Published in final edited form as:

J Mol Biol. 2011 February 18; 406(2): 275–284. doi:10.1016/j.jmb.2010.12.004.

Derepression of Bacterial Transcription-Repair Coupling Factor is Associated with a Profound Conformational Change

Devendra B. Srivastava and Seth A. Darst*

The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

Abstract

Transcription Repair Coupling Factor (TRCF, the product of the *mfd* gene) is a widely conserved bacterial protein that couples DNA repair with transcription. TRCF recognizes RNA polymerase stalled at a non-coding lesion in the DNA template strand, uses the energy from ATP hydrolysis to disrupt the transcription complex, and stimulates DNA repair by recruiting UvrA, a component of the nucleotide excision repair machinery, to the site. TRCF is a large (130 kDa), multi-functional protein with a complex structure/function relationship comprising a compact arrangement of eight structured domains linked by flexible linkers. Through a conserved, intramolecular, interdomain interaction, TRCF is held in a conformation in which its enzymatic activities (ATPase activity, DNA translocase activity) are strongly repressed. Disruption of the repressive interdomain interaction by amino acid substitutions within the interface derepresses ATPase and DNA translocase activities. In this work, we have shown that derepressed TRCF mutants are dramatically sensitized to limited proteolysis compared with repressed TRCF, pointing to an altered conformational state. Analysis of the protease cleavage sites mapped onto the structure of the repressed TRCF conformation indicate that: 1) The cleavage sites tend to cluster at linkers connecting the TRCF structured domains, and 2) Many of the cleavage sites sensitized in the derepressed TRCF are partially or completely buried to protease access in the repressed TRCF structure. We conclude that TRCF derepression is associated with profound conformational changes that primarily involve a reorganization of the interdomain interactions.

Keywords

Limited proteolysis; Mfd; Protein conformational change; Transcription-repair coupling factor

Introduction

Transcription Repair Coupling Factor (TRCF, the product of the *mfd* gene) is a widely conserved bacterial protein that couples DNA repair with transcription^{1–3}. TRCF recognizes RNA polymerase (RNAP) stalled at a non-coding lesion in the DNA template strand and uses the energy from ATP hydrolysis to disrupt the transcription complex. The resulting release of the transcript and the RNAP relieves inhibition of repair due to RNAP occluding the damage site. Moreover, TRCF stimulates repair by recruiting the DNA nucleotide excision repair (NER) machinery to the site.

*Correspondence should be addressed to S.A.D. (darst@rockefeller.edu). Dr. Darst: tel. (212)-327-7479; FAX (212)-327-7477; darst@rockefeller.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

TRCF is a large (130 kDa), multi-functional protein with a complex structure/function relationship^{1; 2}. The 3.2 Å-resolution X-ray crystal structure of *Escherichia coli* (*Eco*) TRCF comprises a compact arrangement of eight structured domains (D1a, D1b, and D2–D7) linked by flexible linkers⁴; Fig. 1). At the N-terminus, domains D1a, D2, and D1b form a structural module with similarity to the namesake domains of the NER component UvrB⁴. This ‘UvrB homology module’ is not required for the RNAP release function of TRCF, but is required for stimulating DNA repair, which it does by directly recruiting the NER machinery through the binding of the NER component UvrA^{3; 5–7}. Following the UvrB homology module is D3, a species-specific domain with unknown function.

A long (more than 40 Å), unstructured linker connects D3 with D4, an RNAP interacting domain (RID) that mediates protein/protein interactions between TRCF and RNAP that are critical for the RNAP release function^{4; 8; 9}. D5/D6 comprise the translocation module, two domains (also called Translocation Domains 1 and 2, or TD1 and TD2) strikingly similar in sequence and structure to the corresponding domains of the double-strand DNA (dsDNA) translocase RecG^{3; 4; 10}. Both the TRCF and RecG translocation modules contain the seven signature sequence motifs of superfamily 2 (SF2) helicases/ATPases¹¹, as well as an additional motif unique to TRCF and RecG termed the TRG (for Translocation in RecG) motif that serves to couple nucleotide hydrolysis with dsDNA translocation^{12; 13}. The translocation module is linked to the C-terminal domain, D7, by a 25 Å extended linker. Interestingly, D7 participates in a conserved, interdomain interaction with D2 in the UvrB homology module that buries the UvrA binding determinant of TRCF (Fig. 1B)^{4; 5}. This structural architecture suggests that TRCF undergoes large-scale conformational changes during the course of its functional cycle⁴.

Structural and biochemical analyses suggest that the static crystal structure represents an inactive, repressed state of TRCF, and that major conformational changes within TRCF are required to perform its functions. While TRCF harbors the SF2 ATPase motifs and has the ability to forward translocate RNAP elongation complexes by translocating on dsDNA¹⁴, TRCF on its own, in the absence of RNAP, shows only very weak ATPase activity (rate of ATP hydrolysis < 10/min), the ATPase activity is not stimulated by DNA, and DNA translocase activity has not been observed¹⁵. In our structural analysis of TRCF, we noted that the UvrA binding surface of TRCF-D2 was occluded through a significant interface (756 Å² buried surface) with D7 (Fig. 1B)⁴. Although the overall sequence of D7 is poorly conserved, D7 appears to be present in all bacterial TRCFs. Interestingly, the residues of D7 that interact with D2 form a small, conserved cluster (corresponding to *Eco* TRCF E1045/D1048/R1049/F1050/G1051), and this comprises the only conserved cluster of residues in D7, highlighting the importance of this interaction⁴. These conserved D7 residues interact with conserved residues in D2 (Fig. 1B). We proposed that this D2/D7 interdomain interaction holds TRCF in the repressed conformation observed in the crystal structure, and that for functions of TRCF to express themselves (such as ATPase activity, DNA translocase activity, and UvrA binding), this autoinhibitory interaction must be disrupted⁴. Following from these structural observations, expression of TRCF functions can be achieved in the absence of RNAP by the disruption of this interaction, either through mutagenesis, or through truncation of the TRCF domains involved^{7; 15; 16}.

Overall, four different types of mutant TRCF proteins in which the D2/D7 interface was disrupted have been investigated:

1. N-terminal deletions, in which the UvrB homology module alone ($\Delta[1-353]$)⁷, or the UvrB homology module and D3 ($\Delta[1-473]$)¹⁶, were deleted.
2. A D2 triple mutant substituting residues involved in the D7 interface (R165A/R181A/F185A, called TRCF-D2***, Fig. 1)⁷.

3. C-terminal deletions, in which D7 was deleted (Δ [998–1148])¹⁵.
4. A D7 triple mutant substituting residues involved in the D2 interface (E1045A/D1048A/R1049A, called TRCF-D7***; Fig. 1)⁷.

Each of these mutant TRCFs, in the absence of any other proteins, hydrolysed ATP at a rate increased by an order of magnitude or more than full-length TRCF^{7; 15}. In an assay for dsDNA translocation measuring displacement of a triplex-forming oligonucleotide from the dsDNA (TFO displacement assay), wt-TRCF is essentially inactive over a period of 1 hour, while four of the mutant enzymes (Δ [1–353]TRCF, Δ [1–473]TRCF, Δ [998–1148]TRCF, TRCF-D7***) displace 95% of the triplex-forming oligo in less than 10 minutes^{7; 15; 16}.

These findings indicate that the conserved D2/D7 interaction not only blocks UvrA binding by TRCF⁴, but ‘locks down’, or stabilizes TRCF in a conformation (i.e., the conformation observed in our crystal structure; Fig. 1B) in which ATP hydrolysis and DNA translocation are strongly repressed. For wt-TRCF during normal transcription-coupled repair, we hypothesize that the interaction of TRCF with RNAP elicits a conformational change in TRCF that derepresses ATP hydrolysis and DNA translocation, which are required for RNAP release. In the absence of RNAP, these activities can be derepressed by disruption of the D2/D7 interface through truncations of either interacting domain (D2 or D7) or through amino acid substitutions of conserved residues involved in the interaction^{7; 15; 16}.

Structural analyses indicate that expression of TRCF functions involves large-scale conformational changes⁴. For instance, comparison of the translocation modules of TRCF (crystallized in the apo state)⁴ and RecG (crystallized in the ADP-bound state)¹⁰ reveals a large rotation ($> 90^\circ$) of TD2 with respect to TD1. In addition, in order for the TRCF UvrB homology module to interact with UvrA, D7 must move to unmask the UvrA binding determinant. Here, we probe these putative conformational changes by comparing limited proteolysis of repressed wt-TRCF and derepressed mutants (TRCF-D2*** and TRCF-D7***) of TRCF. Limited proteolysis has long been used to probe the conformational dynamics of proteins in solution, since proteases such as trypsin or chymotrypsin tend to cleave unstructured, dynamic segments of polypeptides rather than well structured regions¹⁷. We find that the repressed wt-TRCF is highly resistant to proteolysis, while the derepressed mutants are dramatically sensitized. Many of the protease cleavage sites observed in the derepressed mutants are buried to access from the solvent in the structure of the repressed state, strongly indicating that derepression is associated with profound conformational changes of TRCF.

Results

Proteolytic sensitivity of a derepressed TRCF mutant

We have not been able to crystallize wt-TRCF bound to ATP or non-hydrolyzable ATP analogs, presumably because the repressed wt-TRCF binds nucleotides poorly. Therefore, in order to crystallize nucleotide-bound states of TRCF, we sought to purify derepressed mutants of TRCF. We generated an expression construct overexpressing *Eco* TRCF with a C-terminal truncation similar to the previously studied Δ [998–1148]TRCF¹⁵, Δ [994–1148]TRCF (which we will call Δ D7-TRCF) along with a human rhinovirus 3C protease-cleavable N-terminal His₁₀-tag. The Δ D7-TRCF was highly overexpressed in the soluble fraction, and it was purified using essentially the same protocol as used to purify full-length TRCF¹⁸.

Unlike full-length TRCF, Δ D7-TRCF suffered *in situ* proteolytic degradation by an unknown protease during purification. Along with the expected 112 kDa Δ D7-TRCF, some of the final fractions from the gel filtration column were contaminated with a significant

amount of an ~87 kDa band (Fig. S1). N-terminal sequencing of the ~87 kDa band revealed that it had the expected N-terminal sequence of the human rhinovirus 3C-cleaved expression product (Table 1), confirming the ~87 kDa band as a proteolytic degradation product of Δ D7-TRCF. The mass of the proteolytic product was determined by matrix-assisted laser desorption/ionization-mass spectrometry to be 87.2 kDa, indicating that the proteolytic fragment results from cleavage C-terminal to residue 770 (Table 1, band u1), which, interestingly, is near the hinge-like junction between TD1 and TD2. These results revealed a striking difference in proteolytic sensitivity between the repressed wt-TRCF and the derepressed Δ D7-TRCF, and suggested that a more systematic probing of TRCF conformational states by limited proteolysis might be highly informative.

Limited proteolysis of derepressed TRCF mutants

We directly compared limited proteolysis of derepressed TRCF mutants TRCF-D2*** and TRCF-D7*** with repressed wt-TRCF. TRCF-D2*** and TRCF-D7*** are altered by only three point substitutions in each case, so remain the same polypeptide length, and essentially the same mass, as the wt-TRCF, facilitating direct comparison of protease cleavage by polyacrylamide gel electrophoresis (Figs. 2, S2). Moreover, the derepressed phenotype of the TRCF-D2*** and TRCF-D7*** mutants can be more easily attributed to the specific disruption of the D2/D7 intermolecular interface (Fig. 1B), in contrast to the wholesale deletion of large domains that yields the derepressed phenotypes of Δ N-TRCF and Δ D7-TRCF.

Each protein (wt-TRCF, TRCF-D2*** or TRCF-D7***) was incubated with increasing concentrations of protease for a fixed amount of time, then the cleavage reactions were halted by the addition of phenylmethylsulfonyl fluoride and SDS loading buffer, and the reactions were visualized by SDS-polyacrylamide gel electrophoresis and Coomassie staining. Two proteases were tested, trypsin (Figs. 2A, S2) and chymotrypsin (Fig. 2B). The results for TRCF-D2*** are shown in Fig. 2. The results for TRCF-D7*** were essentially identical (Fig. S2). For each protease, wt-TRCF was resistant to cleavage, even at the highest protease concentrations (Fig. 2A, lane 11; Fig. 2B, lane 11; Fig. S2, lane 11). In comparison, the derepressed TRCF mutant was dramatically sensitized to protease cleavage.

Mapping protease cleavage sites onto the TRCF sequence

The N-terminal sequences of the trypsin (Fig. 2A, t0 – t6) and chymotrypsin (Fig. 2B, c0 – c9) cleavage products for TRCF-D2*** were determined. The molecular weight of each band was estimated based on its gel mobility (using the mobilities of the molecular weight markers and of the full-length, uncleaved TRCF as calibrants; Table 1). Due to the error inherent in protein molecular weight determination by SDS-gel electrophoresis (estimated to be $\pm 5\%$) and the distribution of potential protease cleavage sites, the C-terminus for many of the cleavage products could not be determined precisely (Table 1, t1, t3, c1, c3, c7a). For bands t2, t4, t5, c5, and c7b, the C-terminus of the cleavage product was deduced to be the C-terminus of TRCF (residue 1148) since in each case i) this yielded the calculated molecular weight (MW_{calc}) that most closely matched the molecular weight determined from gel mobility (MW_{gel}), and ii) other potential cleavage sites fell within the well-structured, C-terminal D7. For band t6, the C-terminus (residue 953) was precisely determined since there were no other potential cleavage sites that yielded MW_{calc} values within experimental error of the MW_{gel} determination. The closest neighboring potential cleavage site (residue 929) yielded a MW_{calc} of 17.0 kDa (13% error). For bands c2, c4, c6, and c9, the C-terminal cleavage sites were assumed to correspond to precisely determined N-terminal cleavage sites of other cleavage products, since in each case this also yielded the MW_{calc} that most closely matched MW_{gel} . For example, the proteolytic cleavage giving rise

to the C-terminus of c2 (c2C, residue 620) also gives rise to the N-terminus of c9 (c9N, residue 621). Similarly for c4C/c6N, and (c6C, c9C)/(c5N, c7aN).

The overall results are tabulated in Table 1. The precisely determined protease cleavage sites are mapped onto a schematic of the TRCF domain architecture in Fig. 2C.

The mapping of protease cleavage sites onto the TRCF domain architecture (Fig. 2C) provides a number of insights:

- i. The most N-terminal precisely determined protease cleavage site was C-terminal to residue 464 (t2N, Table 1, Fig. 2C), suggesting that the N-terminal 464 residues, comprising domains D1-D3, maintains its protease-resistant structure in the repressed and derepressed TRCF conformations. This is consistent with Murphy et al.¹⁶, who found that the crystal structure of the N-terminal 450 residues of TRCF was identical to the same fragment in the context of full-length TRCF, leading to the conclusion that this TRCF fragment maintains a rigid structure in different contexts. This also indicates that the introduction of the point substitutions in TRCF-D2*** (R165A/R181A/F185A) did not grossly disturb the structure, supporting the conclusion that the effect of these substitutions on the derepression of TRCF can be attributed to the specific disruption of the D2/D7 interface (Fig. 1B).
- ii. The most C-terminal protease cleavage site was C-terminal to residue 972 (c7bN, Table 1, Fig. 2C), indicating that the C-terminal 177 residues, comprising domain D7, maintains its protease-resistant structure in the repressed and derepressed TRCF conformations. This indicates that the introduction of the point substitutions in TRCF-D7*** (E1045A/D1048A/R1049A) did not grossly disturb the D7 structure, supporting the conclusion that the effect of these substitutions on the activation of TRCF can be attributed to the specific disruption of the D2/D7 interface (Fig. 1B).
- iii. Interestingly, the cleavage sites are all located between the D3-RID linker and the C-terminal end of TD2. Thus, the cleavage sites are confined to the TRCF region directly involved in the RNAP release function, i.e. TRCF/RNAP protein/protein interaction (RID) and ATP-hydrolysis/dsDNA translocation activity (TD1/TD2).
- iv. The cleavage sites generally cluster near the domain boundaries (Fig. 2C): D3-RID linker (t3C, t2N), RID-TD1 boundary (c4C, c6N), TD1-TD2 hinge (u1C, t4N, t6N, c5N, c6C, c7aN, c9C), and the C-terminal end of TD2 (t5N, t6C, c7bN). This is consistent with the idea that TRCF derepression largely entails rearrangements of the intact structural domains through conformational changes of hinges or linkers connecting the structural domains, but the domains themselves maintain relatively rigid structures.

Effect of nucleotides on limited proteolysis of derepressed TRCF mutants

We tested the effects of nucleotides on limited trypsinolysis of repressed wt- and derepressed TRCF-D2***. As observed previously (Fig. 2A), in the absence of nucleotide, wt-TRCF was highly resistant to trypsinolysis, while TRCF-D2*** was dramatically more sensitive (Fig. 3, compare lanes 1 and 7 with lanes 4 and 10). The presence of ADP or the slowly hydrolyzable ATP analog ATP- γ S had little effect on wt-TRCF (lanes 2, 3, 8, 9). The full-length, wt-TRCF has very low ATPase activity and likely does not bind nucleotides^{4; 15}. In contrast, ADP and ATP- γ S stabilize TRCF-D2*** significantly against trypsinolysis (Fig. 3, compare lane 4 with 5 and 6, and lane 10 with 11 and 12), indicating effective binding of nucleotide to the derepressed mutant, as expected from the mutants increased ATPase activity^{7; 15}. Interestingly, the TRCF-D2***/ADP complex is nearly as

resistant to trypsinolysis as the wt-TRCF. In the experiments with ATP- γ S at the highest trypsin concentration (Fig. 3, lane 12), there is enough TRCF-D2*** cleavage to see that the cleavage pattern is qualitatively the same as without nucleotide, but there appear to be quantitative differences. It is most informative to compare Fig. 3, lane 4 (TRCF-D2***, no nucleotide, 50:1 TRCF:trypsin) with lane 12 (TRCF-D2***, ATP γ S, 25:1 TRCF:trypsin), where the amount of uncleaved, full-length TRCF-D2*** is similar. The generation of certain cleavage products (t2 and t3) appears to be similarly efficient with or without ATP- γ S, while other cleavage products (for instance, t4) are relatively weak in the presence of ATP- γ S compared with its absence. Interestingly, the cleavage product that is relatively suppressed in the presence of ATP- γ S (t4) results from cleavage near the boundary between TD1 and TD2, the two domains of the translocation module (Fig. 2C). Comparison of the apo-TRCF translocation module with the highly homologous translocation module of RecG/ADP¹⁰ indicated that in different nucleotide states, TD1 and TD2 rotate as relatively rigid bodies with respect to each other, with the boundary region between them serving as a hinge⁴. This is consistent with the view that in the derepressed mutant in the absence of nucleotide, TD1 and TD2 are relatively unconstrained, giving rise to proteolytic sensitivity in the flexible hinge between them. Binding of nucleotide between TD1 and TD2¹⁰ restricts their flexibility with respect to each other, stabilizing the hinge region against proteolysis. The relatively efficient generation of t1 and t3 in the presence of ATP- γ S (both bands result from cleavage in the D3-RID linker; Fig. 2C) indicates that nucleotide binding specifically stabilizes the conformation of the translocation module while other parts of the structure remain relatively open and flexible.

Discussion

Through a conserved, intermolecular, interdomain interaction between D2 and D7, TRCF is held in a repressed conformation represented by the apo-TRCF crystal structure^{4; 15}. In this state, TRCF shows only very weak ATPase activity (rate of ATP hydrolysis < 10/min), the ATPase activity is not stimulated by DNA, and DNA translocase activity is undetectable¹⁵. Disruption of the D2/D7 intermolecular interaction, either by deleting one of the relevant domains, or by amino acid substitutions within the interface, derepresses ATPase and DNA translocase activities^{7; 15; 16}. In this work, we have shown that derepressed TRCF mutants are dramatically sensitized to limited proteolysis compared with the wt, repressed TRCF (Figs. 2A, 2B), suggesting an altered conformational state.

Analysis of the protease cleavage sites mapped onto the structure of the TRCF repressed conformation suggests that the conformational changes associated with derepression are profound (Fig. 4; Table 2). A total of 9 distinct protease cleavage sites, sensitized in the derepressed TRCF, were observed and precisely mapped (Fig. 2C; Table 1). Of these, three sites (t2N, t5N/t6C, and c7bN) were relatively solvent exposed but directly abut adjacent TRCF domains that would appear to make them inaccessible to a protease (Fig. 4; shaded yellow in Table 2). Two sites (u1C, t6N) appear to be mostly or even completely buried in the repressed TRCF structure (Fig. 4; shaded pink in Table 2).

In order for a protease molecule to gain access to these buried sites, the following structural changes appear to be required:

- i. Cleavage site t2N is in the D3-RID linker, sandwiched between D2, TD1, and TD2 (Fig. 4A). For a protease to gain access to t2N, it appears that the D3-RID linker would need to be displaced from its position in the repressed TRCF structure.
- ii. Cleavage site u1C, near the C-terminal end of TD1, is buried underneath the relay-helix, which is at the N-terminal end of TD1 (Fig. 4A). For a protease to gain

access to u1C, it appears that the disposition of the relay-helix with respect to the rest of TD1 would need to change dramatically.

- iii. Cleavage site c5N/c7aN/c6C/c9C, located near the C-terminus of TD1, is nearly completely buried in an interface between TD1 and TD2 (Fig. 4A). For a protease to gain access to c5N/c7aN/c6C/c9C, TD1 and TD2 would have to substantially reorient and separate.
- iv. Cleavage site t6N, located near the N-terminus of TD2, is completely buried beneath the D3-RID linker (Fig. 4A). For a protease to gain access to t6N, it appears that the D3-RID linker would need to be displaced from its location in the repressed TRCF structure.
- v. Cleavage site t5N/t6C, near the C-terminal end of the TRG motif of TD2, is relatively solvent exposed, but directly abuts the RID (Fig. 4B). For a protease to gain access to t5N/t6C, the RID would need to be displaced from TD1.
- vi. Cleavage site c7bN, located near the C-terminus of TD2, is relatively solvent exposed, but is located in a deep trough between the RID, TD1, and D7 (Fig. 4A). For a protease to gain access to c7bN, the RID (and by extension the D3-RID linker), and D7 would need to move away from TD2.

Because of the compact domain architecture of the repressed TRCF, it is difficult to envision that any one of these structural changes could occur without affecting domain/domain interactions in other parts of the structure. Thus, exposure of these inaccessible sites to make them accessible to a protease implies a significant ‘unraveling’ of the compact domain architecture seen in the repressed TRCF structure.

Interestingly, in experiments using even higher concentrations of protease to effect more extensive cleavage of wt-TRCF, the cleavage pattern of wt-TRCF and both derepressed mutants was qualitatively the same (data not shown), supporting the scenario in which wt-TRCF and the derepressed mutants exists in an equilibrium between at least two conformational states, a ‘closed’, repressed conformation that is protease resistant (and likely corresponds to the conformational state of the crystal structure; Fig. 1B), and an ‘open’, derepressed conformation that is protease sensitive. The closed and open conformations for wt-TRCF and the derepressed mutants are the same (or are very similar), but wt-TRCF is strongly biased towards the closed conformation, while the derepressed mutants are biased towards the open conformation.

The presence of nucleotides (ADP or ATP- γ S) stabilized derepressed TRCF-D2*** to protease cleavage (Fig. 3), implying that the nucleotides bind. In the case of ATP- γ S, nucleotide binding appears to preferentially inhibit cleavage of sites located near the TD1/TD2 hinge, which is consistent with binding of the slowly-hydrolyzable ATP analog in the nucleotide binding pocket between TD1 and TD2¹⁰ and fixing or rigidifying the orientation of the two domains with respect to each other.

While the derepression of ATPase and DNA translocase activities of TRCF can be effected by mutations (either truncations or substitutions) that disrupt the D2/D7 intermolecular interaction, these activities are depressed in the wt-TRCF by engaging with the RNAP ternary elongation complex (TEC). Presumably, the D2/D7 interface is broken in this case where the TRCF enzymatic activities are activated through the normal interaction with the TEC. The structural mechanism through which the D2/D7 interface is broken by interactions with the TEC, thereby allowing the TRCF enzymatic activities to displace that very same TEC, is an important area for further research.

Materials and Methods

Protein Preparation

Constructs for the derepressed TRCF mutants (TRCF-D2***, pMFD19RA165RA181FA185; TRCF-D7***, pETMfdT7EA1045DA1048RA1049) were a gift from the laboratory of N. Savery⁷. The relevant mutations were introduced into a pET-based TRCF expression system (pAD6; 4; 18. For TRCF-D2***, a BseRI/BsrGI fragment was transferred from pMFD19RA165RA181FA185 into pAD6, creating pTRCF-D2***. For TRCF-D7***, a BseRI/HindIII fragment was transferred from pETMfdT7EA1045DA1048RA1049 into pAD6, creating pTRCF-D7***. All DNA manipulations were confirmed by DNA sequencing. The wt-TRCF, TRCF-D2***, and TRCF-D7*** were all expressed and purified using the identical protocol as described previously^{4; 18}. Bovine pancreatic trypsin (Sigma, lyophilized powder) was resuspended in 1 mM HCl at a concentration of 10 or 20 μ M. The α -chymotrypsin (Sigma, lyophilized powder) was resuspended in 1 mM HCl, 2 mM CaCl₂, at a concentration of 10 or 20 μ M. The proteases were stored in aliquots at -80°C .

Limited Proteolysis Assay

Protease reactions (total of 10 μ L) containing 50 μ M TRCF (or TRCF-D2*** or TRCF-D7*** in 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5% (v/v) glycerol, 0.1 mM EDTA, 1 mM DTT, were incubated with the amounts of protease indicated in the legend to Fig. 2 at 25°C for 1 hour. The reaction was halted by adding 1 μ L of phenylmethylsulfonyl fluoride stock solution (100 μ M in ethanol). The reactions were then mixed quickly with 5 μ L of 2X SDS loading buffer containing NuPAGE antioxidant (Invitrogen) as well as additional β -mercaptoethanol, boiled for 4 min. at 95°C, then analyzed by SDS-polyacrylamide gel electrophoresis on 4–12% Bis-Tris polyacrylamide gels (Invitrogen) with Coomassie blue staining.

The protease reactions in the presence of nucleotides (ADP or ATP γ S; Fig. 3) were performed using an identical protocol, except that the protein samples were pre-incubated at 25°C for 90 min. in the presence of 2 μ M nucleotide (ADP, Sigma; ATP γ S, Roche) and 10 mM MgCl₂ before addition of the protease.

N-terminal Protein Sequencing

All N-terminal protein sequencing was performed by the University of Texas Medical Branch at Galveston Biomolecular Resource Facility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

N. Savery for the TRCF-D2*** and TRCF-D7*** constructs, which were provided before their publication, as well as for helpful discussion throughout this project; M. Bick for mass spectrometry analysis and other assistance at the beginning stages of this project, as well as members of the Darst laboratory for helpful discussion; B.T. Chait for access to mass-spectrometry equipment. This work was supported by NIH GM073829 to S.A.D.

Abbreviations

dsDNA	double-stranded DNA
NER	nucleotide excision repair

RNAP	RNA polymerase
TD1	translocation domain 1
TD2	translocation domain 2
TEC	ternary elongation complex
TRCF	transcription-repair coupling factor
wt	wild-type

References

1. Deaconescu AM, Savery N, Darst SA. The bacterial transcription repair coupling factor. *Curr Opin Struct Biol.* 2007; 17:96–102.
2. Savery NJ. The molecular mechanism of transcription-coupled DNA repair. *Trends Microbiol.* 2007; 15:326–333. [PubMed: 17572090]
3. Selby CP, Sancar A. Molecular mechanism of transcription-repair coupling. *Science.* 1993; 260:53–8. [PubMed: 8465200]
4. Deaconescu AM, Chambers AL, Smith AJ, Nickels BE, Hochschild A, Savery NJ, Darst SA. Structural basis for bacterial transcription-coupled DNA repair. *Cell.* 2006; 124:507–520. [PubMed: 16469698]
5. Pakotiprapha D, Liu Y, Verdine GL, Jeruzalmski D. A structural model for the damage-sensing complex in bacterial nucleotide excision repair. *J Biol Chem.* 2009; 284:12837–12844. [PubMed: 19287003]
6. Selby CP, Sancar A. Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *J Biol Chem.* 1995; 270:4882–9. [PubMed: 7876261]
7. Manelyte L, Kim T, Smith AJ, Smith RM, Savery NJ. Regulation and rate-enhancement during transcription-coupled DNA repair. *Mol Cell.* 2010 in press.
8. Smith AJ, Savery NJ. RNA polymerase mutants defective in the initiation of transcription-coupled DNA repair. *Nucleic Acids Res.* 2005; 33:755–64. [PubMed: 15687384]
9. Westblade LF, Campbell EA, Pukhrambam C, Padovan JC, Nickels BE, Lamour V, Darst SA. Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic Acids Res.* 2010; 1093/nar/gkq692
10. Singleton MR, Scaife S, Wigley DB. Structural analysis of DNA replication fork reversal by RecG. *Cell.* 2001; 107:79–89. [PubMed: 11595187]
11. Gorbalenya AE, Koonin EV, Domchenko AP, Blinov VM. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* 1989; 17:4713–4729. [PubMed: 2546125]
12. Chambers AL, Smith AJ, Savery NJ. A DNA translocation motif in the bacterial transcription-repair coupling factor, Mfd. *Nucleic Acids Res.* 2003; 31:6409–6418. [PubMed: 14602898]
13. Mahdi AA, Briggs GS, Sharples GJ, Wen Q, Lloyd RG. A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. *EMBO J.* 2003; 22:724–734. [PubMed: 12554672]
14. Park JS, Marr MT, Roberts JW. *E. coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell.* 2002; 109:757–767. [PubMed: 12086674]
15. Smith AG, Szczelkun MD, Savery NJ. Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. *Nucleic Acids Res.* 2007; 35:1802–1811. [PubMed: 17329375]
16. Murphy MN, Gong P, Ralto K, Manelyte L, Savery NJ, Theis K. An N-terminal clamp restrains the motor domains of the bacterial transcription-repair coupling factor Mfd. *Nucleic Acids Res.* 2009; 37:6042–6053. [PubMed: 19700770]

17. Wilson JE. The use of monoclonal antibodies and limited proteolysis in elucidation of structure-function relationships in proteins. *Meth Biochem Anal.* 1991; 35:207–250.
18. Deaconescu AM, Darst SA. Crystallization and preliminary structure determination of *Escherichia coli* Mfd, the transcription-repair coupling factor. *Acta Crystallogr F.* 2006; 61:1062–1064.

Fig. 1.

Structural architecture of repressed *Eco* TRCF.

A. Schematic representation of the TRCF domain architecture⁴. The horizontal bar represents the 1148-residue *Eco* TRCF primary sequence (every 100 residues are marked below the bar). Structural domains are represented as thick, colored bars, thin black bars represent linkers connecting the domains. D1a, D2, and D1b, named according to the same domains in the structurally homologous UvrB RID stands for RNAP Interacting Domain. TD1 and TD2 are Translocation Domains 1 and 2, which contain the seven SF2 ATPase motifs¹¹ and the TRG motif^{12; 13} (shown in white boxes) and comprise the translocation module. On top are shown the locations of the three amino acid substitutions in derepressed TRCF-D2*** and TRCF-D7***⁷.

B. Structure of repressed *Eco* TRCF and the D2/D7 interface. (left) Top view of the α carbon backbone of *Eco* TRCF, color-coded as in the schematic of Figure 1A. The intermolecular, interdomain interface between D2 and D7 is highlighted as molecular surfaces: Atoms of D2 within 4.5 Å of D7 are shown as a cyan surface, atoms of D7 within 4.5 Å of D2 are shown as a red surface. The boxed region is magnified on the right. (right) Magnified view of the D2/D7 interface, except the backbone of D2 and D7 are colored pale-cyan and pink, and the surfaces denoting the D2/D7 interface (4.5 Å cutoff) are transparent. The side chains of the residues substituted in TRCF-D2*** (R165, R181, and F185) are shown with carbon atoms colored cyan. The side chains of the residues substituted in TRCF-D7*** are shown with carbon atoms colored red.

Fig. 2.

Limited proteolysis of repressed wt-TRCF and derepressed TRCF-D2***.

A. Limited trypsinolysis of the 130 kDa TRCF (wt) or TRCF-D2*** (mut) analyzed by SDS-PAGE and Coomassie blue staining on a 4–12% gradient Bis-Tris acrylamide gel. TRCF or TRCF-D2*** at 50 μ M was incubated with no trypsin (lanes 1 and 2), or with increasing concentrations of trypsin (lanes 3, 4 – 0.5 μ M trypsin; lanes 5, 6 – 1 μ M; lanes 7, 8 – 2 μ M; lanes 9, 10 – 5 μ M; lanes 11, 12 – 10 μ M) for 1 hour at 25°C. The molecular weights (in kDa) of the molecular weight marker bands (lane labeled ‘MW’) are shown on the left. The bands used for N-terminal sequencing (Table 1) are labeled on the right.

B. As in Figure 2A, except the protease used was chymotrypsin.

C. Schematic of the TRCF domain architecture (as in Figure 1A). Below are shown the locations of the mapped protease cleavage sites (Table 1). Sites are labeled xyZ: x is u, t, or c, for unknown protease, trypsin or chymotrypsin, respectively; y is the number of the band (as labeled in Figures 2A or 2B and Table 1); Z is N or C, for N-terminus or C-terminus of the labeled band. The sites are color-coded as follows: C-terminus of *in situ* cleavage product from unknown protease (u1, Figure S1), green; N-terminus of trypsin products, red; C-terminus of trypsin products, magenta; N-terminus of chymotrypsin products, blue. The cleavage sites are grouped in clusters: a, D3-RID linker; b, RID-TD1 boundary; c, within TD1; d, TD1-TD2 boundary; e, C-terminus of TD2 and TD2-D7 boundary.

Fig. 3.

Effect of nucleotides on trypsinolysis of wt-TRCF or TRCF-D2***.

Limited trypsinolysis of the 130 kDa TRCF (wt) or TRCF-D2*** (mut) analyzed by SDS-PAGE and Coomassie blue staining on a 4–12% gradient Bis-Tris acrylamide gel. TRCF or TRCF-D2*** at 50 μ M was incubated with trypsin (lanes 1–6, 1 μ M trypsin; lanes 7–12, 2 μ M trypsin) for 90 minutes at 25°C, and with no nucleotide (lanes 1, 4, 7, and 10), 2 mM ADP (lanes 2, 5, 8, and 11), or 2 mM ATP- γ S (lanes 3, 6, 9, and 12). The molecular weights (in kDa) of the molecular weight marker bands (lane labeled 'MW') are shown on the left. The trypsin products (Table 1) are labeled on the right. Addition of ADP to the wt-TRCF caused some precipitation of the protein that was difficult to control, explaining the low intensity of the full-length band (lanes 2 and 8).

Fig. 4.

The mapped protease cleavage sites in the context of the repressed TRCF structure. The structural views of the repressed *Eco* TRCF⁴ are shown as α -carbon backbone worms (color-coded as in the schematic of Figs. 1A and 2C) as well as with a grey, transparent molecular surface. The α -carbons flanking the mapped protease cleavage sites are shown as spheres and color-coded as in Fig. 2C.

A) Top view.

B) Bottom view.

Table 1

Mapping of protease cleavage products within the TRCF sequence.

band	N-term sequence ^a	C-term ^b	(kDa) ^c MW _{MS}	MW _{calc} (kDa)
u1	⁻³ GPHM	770	87.2	87.2

band	N-terminal sequence ^a	C-term ^b	(kDa) ± MW _{gel} 5% ^c	(kDa) MW _{calc}
t1	⁻³ ?PHM	ind. ^d	85.7 ± 4.3	-
t2	⁴⁶⁴ TINPD	1148	74.3 ± 3.7	77.3
t3	⁻³ GPHM	ind., but within D3-RID linker	51.7 ± 2.6	-
t4	⁷⁷² DLSII	1148	42.5 ± 2.1	42.9
t5	⁹⁵⁴ GAGEL	1148	21.7 ± 1.1	22.3
t6	⁷⁸³ RLAVK	953	19.6 ± 1.0	19.5
c1	⁻³ GPHM	ind.	77.2 ± 3.9	-
c2	⁻³ GPHM	620	70.0 ± 3.5	70.3
c3	⁻³ GPHM	ind.	66.5 ± 3.3	-
c4	⁻³ GPHM	550	60.1 ± 3.0	62.4
c5	⁷⁶⁴ NMAMS	1148	43.4 ± 2.2	43.8
c6	⁵⁵¹ SRARQ	763	24.4 ± 1.2	24.2
c7a	⁷⁶⁴ ?MAMS	ind.	20.0 ± 1.0	-
c7b	⁹⁷³ ?LYME	1148	20.0 ± 1.0	20.4
c9	⁶²¹ AMDRL	763	15.7 ± 0.8	16.2

^aThe N-terminal Met residue of the endogenous *Eco* TRCF is denoted residue 1 in the sequence numbering. N-terminal, vector derived residues are numbered ⁻³G⁻²P⁻¹H.

^bThe C-terminal residues were estimated from the experimentally determined mass (either by matrix assisted laser desorption-mass spectrometry, band u1, or from the SDS-PAGE mobility), as well as a consideration of the known protease specificity (for the trypsin and chymotrypsin results).

^cThe molecular weight for band u1 (MW_{MS}) was determined by matrix assisted laser desorption-mass spectrometry. Other molecular weights were estimated from the SDS-PAGE mobility.

^dIndeterminate (ind.) due to error in the MW_{gel} determination.

Table 2

Structural analysis of protease cleavage sites.

Region	Figure	Cleavage site ^a	Status in wt-TRCF structure ^a
D3-RID linker	4A	t2N	Partially buried - abutting D2
		t3C	
RID-TD1 boundary	4B	c6N/c4C	Relatively exposed
Within TD1	4A	c1C	Relatively exposed
	4A	c9N/c2C	Relatively exposed
TD1-TD2 boundary	4A	u1C	Buried
		t4N	Relatively exposed
		c5N/c7aN/c6C/c9C	Nearly completely buried
		t6N	Buried beneath D3-RID linker
C-terminal end of TD2/TD2-D7 boundary	4B	t5N/t6C	Relatively solvent exposed – abutting the RID
	4A	c7bN	Relatively solvent exposed – abutting the RID, TD1, and D7

^a Pink shading indicates the cleavage site is completely buried from protease access within the repressed TRCF structure. Yellow shading indicates the cleavage site is partially buried from protease access within the repressed TRCF structure.