MutL homolog (24), contains a HMG domain.

A hydropathy profile of the HMG domain of Cmb1 revealed that it is extremely hydrophilic, a feature common to other HMG domains (26, 32). A few amino acids are conserved among sequence-specific HMG domain proteins, whereas others are present only in non-sequence-specific proteins (23). In this respect, Cmb1 belongs to the subfamily of non-sequence-specific HMG domain proteins (Fig. 3).

A Second C/C Binding Activity, Cmb2, Is Present in the cmb1 Mutant Cells—The cmb1::his3 disruption strain OL142 was constructed as described under “Experimental Procedures.” Because the haploid mutant is viable, cmb1+ has no essential function. We tested mismatch binding with crude protein extracts from the cmb1 disruption mutant. In contrast to extracts from wild type, in which strong binding to all cytosine-containing mispairs was detected (Fig. 4, lanes 2–5), no binding to C/A,
Two HMG domains of the same protein are distinguished by amino acid binding protein, Cmb2 (marked by an S. cerevisiae hPMS1 (24), mouse tsHMG (25), mouse T160 protein (26), soybean SB11 (27), D that bound to the mismatches C/C, C/A, C/T, C/Cmb1- and Cmb1 expression strains were tested for mismatch binding. Both tide enriched from S. pombe between Cmb1, Cmb1 expression strains were tested for mismatch binding. Both showed). The complex was not observed when uninduced strains only weakly to other mismatches and to homoduplex (data not shown). Noticeably, no differences in gel migration and mismatch specificity be-
or a strain containing the empty expression vector were used.

The full-length protein. 

overexpression of Cmb1 from S. pombe exists in E. coli was extensively purified (see under “Experimental Procedures”). Cmb1Δ41 was first tested for its binding ability to substrates containing base-base mismatches, single unpaired nucleotides and small loops with two or four nucleotide inser-
tions (Fig. 5). A strong protein-DNA complex was formed when C/C or C/Δ mismatches were present (Fig. 5, lanes 2 and 3). Weak or no interaction was found with T/Δ-, G/Δ-, A/Δ-, C/C, Δ-, C/Δ-, and T/Δ-containing substrates (lanes 4–9). Thus, one unpaired cytosine is well recognized, but not an unpaired thyme, guanine, or adenine. On the other hand, recognition is not dependent on the presence of mispaired or unpaired cyto-
sines per se, as two or four cytosines in a loop were only weakly bound by Cmb1Δ41 (Fig. 5, lanes 7 and 8). Cmb1Δ41 was first tested for its binding ability to substrates containing base-base mismatches, single unpaired nucleotides and small loops with two or four nucleotide inser-
tions (Fig. 5). A strong protein-DNA complex was formed when C/C or C/Δ mismatches were present (Fig. 5, lanes 2 and 3). Weak or no interaction was found with T/Δ-, G/Δ-, A/Δ-, C/C, Δ-, C/Δ-, and T/Δ-containing substrates (lanes 4–9). Thus, one unpaired cytosine is well recognized, but not an unpaired thyme, guanine, or adenine. On the other hand, recognition is not dependent on the presence of mispaired or unpaired cyto-
sines per se, as two or four cytosines in a loop were only weakly bound by Cmb1Δ41 (Fig. 5, lanes 7 and 8). Cmb2 does not bind to other mismatches, whereas Cmb1 binds specifically to C/C, C/A, T/C, C/T, or C/Δ mismatches was observed with the mutant extracts (lanes 8–10). Surprisingly, weak binding to C/C persisted (lane 7). Thus a second activity (termed Cmb2) for binding to C/C exists in S. pombe. 

Expression of Cmb1 in E. coli and Binding to Mismatches and Small DNA Loops—The full-length cmb1 ORF and a trun-
cation were expressed in E. coli. The latter version, encoding a 22-kDa protein, Cmb1Δ41, corresponds to the proteolytic peptide enriched from S. pombe cells. Crude extracts of both overexpression strains were tested for mismatch binding. Both Cmb1- and Cmb1Δ41-containing extracts exhibited an activity that bound to the mismatches C/C, C/A, C/T, C/Δ, and T/T but only weakly to other mismatches and to homoduplex (data not shown). The complex was not observed when uninduced strains or a strain containing the empty expression vector were used. No differences in gel migration and mismatch specificity be-

For further binding tests, the recombinant Cmb1Δ41 protein was extensively purified (see under “Experimental Procedures”). Cmb1Δ41 was first tested for its binding ability to substrates containing base-base mismatches, single unpaired nucleotides and small loops with two or four nucleotide inser-
tions (Fig. 5). A strong protein-DNA complex was formed when C/C or C/Δ mismatches were present (Fig. 5, lanes 2 and 3). Weak or no interaction was found with T/Δ-, G/Δ-, A/Δ-, C/C, Δ-, C/Δ-, and T/Δ-containing substrates (lanes 4–9). Thus, one unpaired cytosine is well recognized, but not an unpaired thyme, guanine, or adenine. On the other hand, recognition is not dependent on the presence of mispaired or unpaired cytosines per se, as two or four cytosines in a loop were only weakly bound by Cmb1Δ41 (Fig. 5, lanes 7 and 8). Cmb1Δ41 was first tested for its binding ability to substrates containing base-base mismatches, single unpaired nucleotides and small loops with two or four nucleotide inser-
tions (Fig. 5). A strong protein-DNA complex was formed when C/C or C/Δ mismatches were present (Fig. 5, lanes 2 and 3). Weak or no interaction was found with T/Δ-, G/Δ-, A/Δ-, C/C, Δ-, C/Δ-, and T/Δ-containing substrates (lanes 4–9). Thus, one unpaired cytosine is well recognized, but not an unpaired thyme, guanine, or adenine. On the other hand, recognition is not dependent on the presence of mispaired or unpaired cytosines per se, as two or four cytosines in a loop were only weakly bound by Cmb1Δ41 (Fig. 5, lanes 7 and 8).
much stronger (lane 4). Interestingly, only weak binding was detected to cisplatinated oligonucleotides, where opposite the 1,2 cross-linked guanines, other bases were present instead of the cytosines (lanes 5–8). Even the change of only one cytosine strongly affected complex formation (lane 8). Binding to the compound lesions (intrastrand cross-link combined with mismatches) was as low as, or even lower than, to homoduplex (lane 3) and to a substrate containing a G/T mismatch in the same sequence context (lane 9). No significant affinity to the trans-diamine-dichloroplatinum(II)-GpT/G/CpApC lesion was found (data not shown). This 1,3 intrastrand cross-link is produced by transplatin, an isomer of cisplatin.

*S. pombe* Cells Defective in *cmb1* Are Sensitive to Cisplatin—To investigate the cellular role of Cmb1 in repair of lesions produced by either cisplatin or transplatin, the cytotoxicity of these drugs was tested in wild type and the *cmb1* mutant. The *cmb1* strain was significantly more sensitive to cisplatin than wild type. In the presence of 0.8 mM cisplatin, cell survival was about 17-fold reduced (Fig. 6C). On the other hand, when cells were treated with transplatin at concentrations between 1 and 3 mM, no difference in survival was found between wild type and the *cmb1* mutant (data not shown).

*Cmb1* Binds to *O*6meG:C but Not to *O*6meG/T—The cytotoxic action of MNNG is predominantly due to methylation of guanines. *O*6meG can equally pair with cytosine and thymine. Thus a persisting *O*6meG:C site in DNA is frequently changed during replication to *O*6meG/T, resulting in fixation of a mutation (33). We tested binding of Cmb1 to both types of lesions. When a cytosine was present opposite the modified guanine, a complex was formed that was of comparable strength to that seen with the substrate containing a C/C mismatch (Fig. 7A, lanes 2 and 3). In contrast, binding to *O*6meG/T was as weak as to G:C homoduplex DNA and to G/T containing substrate (lanes 1, 4, and 5). When *S. pombe* cells, defective in *cmb1*, were treated with different amounts of MNNG, no difference from wild type cells was found in survival rates (Fig. 7B).

**DISCUSSION**

*Cmb1* Is a HMG Domain Protein—We have previously described the identification of an activity in *S. pombe* crude extracts that efficiently binds to the cytosine-containing mismatches C/C, C/A, T/C, and C/A and weakly binds to T/T (16). We now report partial purification of the 22-kDa activity, Cmb1, from *S. pombe* wild type cells and cloning of the gene. The ORF of *cmb1* encodes a 26-kDa protein with 223 amino acids, indicating that a proteolytic peptide was enriched from *S. pombe* wild type. Cells defective in *cmb1* are sensitive to cisplatin—

binding to homoduplex (lane 3). Note that the *cis*-DDP-GpG containing substrates run more slowly in the gel than the unmodified substrates. No specific binding to the transplatinated substrate was detected (data not shown). Cmb1 strongly bound to double-stranded substrates containing a *cis*-DDP-GpG lesion when two cytosines were opposite the intrastrand cross-link (lane 4) but only weakly when other bases were present (lanes 5–8). Also, binding to a G/T mismatch in the same sequence context (lane 9) was as low as binding to homoduplex (lane 3). Note that the *cis*-DDP-GpG containing substrates run more slowly in the gel than the unmodified substrates. No specific binding to the transplatinated substrate was detected (data not shown). Cmb1 strongly bound to double-stranded substrates containing a *cis*-DDP-GpG lesion when two cytosines were opposite the intrastrand cross-link (lane 4) but only weakly when other bases were present (lanes 5–8). Also, binding to a G/T mismatch in the same sequence context (lane 9) was as low as binding to homoduplex (lane 3). Note that the *cis*-DDP-GpG containing substrates run more slowly in the gel than the unmodified substrates. No specific binding to the transplatinated substrate was detected (data not shown). Cmb1 strongly bound to double-stranded substrates containing a *cis*-DDP-GpG lesion when two cytosines were opposite the intrastrand cross-link (lane 4) but only weakly when other bases were present (lanes 5–8). Also, binding to a G/T mismatch in the same sequence context (lane 9) was as low as binding to homoduplex (lane 3). Note that the *cis*-DDP-GpG containing substrates run more slowly in the gel than the unmodified substrates. No specific binding to the transplatinated substrate was detected (data not shown). Cmb1 strongly bound to double-stranded substrates containing a *cis*-DDP-GpG lesion when two cytosines were opposite the intrastrand cross-link (lane 4) but only weakly when other bases were present (lanes 5–8). Also, binding to a G/T mismatch in the same sequence context (lane 9) was as low as
Cmb1 binds strongly to O\textsubscript{6}meG:C but only weakly to O\textsubscript{6}meG/T. A, band shift analysis to test binding of Cmb1 to O\textsubscript{6}meG lesions. The reactions contained 3.3 pmol of Cmb1\textsubscript{41} protein, 40 fmol of radiolabeled substrates as indicated, a 120-fold excess of unlabeled homoduplex as competitor, and 120 mM NaCl in a standard reaction buffer. B, cytotoxic effect of MNNG on cell survival of S. pombe wild type (●) and cmb1\textsubscript{1241} (○) strains. The experiment was carried out six times (see under "Experimental Procedures"). Points are the average of three independent experiments; bars represent S.D.

Fig. 7. Cmb1 binds to Cytosine-containing Mismatches—The purification of Cmb1 from S. pombe cells resulted in highly active fractions, which predominantly contained the Cmb1 protein but also a few other proteins (Fig. 1). Although the strength of mismatch binding correlated well with the concentration of the Cmb1 protein in the various fractions from the last purification step, it could not be ruled out that mismatch binding was provided by another protein in the fractions. Efforts to separate Cmb1 from the remaining S. pombe proteins by additional chromatographic steps failed (data not shown). We therefore decided to test substrate binding with recombinant Cmb1 protein purified from E. coli.

In a first step, the full-length cmb1\textsuperscript{+} ORF, as well as a truncated version, Cmb1\textsubscript{41}, which is missing the first 41 amino acids, like the proteolytic peptide purified from yeast, were overexpressed in E. coli and tested for mismatch binding. No difference in specificity was found between the two recombinant Cmb1 peptides and Cmb1 enriched from S. pombe. In the E. coli expression strain containing the empty pT7–7 vector, the specific mismatch-binding complex was not detected. It is concluded that Cmb1 is able to bind to mismatches without the need of other S. pombe proteins.

We then purified Cmb1\textsubscript{41} from E. coli cells and tested binding to mismatches and small DNA loops. Cmb1 recognizes C/C, C/A, C/T, T/T, and a C/Δ insertion but shows no or weak affinity to other types of mismatches and insertions (Fig. 5 and data not shown). Although one unpaired cytosine is strongly bound by Cmb1, two or four unpaired cytosines are poor substrates (Fig. 5). Thus, cytosine-containing base mismatches (and T/T) represent a DNA structure irregularity that is recognized by a new type of HMG domain protein. This binding pattern is different from E. coli MutS and eukaryotic MutS\textalpha, which are able to bind to almost all types of base mismatches (34–38). The difference is most pronounced for C/C mismatches, which are a poor substrate for binding and repair by E. coli and yeast MutS proteins. A human mismatch-binding activity was identified that binds to A/C, C/T, and T/T mismatches without requirement of the MutS homologue hMSH2 (39). This specificity is similar to that of Cmb1 (Fig. 1) except that C/C mismatches were poorly bound by the human activity.

A Second C/C Binding Activity, Cmb2, Exists in S. pombe—A cmb1 gene disruption strain was tested for the presence of the mismatch-binding activity. Specific binding to C/A, C/T, and C/Δ was abolished, but some binding to C/C remained (Fig. 4). We conclude that a second activity, Cmb2, exists in S. pombe that recognizes C/C but not other mismatches. Thus, C/C mismatches are recognized redundantly by Cmb1 and Cmb2.

Cmb1 Specifically Binds to 1,2 GpG Intrastrand Cross-links and to O\textsubscript{6}meG When Cytosines Are Opposite the Damaged Bases—1,2 GpG intrastrand cross-links, produced by cisplatin, are substrates of NER proteins, of MutS-type proteins, and of HMG domain proteins (1, 10, 13, 15). We found that this lesion is also bound by the HMG domain protein Cmb1. Our data further suggest that binding by Cmb1 requires the cytosines opposite the cisplatinated guanines. Consistently, O\textsubscript{6}meG:C but not O\textsubscript{6}meG/T is recognized by Cmb1. Thus, like other HMG domain proteins, MutS-type proteins, and the NER complex, Cmb1 recognizes a series of DNA distortions. It is not well understood how the various recognition proteins discriminate between normally paired bases and the different types of altered DNA structures. In the case of Cmb1, the presence of cytosines in mismatches as well as in DNA lesions might be crucial for recognition. hMutS\textalpha and some HMG domain proteins are thought to confer cellular sensitivity in response to the drug cisplatin. Whereas binding by the HMG domain proteins HMG-1 and Ixr1 is proposed to shield cisplatin adducts from repair (14, 15), it is likely that binding of Cmb1 facilitates repair, as indicated by enhanced sensitivity of the cmb1\textsubscript{+} mutant to cisplatin. Transplatin is an isomer of cisplatin that is inactive as a chemotherapeutic drug, and cross-links produced by transplatin are not bound by HMG domain proteins (13). Consistently, an ixr1 mutant treated with transplatin showed no difference from wild type in cell survival (15). Similarly, Cmb1 does not bind to a 1,3 GpTpG intrastrand cross-link produced by transplatin, and in comparison to wild type, the cmb1\textsubscript{+} mutant is not more sensitive to transplatin (data not shown).

Although Cmb1 specifically binds to an O\textsubscript{6}meG:C lesion (Fig. 7A), no significant difference between wild type and the cmb1\textsubscript{+} mutant was found when cells were treated with MNNG (Fig. 7B). This may be due to inefficient repair of O\textsubscript{6}meG by an O\textsubscript{6}-methylguanine-DNA methyltransferase. However, more experiments should be done to clarify whether Cmb1 has a role in removal of O\textsubscript{6}meG from DNA. Genetic experiments with cmb1 disruption strains revealed changes of the pattern of meiotic mismatch repair but no alteration in mitotic mismatch repair.2 The observed effects on meiotic mismatch repair were rather weak. This may be due to redundancy of Cmb1 with other proteins of similar function. One candidate is Cmb2, because it is able to bind to C/C mismatches (Fig. 4). Cmb1 may act as an
accessory protein binding to chemically modified DNA and mismatches and thereby marking such sites for repair. A candidate repair pathway is the NER system, which can restore bases damaged by cisplatin and MNNG (1, 3). Our recent data show that NER genes of S. pombe are involved in mismatch repair in a MutS/L-independent pathway.2 The key characteristic is the repair of C/C mismatches that cannot be repaired by a MutS/L-type system (40, 41).

Acknowledgments—We thank Wolf-Dietrich Heyer and Primo Schar for helpful discussions and Edgar Kaslin, Ursula Reinhard, and Vladimir Bashkirov for advice on protein purification. We are also grateful to Elmar Maier and Nadja Pohl for kindly providing cosmID/P1 libraries and clones.

REFERENCES

1. Huang, J.-C., Hsu, D. S., Kazantsev, A., and Sancar, A. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 12213–12217
2. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) DNA Repair and Mutagenesis, pp. 191–365, American Society for Microbiology, Washington, D.C.
3. Sancar, A. (1996) Annu. Rev. Biochem. 65, 43–81
4. Modrich, P., and Lahue, R. (1996) Annu. Rev. Biochem. 65, 101–133
5. Kolodner, R. D. (1996) Genes Dev. 10, 1433–1442
6. Jiricny, J. (1994) Trends Genet. 10, 164–168
7. Kolodner, R. D. (1995) Trends Biochem. Sci. 20, 397–401
8. Drummond, J. T., Li, G. M., Longley, M. J., and Modrich, P. (1995) Science, 268, 1909–1912
9. Palombo, F., Gallinaro, P., Iacarino, I., Lettieri, T., Hughes, M., D’Arrigo, A., Truong, O., Hsuan, J. J., and Jiricny, J. (1996) Science, 271, 1789–1796
10. Duckett, D. R., Drummond, J. T., Murchie, A. I., Reardon, J. T., Sancar, A., Truong, O., Hsuan, J. J., and Jiricny, J. (1996) Science, 271, 1789–1796
11. Branch, P., Aquilina, G., Bignami, M., and Karran, P. (1993) Nature 362, 652–654
12. Umar, A., and Kunkel, T. A. (1996) Eur. J. Biochem. 238, 297–307
13. Zambrano, D. B., and Lippard, S. J. (1995) Trends Biochem. Sci. 20, 435–439
14. Huang, J.-C., Zambrano, D. B., Reardon, J. T., Lippard, S. J., and Sancar, A. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 10394–10398
15. Brown, S. J., Kellet, P. J., and Lippard, S. J. (1993) Science, 261, 603–605
16. Fleck, O., Schar, P., and Kohli, J. (1994) Nucleic Acids Res. 22, 5289–5295
17. Wu, W., and Welsh, M. J. (1995) Anal. Biochem. 229, 350–352
18. Barbet, N., Muriel, W. J., and Carr, A. M. (1992) Gene 114, 59–66
19. Hoheisel, J. D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A. V., Schalkwyk, L. C., Ninetic, D., Francis, F., and Lehrach, H. (1995) Cell, 73, 109–120
20. Maier, E., Hoheisel, J. D., McCarthy, L., Mott, R., Grigoriev, A. V., Monaco, A. P., Larin, Z., and Lehrach, H. (1992) Nat. Genet. 1, 273–277
21. Burke, J. D., and Gould, K. L. (1994) Mol. Gen. Genet. 243, 169–176
22. Van Huffel, C., Dubois, E., and Messeguy, F. (1992) Eur. J. Biochem., 205, 33–43
23. Grosschedl, R., Giese, K., and Pagel, J. (1994) Trends Genet. 10, 94–100
24. Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y.-F., Carter, K. C., Fraser, C. M., Adams, M. D., Venter, J. C., Dunlop, M. G., Hamilton, S. R., Petersen, G. M., de la Chapelle, A., Vogelstein, B., and Kinzler, K. W. (1994) Nature 371, 75–80
25. Boissonneault, G., and Lai, Y.-F. C. (1993) Mol. Cell. Biol. 13, 4323–4330
26. Shirakata, M., Huppi, K., Usuda, S., Okazaki, K., Yoshida, K., and Sakano, H. (1991) Mol. Cell. Biol., 11, 4528–4536
27. Laux, T., and Goldberg, R. B. (1991) Nucleic Acids Res. 19, 4769
28. Johns, E. W. (1982) in The HMG Chromosomal Proteins (Johns, E. W., ed) pp. 1–9, Academic Press, London
29. Jantzen, H.-M., Admon, A., Bell, S. P., and Tjian, R. (1990) Nature 344, 830–836
30. Bianchi, M. E., Beltrame, M., and Pasnessa, G. (1989) Science, 243, 1056–1059
31. Hu, C. H., McStay, B., Jeong, S.-W., and Reeder, R. H. (1994) Mol. Cell. Biol. 14, 2871–2882
32. Grass, J. D., and Feix, G. (1991) Nucleic Acids Res. 19, 2573–2577
33. Lindahl, T., Sedgwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988) Annu. Rev. Biochem. 57, 135–157
34. Modrich, P. (1991) Annu. Rev. Genet. 25, 229–253
35. Miret, J. J., Milla, M. G., and Lahue, R. S. (1993) J. Biol. Chem. 268, 3507–3513
36. Alani, E. (1996) Mol. Cell. Biol., 16, 5604–5615
37. Habreken, Y., Sung, P., Prakash, L., and Prakash, S. (1996) Curr. Biol., 6, 1185–1187
38. Iacarino, I., Palombo, F., Drummond, J., Totty, N. F., Hsuan, J. J., Modrich, P., and Jiricny, J. (1996) Curr. Biol., 6, 484–486
39. O’Regan, N. E., Branch, P., Macpherson, P., and Karran, P. (1996) J. Biol. Chem., 271, 1769–1796
40. Schar, P., Muna, P., and Kohli, J. (1993) Genetics 133, 815–824
41. Schar, P., and Kohli, J. (1993) Genetics 133, 825–835