Molecular Mechanism of Transcription-Repair Coupling

Christopher P. Selby and Aziz Sancar

Lesions in the transcribed strand block transcription and are repaired more rapidly than lesions in the nontranscribed (coding) strand which do not block RNA polymerase (RNAP). It has been shown previously that in *Escherichia coli* the *mfd* (mutation frequency decline) gene is necessary for strand-specific repair. The *mfd* gene was cloned and sequenced and the Mfd protein was purified and used to reconstitute strand-specific repair in a completely defined system. The *mfd* gene encodes a protein of 130 kilodaltons and contains the so-called "helicase motifs," a leucine zipper motif, and regions of sequence similarity to UvrB and RecG proteins. The Mfd protein was shown to (i) displace RNAP stalled at a lesion in an adenosine triphosphate–dependent reaction, (ii) bind to the damage recognition subunit (UvrA) of the excision nuclease, and (iii) stimulate the repair of the transcribed strand only when transcription is taking place. Thus, Mfd appears to target the transcribed strand for repair by recognizing a stalled RNAP and actively recruiting the repair enzyme to the transcription blocking lesion as it dissociates the stalled RNAP.

DNA lesions that impede DNA and RNA polymerases may kill cells by interfering with replication in actively growing and dividing cells or by blocking transcription and thus depriving nondividing cells of an essential protein. Thus, targeting DNA repair enzymes to genes that are transcriptionally active is advantageous for cell survival, especially the survival of nondividing cells. Bohr *et al.* (*1*) and Mellon and Hanawalt (*2*) discovered that actively transcribing genes in mammalian cells and in *E. coli*, respectively, were repaired more rapidly than nontranscribed genes or regions of the genome. An important clue to a biochemical understanding of "preferential repair" came from the discovery that in the dihydrofolate reductase gene of CHO cells (*3*) and in the *lacZ* gene of *E. coli* (*2*) the increased rate of repair was confined to the template (transcribed) strand. Although in eukaryotic systems, for certain lesions and in certain genes, enhanced repair occurs in both strands, as a general rule transcription blocking lesions are repaired faster only in the template strand (*4*).

A plausible model for the efficient repair of the transcribed strand is that a stalled RNAP complex creates a high affinity site for the damage recognition subunit of the repair enzyme, which, in the case of *E. coli*, is the UvrA subunit of (A)BC excinuclease (*2*, *5*). However, when this model was tested in vitro in a defined system, it was found that RNAP stalled at a lesion did not stimulate but actually inhibited repair and led to the proposal that cells contain a transcription-repair coupling factor (TRCF) that (i) overcomes the repair inhibitory effect of stalled RNAP and (ii) recruits the repair enzyme to such sites (*5*). An in vitro system for investigating strand-specific repair in *E. coli* has been developed and a putative TRCF was partially purified (*6*). Clues from earlier genetic studies (*7*) led to the discovery that cell-free extracts from an *mfd*$^-$ mutant were incapable of strand-specific repair, and complementation was observed when partially purified TRCF was added to *mfd*$^-$ cell extracts (*8*). It was concluded that the *mfd* gene encoded the TRCF, with the prediction that in *mfd*$^-$ mutants the transcribed strand would be repaired poorly, and most of damage-induced mutations in mfd$^-$ cells should arise from lesions in the transcribed strand (*8*). Two studies have confirmed both of these predictions (*9*, *10*) and thus support the suggestion that *mfd* encodes the TRCF. However, the requirement for an additional factor could not be ruled out. In this article we show that the TRCF is the *mfd* gene product and that it is necessary and sufficient for transcription-repair coupling. A comprehensive model is described for the reaction mechanism on the basis of the properties of highly purified TRCF and its interactions with transcription and repair proteins.

Genetics, cloning, and sequence of the *mfd* gene. The one known *mfd* mutant allele, which had been isolated from an *E. coli* B/r background by Witkin (*11*), is 20 percent cotransducible with *umuCD*, at 26 minutes of the *E. coli* chromosome (*12*). Using transposon transduction, we moved the mutation into the *E. coli* K-12 strain AB1157, mapped the gene at 25.3 minutes, by means of genetic (*13*) and biochemical (*6*, *8*) assays. We then located the gene to phage 238 in the Kohara library mini-set (*14*) by Southern (DNA) hybridization with a 29-nucleotide (nt) degenerate oligomer (*15*), deduced from the amino-terminal sequence of a putative, partially purified TRCF protein band (*8*). The *mfd* gene was subcloned from this phage by screening for ultraviolet-resistant transformants of strain UNCR9 (*mfd*$^-$ recA$^-$) since, as described by Oller *et al.* (*10*), in *E. coli* K-12, the *mfd* mutation confers moderate ultraviolet sensitivity. The plasmid pMFD19, a pIBI25 derivative carrying the *mfd* gene on a 5 kbp Ssp I–Sph I fragment, complements the ultraviolet-sensitive phenotype of *mfd*$^-$ cells (Fig. 1) and confers wild-type strand-specific repair to extracts from transformed *mfd*$^-$ cells (see below).

We sequenced the *mfd* gene by the dideoxy method, using the sonication shotgun sequencing strategy (*16*). The gene encoded a protein of 1148 amino acids. The sequence of the 18 NH_2-terminal residues was identical with the sequence of the partially purified TRCF protein band (*8*). Thus, the Mfd protein and the TRCF are identical and these terms are used interchangeably.

The translated amino acid sequence (Fig. 2A) revealed three regions of special interest. Near the central region (Fig. 2B) were seven motifs that are conserved in the family of known or putative helicases (*17*). A stretch of 387 amino acids that includes the so-called helicase motifs (*17*) showed 38 percent sequence identity with the corresponding helicase motifs region of *E. coli* RecG protein (*18*) (Fig. 2B). The RecG protein promotes branch migration of a synthetic Holliday junction in an adenosine triphosphate (ATP)–dependent manner (*19*). The second region of interest, a stretch of 140 amino acids in the NH_2-terminus (Fig. 2C), showed 22 to 25 percent sequence identity with the three bac-

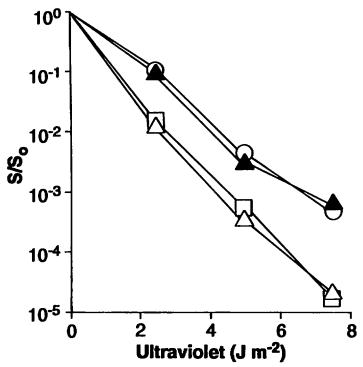

Fig. 1. Effect of ultraviolet on survival S/S_0 of an *E. coli–mfd*$^-$ *recA*$^-$ strain and complementation by the cloned *mfd* gene. Circle, UNCR10 (*recA*$^-$ *mfd*$^+$); open triangle, UNCR9 (*recA*$^-$ *mfd*$^-$); solid triangle, UNCR9/pMFD19; and square, UNCR9/pIBI25. Data represent the average of three experiments.

The authors are in the Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, NC 27599.

A

```
   1  MPEQYRYTLP VKAGEQRLLG ELTGAACATL VAEIAERHAG PVVLIAPDMQ
  51  NALRLHDEIS QFTDQMVMNL ADWETLPYDS FSPHQDIISS RLSTLYQLPT
 101  MQRGVLIVPV NTLMQRVCPH SFLHGHALVM KKGQRLSRDA LRTQLDSAGY
 151  RHVDQVMEHG EYATRGALLD LFPMGSELPY RLDFFDDEID SLRVFDVDSQ
 201  RTLEEVEAIN LLPAHEFPTD KAAIELFRSQ WRDTFEVKRD PEHIYQQVSK
 251  GTLPAGIEYW QPLFFSEPLP PLFSYFPANT LLVNTGDLET SAERFQADTL
 301  ARFENRGVDP MRPLLPPQSL WLRVDELFSE LKNWPRVQLK TEHLPTKAAN
 351  ANLGFQKLPD LAVQRQQKAP LDALRKFLET FDGPVVFSVE SEGRREALGE
 401  LLARIKIAPQ RIMRLDEASD RGRYLMIGAA EHGFVDTVRN LALICESDLL
 451  GERVARRRQD SRRTINPDTL IRNLAELHIG QPVVHLEHGV GRYAGMTTLE
 501  AGGITGEYLM LTYANDAKLY VPVSSLHLIS RYAGGAEENA PLHKLGGDAW
 551  SRARQKAAEK VRDVAAELLD IYAQRAAKEG FAFKHDREQY QLFCDSFPFE
 601  TTPDQAQAIN AVLSDMCQPL AMDRLVCGDV GFGKTEVAMR AAFLAVDNHK
 651  QVAVLVPTTL LAQQHYDNFR DRFANWPVRI EMISRFRSAK EQTQILAEVA
 701  EGKIDILIGT HKLLQSDVKF KDLGLLIVDE EHRFGVRHKE RIKAMRANVD
 751  ILTLTATPIP RTLNMAMSGM RDLSIIATPP ARRLAVKTFV REYDSMVVRE
 801  AILREILRGG QVYYLYNDVE NIQKAAERLA ELVPEARIAI GHGQMREREL
 851  ERVMNDFHHQ RFNVLVCTTI IETGIDIPTA NTIIIERADH FGLAQLHQLR
 901  GRVGRSHHQA YAWLLTPHPK AMTTDAQKRL EAIASLEDLG AGFALATHDL
 951  EIRGAGELLG EEQSGSMETI GFSLYMELLE NAVDALKAGR EPSLEDLTSQ
1001  QTEVELRMPS LLPDDFIPDV NTRLSFYKRI ASAKTENELE EIKVELIDRF
1051  GLLPDPARTL LDIARLRQQA QKLGIRKLEG NEKGGVIEFA EKNHVNPAWL
1101  IGLLQKQPQH YRLDGPTRLK FIQDLSERKT RIEWVRQFMR ELEENAIA*
```

B

```
                                         Motif I                   Motif IA
                                       ***********                 ********
G 266  PFKPTGAQARVVAEIERDMALDVPMMRLVQGDVGSGKTLVAALAALRAIAHGKQVALMAPTELLAEQHAN
       || .|. ||   .   .   ||     ..| ||| |||| ||| ||   ||   |. . ||||.. || |||.|| .
T 598  PFETTPDQAQAINAVLSDMCQPLAMDRLVCGDVGFGKTEVAMRAAFLAVDNHKQVAVLVPTTLLAQQHYD
                                       ++ A oGoGKT                 +++LAPTR

                                                                      Motif II
                                                                      *******
G 336  NFRNWFAPLGIEVGWLAGKQKGKARLAQQEAIASGQVQMIVGTHAIFQEQVQFNGLALVIIDEQHRFGVH
       |||. ||  .. .  ..    .|            .| | .....||| . | .| |  | |.|.||.|||||
T 668  NFRDRFANWPVRIEMISRFRSAKEQTQILAEVAEGKIDILIGTHKLLQSDVKFKDLGLLIVDEEHRFGVR
                                                                      +++DEAH

                            Motif III
                           **********
G 406  QRLALWEKGQQQGFHPHQLIMTATPIPRTLAMTAYADLDTSVIDELPPGRTPVTTVAIPDTRRTDIIDRV
       ..     ||        . . | .||||||||| |.  .  | |.|    |. | .| |  . .     . . .
T 738  HK----ERIKAMRANVDILTLTATPIPRTLNMAMSGMRDLSIIATPPARRLAVKTF-VREYDSMVVREAI
                           ++++SATPPG

                     Motif IV
                     ********
G 476  HHACITEGRQAYWVCTLIEESELLEAQAAEATWEELKLALPELNVGLVHGRMKPAEKQAVMASFKQGELH
           |  | | |..   .|      . | |       ||    .||  ... || |.   | . ||  | .   .
T 803  LRE-ILRGGQVYYLYNDVE-----NIQKAAERLAEL---VPEARIAIGHGQMRERELERVMNDFHHQRFN
                     ++F+oo+o

                Motif V                        Motif VI
         ****************                      ********
G 546  LLVATTVIEVGVDVPNASLMIIENPERLGLAQLHQLRGRVGRGAVASHCVLLYKTP--LSKTAQIRLQV-
       .|| ||.|| |.|.| |   .||| ..    ||||||||||||||.    .    ||    |   ..  || ||.
T 864  VLVCTTIIETGIDIPTANTIIIERADHFGLAQLHQLRGRVGRSHHQAYAWLLTPHPKAMTTDAQKRLEAI
         + To++ooG+o+ o+o                      QR+GR+GR

G 613  --LRDSNDGFVIAQKDLEIRGPGELLGTRQTGNAE  645 G
         | |   || .|  ||||||.|||||   |.|   |
T 934  ASLEDLGAGFALATHDLEIRGAGELLGEEQSGSME  968 T
```

C

```
M 123  SINEEVERLRHSATNALLTRRDVIVVATVSCIYGLGTPEEYIE-QMVTLRRGAEMDRDVLLRRFVQMQY
S 125  SVNDEIDKLRHSATSALLERNDVIVVASVSCIYGLGSPKEYADS-VVSLRPGLEISRDKLLNDLVDIQF
E 114  SVNEHIEQMRLSATKAMLERRDVVVVASVSAIYGLGDPDLYLK-MMLHLTVGMIIDQRAILRRLAELQY

T  82  SPHQDIISSRLSTLYQLPTMQRGVLIVPVNTLMQRVCPHSFLHGHALVMKKGQRLSRDALRTQLDSAGY
  CON  S  o i   R S    l o oo +++  Vo ++    P  +     + +o G  +ord l ool o  y

VRNDVDFHRGTFRVRGDTVEIIPMYE-ELAVRIEFFGDEIESIQTLHPLTGQVVREEEEMYIFPASHYVA  259 M
ERNDIDFQRGRFRVRGDVVEIFPASRDEHAFRVEFFGDEIDRIREVEALTGQVLGEVDHLAIFPATHFVT  262 S
ARNDQAFQRGTFRVRGEVIDIFPAESDDIALRVELFDEEVERLSLFDPLTGQIVSTIPRFTIYPKTHYVT  251 E

RHVDQVMEHGEYATRGALLDLFPMGS-ELPYRLDFFDDEIDSLRVFDVDSQRTLEEVEAINLLPAHEFPT  219 T
   D  +o Go+  RG  +o+fP  o e  +R+ofF dEioo+o +   o o + e    + ++Pa  + t  CON
```

Fig. 2. (**A**) Sequence of the Mfd protein. The leucines of the leucine zipper are in boldface and underlined. (**B**) Sequence similarity between TRCF(T) and RecG(G) and the seven so-called "helicase motifs" (*17*). Conserved hydrophobic (+), charged or polar (O), and specific residues are indicated. Residues that are identical in the two sequences are indicated with vertical lines and dots indicate similar residues. (**C**) Sequence similarity between the TRCF(T) and UvrBs from *E. coli* (*E*), *S. pneumoniae* (*S*), and *M. luteus* (*M*). The consensus (CON) sequence shows where all three (capital letters) or two (lowercase) of the three UvrBs share identical amino acids with TRCF, and shows where all four proteins share hydrophobic (+), or charged or polar (O) residues. In bold are TRCF and UvrB residues that are identical or similar. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

terial UvrB proteins (*E. coli*, *Micrococcus luteus*, and *Streptococcus pneumoniae*) for which sequence information is available (*20*). The UvrB proteins also have the helicase motifs; however, the region of Mfd-UvrB similarity (Fig. 2C) does not include any helicase motif from either protein (*17, 20*). Finally, a leucine zipper motif (*20, 21*) (Fig. 2A) appears in the COOH-terminal region of Mfd. This region may participate in the interaction of TRCF with RNAP (see below).

Purification and properties of Mfd protein (TRCF). Cells carrying the *mfd* plasmid, pMFD19, produced large amounts of TRCF protein (Fig. 3). We followed our earlier purification procedure (*8*) and obtained 4 mg of protein (>95 percent pure) from a 3-liter culture (Fig. 3) and characterized some physical and biochemical properties of the purified protein (Table 1). Our data indicate that TRCF is an adenosine triphosphatase (ATPase) and a DNA binding protein whose modest binding was stimulated by ATPγS; the ATPase activity was modest and was not affected by DNA or by a stalled RNAP complex.

Functionally, a relevant finding is the lack of helicase activity particularly in view of the good match of the protein to the so-called helicase motifs. The TRCF failed to displace either a 71-nt DNA or a 48-nt RNA annealed to single-stranded phage DNA. Since the transcriptional termination factor Rho requires a 5′ single-stranded RNA tail to perform its RNA-DNA helicase activity (*22*), we synthesized an RNA-DNA hybrid with a 222-nt single-stranded RNA tail (*23*). However, even though the Mfd protein has a Rho-like activity (see below), it did not dissociate the partially hybridized RNA. We conclude that TRCF is not an RNA-DNA helicase or a DNA-DNA helicase in the generally accepted sense of the term (*24*).

Our earlier in vitro observations that RNAP stalled at a lesion inhibits repair indicated that the TRCF must (i) overcome the inhibition by RNAP and (ii) target the excision repair enzyme to the damage site (*5*). We therefore investigated the interaction of TRCF with these two enzyme systems.

Interaction of TRCF with RNA polymerase and (A)BC excinuclease. To study the effect of TRCF on transcription, we synthesized two oligonucleotide template-substrates (Fig. 4A), bearing promoters for *E. coli* (*25*) and T7 (*26*) RNAPs and a psoralen-thymine monoadduct located downstream in either the template or coding strand. With these substrates we con-

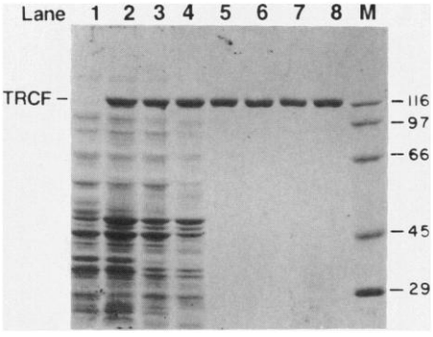

Fig. 3. Purification of Mfd (TRCF). The protein was purified from *E. coli* DH5αF′ cells and pMFD19 plasmids, and purification was followed by SDS–polyacrylamide gel electrophoresis (PAGE) and Coomassie blue staining. (Lane 1) Host cells (DH5αF′, 100 μl); (lane 2) transformed cells (DH5αF′ pMFD19, 100 μl); (lane 3) cell-free extract (1.2 μl); (lane 4) DEAE Bio-Gel peak fractions; (lane 5) blue Sepharose chromatography peak fractions; (lane 6) AcA34 peak fractions; (lane 7) first heparin agarose chromatography; and (lane 8) second heparin agarose chromatography peak fractions. Lanes 4 to 8 each contained about 4 μg of protein; M, molecular weight markers.

Table 1. Properties of *E. coli* TRCF.

Gene	*mfd*
Size (amino acids)*	1148
Mr*	130,067
pI*	6.8
Quaternary structure†	Monomer
Nucleotide binding	Yes
ATPase (k_{cat})‡	
– DNA	2.7 min^{-1}
– ssDNA	3.3 min^{-1}
– dsDNA	2.3 min^{-1}
DNA binding§	Nonspecific
Helicase‖	
DNA-DNA	No
RNA-DNA	No
DNA (5′ tail)-RNA	No

*Based upon the sequence, values include the NH_2-terminal Met residue which is absent in the protein. †Based upon gel filtration analysis. ‡Where indicated, reactions included 0.9 mM TRCF, 200 ng of single-stranded M13 DNA (+ssDNA), or 200 ng of pBR322 (+dsDNA). Values are the averages of three experiments. §Binding to duplex (b) in Fig. 4A was observed in experiments that employed concentrations of TRCF over 90 nM and was stimulated by ATPγS. ‖The substrates were a 71-nt DNA strand annealed to M13 (DNA-DNA), a 48-nt RNA annealed to pGEM-3Zf(+) (RNA-DNA), or a 271-nt RNA, with 49 bases at the 3′ end annealed to pGEMt '5 (*22*) [DNA (5′-tail)-RNA].

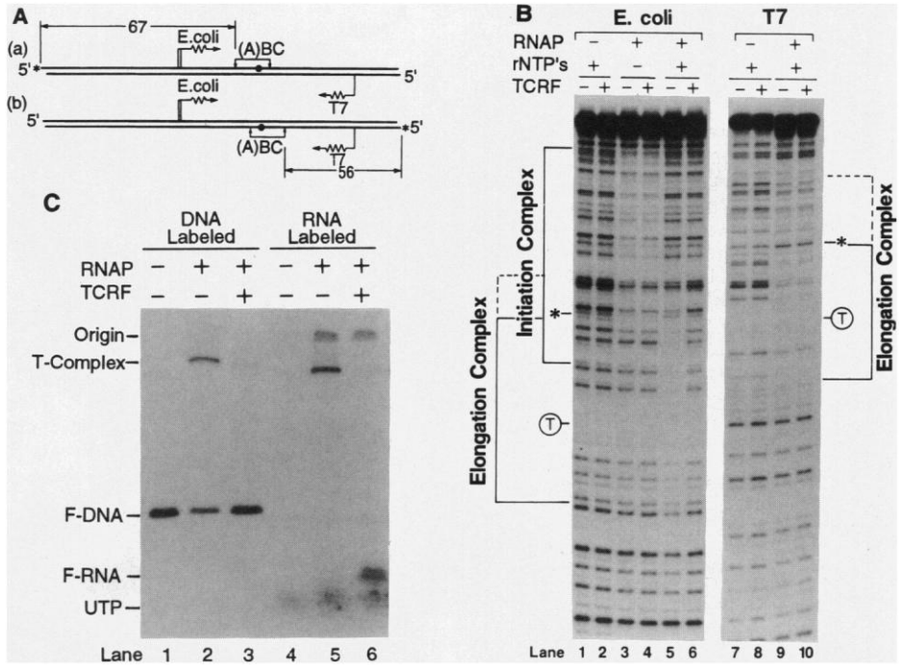

Fig. 4. Release of *E. coli* RNAP and the truncated RNA by TRCF. (**A**) Template-substrates. Closed circles indicate psoralen-modified thymines and asterisks indicate ^{32}P-labeled sites. (A)BC excinuclease incisions are shown with arrows. Incision generates 67-nt and 56-nt ^{32}P-labeled fragments from the two substrates as shown. The transcription initiation sites and direction of transcription for *E. coli* and T7 RNAPs are marked on the coding strands. The adduct in duplex (b) is in the transcribed strand for *E. coli* RNAP and in the nontranscribed strand for T7 RNAP. The reverse is the case in the (a) duplex (*25, 26*). (**B**) Effect of TRCF on footprints of initiation and stalled elongation complexes of *E. coli* and T7 RNAPs. Reactions (50 μl) contained 40 mM Hepes (pH 7.8), 100 mM KCl, 8 mM $MgCl_2$, 2 mM ATP, 5 mM dithiothreitol (DTT), bovine serum albumin at 100 μg/ml and about 1 nM DNA. Where indicated rNTPs were present at 200 μM each, TRCF at 87 nM, and *E. coli* and T7 RNAPs at 0.012, and 5 units per reaction, respectively. After the reactions were incubated at 37°C for 18 minutes (*E. coli* RNAP) or 25 minutes (T7 RNAP), deoxyribonuclease I footprinting was performed (*5*). Duplex (a) was used with T7 RNAP and duplex (b) with *E. coli* RNAP. Asterisks indicate hypersensitive sites in elongation complexes. (**C**) Release of truncated RNA by TRCF. Elongation complexes formed with *E. coli* RNAP and duplex (b) without (lanes 2 and 5) and with TRCF (lanes 3 and 6) were analyzed by gel retardation. In lanes 1 to 3 ^{32}P-labeled duplex (b) and nonradioactive rNTPs were present in the reaction; in lanes 4 to 6 duplex (b) was nonradioactive, and the transcript was labeled with [^{32}P]UTP (3000 Ci/mmol). Reactions were as in (B) except that (i) incubation with only UTP (the first nucleotide in the transcript) at 200 μM (lanes 1 to 3) or [α–^{32}P]UTP at 66 nM (lanes 4 to 6) was performed for 12 minutes at 37°C, and (ii) CTP, GTP, and UTP were added to 200 μM each along with rifampicin at 22 μg/ml, and pBR322 to 2 μg/ml to ensure one round of transcription initiation, and where indicated TRCF to 87 nM. After incubation for an additional 12 minutes the complexes were resolved by electrophoresis at 4°C in a 4 percent polyacrylamide gel and located by autoradiography after washing the gel. F, free DNA or RNA; T, ternary complex; and UTP, unincorporated [^{32}P]UTP.

firmed that a psoralen monoadduct in the template strand blocked *E. coli* RNAP, whereas a lesion in the coding strand did not (*25*). We examined the effect of the TRCF on initiation and elongation complexes by deoxyribonuclease I footprinting (Fig. 4B). The TRCF has no effect on the initiation complex of *E. coli* RNAP (compare lanes 1 and 2 to 3 and 4) but released RNAP from the elongation complex stalled at the adduct site (lanes 5 and 6). This Rho-like effect of TRCF (which we found required ATP hydrolysis and was inhibited by ATPγS) was specific for *E. coli* RNAP because the elongation complex of T7 RNAP was not affected by the TRCF (Fig. 4B, lanes 9 and 10). The lesion was not required for dissociation since removal of a stalled *E. coli* RNAP by TRCF was also observed when transcription was blocked by a protein bound tightly to the template (*27*).

The disposition of the truncated transcript was examined next. Using duplex (b) in Fig. 4A we carried out transcription in the absence and presence of TRCF and analyzed the macromolecular complexes by gel retardation. A ternary complex did form at the lesion site and the TRCF released both RNAP and the transcript (Fig. 4C). Analysis of the labeled RNA on sequencing gels showed that a truncated transcript of 29 nt was synthesized in the absence (*15*) and presence of TRCF, an indication that TRCF does not enable RNAP to transcribe through the lesion or back off and cleave the truncated transcript (*28*).

Having shown the specific interaction of TRCF with a stalled RNAP, we next studied the targeting of the excision repair enzyme, (A)BC excinuclease, to the damage site. (A)BC excinuclease is the enzymatic activity resulting from partially over-

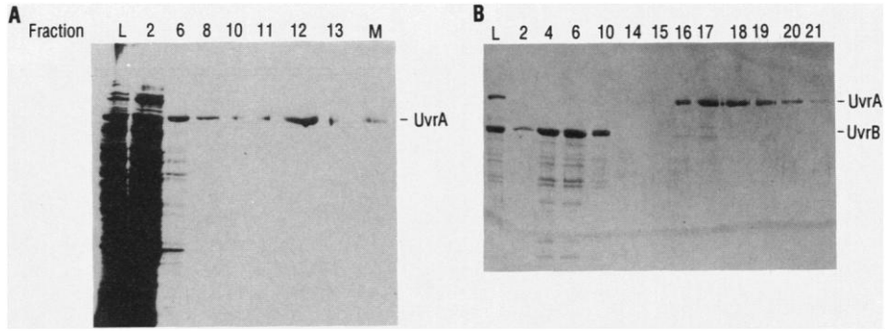

Fig. 5. UvrA binds to a TRCF affinity column. The column was prepared with 270 μl of tresyl-activated agarose and 270 μg of TRCF (*31*). (**A**) UvrA (10 μg) and *E. coli* extract (2.9 mg) in 1 ml of 50 mM tris-HCl (pH 7.6), 50 mM KCl, 1 mM ATP, and 5 percent glycerol was loaded (L) onto the column and, after the column was washed with the same buffer, bound protein was eluted with 900 mM KCl in the same buffer. Fractions of 630 μl were collected, and 100 μl of each was analyzed by SDS-PAGE with subsequent silver staining. Lane M is the UvrA marker. (**B**) UvrA (10 μg) plus UvrB (50 μg) mixture (L) in 1 ml of loading buffer [25 mM tris-HCl (pH 7.6), 25 mM KCl, 1 mM $MgCl_2$, 1 mM ATP, and 5 percent glycerol] was applied to the column and the column was washed with the same buffer collecting 175-μl fractions. The bound protein was eluted with 900 mM KCl (fractions 16 to 21). TRCF did not bind to a control column lacking UvrA.

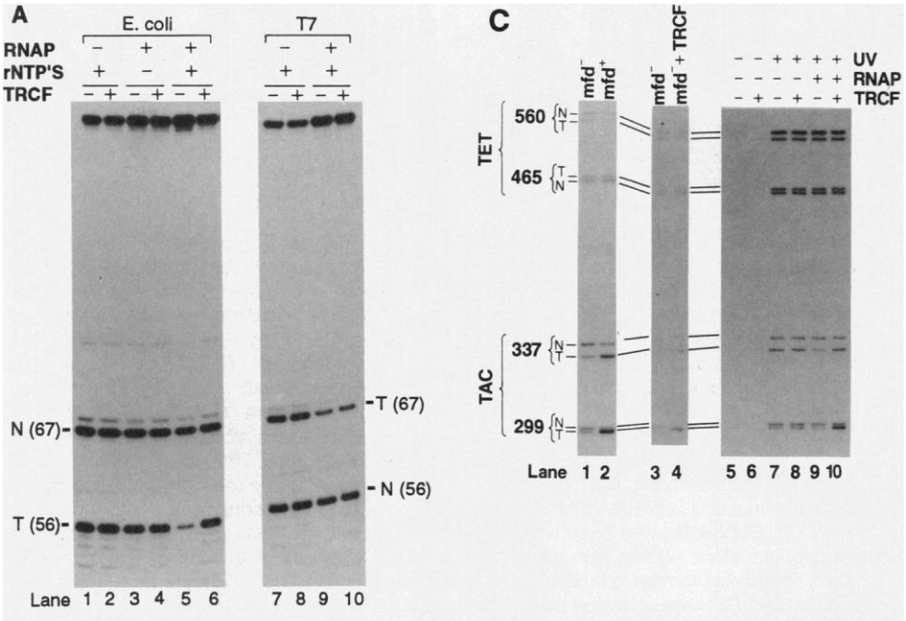

Fig. 6. TRCF overcomes repair inhibition caused by stalled RNAP and specifically stimulates repair of the transcribed strand. (**A**) Suppression of inhibition. A mixture of 137-bp duplexes (a) and (b) were incubated with *E. coli* or T7 RNAP as in Fig. 4B and then UvrA (4 nM), UvrB (120 nM), and UvrC (50 nM) were added and incubation continued for 15 minutes. The reaction products were analyzed on 8 percent polyacrylamide sequencing gels. T and N are the fragments resulting from incisions by (A)BC excinuclease 5′ to the adduct on the transcribed and nontranscribed strands, respectively (Fig. 4A). (**B**) Stimulation of incision of the transcribed strand by RNAP plus TRCF. The template strand of ultraviolet-irradiated pDR3274 (*6*) was labeled with ^{32}P at the 3′ terminus of the Eco RI site 200 bp from the promoter, the DNA was incubated with (A)BC excinuclease at 37°C for 22 minutes under conditions identical to panel A, and the incision products were analyzed on a 3.6 percent polyacrylamide sequencing gel. The *tac* promoter elements are indicated by (−10) and (−35); the arrow on the left indicates the transcriptional start site and direction, and the arrow on the right shows the incision site closest to the promoter which is stimulated by TRCF. (**C**) Strand specificity of transcription-repair coupling. Nonlabeled pDR3274 was irradiated with 225 Jm^{-2} of 254 nm (ultraviolet) (lanes 1 to 4 and 7 to 10) and subjected to transcription-repair reactions in the presence of [α-^{32}P]dCTP to measure repair synthesis (*6*). After incubation with cell-free extracts (lanes 1 to 4) or a defined transcription-repair system (lanes 5 to 10), the DNA was digested with restriction enzymes and resolved on a 3.6 percent sequencing gel to show levels of repair in the transcribed (T) and nontranscribed (N) strands of promoter-proximal fragments from a region transcribed from the strong tac promoter (TAC) and fragments that were essentially nontranscribed (TET). The cell-free extracts were from UNC3610-45 (*mfd*⁻) or UNC3610 (*mfd*⁺), and 87 nM TRCF was included when indicated. The defined system (lanes 5 to 10) contained *E. coli* RNAP (1.2 U/ml), UvrA (4 nM), UvrB (100 nM), UvrC (70 nM), helicase II (5 nM), Pol I (80 U/ml), T4 DNA ligase (48 U/ml), and 350 nM TRCF.

lapping actions of UvrA, UvrB, and UvrC proteins (*29*); UvrA and UvrB form an A_2B_1 complex, which, guided by UvrA's affinity to damaged DNA, can then bind to a damage site; in an ATP hydrolysis–dependent reaction, UvrB is delivered to DNA, with formation of a severe kink in the DNA and dissociation of UvrA from the complex. Binding of UvrC to the UvrB-DNA complex brings about the hydrolysis of the 8th phosphodiester bond 5′ to the lesion and either the 4th or 5th phosphodiester bond 3′ to the lesion. In the overall process, damage recognition and loading of UvrB onto DNA is the rate-limiting step (*30*). Therefore, a plausible role for TRCF would be to target the A_2B_1 complex to the lesion site and facilitate the delivery (loading) of UvrB.

In the absence of transcription, TRCF had no effect on (A)BC excinuclease. Similarly, no TRCF-UvrA interaction was detected by standard gel permeation chromatography. Therefore, we used protein affinity chromatography, which can detect weak, but specific, protein-protein interactions (*31*). An *E. coli* cell-free extract supplemented with UvrA was passed through a column consisting of TRCF coupled to an agarose resin, and the bound proteins were eluted with high salt. UvrA did bind to the TRCF matrix. RNAP, which constitutes about 1 percent of the total protein in the extract, did not bind to the column (Fig. 5A) even though TRCF does interact with a stalled RNAP complex. Since most of UvrA is in the A_2B_1 complex under physiological conditions (*30*) it was of interest to determine whether TRCF bound to this complex as well. When the mixture of UvrA and UvrB, in which all UvrA would be expected to be in A_2B_1 complex was applied to the column, UvrA was retained quantitatively, but UvrB did not bind (Fig. 5B). Thus, TRCF did not directly bind to UvrB, and it appears that binding of TRCF to the A_2B_1 complex promoted dissociation of UvrA from UvrB. These experiments provide an important link to the mechanism of tran-

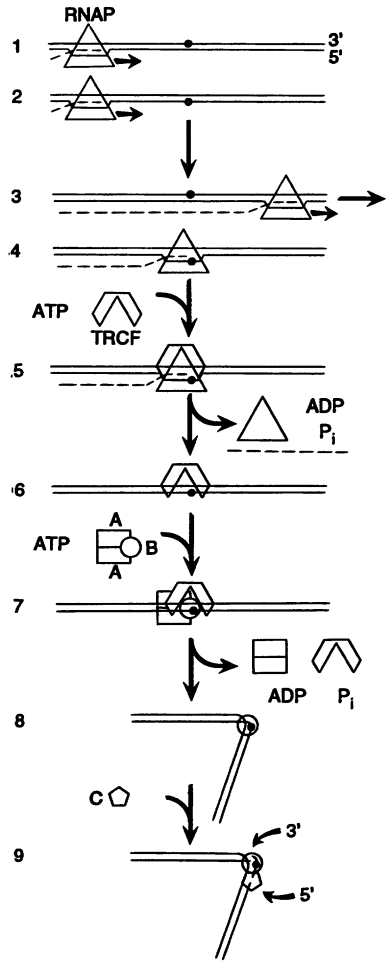

Fig. 7. Model for strand-specific repair in *E. coli*. Closed circle, lesion in DNA; dotted line, RNA; A, UvrA; B, UvrB; and C, UvrC. The TRCF recognizes a stalled RNAP and UvrA in the A_2B_1 complex and delivers UvrB to the lesion in the transcribed strand from the A_2B_1 complex as it releases the stalled RNAP in a sequential (as shown here) or a highly concerted reaction (4 → 8).

scription-repair coupling by revealing that TRCF recruits the damage recognition subunit to a stalled RNAP and thus is capable of increasing the rate of damage recognition at such sites. The results also suggest that the UvrB-TRCF homology region is the UvrA interaction site of UvrB and TRCF. It is possible that they both bind to the same site of UvrA.

Transcription-repair coupling. To find out whether TRCF was necessary and sufficient for transcription-repair coupling we used the duplexes (a) and (b) (Fig. 4A), which contained a psoralen-thymine adduct either in the template or coding strand. Transcription-repair reactions were conducted under conditions similar to those shown in Fig. 4B, except that UvrA, UvrB, and UvrC were also added to the reaction mixtures. Both *E. coli* and T7 RNA polymerases inhibited repair when stalled at a lesion (Fig. 6A). Inclusion of the TRCF had no effect on the inhibition by T7 RNAP, as expected, since it did not displace the polymerase from the elongation complex. In the homologous system, the TRCF did improve the repair rate in the template strand during transcription; however, it did not stimulate repair compared to that obtained in the absence of transcription (lanes 4 and 6 in Fig. 6A). Thus, in this system the TRCF performed only one of the two functions attributed to such a factor; it removed the inhibitory effect by releasing *E. coli* RNAP, but it did not target (A)BC excinuclease to the lesion.

From the above results, we could infer that, even though the TRCF interacted with both components of transcription and repair reactions, another factor may be required for the preferential repair of the transcribed strand. It is also possible that the TRCF had special substrate requirements for exerting its repair stimulatory effect, requirements that were not provided by the 137-bp duplex used in our experiment. We reasoned that a longer substrate might be needed to anchor TRCF onto DNA as it carried out its dual functions, and therefore we labeled a 5.3-kbp fragment from pDR3274 at the 3′ end of the transcribed strand 200 bp upstream from a *tac* promoter (6). The DNA was irradiated with ultraviolet, and a transcription-repair reaction was performed (Fig. 6B). Without transcription, the DNA was nicked at many specific sites as a result of excision by (A)BC excinuclease of photoproducts produced at dipyrimidine hot spots (lane 1). When RNAP was present in the repair reaction, excision was inhibited (lane 2). When the TRCF was also added, the inhibition disappeared and the rate of incision throughout the transcription unit, except the promoter region (lane 3), was actually stimulated. These results, which are consistent with the specific effect of TRCF on elongation but not on initiation complexes (Fig. 4B), demonstrate that TRCF alone is capable of stimulating the rate of incision of the transcribed strand.

To investigate the strand specificity of repair inhibition and stimulation in this system, we conducted repair synthesis assays with closed circular pDR3274 and measured incorporation of label into repair patches in the two strands downstream from the *tac* promoter (6, 8). The preferential repair signal in this assay (Fig. 6C) can be seen by comparing repair synthesis from extracts of mfd^- (lane 1) and mfd^+ (lane 2) cells, and by comparing repair synthesis from mfd^- extract minus (lane 3) and plus (lane 4) the TRCF. Using a defined system consisting of purified transcription and repair proteins in this assay, we found that transcription preferentially inhibited repair of the transcribed strand (lanes 7 and 9) and that added TRCF overcame the inhibitory effect of transcription (lane 9) and stimulated the repair of the transcribed strand above the basal level (lanes 7 and 8) without affecting the repair of the nontranscribed strand (lane 10). Thus, we conclude that the TRCF is necessary and sufficient to specifically increase the rate of repair of the transcribed strand.

Model for transcription-repair coupling. A plausible model for strand-specific repair in *E. coli* is presented in Fig. 7. A lesion in the template but not the coding strand blocks RNAP (**1** to **4**). The RNAP-RNA-DNA ternary complex (**4**) is specifically recognized by the TRCF (**5**), which releases RNAP and the truncated transcript; the TRCF may replace RNAP at the lesion site (**6**) perhaps by simultaneously binding to an anchor site at some distance from the lesion. The DNA-bound TRCF recruits the A_2B_1 complex by its affinity for UvrA (**7**) and aids in the formation of the UvrB-DNA complex by facilitating the dissociation of UvrA, perhaps by binding to UvrA at its UvrB interface. The TRCF dissociates from DNA simultaneously with UvrA leaving behind the preincision UvrB-DNA complex (**8**). UvrC binds to this complex, and the dual incisions are made (**9**). Reactions **4** to **8** might be highly coupled, precluding detection of some of the proposed intermediates by footprinting techniques.

The human genetic disease Cockayne's syndrome appears to be caused by lack of a strand-specific repair mechanism (*32*, *33*). The gene ERCC-6 that specifically corrects this defect has recently been cloned and sequenced (*33*). On the basis of phenotypic similarities between Mfd^- *E. coli* and Cockayne's syndrome cell lines, we predict that the ERCC-6 protein, like Mfd, will specifically interact with its cognate RNAP, RNAP II (*34*), and the damage recognition subunit (or subunits) of human excision nuclease. Both ERCC-6 and Mfd have the "helicase motifs"; however, we have found no striking homology between the two proteins.

REFERENCES AND NOTES

1. V. A. Bohr, C. A. Smith, D. S. Okumoto, P. C. Hanawalt, *Cell* **40**, 359 (1985).
2. I. Mellon and P. C. Hanawalt, *Nature* **342**, 95 (1989).
3. I. Mellon, G. Spivak, P. C. Hanawalt, *Cell* **51**, 241 (1987).
4. P. C. Hanawalt and I. Mellon, *Curr. Biol.* **3**, 67 (1993).
5. C. P. Selby and A. Sancar, *J. Biol. Chem.* **267**, 17688 (1990).
6. _____, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8232 (1991).
7. R. Bockrath and J. E. Palmer, *Mol. Gen. Genet.* **156**, 133 (1977); R. Bockrath, A. Barlow, J. Engstrom, *Mutat. Res.* **183**, 241 (1987).
8. C. P. Selby, E. M. Witkin, A. Sancar, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11574 (1991).

9. S. Kunala and D. E. Brash, *ibid.* **89**, 11031 (1992).
10. A. R. Oller, I. J. Fijalkowska, R. L. Dunn, R. M. Schaaper, *ibid.*, p. 11036.
11. E. M. Witkin, *Science* **152**, 1345 (1966).
12. _____, personal communication (1992).
13. _____, *Annu. Rev. Microbiol.* **23**, 487 (1969).
14. Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
15. The oligonucleotide sequence TAIACIITICCIG-TIAAIGCIGGIGAICA was based on amino acids 6 to 15 of TRCF.
16. A. T. Bankier, K. M. Weston, B. G. Barrell, *Methods Enzymol.* **155**, 51 (1987).
17. A. E. Gorbalenya, E. V. Koonin, A. P. Donchenko, V. M. Blinov, *Nucleic Acids Res.* **17**, 4713 (1989).
18. R. G. Lloyd, *J. Bacteriol.* **173**, 5414 (1991).
19. _____ and G. J. Sharples, *EMBO J.* **12**, 17 (1993); see also R. Holliday, *Genet. Res.* **5**, 282 (1964).
20. N. Sicard, J. Orglia, E. Estevenon, *J. Bacteriol.* **174**, 2412 (1992).
21. W. H. Landshulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
22. C. A. Brennan, A. J. Dombroski, T. Platt, *Cell* **48**, 945 (1987).
23. Substrate B in figure 1 of (*22*); pGEMt '5 was a gift from T. Platt.
24. S. N. Matson and K. A. Kaiser-Rogers, *Annu. Rev. Biochem.* **59**, 289 (1990).
25. Y.-b. Shi, H. Gamper, J. E. Hearst, *Nucleic Acids Res.* **15**, 6843 (1987).
26. _____, *J. Biol. Chem.* **263**, 527 (1988).
27. P. A. Pavco and D. A. Steege, *ibid.* **265**, 9960 (1990).
28. D. Reines, *ibid.* **267**, 3795 (1992).
29. A. Sancar and G. B. Sancar, *Annu. Rev. Biochem.* **27**, 29 (1988); J. J. Lin and A. Sancar, *Mol. Microbiol.* **6**, 2219 (1992).
30. D. K. Orren, C. P. Selby, J. E. Hearst, A. Sancar, *J. Biol. Chem.* **267**, 780 (1992).
31. T. Formosa, R. L. Burke, B. M. Alberts, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2442 (1983).
32. J. Venema, L. H. F. Mullenders, A. T. Natarajan, A. A. van Zeeland, L. V. Mayne, *ibid.* **87**, 4704 (1990).
33. C. Troelstra *et al.*, *Cell* **71**, 939 (1992).
34. J. D. Armstrong and B. A. Kunz, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9005 (1990); S. A. Leadon and D. A. Lawrence, *Mutat. Res.* **255**, 67 (1991); J.-M. H. Vos and E. L. Wauthier, *Mol. Cell. Biol.* **11**, 2245 (1991); L. Lommel and P. C. Hanawalt, *Mutat. Res.* **255**, 183 (1991); F. C. Christians and P. C. Hanawalt, *ibid.* **274**, 93 (1992).

4 November 1992; accepted 26 February 1993

DNA Repair Helicase: A Component of BTF2 (TFIIH) Basic Transcription Factor

Laurent Schaeffer, Richard Roy, Sandrine Humbert, Vincent Moncollin, Wim Vermeulen, Jan H. J. Hoeijmakers, Pierre Chambon, Jean-Marc Egly*

The human BTF2 basic transcription factor (also called TFIIH), which is similar to the δ factor in rat and factor b in yeast, is required for class II gene transcription. A strand displacement assay was used to show that highly purified preparation of BTF2 had an adenosine triphosphate–dependent DNA helicase activity, in addition to the previously characterized carboxyl-terminal domain kinase activity. Amino acid sequence analysis of the tryptic digest generated from the 89-kilodalton subunit of BTF2 indicated that this polypeptide corresponded to the *ERCC-3* gene product, a presumed helicase implicated in the human DNA excision repair disorders xeroderma pigmentosum and Cockayne's syndrome. These findings suggest that transcription and nucleotide excision repair may share common factors and hence may be considered to be functionally related.

Transcription initiation of protein coding genes requires, in addition to RNA polymerase II (B), two sets of transcription factors that control basal and activated transcription, respectively. The basal transcription factors along with RNA polymerase II are required for most if not all promoters (*1*). They function with the minimal promoter, which may or may not contain an initiator element (*2*) in addition to a TATA box, to form a multiprotein complex that, when nucleoside triphosphates (NTPs) are added, initiate transcription. Activated transcription is dependent on regulatory proteins that, by binding to specific cis-acting DNA elements, communicate directly or indirectly with the basal transcription machinery by way of the TATA binding protein (TBP) or through TBP-associated factors (TAFs) to activate or repress RNA synthesis (*3*).

Binding of the factor (TFIID/TBP), which recognizes the TATA box to the minimal promoter is thought to be the first step in the formation of a multiprotein complex containing at least six other general transcription factors TFIIA, TFIIB, TFIIE, TFIIF, TFIIJ/TFIIG, and BTF2 (TFIIH) (*4*). Once formed on the promoter, the conversion of this preinitiation complex to an active initiation complex capable of forming the first phosphodiester bond of the RNA transcript, requires the hydrolysis of the β-γ bond of adenosine triphosphate (ATP) (*5*). In eukaryotes, RNA polymerase II transcription factors or their associated proteins consume ATP. For example, the transcription factor b from yeast (*6*), factor δ in rat (*7*), and the human counterpart BTF2/TFIIH (*8*, *9*) have a DNA-dependent ATPase activity (*7*), whereas the same factors were also found to consume ATP to phosphorylate the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (*10–13*). It was thus postulated that such a phosphorylation of the CTD induces the transition from initiation to elongation (*14*). However, the absolute necessity of ATP hydrolysis for initiation of transcription in RNA polymerase II–dependent systems cannot be entirely accounted for by its consumption in the phosphorylation of the CTD, which can use both ATP and guanosine triphosphate (GTP) as substrates (*12*, *13*), but rather by another ATP-dependent step that must occur during or after the kinase reaction and before the formation of the first phosphodiester bond (*14*). Such an activity could reside in a factor analogous to that responsible for the melting (opening) of the duplex DNA during the formation of the open complex in prokaryotes, a step that requires ATP (*15*). In eukaryotes, an affinity column strategy based on protein-protein interactions was used to show that an ATP-dependent DNA helicase activity was associated either with the transcription factor TFIIF (also called RAP30-RAP74) or with the RNA polymerase II, or both (*16*).

We present evidence that a helicase activity is closely associated with the BTF2/TFIIH transcription factor (*8*, *9*), a multisubunit protein complex, which has a CTD protein kinase activity and includes a 62-kD polypeptide (*17*). Microsequencing of tryptic digests of the largest subunit (the 89-kD polypeptide) of BTF2 resulted in a series of oligopeptides also present in the human ERCC-3 gene product, which has been shown to participate in the nucleotide excision repair process in man, drosophila, and yeast (*18–20*) and whose sequence contains motifs similar to those of RNA and DNA helicases (*21*).

Coelution of helicase, protein kinase, and BTF2 transcription activities. The purification of the basal transcription factor BTF2 was performed as described (*8*), with the last three steps of the purification scheme, including chromatography with phenyl-5PW and hydroxyapatite (HAP) and a subsequent glycer-

L. Schaeffer, R. Roy, S. Humbert, V. Moncollin, P. Chambon, and J.-M. Egly are with the UPR 6520 (CNRS), Unité 184 (INSERM), Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France. W. Vermeulen and J. H. J. Hoeijmakers are in the Department of Cell Biology and Genetics, Medical Genetics Center, Erasmus University Rotterdam, Box 1738, 3000 DR Rotterdam, The Netherlands.

*To whom correspondence should be addressed.