TREX1's Homodimer Structure Has Evolved for Double-Stranded DNA Degradation

Wayne O. Hemphill[1,3,4,], Thomas Hollis[1], Freddie R. Salsbury Jr.[2], Fred W. Perrino[1,*]*

[1]Wake Forest School of Medicine, Department of Biochemistry, Center for Structural Biology
[2]Wake Forest University, Department of Physics
[3]University of Colorado Boulder, Department of Biochemistry & BioFrontiers Institute
[4]Howard Hughes Medical Institute
[*]Corresponding Authors: wayne.hemphill@colorado.edu

ABSTRACT

The 3' → 5' exonuclease TREX1 functions *in vivo* to degrade DNA and prevent chronic immune activation through the cGAS-STING pathway. TREX1 has been described as a nonprocessive enzyme with little preference for ss- versus dsDNA, raising questions about its principle *in vivo* substrate. In addition, there has been minimal data rationalizing TREX1's obligate homodimer structure. We used a combination of biochemical assays, molecular dynamics simulations, and robust kinetic modeling to determine TREX1's relative affinities for ss- versus dsDNA substrates, to quantify its processivity, and to interrogate the possibility of inter-protomer communication within the TREX1 homodimer. Our findings indicate that TREX1 is a semi-processive exonuclease with a predicted several orders of magnitude greater affinity for dsDNA than ssDNA. Furthermore, we find extensively correlated dynamics between TREX1's protomers that involve previously described and unrecognized substrate-adjacent regions of the TREX1 enzyme. We infer that TREX1 has evolved as a semi-processive exonuclease that likely degrades dsDNA *in vivo*, and we propose that its homodimer structure facilitates a mechanism for efficient binding and catabolism of nicked dsDNA. Overall, these studies identify potentially critical regions of the TREX1 enzyme not previously interrogated and contribute to a still-emerging model of TREX1 exonuclease activity.

INTRODUCTION

Three-prime Repair EXonuclease 1 (TREX1) is a 3' → 5' exonuclease[1] that functions to degrade DNA and prevent aberrant DNA-sensing through the cGAS-STING pathway (reviewed in ref[2]), and *TREX1* mutations are associated with a spectrum of autoimmune disorders (reviewed in ref[3]). While it's clear that TREX1 activity in hematopoietic cells is the deciding factor in whether or not discernable pathology occurs[4,5], the source(s) of TREX1's pathogenically relevant biological substrate(s) remains an area of active investigation (ex. replication intermediates[6], retroelements[7], micronuclei[8], enucleated erythroblast DNA[9]). Biochemical studies of TREX1 indicate robust exonuclease activity on single- (ss) and double-stranded (ds) DNA substrates (but not RNA or RNA-DNA duplexes)[1,10–12], and suggest a preference for ss- over dsDNA[10]. Collectively, these findings perpetuate the hypothesis that TREX1's pathogenically relevant *in vivo* substrate is ssDNA. However, the nucleic acid sensor driving pathology in *TREX1* mutants[13–15], cGAS, has specificity for dsDNA[16]. Furthermore, the *TREX1* p.D18N variant is dominant for aggressive autoimmune phenotypes *in vivo*[3,17], but the recombinant TREX1^{D18N} mutant enzymes only recapitulate this dominance *in vitro* on dsDNA (not ssDNA) substrates[17–19]. Collectively, these *in vitro* and *in vivo* data make the contrary suggestion that dsDNA, not ssDNA, is the pathogenically relevant TREX1 substrate. Consequently, the question of TREX1's biochemical preference for ds- versus ssDNA remains highly relevant to its biological function.

Similarly, TREX1's obligate homodimer structure[10,20–22] is likely relevant to its biological function, but this relationship is not fully understood. Obligate homodimerization is not a necessity for catalytic activity in many other exonucleases[23,24], and this implies a function specific to TREX1's biological role. We have previously demonstrated that residues from one protomer communicate across TREX1's dimer interface and contribute to activity in the opposing protomer[21,25]. While these studies establish that TREX1's homodimer structure is critical for its exonuclease activity, they don't explain what biological or evolutionary imperative this serves. One model supported by existing studies[21,25] is that the presence and/or conformation of substrate in one protomer can affect activity in the opposing protomer. Potential explanations for such inter-protomer communication might include inter-protomer regulation of catalysis and/or binding, or an *in vivo* substrate binding mechanism that requires simultaneous engagement of the active-sites. The former explanation is supported by reports about the exonuclease activities of TREX1^{D18N} heterodimers[17,19], which suggest the D18N heterodimer doesn't degrade dsDNA but degrades ssDNA at ~50% wild-type rate. While these collective data have provided an initial framework for understanding TREX1's obligate homodimer structure, elucidating its deeper purpose requires more mechanistic interrogations.

In this work we use a combination of biochemical and computational approaches to interrogate TREX1's preference for single- versus double-stranded DNA and the functional purpose(s) of its obligate homodimer structure. The relative biochemical affinities of TREX1 for ss- versus dsDNA are convincingly demonstrated, and residues contributing to differential binding are proposed. As part of these endeavors, a new robust kinetic model of TREX1 exonuclease activity is also developed and validated from biochemical data. Finally, we explain prior observations about the TREX1 D18N heterodimer's activity, and we further interrogate the phenomenon of inter-protomer communication in the TREX1 homodimer via molecular dynamics simulations.

RESULTS

Determination of TREX1's k_{cat} and K_m:

The catalytic rate (k_{cat}) and Michaelis-Menten (K_m) constants for TREX1 with ss- and dsDNA substrates have been reported once before[10], and suggest comparable affinity for the two substrates (ssDNA: k_{cat} = 22 ± 1.1 s^{-1}, K_m = 19 ± 4.0 nM; dsDNA: k_{cat} = 11 ± 0.77 s^{-1}, K_m = 15 ± 4.4 nM). However, we re-determined TREX1's classic kinetic constants for ss- and dsDNA substrates (Figure 1 and Table 1).

For dsDNA, we developed a novel absorbance-based exonuclease assay. Briefly, recombinant human TREX1$_{1-242}$ enzyme (hT1) is incubated with a large, nicked plasmid substrate, then the reaction is terminated by addition of a small volume of concentrated ethylenediaminetetraacetic acid (EDTA), and the excised deoxynucleoside monophosphates (dNMPs) are separated via a microcentrifugal filter unit. Finally, the concentration of the dNMPs is quantified via spectroscopy and used to calculate activity rate. The sensitivity range, dNMP standard curve, and degradation linearity were all determined for the assay (Figure 1A-B), then initial velocities of hT1 reactions were quantified over a range of substrate concentrations to generate a

Michaelis-Menten curve (Figure 1C). These data indicated that k_{cat} = 5.0 s^{-1} and K_m = 10 nM. However, subsequent findings (see 'Competitive Inhibition…' section in Results) suggest only one active-site of the hT1 homodimer can degrade this ~10-kb dsDNA plasmid at a time, which implies $k_{cat} \approx 10\ s^{-1}$ based on the number of TREX1 active-sites truly available for catalysis under such conditions. Overall, our determinations of TREX1's dsDNA k_{cat} and K_m are generally consistent with prior literature (Table 1 vs ref[10]).

For ssDNA, we employed a reported[26] gel-based assay to quantify hT1 exonuclease activities with high and low ssDNA concentrations (Figure 1D). The hT1 ssDNA k_{cat} and K_m were approximated from these activities and derived equations (Equations 7-8) using Monte-Carlo simulation. Interestingly, while the approximated k_{cat} was in good agreement with the previous report, the approximated K_m was about an order in magnitude greater (Table 1 vs ref[10]). Existing reports of TREX1's ssDNA activity rate ($\frac{v_0}{[E_T]}$) at a fixed substrate concentration[19,21,22,25,27] (1.4-6.3 s^{-1} at 50nM substrate) are more consistent with our kinetic constant values (Equation 1 predicts ~4.7 s^{-1} at 50nM substrate) than those previously published[10] (Equation 1 predicts ~16 s^{-1} at 50nM substrate). Thus, our kinetic constant values for TREX1 suggest a preference for dsDNA over ssDNA, contrary to prior implications.

Comparison of TREX1's ss- and dsDNA Binding Stabilities:

To validate our kinetic assays, we investigated relative ss- and dsDNA binding stabilities for TREX1 via thermal shift assays (Figure 2), using recombinant human (hT1) and mouse (mT1) enzymes. As anticipated, hT1 gained much greater thermostability upon ds- ($\Delta T_m \approx +28.3$ °C) versus ssDNA ($\Delta T_m \approx +13.6$ °C) binding, and the same trend was observed for mT1 ($\Delta T_m \approx +20.1$ vs +7.1 °C). We also note that hT1 and mT1 displayed significantly different melting temperatures as apoenzymes (Figure 2), despite their respective TREX1-DNA complexes having comparable melting temperatures. However, both apoenzymes had melting temperatures too high for direct physiological relevance to humans or mice (T_m > 50 °C).

Comparison of TREX1's ss- and dsDNA Binding Dynamics:

Our biochemical experiments indicate significantly different TREX1-dsDNA versus -ssDNA binding affinities, but prior crystal structures of these interactions only capture relatively static protein-polynucleotide contacts that are highly conserved across the two substrates[22,28]. To better define the amino acid residues that may contribute to the differential TREX1 binding to ss- versus dsDNA, we performed all-atoms molecular dynamics (MD) simulations with mT1 as apoenzyme, bound to 4-mer oligo, and bound to dsDNA. RMSD comparisons of mT1 against initial conditions shows that all simulations quickly equilibrate without indications of instability (Supplemental 1), and visualization of the conformational free energy maps for the mT1 backbone (Cα) in each of these systems reveals stark contrast in the distribution of structural states, particularly between the two mT1-DNA complexes (Supplemental 2). RMSF analysis of the Cα in each system identified three areas of differential fluctuation in the mT1 backbone which have the potential for direct interaction with substrate in the active-site (Figure 3A). Notably, TREX1's flexible loop (aa. 164-175) seems differentially impacted by the two substrates, being significantly rigidified during dsDNA occupation of the active-site (Figure 3A – green markers). We also identified a short region (aa. 23-27) with the capacity to contact the 3'-terminal nucleotides of each substrate, which is mildly rigidified during 4-mer occupation and further rigidified by dsDNA occupation (Figure 3A – purple markers). Finally, another uncharacterized region (aa. 75-88) was identified (Figure 4 – orange markers), and its rigidity appears unchanged by 4-mer occupation of the active-sites while occupation by dsDNA leads to modest rigidification.

We also subjected residues' phi/psi angles across all simulations and systems to a clustering analysis. Pairwise statistical comparisons of each residue's conformation (cluster) proportions were made between the systems and adjusted for false discovery rate (Figure 3B). As expected, several residues' conformations were significantly altered upon both 4-mer and dsDNA occupation of the active-site (Figure 3B – blue markers). These included R174, L19, and aa. 191-193, which are all adjacent to the active-site with the potential for DNA interactions. Several residues' conformations were only altered by dsDNA occupation (Figure 3B – red markers), including S48, S194, and Q209. Of these, S194 is closest to the active-site, and adjacent to the H195 residue previously identified as a residue of interest in TREX1 crystal structures[18,22] and patient data[3]. Curiously, there were several residues whose conformations were affected exclusively during 4-mer occupation of the active-site (Figure 3B – green markers), but only K175 is in any apparent proximity to the active-site.

Kinetic Modeling of TREX1 Exonuclease Activity:

A classic kinetic model fails to account for several potential characteristics of TREX1 including: (1) degradation of a polymeric substrate where the relevant metric for substrate concentration (free 3'-hydroxyls) remains constant between most catalytic events, and (2) non-distributive catalytic activity. We developed a kinetic model for TREX1 (Figure 4A) that addresses these shortcomings, incorporates the idea of processivity as a probabilistic adjustment to the catalytic rate constant, and can be expanded to describe competition between TREX1 variants (Figure 4B). Our model assumes length-independent catalytic rates and irreversible catalysis, which is generally consistent with TREX1 biochemistry[1,10,11,29]. Under our model, reactions are described by systems of nonlinear differential equations (Equations 16.1-17.8), and the number of nucleotides processively excised by TREX1 following DNA binding is described by a discrete probability density function (Equation 18).

We used empirical data to approximate our model's kinetic constant values for hT1 exonuclease activity. Our model (Figure 4A) indicates that turnover rate (Equation 16.2) reduces to $k_2[E_T]$ (classically[30] V_{max}) under saturating substrate conditions if enzyme-product complexing is negligible. Thus, the k_2 (catalytic constant) parameters for both dsDNA and ssDNA have been empirically determined by our initial studies (Table 1). In addition, the previously reported[31] ssDNA K_d value for TREX1 is applicable to our model with the classic[30] relationship $k_{-1} = k_1 K_d$ (Equation 19), but a value for TREX1's dsDNA K_d is unavailable. Using the constraint of Equation 19, our ssDNA exonuclease activity rates (Figure 1D, [S] = 15 nM), and numerical solutions of Equations 16.1-5, we approximated the k_1 (on-rate constant) and k_{-1} (off-rate constant) values that best explained our observed activity rates for a range of Ψ (processivity constant) values (Supplemental 3). Then, we compared our empirical ssDNA degradation patterns (Figure 1D, [S] = 15 nM) to those predicted by the model solutions for each processivity constant value (Figure 5A), and we found $\Psi \approx 0.75$, $k_{-1} \approx 2.2\ s^{-1}$, and $k_1 \approx 5.2 \times 10^7\ M^{-1}s^{-1}$ best recapitulated our data (Table 2).

Since TREX1's dsDNA K_d is unavailable, we used a reported[26] fluorescence-based assay to perform competition experiments with hT1 and recombinant human $TREX1_{1\text{-}242}^{D18N/D18N}$ (hT1-DN) enzymes, then determined the on- and off-rate constant values that best matched our observations for a range of processivity constant values (Supplemental 4). hT1-DN can competitively inhibit wild-type enzyme on ds- but not ssDNA, despite having specifically abolished catalysis, implying this behavior is a manifestation of differing ss- versus dsDNA kinetic constant values[19,25,26,31,32]. To capture this hT1-DN biochemistry in our modeling, the hT1-DN k_{cat} was set to zero and all other kinetic constants were constrained to wild-type values. We found that most processivity constant values had solutions for k_1 and k_{-1} that could recapitulate our data, but for $\Psi \geq 0.90$ there were no solutions that well recapitulated our data (Supplemental 4D-F). Under the additional assumption that $\Psi_{ds} \geq \Psi_{ss}$, consistent with TREX1's greater dsDNA binding stability demonstrated earlier, we infer for the TREX1-dsDNA interaction that $0.75 \leq \Psi \leq 0.90$, $0.020 \leq k_{-1} \leq 0.10\ s^{-1}$, and $k_1 \approx 10^9\ M^{-1}s^{-1}$ (Table 2). However, kinetic constant solutions for $\Psi_{ds} = 0$ were similar ($k_1 = 5.0 \times 10^9\ M^{-1}s^{-1}$, $k_{-1} = 0.30\ s^{-1}$), reinforcing that drastically different TREX1 kinetic constant values for ss- versus dsDNA are necessary to explain empirical data. Strikingly, our inferred TREX1 rate constants (Table 2) indicate dsDNA $K_d \approx$ 20-100 pM, versus ssDNA $K_d \approx 42$ nM, consistent with our earlier empirical data (Figures 1-2).

Finally, we used our inferred TREX1 rate constants (Table 2) to predict (via Equation 18) the number of nucleotides TREX1 is likely to processively excise on ss- versus dsDNA substrates (Figure 5B). Interestingly, our model predicts that TREX1 is unlikely to excise more than ~10 nucleotides processively from either substrate. Notably, while TREX1 processively excising ≥7 nucleotides seems more likely for ds- than ssDNA substrates, the absolute probability of such events is still low.

Competitive Inhibition Between Wild-Type and D18N TREX1 Protomers:

Early biochemical investigations into the exonuclease activities of $TREX1^{D18N}$ variants[17,19,26,32] were consistent with a model where substrate binding by one protomer caused communication across the dimer interface to prevent catalysis in the opposing protomer. However, the dsDNA activity for the heterodimers was only tested under specific substrate conditions: sub-enzyme concentrations of a bulky, nicked plasmid. Thus, the D18N heterodimer inactivity could be alternatively attributed to substrate-mediated steric hindrance of additional binding in the second protomer's active-site, or to a sufficiently low dsDNA K_d allowing saturated substrate binding by the mutant protomers. We directly interrogated these alternative hypotheses by quantifying the hT1, hT1-DN, and recombinant human $TREX1_{1\text{-}242}^{+/D18N}$ (hT1-DNhet) enzymes' exonuclease activities at sub-enzyme concentrations of the original plasmid substrate (Figure 6A) and at saturating

concentrations of a 30-bp dsDNA (Figure 6B). We observed significantly greater exonuclease activities for hT1-DNhet and a 1:1 hT1:hT1-DN mix on the oligo versus plasmid dsDNA (Figure 6C), consistent with the original data being (at least partially) attributable to substrate-mediated steric hinderance of dual active-site occupation and/or saturated binding of substrate. However, it is curious to note that the heterodimer was ~2-fold less active on the oligo dsDNA than the 1:1 hT1:hT1-DN mix, despite these reactions having the same total concentrations of wild-type and D18N protomers. Thus, our results are reconcilable with some degree of catalytic and/or substrate binding suppression via communication across TREX1's dimer interface.

Inter-Protomer Dynamics of the TREX1 Homodimer:

To interrogate the possibility dsDNA binding-mediated regulation across TREX1's dimer interface, we performed all-atoms MD simulations with mT1 apoenzyme and mT1 bound to dsDNA in a single active-site ('Protomer-A') to compare dynamics of the unbound protomer ('Protomer-B') between systems. RMSD comparisons of mT1 against initial conditions shows that all simulations quickly equilibrate without indications of catastrophic instability (Supplemental 1), and visualization of the conformational free energy maps for the mT1 backbone (Cα) in each of these systems reveals stark contrast in the distribution of structural states in each system's unbound protomer (Supplemental 5). An RMSF analysis revealed three distinct regions of the mT1 backbone that were rigidified by dsDNA occupation of the opposing protomer's active-site (Figure 7A). These include TREX1's flexible loop (aa. 164-175), whose corresponding region in the related TREX2 exonuclease has been previously implicated in DNA-binding[33], and two previously unstudied regions surrounding the active-site (aa. 23-27 & 75-88). Curiously, these are the same regions impacted by DNA occupation in the resident protomer (Figure 3A). A phi/psi angle clustering analysis to identify conformationally altered residues echoed these findings, with most identified residues being located on TREX1's flexible loop (Figure 7B).

As a more general interrogation of coordinated activity between TREX1 protomers, Cα correlation analyses were performed for mT1 and an mT1:4-mer complex (Figure 8). We found positively correlated dynamics between the protomers' α5 helices (Figure 8 – correlation 1) and between their β3 strands (Figure 8 – correlation 3), which prior reports suggest are key contributors to TREX1's dimeric stability[18,21,22], and these correlations were independent of 4-mer presence in the active-sites. Interestingly, we also found inter-protomer dynamics correlations that were dependent on the presence of 4-mer in the active-sites. These included positively correlated dynamics between TREX1's flexible loop and α6 helix (Figure 8 – correlation 2) and between the flexible loop and aa. 79-93 (α3-α4) (Figure 8 – correlation 4), as well as negatively correlated dynamics between aa. 79-93 (α3-α4) and TREX1's α5 helix (Figure 8 – correlation 2) and between aa. 79-93 (α3-α4) and aa. 22-31 (β1-α1-β2) (Figure 8 – correlation 6). Prior studies suggest TREX1's flexible loop is involved in DNA binding[19,33], but the other regions implicated in the DNA-dependent correlations have no previously ascribed function(s).

DISCUSSION

TREX1's Substrate Preferences:

Our thermal shift assays (Figure 2) and determinations of k_{cat} and K_m (Table 1) suggest that, despite similar catalytic rates for the two substrates, TREX1 has much greater affinity for dsDNA than ssDNA. The kinetic constant values derived under our robust model for TREX1 exonuclease activity (Table 2) also support this assertion, and place TREX1's dsDNA affinity at ~1000-fold higher than for ssDNA. It's notable that there is over an order in magnitude variation between the substrates' predicted on-rates, and that the predicted dsDNA on-rate, which is consistent with measurements of the classic kinetic constants (Table 1 and ref[10]), seems to strain theoretical diffusion limits[34]. This behavior may be explained if TREX1 is capable of 'scanning' nonspecifically along the DNA backbone to identify free 3'-hydroxyls, which is a proposed capability of the highly related TREX2 enzyme[33,35] and wasn't accounted for in our kinetic model of TREX1 exonuclease activity. If so, our use of a massive, nicked dsDNA species might have exaggerated TREX1's apparent on-rate via a predominantly 1-dimensional versus 3-dimensional search for free 3'-hydroxyls, where diffusion-limited association with the DNA backbone occurs well before 3'-hydroxyl binding and catalysis. Since we used a >300-fold shorter ssDNA species, its apparent on-rate would not be as inflated by this scanning behavior. Regardless, the dsDNA off-rate is still predicted at 20-100X lower than ssDNA, so a multiple order in

magnitude difference in affinity for ds- versus ssDNA is still expected for TREX1 when accounting for scanning-based bias.

Our MD simulations (Figure 3) suggest that this discrepancy in apparent binding affinity is largely attributable to tighter interactions with TREX1's flexible loop, as well as interactions with two uncharacterized regions of the TREX1 structure: its α1 (aa. 23-27) and α3 (aa. 75-88) helices. Implication of the flexible loop corresponds well with prior literature, which demonstrates that mutation of these residues disproportionately reduces TREX1 exonuclease activity on ds- versus ssDNA[19], and that the corresponding region in TREX2 contributes to DNA-binding[33]. Though this function hasn't been directly demonstrated for TREX1's flexible loop, the TREX1^{D18N} variant enzymes' ability to competitively inhibit wild-type enzyme is significantly reduced upon co-mutation of the flexible loop residues[19], suggesting it also serves a DNA-binding role in TREX1. It's fair to note, however, that our MD simulation substrates differed in length, not just strand number. Consequently, a length-dependent interaction of polynucleotides with TREX1's flexible loop could also produce these results, which would be consistent with flexible loop-mediated DNA backbone 'scanning' behavior for TREX1 (as proposed above). While the other two implicated TREX1 regions have not been functionally characterized in prior literature, their positions relative to TREX1's 3'-terminal nucleotide binding pocket indicate probable relevance to exonuclease activity (Figure 3A – orange and purple areas). Thus, they are prime targets for future structure-function studies of TREX1.

Overall, our interrogations of TREX1's ss- versus dsDNA binding provide insights into the likely nature of its biological substrate, and TREX1's role in degrading it. First, TREX1's dramatically higher affinity for ds- versus ssDNA (Tables 1-2 and Figure 2) suggest that TREX1 has evolved to degrade a dsDNA species *in vivo*. It is tempting to posit that TREX1's comparable catalytic rate on, and admittedly still high affinity for, ssDNA suggests it (at least also) functions to degrade ssDNA *in vivo*, which may well be true. However, these observations could also be biophysical consequences of how TREX1 degrades a single strand of dsDNA species. Indeed, in existing crystal structures it is evident that the 3'-terminal nucleotide, and the three subsequent nucleotides, of ss- versus dsDNA are positioned nearly identically for catalysis in the TREX1 active-site[22,28], which would explain why the two substrates display similar catalytic constants. Similarly, TREX1's sequence-independent contacts with these four 3'-terminal nucleotides and its flexible loop's proposed nonspecific interactions with the DNA backbone might necessarily contribute some affinity for ssDNA species. Thus, TREX1's relative affinity for ss- versus dsDNA seems the most appropriate biochemical metric for its substrate preferences. Next, the 'scanning' capabilities of TREX1 suggested by our kinetic model and prior TREX2 literature implicate a large biological substrate, where the benefits of a 1-dimensional search for free 3'-hydroxyls are necessary to degrade insulting nucleic acid at sufficient rates to avoid immune stimulation. Finally, and perplexingly, the processivity predicted by our kinetic model (Figure 5B) suggests that TREX1's biological function only requires it to resect short stretches of the identified DNA strand before dissociating. These biochemical properties of TREX1 seem reminiscent of proofreading exonucleases (ex. ref[36]), but a canonical proofreading function is not consistent with the pathology of *TREX1* mutants (reviewed in refs[2,3]). Finally, it's prudent to acknowledge that our biochemical substrates are unlikely to perfectly recapitulate TREX1's biological substrate(s), and thus the relevant *in vivo* kinetics could differ. Consequently, further interrogation is warranted to validate the biological significance of these biochemical properties.

TREX1's Obligate Homodimer Structure:

Early biochemical and structural studies of TREX1 established the interdependence of protomers in the TREX1 homodimer's exonuclease activity[21,25], and later studies with the TREX1^{D18N} dimeric variants[17,19] spawned the hypothesis that DNA binding in one TREX1 protomer leads to catalytic suppression of the opposing protomer. We recapitulated these initial D18N experiments and expanded them to alternative substrate conditions (Figure 6), and our results are inconsistent with total catalytic suppression of the opposing protomer upon DNA binding. However, they are suggestive of some degree of inter-protomer negative regulation of catalysis and/or binding. Our MD analyses interrogating the effects of dsDNA binding on opposing protomer dynamics (Figure 7) support this conclusion as well, and mechanistically implicate the same regions impacted by DNA binding in the resident protomer (Figure 3): TREX1's flexible loop (aa. 164-175), α1 helix (aa. 23-27), and α3 helix (aa. 75-88). MD correlation analyses (Figure 8) further support this hypothesis and implicate these regions, suggesting that TREX1's flexible loop communicates across the dimer interface with α3-α4 (aa. 79-93) and with α6, and that α3-α4 (aa. 79-93) communicates across the dimer interface with β1-α1-β2 (aa. 22-31). Collectively, these data are consistent with TREX1 possessing flexible loop-dependent negative cooperativity across its dimer interface. This is supported by analogous studies in TREX2 that

suggest its flexible loop is involved in binding cooperativity[33,35], and by the α6 and α3-α4 regions' direct contact with the 3'-terminal nucleotide in existing crystal structures[22,28]. It seems reasonable in TREX1's case that, sequentially, the α6 and/or α3-α4 regions are the 'sensors' of 3'-terminal nucleotide binding and the flexible loop is the 'effector' of negative cooperativity, as opposed to the opposite trend. However, this cannot be definitively stated from these data alone.

When paired with our substrate preference data, these findings support a model where TREX1's homodimer structure has evolved as part of a mechanism to efficiently locate free 3'-hydroxyls on a large, nicked dsDNA substrate and degrade the nicked strand. Mechanistically, we propose that TREX1 is capable of 'scanning' along the dsDNA backbone of its biological substrate in search of free 3'-hydroxyls, with each flexible loop positioned on the backbone of a respective strand. Upon one protomer's recognition of a 3'-terminal nucleotide in its binding pocket by its α6 and/or α3-α4 regions, the opposing protomer's flexible loop is forced to release the non-scissile strand. Then, the scissile strand can be positioned for efficient catalysis in the resident protomer's active-site. Under this model, TREX1's ultra-fast kinetics (Tables 1-2), its flexible loop's apparently disproportionate relevance to dsDNA exonuclease activity[19], and the D18N heterodimer's behavior (Figure 6) are quite expected. This model is, of course, inferential by nature, and warrants rigorous and highly mechanistic future studies to validate it.

Conclusions:

The 3' → 5' exonuclease, TREX1, functions *in vivo* to degrade DNA and prevent chronic immune activation through the cGAS-STING pathway. TREX1 is currently described as a nonprocessive enzyme with little preference for ss- versus dsDNA, ssDNA sources are proposed as relevant TREX1 *in vivo* substrates, and TREX1's obligate homodimer structure is largely unexplained. Our studies demonstrate that TREX1 is predicted to have several orders of magnitude greater affinity for ds- than ssDNA. Furthermore, we find extensively correlated dynamics between TREX1 protomers involving previously described as well as uncharacterized regions of the TREX1 enzyme, all of which have the potential to directly interact with substrate in their resident protomers. We infer that TREX1's obligate homodimer structure is an evolutionary mechanism to efficiently scan for exposed 3'-hydroxyls and facilitate rapid catabolism of a dsDNA biological substrate. Ultimately, these studies identify potentially critical regions of the TREX1 enzyme for future functional characterization, and they provide compelling evidence that the nature of TREX1's biological substrate should be critically reevaluated.

METHODS

Overexpression and Purification of TREX1 Enzymes:

We have published a very detailed protocol for purifying TREX1 enzymes[26], but we will summarize it here. Our experiments used a truncated form (aa. 1-242) of the human and murine TREX1 enzymes which contain only the catalytic domain, referred to in this work as 'hT1' and 'mT1', respectively. Wild-type protomers were expressed as a fusion protein using a pLM303x vector that encodes maltose-binding protein (MBP) linked N-terminally to TREX1. Mutant protomers were expressed as a fusion protein using a pCDF-Duet vector that encodes His-tagged NusA linked N-terminally to TREX1. Both constructs employed a linker with a rhinovirus 3C protease (PreScission Protease) recognition site and were transformed into *Escherichia coli* Rosetta2(DE3) cells (Novagen) for overexpression. Homodimers were purified via amylose column chromatography followed by overnight protease cleavage, then phosphocellulose (p-cell) column chromatography. Heterodimers were purified via amylose column chromatography followed by nickel column chromatography, then overnight protease cleavage followed by phosphocellulose (p-cell) column chromatography. Amylose and nickel columns used one-step elution, and p-cell columns used a salt gradient for elution. All enzymes eluted from p-cell columns at a salt concentration of ~120 mM. Enzyme concentrations of preps were determined by A_{280} on a NanoDrop 2000 (Thermo Fisher) spectrophotometer and validated by SDS-PAGE.

Approximation of TREX1 ssDNA k_{cat} and K_m:

Beginning from classic equations for the initial velocity of uninhibited enzyme[30], we derive Equation 1 where v_0, k_{cat}, $[E_T]$, K_m, and $[S]$ are initial velocity, catalytic rate constant, enzyme concentration, Michaelis-Menten constant, and substrate concentration, respectively.

Equation 1: $$\frac{v_0}{[E_T]} = k_{cat}\frac{[S]}{[S]+K_m}$$

Deriving from classic[30] equations we obtain Equations 2-3, where k_1 represents the association rate of enzyme and substrate (Equation 4.1), k_{-1} represents the dissociation rate of enzyme-substrate complex (Equation 4.2), and k_{cat} represents the catalytic rate constant for conversion of enzyme-substrate to enzyme and product (Equation 4.3).

Equation 2: $$K_m = \frac{k_{-1} + k_{cat}}{k_1}$$

Equation 3: $$K_d = \frac{k_{-1}}{k_1}$$

Equation 4.1: $$[E]\,[S] \xrightarrow{k_1} [ES]$$

Equation 4.2: $$[ES] \xrightarrow{k_{-1}} [E]\,[S]$$

Equation 4.3: $$[ES] \xrightarrow{k_{cat}} [E]\,[P]$$

Beginning from Equation 1, we rearrange to solve for Equations 5 and 6.

Equation 5: $$K_{cat} = \frac{v_0\,([S]+K_m)}{[E_T]\,[S]}$$

Equation 6: $$K_m = \frac{K_{cat}\,[E_T]\,[S]}{v_0} - [S]$$

If the catalytic rate, $v_o / [E_T]$, of the enzyme is determined at two substrate concentrations, then two iterations of Equations 5 and 6 are obtained. Setting these iterations equal to one another and solving for K_m and k_{cat} gives Equations 7 and 8, respectively. We use $[S]_1$ and $[S]_2$ to differentiate the two substrate concentrations, and β_1 and β_2 to indicate the catalytic rates ($v_o / [E_T]$) of the enzyme at those respective substrate concentrations.

Equation 7: $$K_m = \frac{[S]_1\,[S]_2\,(\beta_1 - \beta_2)}{\beta_2\,[S]_1 - \beta_1\,[S]_2}$$

Equation 8: $$k_{cat} = \frac{\beta_1\,\beta_2\,([S]_1 - [S]_2)}{\beta_2\,[S]_1 - \beta_1\,[S]_2}$$

For this study's ssDNA constants in Table 1, reported mean and standard deviation were approximated via Monte-Carlo simulation in R v3.6.1[37]. Small sampling sizes were used and were compensated for via iterative simulation to improve accuracy. We generated vectors of 10^2 values each for the β_1 and β_2 variables. Values were generated by uncorrelated sampling from a normal distribution defined by the wild-type means and standard deviations provided in Figure 1D. We used the substrate concentrations outlined in the 'Polyacrylamide Gel ssDNA Assay' Methods to define the $[S]_1$ and $[S]_2$ parameters. These vectors and parameters were applied pairwise to Equations 7-8 to generate equal length vectors for k_{cat} and K_m. The k_{cat} vector was then subjected to pairwise division by the K_m vector to generate an equal length vector for k_{cat}/K_m. Means and standard deviations of the k_{cat}, K_m, and k_{cat}/K_m vectors were then calculated. For the k_{cat}/K_m values of previous studies, the k_{cat} and K_m vectors used to calculate the k_{cat}/K_m vector were generated by uncorrelated sampling from a normal distribution defined by the previously reported means and standard deviations. For each enzyme we performed 10^5 replicate simulations and reported the median approximations for means and standard deviations of k_{cat}, K_m, and k_{cat}/K_m.

Absorbance-Based dsDNA Exonuclease Assay:

To generate a standard curve, synthetic dAMP nucleotides were serially diluted from 10 mM to 5 μM in 2-fold increments, then all dilutions' absorbances at 260 nm measured on a NanoDrop 2000 (Thermo Fisher) spectrophotometer. The plot of A_{260} vs [dAMP] was fit with linear regression in Prism 7.0 (GraphPad). The

linear fit equation was rearranged and adjusted to the reported[38] extinction coefficient for dAMP to produce Equation 9. In this equation A_{260} indicates the absorbance at 260 nm, and ε_{dNMP} indicates the average extinction coefficient of the nucleotides. The extinction coefficient parameter was approximated as the average of all reported[38] DNA nucleotides' extinction coefficients, 1160 $M^{-1}cm^{-1}$.

Equation 9: $$[\mathrm{dNMP}] = \frac{1.65*\mathrm{A}_{260}+7.19*10^{-3}}{\varepsilon_{\mathrm{dNMP}}}$$

For the activity assays, 50 µL reactions were prepared containing 30 nM hT1, 20 mM TRIS (pH 7.5 @ 25°C), 5 mM $MgCl_2$, 2 mM dithiothreitol (DTT), a variable concentration of dsDNA substrate, and 100 ng/µL bovine serum albumin (BSA). The dsDNA substrate was a ~10-kb Nt.BbvCI-nicked (NEB #R0632) pMYC plasmid used at 2.5, 5, 10, 20, 40, and 75 nM. Reactions were initiated by addition of 10X enzyme diluted in 1 mg/mL BSA and incubated for 10 min at room temperature (~21ºC). Reactions were quenched by the addition of 2.5 µL of 0.5 M EDTA. Quenched reactions were filtered with 10-kDa molecular weight cutoff microcentrifugal filter units according to manufacturer specifications (Amicon Ultra #UFC5010BK), then each flow-through had its absorbance measured at 260 nm on a NanoDrop 2000 (Thermo Fisher) spectrophotometer. Absorbance for each reaction was converted into concentration of excised nucleotides via Equation 9, then used to calculate activity rate as moles of nucleotides excised per second. Plots of activity rate versus substrate concentration were fit with nonlinear regression in Prism 7.0 (GraphPad) consistent with Equation 1 to solve for k_{cat} $[E_T]$ and K_m.

Polyacrylamide Gel ssDNA Assay:

We have published a quite detailed protocol[26] for this assay, but we will summarize it here. 100 µL reactions were prepared containing a variable concentration of hT1, 20 mM TRIS (pH 7.5 @ 25°C), 5 mM $MgCl_2$, 2 mM dithiothreitol (DTT), 15 nM or 515 nM ssDNA substrate, and 100 ng/µL bovine serum albumin (BSA). Reactions were initiated by addition of 10X enzyme diluted in 1 mg/mL BSA and incubated for 20 minutes at room temperature (~21ºC). Reactions were quenched in 400 µL cold ethanol, dried *in vacuo*, resuspended in loading solution, then electrophoresed on a 23% urea-polyacrylamide sequencing gel and imaged on a Typhoon FLA 9500 (GE Healthcare Life Sciences).

For the initial experiment to approximate TREX1's ssDNA k_{cat} and K_m, the substrate was a 5'-FAM-labeled non-annealing 30-mer oligo of the below sequence, used at the indicated concentrations with the indicated concentrations of hT1. For the experiments to optimize the original kinetic model we proposed, the substrate was the same provided below used at 15 nM, and the enzyme was 10 pM of hT1. The optimization experiment included 6 independent reactions. In addition, a single reaction with an increased concentration of wild-type enzyme was used to generate a full ladder of oligo sizes for reference during quantification.

5'-FAM-labeled 30-mer oligo sequence: 5'- [FAM] ATACGACGGT GACAGTGTTG TCAGACAGGT -3'

Quantification of Polyacrylamide Gels:

Polyacrylamide gel images were subjected to 1D densitometry using the GelAnalyzer 2010a software. Lanes were identified automatically with software defaults, and intensity vs migration functions for each lane were subjected to background subtraction by the rolling ball method with a '500' ball radius setting. Next, the peaks in each lane corresponding to various oligo sizes were manually identified using the aforementioned internal ladder as a visual reference for the expected migration positions. Then, peaks were integrated automatically to quantify relative oligo levels for all bands. These band values were subjected to quantification methods[26] we have previously published to calculate the proportion of each substrate size remaining in the reaction, calculate the moles of nucleotides excised, and ultimately determine reaction rates.

Thermal Shift Assays:

For the assay, 100 µL reactions were prepared containing 0.5 mg/mL TREX1, 20 mM TRIS (pH 7.5 @ 25°C), 5 mM $CaCl_2$, 2 mM dithiothreitol (DTT), 5X SYPRO Orange dye (ThermoFisher #S6650), and a variable concentration of substrate. No BSA was included. The dsDNA substrate was a self-annealing 30-mer oligonucleotide, and the ssDNA substrate was a 30-mer oligonucleotide. Both substrates were included at 18 µM in the relevant reactions. 8 replicate reactions per condition/enzyme were created, reconciled into a clear

96-well PCR plate (VWR #76402-848), and sealed with adhesive plate tape. Melting curve experiments were then performed with a 7500 Real-Time PCR System (Applied Biosystems). Melting experiment was conducted over a 25-95°C temperature range, at a rate of 1°C/min. Fluorescence was monitored with the system's TAMRA reporter settings, with 188 fluorescence readings per sample in equal ΔT intervals across the temperature gradient.

Melting curve data was imported to R, then each sample's curve normalized internally to maximum and minimum fluorescence. Curves were then fit with a smoothing spline using the 'smooth.spline()' function. Fit curves were then used to interpolate T_m from the melting curves' inflections points, via the 'deriv()' function. All calculations described in this section were performed in R v3.6.1[37].

30-mer oligo sequence (dsDNA): 5'- GCTCGAGTCA TGACGCGTCA TGACTCGAGC -3'
30-mer oligo sequence (ssDNA): 5'- TTAACCTTCT TTATAGCCTT TGAACAAAGG -3'

Molecular Dynamics Simulations:

The published structure of mTREX1$_{1-242}$ bound to ssDNA (PDB = '2IOC') was modified in PyMOL[39] to remove nucleic acid and ions. The modified structure was then used as a template for MODELLER[40] to model in the flexible loops missing from the original structure. This produced the mT1 model. The original mTREX1 published structure was loaded into PyMOL[39] and aligned to the mT1 model. The calcium ions and ssDNA polymers from the original structure were then superimposed on the mT1 model, and the calcium ions converted to magnesium, to generate the mT1:4-mer model. The published structure of mTREX1$_{1-242}$ bound to dsDNA (PDB = '4YNQ') was loaded into PyMOL[39] and aligned to the mT1:4-mer model. The ssDNA polymers were removed from the mT1:4-mer model, a single dsDNA from the published structure copied into the model, the published structure removed, and then the remaining molecules exported. This generated the mT1:dsDNA model.

Each of the models was converted to pdb and psf files with the 'autopsf' functionality in VMD (v1.9.4)[41], using the CHARMM 'top_all27_prot_lipid_na.inp' force field parameters[42,43]. Explicit water solvent was added to the processed models using a 10 Å padding around the existing models' dimensions. Hydrated models were then ionized with 50 mM $MgCl_2$ and exported. The exported pdb and psf files defined the initial conditions for the four simulation systems, constructed from the four models. Hydration and ionization were also performed in VMD (v1.9.4) using default CHARMM parameter files.

Initial condition files for each system were replicated and used to initiate four MD simulations per system. MD simulations were performed with ACEMD3 (Acellera)[44] on Acellera GPUs. All initial condition files were independently subjected to a 1000-step energy minimization, then simulations were initiated. Simulations were for 1000 ns each, with 4 fs steps. Structure coordinates were exported from simulations every 10 ps in xtc format. Thermostat was set to 300 K, barostat was set to 1.0132 bar, and all other parameters were left to default. Simulations ran for ~10 days each.

Completed simulations were pre-processed in VMD v1.9.4. The xtc trajectory file and psf initial conditions file were loaded into the pdb initial conditions file. Simulation trajectory coordinates were loaded in 20-frame (200 ps) intervals totaling 5000 frames. After loading, simulations were unwrapped using the 'pbc unwrap' function. Unwrapped simulation frames were aligned to the initial conditions frame based on RMSD-minimization between the protein backbones. Pre-processed simulations were exported without solvent molecules to mol2 files. The mol2 files were processed into csv files compatible with later analyses using Bash scripts.

Calculating Phi and Psi Angles of TREX1 Residues:

The python package MDAnalysis was used to calculate phi and psi angles for the TREX1 residues. For each simulation of each system, the respective pre-processed mol2 file was analyzed with the 'Dihedral' and 'Ramachandran' functions to generate phi and psi angles for every time-point (frame) of the simulations. Phi and psi values were then exported as separate csv files compatible with subsequent analyses. Python script for phi/psi angle calculations was run with python v3.7.4.

Calculating RMSD:

The RMSD values for each simulation were calculated via Equation 10. Where: $\boldsymbol{r}_j(t_i)$ is the position vector of the jth atom in the ith time frame and N is the total number of atoms.

Equation 10: $$\mathrm{RMSD_i} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\mathbf{r}_j(t_i) - \mathbf{r}_j(t_1)\right)^2}$$

These calculations provided RMSD values at every time frame for all four simulations per system. The values for all simulations were averaged by time frame to give mean values. All calculations described in this section were performed in R v3.6.1[37].

Calculating RMSF:

The RMSF values in each simulation were calculated via Equation 11. Where: where $\boldsymbol{r}_i(t_j)$ is the instantaneous position vector of the ith atom, $\boldsymbol{r'}_i$ is the mean position of the ith atom, and T is the total number of time frames.

Equation 11: $$\mathrm{RMSF_i} = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\mathbf{r}_i(t_j) - \mathbf{r}_i'\right)^2}$$

The RMSF values for all simulations of a system were grouped by atom to give RMSF means and standard deviations. All calculations described in this section were performed in R v3.6.1[37].

Clustering by Residues' Phi and Psi Angles:

The phi and psi values were first converted into Cartesian coordinates via projection onto a sphere with Equations 12-14. This allowed seamless clustering for residue conformations that had phi (Ψ) or psi (Φ) values spanning the -180º/180º transition.

Equation 12: $$x = \cos(\Psi)\sin(\Phi)$$
Equation 13: $$y = \sin(\Psi)\sin(\Phi)$$
Equation 14: $$z = \cos(\Phi)$$

For each residue, every third time frame of every simulation of every system was clustered by these projected coordinates using the 'dbscan::hdbscan()' package::function[45] in R with a 'minPts=200' parameter to attribute any clusters containing <1% of time frames to noise. For every residue, the proportion of each system's time frames belonging to the identified clusters/conformations were calculated and pairwise comparisons of the proportions between systems were made via a chi-squared test of independence, using the 'chisq.test()' function in R. P-values from these comparisons were converted to q-values by adjusting for false discovery rate with the Benjamini-Hochberg[46] method via the 'p.adjust()' function in R. Residues with significantly different conformation proportions between systems were then identified. All calculations and statistics for this section were performed in R v3.6.1[37].

Conformational Free Energy Maps:

The coordinates of every relevant atom were used as variables in a principal components analysis (PCA) on every time frame of every simulation of every relevant system. The PCA was performed via the 'princomp()' function in R. Each resulting principal component's (PC) proportion of variance (POV) was calculated via Equation 15, where σ_i^2 is the variance for the ith principal component and N is the number of principal components.

Equation 15: $$\mathrm{PV_i} = \frac{\sigma_i^2}{\sum_{j=1}^{N}\sigma_j^2}$$

Values for the first two PCs were extracted and separated by system. Density functions for each system were then approximated using PC-1 and PC-2 with the 'MASS::kde2d()' package::function[47] in R.

Density functions for each system were visualized as contour plots using the 'contour()' function in R. All calculations and statistics for this section were performed in R v3.6.1[37].

Kinetic Modeling:

For the proposed models in Figure 4, and all other equations and figures in this section unless otherwise stated, $k_{\square}$ are respective rate constants in the model, Ψ is the processivity parameter ranging from [0,1] for nonexistent to complete processivity, *[E]* is concentration of free enzyme, $[S_n]$ is concentration of free substrate with *n* excisable nucleotides, *[P]* is concentration of free product (excised dNMPs), and $[\square\square]$ is concentration of a complex of the respective reactants.

For Equations 16.1-5, $S_{\beta:n}$ is all substrates with between β and *n* excisable nucleotides, β is the maximum number of excisable nucleotides on a substrate molecule during the reaction, $[\square]_t$ is concentration of the respective reactant or complex at time *t*, $[\square]_t'$ is rate of change in concentration of the respective reactant or complex at time *t*, and all other previous notation applies. Reference to Figure 4A provides helpful context.

Equation 16.1: $$[\mathrm{E}]_t' = \mathrm{k}_{-3}[\mathrm{EP}]_t + \mathrm{k}_{-1}[\mathrm{ES}_{\beta:1}]_t + \mathrm{k}_2(1-\Psi)\left[\mathrm{ES}_{\beta:2}\right]_t + \mathrm{k}_2[\mathrm{ES}_1]_t - \mathrm{k}_3[\mathrm{E}]_t[\mathrm{P}]_t - \mathrm{k}_1[\mathrm{E}]_t\left[\mathrm{S}_{\beta:1}\right]_t$$

Equation 16.2: $$[\mathrm{P}]_t' = \mathrm{k}_{-3}[\mathrm{EP}]_t + \mathrm{k}_2[\mathrm{ES}_{\beta:1}]_t - \mathrm{k}_3[\mathrm{E}]_t[\mathrm{P}]_t$$

Equation 16.3: $$[\mathrm{S_n}]_t' = \mathrm{k}_{-1}[\mathrm{ES_n}]_t + \mathrm{k}_2(1-\Psi)[\mathrm{ES_{n+1}}]_t - \mathrm{k}_1[\mathrm{E}]_t[\mathrm{S_n}]_t$$

Equation 16.4: $$[\mathrm{ES_n}]_t' = \mathrm{k}_1[\mathrm{E}]_t[\mathrm{S_n}]_t + \mathrm{k}_2\Psi[\mathrm{ES_{n+1}}]_t - (\mathrm{k}_{-1}+\mathrm{k}_2)[\mathrm{ES_n}]_t$$

Equation 16.5: $$[\mathrm{EP}]_t' = \mathrm{k}_3[\mathrm{E}]_t[\mathrm{P}]_t - \mathrm{k}_{-3}[\mathrm{EP}]_t$$

For Equations 17.1-8, *[E]* and *[M]* are concentrations of free wild-type and mutant enzyme, respectively, Ψ and Φ are the wild-type and mutant processivity parameters, respectively, and all other notation is previously established. Reference to Figure 4B provides useful context.

Equation 17.1: $$[\mathrm{E}]_t' = \mathrm{k}_{-3}[\mathrm{EP}]_t + \mathrm{k}_{-1}[\mathrm{ES}_{\beta:1}]_t + \mathrm{k}_2(1-\Psi)\left[\mathrm{ES}_{\beta:2}\right]_t + \mathrm{k}_2[\mathrm{ES}_1]_t - \mathrm{k}_3[\mathrm{E}]_t[\mathrm{P}]_t - \mathrm{k}_1[\mathrm{E}]_t\left[\mathrm{S}_{\beta:1}\right]_t$$

Equation 17.2: $$[\mathrm{M}]_t' = \mathrm{k}_{-c}[\mathrm{MP}]_t + \mathrm{k}_{-a}[\mathrm{MS}_{\beta:1}]_t + \mathrm{k}_b(1-\Phi)\left[\mathrm{MS}_{\beta:2}\right]_t + \mathrm{k}_b[\mathrm{MS}_1]_t - \mathrm{k}_c[\mathrm{M}]_t[\mathrm{P}]_t - \mathrm{k}_a[\mathrm{M}]_t\left[\mathrm{S}_{\beta:1}\right]_t$$

Equation 17.3: $$[\mathrm{P}]_t' = \mathrm{k}_{-3}[\mathrm{EP}]_t + \mathrm{k}_{-c}[\mathrm{MP}]_t + \mathrm{k}_2[\mathrm{ES}_{\beta:1}]_t + \mathrm{k}_b\left[\mathrm{MS}_{\beta:1}\right]_t - \mathrm{k}_3[\mathrm{E}]_t[\mathrm{P}]_t - \mathrm{k}_c[\mathrm{M}]_t[\mathrm{P}]_t$$

Equation 17.4: $$[\mathrm{S_n}]_t' = \mathrm{k}_{-1}[\mathrm{ES_n}]_t + \mathrm{k}_{-a}[\mathrm{MS_n}]_t + \mathrm{k}_2(1-\Psi)[\mathrm{ES_{n+1}}]_t + \mathrm{k}_b(1-\Phi)[\mathrm{MS_{n+1}}]_t - \mathrm{k}_1[\mathrm{E}]_t[\mathrm{S_n}]_t - \mathrm{k}_a[\mathrm{M}]_t[\mathrm{S_n}]_t$$

Equation 17.5: $$[\mathrm{ES_n}]_t' = \mathrm{k}_1[\mathrm{E}]_t[\mathrm{S_n}]_t + \mathrm{k}_2\Psi[\mathrm{ES_{n+1}}]_t - (\mathrm{k}_{-1}+\mathrm{k}_2)[\mathrm{ES_n}]_t$$

Equation 17.6: $$[\mathrm{MS_n}]_t' = \mathrm{k}_a[\mathrm{M}]_t[\mathrm{S_n}]_t + \mathrm{k}_b\Phi[\mathrm{MS_{n+1}}]_t - (\mathrm{k}_{-a}+\mathrm{k}_b)[\mathrm{MS_n}]_t$$

Equation 17.7: $$[\mathrm{EP}]_t' = \mathrm{k}_3[\mathrm{E}]_t[\mathrm{P}]_t - \mathrm{k}_{-3}[\mathrm{EP}]_t$$

Equation 17.8: $$[\mathrm{MP}]_t' = \mathrm{k}_c[\mathrm{M}]_t[\mathrm{P}]_t - \mathrm{k}_{-c}[\mathrm{MP}]_t$$

For Equation 18, $P_\Psi(N)$ is the probability of *N* nucleotides being excised before dissociation with a processivity factor of Ψ, and all other parameters are as defined above.

Equation 18: $$\mathrm{P}_\Psi(\mathrm{N}) = \begin{cases} \frac{\mathrm{k}_{-1}}{\mathrm{k}_{-1}+\mathrm{k}_2}, & \mathrm{N}=0 \\ \frac{\mathrm{k}_2(1-\Psi)(\mathrm{k}_2\Psi)^{\mathrm{N}-1}}{(\mathrm{k}_{-1}+\mathrm{k}_2)^{\mathrm{N}}} + \frac{\mathrm{k}_{-1}(\mathrm{k}_2\Psi)^{\mathrm{N}}}{(\mathrm{k}_{-1}+\mathrm{k}_2)^{\mathrm{N}+1}}, & \mathrm{N}>0 \end{cases}$$

*Assuming $0^0 = 1$, as opposed to being undefined.

For Equation 19, all notation is classic[30] or previously defined.

Equation 19: $$\mathrm{k}_{-1} = \mathrm{k}_1\mathrm{K}_\mathrm{d}$$

We provide detailed accounts of the parameter definitions for every simulated reaction (Supplemental 6), which should allow reproduction of these simulations. Any parameters not provided in the tables were set to '0' under all conditions. Every reaction proposed by the kinetic model was simulated with the R script provided on GitHub.

Fluorescence-Based dsDNA Exonuclease Assay:

We have published a detailed protocol[26] for this assay, but we will summarize it here. 150 µL reactions were prepared containing a variable concentration of hT1, 20 mM TRIS (pH 7.5 @ 25°C), 5 mM $MgCl_2$, 2 mM dithiothreitol (DTT), 5-10 ng/µL dsDNA substrate, and 200 ng/µL bovine serum albumin (BSA). For the '[S]<K_M' experiments, the dsDNA substrate was a ~10-kb Nt.BbvCI-nicked (NEB #R0632) pMYC plasmid used at 5 ng/µL (~0.83 nM), and the hT1 concentration was 15 nM. For the '[S]>>K_M' experiments, the dsDNA substrate was a self-annealing 30-mer oligo used at 1000 nM (~9 ng/µL), and the hT1 concentration was 2.5 nM. Reactions were initiated by addition of 10X enzyme diluted in 1 mg/mL BSA and incubated for 1 hour at room temperature (~21ºC). 20 µL samples were taken at 0, 5, 10, 20, 30, 45, and 60-minute time points and quenched in 20 µL of 15X SYBR Green dye (ThermoFisher #S7563). Quenched samples had their fluorescence measured on a PolarStar Omega microplate reader (BMG LabTech), using excitation/emission of 485/520 nm. For the 1:1:0 enzyme mixes, mix ratios indicate the prevalence of $hT1^{+/+}$, $hT1^{MUT/MUT}$, and $hT1^{MUT/+}$, respectively, where total hT1 concentrations were still as indicated.

30-mer oligo sequence: 5'- GCTCGAGTCA TGACGCGTCA TGACTCGAGC -3'

Quantification of Fluorescence-Based Data:

Each independent enzyme dilution in each experiment was used to create 6 replicate reactions. Plots of fluorescence vs time were generated from the time-course data for every reaction in each experiment. The data points from every reaction replicate were used for fitting with one-phase decay nonlinear regression. This gave 4 fitted datasets per experiment, one for each independent enzyme dilution. The slope of each fit at 0 min was interpolated to give initial velocities, and then initial velocities were normalized within each experiment to the mean of the wild-type fits' initial velocities to calculate relative activities for each variant/mix. Calculations were performed in Excel, and regression and interpolation were performed in Prism 9.0 (GraphPad).

Cα Correlation Analyses:

The Cα coordinates across all time frames and simulations were reconciled for each system, and Equations 20.1-2 were used to calculate the correlation factors between atom pairs. Here, C_{ij} is the correlation factor for atoms *i* and *j*, *N* is the number of frames, $\vec{r_i}^{\beta}$ ($\vec{r_j}^{\beta}$) is the position of atom *i* (*j*) in time frame *β*, and $\langle\vec{r_i}\rangle$ ($\langle\vec{r_j}\rangle$) is the mean position or centroid of atom *i* (*j*) across all time frames. All calculations described in this section were performed in R v3.6.1[37].

Equation 20.1: $$\widetilde{C_{ij}} = \sum_{\beta=1}^{N} \frac{(\vec{r_i}^{\beta} - \langle\vec{r_i}\rangle)\cdot(\vec{r_j}^{\beta} - \langle\vec{r_j}\rangle)}{N}$$

Equation 20.2: $$C_{ij} = \frac{\widetilde{C_{ij}}}{\sqrt{\widetilde{C_{ii}}\widetilde{C_{jj}}}}$$

Availability of Data and Scripts Used for Computational Analyses:

Scripts used for analyses described in these methods are available on GitHub (https://github.com/whemphil/TREX1_Modeling_Scripts.git), pre-processed data input files for MD analysis scripts have been deposited to Zenodo (doi.org/10.5281/zenodo.6347909), and raw files from MD simulations are available from W.O.H. upon request (total data ≈ 0.5 TB).

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Allergy and Infectious Diseases (AI116725 to F.W.P.), National Institute of General Medical Sciences (GM110734 to F.W.P.), Lupus Research Alliance (PI – F.W.P.), Wake Forest Innovations (PI – F.W.P.), Comprehensive Cancer Center of Wake Forest University National Cancer Institute (Center Support Grant, P30CA012197), NIH NRSA Predoctoral Fellowships (T32-GM095440 & T32-AI007401 to W.O.H.), a Sandy Lee Cowgill Memorial Scholarship (Trainee – W.O.H.), an Artom Fellowship (Trainee – W.O.H.), and a Scott Family Fellowship (PI – F.R.S.). W.O.H. has been partially

supported by the Howard Hughes Medical Institute during manuscript writing (PI – Thomas Cech). We'd also like to thank the WFU high-performance computing team for access to the DEAC cluster and technical support.

AUTHOR CONTRIBUTIONS

W.O.H. implemented all described methodologies, studies, and analyses, interpreted data, and wrote the manuscript. T.H. contributed to manuscript revision. F.R.S. supervised all computational studies and contributed to manuscript revision. F.W.P. supervised the project and contributed to manuscript revision.

COMPETING INTERESTS STATEMENT

The authors have no competing interests to declare.

TREX1 Classic Kinetic Constants

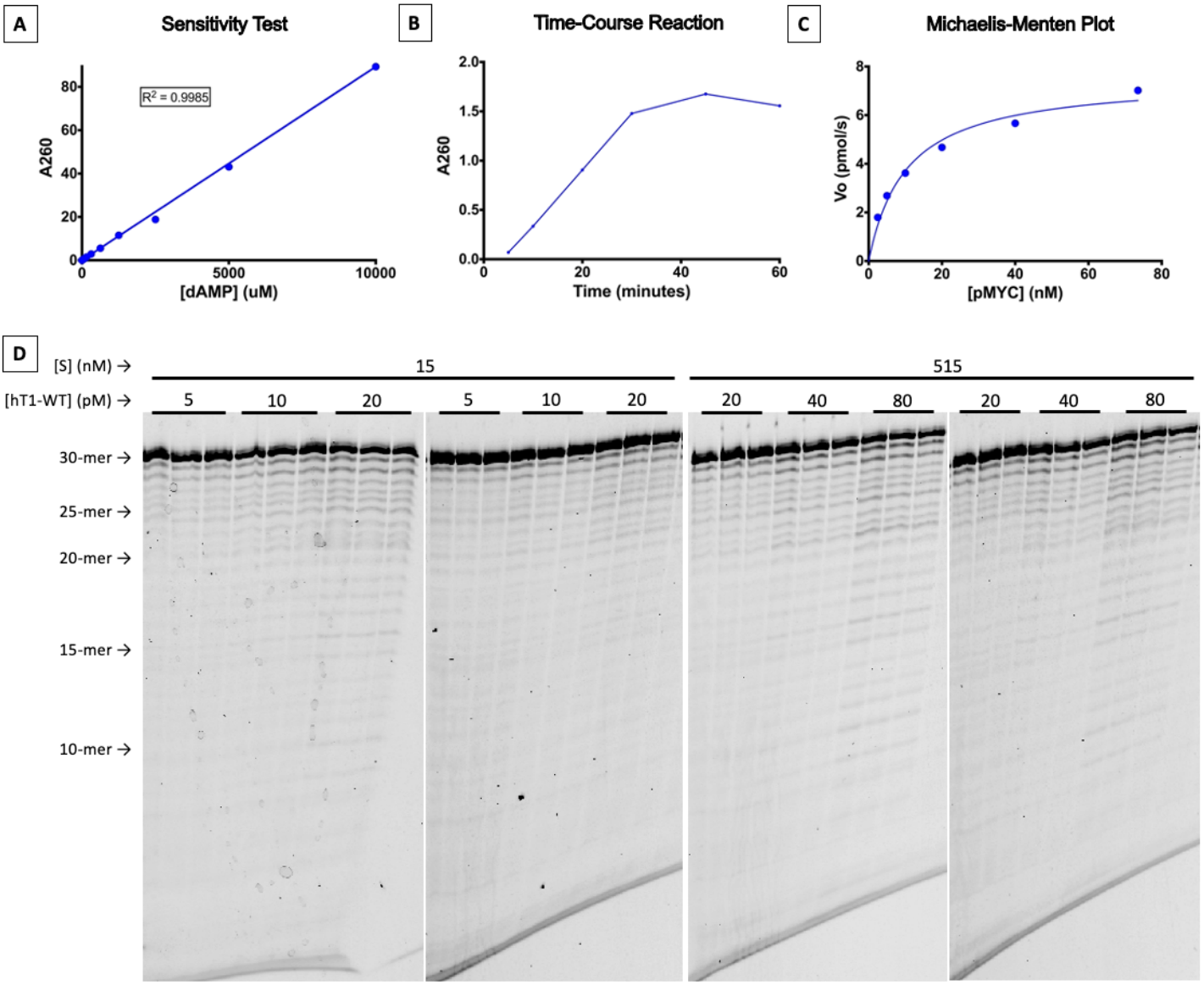
	ssDNA			dsDNA		
	k_{cat} (s^{-1})	K_m (nM)	k_{cat} / K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (nM)	k_{cat} / K_m ($M^{-1}s^{-1}$)
This Study	16 ± 3.2	120 ± 57	1.5 ± 0.40 ($x10^8$)	†10 ± N/A	10 ± N/A	1.0 ($x10^9$)
Reported	22 ± 1.1	19 ± 4.0	1.2 ($x10^9$)	11 ± 0.77	15 ± 4.4	7.3 ($x10^8$)

<u>Table 1</u>: **Classic Kinetic Constants for TREX1.** 'Reported' values are taken from prior work[10]. dsDNA values from this study were calculated empirically using the described absorbance-based exonuclease assay. ssDNA values from this study were approximated via Monte-Carlo simulation using the ssDNA activity rates of wild-type enzyme that were calculated in Figure 1D. Values shown indicate $\mu \pm \sigma$, except for this study's dsDNA values and the reported specific activities. 'N/A' = not applicable. [†]Extrapolation from other data; regression directly indicates $k_{cat} \approx 5.0\ s^{-1}$. Calculations and simulations were performed in R v3.6.1, and table was prepared in Word (Microsoft).

Approximated hTREX1 Kinetic Constants

	k_{-1} (s^{-1})	k_1 ($M^{-1}s^{-1}$)	k_2 (s^{-1})	k_{-3} (s^{-1})	k_3 ($M^{-1}s^{-1}$)	Ψ
ssDNA	2.2	5.2 ($x10^7$)	†16	n.d.	n.d.	0.75
dsDNA	~ [0.10,0.050,0.020]	~1.0 ($x10^9$)	†10			~ [0.75,0.80,0.85]

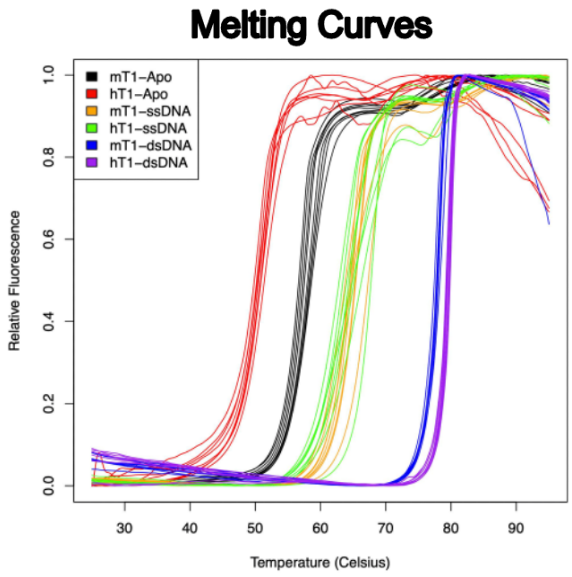
<u>Table 2</u>: **Kinetic Constant Approximations for TREX1 Exonuclease Activity Model.** A table of the values for relevant kinetic constants. Bracketed values are approximation pairs, where the first values in each bracketed vector are true under the same assumptions as one another. Other values are broadly applicable approximations. 'n.d.' is not determined. †Values taken from Table 1, which were constrained in the modeling. Calculations were performed in R v3.6.1, and table was prepared in Word (Microsoft).



$\frac{V_0}{[E_T]}$ = 1.9 ± 0.45 (s^{-1})

$\frac{V_0}{[E_T]}$ = 13 ± 1.8 (s^{-1})

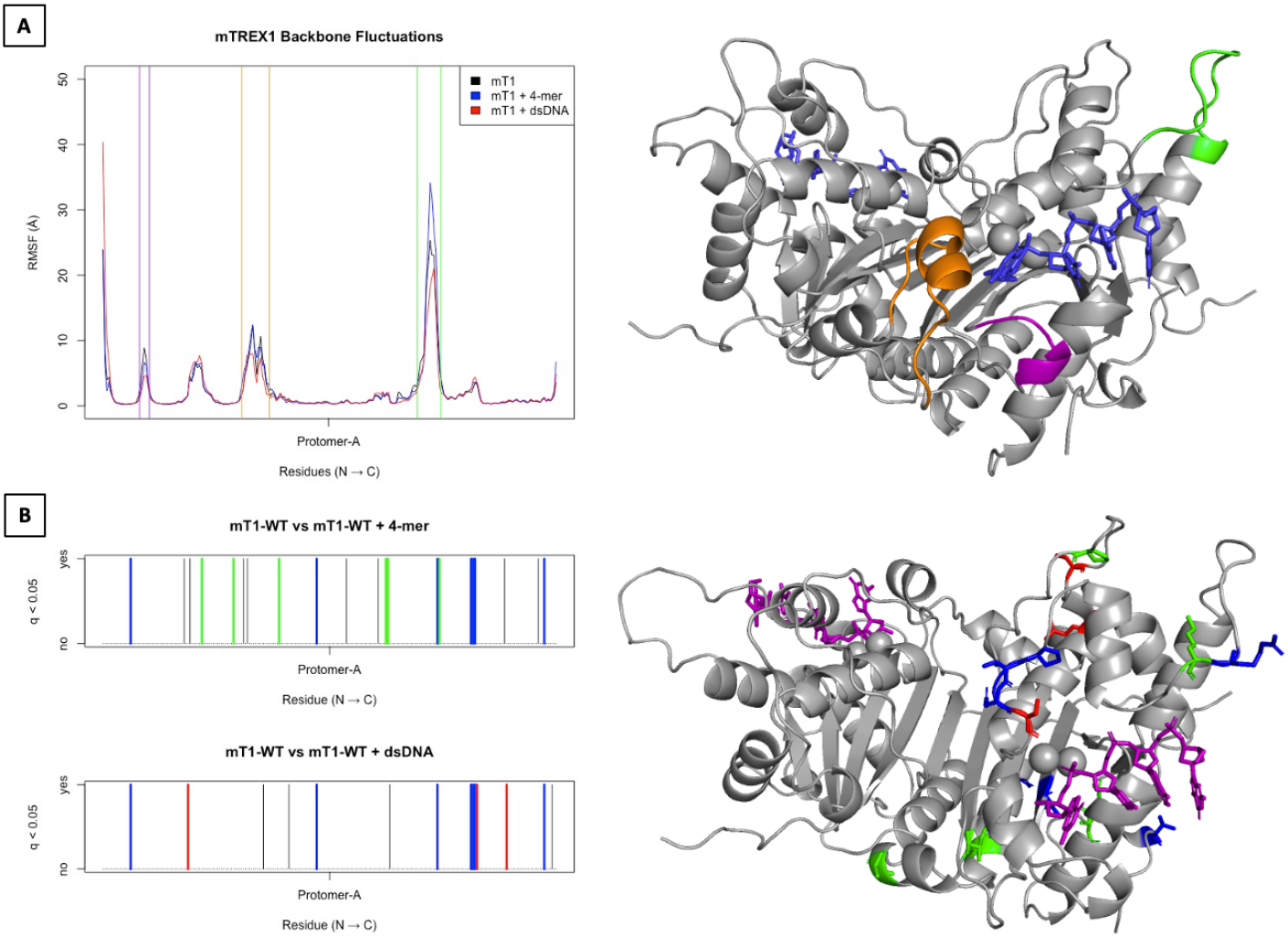
<u>Figure 1</u>: **Determination of TREX1 dsDNA k_{cat} and K_m. [A]** *Test of Absorbance Assay Sensitivity*. A_{260} was measured via NanoDrop for dNMP concentrations from 0.15 μM – 10 mM. Lower limit of nucleotide detection was ~5 μM, and the assay was linear to $A_{260} \geq 89$. **[B]** *Test Reaction with Absorbance Assay*. Time-course reaction was performed as described with standard reaction components, 30 nM hT1, and 60 ng/μL nicked-plasmid. **[C]** *Determination of TREX1 dsDNA Kinetic Constants*. Reactions containing standard components, 30 nM hT1, and the indicated concentrations of nicked dsDNA were incubated 10 minutes at ~21°C, quenched with EDTA, filtered, and their A_{260} measured via NanoDrop. Standard curve in panel A was used to calculate dNMP concentrations used for initial velocity calculations. Data represents a single experiment with single reactions for each substrate concentration. From this data it was calculated that $k_{cat} = 10\ s^{-1}$ and $K_m = 10$ nM. **[D]** *Exonuclease Activities of TREX1 on ssDNA.* Standard exonuclease reactions were prepared with indicated concentrations of hT1, incubated at room temperature for 20 minutes, quenched in ethanol, then visualized on a 23% urea-polyacrylamide gel. Top bands indicate undegraded 30-mer oligonucleotide ('30-mer'), which TREX1 degrades in single-nucleotide units to generate the laddered bands (other '-mer'). Densitometric quantification of these gels was used to calculate activity rates. Values provided at the bottom of each gel set are mean and standard deviation. Activity rates are calculated for each substrate concentration from 18 total reactions each via 6 replicates per 3 concentrations of hT1. Figure was prepared in PowerPoint (Microsoft).



Melting Temperatures (°C)

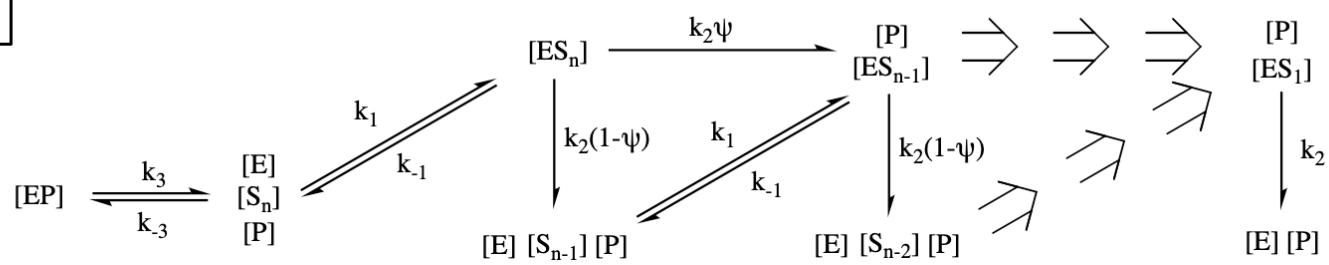
hT1-Apo	=	51.5 ± 0.359
hT1-ssDNA	=	65.1 ± 1.67
hT1-dsDNA	=	79.8 ± 0.236
mT1-Apo	=	58.2 ± 0.594
mT1-ssDNA	=	65.3 ± 1.15
mT1-dsDNA	=	78.3 ± 0.348

<u>Figure 2</u>: **TREX1 Has Significantly More Stable Binding to dsDNA Than ssDNA.** Melting curves and associated temperatures of TREX1 complexes. Indicated apoenzymes and enzyme-substrate complexes were subjected to an increasing temperature gradient in the presence of the nonspecific protein dye, SYPRO Orange, while fluorescence was monitored as a metric of protein unfolding. Data is from a single experiment with 8 replicate reactions per condition. Graph was prepared in R v3.6.1. The temperatures associated with the inflection points of the curves were calculated as a measure of melting temperature. Values provided indicate mean ± standard deviation across the 8 replicate reactions per condition. Calculations were performed in R v3.6.1, and graphic was prepared in PowerPoint (Microsoft).

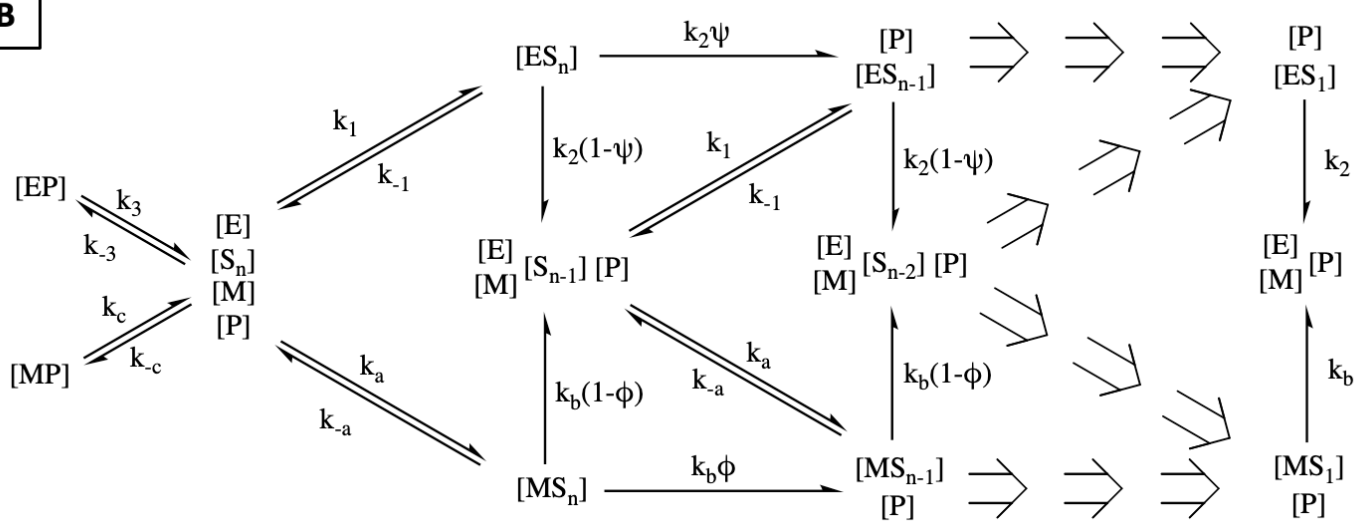

A
mTREX1 Backbone Fluctuations
mT1
mT1 + 4-mer
mT1 + dsDNA
RMSF (Å)
0
10
20
30
40
50
Protomer-A
Residues (N → C)
B
mT1-WT vs mT1-WT + 4-mer
q < 0.05
yes
no
Protomer-A
Residue (N → C)
mT1-WT vs mT1-WT + dsDNA
q < 0.05
yes
no
Protomer-A
Residue (N → C)

<u>Figure 3</u>: **Active-site Occupation by ssDNA Versus dsDNA is Associated with Differential mTREX1 Dynamics. [A]** *Comparative RMSF Plots of TREX1 Backbone Fluctuation*. All TREX1 Cα atoms were used in an RMSF calculation, as described in the relevant methods section. Plot lines displayed are mean RMSF values from the four simulations per system. The purple, orange, and green vertical lines define regions of interest, which are similarly colored in panel E. All calculations and plotting were performed in R v3.6.1. A mT1:4-mer structure modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops is included for reference. Both TREX1 protomers and all magnesium ions are colored grey, both ssDNA polymers are displayed as blue sticks. The dynamic regions identified in the RMSF plot are colored correspondingly. Graphic was constructed in PyMOL[39]. **[B]** *Barcode Graphs of Residues with Significantly Different Phi-Psi Angles*. For each residue, hierarchal clustering was used to determine conformations across all time frames, simulations, and systems. Residues were tested for differing conformation proportions between specific systems via chi-square tests followed by false discovery rate adjustment (q-values). Further details are found in the relevant methods section. Barcode plots show residues with significantly different conformation proportions between systems, with the title indicating the systems compared. Residues with blue markers had their conformations significantly altered by the presence of 4-mer and dsDNA in the active-site. Green markers indicate residues whose conformations were only significantly altered by the presence of 4-mer in the active-site. Red markers indicate residues whose conformations were only significantly altered by the presence of dsDNA in the active-site. Analyses and graphing were performed in R v3.6.1. A mT1:4-mer structure modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops is included for reference. Both TREX1 protomers and all magnesium ions are colored grey, both ssDNA polymers are displayed as purple sticks. The residues identified in the barcode graphs by colored markers are shown as correspondingly colored sticks. Graphic was constructed in PyMOL[39].

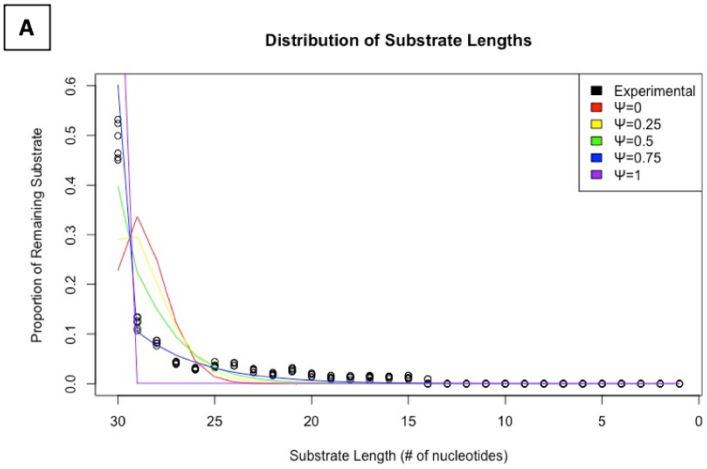
A

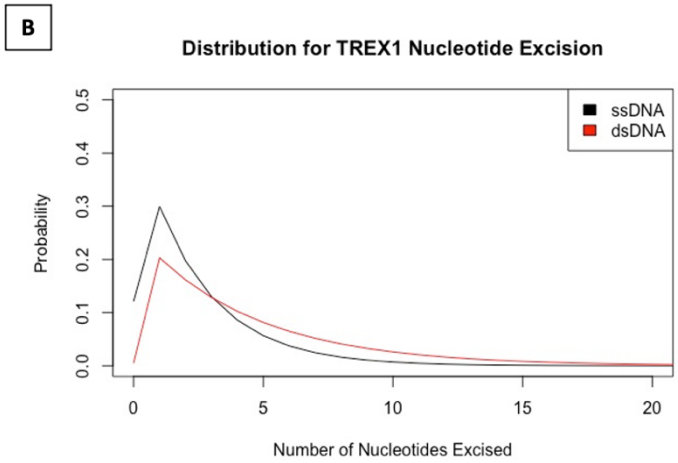

[EP]
k3
k-3
[E]
[Sn]
[P]
k1
k-1
[ESn]
k2ψ
[P]
[ESn-1]
[P]
[ES1]
k2(1-ψ)
k1
k-1
k2(1-ψ)
k2
[E] [Sn-1] [P]
[E] [Sn-2] [P]
[E] [P]

B

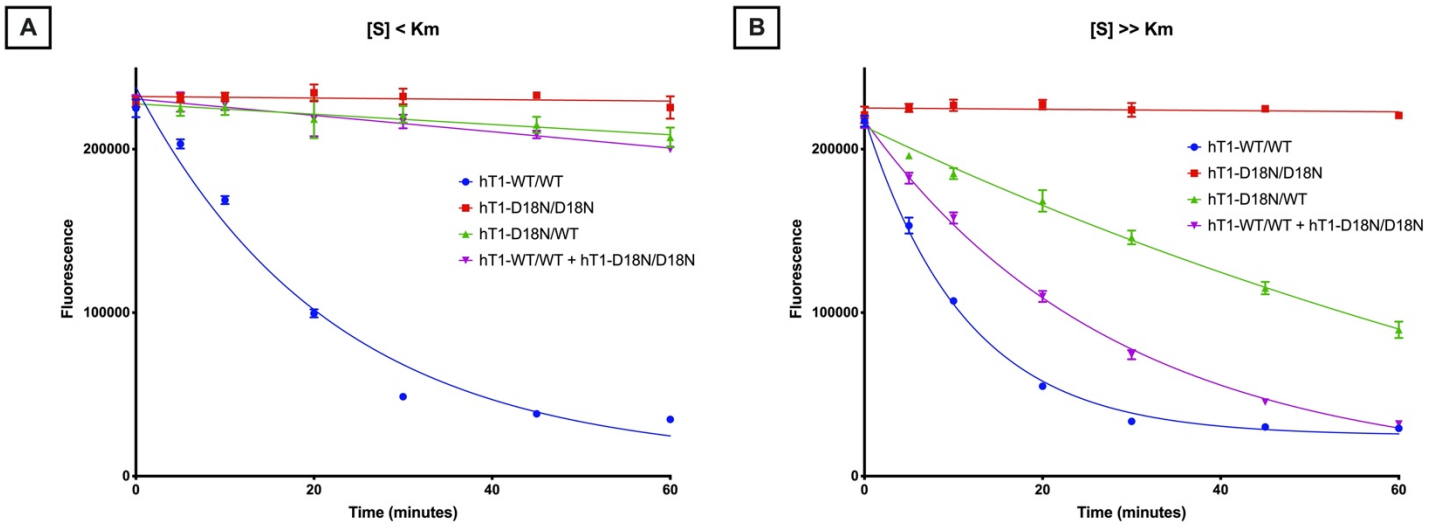

[EP]
k3
k-3
[E]
[Sn]
[M]
[P]
kc
k-c
[MP]
k1
k-1
[ESn]
k2ψ
[P]
[ESn-1]
[P]
[ES1]
k2(1-ψ)
k1
k-1
k2(1-ψ)
k2
[E]
[M]
[Sn-1] [P]
[E]
[M]
[Sn-2] [P]
[E]
[M]
[P]
ka
k-a
ka
k-a
kb(1-ϕ)
kb(1-ϕ)
kb
[MSn]
kbϕ
[MSn-1]
[P]
[MS1]
[P]

<u>Figure 4</u>: Proposed Kinetic Models for TREX1 Exonuclease Activity and Competition Between TREX1 Variants. Models are provided for exonuclease activity of wild-type TREX1 only (panel A), and for gross activity of TREX1 variants in competition (panel B). In both models, $k_{\square}$ are respective rate constants in the models, Ψ and Φ are the wild-type and mutant processivity parameters, respectively, *[E]* is concentration of free wild-type enzyme, *[M]* is concentration of free mutant enzyme, $[S_n]$ is concentration of free substrate with *n* excisable nucleotides, *[P]* is concentration of free product (excised dNMPs), and *[□-□]* is concentration of a complex of the respective reactants.


A
Distribution of Substrate Lengths
Experimental
Ψ=0
Ψ=0.25
Ψ=0.5
Ψ=0.75
Ψ=1
Proportion of Remaining Substrate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
30
25
20
15
10
5
0
Substrate Length (# of nucleotides)


B
Distribution for TREX1 Nucleotide Excision
ssDNA
dsDNA
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
Number of Nucleotides Excised

<u>Figure 5</u>: **TREX1 ssDNA Degradation is Consistent with Modest Processivity. [A]** A polyacrylamide gel-based TREX1 ssDNA degradation assay was performed as described in the relevant methods section, and the banding data quantified to give the proportion of total substrate concentration for each length substrate remaining. Substrate length proportions for all 6 replicate reactions are displayed ('Experimental'). The five optimized sets of Ψ, k_1, and k_{-1} parameter values identified in Supplemental 3 were used to simulate their predicted results for the experiment, which are indicated in the graph legend by their corresponding Ψ value. Simulated reactions are defined in Supplemental 6. **[B]** The TREX1 kinetic constants in Table 2 were used with Equation 18 to generate the predicted probability distribution for the number of nucleotides TREX1 excises from dsDNA and ssDNA substrates after initial enzyme-substrate complexing. For dsDNA, the matched kinetic constant values associated with $\Psi = 0.80$ were used. Calculations and plotting were performed in R v3.6.1.



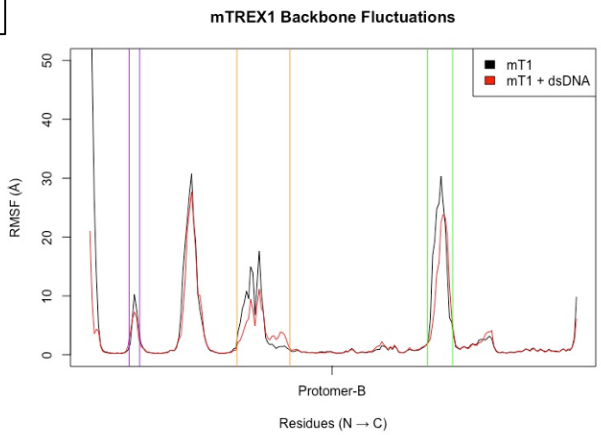
C

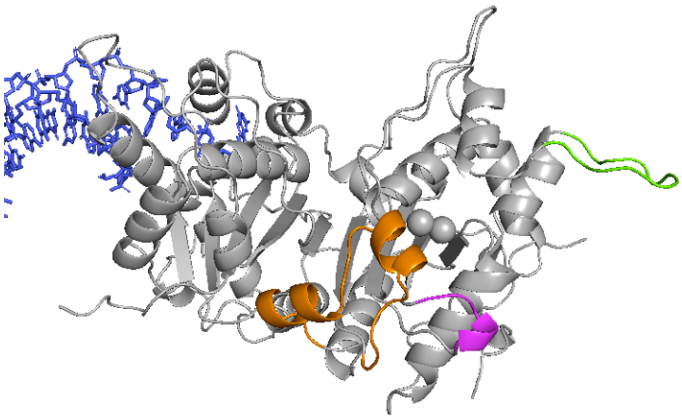
Relative Activities

		$[S] < K_m$:	$[S] >> K_m$:
hT1-WT/WT	=	1.0	1.0
hT1-D18N/D18N	=	0.0051	0.0028
hT1-D18N/WT	=	0.034	0.19
1:1:0 MIX	=	0.055	0.50

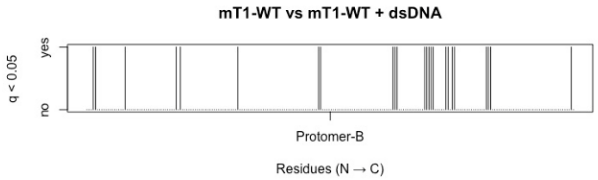
<u>Figure 6</u>: TREX1^{D18N} Heterodimer's Deficiency in dsDNA Exonuclease Activity is Dependent on Substrate Conditions. [A-B] *Fluorescence-Based Quantification of TREX1 dsDNA Degradation*. Standard exonuclease reactions were prepared with equimolar concentrations of the indicated enzymes, incubated at room temperature for the indicated times, quenched in SYBR Green, and dsDNA content measured by fluorescence. Plots of fluorescence vs time were generated and fit with one-phase decay nonlinear regression in Prism 9.0 (GraphPad). Plots are composites of 6 different reactions and representative of three experiments. Data points indicate mean, and error bars represent standard deviation. **[C]** *Activity Rates of TREX1 Variants*. Initial velocities were quantified from the respective regression lines in panels A-B and normalized to wild-type initial velocity to calculate relative activity. '[S] < K_m' refers to ~1 nM of a 10-kb dsDNA plasmid, and '[S] >> K_m' refers to ~1 μM of a 30-bp self-annealing oligo. For the 1:1:0 enzyme mix, mix ratios indicate the prevalence of hT1$^{+/+}$, hT1$^{MUT/MUT}$, and hT1$^{MUT/+}$, respectively. All reactions have equimolar concentrations of total hT1, even mixes.

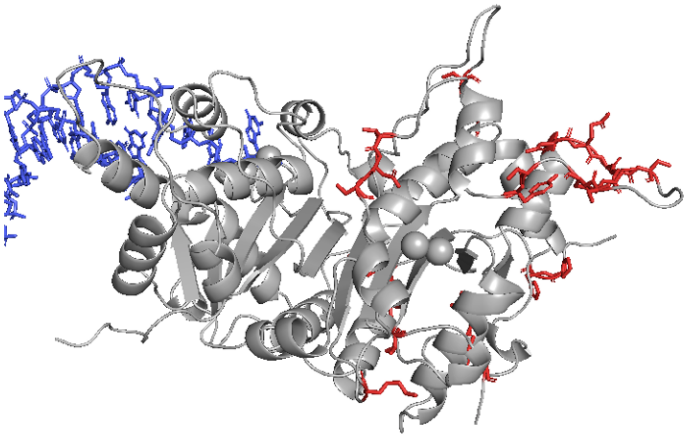
A


mTREX1 Backbone Fluctuations
mT1
mT1 + dsDNA
RMSF (Å)
0
10
20
30
40
50
Protomer-B
Residues (N → C)

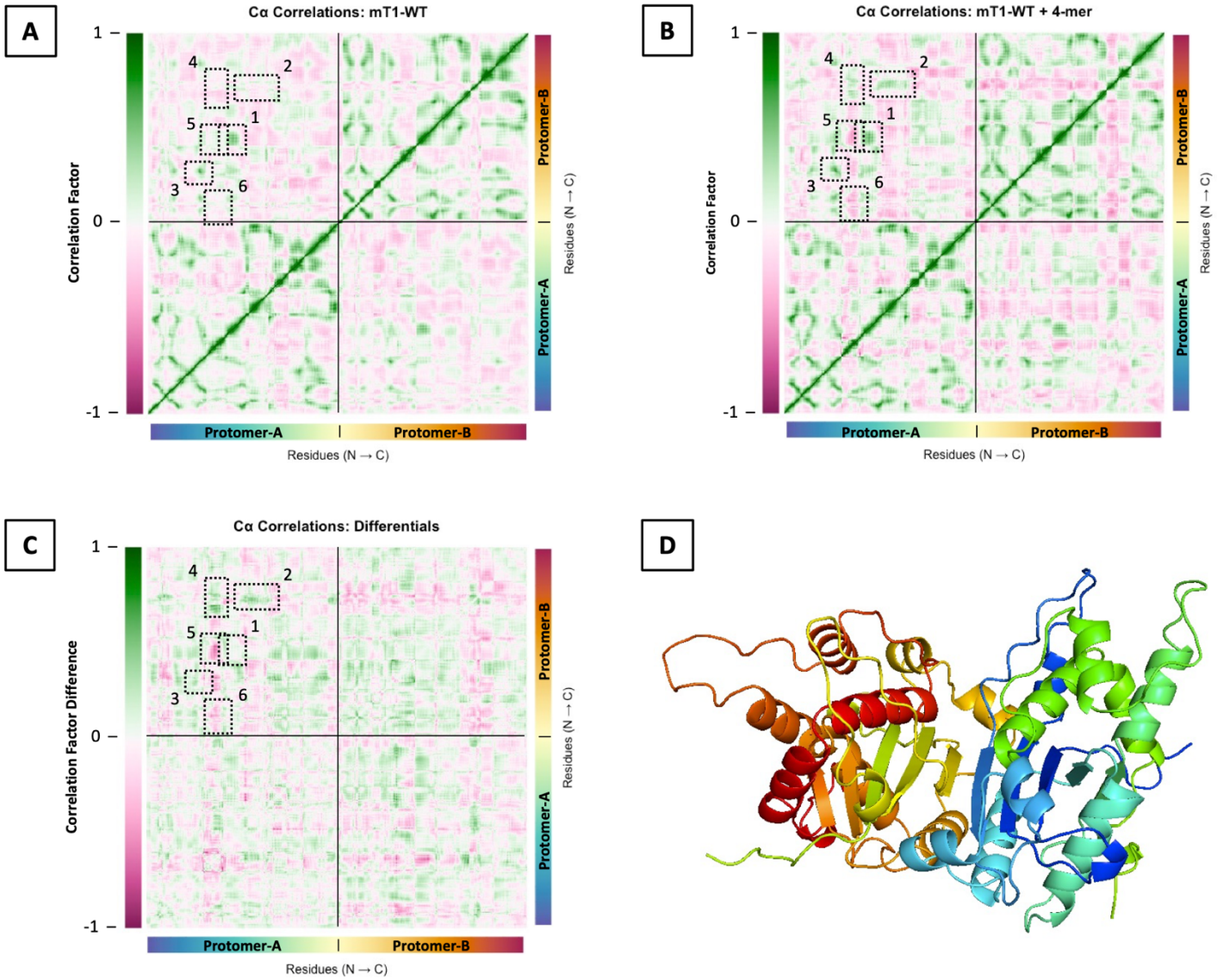


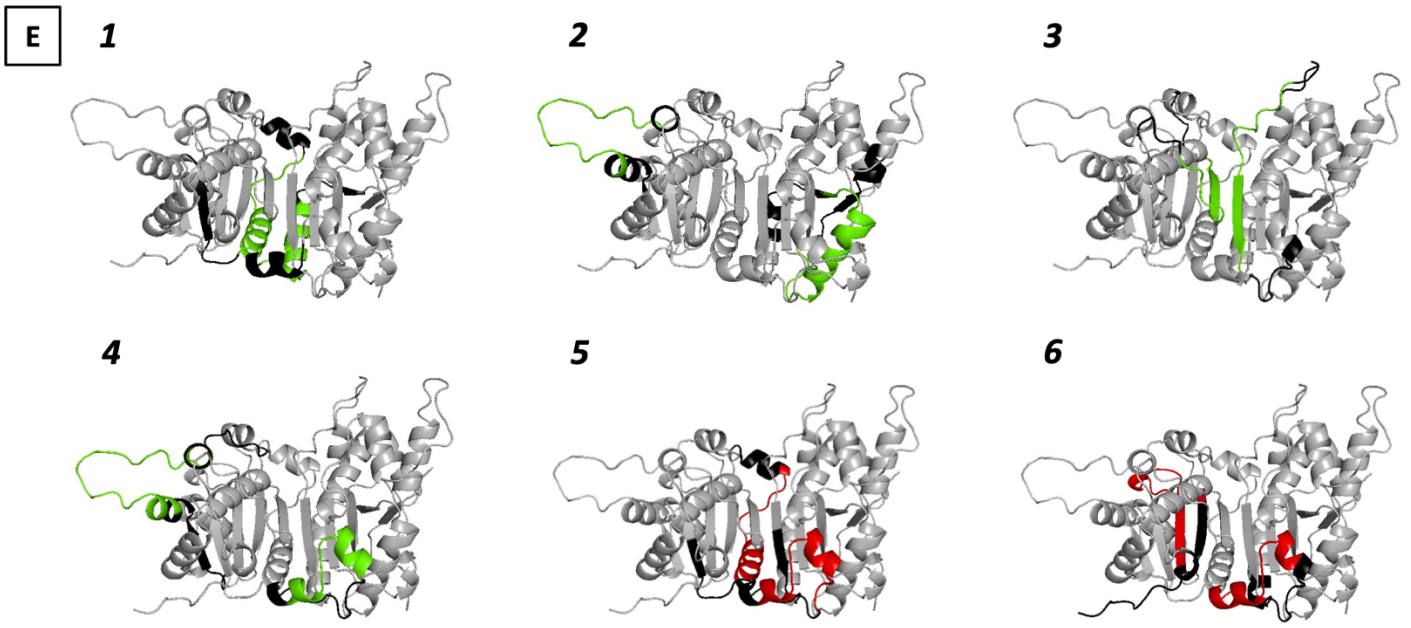
B


mT1-WT vs mT1-WT + dsDNA
q < 0.05
yes
no
Protomer-B
Residues (N → C)

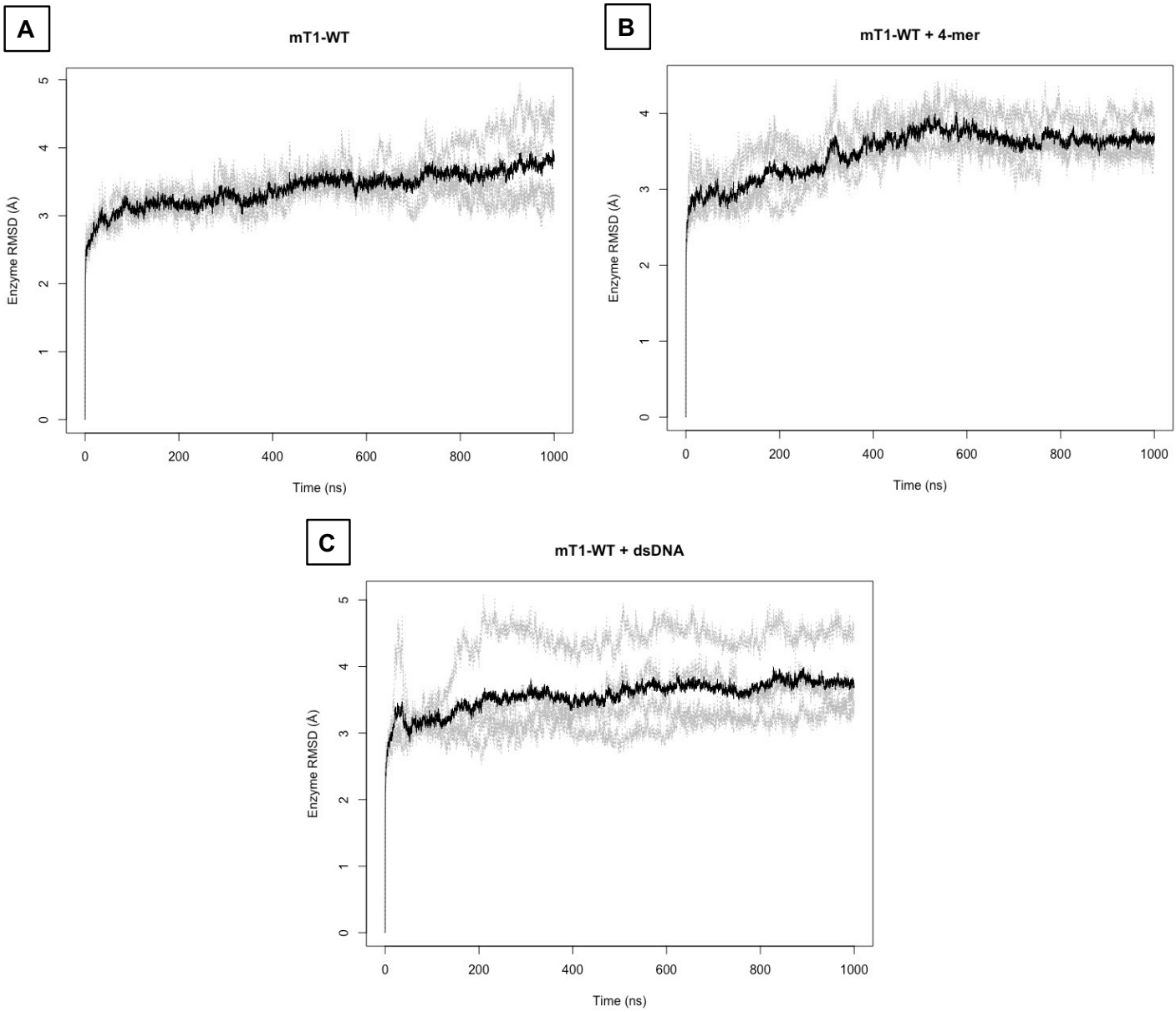


<u>Figure 7</u>: **Active-site Occupation by dsDNA is Associated with Differential Dynamics in the Opposing Protomer. [A]** *Comparative RMSF Plots of mTREX1 Backbone*. All TREX1 Cα atoms in the unoccupied protomer of each system ('Protomer-B') were used in an RMSF calculation, as described in the relevant methods section. Plot lines displayed are mean RMSF values from the four simulations per system. The purple, orange, and green vertical lines define regions of interest, which are similarly colored in the reference mT1 crystal structure (PDB = '2IOC') modeled to include disordered flexible loops. Both TREX1 protomers and all magnesium ions are colored grey and the dsDNA polymers are displayed as blue sticks. Graphic was constructed in PyMOL[39]. **[B]** *Barcode Graph of Residues with Significantly Different Phi-Psi Angles*. For each residue, hierarchal clustering was used to determine conformations across all time frames, simulations, and systems. Residues were tested for differing conformation proportions between specific systems via chi-square tests followed by false discovery rate adjustment (q-values). Further details are found in the relevant methods section. Barcode plots show residues with significantly different conformation proportions between systems, with the title indicating the systems compared. Analyses and graphing were performed in R v3.6.1. A mT1 structure modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops is included for reference. Both TREX1 protomers and all magnesium ions are colored grey. The residues identified in the barcode graph are shown as red colored sticks. Graphic was constructed in PyMOL[39].

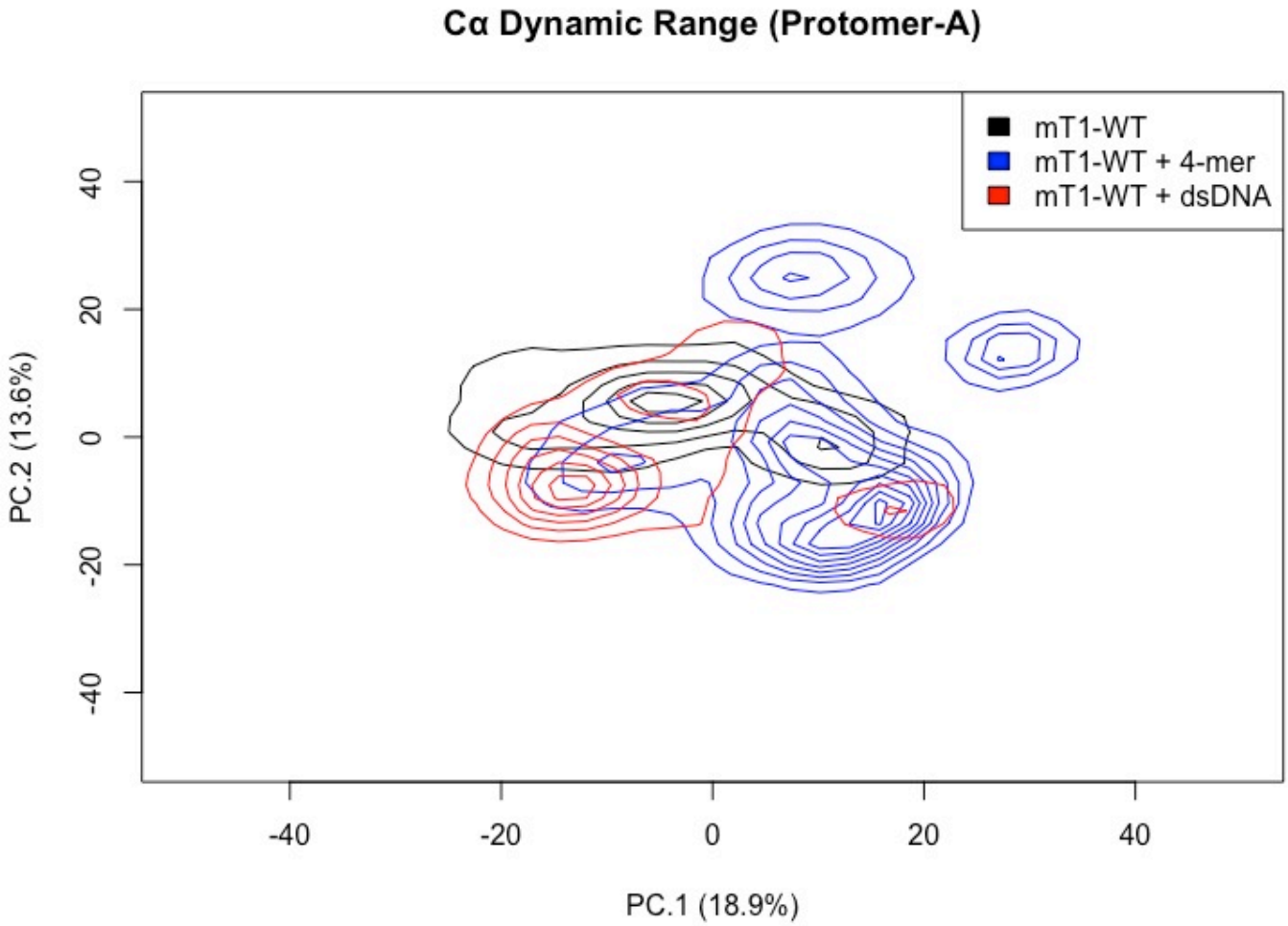

A
Cα Correlations: mT1-WT
Correlation Factor
1
0
-1
4
2
5
1
3
6
Protomer-A
Protomer-B
Residues (N → C)
B
Cα Correlations: mT1-WT + 4-mer
C
Cα Correlations: Differentials
Correlation Factor Difference
D


E
1
2
3
4
5
6

<u>Figure 8</u>: Regions of the mTREX1 Structure Associated with Exonuclease Activity Have Correlated Dynamics Between Protomers. The α-carbons of every residue in the mT1 structure were subjected to correlation analyses against all other α-carbons in the mT1 structure, with comparisons across every time frame in every simulation, for each of the mT1 and mT1:4-mer systems. **[A-B]** *Correlation Heat Maps*. Bottom-x and right-y axes indicate the residues being compared, and the color gradient for these axes' labels indicate the corresponding region of the mT1 structure shown in panel D. Vertical color legends to the left of each graph indicate the correlation factor between residues on the map, where '1' is a perfect positive correlation of dynamics and '-1' is a perfect negative correlation of dynamics. Solid black lines separate protomers, and dashed black boxes define regions of interest that are correspondingly numbered in panel E. Calculations and graphing were performed in R v3.6.1, and all other features were added in PowerPoint (Microsoft). **[C]** *Differential Correlation Heat Map*. A heat map of the difference in correlation factors achieved by subtraction of panel A from panel B. All other features are as discussed for panels A-B. **[D]** *Color Legend of TREX1 Structure*. A mT1 structure modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops. Backbone colors correspond to the bottom-x and right-y axes in panels A-C. Graphic was prepared in PyMOL[39]. **[E]** *Structural Representations of Regions of Interest*. mT1 structures modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops. Italicized numbers indicate the correlated regions from panels A-C that are illustrated. The mT1 structures are generally shown as grey cartoons. The inner 50% of correlated areas from panels A-C are colored green or red for positively or negatively correlated regions, respectively. The peripheral 50% of correlated areas from panels A-C are colored black. Graphics were prepared in PyMOL[39].

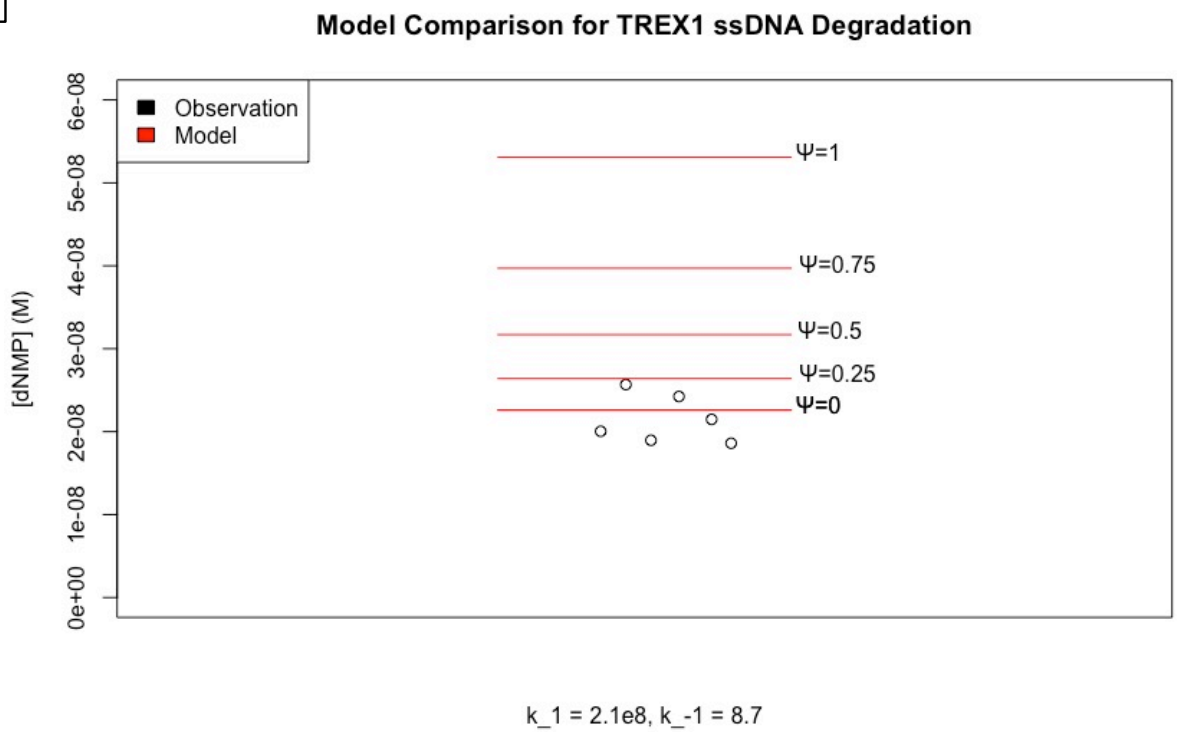

A
mT1-WT
Enzyme RMSD (Å)
Time (ns)
B
mT1-WT + 4-mer
Enzyme RMSD (Å)
Time (ns)
C
mT1-WT + dsDNA
Enzyme RMSD (Å)
Time (ns)

<u>Supplemental 1</u>: **RMSD of mTREX1 Structure from Initial Conditions.** All TREX1 atoms were used in an RMSD calculation as described in the relevant methods section. RMSD from initial conditions was plotted as a function of time. Grey lines indicate individual simulations, and black lines indicate averages for the 4 replicate simulations of each system. All calculations and graphing were performed in R v3.6.1.

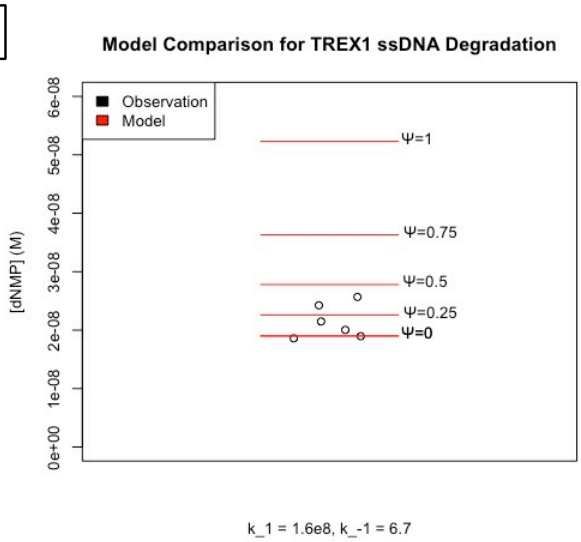

Cα Dynamic Range (Protomer-A)
mT1-WT
mT1-WT + 4-mer
mT1-WT + dsDNA
PC.2 (13.6%)
PC.1 (18.9%)
40
20
0
-20
-40
-40
-20
0
20
40

<u>Supplemental 2</u>: **Conformational Free Energy Maps of mTREX1 Backbone.** Cartesian coordinates of every Cα in mT1 were used to calculate map density functions as described in the relevant methods section. Density functions were plotted as contour maps, with innermost rings indicating the areas of highest density for respective systems. All calculations and graphing were performed in R v3.6.1.

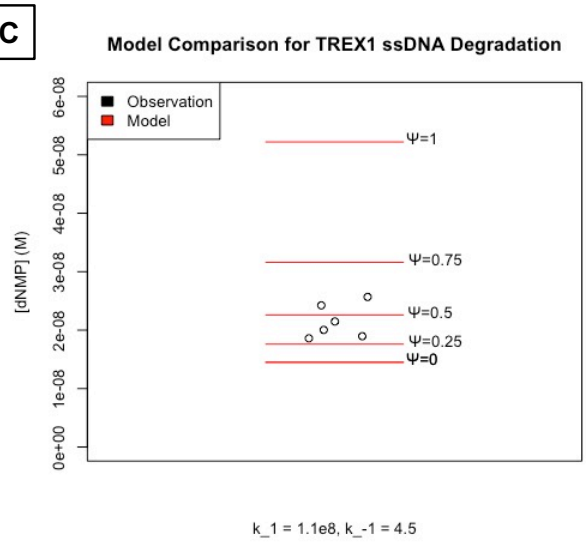
A


Model Comparison for TREX1 ssDNA Degradation
Observation
Model
Ψ=1
Ψ=0.75
Ψ=0.5
Ψ=0.25
Ψ=0
[dNMP] (M)
0e+00
1e-08
2e-08
3e-08
4e-08
5e-08
6e-08
k_1 = 2.1e8, k_-1 = 8.7

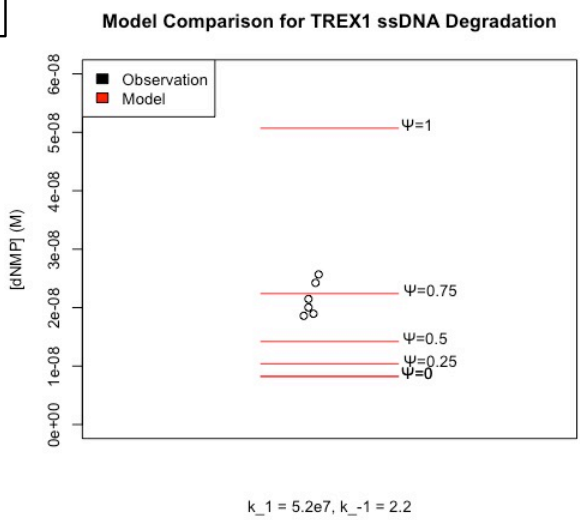
B


Model Comparison for TREX1 ssDNA Degradation
Observation
Model
Ψ=1
Ψ=0.75
Ψ=0.5
Ψ=0.25
Ψ=0
[dNMP] (M)
0e+00
1e-08
2e-08
3e-08
4e-08
5e-08
6e-08
k_1 = 1.6e8, k_-1 = 6.7

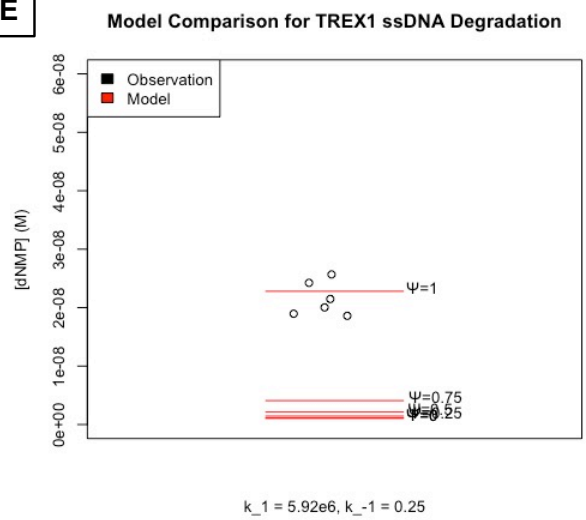
C


Model Comparison for TREX1 ssDNA Degradation
Observation
Model
Ψ=1
Ψ=0.75
Ψ=0.5
Ψ=0.25
Ψ=0
[dNMP] (M)
0e+00
1e-08
2e-08
3e-08
4e-08
5e-08
6e-08
k_1 = 1.1e8, k_-1 = 4.5

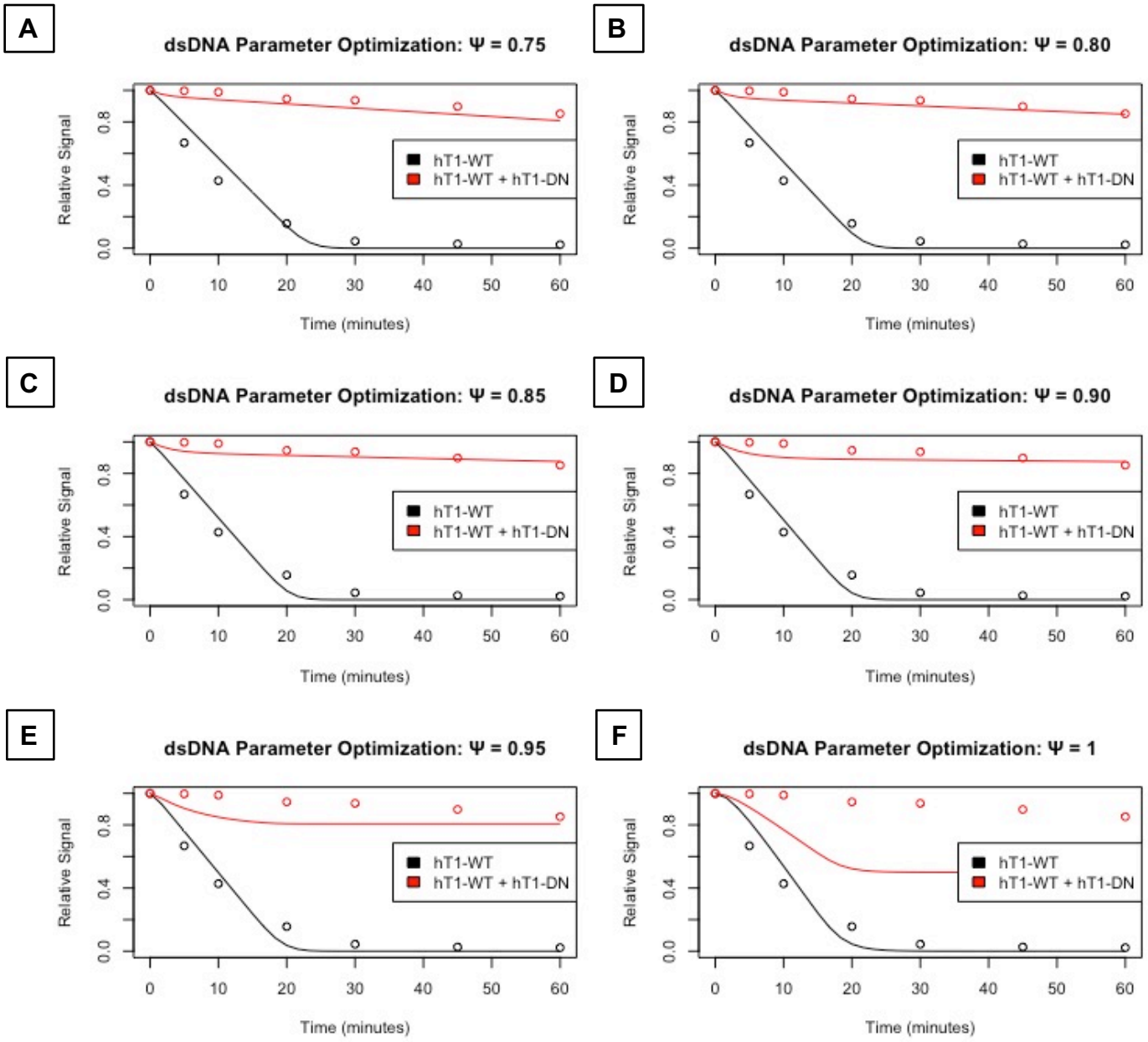
D


Model Comparison for TREX1 ssDNA Degradation
Observation
Model
Ψ=1
Ψ=0.75
Ψ=0.5
Ψ=0.25
Ψ=0
[dNMP] (M)
0e+00
1e-08
2e-08
3e-08
4e-08
5e-08
6e-08
k_1 = 5.2e7, k_-1 = 2.2

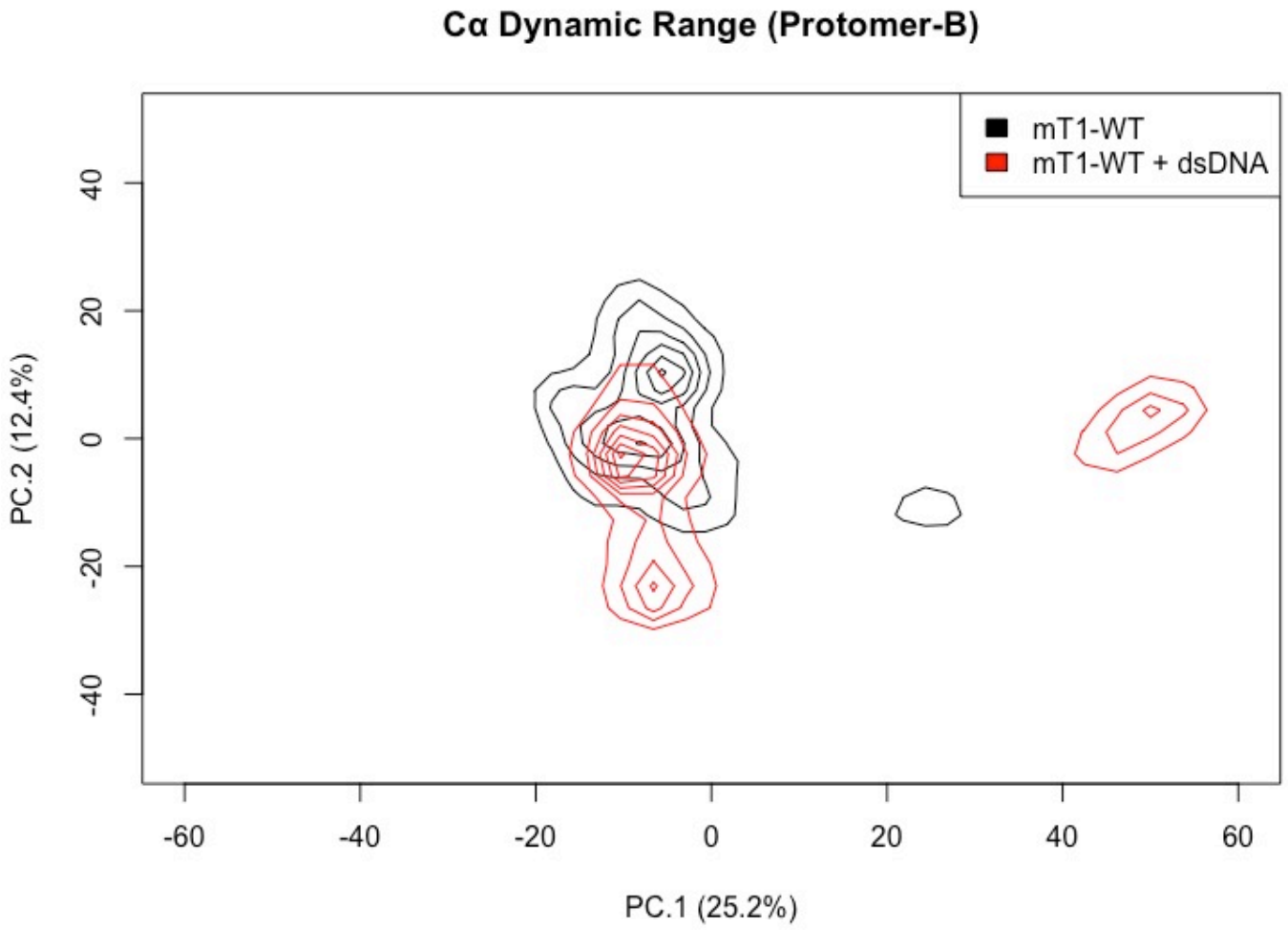
E


Model Comparison for TREX1 ssDNA Degradation
Observation
Model
Ψ=1
Ψ=0.75
Ψ=0.5
Ψ=0.25
Ψ=0
[dNMP] (M)
0e+00
1e-08
2e-08
3e-08
4e-08
5e-08
6e-08
k_1 = 5.92e6, k_-1 = 0.25

<u>Supplemental 3</u>: **Optimization of Parameters for TREX1 ssDNA Kinetic Model.** A polyacrylamide gel-based TREX1 ssDNA degradation assay was performed as described in the relevant methods section, and the banding data quantified to give the concentration of product in the reaction. Product concentrations for all 6 replicate reactions are displayed ('Observation'). Any variability along the x-axis in the empirical data points between graphs is not indicative of a different data set but an artifact of R plotting code. The kinetic model for hTREX1 ssDNA exonuclease activity was optimized at several processivity parameter values ('Ψ') to determine the matching kinetic constant values ('k_1', 'k_-1') that gave the best fit to the empirical data. The 'A-E' plots correspond to optimization for Ψ values of 0, 0.25, 0.5, 0.75, and 1, respectively. Labels on the bottom-x axis of each graph indicate the model's relevant reaction parameters, and the labeled red lines on each graph indicate the predicted product concentration for indicated Ψ values under those parameters. All calculations and plotting were performed in R v3.6.1, and figure preparation was performed in PowerPoint (Microsoft). Simulated reactions are defined in Supplemental 6.


A
dsDNA Parameter Optimization: Ψ = 0.75
B
dsDNA Parameter Optimization: Ψ = 0.80
C
dsDNA Parameter Optimization: Ψ = 0.85
D
dsDNA Parameter Optimization: Ψ = 0.90
E
dsDNA Parameter Optimization: Ψ = 0.95
F
dsDNA Parameter Optimization: Ψ = 1
Relative Signal
0.0
0.4
0.8
Time (minutes)
0
10
20
30
40
50
60
hT1-WT
hT1-WT + hT1-DN

<u>Supplemental 4</u>: **Optimization of Parameters for TREX1 dsDNA Kinetic Model.** Standard exonuclease reactions were prepared with 15 nM hT1 (black data) or 15 nM of a 1:1 mix of hT1:hT1-DN (red data), incubated at room temperature for the indicated times, quenched in SYBR Green, and dsDNA content measured by fluorescence. Plots of fluorescence vs time were generated and normalized to maximum and minimum fluorescence. Data points on the graph correspond to these empirical data, and indicate means of 18 reactions per condition across three experiments. The parameters of the TREX1 dsDNA kinetic model were optimized to fit these empirical data for each of the processivity values listed in respective plot titles. Optimization was carried out by starting with the ssDNA k_1 and k_{-1} values, increasing k_1 until the predicted wild-type degradation rate (black lines) matched observation, then decreasing k_{-1} until the predicted level of enzyme competition (red lines) matched observation. Calculations and plotting were performed in R v3.6.1. Simulated reactions are defined in Supplemental 6. The black and red data points here are derived from the same data as the blue and purple data points in Figure 6A, respectively, after normalization to maximum and minimum fluorescence.


Cα Dynamic Range (Protomer-B)
mT1-WT
mT1-WT + dsDNA
PC.2 (12.4%)
PC.1 (25.2%)
40
20
0
-20
-40
-60
-40
-20
0
20
40
60

<u>Supplemental 5</u>: **Conformational Free Energy Maps of mTREX1 Backbone.** Cartesian coordinates of every Cα in the unbound protomer ('Protomer-B') of mT1 in every time frame in every simulation in both systems were used to calculate map density functions as described in the relevant methods section. Density functions were plotted as contour maps, with innermost rings indicating the areas of highest density for respective systems. All calculations and graphing were performed in R v3.6.1.

Simulations Legend

Reaction	Description
R1	Supplemental 3A, ‘Ψ=0’
R2	Supplemental 3A, ‘Ψ=0.25’
R3	Supplemental 3A, ‘Ψ=0.5’
R4	Supplemental 3A, ‘Ψ=0.75’
R5	Supplemental 3A, ‘Ψ=1’
R6	Supplemental 3B, ‘Ψ=0’
R7	Supplemental 3B, ‘Ψ=0.25’
R8	Supplemental 3B, ‘Ψ=0.5’
R9	Supplemental 3B, ‘Ψ=0.75’
R10	Supplemental 3B, ‘Ψ=1’
R11	Supplemental 3C, ‘Ψ=0’
R12	Supplemental 3C, ‘Ψ=0.25’
R13	Supplemental 3C, ‘Ψ=0.5’
R14	Supplemental 3C, ‘Ψ=0.75’
R15	Supplemental 3C, ‘Ψ=1’
R16	Supplemental 3D, ‘Ψ=0’
R17	Supplemental 3D, ‘Ψ=0.25’
R18	Supplemental 3D, ‘Ψ=0.5’
R19	Supplemental 3D, ‘Ψ=0.75’
R20	Supplemental 3D, ‘Ψ=1’
R21	Supplemental 3E, ‘Ψ=0’
R22	Supplemental 3E, ‘Ψ=0.25’
R23	Supplemental 3E, ‘Ψ=0.5’
R24	Supplemental 3E, ‘Ψ=0.75’
R25	Supplemental 3E, ‘Ψ=1’
R26	Figure 5A, red line
R27	Figure 5A, yellow line
R28	Figure 5A, green line

R29	Figure 5A, blue line
R30	Figure 5A, purple line
R31	Supplemental 4A, black line
R32	Supplemental 4A, red line
R33	Supplemental 4B, black line
R34	Supplemental 4B, red line
R35	Supplemental 4C, black line
R36	Supplemental 4C, red line
R37	Supplemental 4D, black line
R38	Supplemental 4D, red line
R39	Supplemental 4E, black line
R40	Supplemental 4E, red line
R41	Supplemental 4F, black line
R42	Supplemental 4F, red line

Simulations Parameters

	R1	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**R9**	**R10**
t (min)	20	20	20	20	20	20	20	20	20	20
Δt (ms)	1	1	1	1	1	1	1	1	1	1
$[E]_0$ (M)	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}
$[M]_0$ (M)	0	0	0	0	0	0	0	0	0	0
β	30	30	30	30	30	30	30	30	30	30
$[S_\beta]_0$ (M)	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9
Ψ	0	0.25	0.5	0.75	1	0	0.25	0.5	0.75	1
k_{-1} (s^{-1})	8.7	8.7	8.7	8.7	8.7	6.7	6.7	6.7	6.7	6.7
k_1 ($M^{-1}s^{-1}$)	2.1e8	2.1e8	2.1e8	2.1e8	2.1e8	1.6e8	1.6e8	1.6e8	1.6e8	1.6e8
k_2 ($M^{-1}s^{-1}$)	16	16	16	16	16	16	16	16	16	16
Φ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ
k_{-a} (s^{-1})	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}
k_a ($M^{-1}s^{-1}$)	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1
k_b ($M^{-1}s^{-1}$)	0	0	0	0	0	0	0	0	0	0

	R11	**R12**	**R13**	**R14**	**R15**	**R16**	**R17**	**R18**	**R19**	**R20**
t (min)	20	20	20	20	20	20	20	20	20	20
Δt (ms)	1	1	1	1	1	1	1	1	1	1
$[E]_0$ (M)	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}
$[M]_0$ (M)	0	0	0	0	0	0	0	0	0	0
β	30	30	30	30	30	30	30	30	30	30
$[S_\beta]_0$ (M)	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9
Ψ	0	0.25	0.5	0.75	1	0	0.25	0.5	0.75	1
k_{-1} (s^{-1})	4.5	4.5	4.5	4.5	4.5	2.2	2.2	2.2	2.2	2.2
k_1 ($M^{-1}s^{-1}$)	1.1e8	1.1e8	1.1e8	1.1e8	1.1e8	5.2e7	5.2e7	5.2e7	5.2e7	5.2e7

k_2 ($M^{-1}s^{-1}$)	16	16	16	16	16	16	16	16	16	16
Φ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ
k_{-a} (s^{-1})	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}
k_a ($M^{-1}s^{-1}$)	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1
k_b ($M^{-1}s^{-1}$)	0	0	0	0	0	0	0	0	0	0

	R21	**R22**	**R23**	**R24**	**R25**	**R26**	**R27**	**R28**	**R29**	**R30**
t (min)	20	20	20	20	20	20	20	20	20	20
Δt (ms)	1	1	1	1	1	1	1	1	1	1
$[E]_0$ (M)	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}
$[M]_0$ (M)	0	0	0	0	0	0	0	0	0	0
β	30	30	30	30	30	30	30	30	30	30
$[S_\beta]_0$ (M)	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9	15e-9
Ψ	0	0.25	0.5	0.75	1	0	0.25	0.5	0.75	1
k_{-1} (s^{-1})	0.25	0.25	0.25	0.25	0.25	8.7	6.7	4.5	2.2	0.25
k_1 ($M^{-1}s^{-1}$)	5.9e6	5.9e6	5.9e6	5.9e6	5.9e6	2.1e8	1.6e8	1.1e8	5.2e7	5.9e6
k_2 ($M^{-1}s^{-1}$)	16	16	16	16	16	16	16	16	16	16
Φ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ
k_{-a} (s^{-1})	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}
k_a ($M^{-1}s^{-1}$)	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1
k_b ($M^{-1}s^{-1}$)	0	0	0	0	0	0	0	0	0	0

	R31	**R32**	**R33**	**R34**	**R35**	**R36**	**R37**	**R38**	**R39**	**R40**
t (min)	60	60	60	60	60	60	60	60	60	60
Δt (ms)	1	1	1	1	1	1	1	1	1	1
$[E]_0$ (M)	7.5e-9	3.75e-9	7.5e-9	3.75e-9	7.5e-9	3.75e-9	7.5e-9	3.75e-9	7.5e-9	3.75e-9
$[M]_0$ (M)	0	3.75e-9	0	3.75e-9	0	3.75e-9	0	3.75e-9	0	3.75e-9
β	1e4	1e4	1e4	1e4	1e4	1e4	1e4	1e4	1e4	1e4
$[S_\beta]_0$ (M)	0.83e-9	0.83e-9	0.83e-9	0.83e-9	0.83e-9	0.83e-9	0.83e-9	0.83e-9	0.83e-9	0.83e-9
Ψ	0.75	0.75	0.80	0.80	0.85	0.85	0.90	0.90	0.95	0.95
k_{-1} (s^{-1})	0.10	0.10	0.050	0.050	0.020	0.020	0.010	0.010	0	0
k_1 ($M^{-1}s^{-1}$)	1e9	1e9	1e9	1e9	1e9	1e9	1e9	1e9	1e9	1e9
k_2 ($M^{-1}s^{-1}$)	10	10	10	10	10	10	10	10	10	10
Φ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ	Ψ
k_{-a} (s^{-1})	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}	k_{-1}
k_a ($M^{-1}s^{-1}$)	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1	k_1
k_b ($M^{-1}s^{-1}$)	0	0	0	0	0	0	0	0	0	0

	R41	**R42**
t (min)	60	60
Δt (ms)	1	1
$[E]_0$ (M)	7.5e-9	3.75e-9
$[M]_0$ (M)	0	3.75e-9
β	1e4	1e4
$[S_\beta]_0$ (M)	0.83e-9	0.83e-9
Ψ	1	1
k_{-1} (s^{-1})	0	0

k_1 ($M^{-1}s^{-1}$)	1e9	1e9
k_2 ($M^{-1}s^{-1}$)	10	10
Φ	Ψ	Ψ
k_{-a} (s^{-1})	k_{-1}	k_{-1}
k_a ($M^{-1}s^{-1}$)	k_1	k_1
k_b ($M^{-1}s^{-1}$)	0	0

<u>Supplemental 6</u>: **Kinetic Parameters for Simulated Reactions in These Studies.** A table of the parameters used to simulate TREX1 exonuclease activity. A legend identifying the simulations is also provided. Tables were prepared in Word (Microsoft).

REFERENCES

1. Perrino, F. W., Miller, H. & Ealey, K. A. Identification of a 3'-->5'-exonuclease that removes cytosine arabinoside monophosphate from 3' termini of DNA. *J. Biol. Chem.* **269**, 16357–16363 (1994).
2. Simpson, S. R., Hemphill, W. O., Hudson, T. & Perrino, F. W. TREX1 – Apex predator of cytosolic DNA metabolism. *DNA Repair* 102894 (2020) doi:10.1016/j.dnarep.2020.102894.
3. Rice, G. I., Rodero, M. P. & Crow, Y. J. Human Disease Phenotypes Associated With Mutations in TREX1. *J. Clin. Immunol.* **35**, 235–243 (2015).
4. Peschke, K. *et al.* Loss of Trex1 in Dendritic Cells Is Sufficient To Trigger Systemic Autoimmunity. *J. Immunol. Baltim. Md 1950* **197**, 2157–2166 (2016).
5. Simpson, S. R. *et al.* T Cells Produce IFN-α in the TREX1 D18N Model of Lupus-like Autoimmunity. *J. Immunol.* **204**, 348–359 (2020).
6. Yang, Y.-G., Lindahl, T. & Barnes, D. E. Trex1 Exonuclease Degrades ssDNA to Prevent Chronic Checkpoint Activation and Autoimmune Disease. *Cell* **131**, 873–886 (2007).
7. Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**, 587–598 (2008).
8. Mohr, L. *et al.* ER-directed TREX1 limits cGAS activation at micronuclei. *Mol. Cell* S1097276520309588 (2021) doi:10.1016/j.molcel.2020.12.037.
9. Rego, S. L. *et al.* TREX1 D18N mice fail to process erythroblast DNA resulting in inflammation and dysfunctional erythropoiesis. *Autoimmunity* **51**, 333–344 (2018).
10. Mazur, D. J. & Perrino, F. W. Excision of 3′ Termini by the Trex1 and TREX2 3′→5′ Exonucleases CHARACTERIZATION OF THE RECOMBINANT PROTEINS. *J. Biol. Chem.* **276**, 17022–17029 (2001).
11. Belyakova, N. V. *et al.* Proof-reading 3'-->5' exonucleases isolated from rat liver nuclei. *Eur. J. Biochem.* **217**, 493–500 (1993).
12. Mazur, D. J. & Perrino, F. W. Identification and Expression of the TREX1 and TREX2 cDNA Sequences Encoding Mammalian 3′→5′ Exonucleases. *J. Biol. Chem.* **274**, 19655–19660 (1999).
13. Ablasser, A. *et al.* TREX1 Deficiency Triggers Cell-Autonomous Immunity in a cGAS-Dependent Manner. *J. Immunol.* **192**, 5993–5997 (2014).

14. Gray, E. E., Treuting, P. M., Woodward, J. J. & Stetson, D. B. cGAS is required for lethal autoimmune disease in the Trex1-deficient mouse model of Aicardi-Goutieres Syndrome. *J. Immunol. Baltim. Md 1950* **195**, 1939–1943 (2015).
15. Xiao, N. *et al.* cGAS activation causes lupus-like autoimmune disorders in a TREX1 mutant mouse model. *J. Autoimmun.* **100**, 84–94 (2019).
16. Wu, J. *et al.* Cyclic-GMP-AMP Is An Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. *Science* **339**, (2013).
17. Rice, G. *et al.* Heterozygous Mutations in TREX1 Cause Familial Chilblain Lupus and Dominant Aicardi-Goutières Syndrome. *Am. J. Hum. Genet.* **80**, 811–815 (2007).
18. Bailey, S. L., Harvey, S., Perrino, F. W. & Hollis, T. Defects in DNA degradation revealed in crystal structures of TREX1 exonuclease mutations linked to autoimmune disease. *DNA Repair* **11**, 65–73 (2012).
19. Fye, J. M., Orebaugh, C. D., Coffin, S. R., Hollis, T. & Perrino, F. W. Dominant Mutations of the TREX1 Exonuclease Gene in Lupus and Aicardi-Goutières Syndrome. *J. Biol. Chem.* **286**, 32373–32382 (2011).
20. Orebaugh, C. D. *et al.* The TREX1 C-terminal Region Controls Cellular Localization through Ubiquitination. *J. Biol. Chem.* **288**, 28881–28892 (2013).
21. Orebaugh, C. D., Fye, J. M., Harvey, S., Hollis, T. & Perrino, F. W. The TREX1 Exonuclease R114H Mutation in Aicardi-Goutières Syndrome and Lupus Reveals Dimeric Structure Requirements for DNA Degradation Activity♦. *J. Biol. Chem.* **286**, 40246–40254 (2011).
22. Silva, U. de *et al.* The Crystal Structure of TREX1 Explains the 3′ Nucleotide Specificity and Reveals a Polyproline II Helix for Protein Partnering. *J. Biol. Chem.* **282**, 10537–10543 (2007).
23. Huang, K.-W. *et al.* Identification of Inhibitors for the DEDDh Family of Exonucleases and a Unique Inhibition Mechanism by Crystal Structure Analysis of CRN-4 Bound with 2-Morpholin-4-ylethanesulfonate (MES). *J. Med. Chem.* **59**, 8019–8029 (2016).
24. Mason, P. A. & Cox, L. S. The role of DNA exonucleases in protecting genome stability and their impact on ageing. *Age* **34**, 1317–1340 (2012).
25. Fye, J. M., Coffin, S. R., Orebaugh, C. D., Hollis, T. & Perrino, F. W. The Arg-62 residues of the TREX1 exonuclease act across the dimer interface contributing to catalysis in the opposing protomers. *J. Biol. Chem.* **289**, 11556–11565 (2014).

26. Hemphill, W. O. & Perrino, F. W. Measuring TREX1 and TREX2 exonuclease activities. *Methods Enzymol.* **625**, 109–133 (2019).

27. Lehtinen, D. A., Harvey, S., Mulcahy, M. J., Hollis, T. & Perrino, F. W. The TREX1 Double-stranded DNA Degradation Activity Is Defective in Dominant Mutations Associated with Autoimmune Disease. *J. Biol. Chem.* **283**, 31649–31656 (2008).

28. Grieves, J. L. *et al.* Exonuclease TREX1 degrades double-stranded DNA to prevent spontaneous lupus-like inflammatory disease. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5117–5122 (2015).

29. Lindahl, T., Gally, J. A. & Edelman, G. M. Properties of Deoxyribonuclease III from Mammalian Tissues. *J. Biol. Chem.* **244**, 5014–5019 (1969).

30. Ault, A. An introduction to enzyme kinetics. *J. Chem. Educ.* **51**, 381 (1974).

31. de Silva, U. STRUCTURAL AND BIOCHEMICAL STUDIES OF TREX - THREE PRIME REPAIR EXONUCLEASES. (WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES, 2007).

32. Bailey, S. L. THE TREX1 3 EXONUCLEASE AND AUTOIMMUNE DISEASE: STRUCTURAL AND BIOCHEMICAL ANALYSIS OF DISEASE MUTANTS INVOLVED IN AUTOIMMUNE DYSFUNCTION. 145.

33. Perrino, F. W. *et al.* Cooperative DNA Binding and Communication across the Dimer Interface in the TREX2 3′ → 5′-Exonuclease. *J. Biol. Chem.* **283**, 21441–21452 (2008).

34. Kuo-Chen, C. The kinetics of the combination reaction between enzyme and substrate. *Sci. Sin.* **19**, 505–528 (1976).

35. de Silva, U., Perrino, F. W. & Hollis, T. DNA binding induces active site conformational change in the human TREX2 3′-exonuclease. *Nucleic Acids Res.* **37**, 2411–2417 (2009).

36. Fazlieva, R. *et al.* Proofreading exonuclease activity of human DNA polymerase δ and its effects on lesion-bypass DNA synthesis. *Nucleic Acids Res.* **37**, 2854–2866 (2009).

37. R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2021).

38. Onidas, D., Markovitsi, D., Marguet, S., Sharonov, A. & Gustavsson, T. Fluorescence Properties of DNA Nucleosides and Nucleotides: A Refined Steady-State and Femtosecond Investigation. *J. Phys. Chem. B* **106**, 11367–11374 (2002).

39. Schrodinger, L. & DeLano, W. *PyMOL*. (2020).

40. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **54**, 5.6.1-5.6.37 (2016).

41. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

42. Foloppe, N. & MacKerell, A. D., Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**, 86–104 (2000).

43. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

44. Harvey, M. J., Giupponi, G. & Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **5**, 1632–1639 (2009).

45. Hahsler, M., Piekenbrock, M. & Doran, D. dbscan: Fast Density-Based Clustering with R. *J. Stat. Softw.* **91**, 1–30 (2019).

46. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

47. Venables, W.N. & Ripley, B.D. *Modern Applied Statistics with R*. (Springer New York, 2002).