Towards a New Model for the TREX1 Exonuclease

Wayne O. Hemphill[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
[*]Corresponding Authors: wayne.hemphill@colorado.edu & fperrino@wakehealth.edu

ABSTRACT

The 3' → 5' exonuclease, TREX1, functions *in vivo* to degrade DNA and prevent chronic immune activation by the cGAS-STING pathway. Current literature describes TREX1 as a nonprocessive enzyme with little preference for ssDNA versus dsDNA, and several reports propose sources of ssDNA as TREX1's principle *in vivo* substrate. In addition, there exist minimal published 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 in the TREX1 homodimer. Our findings demonstrate that TREX1 is a semi-processive exonuclease with significantly greater affinity for dsDNA than ssDNA, predicted at several orders of magnitude. Furthermore, we find extensively correlated dynamics between TREX1's protomers that involve previously described as well as unstudied substrate-adjacent regions of the TREX1 enzyme. We infer from these and other findings that TREX1 has evolved as a semi-processive exonuclease to degrade a dsDNA species *in vivo*, and we propose that its dimeric structure facilitates a mechanism for efficient binding and catabolism of nicked dsDNA. Overall, these studies identify potentially critical regions of the TREX1 enzyme hitherto unstudied, and they provide compelling evidence that the prevailing model of TREX1 exonuclease activity should be critically re-evaluated.

INTRODUCTION

Three-prime Repair EXonuclease 1 (TREX1) is a 3' → 5' exonuclease[1], which functions *in vivo* to degrade DNA and prevent aberrant DNA-sensing through the cGAS-STING pathway (reviewed in ref[2]). More than sixty *TREX1* mutations have now been associated with a spectrum of autoimmune phenotypes in humans (reviewed in ref[3]), which exhibit dominant and recessive genetics and occur as inherited or *de novo* mutations. *TREX1* disease alleles include missense mutations, insertions, duplications, and frame shifts that locate to positions throughout the 314-amino acid-coding gene. There are general correlates between the positions of *TREX1* mutations and the observed clinical phenotype. Most of the *TREX1* mutations affecting the catalytic domain are recessive and are largely associated with Aicardi-Goutières Syndrome (AGS) or Familial Chilblains Lupus (FCL)[3]. The dominant *TREX1* mutations produce enzyme that competitively inhibits wild-type enzyme activity on bulky dsDNA substrates[4–7]. *TREX1* mutations that cause Retinal Vasculopathy with Cerebral Leukodystrophy (RVCL) exhibit dominant inheritance and are exclusively frame-shift mutations in the C-terminal tail region of the enzyme[3,8,9]. Additional frame-shift mutations in the C-terminal region result in recessive AGS[3]. Interestingly, the phenotype for total knockout of TREX1 appears distinct from that caused by either catalytic deficiency or truncation of the C-terminal tail[5,7,9,10] (reviewed in refs[2,3]). Together, these observations indicate a complex relationship between TREX1 structure and function(s), genetic background, and clinical diagnosis. The Perrino lab[11] and others[12] have also demonstrated that TREX1's activity in hematopoietic cells is the deciding factor in whether or not discernable pathology occurs. However, studies to determine TREX1's biological substrate(s) have mixed conclusions, and have included DNA replication intermediates[13], retroelements[14], and enucleated erythroblast DNA[15] as proposed sources of TREX1 substrate *in vivo*. It's also notable that cGAS appears to be specific for dsDNA[16], and this suggests TREX1's pathologically relevant *in vivo* substrate is dsDNA. Ultimately, the question of TREX1's biological substrate(s) remains an area of active investigation.

Early biochemical investigations of TREX1 established considerable *in vitro* activity rates on single and double-stranded DNA substrates but negligible activity on RNA and RNA-DNA duplexes, and they proposed that TREX1 is a nonprocessive exonuclease[1,17–19]. The only published study on classical kinetics of TREX1[17] suggested that the enzyme is better suited for ssDNA degradation than dsDNA, if any discrepancy exists at all. In fact, these and other studies into TREX1's *in vivo* substrate(s)[13,14] have perpetuated the hypothesis that the relevant *in vivo* substrate of TREX1 is ssDNA, or that ssDNA is at least as relevant a substrate for TREX1 as dsDNA. This is perplexing because if TREX1's relevant biological substrate is indeed ssDNA, then the dsDNA-specific DNA-sensor cGAS[16] would not be expected to drive relevant pathology in TREX1 mutants[20–22]. Furthermore, while *hTREX1*$^{D18N/+}$ patients develop aggressive autoimmune phenotypes[3,5] indicating the D18N mutation is dominant *in vivo*, the recombinant TREX1^{D18N} variant enzymes only recapitulate this dominance *in vitro* on dsDNA (not ssDNA) substrates[5,7,23]. Together, these *in vitro* and *in vivo* D18N data suggest that dsDNA is the pathogenically relevant TREX1 substrate, contrary to the prevailing hypothesis in the literature. Thus, the question of TREX1's biochemical preference between these two DNA species remains highly relevant to its biological function, and it is ultimately unresolved in the literature.

Structurally, TREX1 is a homodimer of identical 314 amino acid protomers. Each of TREX1's composing protomers contain an N-terminal catalytic domain (aa. 1-242) and C-terminal tail region (aa. 243-314)[24]. While the catalytic domain is responsible for TREX1's exonuclease activity[17], the tail region facilitates the enzyme's perinuclear localization in cells[25]. The enzyme's protomers are connected by a highly stable network of hydrogen bonds and hydrophobic interactions, such that the homodimer does not measurably dissociate after initial formation[4]. This obligate homodimer structure of TREX1 is unique among exonucleases and implies a key function in TREX1's biological role. The Perrino and Hollis labs have previously demonstrated that residues from one protomer communicate across TREX1's dimer interface and contribute to activity in the opposing protomer[4,26]. While these studies make it clear that TREX1's homodimer structure is critical for its exonuclease activity, and explain why its existing homodimer structure is obligatory, they don't explain what biological or evolutionary imperative TREX1's homodimer structure is meant to satisfy. After all, other exonucleases exist and most do not function as homodimers[27,28]. This suggests that TREX1's obligate homodimer structure is not a physical necessity for exonuclease activity, but rather serves a purpose unique to TREX1's function *in vivo*. One model supported by existing studies[4,26] is that the presence and/or conformation of substrate in one protomer can affect activity in the opposing protomer. Under this model, potential roles for this inter-protomer communication include necessary coordination of protomers' catalysis, a requirement for both protomers' engagement to properly interact with TREX1's *in vivo* substrate, or as a mechanism to enforce some level of cooperativity. This model is supported by reports about the exonuclease activities of TREX1^{D18N}

heterodimers[5,7], which suggest the D18N heterodimer doesn't degrade dsDNA but degrades ssDNA at ~50% wild-type rate. Ultimately, the deeper purpose of TREX1's homodimer structure remains a mystery.

In this work we use a combination of biochemical and computational approaches to interrogate several of these unresolved properties of TREX1, and to ultimately propose a new model for TREX1 exonuclease function. The relative biochemical affinities of TREX1 for ssDNA versus dsDNA are convincingly demonstrated by a combination of novel exonuclease assays, kinetic modeling, and thermal shift DNA-binding assays, and the residues involved in binding of these respective substrates are interrogated via molecular dynamics simulations. As part of determining these affinities, and to interrogate TREX1's existing designation as a nonprocessive enzyme, a new robust kinetic model of TREX1 exonuclease activity is developed and validated from biochemical data. Finally, we use biochemical assays to explain prior observations that the $TREX1^{D18N}$ heterodimer is specifically inactive on dsDNA, and we employ molecular dynamics simulations to further interrogate inter-protomer communication in the TREX1 homodimer.

RESULTS & DISCUSSION

Comparison of TREX1's ssDNA and dsDNA Binding:

TREX1's catalytic rate (k_{cat}) and Michaelis-Menten (K_m) constants for ss- and dsDNA have been previously reported[17], and suggest comparable affinity for the two substrates (ssDNA: $k_{cat} = 22 \pm 1.1\ s^{-1}$, $K_m = 19 \pm 4.0$ nM; dsDNA: $k_{cat} = 11 \pm 0.77\ s^{-1}$, $K_m = 15 \pm 4.4$ nM). However, a critical review of the literature identifies a number of subsequent studies that report TREX1's ssDNA activity rate ($v_o / [E_T]$) as 5.3 (s^{-1})[24], 1.6 and 1.7 (s^{-1})[6], 1.6 (s^{-1})[7], 6.3 (s^{-1})[4], and 1.4 and 1.9 (s^{-1})[26], all with a ssDNA concentration of 50 nM. By contrast, Equation 1 (see Methods) dictates that given the reported k_{cat} and K_m[17] these values should have been ~16 s^{-1}, which represents a significant observation-theory deviation ($p < 0.001$). In addition, it has been reported for TREX1 that the ssDNA dissociation constant (K_d) is 42 nM[29], implying the reported ssDNA K_m is in direct violation of the $K_d < K_m$ demands of classical kinetics (compare Equations 2 and 3). Collectively, these discrepancies warranted revisiting TREX1's kinetics constants for the two substrates.

To calculate TREX1's dsDNA kinetic constants, we developed a novel absorbance-based exonuclease assay. Briefly, recombinant human $TREX1_{1\text{-}242}^{WT/WT}$ enzyme (hT1) is incubated with a large, nicked plasmid substrate under standard reaction conditions. Then, the reaction is terminated by addition of a small volume of concentrated ethylenediaminetetraacetic acid (EDTA), and the excised deoxynucleoside monophosphates (dNMPs) 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 1 A-B). Then, initial velocities of hT1 reactions were quantified over a range of substrate concentrations and plotted as a Michaelis-Menten curve (Figure 1 C). Regression and interpolation of the curve, followed by adjustment to the moles of hT1, indicated that $k_{cat} = 5.0\ s^{-1}$ and $K_m = 10$ nM. However, our later studies will suggest that only one active site of the hT1 homodimer can degrade this ~10-kb dsDNA plasmid at a time. Based on this, adjustment for the number of TREX1 active sites available for catalysis under saturating substrate conditions would effectively double the apparent k_{cat} to 10 s^{-1}. Comparing these values to those previously reported, we find no significant discrepancies for either dsDNA classic kinetic constant.

To determine TREX1's ssDNA k_{cat} and K_m, a gel-based exonuclease assay was used to determine the exonuclease activities of hT1 under high and low ssDNA concentrations (Figure 2). Then, ssDNA kinetic constants were approximated from these activities using equations derived to relate the kinetic constants to activity rates at two different substrate concentrations (see Methods). The activity rates from Figure 2 and these equations were used with Monte-Carlo simulation methods to approximate a mean and standard deviation for the hT1 ssDNA k_{cat} and K_m. These approximations are provided in Table 1. Interestingly, while the approximated k_{cat} was in good agreement with what was previously reported[17], the approximated K_m was almost an order in magnitude greater. These data, paired with the inconsistencies described at the beginning of this section, demonstrate that TREX1's K_m for ssDNA is much higher than initially reported. By extension, they suggest that TREX1 does not have comparable affinity for dsDNA and ssDNA, but instead prefers dsDNA.

To test relative binding affinities, we performed thermal shift assays with TREX1 apoenzyme, a TREX1-ssDNA complex, or a TREX1-dsDNA complex (Figure 3), utilizing both hT1 and recombinant murine $TREX1_{1\text{-}242}^{WT/WT}$ enzyme (mT1). As expected, TREX1 gained much greater thermostability upon dsDNA versus ssDNA

binding, which is indicative of a stronger association between TREX1-dsDNA than TREX1-ssDNA. It is curious to note, outside the intended scope of these studies, that hT1 and mT1 displayed significantly different melting temperatures as apoenzymes. For both enzymes the T_m is greater than 50°C, so it is unlikely their differing thermostabilities are directly physiologically relevant. However, the discrepancy does suggest the networks of bonds within each enzyme differ much more than previously seemed likely, and it raises the question of what other discrepancies may exist between the two enzymes. Ultimately, these data empirically demonstrate that TREX1 has a significantly greater binding affinity for dsDNA than ssDNA, and they provide further biochemical evidence that TREX1's relevant *in vivo* substrate is dsDNA.

These data also compel interrogation of the residues that contribute to these differential TREX1 affinities between ss- and dsDNA. We performed all-atoms molecular dynamics 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 catastrophic instability (Supplemental 1). RMSF of the Cα in each system identified three potential areas of differential fluctuation in the mT1 backbone (Figure 4 A; see purple, orange, and green marked areas), all of which have the potential for direct interaction with substrate in the active site (Figure 4 B). Of immediate interest is the observation that the flexible loop (aa. Q164-K175) seems differentially impacted by the two substrates, being rigidified during dsDNA occupation of the active site (Figure 4, green markers). This corresponds well with prior literature, which demonstrates mutation of TREX1 flexible loop residues disproportionately reduces TREX1 exonuclease activity on dsDNA[7]. Additional studies with the corresponding region in the highly related three-primer repair exonuclease 2 (TREX2) enzyme suggest that the flexible-loop contributes to DNA-binding[30]. Though this hasn't been recapitulated for TREX1, other reports demonstrate that the $TREX1^{D18N}$ variant enzymes' ability to competitively inhibit wild-type enzyme is significantly reduced upon co-mutation of the flexible loop residues[7], suggesting the flexible loop serves a similar role in TREX1 and TREX2. Thus, our finding corroborates the published literature and suggests TREX1's greater apparent binding affinity for dsDNA is at least partially attributable to the flexible loop.

We also identified a short region (aa. G23-S27) 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 4, purple markers). This region has not been studied in the existing TREX1 literature, and it may represent a key interaction in TREX1 DNA-binding or catalysis. Finally, another previously unstudied region (aa. K75 – S88) 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. It's worth noting that the RMSF analyses would miss changes in the conformation or location of the TREX1 backbone if its relative fluctuations were comparable. Visualization of the conformational free energy maps for the mT1 backbone in each of these systems reveals stark contrast in the distribution of structural states, particularly between the two mT1-DNA complexes, and suggests such oversight could have occurred here (Supplemental 2).

To further interrogate the effects of DNA-binding on mT1 dynamics, residues' phi and psi values across all simulations and systems were subjected 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 5 A-B). As expected, several residues' conformations were significantly altered upon both 4-mer and dsDNA occupation of the active site (Figure 5, blue markers). These included R174, L19, and P191-D193, which are all adjacent to the active site with the potential for DNA interactions. The identification of R174 is not entirely surprising, since reported biochemical data indicates that mutation of the R174 residue has the capacity to impact activity on both substrates[7]. In addition, several other residues' conformations were only altered by dsDNA occupation (Figure 5, red markers). These were S48, S194, and Q209. S194 is closest to the active site, located directly next to the H195 residue previously identified as a residue of interest in TREX1 crystal structures[23,24] and patient data[3]. In addition, it is directly next to three other residues whose conformations were impacted by either substrate's occupation of the active site. Collectively, these data paired with the region's mild RMSF (Figure 4 A) suggest it could serve a key role in TREX1 activity.

Curiously, there were several residues whose conformations were only affected during 4-mer occupation of the active site (Figure 5, green markers). K175 is the only of these residues in any apparent proximity to the active site, and it is part of the flexible loop suspected to contribute to DNA-binding. It's noteworthy that: (1) so many residues' conformations were altered by 4-mer only despite dsDNA overlaying well with ssDNA in prior crystal structures[23,24,31], and (2) residues whose conformational states are only affected by a specific substrate are predominated by residues distant from the active site. The first observation may be explained by additional mT1-dsDNA contacts acting to stabilize regions of the enzyme that are otherwise perturbed by the smaller mT1-ssDNA binding footprint, or it may be otherwise attributable to

differences between the 4-mer and dsDNA ligands' internal structure rigidity or molecular mass. The former hypothesis is supported by the RMSF and thermal shift data, which suggest dsDNA facilitates 'tighter' association within the mT1 structure. It's prudent to note that the phi/psi angle clustering analysis would miss residues whose locations are altered without changes to their conformations, and it's thus possible several residues involved in TREX1 activity were not identified by this strategy.

Ultimately, these analyses are meant to identify potential contributors to TREX1's relatively high dsDNA affinity, and to guide new directions of research regarding TREX1's structure-activity relationships. In that respect, these data are consistent with prior studies' implications that the flexible loop is disproportionately involved in TREX1's dsDNA vs ssDNA binding. Furthermore, they identify previously unstudied regions of the enzyme that may also be involved in TREX1 activity (aa. K75-S88 and P191-S194).

Kinetic Modeling of TREX1 Exonuclease Activity:

A classic kinetic model fails to account for several documented and proposed characteristics of TREX1. First, TREX1 degrades a polymeric substrate upon which catalysis and product formation can occur without a necessary reduction in the relevant metric for substrate concentration: 3'-hydroxyls. Second, TREX1 may possess some semi-processive capacity, while classical kinetics are derived from assumptions of distributive enzyme function. A more robust model would allow more accurate prediction of TREX1 activity, the ability to simulate effects of TREX1 mutations on activity, and estimation from empirical data of other properties like processivity.

The model we propose (Figure 6) allows for intermediate-length DNA species during degradation and incorporates the idea of processivity as a probabilistic adjustment to the catalytic rate constant. Our model also assumes equivalent kinetic constants for practically all length substrates and irreversible catalysis, which is generally consistent with TREX1 biochemistry[1,17,18,32]. Under this model, the number of nucleotides excised by TREX1 after initial enzyme-substrate complexing is given by a discrete probability distribution function (Equation 18). More broadly, a reaction is described by a system of nonlinear differential equations (Equations 19.1-5), which have no analytical solution but can be approximated numerically.

First, we estimate the kinetic constants that parameterize these equations. From our model's function for turnover rate (Equation 19.2) it is evident that in the absence of any enzyme-product complexing and under saturating conditions the turnover rate would reduce to $k_2[E_T]$, which is classically[33] V_{max}. Thus, the k_2 parameters for both dsDNA and ssDNA have been empirically determined by our classical kinetics studies (Table 2). Next, under our model an equation describing steady-state binding in the absence of catalytic activity would remain unchanged from classic proposals[33], so the previously reported ssDNA K_d value for TREX1[29] gives the parameter relationship $K_d = \frac{k_{-1}}{k_1}$ (Equation 24). The classical relationship of $K_m = \frac{k_{-1}+k_2}{k_1}$, where K_m is the substrate concentration at which $[P]' = \frac{1}{2}k_2[E_T]$, does not hold true for our model when $\Psi \neq 0$. However, since ssDNA degradation patters for TREX1 are not consistent with complete processivity (compare Figure 2 to examples in ref[34]), using the classical relationship could provide initial values for k_1 and k_{-1}. Under these assumptions, we derive equations for k_1 and k_{-1} (Equations 21.1-2) whose solutions from existing TREX1 data (Table 1 and ref[29]) provide initial values for ssDNA. However, the absence of a dsDNA K_d for TREX1 prohibits a similar solution of initial values for dsDNA.

To approximate the model's ssDNA kinetic constants, we compared the concentration of excised nucleotides produced by reactions in Figure 2 ([hT1] = 10 pM, [S] = 15 nM) to the model's predictions under various processivity factors (Supplemental 3 A), using our initial estimations of the other kinetic parameters. Next, for various alternative processivity constant values we found solutions to Equation 20 that best fit the experimental data (Supplemental 3 B-E). During these reaction simulations the kinetic constants associated with product inhibition were set to zero, since experimentally observed product inhibition is negligible. Then, the banding patterns from the ssDNA degradation gels associated with the training reactions (Figure 2, [hT1] = 10 pM & [S] = 15 nM) were quantified and compared to predictions by our model under each of the various optimized kinetic constant settings (Figure 7). Strikingly, the kinetic constant values associated with $\Psi = 0.75$ gave a great fit to the experimental data, and they are reported in Table 2.

For dsDNA, the model was expanded to account for competitive effects by other TREX1 variants (Figure 8). Like before, reactions under this model are described by a system of nonlinear differential equations (Equations 23.1-8), whose solutions are approximated numerically. Mutations to TREX1's metal coordinating residues, like D18 and D200, produce a catalytically-inactive enzyme that maintains DNA-binding capability[7,26,29,35,36]. In addition, these enzyme variants are reported to competitively inhibit wild-type enzyme in

a reaction with dsDNA (but not ssDNA) substrates[7,26,29,35,36]. Since the D18N-like mutants have their catalytic activity specifically abolished, their observed substrate-specific competition with wild-type enzyme must be a consequence of differences in TREX1's binding rate constants for dsDNA versus ssDNA. Thus, D18N-like variants are a convenient way to ascertain $TREX1^{WT}$ binding rate constants. For the purposes of modeling, we generally represent TREX1 variants like human $TREX1_{1\text{-}242}^{D18N/D18N}$ (hT1-DN) as enzymes with zero catalytic activity ($k_2 = 0$) but otherwise identical kinetic constants, consistent with prior biochemical studies on these mutants.

To determine TREX1's dsDNA kinetic constants, we performed dsDNA degradation assays on low concentrations of nicked dsDNA plasmid with equimolar concentrations of hT1 (Supplemental 4, black data points) or a 1:1 mix of hT1:hT1-DN (Supplemental 4, red data points). We also simulated these reactions with our TREX1 dsDNA kinetic model with iterative adjustments to the parameter values until they best reproduced these data. Parameters were initially constrained to $\Psi_{ds} \geq \Psi_{ss}$, where Ψ_{ds} and Ψ_{ss} are the dsDNA and ssDNA processivity factors, respectively. During parameter optimization the ssDNA k_1 and k_{-1} values were used as initial values for dsDNA, and k_1 increased until the predicted wild-type degradation rate best matched observation, then k_{-1} decreased until the predicted inter-enzyme competition best matched observation. Upon optimizing the model's other parameters at several Ψ_{ds} values across the constrained range (Supplemental 4), it is evident that $\Psi_{ds} \geq 0.90$ is incapable of a desirable fit with the empirical data. Thus, our model of TREX1 dsDNA exonuclease activity suggests $0.75 \leq \Psi_{ds} \leq 0.85$ if we assume TREX1 is at least as processive on dsDNA as on ssDNA. Our approximated dsDNA kinetic constants are reported in Table 2. We also optimized the other model parameters to fit the empirical data with $\Psi_{ds} = 0$ (data not shown), and we found good model-observation agreement if $k_1 = 5.0(10^9)\ M^{-1}s^{-1}$ and $k_{-1} = 0.30\ s^{-1}$. Thus, it is evident that for any value of the processivity factor, our model predicts significantly different binding rate constants for ss- versus dsDNA.

Finally, we used Equation 18 with the approximated ssDNA kinetic constants and $\Psi_{ds} = 0.80$ dsDNA kinetic constants in Table 2 to generate a probability distribution for the number of consecutive nucleotides our model theorizes TREX1 should excise from dsDNA versus ssDNA substrates (Figure 9). Surprisingly, our model suggests that TREX1 does not commonly excise more than ~10 nucleotides from either substrate during a single association event. Given the high Ψ values, this is more than likely attributable to TREX1's k_{-1} values. Consequently, it is unsurprising that our model also predicts TREX1 excises more nucleotides from dsDNA than ssDNA during their average semi-processive events, since our studies suggest a much smaller k_{-1} value for dsDNA than ssDNA (Table 2).

Notably, our proposed model has several implications. First, comparison of our model's predictions to multiple empirical data suggest TREX1 is quite capable of semi-processive cleavage events, and that continuing to identify it as 'nonprocessive' is misleading of its capabilities as an exonuclease. Next, our earlier studies convincingly demonstrated a tighter association for hT1-dsDNA than hT1-sDNA, and our model both affirms this observation and quantifies the magnitude of the discrepancy in binding affinities. Strikingly, while the reported and modeled TREX1 ssDNA $K_d = 42$ nM, the range of binding rate constants across all Ψ values reconcilable with our model suggests dsDNA $K_d \approx$ 20-100 pM. Thus, our model predicts from empirical data that TREX1 should have ~1000-fold greater binding affinity for dsDNA than ssDNA. If true, this is the most compelling biochemical evidence to date that TREX1 has primarily evolved to degrade double-stranded DNA, and by extension that dsDNA is its relevant *in vivo* substrate.

We conclude this section by discussing some of our models' potential limitations. Primarily, it does not fully account for TREX1's dimeric nature. Consequently, robust simulation of TREX1 mutant heterodimers is not possible with our model. Furthermore, if both protomers could be bound simultaneously and displayed cooperativity, then our model does not account for their differential binding or catalysis. Fortuitously, our studies largely avoid applicability of these limitations. First, only homodimers were used for the empirical data our model was optimized from. Next, prior studies refute cooperativity during TREX1 ssDNA-binding[29], affirming that our model's ssDNA kinetic constants are representative of TREX1 protomers regardless of the opposing protomers' binding states. Finally, cooperativity has not been similarly ruled out for TREX1 dsDNA-binding, but subsequent studies within this work suggest that the nicked plasmid substrate used to optimize our model's dsDNA kinetic parameters limits binding to a single protomer in each TREX1 dimer. Thus, our model's dsDNA kinetic parameters are representative of TREX1 protomers whose opposing protomer is unbound.

Inter-Protomer Communication in the TREX1 Homodimer:

Early biochemical investigations into the exonuclease activities of $TREX1^{D18N}$ variants[5,7,35,36] 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, a critical review of the literature reveals that the dsDNA activity for the heterodimers was only ever tested under specific substrate conditions: sub-enzyme concentrations of a bulky nicked plasmid. Thus, another model to explain the D18N heterodimer activity is that the nicked plasmid substrate's bulk caused such significant steric hindrance that both protomers of TREX1 could not feasibly bind substrate at the same time. Alternatively, if TREX1 has a sufficiently small dsDNA K_d, then the sub-enzyme plasmid concentrations could have allowed the D18N protomers to bind all available substrate. If either of these models were true then the D18N heterodimer would be active in the presence of saturating concentrations of a small dsDNA substrate, but if the coordinated catalysis model were true then activity should remain unaffected by the size and concentration of the substrate.

We directly interrogated this by quantifying the $TREX1^{D18N}$ variants' exonuclease activities at low concentrations of the original bulky plasmid substrate (Figure 10 A, '[S] < K_m') and at saturating concentrations of a 30-bp dsDNA (Figure 10 B, '[S] >> K_m'). We observed partial activity under the latter conditions, consistent with the steric hindrance and low-K_d theories for heterodimer inactivity. However, it is curious to note that if the wild-type protomer activity in the heterodimer was completely uncorrelated from the D18N protomer, then it should have displayed ~50% wild-type enzyme activity like the equal mix of WT and D18N homodimers (Figure 10 B-C). Instead, it was only ~20% active compared to wild-type enzyme. It's plausible that the 30-bp dsDNA still causes some degree of steric hindrance that impedes substrate binding in the opposing protomer, but this observation would also be consistent with regulation of activity across the dimer interface. The latter hypothesis seems to require catalytic regulation, as opposed to just binding affinity regulation, since the substrate concentration was likely saturating (~100-fold the wild-type K_m). However, we cannot rule out negative cooperativity that reduced the opposing protomers' binding affinities too low for their active sites to saturate at the utilized substrate concentration. Ultimately, these data refute a model where dsDNA occupation of one protomer enforces total suppression of catalysis across the dimer interface, and they suggest the apparent catalytic inactivity of the D18N heterodimer reported in prior studies could be attributable to the bulky dsDNA substrate used. This possibility of steric hinderance by bulky dsDNA substrates raises other questions. If only one protomer can bind/degrade a bulky dsDNA substrate at a time, what biological implications might that have? If TREX1's relevant dsDNA substrate(s) *in vivo* is sterically hindering (ex. large segments of chromatin), then it's puzzling that the enzyme would evolve with an obligate homodimer structure incapable of fully utilizing both protomers on that substrate. By contrapositive, it's possible these data indicate that TREX1's *in vivo* dsDNA substrate(s) is significantly smaller than nicked plasmid, perhaps like small fragments of host DNA.

To further interrogate coordinated activity between TREX1 protomers, we performed all-atoms molecular dynamics simulations with mT1 as apoenzyme, bound to 4-mer oligo, and bound to dsDNA. The 4-mer was bound in both active sites, while the dsDNA was bound in a single active site ('Protomer-A') to facilitate study of the impacts on the unoccupied opposing protomer. RMSD comparisons of mT1 against initial conditions shows that all simulations quickly equilibrate without indications of catastrophic instability (Supplemental 1). An RMSF comparison between the mT1 and mT1-dsDNA systems' α-carbons in each unbound protomer ('Protomer-B') revealed differential dynamics in the mT1 backbone (Figure 11 A). This implies that the presence of dsDNA in one active site impacts the structure of the other. The regions affected are the same identified earlier as being impacted during substrate occupation of the active site, which suggests the regions of the unbound protomer affected by substrate binding in the opposing protomer are also relevant to mT1 DNA-binding. Furthermore, conformational free energy maps of the unbound protomers' backbones revealed modestly different distributions of structural states between systems (Supplemental 8). Collectively, these data suggest coordinated activity between TREX1 protomers.

Like before, we also subjected the unbound protomer residues' phi and psi values 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 12 A). Curiously, most residues with differential conformation proportions were located on the flexible loop (Figure 12 B). The other residues affected were aa. 192-194, and so the residues with altered conformations versus altered backbone dynamics seem to overlap significantly. In general, the residues' conformation data are consistent with the same two hypotheses as the RMSF data: coordinated activity or artifactual variation between simulation systems. It's worth emphasizing that the mT1 structures of every simulation in every system were identical at the start of each simulation; the respective DNA was merely super-imposed on the active site(s) via mT1 backbone-alignment with the relevant published crystal co-structures. This countermeasure, paired with the use of multiple different randomized simulations per system, makes the artifactual variation hypothesis much less likely.

As a final interrogation of coordinated activity, correlation analyses were performed on all α-carbons in both protomers of the mT1 and mT1:4-mer systems (Figure 13 A-D). Several region pairs with correlated fluctuations across the dimer interface were identified. First, the alpha helices (Figure 13 A-E, region '1') and the beta strands (Figure 13 A-D and G, region '3') at the dimer interface had positively correlated fluctuations in both the apoenzyme and enzyme-substrate systems. This was expected based on prior structural studies[4,23,24], and so contributes to the credibility of the other pairs' correlations. Interestingly, the other region pairs were only correlated when substrate occupied the protomers' active sites. There was positively correlated fluctuation between the flexible loop in one protomer and the α6 helix of the opposing protomer (Figure 13 A-D and F, region '2'), and positively correlated fluctuation between the flexible loop in one protomer and aa. 79-93 in the opposing protomer (Figure 13 A-D and F, region '4'). In addition, negatively correlated fluctuations were discovered between aa. 79-93 in one protomer and the dimer-interface β-strands, dimer-interface α-helices, and aa. 22-31 on the opposing protomer.

These are particularly important observations for several reasons. First, *two* of these correlated pairs directly link the flexible loop of one protomer, which has been previously linked to DNA-binding, to previously unstudied substrate-adjacent residues in the opposing protomer. Notably, both substrate-adjacent regions (α6 helix and aa. 79-93) were implicated by earlier studies in this chapter as regions impacted during substrate-occupation of the mT1 active site. One of these regions (aa. 79-93) was also negatively correlated to the β-strands and α-helices at the dimer interface, and to another active site-adjacent region in the opposing protomer that was implicated in the earlier studies (Figure 4, purple area). Thus, the concerted implication of these observations is that DNA-binding in one TREX1 protomer may impact TREX1 regions across the dimer interface and regulate activity in the opposing protomer.

We favor cooperativity by the TREX1 enzyme as the most likely explanation for this apparent coordination across the dimer interface. Prior fluorescence anisotropy experiments interrogating TREX1 binding to ssDNA yielded non-sigmoidal binding curves with Hill coefficients of ~1, which is inconsistent with cooperativity during TREX1 ssDNA-binding[29]. However, inspection of the binding curves for similar published[26] fluorescence anisotropy experiments interrogating TREX1 dsDNA-binding show binding curves that don't achieve half-maximal anisotropy signal, the traditional measure of K_d, until a concentration of hT1 over 20-fold greater than the reported ssDNA K_d[29]. Since our thermal shift data convincingly demonstrated a significantly tighter binding for dsDNA than ssDNA, it's unlikely this metric represents the dsDNA K_d. These anomalies in the published anisotropy data could be explained by cooperativity in the TREX1 homodimer[37]. In addition, the highly-related TREX2 exonuclease demonstrates cooperativity in a manner dependent on its corresponding flexible loop[30]. Furthermore, the suggestion by our molecular dynamics simulations (Figure 4) and published biochemical data[7] that the flexible loop disproportionately interacts with ds- versus ssDNA may well explain why TREX1 cooperativity would be restricted to dsDNA-binding, which is additional evidence that TREX1's relevant *in vivo* substrate is dsDNA.

A subsequent question is whether substrate binding is sensed by the resident flexible loop and cooperativity enforced through these unstudied regions in the opposing protomer, or if the opposite direction of influence occurs. From our studies into the effects of dsDNA-binding on the unbound opposing TREX1 protomer, we see that residue conformations are primarily affected on the flexible loop (Figure 12). Thus, it seems more likely that these unstudied regions near the TREX1 active site communicate the presence of DNA in their resident protomer across the dimer interface to the flexible loop, which then regulates DNA-binding.

Based on the preponderance of data in the literature and these studies, we propose an interesting *in vivo* role for TREX1's obligate homodimer structure. We suspect the TREX1 homodimer's flexible loops each interact nonspecifically with a respective strand on *in vivo* dsDNA substrates to facilitate simultaneous scanning of both strands for exposed 3'-hydroxyls. Then, recognition and active site-loading of a 3'-terminal nucleotide promotes negative cooperativity that causes the opposing flexible loop to release the non-scissile strand so it can initiate efficient catabolism of the identified scissile strand. Under this model, TREX1's well-established preference for nicked dsDNA[1,17,18,31,38], single-strand mechanism for catabolizing dsDNA substrates[31], implied[26,30] flexible loop-mediated 'scanning', and incredibly fast kinetics[17] are all quite sensical. This model would also explain why only one TREX1 protomer might catabolize bulky dsDNA substrates at a time, as our earlier data suggested. Finally, this model's implicit requirement that TREX1's proposed cooperativity on dsDNA be negative in nature is consistent with our earlier D18N heterodimer data and published fluorescence polarization binding curves for the TREX1-dsDNA interaction[26].

Ultimately, we have demonstrated correlated dynamics between protomers in the TREX1 homodimer. Based on our and others' data, we propose a model where TREX1's obligate homodimer structure has evolved

with flexible loop-mediated negative cooperativity to facilitate efficient scanning along both strands of dsDNA substrates. These features would likely contribute to TREX1's ultra-fast kinetics.

Conclusions:

The 3' → 5' exonuclease, TREX1, functions *in vivo* to degrade DNA and prevent chronic immune activation by the cGAS-STING pathway. Current literature largely describes TREX1 as a nonprocessive enzyme with little preference for ssDNA versus dsDNA, and several reports propose sources of ssDNA as TREX1's principle *in vivo* substrate. In addition, there were minimal reports to interrogate or rationalize TREX1's obligate homodimer structure. By contrast, our studies demonstrate that TREX1 is a semi-processive exonuclease, with significantly greater affinity for dsDNA than ssDNA that is predicted to be several orders in magnitude. This considerable biochemical preference likely reflects the nature of its relevant *in vivo* substrate. Furthermore, we find extensively correlated dynamics between TREX1 protomers involving previously described as well as unstudied regions of the TREX1 enzyme, all of which have the potential to directly interact with substrate in their resident protomers. We infer from these and other data that TREX1's obligate homodimer structure has evolved as part of a mechanism to efficiently scan for exposed 3'-hydroxyls on dsDNA and facilitate rapid biological substrate catabolism. Overall, these studies identify potentially critical regions of the TREX1 enzyme unstudied to date, and they provide compelling evidence that the prevailing model of TREX1 exonuclease activity should be critically re-evaluated.

METHODS

Overexpression and Purification of TREX1 Enzymes:

We have published a very detailed protocol for purifying TREX1 enzymes[35], 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* BL21(DE3) Rosetta 2 cells (Novagen) for overexpression. Homodimers were purified via amylose column chromatography followed by phosphocellulose (p-cell) column chromatography. Heterodimers were purified via amylose column chromatography followed by nickel column chromatography then 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 canonical equations for the initial velocity of uninhibited enzyme[33], 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 classical[33] equations we obtain Equations 2 and 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. Small sampling sizes were used to limit skewing by severe outliers 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 3.5. We used the substrate concentrations outlined in the 'Polyacrylamide Gel ssDNA Assay' methods section of Chapter 3 to define the $[S]_1$ and $[S]_2$ parameters. These vectors and parameters were applied pairwise to Equations 7 and 4.8 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. R scripts for these analyses are available upon request.

P values for observation-theory comparisons were calculated by one-sample two-tailed Student's t-test. Test statistics were calculated via Equation 9, and degrees of freedom via Equation 10. Test statistics and degrees of freedom were used with the 'pt()' function in R to calculate one-tailed probabilities, which were doubled to achieve two-tailed p-values. Sample sizes were calculated as the number of mean activities cited from the literature. We use t, v, μ, μ_0, σ, and n to indicate the test statistic, degrees of freedom, empirical relative activity mean, theoretical relative activity mean, empirical relative activity standard deviation, and sample size, respectively.

Equation 9: $t = \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

Equation 10: $v = n - 1$

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[39] extinction coefficient for dAMP to produce Equation 11. 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[39] DNA nucleotides' extinction coefficients, 1160 cm^{-1} M^{-1}.

Equation 11: $$[dNMP] = \frac{1.65 * A_{260} + 7.19 * 10^{-3}}{\varepsilon_{dNMP}}$$

For the activity assays, 50 µL reactions were prepared containing 30 nM hT1, 20 mM Tris (pH 7.5), 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 ~10kb 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 11, 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.

Thermal Shift Assays:

For the assay, 100 µL reactions were prepared containing 0.5 mg/mL TREX1, 20 mM Tris (pH 7.5), 5 mM $CaCl_2$, 2 mM dithiothreitol (DTT), 5X SYPRO Orange dye (Promega), 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 v3.6.1, 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, and the scripts are available upon request.

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 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) to remove nucleic acid and ions. The modified structure was then used as a template for MODELLER to model in the flexible loops missing from the original structure. This produced the mT1 model. This was repeated, but the R174A and K175A mutations were additionally modeled into each protomer. This produced the mT1-RK model. The original mTREX1 published structure was loaded into PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) 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 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) 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. The Python scripts used for MODELLER are available upon request.

Each of the models was converted to pdb and psf files with the 'autopsf' functionality in VMD (v1.9.4), using the CHARMM 'top_all27_prot_lipid_na.inp' force field parameters. 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) 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. Simulation scripts are available upon request.

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, which are available upon request.

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 and is available upon request.

Calculating RMSD:

The RMSD values for each simulation were calculated via Equation 12. 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 12: $$RMSD_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\boldsymbol{r}_j(t_i) - \boldsymbol{r}_j(t_1)\right)^2}$$

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, and the scripts are available upon request.

Calculating RMSF:

The RMSF values in each simulation were calculated via Equation 13. 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 13: $$RMSF_i = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\boldsymbol{r}_i(t_j) - \boldsymbol{r}'_i\right)^2}$$

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, and the scripts are available upon request.

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 14-16. This allowed seamless clustering for residue conformations that had phi (Ψ) or psi (Φ) values spanning the -180º/180º transition.

Equation 14: $x = \cos(\Psi)\sin(\Phi)$
Equation 15: $y = \sin(\Psi)\sin(\Phi)$
Equation 16: $z = \cos(\Phi)$

For each residue, every third time frame of every simulation of every system was clustered by these projected coordinates using the 'dbscan::hdbscan()' function 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 v3.6.1. P-values from these comparisons were converted to q-values by adjusting for false discovery rate with the Benjamini-Hochberg method via the 'p.adjust()' function in R v3.6.1. 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, and scripts are available upon request.

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 v3.6.1. Each resulting principal component's (PC) proportion of variance (POV) was calculated via Equation 17, where σ_i^2 is the variance for the ith principal component and N is the number of principal components.

Equation 17: $$PV_i = \frac{\sigma_i^2}{\sum_{j=1}^{N} \sigma_j^2}$$

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()' function in R v3.6.1. Density functions for each system were visualized as contour plots using the 'contour()' function in R v3.6.1. All calculations and statistics for this section were performed in R v3.6.1, and scripts are available upon request.

Polyacrylamide Gel ssDNA Assay:

We have published a quite detailed protocol[35] 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), 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 400µL 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[35] 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.

Kinetic Modeling Equations:

For the proposed model in Figure 6, 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 *[□□]* is concentration of a complex of the respective reactants.

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: $$P_\Psi(N) = \begin{cases} \frac{k_{-1}}{k_{-1}+k_2}, & N = 0 \\ \frac{k_2(1-\Psi)(k_2\Psi)^{N-1}}{(k_{-1}+k_2)^N} + \frac{k_{-1}(k_2\Psi)^N}{(k_{-1}+k_2)^{N+1}}, & N > 0 \end{cases}$$

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

For Equations 19.1-5, $S_{\beta:n}$ 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 6 provides helpful context.

Equation 19.1: $$[E]_t' = k_{-3}[EP]_t + k_{-1}[ES_{\beta:1}]_t + k_2(1-\Psi)\left[ES_{\beta:2}\right]_t + k_2[ES_1]_t - k_3[E]_t[P]_t - k_1[E]_t\left[S_{\beta:1}\right]_t$$

Equation 19.2: $$[P]_t' = k_{-3}[EP]_t + k_2[ES_{\beta:1}]_t - k_3[E]_t[P]_t$$

Equation 19.3: $$[S_n]_t' = k_{-1}[ES_n]_t + k_2(1-\Psi)[ES_{n+1}]_t - k_1[E]_t[S_n]_t$$

Equation 19.4: $$[ES_n]_t' = k_1[E]_t[S_n]_t + k_2\Psi[ES_{n+1}]_t - (k_{-1}+k_2)[ES_n]_t$$

Equation 19.5: $$[EP]_t' = k_3[E]_t[P]_t - k_{-3}[EP]_t$$

For Equations 20-22, K_m and K_d are the classically defined constants[33] and all other notation is previously defined.

Equation 20: $$k_{-1} = k_1 K_d$$

Equation 21.1: $$k_{-1} = \frac{k_2 K_d}{K_m - K_d}.$$

Equation 21.2: $$k_1 = \frac{k_2}{K_m - K_d}$$

Equation 22: $$k_1 = \frac{k_{-1}}{K_m} + \frac{k_2}{K_m}$$

For Equations 23.1-8, *[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 8 provides useful context.

Equation 23.1: $$[E]_t' = k_{-3}[EP]_t + k_{-1}[ES_{\beta:1}]_t + k_2(1-\Psi)\left[ES_{\beta:2}\right]_t + k_2[ES_1]_t - k_3[E]_t[P]_t - k_1[E]_t\left[S_{\beta:1}\right]_t$$

Equation 23.2: $$[M]_t' = k_{-c}[MP]_t + k_{-a}[MS_{\beta:1}]_t + k_b(1-\Phi)\left[MS_{\beta:2}\right]_t + k_b[MS_1]_t - k_c[M]_t[P]_t - k_a[M]_t\left[S_{\beta:1}\right]_t$$

Equation 23.3: $$[P]_t' = k_{-3}[EP]_t + k_{-c}[MP]_t + k_2[ES_{\beta:1}]_t + k_b\left[MS_{\beta:1}\right]_t - k_3[E]_t[P]_t - k_c[M]_t[P]_t$$

Equation 23.4: $$[S_n]_t' = k_{-1}[ES_n]_t + k_{-a}[MS_n]_t + k_2(1-\Psi)[ES_{n+1}]_t + k_b(1-\Phi)[MS_{n+1}]_t - k_1[E]_t[S_n]_t - k_a[M]_t[S_n]_t$$

Equation 23.5: $$[ES_n]_t' = k_1[E]_t[S_n]_t + k_2\Psi[ES_{n+1}]_t - (k_{-1}+k_2)[ES_n]_t$$

Equation 23.6: $$[MS_n]_t' = k_a[M]_t[S_n]_t + k_b\Phi[MS_{n+1}]_t - (k_{-a}+k_b)[MS_n]_t$$

Equation 23.7: $$[EP]_t' = k_3[E]_t[P]_t - k_{-3}[EP]_t$$

Equation 23.8: $$[MP]_t' = k_c[M]_t[P]_t - k_{-c}[MP]_t$$

Reaction Simulations for Kinetic Model:

We provide detailed accounts of the parameter definitions for every reaction, which should allow reproduction of these simulations. A comprehensive list of simulated reactions is shown in Supplemental 5, and their associated parameter values are provided in Supplemental 6. 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 in Supplemental 7.

Fluorescence-Based dsDNA Exonuclease Assay:

We have published a detailed protocol[35] 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), 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 ~10kb 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). 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 24.1-2 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, and the scripts are available upon request.

Equation 24.1: $$\widetilde{C_{ij}} = \sum_{\beta=1}^{N} \frac{(\vec{r_i}^{\beta} - \langle\vec{r_i}\rangle)\cdot(\vec{r_j}^{\beta} - \langle\vec{r_j}\rangle)}{N}$$

Equation 24.2: $$C_{ij} = \frac{\widetilde{C_{ij}}}{\sqrt{\widetilde{C_{ii}}\widetilde{C_{jj}}}}$$

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Allergy and Infectious Diseases (AI116725), National Institute of General Medical Sciences (GM110734), Lupus Research Alliance, Wake Forest Innovations, Comprehensive Cancer Center of Wake Forest University National Cancer Institute (Center

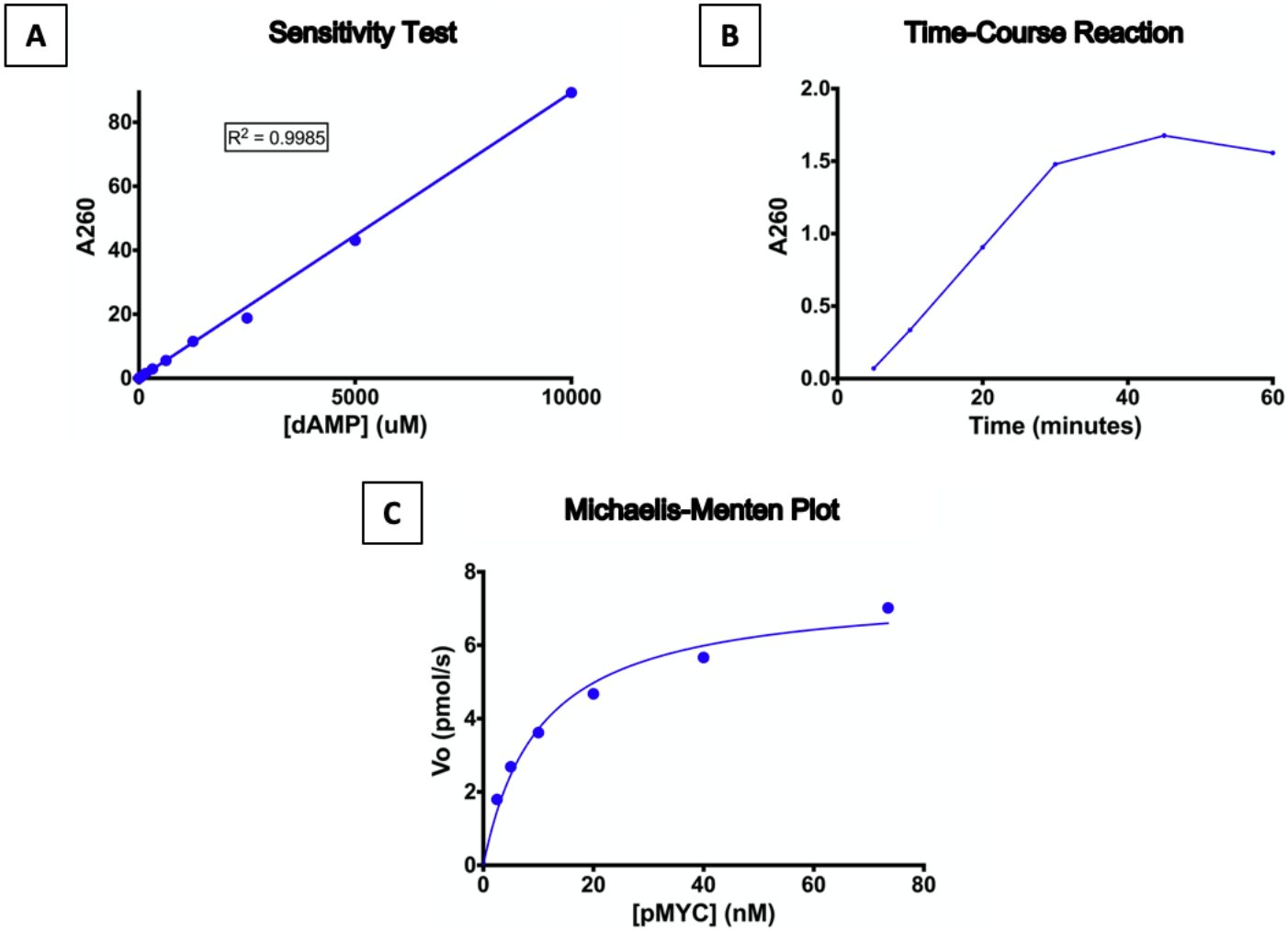
Support Grant, P30CA012197), NIH NRSA Predoctoral Fellowships (T32-GM095440 & T32-AI007401), a Sandy Lee Cowgill Memorial Scholarship, an Artom Fellowship, and a Scott Family Fellowship. We'd also like to thank the WFU high-performance computing team for access to the DEAC cluster and technical support.

AUTHOR CONTRIBUTIONS

W.O.H. performed all described methodologies, studies, and analyses, interpreted data, and wrote the manuscript. 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.

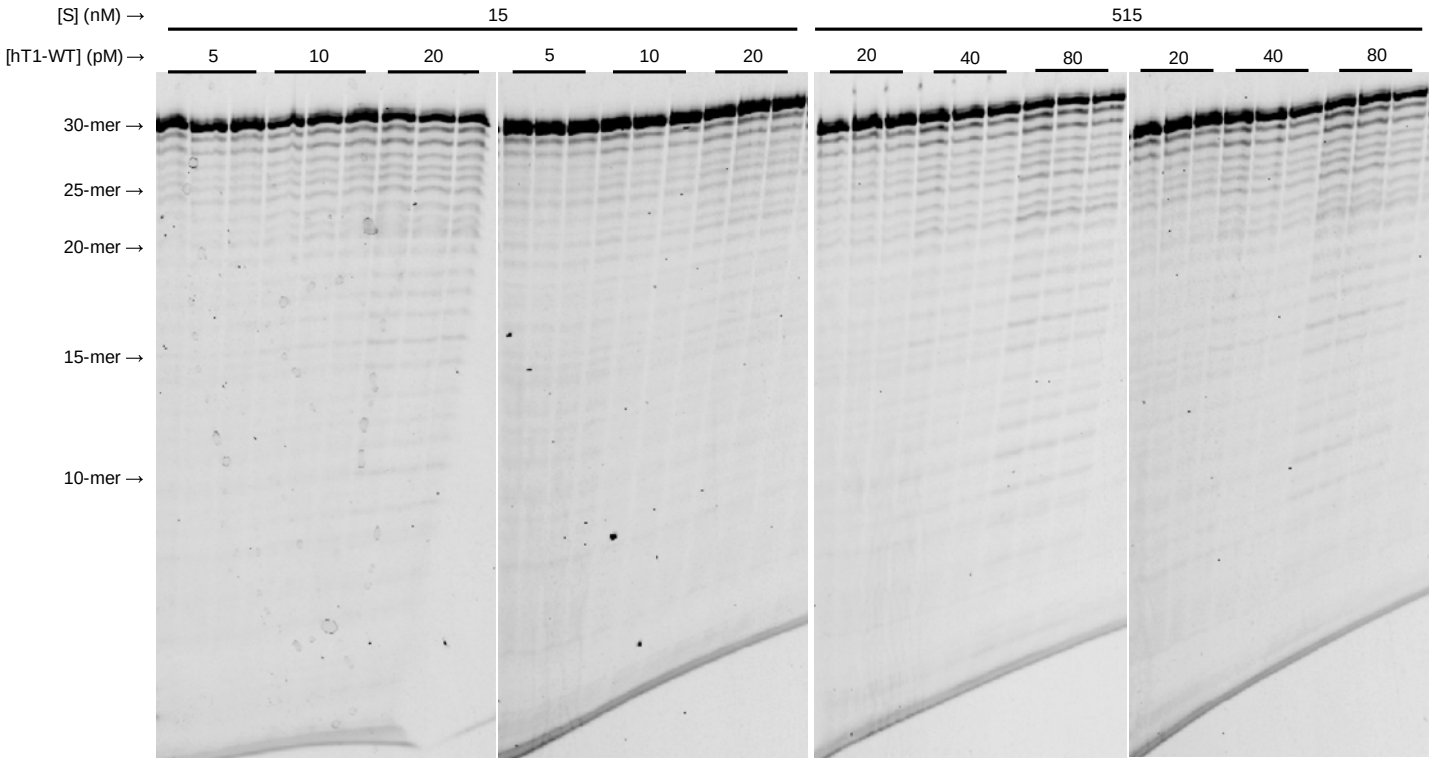

A
Sensitivity Test
R2 = 0.9985
A260
[dAMP] (uM)
0
20
40
60
80
0
5000
10000
B
Time-Course Reaction
A260
Time (minutes)
0.0
0.5
1.0
1.5
2.0
0
20
40
60
C
Michaelis-Menten Plot
Vo (pmol/s)
[pMYC] (nM)
0
2
4
6
8
0
20
40
60
80

<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 assay maintained linearity to at least A_{260} ~ 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 '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.

TREX1 Classical Kinetic Constants

	ssDNA			dsDNA		
	k_{cat} (s^{-1})	K_m (nM)	k_{cat} / K_m (s^{-1} M^{-1})	k_{cat} (s^{-1})	K_m (nM)	k_{cat} / K_m (s^{-1} M^{-1})
This Study	16 ± 3.2	120 ± 57	1.5 ± 0.40 (x10^8)	10	10	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)

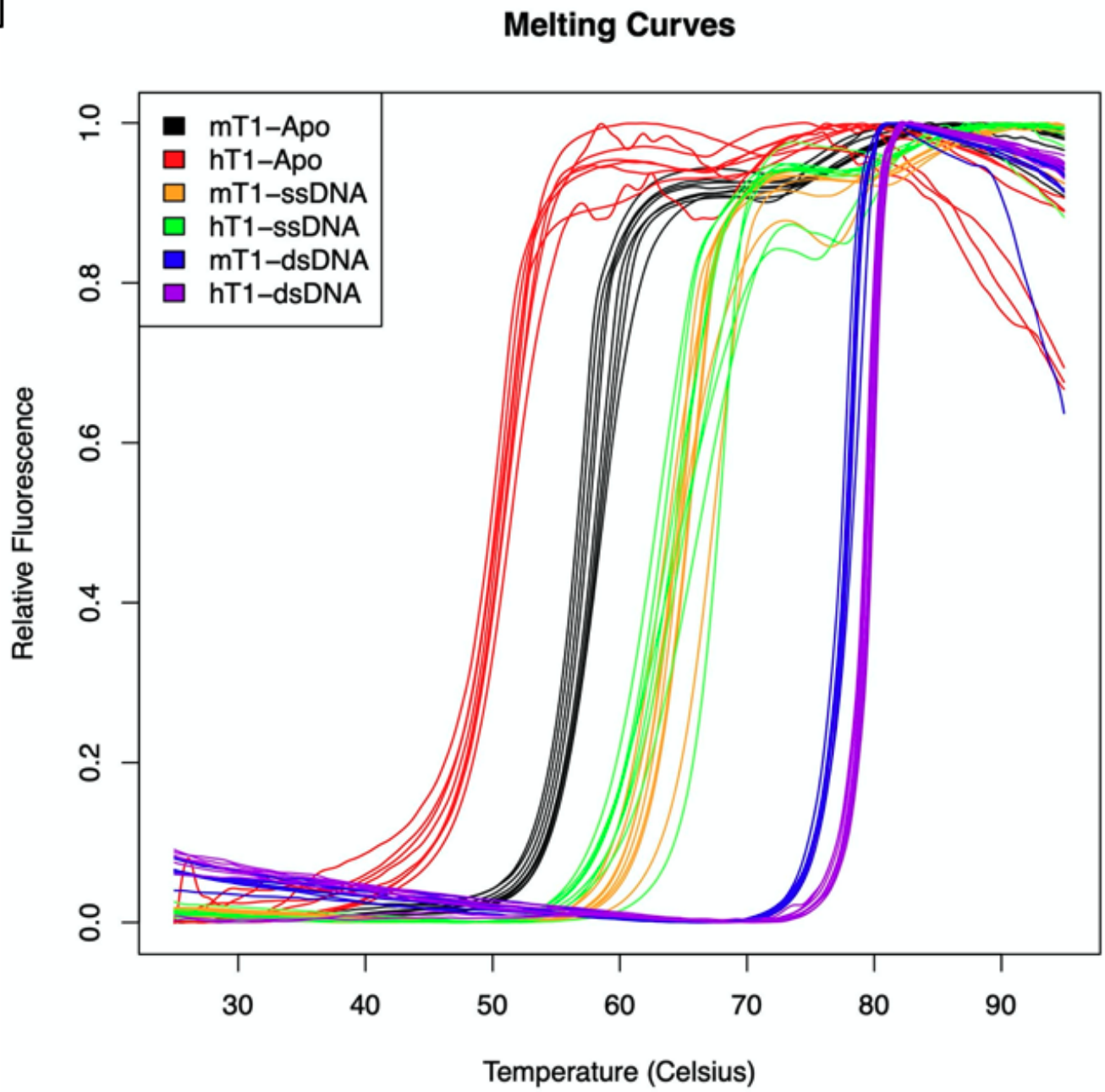
<u>Table 1</u>: **Classical Kinetic Constants for TREX1.** 'Reported' values are taken from prior work[17]. dsDNA values from this study were calculated empirically using the described absorbance-based exonuclease assay. ssDNA values from this study were approximated via Monte-Carlo simulation using the ssDNA activity rates of wild-type enzyme that were calculated in Figure 2. Values shown indicate $\mu \pm \sigma$, except for this study's dsDNA values and the reported specific activities. Calculations and simulations 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 2</u>: **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).

A

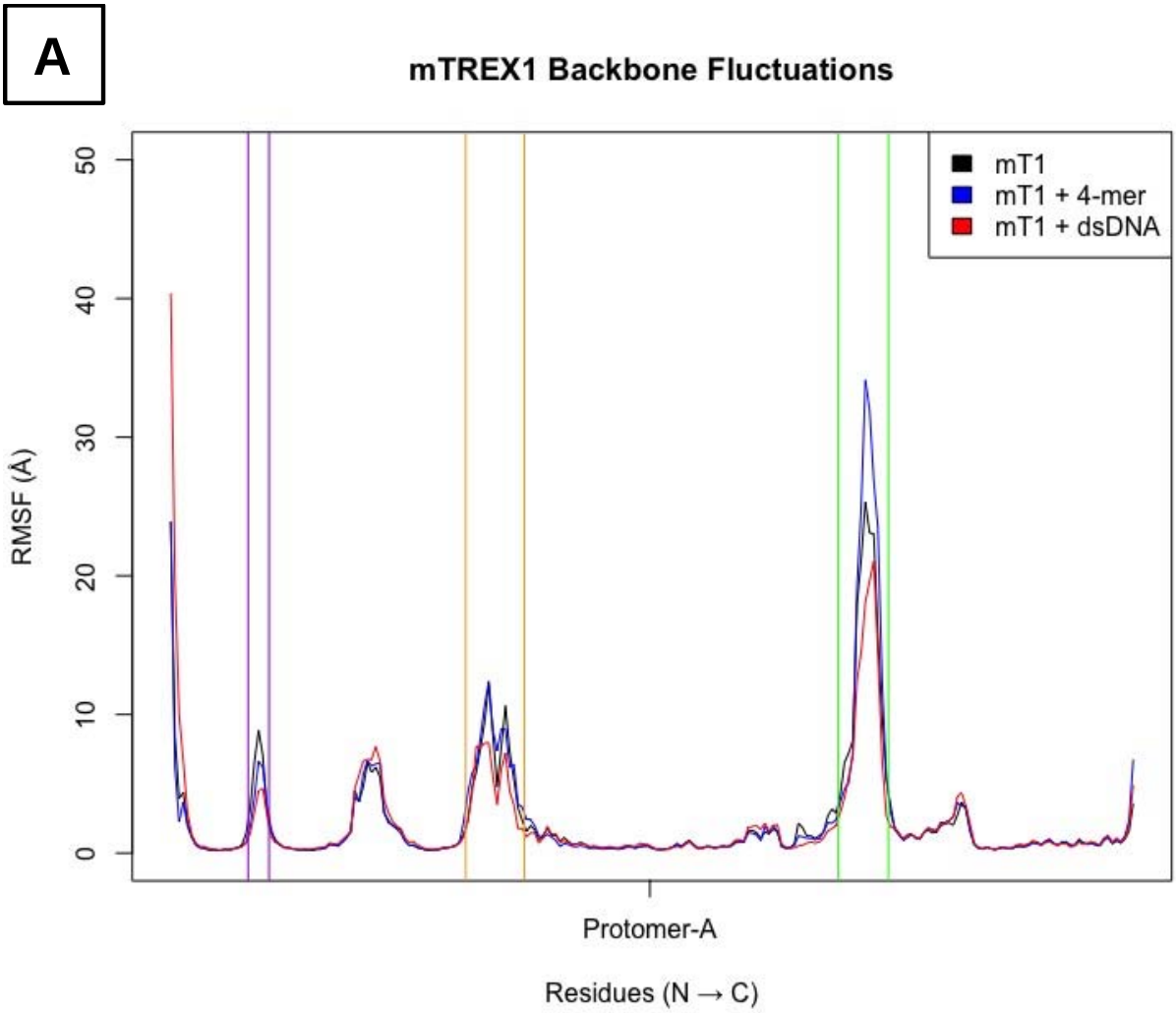


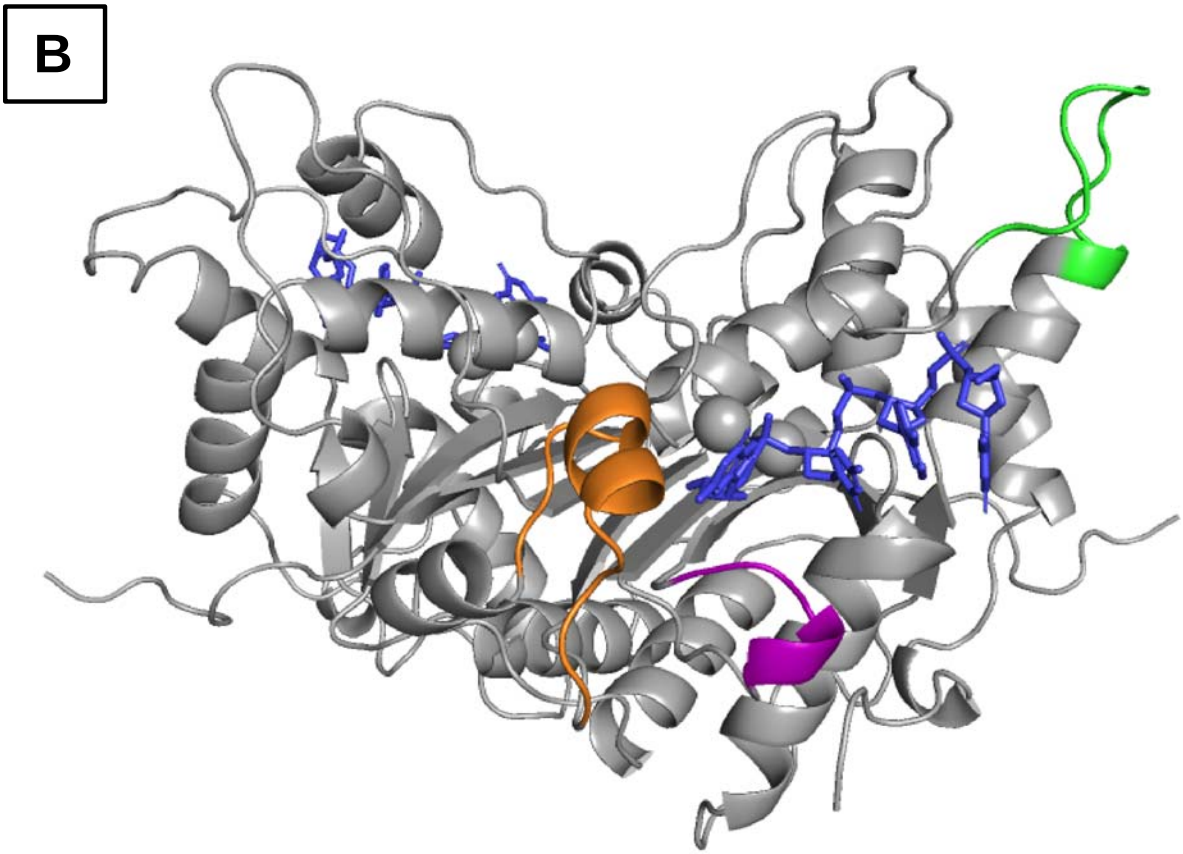
B

Melting Temperatures (°C)

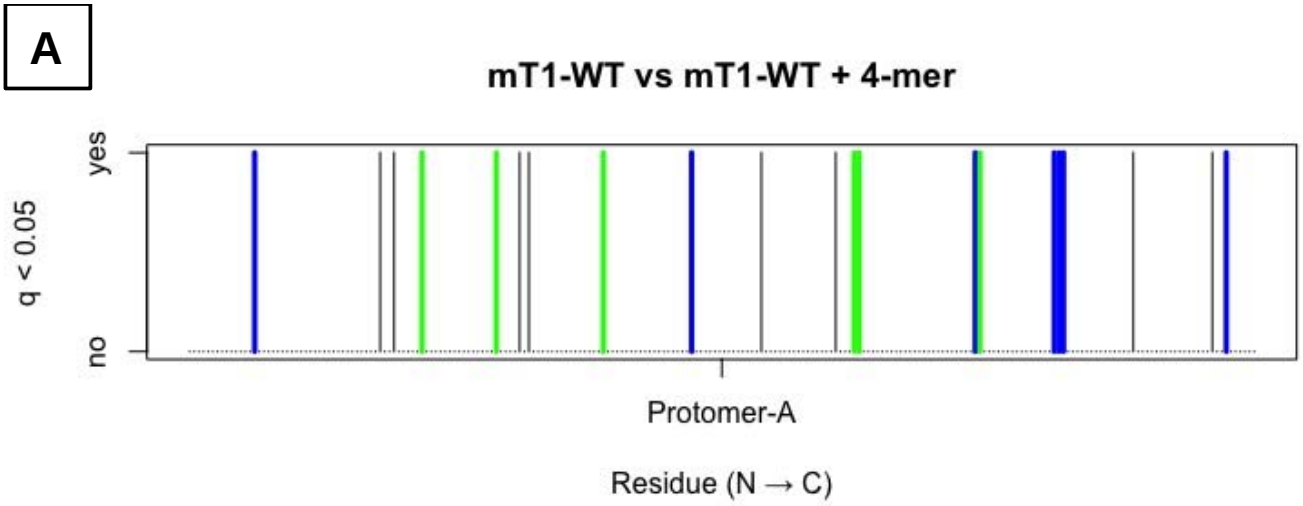
hT1-Apo	=	51.5 ± 0.359
hT1-ssDNA	=	65.1 ± 1.67
hT1-dsDNA	=	79.8 ± 0.236
mT1-Apo	=	58.2 ± 0.594
mT1-ssDNA	=	65.3 ± 1.15
mT1-dsDNA	=	78.3 ± 0.348

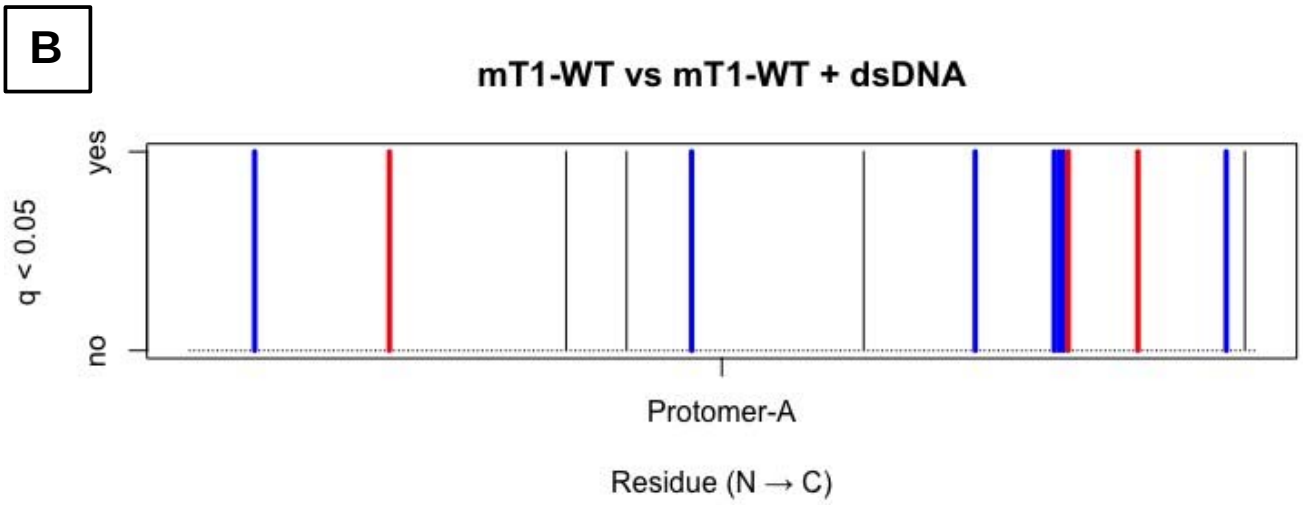
<u>Figure 3</u>: **TREX1 Has Significantly Higher Affinity for dsDNA Than ssDNA.** [A] *Melting Curves of TREX1 Complexes*. Indicated apoenzymes and enzyme-substrate complexes were subjected to an increasing temperature gradient in the presence of the nonspecific protein dye, SYPRO Orange, while fluorescence 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. [B] *Melting Points of TREX1 Complexes*. The temperatures associated with the inflection points of the curves in 'A' 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).

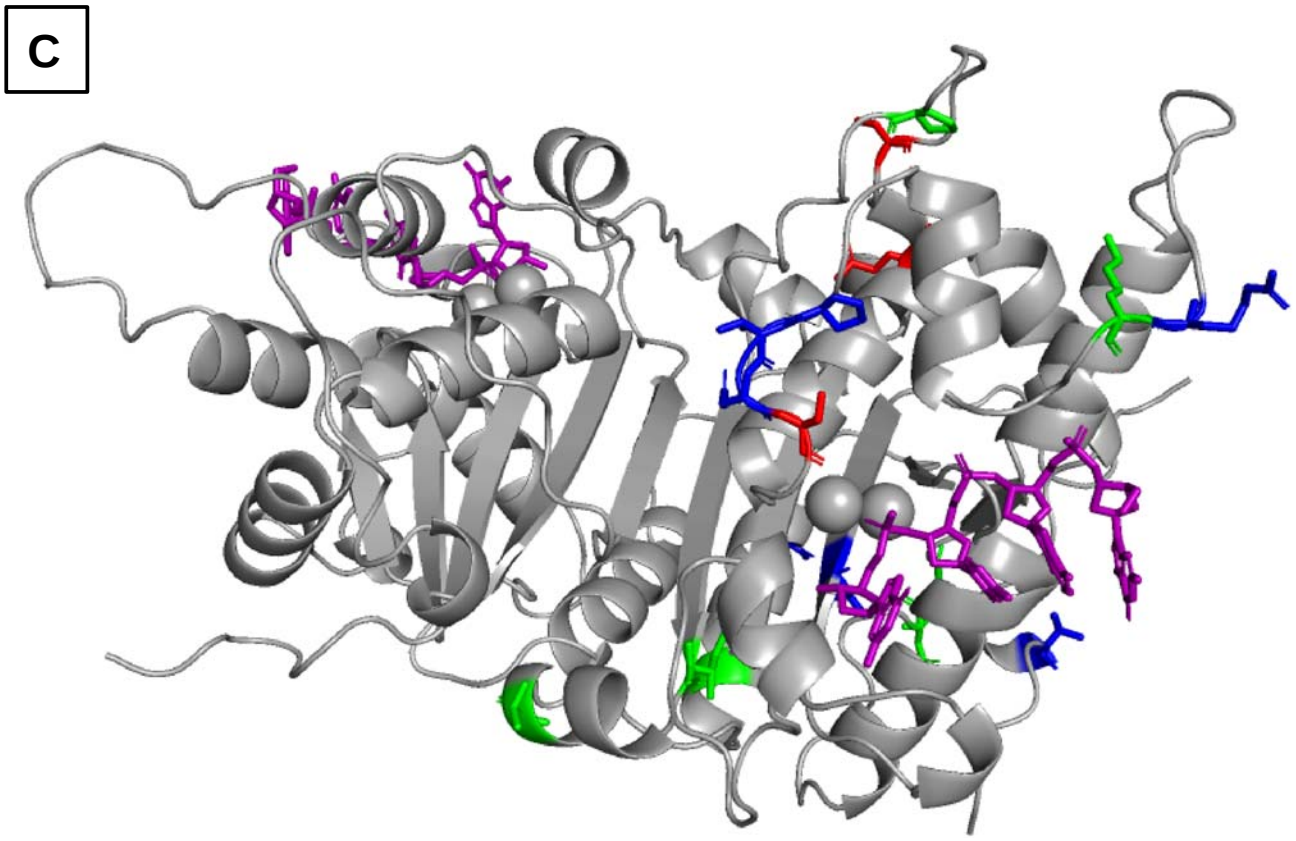

A
mTREX1 Backbone Fluctuations
mT1
mT1 + 4-mer
mT1 + dsDNA
RMSF (Å)
0
10
20
30
40
50
Protomer-A
Residues (N → C)


B

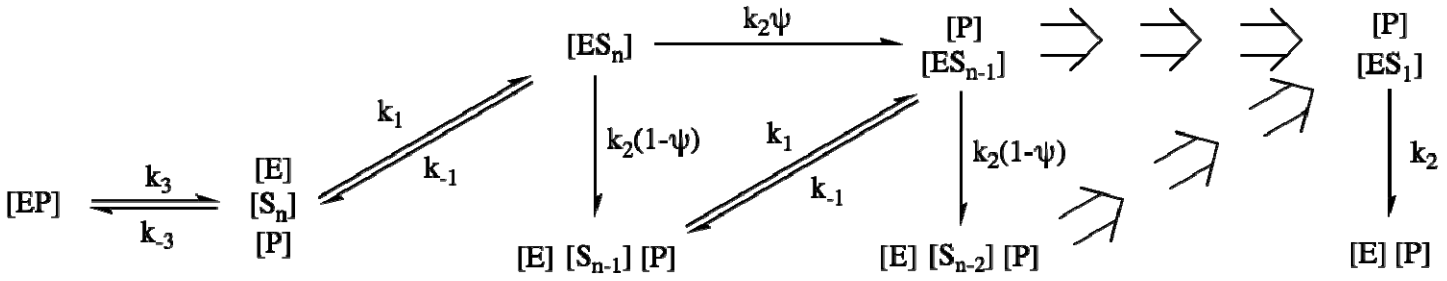
<u>Figure 4</u>: **Active Site Occupation by ssDNA or dsDNA is Associated with Differential Fluctuations in the mTREX1 Backbone.** [A] *Comparative RMSF Plots of TREX1 Simulation Systems*. All TREX1 Cα atoms were used in an RMSF calculation, as described in the relevant methods section. Plot lines displayed 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 'E'. All calculations and plotting were performed in R v3.6.1. [B] *Visualization of Dynamic hTREX1 Backbone Region*. A mT1:4-mer structure modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops. Both TREX1 protomers and all magnesium ions are colored grey, both ssDNA polymers are displayed as blue sticks. The dynamic regions identified in 'A' are colored correspondingly. Graphic was constructed in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).


A
mT1-WT vs mT1-WT + 4-mer
q < 0.05
yes
no
Protomer-A
Residue (N → C)

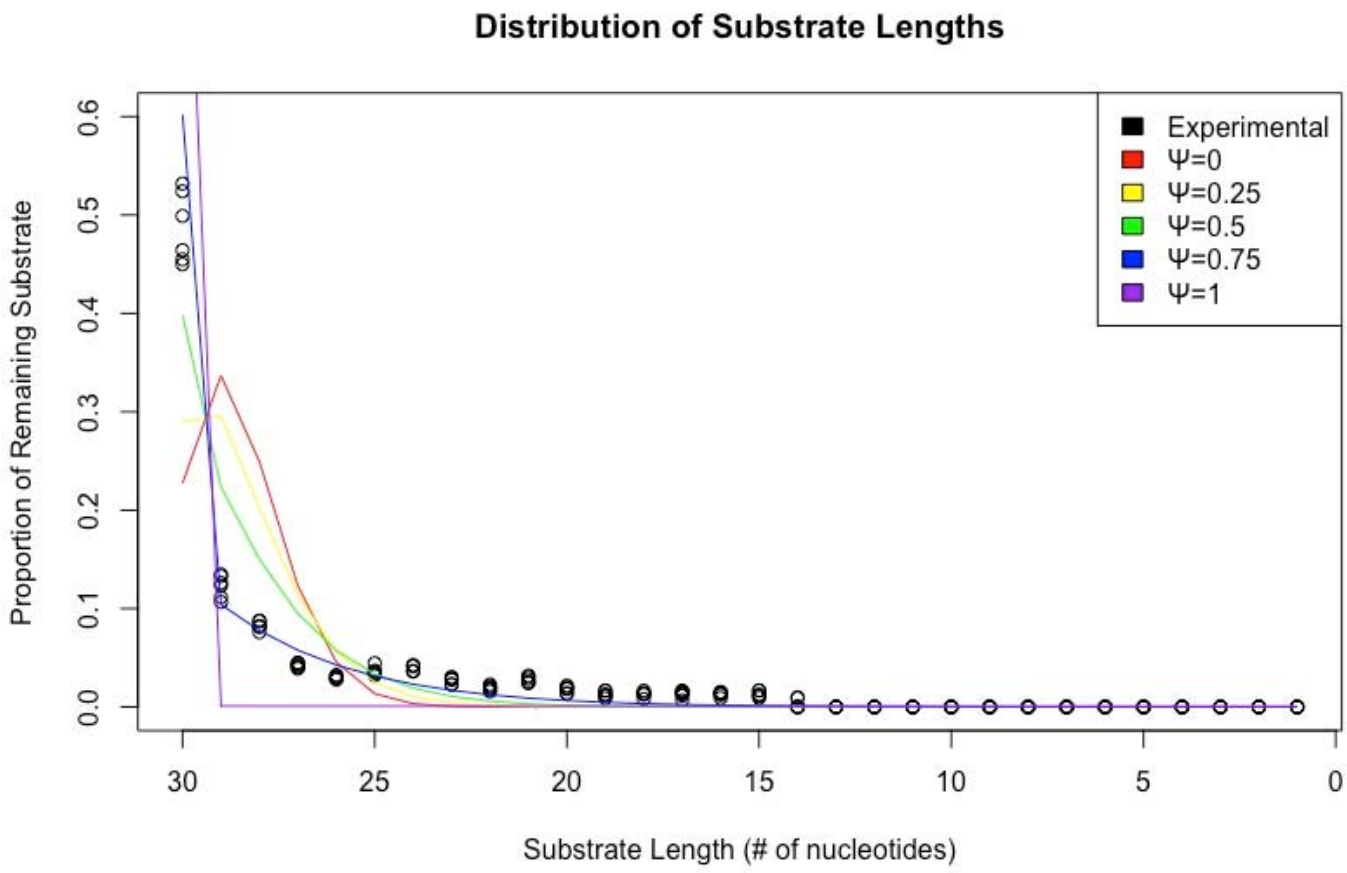

B
mT1-WT vs mT1-WT + dsDNA
q < 0.05
yes
no
Protomer-A
Residue (N → C)


C

<u>Figure 5</u>: **ssDNA and dsDNA Differentially Alter mTREX1 Residue Conformations.** [A-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. Residues with blue markers were independently identified in the analyses of both 'A' and 'B'. Green markers indicate residues identified in 'A' but not 'B', with conformation proportions significantly different between mT1:4-mer and mT1:dsDNA. Red markers indicate residues identified in 'B' but not 'A', with conformation proportions significantly different between mT1:4-mer and mT1:dsDNA. Analyses and graphing were performed in R v3.6.1. [C] *Visualization of Affected Residues*. A mT1:4-mer structure modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops. Both TREX1 protomers and all magnesium ions are colored grey, both ssDNA polymers are displayed as purple sticks. The residues identified in 'A-B' by markers are shown as correspondingly colored sticks. Graphic was constructed in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).


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

<u>Figure 6</u>: **Proposed Kinetic Model for TREX1 Exonuclease Activity.** Here, $k_{\square}$ are respective rate constants in the model, Ψ is the processivity parameter ranging from [0,1] for zero 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 *[□□]* is concentration of a complex of the respective reactants.

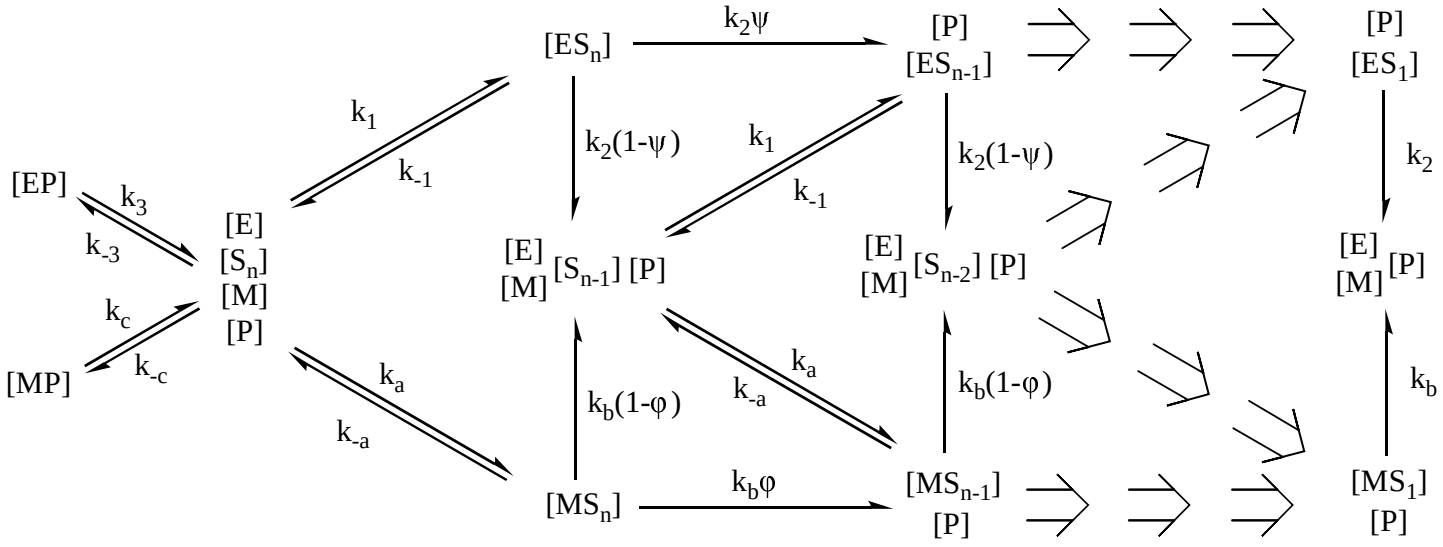

Distribution of Substrate Lengths
Experimental
Ψ=0
Ψ=0.25
Ψ=0.5
Ψ=0.75
Ψ=1
Proportion of Remaining Substrate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
30
25
20
15
10
5
0
Substrate Length (# of nucleotides)

<u>Figure 7</u>: **TREX1 ssDNA Degradation is Consistent with Semi-Processivity.** 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 4 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 Supplementals 5-6.

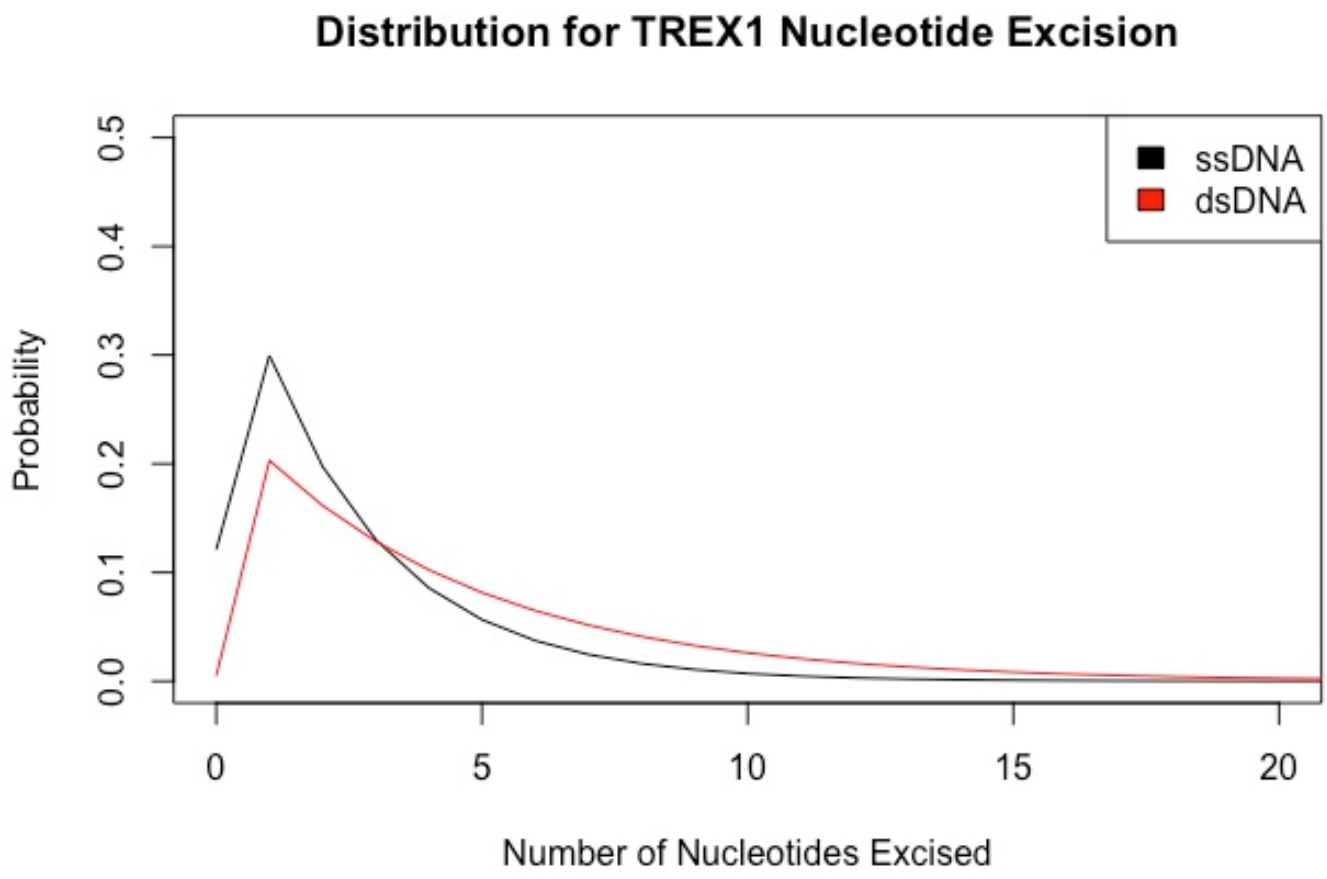
Approximated hTREX1 Kinetic Constants

	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 (10^7)	16	n.d.	n.d.	0.75
dsDNA	~ [0.10,0.050,0.020]	~1.0 (10^9)	10			~ [0.75,0.80,0.85]

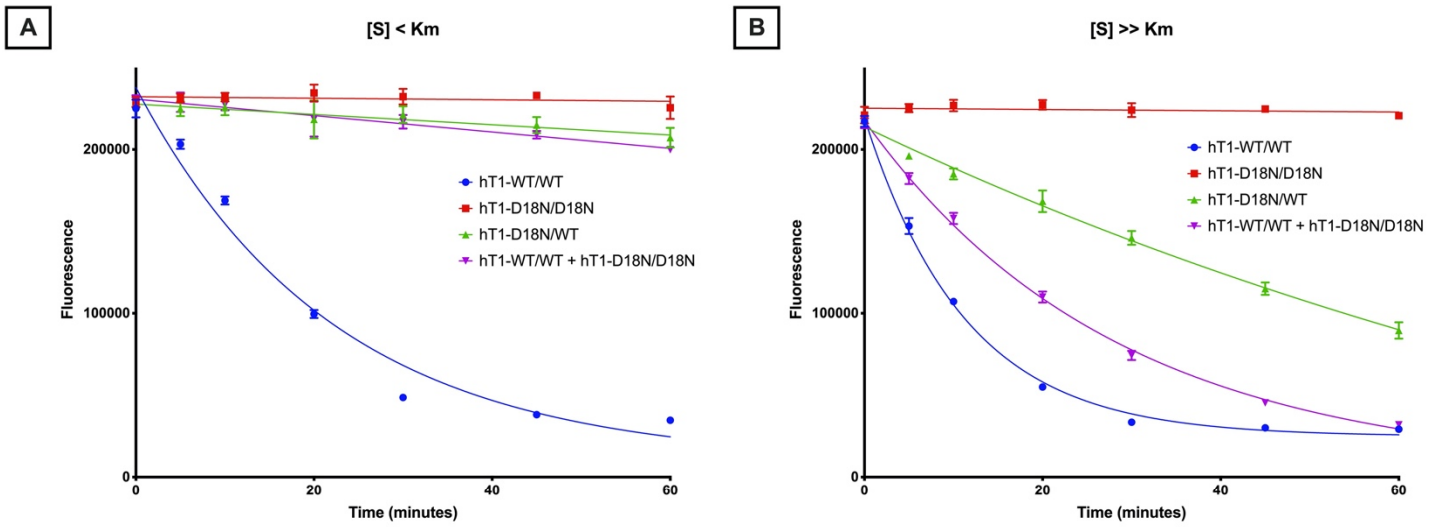
<u>Table 2</u>: **Kinetic Constant Approximations for TREX1 Exonuclease Activity Model.** A table of the values for relevant kinetic constants. Bracketed values are approximation pairs, where the first values in each bracketed vector are true under the same assumptions as one another. Other values are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


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

<u>Figure 8</u>: **Proposed Kinetic Model for Competition Between TREX1 Variants.** Here, $k_{\square}$ are respective rate constants in the model, 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.


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 9</u>: **TREX1 Typically Excises More Nucleotides from dsDNA than ssDNA During Semi-Processive Events.** 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.



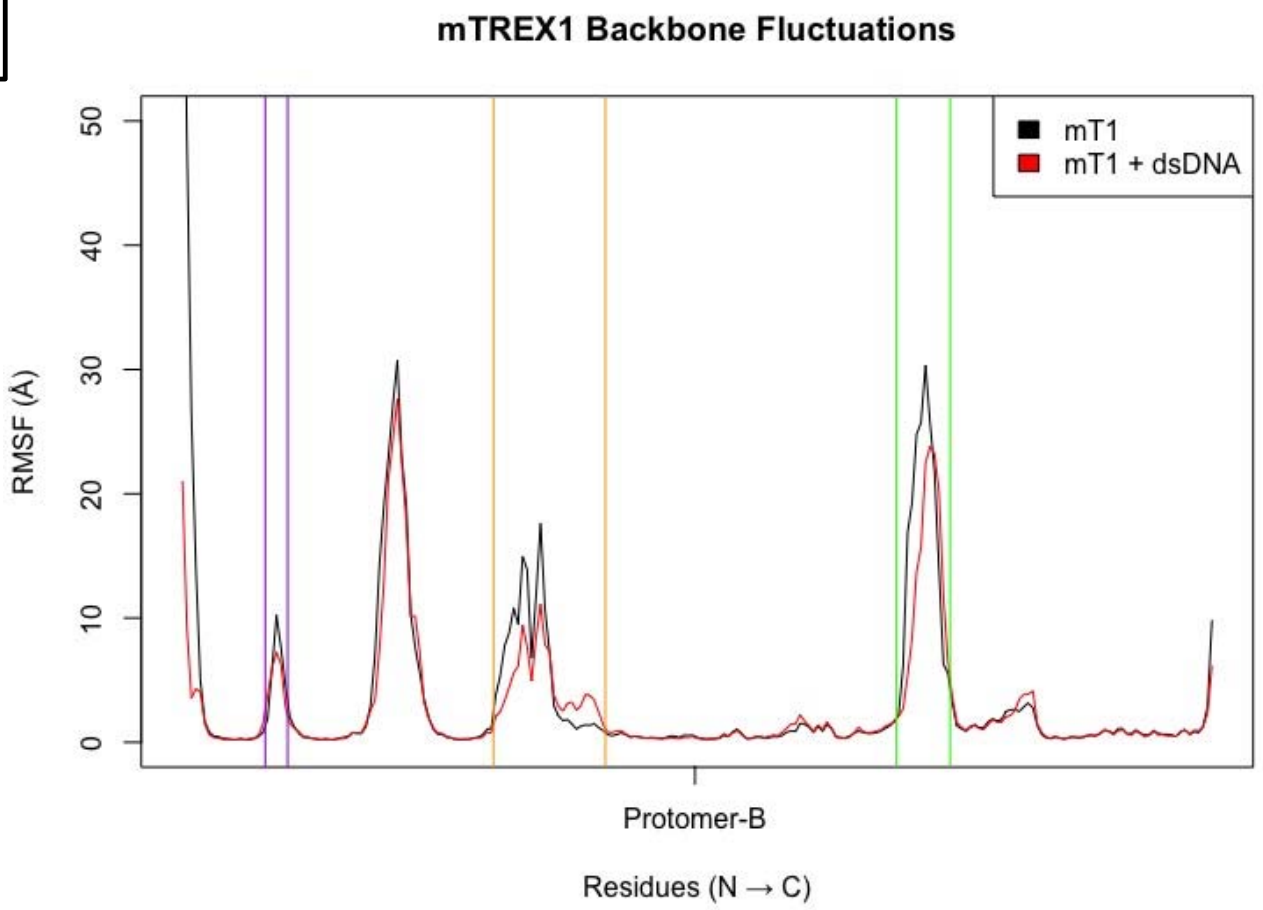
C

Relative Activities

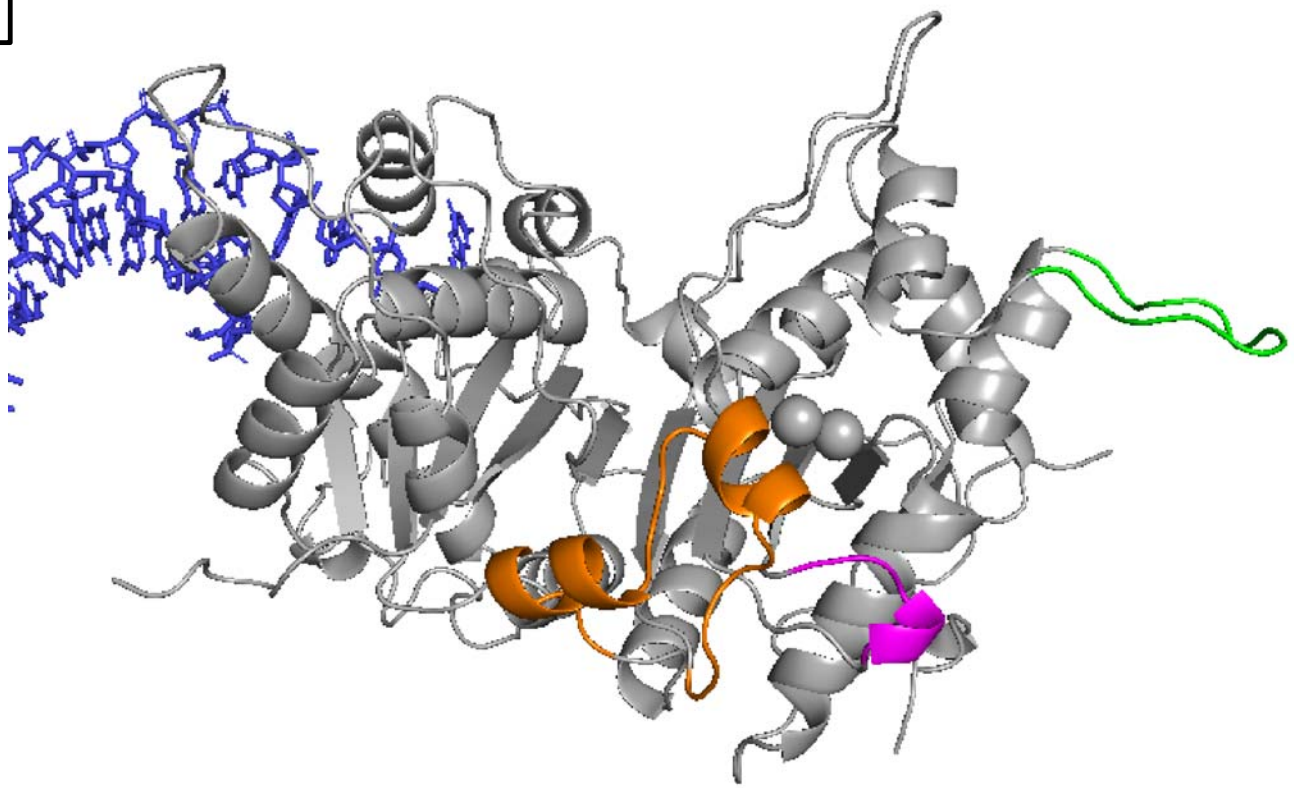
		$[S] < K_m$:	$[S] >> K_m$:
hT1-WT/WT	=	1.0	1.0
hT1-D18N/D18N	=	0.0051	0.0028
hT1-D18N/WT	=	0.034	0.19
1:1:0 MIX	=	0.055	0.50

<u>Figure 10</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 graphic '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

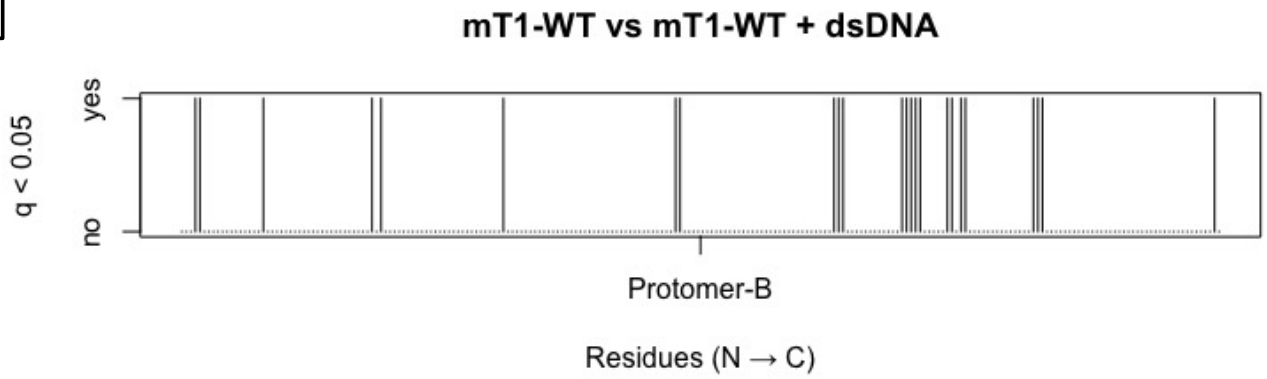

mTREX1 Backbone Fluctuations
mT1
mT1 + dsDNA
RMSF (Å)
0
10
20
30
40
50
Protomer-B
Residues (N → C)

B

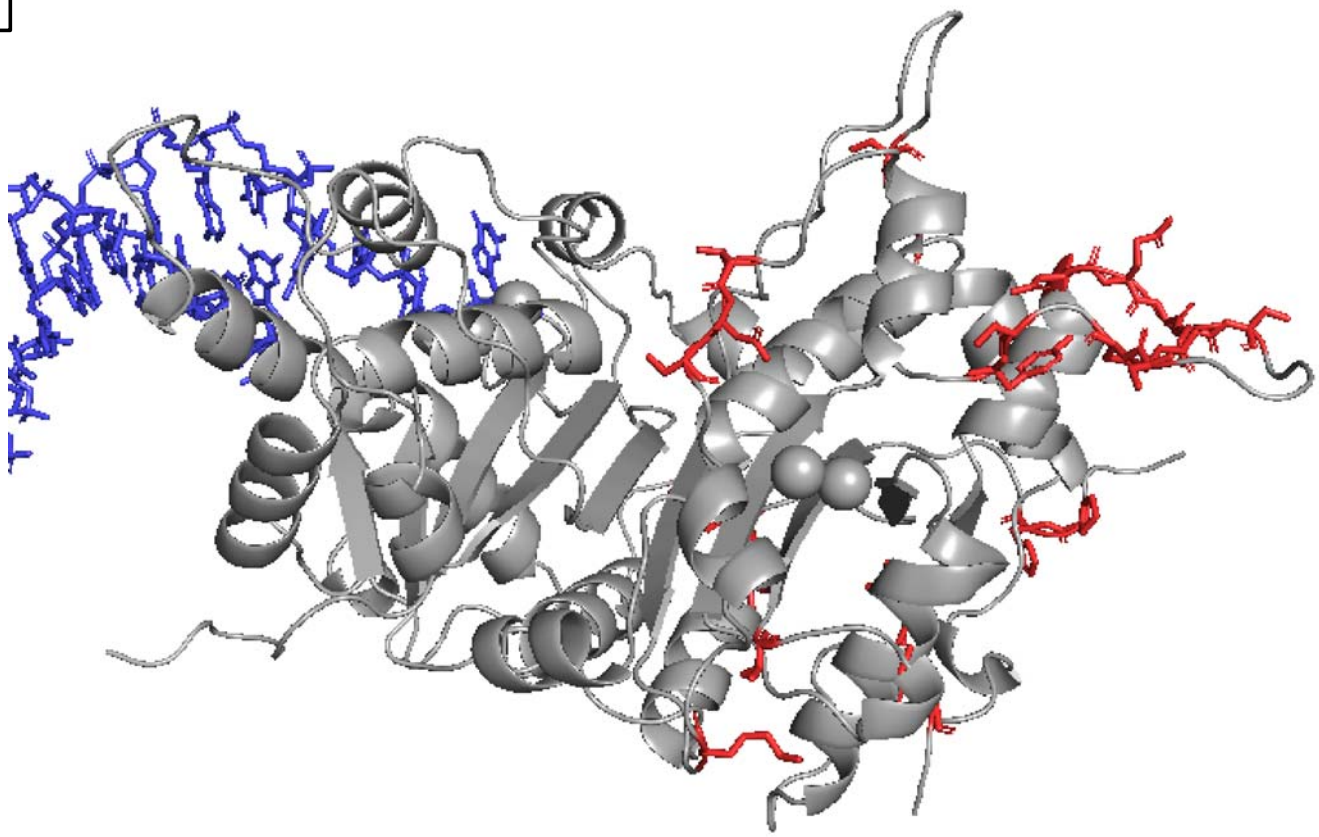


<u>Figure 11</u>: **Active Site Occupation by dsDNA is Associated with Differential Backbone Fluctuations in the Opposing Protomer.** [A] *Comparative RMSF Plots of TREX1 Simulation Systems*. All TREX1 Cα atoms in the unoccupied protomer of each system ('Protomer-B') were used in an RMSF calculation, 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 'B'. All calculations and plotting were performed in R v3.6.1. [E] *Visualization of Dynamic hTREX1 Backbone Region*. A 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 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

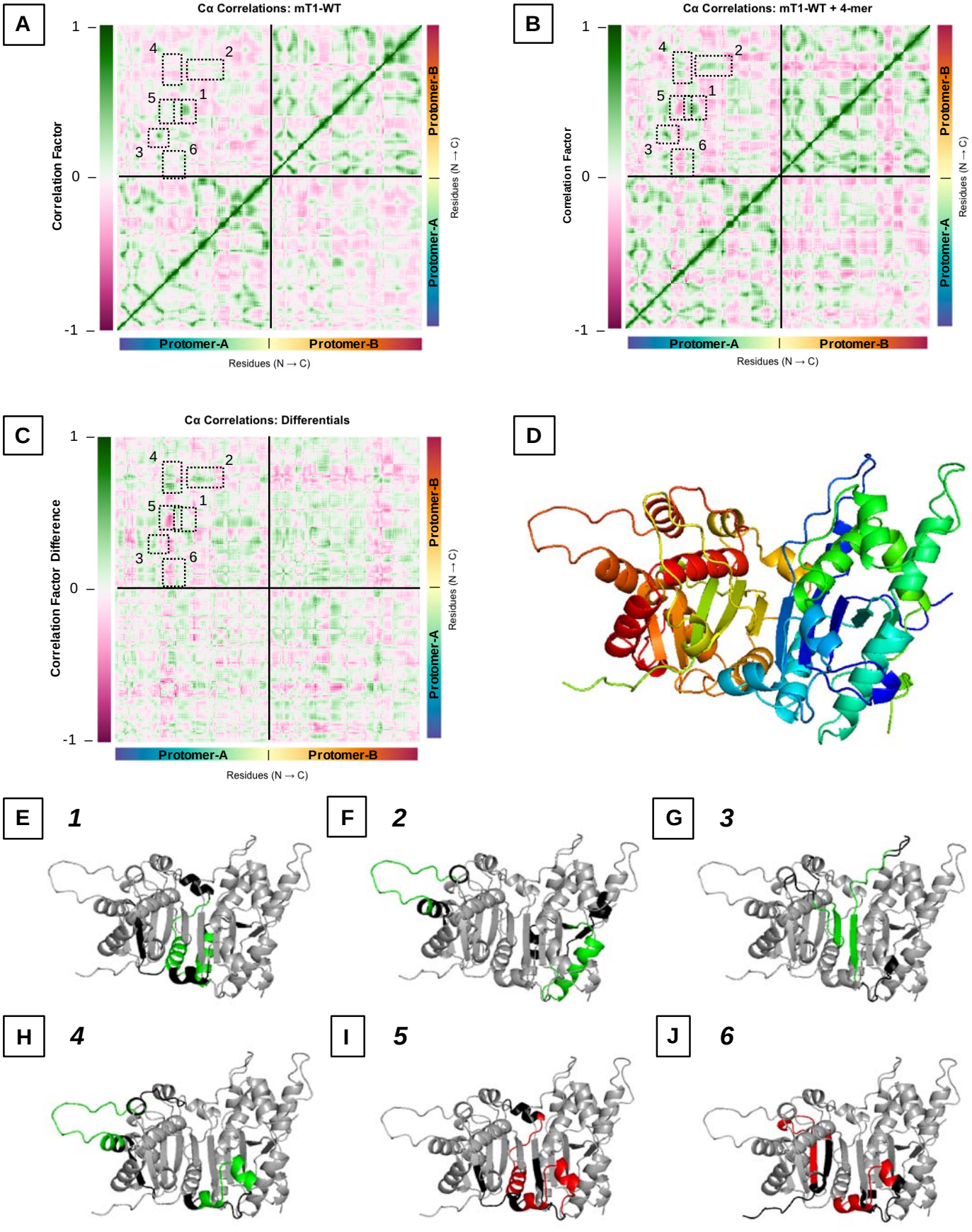
A



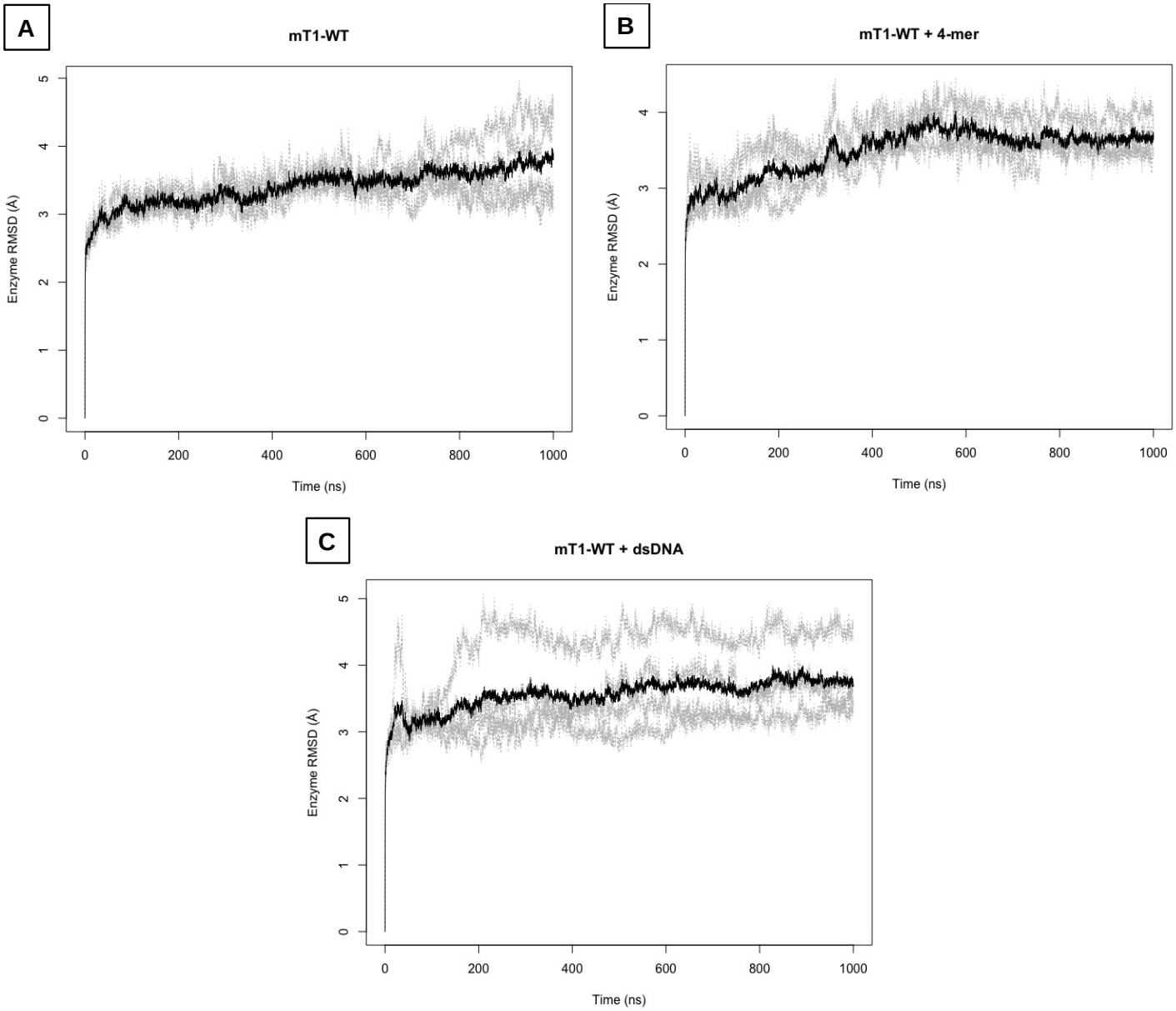
B



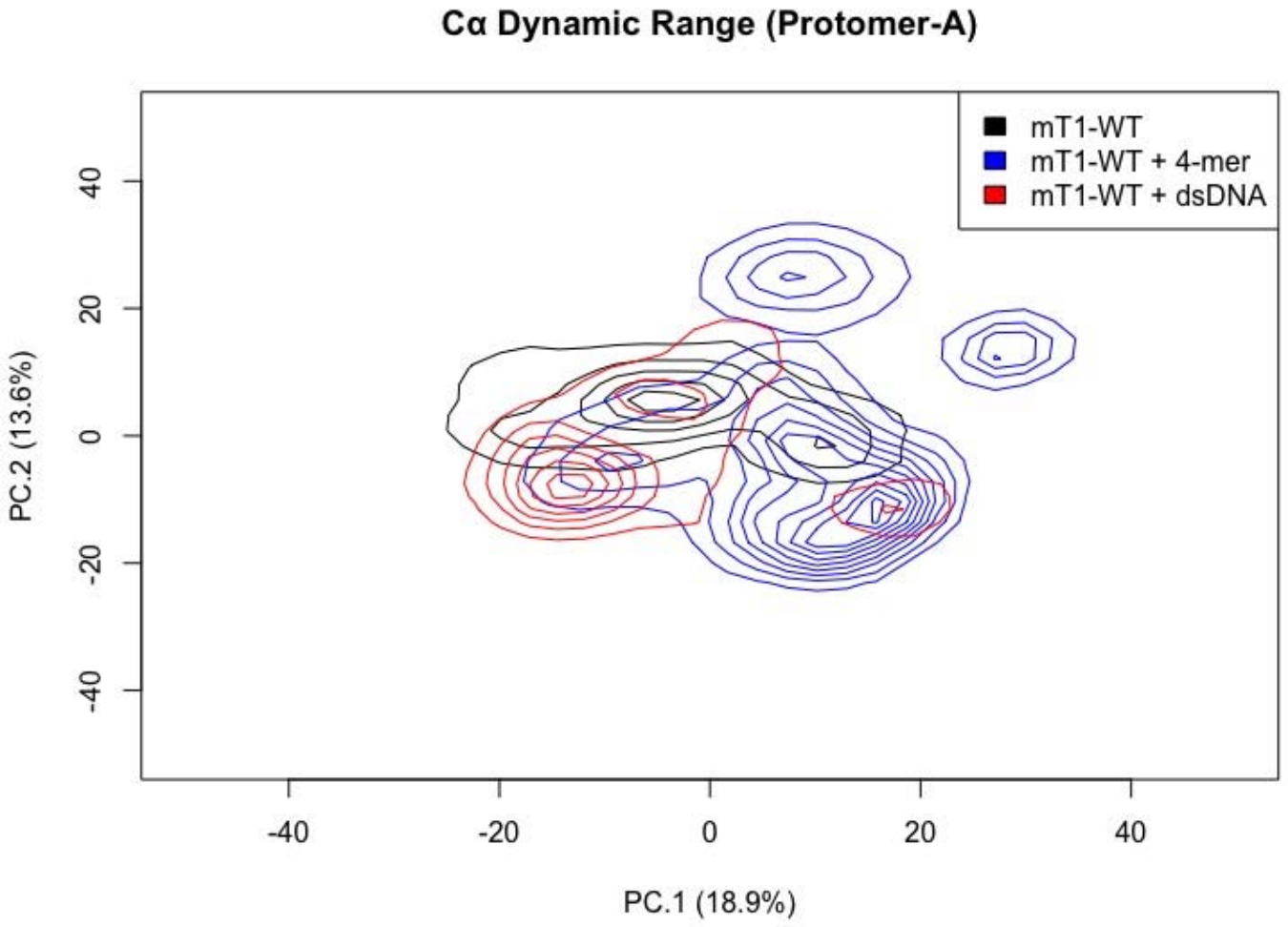
<u>Figure 12</u>: **Occupation of mT1 Active Site by dsDNA is Associated with Altered Residue Conformations in the Opposing Protomer.** [A] *Barcode Graph of Residues with Significantly Different Phi-Psi Angles*. For each residue, hierarchal clustering was used to determine conformations across all time 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. [B] *Visualization of Affected Residues*. A mT1 structure modeled from a mT1 crystal structure (PDB = '2IOC') to include disordered flexible loops. Both TREX1 protomers and all magnesium ions are colored grey. The residues identified in 'A' by markers are shown as red colored sticks. Graphic was constructed in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).


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
F
2
G
3
H
4
I
5
J
6

<u>Figure 13</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 '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 'E-J'. 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 'A' from 'B'. All other features are as discussed for '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 'A-C'. Graphic was prepared in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). [E-J] *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 in 'A-C' that are illustrated. The mT1 structures are generally shown as grey cartoons. The inner 50% of areas of interest in 'A-C' are colored green or red for positively or negatively correlated regions, respectively. The peripheral 50% of areas of interest in 'A-C' are colored black. Graphics were prepared in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

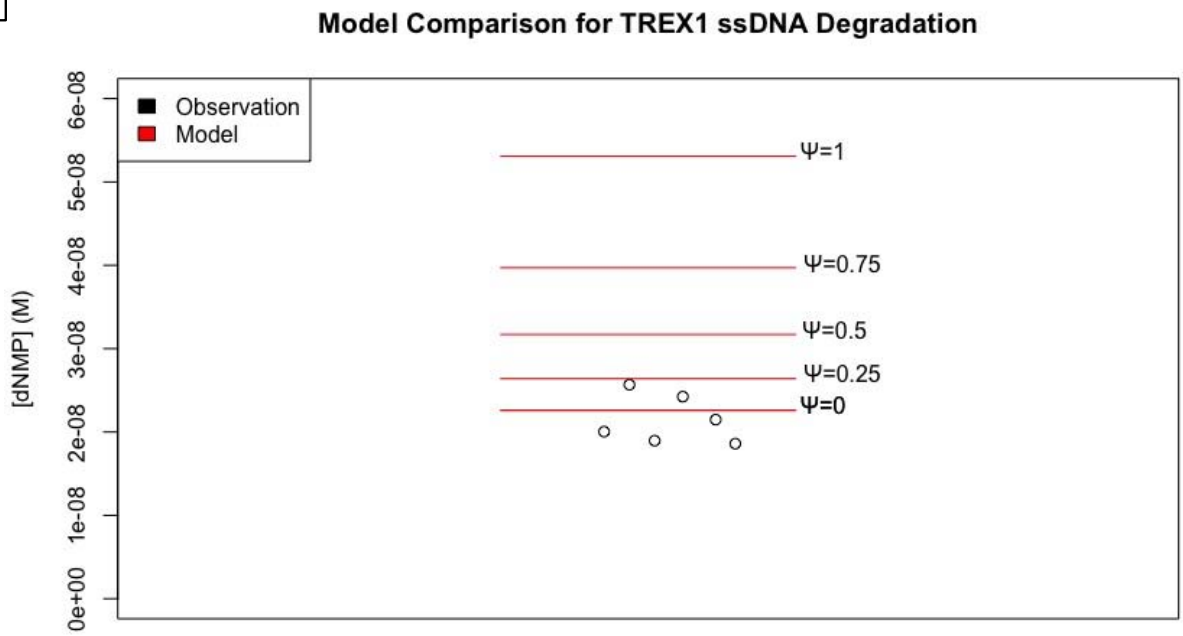

A
mT1-WT
Enzyme RMSD (Å)
Time (ns)
B
mT1-WT + 4-mer
Enzyme RMSD (Å)
Time (ns)
C
mT1-WT + dsDNA
Enzyme RMSD (Å)
Time (ns)

<u>Supplemental 1</u>: **RMSD of mTREX1 Structure from Initial Conditions.** [A-C] 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

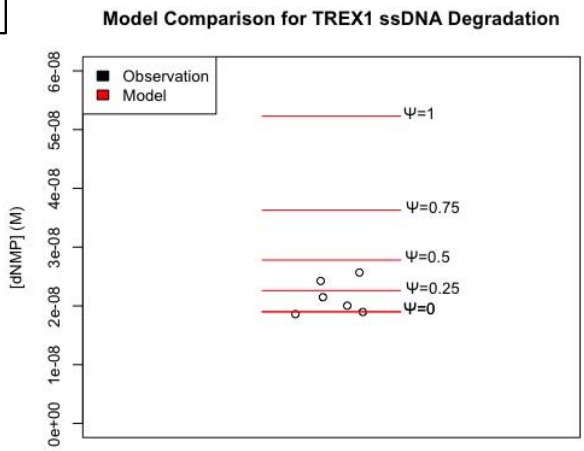
<u>Supplemental 2</u>: **Conformational Free Energy Maps of mTREX1 Backbone.** Cartesian coordinates of every Cα in mT1 were used to calculate map density functions as described in the relevant methods section. Density functions were plotted as contour maps, with innermost rings indicating the areas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

A



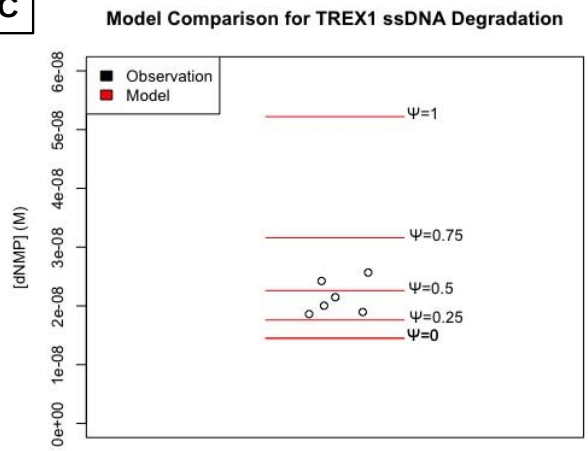
k_1 = 2.1e8, k_-1 = 8.7

B



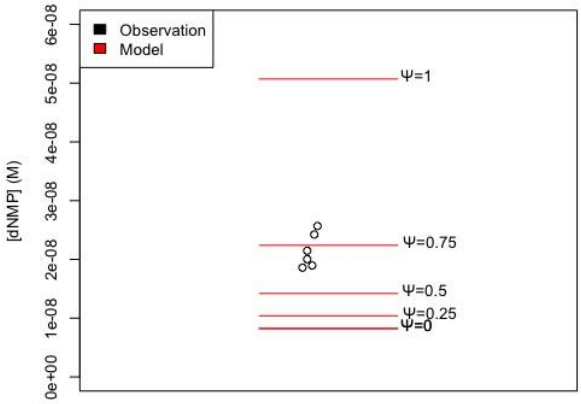
k_1 = 1.6e8, k_-1 = 6.7

C



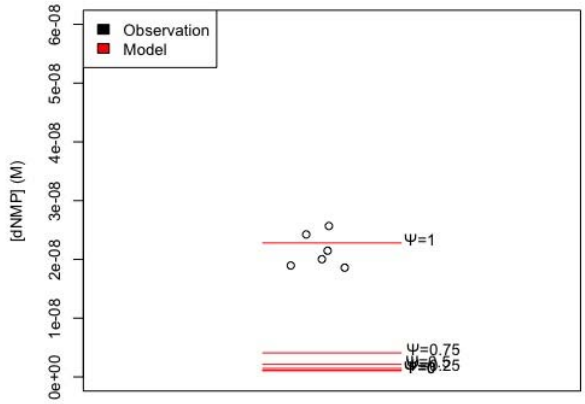
k_1 = 1.1e8, k_-1 = 4.5

D



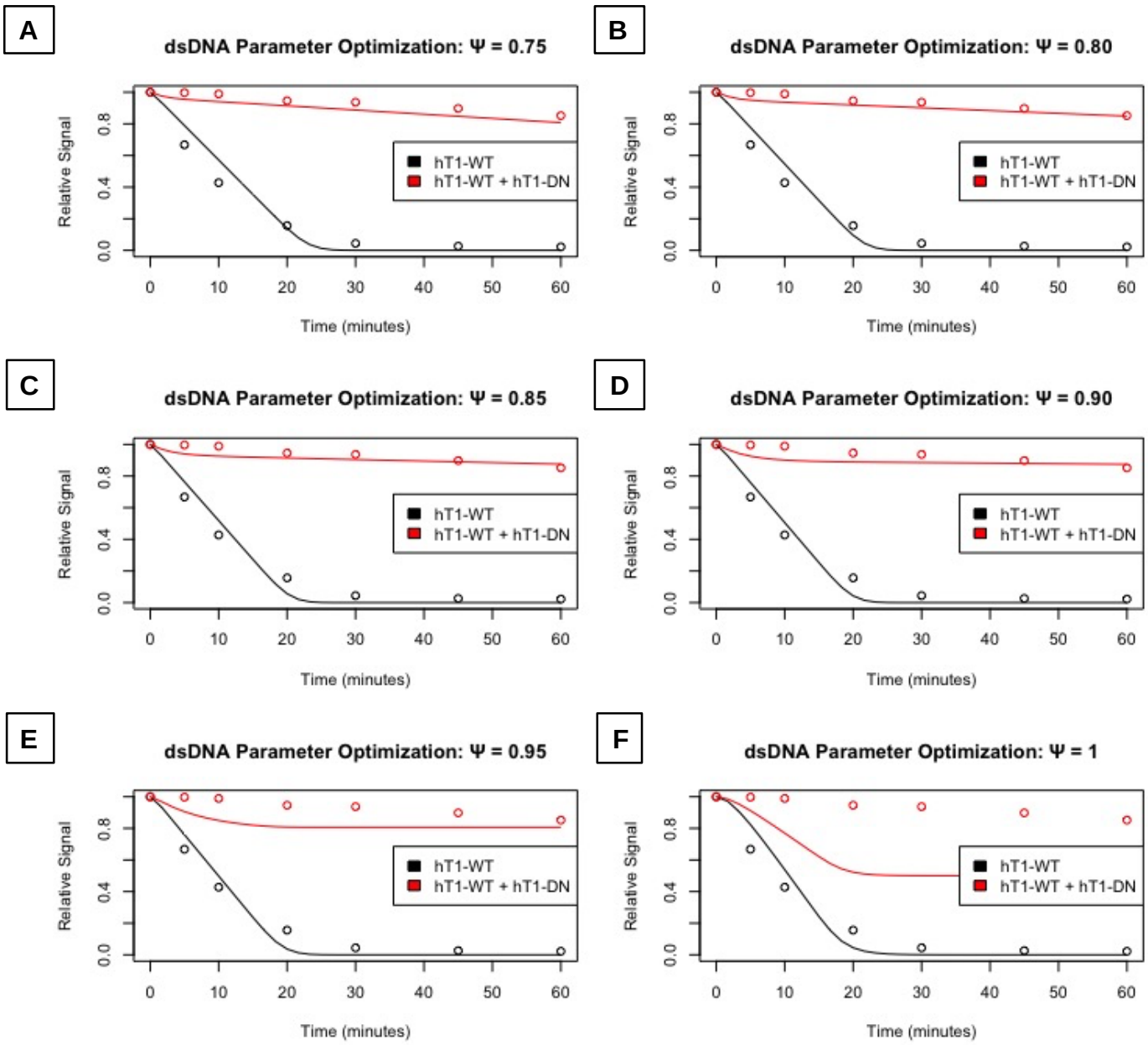
k_1 = 5.2e7, k_-1 = 2.2

E



k_1 = 5.92e6, k_-1 = 0.25

<u>Supplemental 3</u>: **Optimization of Parameters for TREX1 ssDNA Kinetic Model.** [A-E] 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 Supplementals 5-6.


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.** [A-F] 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 Supplementals 5-6. The black and red data points here are derived from the same data as the blue and purple data points in Figure 10 A, respectively, after normalization to maximum and minimum fluorescence.

Simulations Legend

Reaction	Description
R1	Supplemental 3 A, ‘Ψ=0’
R2	Supplemental 3 A, ‘Ψ=0.25’
R3	Supplemental 3 A, ‘Ψ=0.5’
R4	Supplemental 3 A, ‘Ψ=0.75’
R5	Supplemental 3 A, ‘Ψ=1’
R6	Supplemental 3 B, ‘Ψ=0’
R7	Supplemental 3 B, ‘Ψ=0.25’
R8	Supplemental 3 B, ‘Ψ=0.5’
R9	Supplemental 3 B, ‘Ψ=0.75’
R10	Supplemental 3 B, ‘Ψ=1’
R11	Supplemental 3 C, ‘Ψ=0’
R12	Supplemental 3 C, ‘Ψ=0.25’
R13	Supplemental 3 C, ‘Ψ=0.5’
R14	Supplemental 3 C, ‘Ψ=0.75’
R15	Supplemental 3 C, ‘Ψ=1’
R16	Supplemental 3 D, ‘Ψ=0’
R17	Supplemental 3 D, ‘Ψ=0.25’
R18	Supplemental 3 D, ‘Ψ=0.5’
R19	Supplemental 3 D, ‘Ψ=0.75’
R20	Supplemental 3 D, ‘Ψ=1’
R21	Supplemental 3 E, ‘Ψ=0’
R22	Supplemental 3 E, ‘Ψ=0.25’
R23	Supplemental 3 E, ‘Ψ=0.5’
R24	Supplemental 3 E, ‘Ψ=0.75’
R25	Supplemental 3 E, ‘Ψ=1’
R26	Figure 7, red line
R27	Figure 7, yellow line
R28	Figure 7, green line

R29	Figure 7, blue line
R30	Figure 7, purple line
R31	Supplemental 4 A, black line
R32	Supplemental 4 A, red line
R33	Supplemental 4 B, black line
R34	Supplemental 4 B, red line
R35	Supplemental 4 C, black line
R36	Supplemental 4 C, red line
R37	Supplemental 4 D, black line
R38	Supplemental 4 D, red line
R39	Supplemental 4 E, black line
R40	Supplemental 4 E, red line
R41	Supplemental 4 F, black line
R42	Supplemental 4 F, red line

<u>Supplemental 5</u>: **Reactions Simulated in These Studies.** A legend corresponding to the reaction parameters outlined in Supplemental 6. Table was prepared in Word (Microsoft).

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-	7.5e-9	3.75e-	7.5e-9	3.75e-	7.5e-9	3.75e-	7.5e-9	3.75e-

		9		9		9		9		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

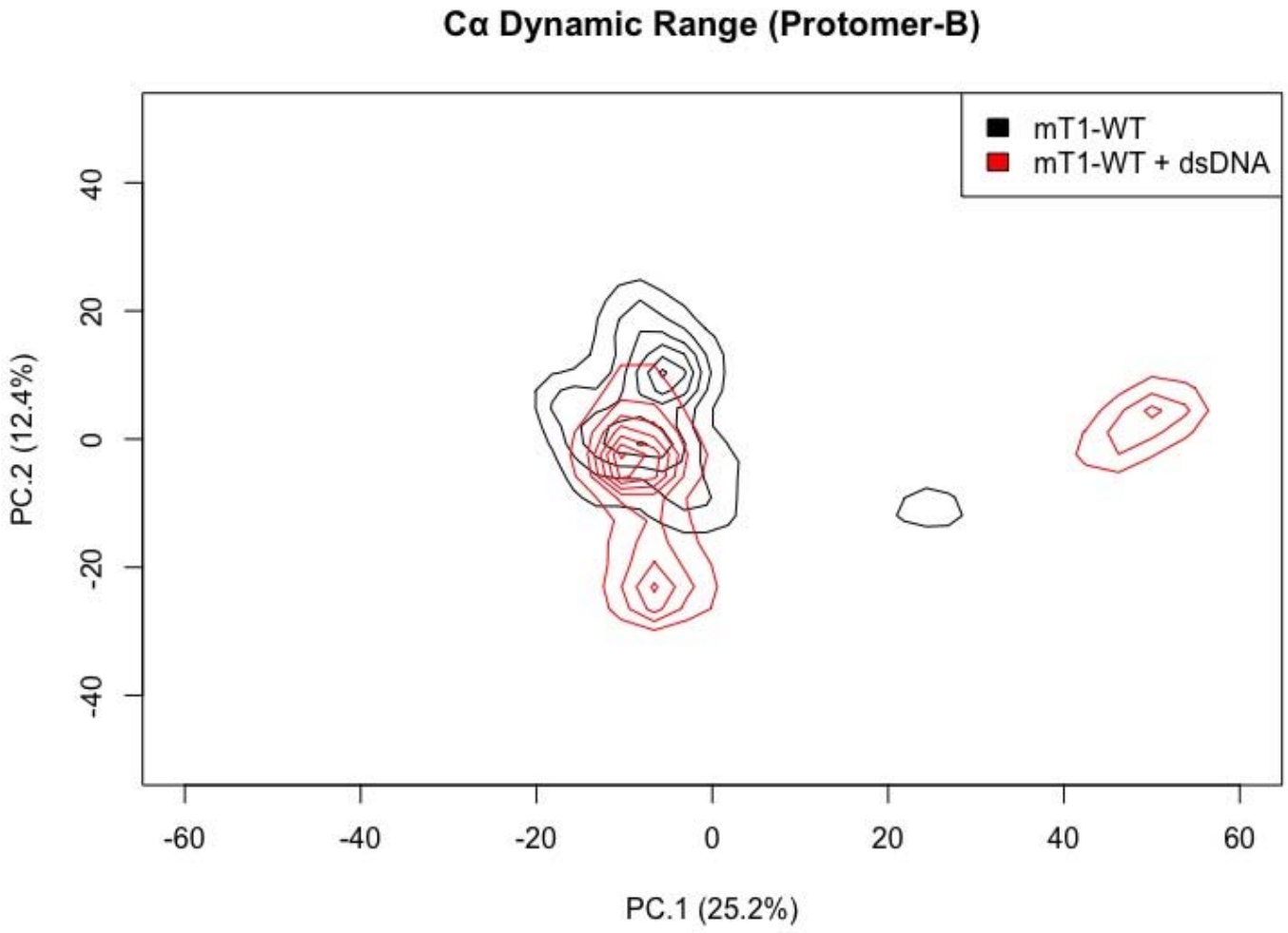
	R41	**R42**
t (min)	60	60
Δt (ms)	1	1
$[E]_0$ (M)	7.5e-9	3.75e-9
$[M]_0$ (M)	0	3.75e-9
β	1e4	1e4
$[S_\beta]_0$ (M)	0.83e-9	0.83e-9
Ψ	1	1
k_{-1} (s^{-1})	0	0
k_1 ($M^{-1}s^{-1}$)	1e9	1e9
k_2 ($M^{-1}s^{-1}$)	10	10
Φ	Ψ	Ψ
k_{-a} (s^{-1})	k_{-1}	k_{-1}
k_a ($M^{-1}s^{-1}$)	k_1	k_1
k_b ($M^{-1}s^{-1}$)	0	0

<u>Supplemental 6</u>: **Kinetic Parameters for Simulated Reactions in These Studies.** A table of the parameters used with the R Script in Appendix IV to simulate reactions. A legend identifying the reactions is given in Supplemental 5. Table was prepared in Word (Microsoft).

```
#### BEGIN SCRIPT
## Define Parameters
# Simulation
time=20*60 # s
save=1 # s
dt=1e-3 # s
# Reaction Components
beta=30 # [nt]/[S]
E.0=10e-12 # M
M.0=0e-9 # M
S.0=15e-9 # M
P.0=0 # M
ES.0=0 # M
EP.0=0 # M
MS.0=0 # M
MP.0=0 # M
# WT Enzyme Constants
psi=0.75 # processivity factor (ss= 0.75;ds~ [0.00;0.75,0.80,0.85])
k.n1=2.2 # s^-1 (ss= 2.2;ds~ [0.30;0.10,0.050,0.020]) # [ES]->[E][S] rate constant
k.1=5.2e7 # s^-1 (ss= 5.2e7;ds~ [5.0e9;1.0e9]) # [E][S]->[ES] rate constant
k.2=16 # s^-1 (ss= 16;ds= 10) # [ES]->[E][P]([S]) rate constant + [ES]->[ES][P] rate constant
k.n3=0 # s^-1 (??) # [EP]->[E][P] rate constant
k.3=0 # s^-1 (??) # [E][P]->[EP] rate constant
# MUT Enzyme Constants
phi=psi # processivity factor
k.na=k.n1 # s^-1 # [MS]->[M][S] rate constant
k.a=k.1 # s^-1 # [M][S]->[MS] rate constant
k.b=0 # s^-1 # [MS]->[M][P]([S]) rate constant + [MS]->[MS][P] rate constant
k.nc=k.n3 # s^-1 # [MP]->[M][P] rate constant
k.c=k.3 # s^-1 # [M][P]->[MP] rate constant
## Generate Initial Conditions
# Create Vectors
t=0
E=E.0
M=M.0
S=matrix(0,nrow=beta,ncol=(time/save+1))
S[beta,1]=S.0
P=P.0
ES=matrix(0,nrow=beta,ncol=(time/save+1))
ES[beta,1]=ES.0
MS=matrix(0,nrow=beta,ncol=(time/save+1))
MS[beta,1]=MS.0
EP=EP.0
MP=MP.0
# Initial Concentrations
t.now=0
E.now=E.0
M.now=M.0
S.now=rep(0,times=beta)
S.now[beta]=S.0
P.now=P.0
ES.now=rep(0,times=beta)
ES.now[beta]=ES.0
MS.now=rep(0,times=beta)
MS.now[beta]=MS.0
EP.now=EP.0
```

```
MP.now=MP.0
# Initial Rates
d.E=k.n3*EP.now+k.n1*sum(ES.now)+k.2*(1-psi)*sum(ES.now[2:beta])+k.2*ES.now[1]-k.3*E.now*P.now-
k.1*E.now*sum(S.now)
d.M=k.nc*MP.now+k.na*sum(MS.now)+k.b*(1-phi)*sum(MS.now[2:beta])+k.b*MS.now[1]-k.c*M.now*P.now-
k.a*M.now*sum(S.now)
d.P=k.n3*EP.now+k.nc*MP.now+k.2*sum(ES.now)+k.b*sum(MS.now)-k.3*E.now*P.now-k.c*M.now*P.now
d.S=k.n1*ES.now+k.na*MS.now+k.2*(1-psi)*c(ES.now[2:beta],0)+k.b*(1-phi)*c(MS.now[2:beta],0)-
k.1*E.now*S.now-k.a*M.now*S.now
d.ES=k.1*E.now*S.now+k.2*psi*c(ES.now[2:beta],0)-(k.n1+k.2)*ES.now
d.EP=k.3*E.now*P.now-k.n3*EP.now
d.MS=k.a*M.now*S.now+k.b*phi*c(MS.now[2:beta],0)-(k.na+k.b)*MS.now
d.MP=k.c*M.now*P.now-k.nc*MP.now
## Start Simulation
for (i in 1:(time/dt)){
	t.now=t.now+dt
	E.now=E.now+d.E*dt
	M.now=M.now+d.M*dt
	S.now=S.now+d.S*dt
	P.now=P.now+d.P*dt
	ES.now=ES.now+d.ES*dt
	EP.now=EP.now+d.EP*dt
	MS.now=MS.now+d.MS*dt
	MP.now=MP.now+d.MP*dt
	d.E=k.n3*EP.now+k.n1*sum(ES.now)+k.2*(1-psi)*sum(ES.now[2:beta])+k.2*ES.now[1]-
	k.3*E.now*P.now-k.1*E.now*sum(S.now)
	d.M=k.nc*MP.now+k.na*sum(MS.now)+k.b*(1-phi)*sum(MS.now[2:beta])+k.b*MS.now[1]-
	k.c*M.now*P.now-k.a*M.now*sum(S.now)
	d.P=k.n3*EP.now+k.nc*MP.now+k.2*sum(ES.now)+k.b*sum(MS.now)-k.3*E.now*P.now-
	k.c*M.now*P.now
	d.S=k.n1*ES.now+k.na*MS.now+k.2*(1-psi)*c(ES.now[2:beta],0)+k.b*(1-phi)*c(MS.now[2:beta],0)-
	k.1*E.now*S.now-k.a*M.now*S.now
	d.ES=k.1*E.now*S.now+k.2*psi*c(ES.now[2:beta],0)-(k.n1+k.2)*ES.now
	d.EP=k.3*E.now*P.now-k.n3*EP.now
	d.MS=k.a*M.now*S.now+k.b*phi*c(MS.now[2:beta],0)-(k.na+k.b)*MS.now
	d.MP=k.c*M.now*P.now-k.nc*MP.now
	if (sum(seq(0,time/dt,save/dt)==i)>=1){
		t=append(t,t.now,after = length(t))
		E=append(E,E.now,after = length(E))
		M=append(M,M.now,after = length(M))
		S[,(t.now/save+1)]=S.now
		P=append(P,P.now,after = length(P))
		ES[,(t.now/save+1)]=ES.now
		EP=append(EP,EP.now,after = length(EP))
		MS[,(t.now/save+1)]=MS.now
		MP=append(MP,MP.now,after = length(MP))
		show(paste("Progress = ",round(i*dt/time*100,1),"%",sep=""))
	}
}
#### END SCRIPT
```

<u>Supplemental 7</u>: R Script for Simulating Reactions with the TREX1 Kinetic Model.


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 8</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.

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. 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).
5. 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).
6. 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).
7. 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).
8. Hasan, M. *et al.* Cytosolic nuclease TREX1 regulates oligosaccharyltransferase activity independent of nuclease activity to suppress immune activation. *Immunity* **43**, 463–474 (2015).
9. Richards, A. *et al.* C-terminal truncations in human 3′-5′ DNA exonuclease TREX1 cause autosomal dominant retinal vasculopathy with cerebral leukodystrophy. *Nat. Genet.* **39**, 1068–1070 (2007).
10. Günther, C., Meurer, M., Stein, A., Viehweg, A. & Lee-Kirsch, M. A. Familial Chilblain Lupus – A Monogenic Form of Cutaneous Lupus Erythematosus due to a Heterozygous Mutation in TREX1. *Dermatology* **219**, 162–166 (2009).
11. Simpson, S. R. *et al.* T Cells Produce IFN-α in the TREX1 D18N Model of Lupus-like Autoimmunity. *J. Immunol.* (2019) doi:10.4049/jimmunol.1900220.
12. 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).
13. 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).
14. Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**, 587–598 (2008).
15. Rego, S. L. *et al.* TREX1 D18N mice fail to process erythroblast DNA resulting in inflammation and dysfunctional erythropoiesis. *Autoimmunity* **51**, 333–344 (2018).
16. Wu, J. *et al.* Cyclic-GMP-AMP Is An Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. *Science* **339**, (2013).
17. 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).
18. Belyakova, N. V. *et al.* Proof-reading 3'-->5' exonucleases isolated from rat liver nuclei. *Eur. J. Biochem.* **217**, 493–500 (1993).
19. 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).
20. Ablasser, A. *et al.* TREX1 Deficiency Triggers Cell-Autonomous Immunity in a cGAS-Dependent Manner. *J. Immunol.* **192**, 5993–5997 (2014).
21. 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).
22. Xiao, N. *et al.* cGAS activation causes lupus-like autoimmune disorders in a TREX1 mutant mouse model. *J. Autoimmun.* **100**, 84–94 (2019).
23. 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).
24. 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).
25. Orebaugh, C. D. *et al.* The TREX1 C-terminal Region Controls Cellular Localization through Ubiquitination. *J. Biol. Chem.* **288**, 28881–28892 (2013).

26. 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).
27. 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).
28. 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).
29. de Silva, U. STRUCTURAL AND BIOCHEMICAL STUDIES OF TREX - THREE PRIME REPAIR EXONUCLEASES. (WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES, 2007).
30. 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).
31. 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).
32. Lindahl, T., Gally, J. A. & Edelman, G. M. Properties of Deoxyribonuclease III from Mammalian Tissues. *J. Biol. Chem.* **244**, 5014–5019 (1969).
33. Ault, A. An introduction to enzyme kinetics. *J. Chem. Educ.* **51**, 381 (1974).
34. Processivity - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/processivity.
35. Hemphill, W. O. & Perrino, F. W. Measuring TREX1 and TREX2 exonuclease activities. *Methods Enzymol.* **625**, 109–133 (2019).
36. Bailey, S. L. THE TREX1 3 EXONUCLEASE AND AUTOIMMUNE DISEASE: STRUCTURAL AND BIOCHEMICAL ANALYSIS OF DISEASE MUTANTS INVOLVED IN AUTOIMMUNE DYSFUNCTION. 145.
37. Neet, K. E. [7] Cooperativity in enzyme function: Equilibrium and kinetic aspects. in *Methods in Enzymology* (ed. Purich, D. L.) vol. 64 139–192 (Academic Press, 1980).
38. Chowdhury, D. *et al.* The Exonuclease TREX1 Is in the SET Complex and Acts in Concert with NM23-H1 to Degrade DNA during Granzyme A-Mediated Cell Death. *Mol. Cell* **23**, 133–142 (2006).
39. 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).