Supplemental Material can be found at:
http://www.jbc.org/content/suppl/2011/01/03/M110.205088.DC1.html

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 10, pp. 8240–8251, March 11, 2011
© 2011 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

Creating Directed Double-strand Breaks with the Ref Protein

A NOVEL RecA-DEPENDENT NUCLEASE FROM BACTERIOPHAGE P1[*][S]

Received for publication, November 19, 2010 Published, JBC Papers in Press, December 30, 2010, DOI 10.1074/jbc.M110.205088

Marielle C. Gruenig[‡], Duo Lu[§], Sang Joon Won[‡], Charles L. Dulberger[‡], Angela J. Manlick[‡], James L. Keck[§], and Michael M. Cox[‡1]

From the [‡]Department of Biochemistry, University of Wisconsin and the [§]Department of Biomolecular Chemistry, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin 53706

The bacteriophage P1-encoded Ref protein enhances RecA-dependent recombination *in vivo* by an unknown mechanism. We demonstrate that Ref is a new type of enzyme; that is, a RecA-dependent nuclease. Ref binds to ss- and dsDNA but does not cleave any DNA substrate until RecA protein and ATP are added to form RecA nucleoprotein filaments. Ref cleaves only where RecA protein is bound. RecA functions as a co-nuclease in the Ref/RecA system. Ref nuclease activity can be limited to the targeted strands of short RecA-containing D-loops. The result is a uniquely programmable endonuclease activity, producing targeted double-strand breaks at any chosen DNA sequence in an oligonucleotide-directed fashion. We present evidence indicating that cleavage occurs in the RecA filament groove. The structure of the Ref protein has been determined to 1.4 Å resolution. The core structure, consisting of residues 77–186, consists of a central 2-stranded β-hairpin that is sandwiched between several α-helical and extended loop elements. The N-terminal 76 amino acid residues are disordered; this flexible region is required for optimal activity. The overall structure of Ref, including several putative active site histidine residues, defines a new subclass of HNH-family nucleases. We propose that enhancement of recombination by Ref reflects the introduction of directed, recombinogenic double-strand breaks.

The bacterial recombinase, RecA protein, plays a central role in maintenance of genome stability. It generally functions as a RecA nucleoprotein filament formed on DNA (1). RecA has three identified functions in the cell. First, it participates directly in all recombination processes. Filaments of *Escherichia coli* RecA protein form most readily on ssDNA; these filaments can then catalyze DNA pairing and strand exchange with a homologous duplex DNA (2). A RecA filament complexed with an oligonucleotide can invade duplex DNA and pair the oligonucleotide with a complementary sequence within a much longer duplex, forming a displacement loop (D-loop) (2). Second, RecA filaments play a central role in the induction of the SOS response (3). In brief, RecA filaments assembled on ssDNA generated by stalled replication forks bind and stimulate the autocatalytic cleavage of the LexA repressor. This co-protease function for RecA inactivates LexA, which leads to the induction of SOS. Third, RecA filaments play two roles in activating the translesion DNA synthesis function of DNA polymerase V. If cellular genomic replication is not restored after 30–60 min of SOS response, polymerase V opens a final, mutagenic phase of SOS. The UmuD subunit of polymerase V is autocatalytically cleaved, again facilitated by interaction with RecA filaments, to create UmuD′. The weakly active UmuD′$_2$·UmuC complex undergoes a final activation step with the transfer of a RecA subunit from the 3′-proximal end of a RecA filament to form the activated polymerase V enzyme consisting of UmuD′$_2$·UmuC/RecA (4, 5). The presence of RecA as a subunit of active polymerase V is the only known activity where RecA exhibits a function when it is not part of a filament.

Many other proteins interact with RecA protein filaments on DNA, and many of these serve to regulate almost every aspect of RecA function (6). The regulators include DinI (7), RecX (8–10), RdgC (11), PsiB (12), RecFOR (13–15), UvrD (16), and RecBCD (17–19), a list that will doubtlessly grow. In all, more than a dozen known proteins interact with RecA and help coordinate its functions with many aspects of DNA metabolism. A number of RecA partner proteins bind within the helical groove of active RecA nucleoprotein filaments. These include LexA (20, 21), UmuD (22), RecX (20), and DinI.[2] The LexA and UmuD proteins (along with the bacteriophage λ repressor (23)) undergo autocatalytic cleavage in this environment. The regulation of RecA is not limited to bacteria-encoded functions. The PsiB protein is a product of conjugative F-plasmids, expressed early in conjugation to suppress the SOS response in the recipient cell (12, 24, 25). Additional proteins that interact with RecA are encoded by bacteriophages, including the bacteriophage P1.

Almost from the moment it was described in 1951 (26), bacteriophage P1 has been a workhorse of molecular biology. It is now used largely for generalized transduction applied to strain construction (27, 28). Its genome (93.6 kbp; ~117 genes (27)) is packaged in phage particles as a terminally redundant linear DNA molecule that is cyclized once it enters the cell. The cyclization can use the phage-encoded Cre-lox site-specific recombination system. Cyclization can also utilize homologous

* This work was supported, in whole or in part, by National Institutes of Health Grants GM32335 (to M. M. C.) and GM068061 (NIGMS; to J. L. K.). This work was also supported by a William R. and Dorothy E. Sullivan Wisconsin Distinguished Graduate Fellowship (to M. C. G.).

The atomic coordinates and structure factors (code 3PLW) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–5.

[1] To whom correspondence should be addressed: Dept. of Biochemistry, University of Wisconsin, 433 Babcock Dr., Madison, WI 53575. Tel.: 608-262-1181; Fax: 608-265-2603; E-mail: cox@biochem.wisc.edu.

[2] M. Cox and E. Egelman, unpublished data.

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

recombination involving the terminally redundant DNA ends, catalyzed by the host RecA protein (27, 28).

A search for new modes of RecA regulation led us to a recombination enhancement function gene, *ref*, encoded by bacteriophage P1. The Ref protein product of this gene (186 amino acid residues; predicted M_r 21,329 including the initiating methionine) increases chromosomal recombination, plasmid recombination, and the excision of an IS1 element in *E. coli*, all in a RecA-dependent fashion (29–31), suggesting that Ref might be a positive RecA regulator. Whereas it is reasonable to presume that Ref has a role in enhancing recombination events important to the bacteriophage, such as genome cyclization, *ref* gene deletion has little effect on the P1 life cycle. Thus, the role of the Ref protein in bacteriophage P1 biology remains unclear (29, 30, 32). In this work we have purified and characterized the Ref protein, finding no evidence that it functions as a RecA regulator. Instead, Ref protein is a new kind of enzyme; that is, a RecA-dependent nuclease.

EXPERIMENTAL PROCEDURES

Proteins—The native *E. coli* wild type RecA and single-stranded DNA binding proteins were purified as described previously (12). The concentration of the purified RecA and single-stranded DNA binding proteins was determined from the absorbance at 280 nm using the extinction coefficients of 2.23×10^4 M^{-1} cm^{-1} (33) and 2.83×10^4 M^{-1} cm^{-1}, respectively (34). The non-cleavable UmuD1 protein (35) was a generous gift from M. Goodman. LexA S119A (36) was a generous gift from J. Little. Wild type LexA was purified as described (37). DinI was purified as described (7).

The RecA E38K mutant was purified like wild type RecA with the following modifications. The polyethyleneimine pellet containing RecA E38K was washed with R buffer (20 mM Tris Cl (80% cation; pH 7.5) 0.1 mM EDTA 1 mM DTT 10% glycerol), not R plus 150 mM $(NH_4)_2SO_4$. The protein was then purified by successive chromatographic steps using a butyl-Sepharose column, an SP-Sepharose column, a ceramic hydroxyapatite column, and a Sephacryl S-300 gel filtration column. RecA E38K K72R and RecA K72R were purified as described previously (37). The concentrations of all three RecA mutant proteins were determined from the absorbance at 280 nm using the extinction coefficient 2.23×10^4 M^{-1} cm^{-1}, and they are stored in R buffer.

DNA Substrates—The circular ssDNA from bacteriophage M13mp18 (7249 nucleotides) was prepared essentially as described (38, 39). The linear single-stranded DNA was prepared by annealing an oligonucleotide complementary to the BamHI site followed by a BamHI restriction digest. The linear ssDNA was cleaned up by removal of the oligonucleotide using a YM-100 Centricon. The concentration of circular and linear ssDNA was determined by absorbance at 260 nm using 36 μg ml^{-1} A_{260}^{-1} as the conversion factor. The M13mp18 circular dsDNA was prepared as described in Refs. 38–40. The M13mp18 linear dsDNA substrate was prepared by digesting M13mp18 circular dsDNA with PstI. All DNA concentrations are given in total nucleotides. Oligonucleotides were purchased from Integrated DNA Technologies. Sequences of oligonucleotides used in this study are presented in Table 1.

Cloning, Overexpression, and Purification of Ref and Ref Variants—The P1 *ref* gene was obtained by PCR amplification, cloned, and expressed using the vector pET21A (Novagen). The resulting plasmid was designated pEAW584. The presence of wt P1 *ref* was confirmed by direct sequencing.

Standard methods were used to alter the *ref* gene in plasmid pEAW584, changing the CAT bases at 457–459 (His) at amino acid 153 to GCT (Ala). The resulting plasmid was designated pEAW665. Ref ΔN76 was constructed beginning with pEAW584. A PCR primer was constructed to amplify codons 77–185 of the Ref gene in pEAW584, adding a new Met initiator codon on the altered N terminus. The upstream primer consisted of an NdeI site and bases 229–270 of the P1 *ref* gene. The ATG of the NdeI site codes for the Met. For better codon usage in *E. coli*, the GGG coding for Gly at amino acid (aa) 2 (codon 77) was changed to GGT, the AGA coding for Arg at aa 3 was changed to CGT, the ACA coding for Thr at aa 4 was changed to ACC, the ACG coding for Thr at aa 6 was changed to ACC, and the CGG coding for Arg at aa 10 was changed to CGT. The downstream primer was the same used to clone pEAW584. The PCR product was digested with NdeI and BamHI and ligated to pET21A (Novagen) digested with the same enzymes. The resulting plasmid was designated pEAW685. The structure of both mutant *ref* genes was confirmed by direct sequencing.

Competent cells of *E. coli* strain BL21(DE3) were transformed with plasmid pEAW584. Ten liters of culture were grown in LB broth to an A_{600} of 0.51. Ref protein expression was induced by the addition of isopropyl 1-thio-β-D-galactopyranoside to 0.4 mM. After a 3-h 10-min outgrowth at 37 °C, 24.6 g of cells were harvested by centrifugation, flash-frozen in liquid N_2, and stored at −80 °C. The protein expressed is the native, 186-amino acid polypeptide without tags.

All purification steps were carried out at 4 °C. Purification entailed polyethyleneimine precipitation and pellet extraction, precipitation with $(NH_4)_2SO_4$, and chromatography successively using butyl-Sepharose, Source 15 Q, ceramic hydroxyapatite columns, and Sephacryl S-100 gel filtration columns. This was followed by another butyl-Sepharose chromatography step. The protein was concentrated using Amicon Centricon-Plus 20 and dialyzed against Ref storage buffer (R plus 200 mM potassium glutamate), flash-frozen in liquid N_2, and stored at 80 °C. The protein was >99% pure and free of detectable nuclease activity when incubated at 37 °C for 2 h with different DNA substrates (circular ssDNA, linear and supercoiled dsDNA, and labeled oligonucleotides) in a buffer containing 25 mM Tris-OAc (80% cation, pH 7.6), 1 mM DTT, 3 mM potassium glutamate, 10 mM $Mg(OAc)_2$, and 5% (w/v) glycerol (buffer A). A more detailed purification protocol is available on request.

The concentration of the Ref protein was determined from the absorbance at 280 nm using the extinction coefficient 2.851×10^4 M^{-1} cm^{-1}. The Ref extinction coefficient ($\epsilon_{nat,280\ nm} = 2.851 \times 10^4 \pm 0.108$ M^{-1} cm^{-1}) was determined during the course of the present work using procedures described elsewhere (41, 42). The identity of the purified protein was confirmed by mass spectrometry. The measured mass of the protein was 21,326 Da, in very good agreement with the calculated mass of Ref protein of 21,329 Da (with the initiator Met residue still present). In the course of these studies, a

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

TABLE 1
Primers and oligonucleotides
FAM, 6-carboxyfluorescein.

Name	Length	Sequence	Complementary to M13mp18 at bases
	nt		
rlb1	150	TTTTGGTTTTTATCGTCGTCTGGTAAACGAGGGTTATGATAGTGTTGCTCTTACTATGCCTCGTAA TTCCTTTTGGCGTTATGTATCTGCATTAGTTGAATGTGGTATTCCTAAATCTCAACTGATGAATCT TTCTACCTGTAATAATGT	597–746
msc1	60	ATTCTTACGCTTTCAGGTCAGAAGGGTTCTATCTCTGTTGGCCAGAATGTCCCTTTTATT	5041–5100
den7	50	GGCCTCGCGGTAGCTGAGCTCGGAGCGCACGATTCGCACTGCTGATGTTC[36]-FAM	N/A
galK	150	GCATTTGGCTACCCTGCCACTCACACCATTCAGGCGCCTGGCCGCGTGAATTTGATTGGTGAACA CACCGACTACAACGACGGTTTCGTTCTGCCCTGCGCGATTGATTATCAAACCGTGATCAGTTGTG CACCACGCGATGACCGTAA	N/A
mcg1	54	FAM-[56]TAACATCAGCAGTGCGAATCGTGCGCTCCGAGCTCAGCTACCGCGAGGCCTGCA	N/A
den10	50	GAACATCAGCAGTGCGAATCGTGCGCTCCGAGCTCAGCTACCGCGAGGCC	N/A
mcg2	54	GAACATCAGCAGTGCGAATCGTGCGCTCCGAGCTCAGCTACCGCGAGGCCTGCA	N/A
mcg3	150	TCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCA GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAG GAAACAGCTATGA	6070–6219
HJ1	41	FAM-[56]CCCGTGATCACCAATGCAGATTGACGAACCTTTGCCCACGT	
HJ2	41	GACGTGGGCAAAGGTTCGTCAATGGACTGACAGCTGCATGG	
HJ3	41	GCCATGCAGCTGTCAGTCCATTGTCATGCTAGGCCTACTGC	
HJ4	41	GGCAGTAGGCCTAGCATGACAATCTGCATTGGTGATCACGG	
ANP003	19	GACGTGGGCAAAGGTTCGT	
ANP004	18	TCATGCTAGGCCTACTGC	

higher molecular weight band on SDS-PAGE gels was observed that corresponded to the approximate size of a Ref dimer. The identity of this band was confirmed to be Ref by mass spectrometry.

The Ref H153A and ΔN76 mutant proteins were purified with procedures using very similar growth, induction, cell harvesting, and early fractionation steps. Ref 153A was purified with successive chromatographic steps employing butyl-Sepharose, ceramic hydroxyapatite, and Sephacryl S-100 followed by additional ceramic hydroxyapatite and butyl-Sepharose steps. The protein was >99% pure by SDS-PAGE and free of any detectable nuclease contamination. Ref H153A is folded the same as wild type Ref as confirmed by CD spectroscopy (supplemental Fig. S1). The Ref ΔN76 protein was purified with successive butyl-Sepharose, ceramic hydroxyapatite, Sephacryl S-100, and Source 15 Q-Sepharose. The protein was >99% pure and free of detectable nuclease activity. The concentration of the Ref ΔN76 protein was determined from the absorbance at 280 nm using a calculated extinction coefficient of 15,220 M^{-1} cm^{-1}. The identity of the Ref ΔN76 protein was confirmed by excising the protein band from a gel and digesting with trypsin. The products were subjected to MALDI-TOF/TOF mass spectrometry (Applied Biosystems/MDS SCIEX 4800) for identification of peptides. Peptides detected and sequenced (many repeatedly) were all consistent with the predicted sequence of Ref ΔN76 and included 92% of the amino acid residues in the protein.

Electrophoretic Mobility Shift Assays—A single-stranded 50-mer oligonucleotide, 3′ 6-carboxyfluorescein-labeled den7, was purchased from Integrated DNA Technologies (Table 1). To generate labeled blunt-ended double-stranded DNA, the labeled 50-mer was annealed to an unlabeled complementary 50-mer, den10 (Table 1). The two labeled DNA substrates were used at 2.5 or 5 μM (in total nucleotides) for the ssDNA or dsDNA oligos, respectively. The DNA binding reactions also contained buffer A. The reactions (10 μl) containing the indicated concentrations of Ref protein (replaced in controls with Ref storage buffer) were incubated at 37 °C for 10 min. Then 5 μl of 6× loading buffer (18% (w/v) Ficoll, 20 mM Tris-OAc 80% cation) was added to 10 μl of the reactions, and the reactions were loaded onto a native 8% polyacrylamide gel and subjected to electrophoresis in TBE buffer (90 mM Tris borate and 10 mM EDTA). The longer M13mp18 circular ssDNA or linear dsDNA substrates were used at a concentration of 10 μMnt[3] in reactions (40 μl) containing otherwise the same components as described above. Various concentrations of Ref were added to the reactions and incubated for 40 min at 37 °C. Then 5 μl of 6× loading buffer (18% (w/v) Ficoll, 20 mM Tris-OAc 80% cation) was added to the reactions, and the entire reactions were loaded onto a native 0.5% agarose gel and subjected to electrophoresis in TE buffer (10 mM Tris acetate (80% cation) and 1 mM EDTA).

Nuclease Assay—The reactions were carried out at 37 °C and contained buffer A, an ATP regeneration system (10 units/ml pyruvate kinase and 3.5 mM phosphoenolpyruvate), and 4 μMnt M13mp18 circular ssDNA. The aforementioned components were incubated with 2.4 μM RecA for 10 min. Three mM ATP was added followed by a 15-min incubation to allow for RecA filament formation before Ref was added at the concentrations indicated. After 20 min, 20 μl of the reaction was stopped by incubation with 5 μl of 20 mg/ml proteinase K for 60 min at 37 °C followed by the addition of 10 μl of a solution containing 9% Ficoll, 0.25% bromphenol blue, 0.25% xylene cyanol, and 4% SDS and another 60-min incubation at 37 °C. Samples were subjected to electrophoresis in 0.8% agarose gels with TAE buffer (40 mM Tris-Acetate 1 mM EDTA), stained with SYBR-Gold nucleic acid stain (Invitrogen), and exposed to UV light.

Nuclease Site-specific Targeting Assay—The reactions were carried out at 37 °C and contained buffer A and an ATP regeneration system (10 units/ml pyruvate kinase and 3.5 mM phosphoenolpyruvate). Four μMnt of a 150-nt long oligonucleotide (rlb1) and 0.67 μM RecAE38K were incubated with the components mentioned above for 10 min followed by the addition of 3

[3] The abbreviations used are: nt, nucleotide(s); ATPγS, adenosine 5′-*O*-(thiotriphosphate).

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

mM ATP and a 20-min incubation. Eight micromolar nucleotides M13mp18 circular dsDNA were added, and the reactions were incubated for another 20 min. Then 48 nM Ref was added. The reactions were stopped by phenol chloroform extraction and ethanol precipitation after 3 h. The resulting pellet was resuspended in EcoRI buffer and cut with EcoRI at 37 °C for 3 h. The digest was stopped by the addition of 10 μl of loading dye (9% Ficoll, 0.25% bromphenol blue, 0.25% xylene cyanol, and 4% SDS) followed by another 30-min incubation at 37 °C. Samples were subjected to electrophoresis in 0.8% agarose gels with TAE buffer, stained with SYBR-Gold nucleic acid stain (Invitrogen), and exposed to UV light.

Nuclease Site-specific Targeting Assay to Define Cut Sites—Targeting assays were carried out as described above up until the restriction digest, with the following exception. The 150-nt oligonucleotide used (mcg3) is homologous to bases 6070–6219 of M13mp18 DNA and is 54 nt (top strand) and 50 nt (bottom strand) away from the PstI restriction site. Reactions were digested with PstI at 37 °C for 3 h. The digest was stopped by phenol chloroform extraction and ethanol precipitation. The resulting pellet was resuspended in T4 DNA ligase buffer. Ligations to fluorescently labeled linkers (0.65 pmol) were carried out using 2 μl of T4 DNA ligase and incubating for 1 h at room temperature. Linker 1 (L1) consisted of den7 annealed to mcg2, and linker 2 (L2) consisted of den4 annealed to mcg1. Den7 and den4 were phosphorylated at the 5′ OH using polynucleotide kinase (Promega) at 37 °C for 30 min according to the manufacturer's instructions before annealing. The annealing reactions were carried out as described under electrophoretic mobility shift assays. The linkers were designed to have a sticky end complementary to the end created by PstI and a label on the blunt end. The ligation reactions were stopped by ethanol precipitation of the samples. The resulting pellets were resuspended in 90% formamide, 10% EDTA, heated at 95 °C for 10 min, and quick-cooled in an ice-water slurry for 10 min. The samples were then loaded on a 10% denaturing acrylamide sequencing gel and run at 30 watts (1600 V) for 4.5 h in TBE buffer. The fluorescently labeled DNA was then visualized using a Typhoon 9410 Variable Mode Imager (Amersham Biosciences) blue laser at 488 nm.

Structure Determination—P1 Ref (22 mg/ml in 20 mM Tris-HCl, pH 8.0, 200 mM NaCl) was mixed with mother liquor (0.2 M ammonium nitrate, 20% PEG 3350) in a 1:1 ratio and suspended over 1 ml of mother liquor in hanging-drop vapor diffusion experiments to generate crystals. Crystals were transferred to a cryoprotectant solution (0.15 M ammonium nitrate, 22% PEG 3350, 25% ethylene glycol) and flash-frozen in liquid nitrogen before data collection. The structure of the Ref was determined by single-wavelength anomalous dispersion phasing that took advantage of the anomalous scattering of the bound Zn^{2+} ions (Table 2). Diffraction data were collected at a suboptimal wavelength for Zn^{2+} anomalous scattering because it preceded discovery that Ref binds Zn^{2+}. However the single-wavelength anomalous dispersion phases calculated from the dataset were sufficient to produce an excellent experimental electron density map for model building. The data were indexed and scaled using HKL2000 (43). Zinc positions were identified, and an initial structure was built using Phenix (44).

TABLE 2
Diffraction data and crystal structure solution

Data collection	
Wavelength, Å	0.97856
Space group	$P3_221$
Unit cell (a, b, c (Å)/α, β, γ (°))	71.73, 71.73, 54.24/90, 90, 120
Resolution (last shell), Å	50-1.4 (1.42-1.40)
Reflection measured/unique	401,946/30,636
Multiplicity	13.1 (5.6)
Completeness (last shell), %	95.9 (69.4)
R_{sym}[a] (last shell), %	8.6 (43.5)
I/σ (last shell)	29.2 (2.1)
Phasing statistics	
Resolution, Å	50-1.4
Figure of merit (before/after density modification)	0.48/0.66
Refinement	
Resolution, Å	40-1.4
R_{work}/R_{free}[b] %	16.3/17.2
Atoms, number/⟨B factor⟩	
Protein[c]	925/13.5
Waters	127/32.0
Ligands (2 Zn^{2+} and 1 SO_4)	7/17.6
Root mean square deviation bond lengths, Å	0.008
Root mean square deviation bond angles, °	1.18
Ramachandran statistics (% most favored/allowed/additionally allowed/disallowed)	88.3/10.6/1.1/0

[a] $R_{sym} = \Sigma\Sigma j|Ij - \langle I\rangle|\Sigma Ij$, where Ij is the intensity measurement for reflection, and $\langle I\rangle$ is the mean intensity for multiply recorded reflections.
[b] $R_{work}/R_{free} = \Sigma||F_{obs}| - |F_{calc}||/|F_{obs}|$, where the working and free R factors are calculated by using the working and free reflection sets, respectively. The free R reflections (5% of the total) were held aside throughout refinement.
[c] Several side chains were modeled in multiple rotamers/conformations; in cases where atoms have been modeled in multiple conformations, each modeled position is counted in the number of protein atoms (*i.e.* if a given atom is modeled in two positions it is counted as two protein atoms).

The structure was improved by rounds of manual fitting using Coot (45) and refinement against the native data set using REFMAC5 (46). Coordinate and structure factor files have been deposited at the Protein Data Bank (PDB ID 3PLW).

Assay for Quantification of Zn^{2+} Bound to Ref—The Zn^{2+} content of wt Ref and of Ref H153A was measured using 4-(2-pyridylazo)resorcinol. 4-(2-Pyridylazo)resorcinol binding to Zn^{2+} causes an increase in absorbance at 490 nm (47). Protein samples were dialyzed against Ref storage buffer (R plus 200 mM potassium glutamate but not containing EDTA) overnight at 4 °C. 100 μl of either wt Ref or Ref H153A were incubated with 60 μg of proteinase K for 1 h at 37 °C. Then 78.4 μl of that sample was added to 1.6 μl of 5 mM 4-(2-pyridylazo)resorcinol and incubated at room temperature for 20 min, and the absorbances at 490 nm were measured. These measurements were compared with a standard curve of $ZnCl_2$ samples ranging from 1 to 10 μM $ZnCl_2$ in the same buffer.

RESULTS

To examine the mechanism underlying Ref protein functions with RecA, recombinant Ref protein was expressed and purified by standard methods. The only activity identified for pure Ref protein alone was DNA binding. Ref binds to (but does not cleave) both ssDNA and dsDNA in the absence of any other proteins or nucleotide cofactors (Fig. 1), consistent with results published previously (30).

Ref Is a RecA-dependent Nuclease—All proteins used in this work, including all Ref protein and Ref variant preps, were tested carefully by themselves for exo- and endonuclease activities on both ssDNA and dsDNA. The screen employs protein concentrations at or above the highest used in this work and

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

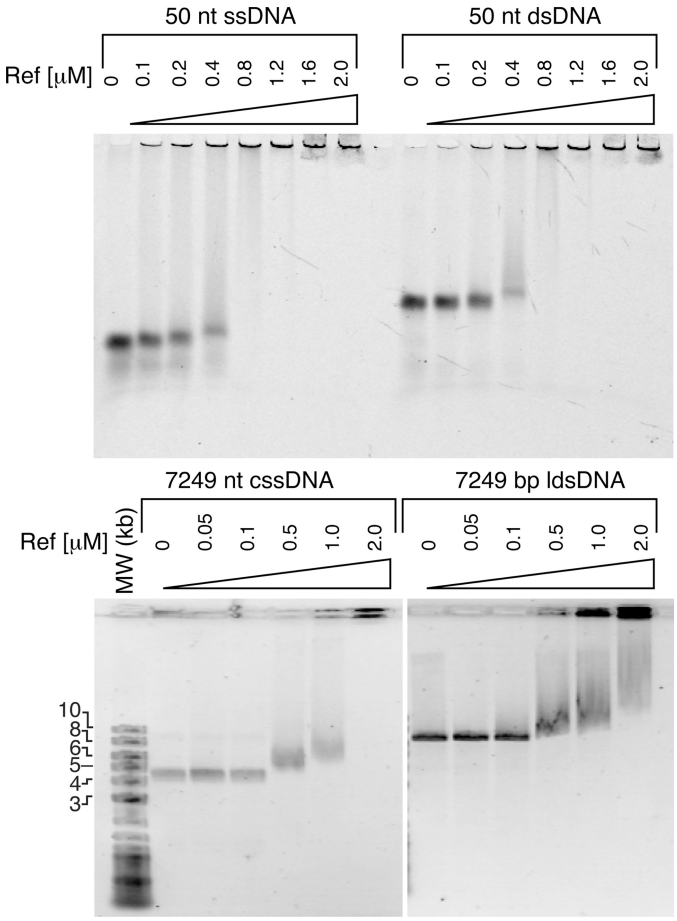


FIGURE 1. **Ref binds ssDNA and dsDNA.** Electrophoretic mobility shift assays were run on an 8% native acrylamide gel for the 50-nt DNA substrates and on a 0.8% agarose gel for the 7249-nt DNA. Ref was added at the concentrations indicated in the figure. At high Ref concentrations the complex formed was too large to enter the gel.

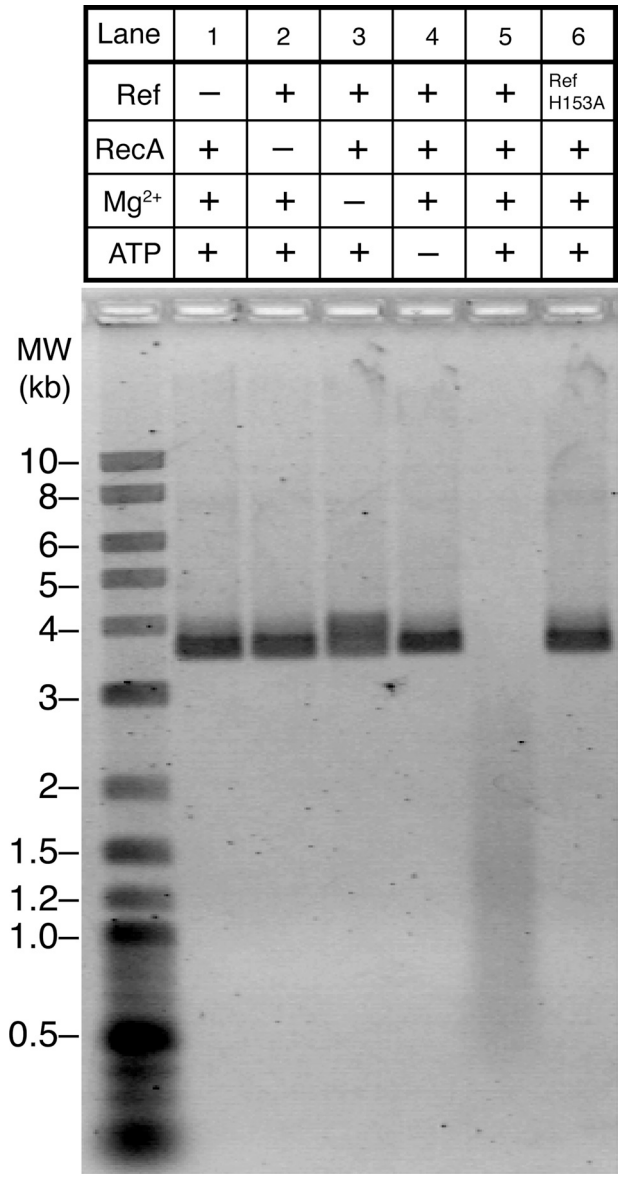


FIGURE 2. **Ref degrades ssDNA in a RecA, ATP, and Mg^{2+}-dependent manner.** Electrophoretic analysis of M13mp18 circular ssDNA incubated with indicated components is shown. RecA (2.4 μM) was added to reaction mixtures containing RecA buffer (see "Experimental Procedures") and 4 μMnt DNA for 10 min followed by the addition of 3 mM ATP. Ref protein (24 nM) was added 15 min later. The reactions were stopped 20 min after Ref addition. In *lane 6* Ref H153A replaced wt Ref. Here and in other figures illustrating reactions with DNA, the 2-log ladder of duplex DNA fragments from New England Biolabs is used in the marker lane to provide a point of reference between figures.

incubation times at least equivalent to the longest employed. No nuclease was detected under standard reaction conditions, which includes 24 nM Ref protein. With the wild type Ref, we did note a minor endonuclease activity that nicked about 30% of the supercoiled dsDNA prep but only when we used a 100-fold higher (2.4 μM) concentration and a 2-h incubation (data not shown). No other exo- or endonuclease activities were detectable even at these higher levels of Ref. We further tested a wide range of branched and partially single-stranded DNA substrates for cleavage by wild type Ref protein under standard reaction conditions (supplemental Fig. S2). No DNA cleavage was detected for any of these, indicating that Ref does not recognize and cleave a particular type of DNA structure.

When Ref and RecA were introduced to reactions together, the results changed. Incubation of Ref with bacteriophage M13mp18 circular ssDNA in the presence of RecA protein, ATP, and Mg^{2+} unexpectedly produced extensive degradation of the ssDNA (Fig. 2). All four of the components in the reconstituted system were required for nuclease activity. In particular, the nuclease activity depended on both Ref and active RecA filaments (Fig. 2). ATP is needed to form active RecA nucleoprotein filaments, whereas Mg^{2+} may be used by both RecA and Ref.

The observed RecA-dependent nuclease activity was due to the Ref protein. The activity co-eluted precisely with Ref protein-containing fractions on a gel filtration column (supplemental Fig. S3), indicating that the activity was catalyzed by Ref and not a trace contaminant. In addition, as described below, Ref has structural characteristics that identify it as an HNH endonuclease.

The size distribution of the degraded ssDNA in the RecA-dependent reaction suggested random cleavage, and the average length of fragments decreased with time. After 20 min of incubation, the average fragment size was about 1.2 kb, indicating that about 6 cleavage events have occurred per DNA molecule

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

(3 nM total cleavage events) in the presence of 24 nM Ref protein. This indicates a turnover of just under 0.01 cleavage events per minute.

Ref-mediated DNA cleavage in the presence of RecA protein was selective in that ssDNA was cleaved but RecA-bound dsDNA was not. No degradation of dsDNA was detected under any set of conditions, including conditions (low pH or longer preincubation with RecA to allow filament nucleation) in which saturating levels of RecA protein were demonstrably bound as indicated by RecA-mediated ATP hydrolysis (data not shown).

Ref nuclease activity on ssDNA was observed under all conditions under which RecA filaments formed. Optimal concentrations of Ref protein are far substoichiometric to both RecA and ssDNA nucleotides (about one Ref monomer to 100 RecA monomers) (Fig. 3*A*). At Ref concentrations approaching RecA concentrations, nuclease activity is inhibited (Fig. 3*A*). The direct binding of Ref to DNA could possibly inhibit RecA filament formation and preclude Ref activation. If the RecA protein concentration is reduced to levels insufficient for normal RecA filament formation but equivalent to the standard concentrations of Ref protein we used, Ref nuclease activity was again suppressed (Fig. 3*B*).

The RecA dependence of the nuclease activity implies an interaction between RecA and Ref. In principle, the RecA-dependent nuclease reaction could occur at a RecA filament end. Alternatively, it could involve DNA strands exposed in the RecA filament groove. We first examined the effect of RecA-mediated ATP hydrolysis. DNA degradation is reduced when ATP is not hydrolyzed by RecA. This is true for wild type RecA protein incubated with the non-hydrolyzable ATP analog, ATPγS, (Fig. 4, *lane 9*), and when ATPase-deficient RecA mutants such as RecA K72R and RecA E38K K72R are used (Fig. 4, *lanes 10–11*). Inasmuch as ATP hydrolysis is coupled to RecA dissociation at the 5′-proximal filament end, this could indicate that Ref-mediated cleavage occurs at the disassembling end of a filament. However, the RecA E38K mutant protein (also called *recA*730 (48)), which disassembles much less than the wild type protein (supplemental Fig. S4), promoted the Ref cleavage reaction at least as well and sometimes better (20–30% increase in some assays (data not shown)) than the wild type RecA. The RecA E38K mutant, thus, replaced the wild type RecA in many assays with Ref. We also examined the effect of proteins that bind in the RecA filament groove. Ref-mediated nuclease activity is blocked by a non-cleavable variant of the LexA protein (S119A) that binds well enough to active RecA filaments to inhibit DNA strand exchange (49), but not by the DinI protein, the non-cleavable UmuD1 protein, or the wild type LexA protein (Fig. 5). All of these proteins appear to bind in distinct (and sometimes transient) ways to the major groove of a RecA filament (20). When single-stranded DNA binding protein is incubated with DNA before RecA is added, so as to inhibit RecA binding, Ref does not exhibit nuclease activity (Fig. 4, *lane 12*).

The X-ray Crystal Structure of Ref Reveals a Novel HNH-family Nuclease—A sequence data base search revealed only a small number of uncharacterized Ref homologs. These included six close homologs (>90% identity) encoded by bacteriophages or prophages related to P1 and three additional

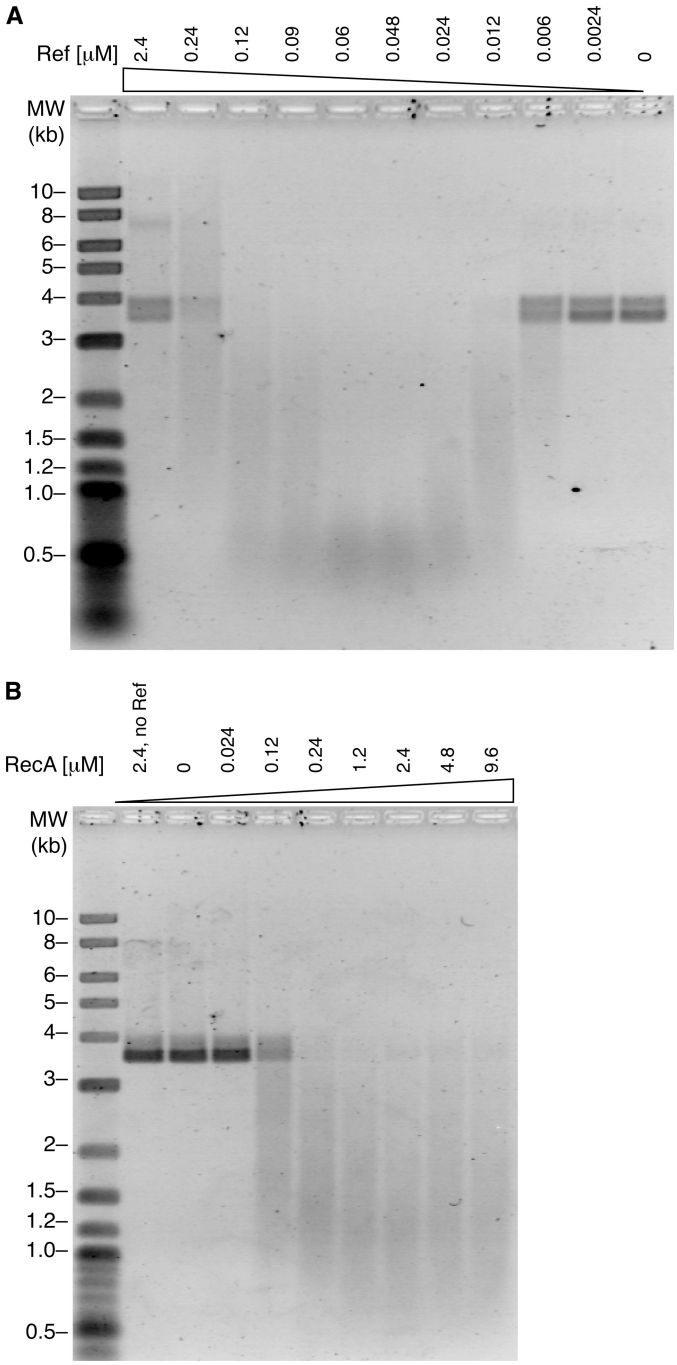


FIGURE 3. **Dependence of Ref ssDNA nuclease activity on RecA and Ref concentrations.** Nuclease assays were carried out as described under "Experimental Procedures." *A*, RecA concentration was constant at 2.4 μM, and Ref was added as indicated in the figure. *B*, RecA concentrations were varied as indicated in the figure, and 48 nM Ref were added.

bacterial genes encoding more distantly related proteins. A sampling of these is presented in Fig. 6*A*. The bacterial proteins, such as those from *Salmonella enterica* subsp. *enterica* serovar Newport strain SL317, ~65% identical, and *Bordetella avium* may be parts of cryptic prophages. It is not known if any of the detected homologs are active.

Comparison of this small sampling of Ref sequences revealed the presence of invariant Cys-Xaa-Xaa-Cys, Cys-Xaa-Xaa-His, and His-His motifs. Similar motifs are known to be involved in

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

Lane	1	2	3	4	5	6	7	8	9	10	11	12
Ref	–	+	–	+	–	+	–	+	+	+	+	+
RecA	wt	wt	wt	wt	wt	wt	wt	wt	wt	E38K K72R	K72R	wt
DNA	cds	cds	lds	lds	css	css	lss	lss	css	css	css	css
ATP	+	+	+	+	+	+	+	+	ATPγS	+	+	+
SSB	–	–	–	–	–	–	–	–	–	–	–	+

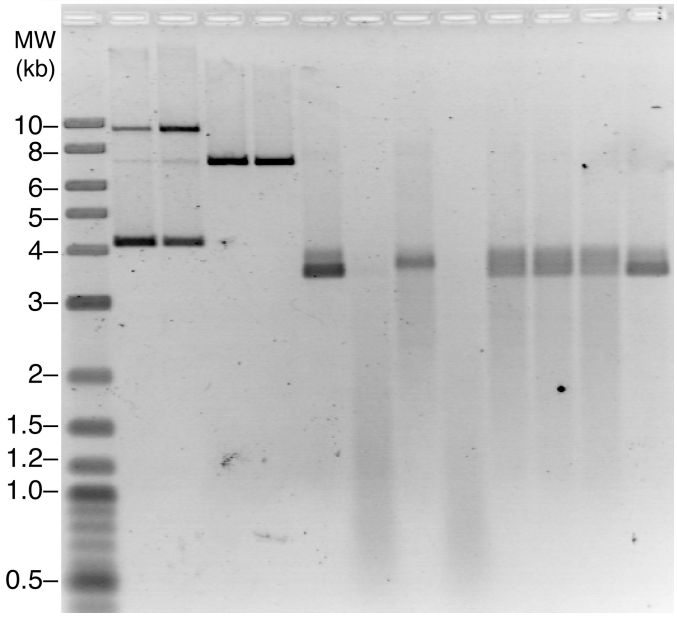


FIGURE 4. **Activity of Ref/RecA on different DNA substrates and with ATPase-deficient RecA variants.** Linear dsDNA (*lds*), supercoiled dsDNA (*cds*), circular ssDNA (*css*), and linear ssDNA (*lss*) substrates and RecA variants were used as indicated. Reaction conditions and protocols are as described in Fig. 2. In *lane 12*, 0.4 μM single-stranded DNA binding (*SSB*) was added to the DNA immediately before RecA was added. The linear ssDNA (*lane 7* and 8) was prepared by restriction digestion of circular M13mp18 ssDNA to which a short oligonucleotide had been annealed at the restriction site. The extensive secondary structure of M13mp18 ssDNA led to some nonspecific restriction digestion at other sites, producing some smearing of the DNA band.

Lane	1	2	3	4	5	6	7	8	9	10
Other	–	–	UmuD1	UmuD1	LexA	LexA	LexA S119A	LexA S119A	DinI	DinI
Ref	–	+	–	+	–	+	–	+	–	+

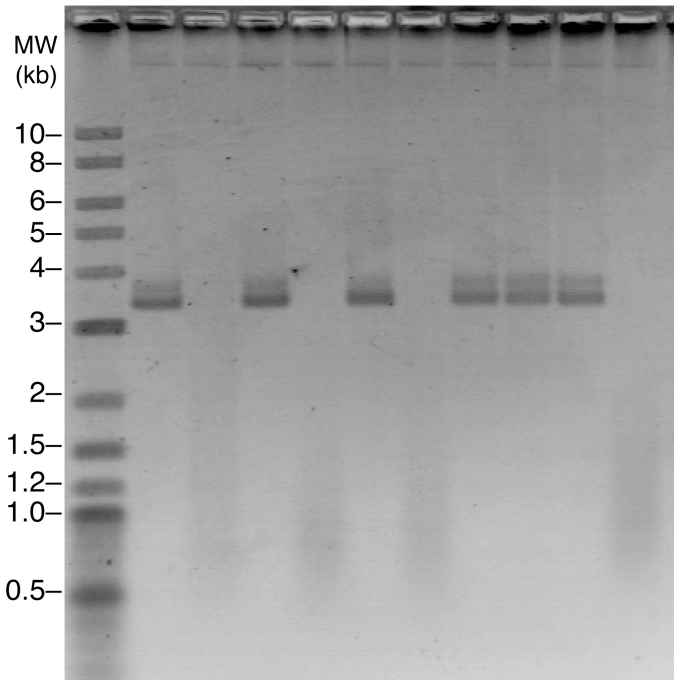


FIGURE 5. **Non-cleavable LexA mutant inhibits Ref nuclease activity.** Nuclease assays were carried out as described under "Experimental Procedures" with 24 nM Ref protein. UmuD1 (7.2 μM), LexA (7.2 μM), LexA S119A (7.2 μM), or DinI (16.8 μM) were added immediately before RecA was added. The presence or absence of Ref is indicated in the figure.

divalent metal binding (generally Zn^{2+}) in other proteins, which led us to test whether Ref binds Zn^{2+}. Zn^{2+} was associated with purified Ref at a 1.6:1 metal:protein ratio (Table 3). Because chelating agents were included during purification of Ref, these Zn^{2+} ions are likely to be stably bound to the protein.

A crystallographic approach was taken to better understand the structure and function of Ref. Crystals of Ref that diffracted to 1.4 Å resolution were produced, and the structure was determined by single-wavelength anomalous dispersion phasing that took advantage of anomalous scattering by the bound Zn^{2+} ions (Table 2). Consistent with the solution Zn^{2+} binding studies, the crystallographic asymmetric unit contained a single Ref protein bound to 2 Zn^{2+} ions (Fig. 6, *B* and C).

Ref folds as a globular protein, with a central two-stranded β-hairpin that is sandwiched between several α-helical and extended loop elements (Fig. 6*B*). Electron density for the N-terminal 76 residues was absent, indicating that it may be a flexible element. A careful search for difference density did not reveal any additional structural information about the missing N-terminal amino acids. To test whether the N terminus of the protein remained intact after crystallization, we washed and dissolved several crystals and subjected them to SDS acrylamide electrophoresis. The crystallized protein was completely intact (supplemental Fig. S5).

To test the functional importance of the N terminus, we constructed a deletion mutant of Ref that lacked the 76 N-terminal amino acids and purified it. It proved to be devoid of Ref-mediated nuclease activity on ssDNA under normal reaction conditions (Fig. 7, *lanes 11–13*). Some RecA-dependent ssDNA cleavage was seen at much higher concentrations of the mutant Ref protein (10× and 100×; Fig. 7, *lanes 14–21*). Ref ΔN76 was completely deficient in DNA binding to a 50-mer oligonucleotide up to a concentration of 2.4 μM using the assay shown in Fig. 1 (data not shown). This indicates that the disordered N-terminal domain plays an essential role the DNA binding activity of Ref and also enhances the RecA-dependent nuclease activity. We note that these N-terminal 76 amino acid residues represent a very highly charged part of the protein. Of the 76, 25 are amino acids with a positive charge (Arg or Lys (Fig. 6*A*)), and another 9 are negatively charged (Glu or Asp). Interestingly, the apparent homolog from *B. avium* has lost 75 of these 76 amino acid residues (Fig. 6*A*).

The central β-hairpin element of Ref is threaded through the core of the protein and presents ligands that define the Zn^{2+}-binding sites. The first binding site is composed of three His residues that are presented by the first β-strand of the hairpin, an adjacent α-helix, and a loop that is C-terminal to the hairpin

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

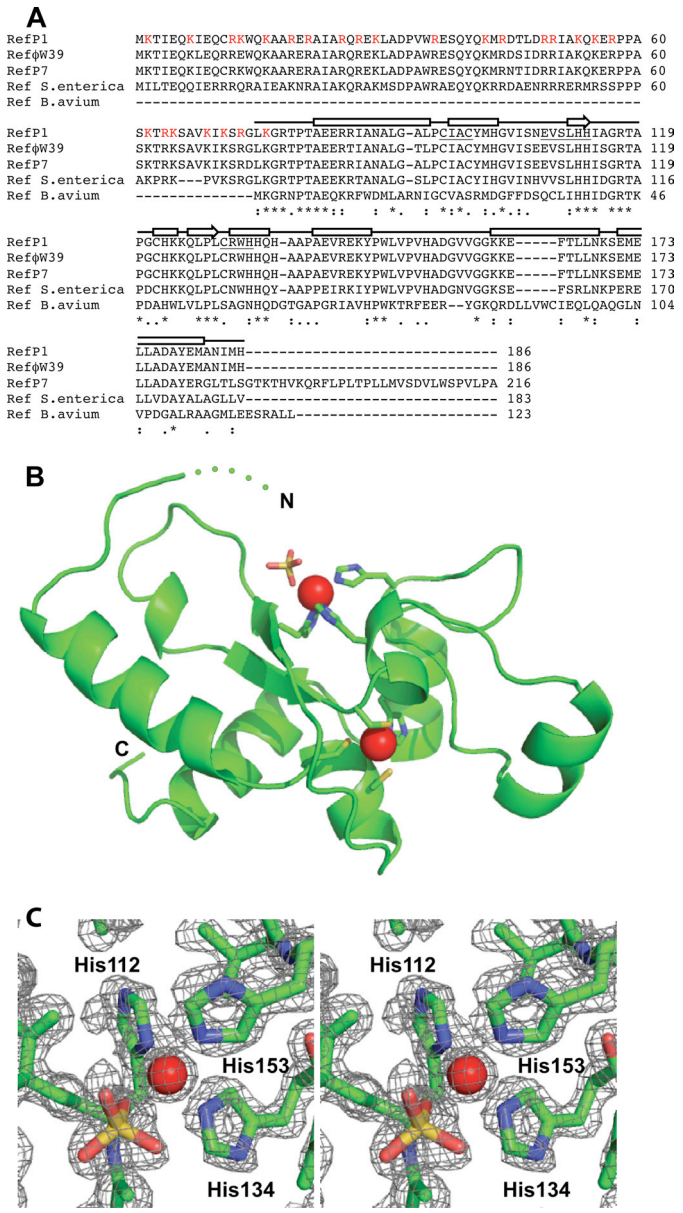


FIGURE 6. **The structure of bacteriophage P1 Ref protein.** *A*, sequence alignment of Ref protein homologs is shown. The homologs shown are from bacteriophages P7 and φW39 and from the bacteria *S. enterica* subsp. *enterica* serovar Newport strain SL317 and *B. avium*. Alignments were carried out with the program ClustalW (63). Invariant (*) and conserved (: or .) residues are marked. Electropositive residues in the P1 Ref N terminus are colored *red*, and sequences of importance are *underlined*. Elements of secondary structure are noted in the diagram *above the P1 Ref sequence*, with *open bars* representing α-helical segments and *open arrows* representing β conformation. *B*, the overall structure of the Ref protein (Zn^{2+} ions are in *red*) is shown. *C*, a sample of electron density showing amino acid side chains in the putative Ref active site is presented in stereo.

TABLE 3
Zn^{2+} quantification

Protein Sample	Protein	Zn^{2+}	Zn^{2+}:Protein	Average Zn^{2+}:protein[a]
	μM	μM		
wt Ref	2.66	4.63	1.74	1.59 (±0.16)
	3.37	4.79	1.42	
	2.39	3.84	1.61	
Ref H153A	2.47	4.37	1.77	1.55 (±0.20)
	3.45	5.19	1.50	
	2.84	3.93	1.38	

[a] Average of three independent experiments as represented in the three left columns.

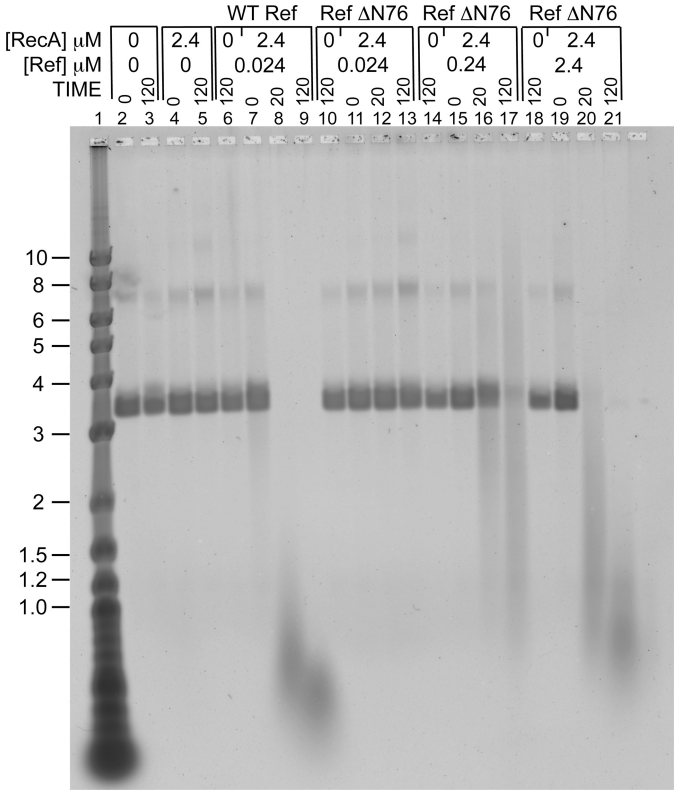


FIGURE 7. **The Ref ΔN76 mutant protein exhibits reduced activity for RecA-dependent cleavage of ssDNA.** Reactions were carried out as described under "Experimental Procedures" with the concentrations of RecA and Ref protein indicated at the *top of the figure*. The indicated incubation times are in minutes. As in other figures illustrating reactions with DNA, the 2-log ladder of duplex DNA fragments from New England Biolabs is used in the marker lane to provide a point of reference between figures.

(Fig. 6*B*). Interestingly, this Zn^{2+} is also liganded by a SO_4^{2-} ion, indicating that the Zn^{2+} bound at this site is solvent-exposed. The second site is composed of three Cys and one His residues from helical and loop elements in the structure. This second Zn^{2+} ion is entirely buried within the protein core.

Comparison of the Ref structure to other proteins in the Protein Data Bank using automated structure-comparison software failed to reveal any similar structures. However, we noticed that a motif in Ref bore similarity to an element in the HNH family of nucleases. HNH family members are defined by the presence of ββα-metal core elements in which residues from the β-hairpin/α-helical structure form a metal-binding site. Outside of this core, HNH enzymes are structurally diverse. The HNH family includes a number of bacteriophage-encoded homing endonucleases among others (50). Overlaying the β-hairpin core of Ref and the colicin E9 HNH DNase (51) revealed a striking similarity in the ββα-metal core between the two proteins (Fig. 8*A*). In both cases, the three Zn^{2+} binding His residues superimpose nearly identically, and a fourth Zn^{2+} ligand (SO_4^{2-} or PO_4^{3-}) is also nearly identically positioned. Outside of the ββα-metal core, the Ref and the E9 DNase fold lack tertiary or topological similarity. Even within the ββα-metal core there are differences. For E9 DNase, the second and third Zn^{2+} binding His residues are both presented from the α-helix of the ββα fold. In contrast, for Ref the first of these His residues is from the helix, but the second (His-153) is presented

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

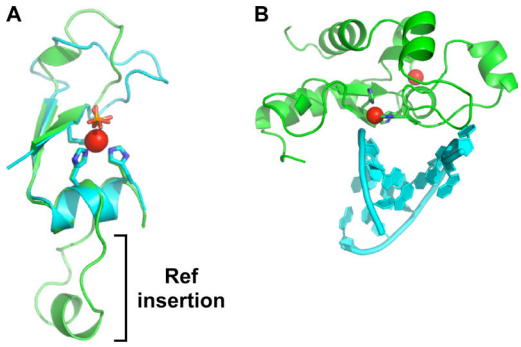


FIGURE 8. **The Ref protein active site.** *A*, the Ref crystal structure reveals an active site similar to HNH family endonucleases. Comparison of the $\beta\beta\alpha$ folds of Ref (*green*) and E9 DNase (*cyan*) is shown. An insertion in Ref relative to E9 DNase is labeled. *B*, superposition of DNA from the E9 DNase/DNA co-crystal structure (PDB ID 1V14) onto Ref is shown. This model was produced by superimposing the $\beta\beta\alpha$ folds of the two proteins.

by a extended loop element. This split arrangement of His residues distinguishes Ref from other HNH family members, partially explaining the lack of previous recognition of a possible nuclease function for Ref.

Superposition of the $\beta\beta\alpha$-metal core folds of Ref and the DNA-bound form of the E9 DNase allowed a DNA-bound structure of Ref to be modeled (Fig. 8*B*). This model predicts that the scissile phosphate would be positioned where the SO_4^{2-} ion is observed in the Ref structure and that the Zn^{2+} binding His residues also form portions of the nuclease active site. To test this model, a Ref variant in which His-153 is substituted with Ala was constructed, purified, and found to lack all nuclease activity (Fig. 2). The inactivity of this Ref H153A mutant is not due to a loss of the Zn^{2+} ion in the active site (Table 3), suggesting that the His-153 residue itself is important for catalysis. In addition, Ref H153A appears to be properly folded, as determined by circular dichroism (supplemental Fig. S1). Interestingly, elements that are important for DNA binding in E9 (52) are absent in the core Ref structure, further suggesting that the DNA binding activity is embedded in the N-terminal amino acids.

HNH-family nucleases are defined by the $\beta\beta\alpha$-metal core fold in which a central His or Asp residue is immediately flanked by an N-terminal Asp or His residue and at some distance by a C-terminal His, Asp, or Glu residue (53). For P1 Ref, this sequence is -His-His-(9 residues)-His- for residues 112–123. HNH-family members generally also have conserved Cys-Xaa-Xaa-Cys sequence motifs N- and C-terminal to this central sequence cluster that are used for binding metals; a His residue can substitute for one of the Cys residues in these motifs. P1 Ref also has such motifs (Cys-Ile-Ala-Cys (residues 96–99) and Cys-Arg-Trp-His (residues 130–133)). We note that the apparent homolog encoded by *B. avium* (Fig. 6*A*) has replaced a Cys with a Ser residue in both of these motifs. Mehta *et al.* (53) have defined 8 subclasses of HNH-family nucleases by comparing the sequences of 323 proteins using ClustalW. Comparing P1 Ref to these eight subclasses indicated that Ref did not fit into any of the published categories. Thus, we propose that Ref defines a new subclass of HNH-family nucleases.

Ref Activity Can Be Restricted to D-loops Created by RecA; Oligonucleotide-directed Introduction of Double-strand Breaks—During DNA strand exchange, the RecA protein can bind up to three strands of DNA (2, 54). As shown in Fig. 9, Ref will cleave both strands of a duplex DNA at the site of a D-loop formed by a RecA-coated oligonucleotide complementary to that site. In this experiment, the RecA E38K mutant protein replaced the wild type RecA protein as it consistently produced a 20–30% enhancement of activity relative to the wild type RecA protein (data not shown). *Lanes 9–12* of Fig. 9, employing four oligonucleotides targeted to three different sites in the duplex DNA circle, make two points. First, restriction enzyme analysis demonstrates that the Ref-induced breaks in the circular double-stranded DNA occur where the RecA-bound oligonucleotide invaded the duplex DNA to form a D-loop. This result demonstrates the targeting potential of the Ref/RecA system. It also provides additional evidence that Ref is activated for cleavage by direct interaction with RecA protein filaments. Second, the length of the oligonucleotide influences RecA activation of Ref. The best cleavage is seen with the two 150-mers. A 100-mer targeted to the same site as one of the 150-mers is nearly as effective. Cleavage efficiency drops off substantially for the 60-mer, corresponding to the decreased stability of RecA filaments on the shorter oligonucleotide (55). Like the ssDNA nuclease activity, targeted dsDNA nuclease activity depends on RecA, ATP, Mg^{2+}, and the presence of a homologous oligonucleotide (Fig. 9). The Ref H153A variant was again inactive.

We determined sites of DNA cleavage within the region targeted by one of the 150-mer oligonucleotides, cutting the reaction product at a nearby restriction site (PstI) and ligating a labeled linker to the resulting fragment. A gel allowing dideoxy sequencing-like resolution reveals that Ref-mediated cuts are found at multiple locations, all relatively close to the 3′ end of the invading oligonucleotide (Fig. 10). We note that the method employed would eliminate signals from distal cleavage sites for any DNA that was cleaved twice on the same strand of the D-loop and, thus, may bias the results toward detection of 3′ end-proximal events. However, the results clearly indicate that cleavage can occur at multiple locations within the D-loop. Preferred cleavage sites were evident, although the mechanistic basis of the pattern is not yet understood. The patterns also indicate that Ref-mediated introduction of double-strand breaks does not necessarily generate blunt ends.

DISCUSSION

There are four major conclusions to this work. First, the bacteriophage P1-encoded Ref protein is an endonuclease, with the novel property that nuclease function is entirely dependent on the presence of active RecA nucleoprotein filaments. Second, Ref is not simply activated by RecA. Where cleavage locations can be correlated with RecA locations (Fig. 9), Ref cleaves DNA where RecA protein is bound. Thus, the work defines a new function for RecA nucleoprotein filaments, that of co-nuclease. Third, our structural analysis indicates that Ref protein defines a new subfamily of HNH endonucleases. Finally, the Ref/RecA system can be used to introduce targeted double-strand breaks at any chosen location in an oligonucleotide-directed manner. The enhanced RecA-dependent recombination observed *in*

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

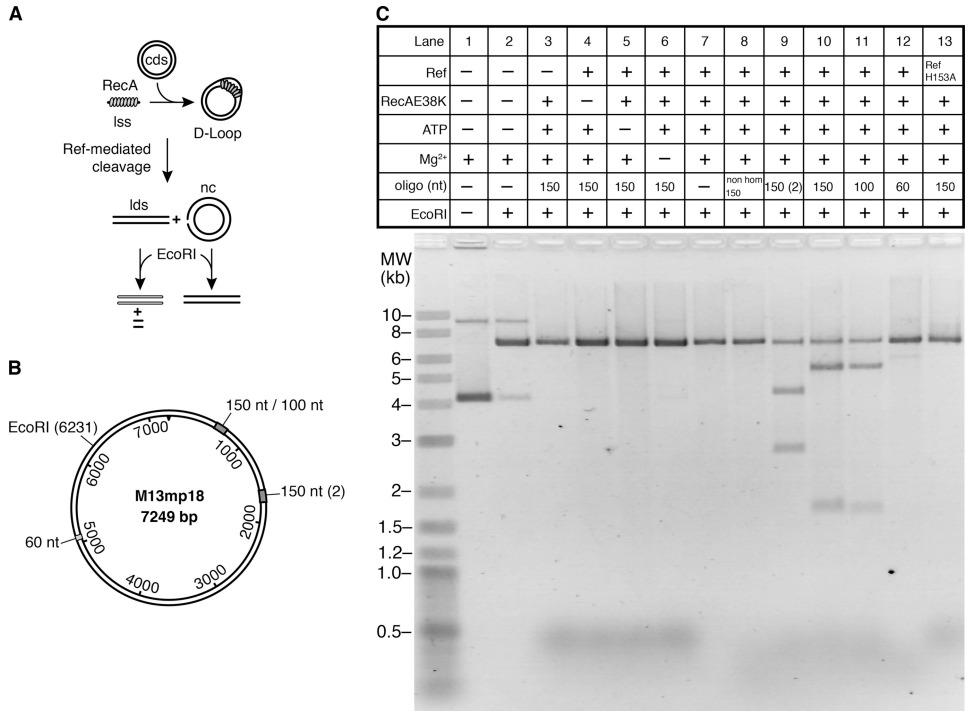


FIGURE 9. **RecA-mediated D-loop formation results in directed dsDNA cleavage by Ref.** *A*, the reaction scheme is shown. RecA (0.67 μM) was incubated with oligonucleotide (4 μMnt) for 20 min with 3 mM ATP to allow RecA filament formation. Circular dsDNA (8 μMnt) was added, and D-loop formation occurred. After 20 min, Ref (48 nM) was added to the reactions followed by a 180-min incubation producing linear dsDNA and leaving some uncleaved dsDNA. The reactions were stopped, and the DNA was digested using EcoRI. This produced cleavage products of expected size when the DNA had been linearized by Ref. *lds*, linear dsDNA; *lss*, linear ssDNA; *cds*, supercoiled dsDNA. *B*, a map of M13mp18 circular dsDNA with locations of oligonucleotides and restriction site is shown. *C*, shown is electrophoretic analysis of reactions carried out as described in *A*. Components were omitted from the reactions as indicated *above the pane;*. RecA E38K was used instead of wt RecA because it produced a slightly increased product yield. In *lane 8* a 150-nt oligo that was not homologous to M13mp18 was used. In *lanes 9* and *10*, two 150-nt oligos complementary to M13mp18 at different locations were used (see B). In *lane 11*, a 100-nt oligo complementary to M13mp18 at the same location as the 150-nt oligo in *lane 10* was used. In *lane 12*, a 60-nt oligo homologous to M13mp18 at a different location replaced the 150-mer. In *lane 13*, Ref H153A replaced wt Ref. *Arrows* point to the cleavage products of expected size detectable in *lanes 9–12*. *nc*, nicked double strand circle.

vivo when Ref is present (29–31) does not reflect regulation of RecA. Instead, through Ref, bacteriophage P1 employs a strategy for the stimulation of homologous recombination that is seen during meiotic recombination in eukaryotes, the directed generation of double-strand breaks.

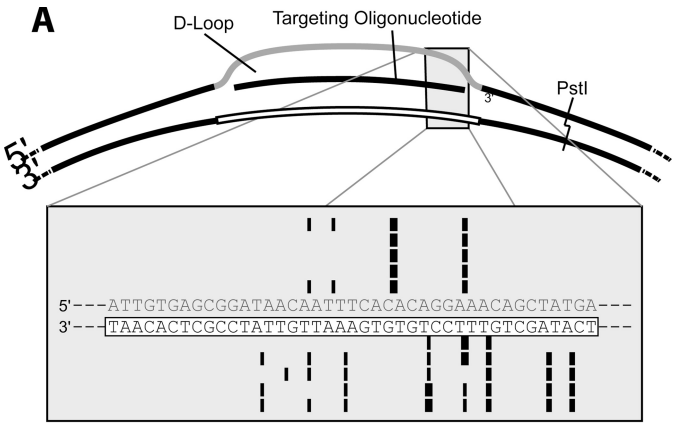
Multiple experiments indicate that the observed nuclease activity is due to Ref and not to a contaminating nuclease activity. All protein preparations used in this study were carefully tested and found free of exo- or endonuclease activity on circular or linear single- or double-stranded DNA under standard reaction conditions. The observed nuclease activity co-elutes with the Ref protein on a size exclusion column. Finally, the structure of the Ref protein identifies it as a novel HNH endonuclease.

These results further broaden the already long list of functions for RecA filaments to include co-nuclease. The Ref protein cleaves only where RecA protein is bound to DNA. The reaction is enhanced if the RecA filament actively hydrolyzes ATP. In principle, Ref could cleave at a filament end or in the RecA filament groove. The enhancement by ATP hydrolysis might suggest a link to end-dependent RecA filament dissociation. However, we currently favor a mechanism in which cleavage occurs in the filament groove for two major reasons. First, the Ref-mediated cleavage reaction is also enhanced when the RecA E38K mutant protein replaces the wild type RecA. The filaments formed by the E38K mutant protein are much less dynamic than those of the wild type protein, exhibiting no measurable dissociation in a standard challenge assay when they are bound to ssDNA. Second, the Ref-mediated cleavage reaction is completely inhibited by the LexA S119A mutant protein, which binds stably in the RecA filament groove. RecA-mediated ATP hydrolysis is not limited to the ends of RecA filaments (56) but occurs throughout the filament. Conformation changes associated with the ATP hydrolytic cycle may play some role in the Ref nuclease reaction. A structure of RecA bound to DNA in the presence of ATP has appeared (57). However, the core domain of RecA is closely related to helicases in which substantial conformational changes are associated with ATP hydrolysis (58–61). For RecA, little is known about the structural changes associated with ATP hydrolysis or the status of bound DNA strands at different stages of that cycle.

The Ref protein features a core structure with a clear relationship to HNH endonucleases. Elimination of an active site His residue (His-153) eliminated Ref function. The 76 N-terminal residues of Ref were disordered in the structure. This region of the protein features 34 charged amino acid side chains (25 Lys or Arg), and results so far indicate that it is the region responsible for the DNA binding activity of Ref.

Ref will cleave RecA-bound ssDNA as well as DNA to which RecA nucleoprotein filaments are paired. When RecA filaments are restricted to oligonucleotides and used to form D-loops, the Ref/RecA system becomes a uniquely programma-

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011



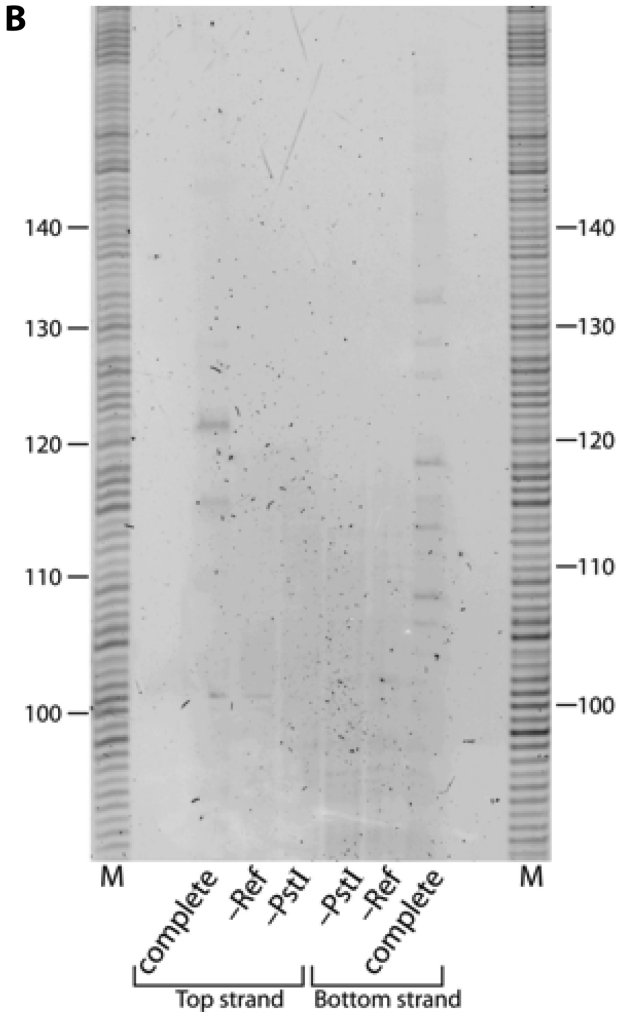


FIGURE 10. **Ref cleaves DNA at multiple sites within the D-loop.** *A*, cleavage sites detected in five independent targeted cleavage experiments are shown. The duplex DNA sequences corresponding to the 3′ end of the 150-mer targeting oligonucleotide are shown in the *gray box*, corresponding to the region containing the detected Ref-mediated cuts in the M13mp18 DNA. The thickness of the marks corresponds approximately to the intensity of the bands on the sequencing gel. *Each row* of marks represents one independent experiment. *B*, shown is an example of site-determination experiment. Reactions were carried out as described under "Experimental Procedures." After Ref-mediated cleavage of targeted DNA within D-loops, the DNA was isolated and cleaved with PstI restriction endonuclease. A labeled oligonucleotide was ligated to the cleaved end followed by sizing of the products on a sequencing gel.

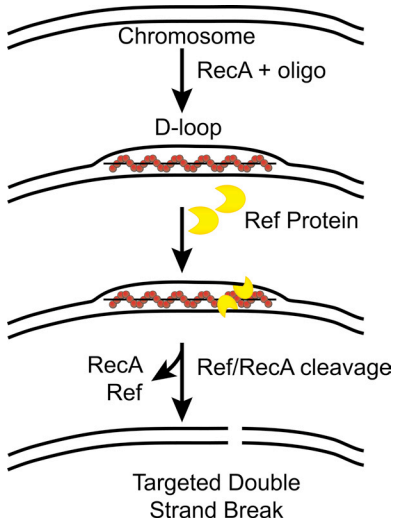


FIGURE 11. **Proposed scheme for the targeted cleavage of duplex DNA within RecA-mediated D-loops by the Ref protein of bacteriophage P1.**

ble nuclease system. Ref cleaves both strands of the targeted duplex DNA within the D-loop (Figs. 9 and 10), introducing a double-strand break within a small target area. Cleavage can, thus, be introduced in an oligonucleotide-directed fashion. Used this way, Ref/RecA may be considered a universal restriction enzyme. The targeting reaction is illustrated in Fig. 11.

Cleavage of RecA-created D-loops has previously been reported for the S1 and Bal-31 nucleases (62), both normally specific for single-stranded DNA. The unwound DNA at the ends of D-loops offers a target for such nucleases. The effects of the Ref nuclease differ from this earlier report in several important ways. First, neither S1 nor Bal-31 has any RecA-dependent phenotype *in vivo*. Second, both S1 and Bal31 will cleave single-stranded DNA wherever it occurs, including at nicks and small gaps; Ref will cleave only where RecA protein is bound. It is likely that Ref evolved to work with RecA filaments. There is no such indication for the S1 and Bal-31 nucleases. Finally, the more proficient of the two nucleases, S1, requires a non-physiological set of reaction conditions to carry out its reaction efficiently.

The efficiency of oligonucleotide-directed Ref cleavage of DNA appears to be strongly dependent on the efficiency of RecA-mediated D-loop formation. This efficiency in turn can be modulated by the length of the DNA oligonucleotide, RecA mutant proteins with enhanced DNA binding and/or DNA pairing properties (the RecA E38K mutant protein is the most effective variant we have tested to date) and likely other factors that remain to be elucidated. Efficient targeted cleavage by the Ref/RecA system may eventually provide a convenient and inexpensive method to introduce targeted double-strand breaks in a range of biotechnology applications.

Acknowledgments—We thank E. A. Wood for cloning the wild type ref gene and the ref H153A mutant, D. R. McCaslin for carrying out the CD spectroscopy, and Grzegorz Sabat of the mass spectrometry facility in the University of Wisconsin Biotechnology Center for assistance with protein identity confirmation.

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011

REFERENCES

1. Yu, X., and Egelman, E. H. (1992) *J. Mol. Biol.* **227,** 334–346
2. Lusetti, S. L., and Cox, M. M. (2002) *Ann Rev Biochem.* **71,** 71–100
3. Walker, G. C., Smith, B. T., and Sutton, M. D. (2000) in *Bacterial Stress Responses* (Storz, G., and HenggeAronis, R., eds) pp. 131–144, American Society for Microbiology, Washington, D. C.
4. Jiang, Q., Karata, K., Woodgate, R., Cox, M. M., and Goodman, M. F. (2009) *Nature* **460,** 359–363
5. Patel, M., Jiang, Q., Woodgate, R., Cox, M. M., and Goodman, M. F. (2010) *Crit. Rev. Biochem. Mol. Biol.* **45,** 171–184
6. Cox, M. M. (2007) *Crit. Rev. Biochem. Mol. Biol.* **42,** 41–63
7. Lusetti, S. L., Voloshin, O. N., Inman, R. B., Camerini-Otero, R. D., and Cox, M. M. (2004) *J. Biol. Chem.* **279,** 30037–30046
8. Drees, J. C., Lusetti, S. L., and Cox, M. M. (2004) *J. Biol. Chem.* **279,** 52991–52997
9. Drees, J. C., Lusetti, S. L., Chitteni-Pattu, S., Inman, R. B., and Cox, M. M. (2004) *Mol. Cell* **15,** 789–798
10. Stohl, E. A., Brockman, J. P., Burkle, K. L., Morimatsu, K., Kowalczykowski, S. C., and Seifert, H. S. (2003) *J. Biol. Chem.* **278,** 2278–2285
11. Drees, J. C., Chitteni-Pattu, S., McCaslin, D. R., Inman, R. B., and Cox, M. M. (2006) *J. Biol. Chem.* **281,** 4708–4717
12. Petrova, V., Chitteni-Pattu, S., Drees, J. C., Inman, R. B., and Cox, M. M. (2009) *Mol. Cell* **36,** 121–130
13. Hobbs, M. D., Sakai, A., and Cox, M. M. (2007) *J. Biol. Chem.* **282,** 11058–11067
14. Morimatsu, K., and Kowalczykowski, S. C. (2003) *Mol. Cell* **11,** 1337–1347
15. Sakai, A., and Cox, M. M. (2009) *J. Biol. Chem.* **284,** 3264–3272
16. Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M. A. (2005) *EMBO J.* **24,** 180–189
17. Churchill, J. J., Anderson, D. G., and Kowalczykowski, S. C. (1999) *Genes Dev.* **13,** 901–911
18. Churchill, J. J., and Kowalczykowski, S. C. (2000) *J. Mol. Biol.* **297,** 537–542
19. Spies, M., Dillingham, M. S., and Kowalczykowski, S. C. (2005) *J. Biol. Chem.* **280,** 37078–37087
20. VanLoock, M. S., Yu, X., Yang, S., Galkin, V. E., Huang, H., Rajan, S. S., Anderson, W. F., Stohl, E. A., Seifert, H. S., and Egelman, E. H. (2003) *J. Mol. Biol.* **333,** 345–354
21. Yu, X., and Egelman, E. H. (1993) *J. Mol. Biol.* **231,** 29–40
22. Frank, E. G., Cheng, N., Do, C. C., Cerritelli, M. E., Bruck, I., Goodman, M. F., Egelman, E. H., Woodgate, R., and Steven, A. C. (2000) *J. Mol. Biol.* **297,** 585–597
23. Kim, B., and Little, J. W. (1993) *Cell* **73,** 1165–1173
24. Bagdasarian, M., Bailone, A., Angulo, J. F., Scholz, P., Bagdasarian, M., and Devoret, R. (1992) *Mol. Microbiol.* **6,** 885–893
25. Bailone, A., Bäckman, A., Sommer, S., Célérier, J., Bagdasarian, M. M., Bagdasarian, M., and Devoret, R. (1988) *Mol. Gen. Genet.* **214,** 389–395
26. Bertani, G. (1951) *J. Bacteriol.* **62,** 293–300
27. Łobocka, M. B., Rose, D. J., Plunkett, G., 3rd, Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M. B., and Blattner, F. R. (2004) *J. Bacteriol.* **186,** 7032–7068
28. Yarmolinsky, M. B. (2004) *J. Bacteriol.* **186,** 7025–7028
29. Windle, B. E., and Hays, J. B. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83,** 3885–3889
30. Lu, S. D., Lu, D., and Gottesman, M. (1989) *J. Bacteriol.* **171,** 3427–3432
31. Laufer, C. S., Hays, J. B., Windle, B. E., Schaefer, T. S., Lee, E. H., Hays, S. L., and McClure, M. R. (1989) *Genetics* **123,** 465–476
32. Windle, B. E., Laufer, C. S., and Hays, J. B. (1988) *J. Bacteriol.* **170,** 4881–4889
33. Craig, N. L., and Roberts, J. W. (1981) *J. Biol. Chem.* **256,** 8039–8044
34. Lohman, T. M., and Overman, L. B. (1985) *J. Biol. Chem.* **260,** 3594–3603
35. Koch, W. H., Ennis, D. G., Levine, A. S., and Woodgate, R. (1992) *Mol. Gen. Genet.* **233,** 443–448
36. Slilaty, S. N., and Little, J. W. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84,** 3987–3991
37. Gruenig, M. C., Renzette, N., Long, E., Chitteni-Pattu, S., Inman, R. B., Cox, M. M., and Sandler, S. J. (2008) *Mol. Microbiol.* **69,** 1165–1179
38. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78
39. Neuendorf, S. K., and Cox, M. M. (1986) *J. Biol. Chem.* **261,** 8276–8282
40. Haruta, N., Yu, X., Yang, S., Egelman, E. H., and Cox, M. M. (2003) *J. Biol. Chem.* **278,** 52710–52723
41. Marrione, P. E., and Cox, M. M. (1995) *Biochemistry* **34,** 9809–9818
42. Robu, M. E., Inman, R. B., and Cox, M. M. (2004) *J. Biol. Chem.* **279,** 10973–10981
43. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
44. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221
45. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
46. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr. D Biol. Crystallogr.* **57,** 122–133
47. Hunt, J. B., Neece, S. H., Schachman, H. K., and Ginsburg, A. (1984) *J. Biol. Chem.* **259,** 14793–14803
48. Lavery, P. E., and Kowalczykowski, S. C. (1992) *J. Biol. Chem.* **267,** 20648–20658
49. Harmon, F. G., Rehrauer, W. M., and Kowalczykowski, S. C. (1996) *J. Biol. Chem.* **271,** 23874–23883
50. Stoddard, B. L. (2005) *Q. Rev. Biophys.* **38,** 49–95
51. Maté, M. J., and Kleanthous, C. (2004) *J. Biol. Chem.* **279,** 34763–34769
52. Garinot-Schneider, C., Pommer, A. J., Moore, G. R., Kleanthous, C., and James, R. (1996) *J. Mol. Biol.* **260,** 731–742
53. Mehta, P., Katta, K., and Krishnaswamy, S. (2004) *Protein Sci.* **13,** 295–300
54. Cox, M. M. (2003) *Annu. Rev. Microbiol.* **57,** 551–577
55. McIlwraith, M. J., and West, S. C. (2001) *J. Mol. Biol.* **305,** 23–31
56. Brenner, S. L., Mitchell, R. S., Morrical, S. W., Neuendorf, S. K., Schutte, B. C., and Cox, M. M. (1987) *J. Biol. Chem.* **262,** 4011–4016
57. Chen, Z., Yang, H., and Pavletich, N. P. (2008) *Nature* **453,** 489–494
58. Geiselmann, J., Wang, Y., Seifried, S. E., and von Hippel, P. H. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90,** 7754–7758
59. Singleton, M. R., Sawaya, M. R., Ellenberger, T., and Wigley, D. B. (2000) *Cell* **101,** 589–600
60. Soultanas, P., and Wigley, D. B. (2000) *Curr. Opin. Struct. Biol.* **10,** 124–128
61. Yong, Y., and Romano, L. J. (1995) *J. Biol. Chem.* **270,** 24509–24517
62. Shigemori, Y., and Oishi, M. (2004) *Nucleic Acids Res.* **32,** e4
63. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680

Downloaded from www.jbc.org at University of Wisconsin-Madison, on April 12, 2011