Chemical Research in Toxicology

pubs.acs.org/crt

Article

Biochemical Activity of 17 Cancer-Associated Variants of DNA Polymerase Kappa Predicted by Electrostatic Properties

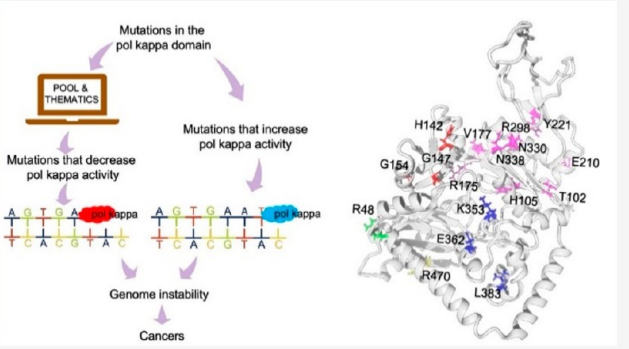
Lakindu S. Pathira Kankanamge, Alexandra Mora, Mary Jo Ondrechen,* and Penny J. Beuning*

Cite This: *Chem. Res. Toxicol.* 2023, 36, 1789−1803

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: DNA damage and repair have been widely studied in relation to cancer and therapeutics. Y-family DNA polymerases can bypass DNA lesions, which may result from external or internal DNA damaging agents, including some chemotherapy agents. Overexpression of the Y-family polymerase human pol kappa can result in tumorigenesis and drug resistance in cancer. This report describes the use of computational tools to predict the effects of single nucleotide polymorphism variants on pol kappa activity. Partial Order Optimum Likelihood (POOL), a machine learning method that uses input features from Theoretical Microscopic Titration Curve Shapes (THEMATICS), was used to identify amino acid residues most likely involved in catalytic activity. The $\mu 4$ value, a metric obtained from POOL and THEMATICS that serves as a measure of the degree of coupling between one ionizable amino acid and its neighbors, was then used to identify which protein mutations are likely to impact the biochemical activity. Bioinformatic tools SIFT, PolyPhen-2, and FATHMM predicted most of these variants to be deleterious to function. Along with computational and bioinformatic predictions, we characterized the catalytic activity and stability of 17 cancer-associated DNA pol kappa variants. We identified pol kappa variants R48I, H105Y, G147D, G154E, V177L, R298C, E362V, and R470C as having lower activity relative to wild-type pol kappa; the pol kappa variants T102A, H142Y, R175Q, E210K, Y221C, N330D, N338S, K353T, and L383F were identified as being similar in catalytic efficiency to WT pol kappa. We observed that POOL predictions can be used to predict which variants have decreased activity. Predictions from bioinformatic tools like SIFT, PolyPhen-2, and FATHMM are based on sequence comparisons and therefore are complementary to POOL but are less capable of predicting biochemical activity. These bioinformatic and computational tools can be used to identify SNP variants with deleterious effects and altered biochemical activity from a large data set.



INTRODUCTION

DNA damage can occur due to endogenous and exogenous factors and can play an important role in aging and disease.[1] DNA damage can lead to the chemical modification of the DNA bases, and this may cause collapse or stalling of the replication fork. A chemically modified DNA base can hinder normal base pairing and base stacking, leading to instability of the DNA.[2] Y-family DNA polymerases (pols) are specialized in translesion synthesis (TLS), allowing them to bypass certain types of damaged DNA sites.[3] Y-family DNA polymerases are more error prone compared with the replicative DNA polymerases, as the Y-family polymerases make limited and nonspecific contacts with the replicating base pair, relaxing base selection.[4−7]

Y-family DNA polymerases are conserved in all domains of life and can be divided into six major groups depending on their sequence.[8,9] Among them, DinB (bacterial), Dpo4 and Dbh (archaeal homologues), and pol κ (eukaryotic) have the ability to bypass bulky DNA adducts.[6,10,11] DNA pol κ is one of the four Y-family polymerases in humans. Human DNA pol κ consists of 870 amino acids, but it was determined from previous studies[12] that a construct with residues 19−526 is more stable in vitro and showed polymerase activity similar to that of full-length DNA pol κ. This construct includes the pol κ polymerase domains N-clasp, thumb, fingers, palm, and little finger/PAD. Human pol κ can add up to ∼20−30 nucleotides per binding event.[13] DNA pol κ can bypass minor groove N^2-dG adducts such as N^2-furfuryl-dG (N^2ffdG), N^2-(1-carboxyethyl)-dG, and bulkier adducts like N^2-benzo[a]pyrene diol epoxide-dG, in an error-free manner but is inhibited by major groove adducts.[14−20]

A single nucleotide polymorphism (SNP) is the substitution of a single nucleotide for another.[21] Most SNPs arise in noncoding regions, and the SNPs that arise in coding regions

Received: August 6, 2023
Revised: October 3, 2023
Accepted: October 4, 2023
Published: October 26, 2023

can lead to missense mutations, silent mutations, and nonsense mutations. Missense mutations, which change coded amino acids (nonsynonymous mutations (nsSNPs)), may produce defects in the protein and could be subjected to natural selection.[22] Almost half of the gene lesions that are associated with inherited diseases are because of amino acid substitutions.[23] Y-family DNA polymerases are responsible for point mutations that occur in cells in part due to the TLS mechanism.[24] A few SNPs of pol κ were found to be associated with breast cancer risk.[25] Some missense mutations of pol κ have also been found in nonsmall cell lung cancer patients[26] and in prostate, melanoma, lung, and large intestine cancers.[27,28] SNPs can affect protein function by either increasing or decreasing protein activity.[27] Mutations that decrease the activity of DNA pol κ on damaged DNA could make cells more sensitive to DNA damaging agents like chemotherapeutics, while mutations that show increased activity of DNA pol κ could play a role in resistance to DNA damaging chemotherapy.[13,25,26,28]

In this study, we used COSMIC[29] and Ensembl[30] databases to identify 857 cancer-associated DNA pol κ SNPs. POOL and THEMATICS[31−36] were used to identify catalytically active residues of pol κ and whether the SNP-encoded mutation alters the chemical and electrostatic properties of these catalytically active residues, as measured by the computed μ4 values of the catalytic amino acids for each protein variant. The μ4 value of an amino acid in an enzyme is a measure of the degree of coupling of its protonation equilibrium with those of nearby amino acids and is used successfully to predict the likelihood that this residue is biochemically active.[32,33,37−40]

THEMATICS is based on a solution to the Poisson−Boltzmann (P−B) equations[41,42] to obtain the electrical potential function of the 3D structure of the protein and then uses a hybrid procedure[43] to evaluate the average proton occupations of each ionizable amino acid as a function of pH. The residues that are involved in catalysis show different chemical properties compared to other residues,[34] namely expanded buffer ranges and theoretical titration curves that deviate most strongly from Henderson−Hasselbalch behavior. To identify the active site residues from the 3D structure of the query protein, THEMATICS uses the titration curve shapes that are described by the moments of their first derivative functions. These derivative functions are also probability density functions, and their central moments are used to define the asymmetry and kurtosis of the titration curves. The fourth central moment, μ4, kurtosis, has been shown to be the best predictor of biochemical activity.[32] We used the fourth central moments of the catalytically active residues as metrics to identify which SNP variants would most substantially alter the proton occupancy behavior of these biochemically active residues when compared with those of WT pol κ;[32−34,44] 17 SNP variants were selected. The bioinformatics tools Sorting Intolerant From Tolerant (SIFT),[45] Polymorphism Phenotyping (PolyPhen-2),[46] and Functional Analysis Through Hidden Markov Models (FATHMM),[47] which are based on amino acid sequence and conserved amino acids, were used to predict whether an amino acid substitution can cause deleterious effects. We tested the activity of the variants using primer extension, misincorporation, steady-state kinetics, and thermal shift assays with undamaged and damaged DNA substrates. The variants R48I, H105Y, G147D, G154E, V177L, R298C, E362V, and R470C showed lower activity than WT pol κ, while the others tested showed only minor or no differences in activity. We find that changes in μ4 values are better predictors of changes in activity than the other prediction methods used here.

EXPERIMENTAL PROCEDURES

Bioinformatic and Computational Analyses. Cancer-associated DNA pol κ variants were identified by using COSMIC[29] and ENSEMBL[30] databases. The FASTA sequence for DNA polymerase kappa was obtained from Uniprot (Q9UBT6).[48] A 2.0 Å crystal structure is reported for human pol kappa (PDB ID 6CST[48]); however, because this structure has a missing loop, a homology model was created using the homology module in YASARA.[49,50] Model quality was assessed using the servers QMEAN,[51] PROCHECK,[52] and ANOLEA.[53] We identified 857 protein variants associated with cancer, and 421 of these are located in the polymerase domain; SNP mutations were created in the homology models for the 421 variants using Schrödinger's Maestro.[54] Energy minimization for each of the protein variants was done using YASARA.[55] Biochemically active residues of DNA pol κ (Uniprot ID Q9UBT6) and respective μ4 values were identified by POOL and THEMATICS.[31−34] Seventeen variants that showed at least a ±20% change in THEMATICS calculated μ4 values for the catalytically active residues relative to WT pol κ were chosen (Table S1) for biochemical characterization. SIFT,[45] PolyPhen-2,[46] and FATHMM[47] scores were also used to predict the effect of SNP variants on function.

Schrödinger's Maestro[54] was used to create homology models of each of the pol κ SNP variants. It also created an output file with the change in stability of each SNP variant calculated using the Prime energy function with an implicit solvent term. This stability is the free energy difference between the unfolded and folded state. A negative value in the stability means that the protein is more stable than the WT, while a positive value indicates lower stability of the protein.[56]

Proteins and DNA. The truncated pol κ construct with residues 19−526 was expressed and purified as previously described.[27,57] This construct is more stable in vitro and shows similar polymerase activity to full-length DNA pol κ.[12] DNA pol κ variants were created using Quickchange site directed mutagenesis kits (Agilent). The presence of mutations in these constructs was confirmed by DNA Sequencing (Eton Bioscience, Charlestown, MA).

The DNA template containing a single N^2ffdG was prepared as described previously.[16,58] The undamaged DNA (T25) and Primers (P13, MatchC, and MatchT) were from Eurofins Operon or Integrated DNA Technologies (IDT) (Table 1). DNA was purified using urea-polyacrylamide gel electrophoresis and the crush and soak method.[59] DNA Primer13, MatchC, and MatchT were labeled with ^{32}P as previously described.[59]

Table 1. DNA Template and Primer Sequences

DNA	Length	Sequence
T25[a]	25 mer	5′ TGCGAAGAACTXGCGTCCGGCAAGC 3′
ddP13[b]	13 mer	5′ GCTTGCCGGACGC 3′
P13	13 mer	5′ GCTTGCCGGACGC 3′
MatchC	14 mer	5′ GCTTGCCGGACGCC 3′
MatchT	14 mer	5′ GCTTGCCGGACGCT 3′

[a]X = G, N^2ffdG. [b]dd = dideoxy-terminated.

Primer Extension and Misincorporation Assays. DNA template T25 (with and without N^2ffdG) was combined with ^{32}P-labeled primers in a 1:1 ratio (500 nM) and annealed in annealing buffer (20 mM HEPES, pH 7.5, 5 mM $Mg(OAc)_2$) by heating for 2 min at 95 °C, incubating 1 h at 50 °C, and cooling to 37 °C. Primer extension assay reactions were carried out in 1× reaction buffer (30 mM HEPES, pH 7.5, 20 mM NaCl, 7.5 mM $MgSO_4$, 2 mM β-mercaptoethanol, and 1% (w/v) bovine serum albumin) with 100 nM labeled ^{32}P primer/template, 2 nM pol κ, and 500 μM dNTPs. Misincorporation assay reactions were carried out in 1× reaction buffer with 100 nM ^{32}P-labeled primer/template, 2 nM pol κ, 1 mM

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

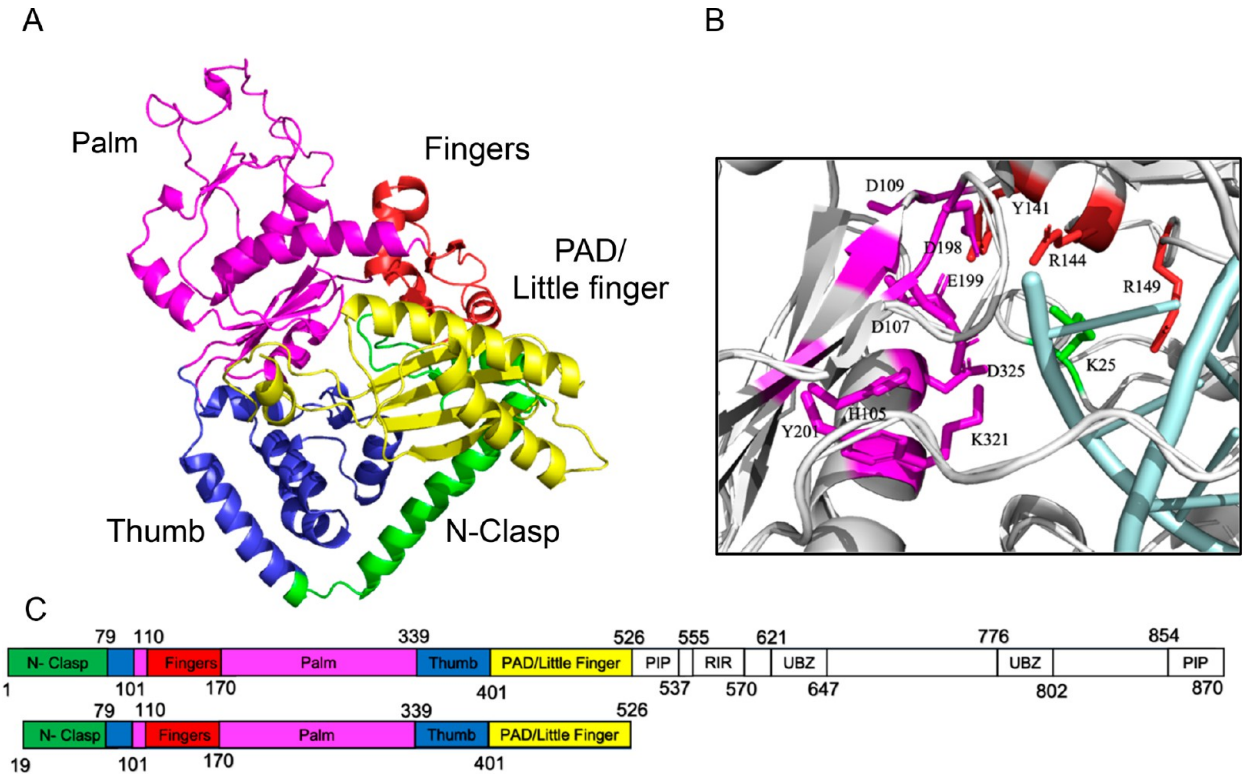


Figure 1. DNA pol κ domains and catalytically active residues. Homology model of DNA pol κ (PDB ID: 6cst served as template) modeled using YASARA (A) with the five polymerase domains indicated, N-clasp (green), Fingers (red), Thumb (blue), PAD/Little Finger domain (yellow), and Palm (pink). (B) POOL-predicted active site residues of DNA pol κ. DNA is in cyan. (C) DNA pol κ construct from 19 to 526 residues, which is more stable in vitro compared with the full-length DNA pol κ.

Table 2. DNA pol κ SNP Association in Cancer

ID	Amino acid mutation	Domain	Tumor site	Reference[a]
COSV104385564	R48I	N-Clasp	Large intestine	64
COSV54039112	T102A	Palm	Breast	N/A
COSV54040524	H105Y	Palm	Large intestine	65
COSV54037156	H142Y	Fingers	Large intestine, Upper aerodigestive tract	66, 67
COSV54035583	G147D	Fingers	Large intestine	68
COSV54033403	G154E	Fingers	Stomach	N/A
COSV54036288	R175Q	Palm	Prostate, Small intestine	69, 70
COSV99062812	V177L	Palm	Lung	71
COSV54034453	E210K	Palm	Kidney	72
COSV54037397	Y221C	Palm	Endometrium	N/A
COSV54034412	R298C	Palm	Skin, Endometrium, Large intestine	73−75
COSV54031385	N330D	Palm	Skin	76
COSV54032071	N338S	Palm	Malignant melanoma	77
COSV54037023	K353T	Thumb	Endometrium	N/A
COSV54039187	E362V	Thumb	Esophagus	N/A
COSV54033877	L383F	Thumb	Endometrium	N/A
COSV54031865	R470C	Little Finger	Endometrium, Skin	N/A

[a]N/A − References for these were not available

dATP, dCTP, dGTP, and dTTP. Reaction products were separated by denaturing 16% 8 M urea polyacrylamide gels, detected by phosphorimaging on a Typhoon (GE), and analyzed with ImageQuant TL1D gel analysis software to determine percent nucleotide incorporation. The percentages were calculated relative to the amount extended by WT pol κ at the last time point.

Steady-State Kinetic Assays. Steady-state kinetic assays were carried out using the initial rates method[60] with 100 nM ^{32}P-labeled primer/template and primer/N^2ffdG-containing DNA template, 2 nM WT pol κ or protein variants, and varying concentrations of dCTP (0.25−800 μM) as the incoming nucleotide. All reactions were performed in 1× reaction buffer. Six time points for each dCTP concentration were taken, including the zero-time point. All assays were carried out within the linear range and conducted in triplicate. Reaction velocities were plotted as a function of [dCTP], and data were fit using GraphPad Prism 5.02 for Windows (GraphPad Software San Diego, CA, USA) to the Michaelis−Menten equation. Errors reported are standard deviations.

Thermal Shift Assay. Thermal shift assays were conducted using a Bio-Rad CFX 96 Real-time PCR instrument as previously

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

Table 3. Bioinformatic and Computational Tools Predictions for Cancer-Associated DNA Polymerase κ SNP Variants[d]

Protein Variant	SIFT	PolyPhen-2	FATHMM	Overall prediction[a]	Observed Δμ4 (%)[b]			ΔStability[c] (Solvated)
					D107	D198	E199	
R48I	Affect Protein Function	Possibly damaging	Pathogenic	Deleterious	-59%	-18%	-2%	13.3
T102A	Tolerated	Benign	Pathogenic	Not deleterious	7%	8%	2%	-3.39
H105Y	Tolerated	Possibly damaging	Pathogenic	Deleterious	-14%	59%	58%	998
H142Y	Tolerated	Benign	Pathogenic	Not deleterious	20%	8%	3%	-1.55
G147D	Affect Protein Function	Possibly damaging	Pathogenic	Deleterious	52%	6%	5%	685
G154E	Tolerated	Possibly damaging	Pathogenic	Deleterious	-58%	-16%	-8%	34.2
R175Q	Tolerated	Benign	Pathogenic	Not deleterious	14%	11%	-5%	2.98
V177L	Affect Protein Function	Possibly damaging	Pathogenic	Deleterious	11%	52%	2%	90.3
E210K	Tolerated	Possibly damaging	Pathogenic	Deleterious	10%	4%	3%	4.61
Y221C	Tolerated	Possibly damaging	Pathogenic	Deleterious	22%	10%	10%	8.70
R298C	Affect Protein Function	Possibly damaging	Pathogenic	Deleterious	56%	15%	8%	240
N330D	Tolerated	Possibly damaging	Pathogenic	Deleterious	23%	10%	8%	-5.85
N338S	Tolerated	Benign	Neutral	Not deleterious	7%	2%	11%	-5.27
K353T	Tolerated	Possibly damaging	Pathogenic	Deleterious	20%	19%	2%	-6.09
E362V	Tolerated	Possibly damaging	Pathogenic	Deleterious	54%	1%	-58%	235
L383F	Tolerated	Possibly damaging	Pathogenic	Deleterious	10%	12%	-7%	-4.38
R470C	Affect Protein Function	Possibly damaging	Pathogenic	Deleterious	5%	55%	-1%	312

[a]Overall prediction is a "majority rules" outcome from the three bioinformatics methods, SIFT, PolyPhen-2, and FATHMM. SIFT score is a normalized probability of having a different amino acid at that position, where the score ranges from 0 to 1. A score between 0 and 0.05 predicts an effect on protein function and is tolerated when the obtained score is ≥0.05.[45,47] PolyPhen-2 score is the probability of a substitution being damaging. PolyPhen-2 scores ranging from 0 to 0.15 are predicted to be benign, scores ranging from 0.15 to 0.85 are predicted to be possibly damaging, and scores from 0.85 to 1.0 are predicted to be damaging.[80,81] FATHMM score represents the functional scores obtained for each mutation from the FATHMM-MKL model and are in the form of a p-value which ranges from 0 to 1. Scores >0.5 are deleterious, >0.7 are pathogenetic (i.e., highly deleterious), and ≤0.5 are neutral.[47,82] [b]Change in μ4 for the POOL predicted and known active site residues (complete data set is in Table S1) D107, D198, and E199, compared with WT pol κ for each SNP variant. [c]Solvated change in stability was predicted by Schrödinger Maestro.[56] [d]Variants that have at least 50% change in at least one catalytic residue are highlighted in green.

described.[61−63] The assays were conducted with 10 μM dideoxy-terminated primer13 and undamaged T25 or damaged T25 with N^2ffdG template DNA (Table 1), 5 μM WT pol κ or protein variants, and 1 mM dNTP (dCTP, dTTP) in 15 μL of 1× reaction buffer. Reactions were incubated at room temperature (22 °C) before the addition of SYPRO Orange (Invitrogen) to a final concentration of 20×. Fluorescence was detected over a temperature range of 10−90 °C heated at a rate of 0.2 °C/min for all assays. All reactions were carried out in triplicate, and the error bars represent the standard deviation.

RESULTS

We used computational tools first to identify the most catalytically important residues in pol κ and then applied these computational tools to identify 17 pol κ SNP variants that are associated with cancer and are predicted to have altered biochemical functions, comparing these predictions with those from bioinformatic tools. We evaluated the activity of SNP variants in vitro with primer extension assay, misincorporation assay, steady-state kinetics assay, and thermal stability assay and categorized the SNP variants according to their effects on biochemical function relative to those of WT pol κ.

Bioinformatic and Computational Analyses. POOL and THEMATICS identified K25, H105, D107, D109, Y141, R144, R149, D198, E199, Y201, E296, K321, D325, K328, and K353 as catalytically important residues of DNA pol κ, and we note that most of the POOL-predicted active residues are in the palm region of the protein (Figure 1). Of these, D107, D198, and E199 have been identified previously as catalytically active site residues.[2] The μ4 values of all 15 predicted catalytically active residues were calculated for each of the 421 cancer-associated SNP variants located in the polymerase domain. We sorted variants that showed 20% or greater change in the μ4 value in at least one POOL-predicted active site residue compared with the values for these active residues in WT pol κ. Surprisingly, out of 421 SNP variants, only 17 showed a 20% or greater change in the μ4 value of a predicted biochemically active amino acid (Table S1). The 17 protein variants are R48I, T102A, H105Y, H142Y, G147D, G154E, R175Q, V177L, E210K, Y221C, R298C, N330D, N338S, K353T, E362V, L383F, and R470C. These mutations are distributed in the DNA pol κ coding region and are associated with various cancers (Table 2). The likely effects on the activity of these 17 mutations were then analyzed with bioinformatic tools. SIFT and PolyPhen-2 use sequence similarity of related proteins to predict whether an amino acid substitution is likely to affect function based on the degree of conservation of the affected amino acid throughout evolution.[78] FATHMM uses Hidden Markov Models (HMM) to predict functional consequences of amino acid substitutions[47] using both sequence and structural information.[79] SIFT, PolyPhen-2, and FATHMM predictions for each of the 17 pol κ variants are listed in Table 3. SIFT predicted that R48I, G147D, V177L, R298C, and R470C could affect pol κ function. PolyPhen-2 predicted that except for T102A, H142Y, R175Q, and N338S, all the other mutations are

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

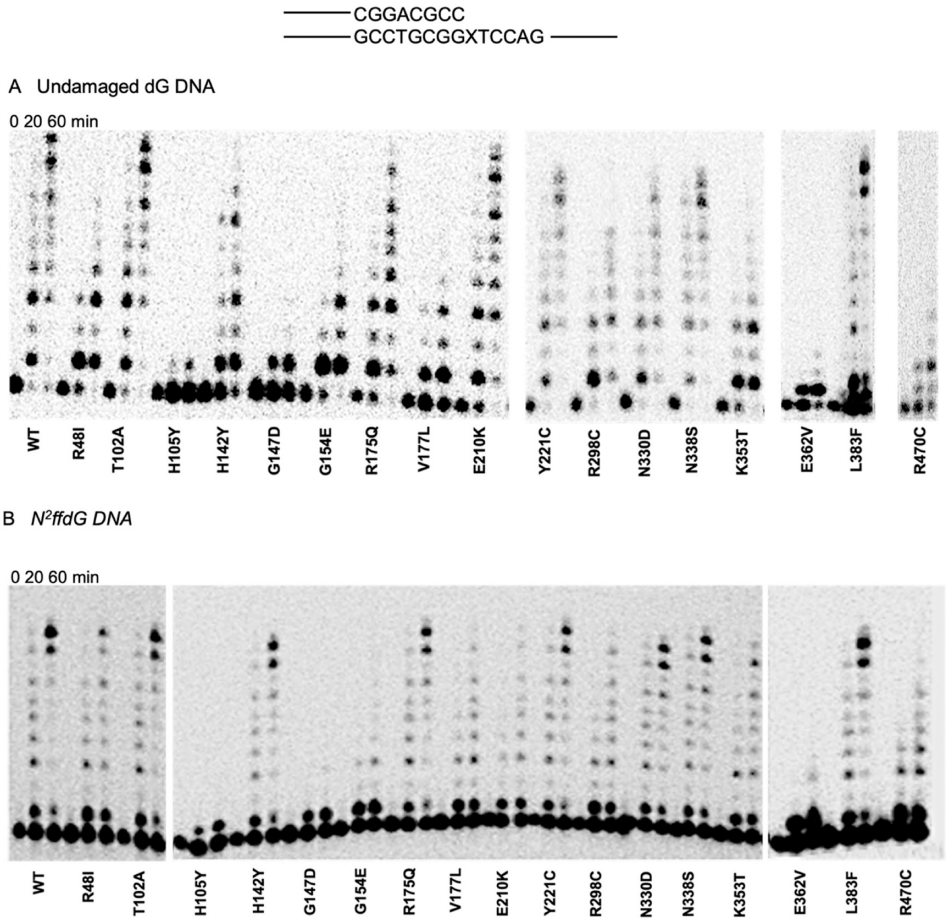


Figure 2. Primer extension activity of pol κ variants. DNA pol κ SNP variants show different extension patterns in the presence of both undamaged (A) and minor groove damaged N^2ffdG (B) DNA templates. Variants T102A, H142Y, R175Q, E210K, Y221C, N330D, N338S, K353T, and L383F showed a similar primer extension as WT pol κ on undamaged DNA templates and damaged DNA templates, while variants R48I, H105Y, G147D, G154E, V177L, R298C, E362V, and R470C showed less primer extension compared to WT on undamaged pol κ. Primer extension was carried out in the presence of dNTP mix. Time points are 0, 20, and 60 min for each variant; quantitation is reported in Table S2 and Figure S1.

possibly damaging. FATHMM predicted that all except N338S are pathogenic. Compiled results, based on at least two of the three bioinformatic tools, suggest that SNPs encoding pol κ variants T102A, H142Y, R175Q, and N338S are not deleterious or do not affect the protein function, while other pol κ variants R48I, H105Y, G147D, G154E, V177L, E210K, Y221C, R298C, N330D, K353T, E362V, L383F, and R470C could affect protein function (Table 3).

POOL and THEMATICS generated the fourth central moments for all of the catalytically active residues; μ4 was analyzed for WT pol κ and all 17 pol κ variants (Table S1). The pol κ variants R48I, H105Y, G147D, G154E, V177L, R298C, E362V, and R470C have more than a 50% change in μ4 values in at least one catalytically active residue when compared with WT pol κ; these are the largest changes in μ4 values among the pol κ variants (Table 3). A change in fourth central moment μ4 value in the catalytic residues D107, D198, and E199 could directly impact catalytic function. Stability analysis shows that pol κ variants H105Y, G147D, R298C, E362V, and R470C have high positive values, indicating their relative loss of stability, whereas pol κ variants T102A, H142Y, N330D, N338S, K353T, and L383F have negative values, indicating that they are not destabilized relative to WT pol κ (Table 3).

All Eight Protein Variants with the Largest Changes in μ4 Are Less Active than WT pol κ. The primer extension and insertion ability of pol κ variants in the presence of undamaged and damaged (N^2ffdG) DNA were determined with Primer13:Template25. The Primer13 3′ end binds Template25 one base before the undamaged or damaged site "X", and synthesis starts there in the presence of DNA pol κ. At the 60 min time point, SNP variants pol κ T102A, H142Y, R175Q, E210K, Y221C, N330D, N338S, K353T, and L383F showed similar activity as WT pol κ with undamaged DNA (Figure 2). The pol κ SNP variants R48I, V177L, R298C, K353T, and R470C were observed to be more efficient in the presence of minor groove adduct N^2ffdG DNA than in the presence of undamaged DNA (Figure 2, Figure S1, and Table S2). The pol κ SNP variants H105Y, G147D, G154E, V177L, R298C, E362V, and R470C showed lower activity than that of WT pol κ with both undamaged dG and damaged N^2ffdG. DNA pol κ variants H105Y, G147D, and E362V were only able to add 1−3 nucleotides efficiently, which was observed in the presence of either undamaged DNA or DNA harboring the N^2ffdG adduct (Figure 2).

Next, the extension ability of WT pol κ and pol κ SNP variants was determined using the MatchC primer, which harbors the correct pair of a C with a template damaged base N^2ffdG and measures the ability of pol κ variants to extend

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

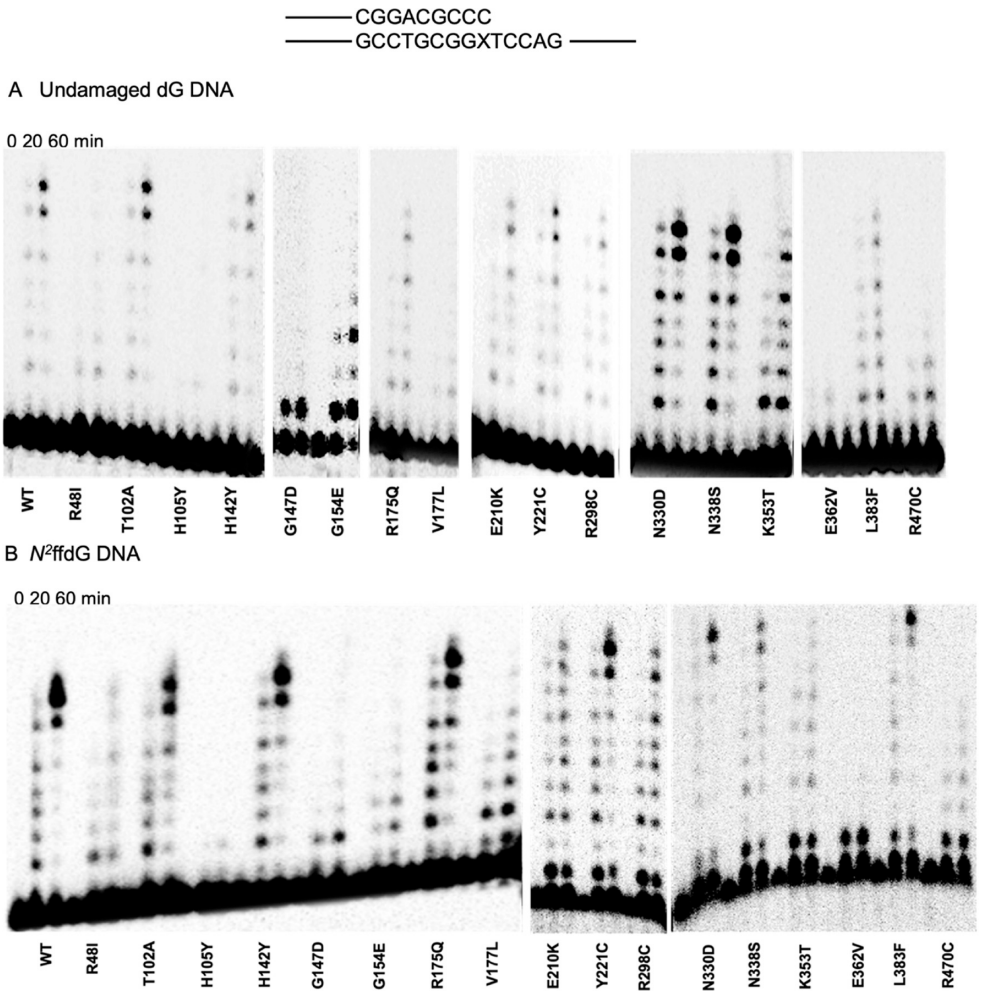


Figure 3. Primer extension in the presence of MatchC primer. DNA pol κ SNP variants except H105Y, G147D, V177L, E362V, and R470C show similar extension patterns in the presence of both undamaged (A) and minor groove damaged N^2ffdG (B) DNA templates. Time points are 0, 20, and 60 min for each variant; quantitation is reported in Table S3, Figure S2.

DNA replication following a base pair containing a damaged base. At the 60 min time point, the pol κ protein variants R48I, H105Y, G147D, G154E, V177L, E362V, and R470C, while active, failed to extend the primer fully in the presence of undamaged DNA (Figure 3, Table S3, Figure S2). The pol κ variant H105Y, however, had low activity with all DNA constructs tested.

Finally, the extension ability of WT pol κ and the 17 pol κ variants was tested with MatchT primer, in which the 3′ T is paired with a template dG or damaged template base N^2ffdG. We observed primer extension by WT pol κ and pol κ T102A, H142Y, Y221C, N330D, N338S, and L383F in the presence of undamaged and damaged DNA (Figure 4, Table S4, and Figure S3). In addition, pol κ variants R175Q, E210K, and K353T exhibited primer extension activity in the presence of damaged base N^2ffdG but not undamaged DNA (Figure 4, Table S4, and Figure S3), indicating that these variants have a specialized ability to extend from a mismatched, damaged base pair.

SNP Variants Showed a Similar Nucleotide Incorporation Pattern as WT pol κ. In the presence of Primer13 and undamaged dG template DNA, we observed primarily the correct incorporation of dC for all protein variants. Misincorporation of dA, a second and a third dC, a low level of dG, and dT was observed with WT pol κ. Misincorporation of dA was observed with pol κ variants T102A, H142Y, G154E, R175Q, V177L, E210K, Y221C, R298C, N330D, and L383F. Misincorporation of dT was observed with all pol κ variants, except R48I, E362V, and R470C. DNA pol κ variants H142Y, R175Q, Y221C, N338S, and K353T misincorporated dG (Figure 5A)

In the presence of Primer13 DNA with N^2ffdG as the template base, we observed the same correct incorporation of dC and misincorporation of dT opposite N^2ffdG in all variants. WT pol κ misincorporated dT and weakly misincorporated two dA nucleotides as well as dG. Protein variants T102A, H142Y, G154E, R175Q, V177L, E210K, Y221C, R298C, N330D, N338S, K353T, and L383F misincorporated dA. Weak misincorporation of dG was observed with pol κ variants H142Y, R175Q, and Y221C (Figure 5B). Most pol κ variants displayed higher misincorporation opposite N^2ffdG than that with undamaged template DNA, similar to WT pol κ (Figure 5). The pol κ variant H142Y had the greatest extent of misincorporation on both undamaged and damaged DNA.

The pattern of incorporation immediately following a correct base pair containing an undamaged dG or lesion N^2ffdG was examined by using the primer MatchC with the respective DNA templates. The next nucleotide base after the dG/N^2ffdG lesion site in the DNA template is dT; therefore, the correct incorporation opposite dT is dA. WT pol κ and the

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

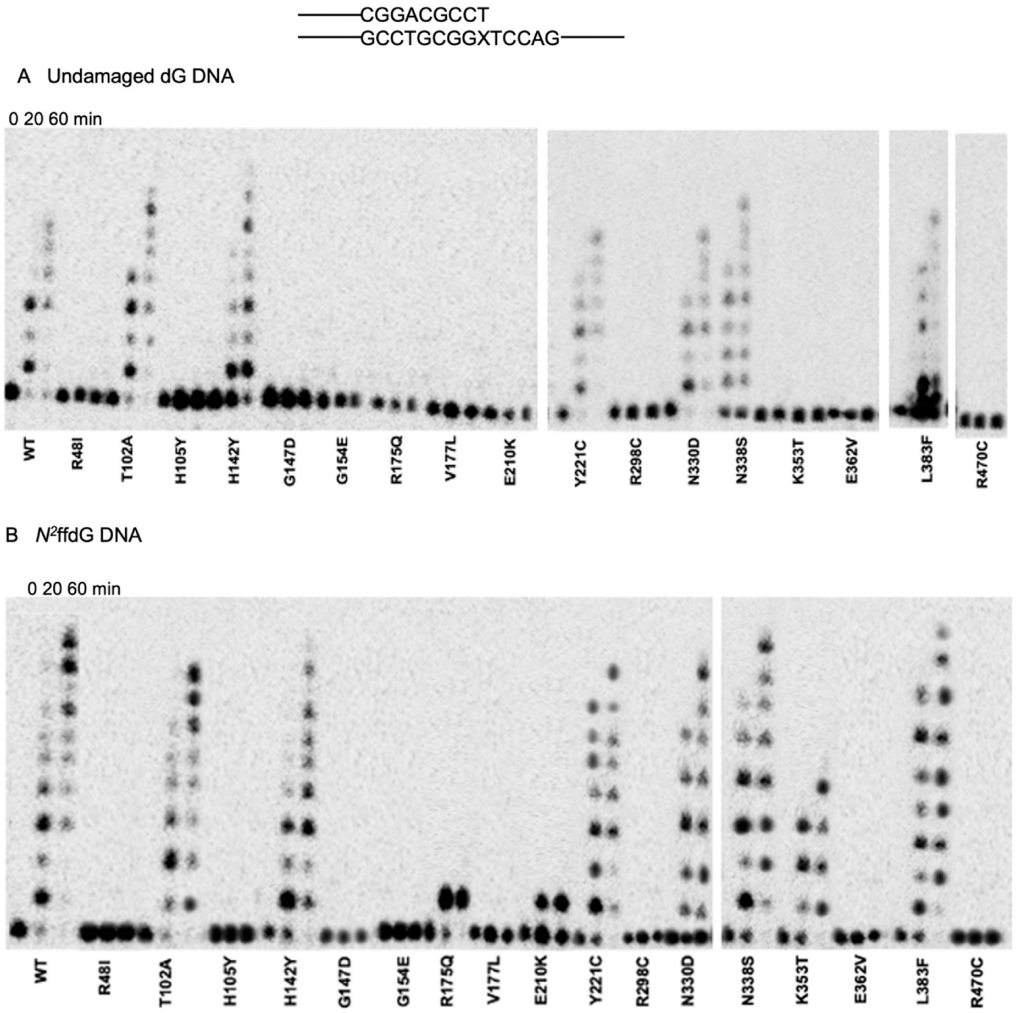


Figure 4. Primer extension in the presence of MatchT primer. DNA pol κ SNP variants T102A, H142Y, Y221C, N330D, N338S, and L383F showed similar primer extension patterns in the presence of both undamaged (A) and minor groove damaged N^2ffdG (B) DNA templates. Variants R175Q, E210K, and K353T showed primer extension only in the presence of damaged N^2ffdG. Time points are 0, 20, and 60 min for each variant; quantitation is reported in Table S4, Figure S3.

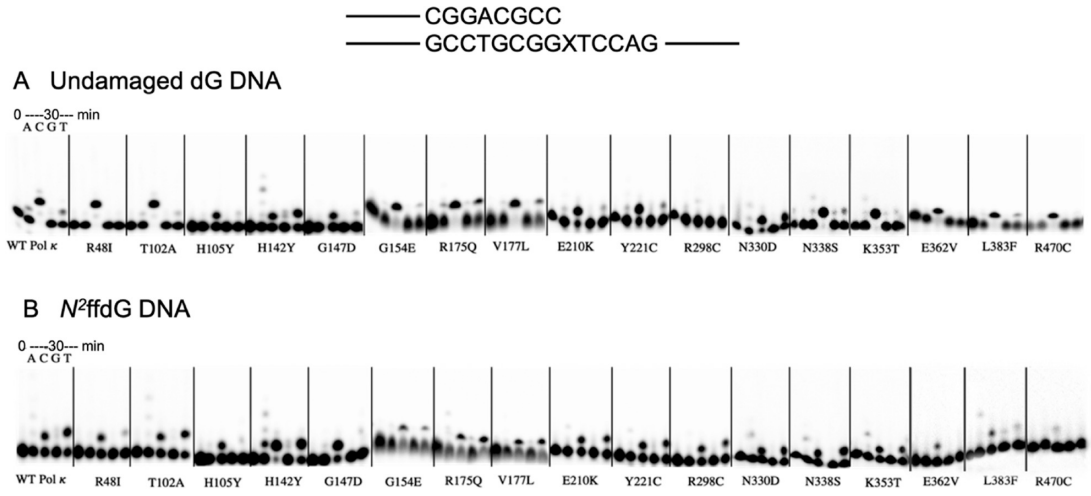


Figure 5. Correct incorporation of dC and misincorporation of nucleotides by WT pol κ and protein variants in the presence of Primer13 and DNA substrate. Incorporation and misincorporation by WT pol κ and pol κ SNP variants in the presence of (A) undamaged dG and (B) DNA harboring damaged N^2ffdG. DNA pol κ protein variants incorporate the correct nucleotide dC opposite the dG position of undamaged and damaged DNA template.

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

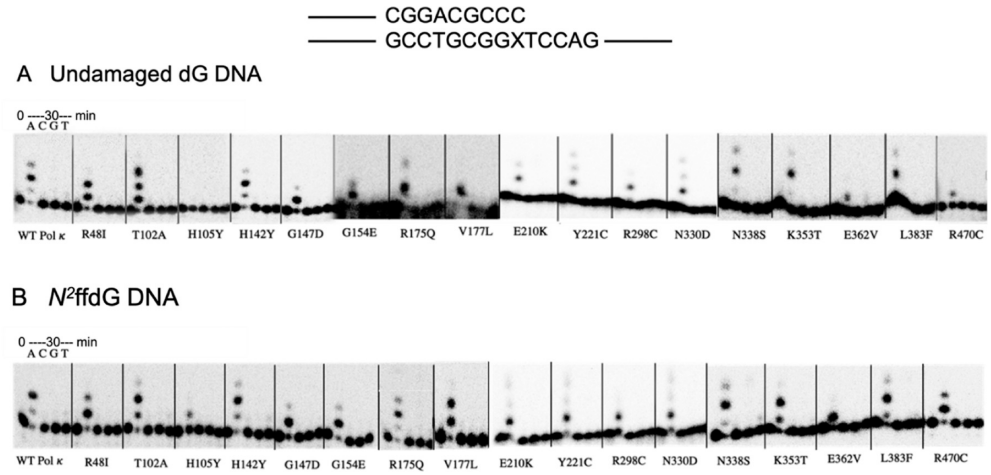


Figure 6. Incorporation and misincorporation by pol κ variants in the presence of MatchC primer and DNA substrates. WT pol κ and pol κ SNP variants incorporate the correct nucleotide dA opposite the dT position in (A) undamaged and (B) damaged DNA template harboring N^2ffdG.

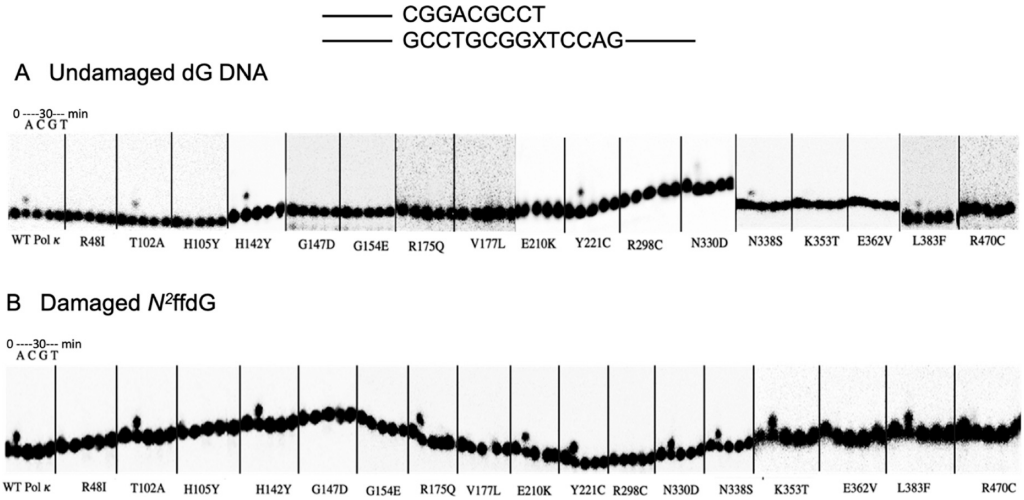


Figure 7. Incorporation by pol κ protein variants in the presence of MatchT primer and DNA substrates. WT pol κ and some pol κ SNPs in the presence of mismatch base pair dT:dG or dT:N^2ffdG incorporate the correct nucleotide dA opposite the dT position of (A) undamaged and (B) damaged DNA template harboring N^2ffdG.

SNP variants correctly inserted dA, but in most cases, multiple dAs were incorporated (Figure 6). Lower frequency of incorporation of dA was apparent with pol κ SNP variants G147D, G154E, V177L, R298C, and R470C. DNA pol κ SNP variants H105Y and E362V were observed to add only one dA in the presence of the damaged N^2ffdG DNA template (Figure 6).

When incorporation starts from a mismatch base pair (dT:dG or dT:N^2ffdG), WT pol κ was able to incorporate the correct next base dA, which has been observed previously.[27,83−85] Likewise, other pol κ SNP variants T102A, H142Y, Y221C, N330D, N338S, and L383F were able to do so in the presence of undamaged DNA, some of them more robustly than WT pol κ (Figure 7A). In the presence of N^2ffdG, pol κ SNP variants T102A, H142Y, R175Q, E210K, Y221C, N330D, N338S, K353T, and L383F were able to incorporate the correct dA base (Figure 7B). Only a few of the pol κ SNP variants are capable of extension from the mismatch base pair; extension from a mismatch containing a damaged DNA base could result in a mutation being fixed in the genome upon repair.

Steady-State Kinetics Analyses. WT pol κ and variants were titrated with different dCTP concentrations varying from 0.25−800 μM until they reached a saturation level to carry out steady-state kinetics analysis (Tables 4 and 5; Figures S4−S6). DNA pol κ variants R48I, H105Y, G147D, G154E, V177L, R298C, E362V, and R470C show lower catalytic efficiency, while other variants have efficiency similar to WT pol κ with undamaged DNA. In general, pol κ and the SNP variants showed increased catalytic efficiency in the presence of damaged N^2ffdG adduct. The pol κ SNP variants with reduced activity on undamaged DNA also showed reduced activity in the presence of the DNA adduct N^2ffdG.

Stability of DNA Polymerase κ Variants with Preferred Substrates. The overall stability of WT pol κ and variants in the presence of their substrates was assessed by using a thermal stability assay. 10 of the 17 DNA pol κ SNP variants had a similar melting temperature (T_m) as WT pol κ (42.1 ± 0.6 °C), while pol κ variants R48I (39.3 ± 0.4 °C), H105Y (38.5 ± 0.5 °C), G147D (38.6 ± 0.4 °C), R298C (33.7 ± 0.3 °C), E362V (38.7 ± 0.2 °C), and R470C (38.9 ± 0.5 °C) (Figure 8, Table S5) showed a lower stability of ≥2.8 °C compared to WT pol κ. DNA pol κ variant R298C showed

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

Table 4. Kinetics Analysis for WT pol κ and 17 Variants in the Presence of Undamaged DNA Template

Protein	Undamaged dG										Fold decrease
	V_{max} (μMmin^{-1})			k_{cat} (s^{-1})			K_m (μM)			Catalytic efficiency[a] ($M^{-1}s^{-1}$)	
WT	1.12E-01	±	1.83E-03	9.31E-01	±	1.52E-02	3.01E+01	±	1.98E+00	3.04E+04 ± 1.65E+03	1
R48I	8.10E-02	±	1.11E-03	6.71E-01	±	9.17E-03	6.67E+01	±	2.96E+00	1.01E+04 ± 7.33E+02	3.0
T102A	1.32E-01	±	4.89E-03	1.11E+00	±	4.01E-02	2.81E+01	±	3.13E+00	3.93E+04 ± 2.98E+03	0.7
H105Y	8.30E-02	±	3.11E-03	6.93E-01	±	2.59E-02	1.04E+02	±	1.29E+01	6.65E+03 ± 6.88E+02	4.5
H142Y	1.35E-01	±	1.23E-03	1.13E+00	±	1.43E-02	2.84E+01	±	3.35E+00	3.95E+04 ± 1.13E+03	0.7
G147D	1.22E-01	±	6.84E-03	1.02E+00	±	5.71E-02	1.09E+02	±	1.71E+01	9.38E+03 ± 8.99E+02	3.2
G154E	1.11E-01	±	9.76E-03	9.25E-01	±	8.13E-02	9.01E+01	±	1.99E+01	1.02E+04 ± 1.49E+03	3.0
R175Q	1.21E-01	±	1.43E-03	1.01E+00	±	1.19E-02	3.27E+01	±	5.07E+00	3.06E+04 ± 4.06E+03	0.9
V177L	1.07E-01	±	7.61E-03	8.89E-01	±	6.34E-02	8.08E+01	±	9.01E+00	1.01E+04 ± 7.05E+03	3.0
E210K	1.25E-01	±	7.13E-03	1.04E+00	±	5.94E-02	3.39E+01	±	4.64E+00	3.07E+04 ± 3.05E+03	0.9
Y221C	1.31E-01	±	7.51E-03	1.09E+00	±	6.25E-02	2.75E+01	±	1.69E+00	3.97E+04 ± 2.99E+03	0.7
R298C	1.36E-01	±	1.42E-02	1.13E+00	±	1.18E-01	8.54E+01	±	1.59E+01	1.06E+04 ± 1.54E+03	2.9
N330D	1.31E-01	±	1.36E-02	1.09E+00	±	1.13E-01	2.46E+01	±	9.89E+00	4.42E+04 ± 1.14E+04	0.7
N338S	1.41E-01	±	1.33E-02	1.17E+00	±	1.11E-01	2.61E+01	±	6.21E+00	4.50E+04 ± 8.05E+03	0.7
K353T	1.29E-01	±	7.45E-03	1.08E+00	±	6.21E-02	3.61E+01	±	5.07E+00	2.98E+04 ± 2.40E+03	0.9
E362V	1.05E-01	±	3.37E-03	8.73E-01	±	2.81E-02	1.07E+02	±	1.29E+01	8.15E+03 ± 6.88E+02	3.8
L383F	1.36E-01	±	2.61E-03	1.13E+00	±	2.06E-02	2.51E+01	±	2.75E+00	4.51E+04 ± 4.23E+03	0.7
R470C	1.15E-01	±	2.44E-02	9.60E-01	±	1.02E-01	8.36E+01	±	1.01E+01	1.05E+04 ± 1.10E+03	2.9

[a]Catalytic efficiency is k_{cat}/K_m. Variants that show lower catalytic activity (2.9-fold or more) are highlighted in pale red.

the lowest stability of all at >8 °C lower than WT pol κ. WT pol κ and the pol κ SNP variants showed an increase in T_m (2.7−7.1 °C) with the addition of any primer:template DNA[61,62,86] and greater stability with the addition of correct nucleotide dCTP (10.9−19.9 °C) for all protein variants. The addition of an incorrect nucleotide, dTTP, led to a modest stabilization relative to the respective pol κ:DNA binary complexes (Figure 8, Table S5).

DISCUSSION

DNA pol κ is involved in nucleotide excision repair (NER) in mammalian cells and is known to bypass minor groove DNA adducts, whereas it is inhibited by major groove DNA adducts.[14−20,87,88] In this work, we applied bioinformatic methods and a structure-based method that computes electrostatic and chemical properties to identify 17 cancer-associated pol κ variants that are most likely to impact biochemical activity (Figure 9). We categorized these variants into two groups, the variants that showed lower activity than WT pol κ and the variants that showed similar activity as WT pol κ. We identified pol κ SNP variants R48I, H105Y, G147D, G154E, V177L, R298C, E362V, and R470C as variants with lower activity, whereas pol κ T102A, H142Y, R175Q, E210K, Y221C, N330D, N338S, K353T, and L383F showed activity similar to WT pol κ. Our analysis did not identify variants with higher activity than WT pol κ, although previous work did identify one such variant, pol κ S423R.[27,89] Primer extension and steady-state kinetics analysis show that the variants have increased activity in the presence of the N^2ffdG DNA adduct relative to undamaged DNA. The eight variants for which lower activity was observed with undamaged or damaged DNA or both coincide exactly with the eight variants that have a change in μ_4 value of 50% or more from that of WT for one or more catalytic residues.

All the DNA polymerase variants showed similar patterns in their thermal stability data when substrate or ligands are present, with the presence of DNA substrate increasing the melting temperature. The presence of the correct incoming nucleotide provides additional stabilization. In general, variants that showed lower activity (R48I, H105Y, G147D, G154E, V177L, R298C, E362V, and R470C) also showed a thermal stability lower than that of WT pol κ to varying degrees.

The DNA pol κ variant R48I resides in the N-clasp, which forms additional contacts between the polymerase and DNA.[90] Based on homology models, both R48 and I48 make contacts with S47, F49, Y50, G51, N52, V463, N464, and F465 while R48 makes an extra contact with R507 (Figure 10). The pol κ R48I variant shows diminished extension capability and somewhat lower activity with undamaged DNA compared to WT pol κ. When the damaged site is paired with the correct nucleotide base, pol κ R48I was able to extend the primer. SIFT, PolyPhen-2, and FATHMM predicted correctly that R48I mutation affects protein function or is damaging/deleterious. Also, THEMATICS showed >50% change in the

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

Table 5. Kinetics Analysis for WT pol κ and 17 Variants in the Presence of Damaged DNA Template

Protein	Damaged N^2ffdG			
	V_{max} (μMmin^{-1})	k_{cat} (s^{-1})	K_m (μM)	Catalytic efficiency[a] ($M^{-1}s^{-1}$)	Fold decrease
WT	1.38E-01 ± 6.94E-03	1.15E+00 ± 5.78E-02	2.14E+01 ± 2.06E+00	5.49E+04 ± 3.94E+03	1
R48I	9.20E-02 ± 3.76E-03	7.70E-01 ± 3.13E-02	3.25E+01 ± 3.61E+00	2.01E+04 ± 1.67E+03	2.7
T102A	1.44E-01 ± 5.44E-03	1.19E+00 ± 4.53E-02	2.05E+01 ± 1.99E+00	5.85E+04 ± 3.50E+03	0.9
H105Y	1.03E-01 ± 6.43E-03	8.61E-01 ± 5.35E-02	1.27E+02 ± 2.12E+01	6.76E+03 ± 7.35E+02	8.1
H142Y	1.41E-01 ± 2.33E-03	1.17E+00 ± 4.31E-02	1.91E+01 ± 2.31E+00	6.16E+04 ± 8.90E+02	0.9
G147D	1.34E-01 ± 4.17E-03	1.12E+00 ± 3.48E-02	1.17E+02 ± 3.08E+00	9.52E+03 ± 1.98E+02	5.7
G154E	1.32E-01 ± 6.95E-03	1.11E+00 ± 5.79E-02	1.04E+02 ± 1.77E+01	1.06E+04 ± 1.34E+03	5.2
R175Q	1.43E-01 ± 6.63E-03	1.19E+00 ± 5.52E-02	3.59E+01 ± 5.07E+00	3.31E+04 ± 3.08E+03	1.6
V177L	1.27E-01 ± 3.53E-03	1.06E+00 ± 2.94E-02	8.46E+01 ± 4.15E+00	1.25E+04 ± 3.26E+02	4.4
E210K	1.31E-01 ± 4.41E-03	1.09E+00 ± 3.67E-02	2.88E+01 ± 1.21E+00	3.79E+04 ± 2.15E+03	1.5
Y221C	1.37E-01 ± 8.25E-03	1.14E+00 ± 6.87E-02	2.77E+01 ± 3.28E+00	4.11E+04 ± 2.55E+03	1.4
R298C	1.28E-01 ± 1.31E-02	1.06E+00 ± 1.09E-01	6.76E+01 ± 1.66E+01	1.58E+04 ± 2.09E+03	3.5
N330D	1.38E-01 ± 3.96E-03	1.15E+00 ± 3.31E-02	2.36E+01 ± 2.13E+00	4.86E+04 ± 3.37E+03	1.1
N338S	1.39E-01 ± 1.27E-02	1.16E+00 ± 1.06E-01	1.75E+01 ± 7.44E+00	6.65E+04 ± 3.05E+03	0.8
K353T	1.34E-01 ± 6.11E-04	1.11E+00 ± 5.09E-03	2.95E+01 ± 9.30E-01	3.78E+04 ± 1.15E+03	1.5
E362V	1.23E-01 ± 6.93E-03	1.03E+00 ± 5.77E-02	1.17E+02 ± 1.83E+01	8.76E+03 ± 8.04E+02	6.2
L383F	1.34E-01 ± 1.51E-02	1.12E+00 ± 1.24E-01	1.58E+01 ± 6.11E+00	7.07E+04 ± 2.07E+03	0.8
R470C	1.07E-01 ± 5.31E-03	8.95E-01 ± 4.32E-02	4.81E+01 ± 5.32E+00	1.86E+04 ± 6.08E+02	3.0

[a]Catalytic efficiency is k_{cat}/K_m. Variants that show lower catalytic activity (3-fold or more) are highlighted in pale red.

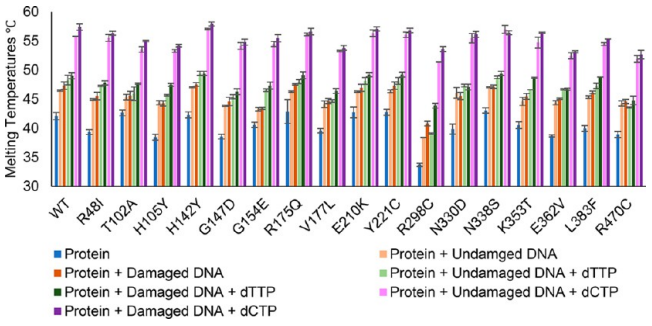


Figure 8. WT pol κ and variants showed an increase in the melting temperature in the presence of DNA substrates and the correct nucleotide dCTP. Stabilization of variants is observed in the presence of both undamaged or damaged N^2ffdG DNA with the correct nucleotide dCTP, and less stabilization is observed with the incorrect nucleotide dTTP.

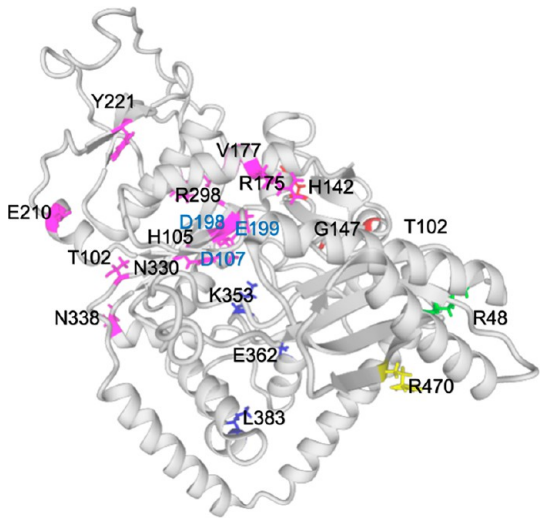


Figure 9. WT DNA polymerase kappa structure showing location of mutations and three previously identified catalytic residues. Residues in the N-clasp region are colored in green, palm region in magenta, fingers region in red, thumb region in blue, and PAD/little finger domain in yellow. Three previously identified catalytic resides are labeled in blue.

μ_4 values of catalytic residue D107 (Table 3) and K328 (Table S1), indicating distortion of the proton binding equilibria of the active site residues D107 and K328 in the pol κ R48I variant compared to WT pol κ.

The protein variant pol κ H105Y shows the lowest activity of all, with 5-fold and 8-fold lower catalytic efficiency for undamaged and damaged DNA, respectively, relative to WT pol κ; poor extension with both undamaged and damaged DNA; and a lower melting point (38.5 ± 0.5 °C; ΔT_m = 3.6 °C) compared with WT pol κ. H105 is one of the POOL-predicted functionally important residues with a POOL rank of 4 (Table S1) and is strongly coupled[37,91] to the catalytic residues D107 and E199. The H105Y substitution causes large distortions in μ4 values of D198, E199, and K328, which are known catalytically active residues that are responsible for nucleotidyl transfer in pol κ.[90] Residue H105 makes contact with E199, G312, and S324 within 4 Å. Y105 no longer makes

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

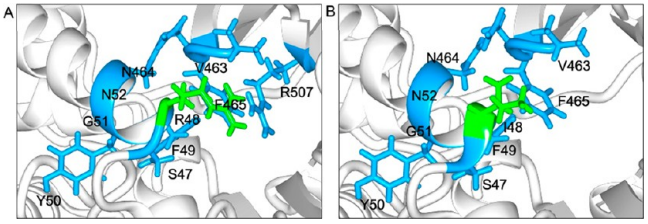


Figure 10. Interactions within the 4 Å region around (A) R48 and (B) I48 residues. R48 and I48 are colored in green, and residues that interact with residue 48 within 4 Å are colored in blue. While most of the contact residues remain the same in both R48 and I48, R48 makes additional contact with R507. Images were prepared using Pymol version 2.5.4.

contact with the catalytic E199 but instead interacts with D107 and K321 (Figure 11). The bioinformatic prediction for H105Y by SIFT was that this mutation is tolerated, but the other methods predicted it to be damaging/deleterious.

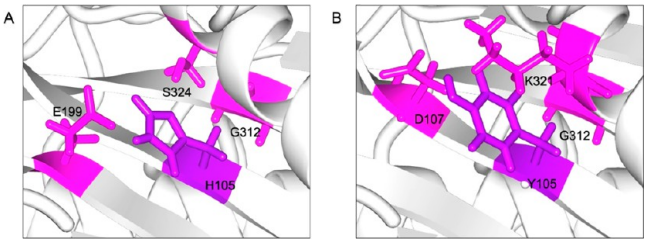


Figure 11. Interactions within the 4 Å region around (A) H105 and (B) Y105 residues. H105 and Y105 are colored in purple, and residues that interact with residue 105 within 4 Å are colored in pink. H105 makes interactions with E199, S324, and G312. While Y105 also interacts with G312, it acquires two new interactions with D107 and K321 instead of E199 and S324. Images were prepared using Pymol version 2.5.4.

Other variants that showed decreased catalytic activity relative to WT pol κ are G147D and G154E, which reside in the finger domain; V177L and R298C in the palm region; E362V in the thumb region; and R470C in the PAD/little finger domain. These variants also showed lower stability in the thermal shift assay, although in most cases, the difference was less than 3 °C. The variant G147D showed diminished insertion and extension activity relative to other variants. The DNA pol κ G154E variant was previously shown to have decreased DNA polymerizing activity by lowering the efficiency of nucleotide incorporation for dGTP opposite an abasic site compared to WT pol κ.[28] Even in the presence of the minor groove adduct N^2ffdG, the primer extension activity of pol κ G154E is low. The pol κ R298C variant showed the lowest stability of all variants (8.4 °C lower than WT), with increases in T_m in the presence of DNA and correct dCTP to within 3.8−4.4 °C of WT pol κ with the same substrates. The decrease in activity of the pol κ R298C variant could be due to this lower stability. Residue R298 forms hydrogen bonds with V294, I301, E302, N330, and Q332 within 4 Å. C298 would no longer form these interactions with N330 and Q332 (Figure 12). DNA pol κ residues 320−332 are important for activity, as they are involved in inducing conformation changes in pol κ to align active site residues with the incoming dNTP.[62] A similar pattern of lower activity and stability was noted with the R298H variant previously.[27,92] In spite of the difference between cysteine and histidine (R298H and R298C), where

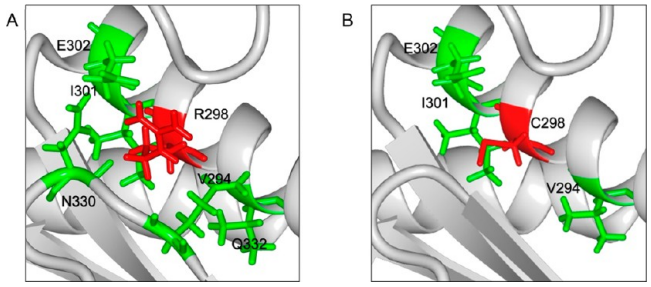


Figure 12. Interactions within the 4 Å region around (A) R298 and (B) C298 residues. R298 and C298 are colored in red, and residues that interact with residue 298 within 4 Å are colored in green. Both R298 and C298 interact with V294, I301, and E302. R298 makes two additional interactions with N330 and Q332; those interactions are not observed in pol κ C298. Images were prepared using Pymol version 2.5.4.

cysteine has a thiol group that can form a negatively charged thiolate, while histidine has an imidazole group that can form a positively charged side chain, similar negative effects on activity and stability are observed for pol κ variants R298H and R298C.

The thumb residue E362 is in a region that is strongly protected from deuterium exchange in the presence of the correct dNTP.[62] The pol κ E362V variant showed lower stability with a melting temperature of 38.7 ± 0.2 °C. It also showed a somewhat lower catalytic efficiency. DNA pol κ E362V had a low primer extension activity regardless of the DNA construct tested. On the other hand, SNP variant pol κ R470C showed low primer extension activity with both primer13 and primer14, with more robust activity in the presence of damaged N^2ffdG DNA, albeit still reduced relative to WT pol κ. These protein variants showed lower activity compared with WT pol κ, showed lower stability, and did not insert a single base in the presence of mismatch pairs (dT:dG or dT:N^2ffdG). THEMATICS-computed μ4 values for at least one of the active site residues of these less active variants showed a >50% difference relative to those of WT pol κ for the previously identified catalytically active site residues D107, D198, and/or E199. These changes in the computed protonation equilibria suggest impairment of catalytic function, and indeed, this is observed. Thus, mutations that distort significantly the protonation behavior of catalytically active residues are associated with changes in protein function. This suggests that THEMATICS data can assist in predicting whether a SNP causes alteration in protein function (Table S1).

POOL and THEMATICS correctly predicted known catalytically active residues and other residues of DNA pol κ that are known to be involved in the catalytic mechanism. This method has been previously used to identify the effects of noncatalytic distal residues on the catalytic activity of *E. coli* Y-family DNA polymerase DinB.[38] In the case of DinB, mutation of the POOL-predicted distal residues had a specific effect on the extension step of translesion synthesis. For the pol κ variants tested here, a specific defect in extension was not observed, even for mutations of highly POOL-ranked residues, such as H105, Y221, and K353; instead, pol κ SNP variants with effects on activity showed decreased activity generally.

Five of the eight pol κ SNP variants that showed lower activity were predicted by SIFT to affect protein function, and the remaining three variants with lower activity (pol κ H105Y,

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

G154E, and E362V) were classified as tolerated by SIFT. Notably, pol κ H105Y and E362V variants have very low primer extension activity and are among the lower stability variants. PolyPhen-2 correctly classifies all variants with decreased activity as harboring possibly damaging mutations but also classifies five active and relatively stable variants as possibly damaging. FATHMM also correctly classifies all variants with decreased activity as pathogenic but lists only one variant (active variant pol κ N338S) as neutral. It is important to note that variants may be harmful for reasons other than impaired catalysis, such as altered stability or perturbed protein−protein interactions.[93] In this work, we focused on the catalytic activity and protein stability. We chose the majority of predictions from the three bioinformatics tools as the basis of our overall prediction, which resulted in a more accurate prediction overall but with variants still incorrectly classified: five pol κ variants E210K, Y221C, N330D, K353T, and L383F were predicted to have impaired function with this criterion but were active. Applying the majority rule criterion correctly classified four of the active pol κ variants. SIFT, PolyPhen-2, and FATHMM use conserved regions, probability methods, and mathematical models,[82,94−96] and these analyses are based on protein amino acid sequence only. Considering the 3D structure of a protein reveals the position of the catalytically active residues, surface amino acids, hydrophobicity, H-bonds, and solvation energy, which can affect the protein function. Adding THEMATICS/POOL calculations of μ4 values of the active site residues for each mutation greatly increased the quality of the predictions. All variants with at least one of three active site residues (D107, D198, and E199) having a μ4 value change of at least 50% relative to WT pol κ were less active variants. The variants with activity similar to WT all had much smaller changes of <25% in the μ4 values of the active site residues D107, D198, and E199. Coulther[37] has reported evidence that, for highly evolved enzymes, the degree of coupling of the protonation equilibria of the catalytic residues has been optimized, such that any significant deviation, as measured by large changes in the μ4 values of the active residues, results in lower activity. This has been demonstrated here to be true for human Pol κ. Therefore, combining sequence-based predictions with structure-based predictions that incorporate perturbations in the protonation equilibria improves the accuracy of the predictions of altered biochemical activity of SNP variants.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemrestox.3c00233.

Quantitation of primer extension experiments; steady-state kinetics data and fits; melting temperature data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Penny J. Beuning − *Department of Chemistry and Chemical Biology and Department of Bioengineering, Northeastern University, Boston, Massachusetts 02115, United States;* orcid.org/0000-0002-7770-022X; Email: beuning@neu.edu

Mary Jo Ondrechen − *Department of Chemistry and Chemical Biology and Department of Bioengineering, Northeastern University, Boston, Massachusetts 02115, United States;* orcid.org/0000-0003-2456-4313; Email: mjo@neu.edu

Authors

Lakindu S. Pathira Kankanamge − *Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, United States;* orcid.org/0000-0002-2836-9489

Alexandra Mora − *Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrestox.3c00233

Author Contributions

CRediT: **Lakindu S Pathira Kankanamge** conceptualization, formal analysis, investigation, methodology, writing-original draft, writing-review & editing; **Alexandra Mora** formal analysis, investigation, writing-review & editing; **Mary Jo Ondrechen** conceptualization, funding acquisition, resources, supervision, writing-review & editing; **Penny J. Beuning** conceptualization, funding acquisition, supervision, writing-original draft, writing-review & editing.

Funding

This research is supported by the National Science Foundation under grant numbers MCB-1517290, CHE-1905214, and MCB-2147498.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

POOL, Partial Order Optimum Likelihood; THEMATICS, Theoretical Microscopic Titration Curve Shapes; SNP, Single nucleotide polymorphism; SIFT, Sorting Intolerant from Tolerant; PolyPhen-2, Polymorphism Phenotyping v2; FATHMM, Functional Analysis Through Hidden Markov Models; N^2ffdG, N^2-furfuryl-dG

■ REFERENCES

(1) Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709−715.

(2) Yang, W. An overview of Y-Family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* **2014**, *53*, 2793−2803.

(3) Livneh, Z.; Z, O.; Shachar, S. Multiple two-polymerase mechanisms in mammalian translesion DNA synthesis. *Cell Cycle* **2010**, *9*, 729−735.

(4) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. Crystal structure of a Y-Family DNA polymerase in action. *Cell* **2001**, *107*, 91−102.

(5) Pata, J. D. Structural diversity of the Y-family DNA polymerases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2010**, *1804*, 1124−1135.

(6) Yang, W.; Woodgate, R. What a difference a decade makes: Insights into translesion DNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15591−15598.

(7) Chilkova, O.; Stenlund, P.; Isoz, I.; Stith, C. M.; Grabowski, P.; Lundstrom, E. B.; Burgers, P. M.; Johansson, E. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res.* **2007**, *35*, 6588−6597.

(8) Sale, J. E.; Lehmann, A. R.; Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 141−152.
(9) Ohmori, H.; Friedberg, E. C.; Fuchs, R. P. P.; Goodman, M. F.; Hanaoka, F.; Hinkle, D.; Kunkel, T. A.; Lawrence, C. W.; Livneh, Z.; Nohmi, T.; Prakash, L.; Prakash, S.; Todo, T.; Walker, G. C.; Wang, Z.; Woodgate, R. The Y-family of DNA polymerases. *Mol. Cell* **2001**, *8*, 7−8.
(10) Bauer, J.; Xing, G.; Yagi, H.; Sayer, J. M.; Jerina, D. M.; Ling, H. A structural gap in Dpo4 supports mutagenic bypass of a major benzo[a]pyrene dG adduct in DNA through template misalignment. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14905−14910.
(11) Wu, Y.; Wilson, R. C.; Pata, J. D. The Y-family DNA polymerase Dpo4 uses a template slippage mechanism to create single-base deletions. *J. Bacteriol.* **2011**, *193*, 2630−2636.
(12) Uljon, S. N.; Johnson, R. E.; Edwards, T. A.; Prakash, S.; Prakash, L.; Aggarwal, A. K. Crystal structure of the catalytic core of human DNA polymerase kappa. *Structure* **2004**, *12*, 1395−1404.
(13) Stern, H. R.; Sefcikova, J.; Chaparro, V. E.; Beuning, P. J. Mammalian DNA polymerase kappa activity and specificity. *Molecules* **2019**, *24*, 2805.
(14) Rechkoblit, O.; Zhang, Y.; Guo, D.; Wang, Z.; Amin, S.; Krzeminsky, J.; Louneva, N.; Geacintov, N. E. trans-Lesion synthesis past bulky benzo[a]pyrene diol epoxide N^2-dG and N^6-dA lesions catalyzed by DNA bypass polymerases. *J. Biol. Chem.* **2002**, *277*, 30488−30494.
(15) Choi, J. Y.; Angel, K. C.; Guengerich, F. P. Translesion synthesis across bulky N^2-alkyl guanine DNA adducts by human DNA polymerase κ. *J. Biol. Chem.* **2006**, *281*, 21062−21072.
(16) Jarosz, D. F.; Godoy, V. G.; Delaney, J. C.; Essigmann, J. M.; Walker, G. C. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature* **2006**, *439*, 225−228.
(17) Yuan, B.; Cao, H.; Jiang, Y.; Hong, H.; Wang, Y. Efficient and accurate bypass of N^2-(1-carboxyethyl)-2′-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8679−8684.
(18) Zhang, Y. Activities of human DNA polymerase κ in response to the major benzo[a]pyrene DNA adduct: error-free lesion bypass and extension synthesis from opposite the lesion. *DNA Repair* **2002**, *1*, 559−569.
(19) Choi, J.-Y.; Guengerich, F. P. Kinetic analysis of translesion synthesis opposite bulky N^2- and O^6-alkylguanine DNA adducts by human DNA polymerase REV1. *J. Biol. Chem.* **2008**, *283*, 23645−23655.
(20) Walsh, J. M.; Ippoliti, P. J.; Ronayne, E. A.; Rozners, E.; Beuning, P. J. Discrimination against major groove adducts by Y-family polymerases of the DinB subfamily. *DNA Repair* **2013**, *12*, 713−722.
(21) Miller, R.; Phillips, M.; Jo, I.; Donaldson, M.; Studebaker, J.; Addleman, N.; Alfisi, S.; Ankener, W.; Bhatti, H.; Callahan, C. High-density single-nucleotide polymorphism maps of the human genome. *Genomics* **2005**, *86*, 117−126.
(22) Shastry, B. S. SNPs in disease gene mapping, medicinal drug development and evolution. *J. Hum. Genet.* **2007**, *52*, 871−880.
(23) Cooper, D. The human gene mutation database. *Nucleic Acids Res.* **1998**, *26*, 285−287.
(24) Lange, S. S.; Takata, K.-I.; Wood, R. D. DNA polymerases and cancer. *Nat. Rev. Cancer* **2011**, *11*, 96−110.
(25) Dai, Z.-J.; Liu, X.-H.; Ma, Y.-F.; Kang, H.-F.; Jin, T.-B.; Dai, Z.-M.; Guan, H.-T.; Wang, M.; Liu, K.; Dai, C.; Yang, X.-W.; Wang, X.-J. Association between single nucleotide polymorphisms in DNA polymerase kappa gene and breast cancer risk in chinese han population: a STROBE-compliant observational study. *Medicine* **2016**, *95*, No. e2466.
(26) Shao, M.; Jin, B.; Niu, Y.; Ye, J.; Lu, D.; Han, B. Association of POLK polymorphisms with platinum-based chemotherapy response and severe toxicity in non-small cell lung cancer patients. *Cell Biochem. Biophys.* **2014**, *70*, 1227−1237.
(27) Antczak, N. M.; Walker, A. R.; Stern, H. R.; Leddin, E. M.; Palad, C.; Coulther, T. A.; Swett, R. J.; Cisneros, G. A.; Beuning, P. J. Characterization of nine cancer-associated variants in human DNA polymerase κ. *Chem. Res. Toxicol.* **2018**, *31*, 697−711.
(28) Yadav, S.; Mukhopadhyay, S.; Anbalagan, M.; Makridakis, N. Somatic mutations in catalytic core of *POLK* reported in prostate cancer alter translesion DNA synthesis. *Hum. Mutat.* **2015**, *36*, 873−880.
(29) Tate, J. G.; Bamford, S.; Jubb, H. C.; Sondka, Z.; Beare, D. M.; Bindal, N.; Boutselakis, H.; Cole, C. G.; Creatore, C.; Dawson, E.; Fish, P.; Harsha, B.; Hathaway, C.; Jupe, S. C.; Kok, C. Y.; Noble, K.; Ponting, L.; Ramshaw, C. C.; Rye, C. E.; Speedy, H. E.; Stefancsik, R.; Thompson, S. L.; Wang, S.; Ward, S.; Campbell, P. J.; Forbes, S. A. COSMIC: the catalogue of somatic mutations In cancer. *Nucleic Acids Res.* **2019**, *47*, D941−D947.
(30) Flicek, P.; Amode, M. R.; Barrell, D.; Beal, K.; Brent, S.; Chen, Y.; Clapham, P.; Coates, G.; Fairley, S.; Fitzgerald, S.; Gordon, L.; Hendrix, M.; Hourlier, T.; Johnson, N.; Kähäri, A.; Keefe, D.; Keenan, S.; Kinsella, R.; Kokocinski, F.; Kulesha, E.; Larsson, P.; Longden, I.; McLaren, W.; Overduin, B.; Pritchard, B.; Riat, H. S.; Rios, D.; Ritchie, G. R. S.; Ruffier, M.; Schuster, M.; Sobral, D.; Spudich, G.; Tang, Y. A.; Trevanion, S.; Vandrovcova, J.; Vilella, A. J.; White, S.; Wilder, S. P.; Zadissa, A.; Zamora, J.; Aken, B. L.; Birney, E.; Cunningham, F.; Dunham, I.; Durbin, R.; Fernández-Suarez, X. M.; Herrero, J.; Hubbard, T. J. P.; Parker, A.; Proctor, G.; Vogel, J.; Searle, S. M. J. Ensembl 2011. *Nucleic Acids Res.* **2011**, *39*, D800−D806.
(31) Tong, W.; Wei, Y.; Murga, L. F.; Ondrechen, M. J.; Williams, R. J. Partial order optimum likelihood (POOL), Maximum likelihood prediction of protein active site residues using 3D structure and sequence properties. *PLoS Comp. Biol.* **2009**, *5*, No. e1000266.
(32) Ko, J.; Murga, L. F.; André, P.; Yang, H.; Ondrechen, M. J.; Williams, R. J.; Agunwamba, A.; Budil, D. E. Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. *Proteins: Struct. Funct. Bioinform.* **2005**, *59*, 183−195.
(33) Wei, Y.; Ko, J.; Murga, L. F.; Ondrechen, M. J. Selective prediction of interaction sites in protein structures with THEMATICS. *BMC Bioinformatics* **2007**, *8*, 119.
(34) Ondrechen, M. J.; Clifton, J. G.; Ringe, D. THEMATICS: A simple computational predictor of enzyme function from structure. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12473−12478.
(35) Somarowthu, S.; Ondrechen, M. J. POOL server: machine learning application for functional site prediction in proteins. *Comput. Appl. Biosci.* **2012**, *28*, 2078−2079.
(36) Somarowthu, S.; Yang, H.; Hildebrand, D. G. C.; Ondrechen, M. J. High-performance prediction of functional residues in proteins with machine learning and computed input features. *Biopolymers* **2011**, *95*, 390−400.
(37) Coulther, T. A.; Ko, J.; Ondrechen, M. J. Amino acid interactions that facilitate enzyme catalysis. *J. Chem. Phys.* **2021**, *154*, 195101.
(38) Walsh, J. M.; Parasuram, R.; Rajput, P. R.; Rozners, E.; Ondrechen, M. J.; Beuning, P. J. Effects of non-catalytic, distal amino acid residues on activity of *E. coli* DinB (DNA polymerase IV). *Environ. Mol. Mutag.* **2012**, *53*, 766−776.
(39) Parasuram, R.; Coulther, T. A.; Hollander, J. M.; Keston-Smith, E.; Ondrechen, M. J.; Beuning, P. J. Prediction of active site and distal residues in *E. coli* DNA polymerase III alpha polymerase activity. *Biochemistry* **2018**, *57*, 1063−1072.
(40) Brodkin, H. R.; Delateur, N. A.; Somarowthu, S.; Mills, C. L.; Novak, W. R.; Beuning, P. J.; Ringe, D.; Ondrechen, M. J. Prediction of distal residue participation in enzyme catalysis. *Protein Sci.* **2015**, *24*, 762−778.
(41) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144−1149.
(42) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L.R.; McCammon, J.A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian Dynamics

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803

program: Biomolecular simulations. *Comput. Phys. Commun.* **1995**, *91*, 57−95.
(43) Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; Gilson, M. K.; McCammon, J. A. Computing ionization states of proteins with a detailed charge model. *J. Comput. Chem.* **1996**, *17*, 1633−1644.
(44) Tong, W.; Williams, R. J.; Wei, Y.; Murga, L. F.; Ko, J.; Ondrechen, M. J. Enhanced performance in prediction of protein active sites with THEMATICS and support vector machines. *Protein Sci.* **2008**, *17*, 333−341.
(45) Sim, N. L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P. C. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452−W457.
(46) Adzhubei, I.; Jordan, D. M.; Sunyaev, S. R., Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc Hum Genet* **2013**, *76*, Chapter 7, Unit 7. DOI: 10.1002/0471142905.hg0720s76.
(47) Shihab, H. A.; Gough, J.; Cooper, D. N.; Stenson, P. D.; Barker, G. L. A.; Edwards, K. J.; Day, I. N. M.; Gaunt, T. R. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using Hidden Markov Models. *Hum. Mutat.* **2013**, *34*, 57−65.
(48) Jha, V.; Ling, H. 2.0 Å resolution crystal structure of human polκ reveals a new catalytic function of N-clasp in DNA replication. *Sci. Rep.* **2018**, *8*, 15125−15129.
(49) Land, H.; Humble, M. S. YASARA: A tool to obtain structural guidance in biocatalytic investigations. *Methods Mol. Biol.* **2018**, *1685*, 43−67.
(50) Krieger, E.; Koraimann, G.; Vriend, G. Increasing the precision of comparative models with YASARA NOVA-a self-parameterizing force field. *Proteins: Struct. Funct. Bioinform.* **2002**, *47*, 393−402.
(51) Benkert, P.; Tosatto, S. C. E.; Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Struct. Funct. Bioinform.* **2008**, *71*, 261−277.
(52) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(53) Melo, F.; Devos, D.; Depiereux, E.; Feytmans, E. ANOLEA: A www server to assess protein structures. *Proc Int Conf Intell Syst Mol Biol.* **1997**, *5*, 187−190.
(54) Salam, N. K.; Adzhigirey, M.; Sherman, W.; Pearlman, D. A. Structure-based approach to the prediction of disulfide bonds in proteins. *Protein Eng. Des. Sel.* **2014**, *27*, 365−374.
(55) Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins* **2004**, *57*, 678−683.
(56) Laimer, J.; Hofer, H.; Fritz, M.; Wegenkittl, S.; Lackner, P. MAESTRO - Multi agent stability prediction upon point mutations. *BMC Bioinformatics* **2015**, *16*, 116.
(57) Irimia, A.; Eoff, R. L.; Guengerich, F. P.; Egli, M. Structural and functional elucidation of the mechanism promoting error-prone synthesis by human DNA polymerase κ opposite the 7,8-dihydro-8-oxo-2′-deoxyguanosine adduct. *J. Biol. Chem.* **2009**, *284*, 22467−22480.
(58) Decorte, B. L.; Tsarouhtsis, D.; Kuchimanchi, S.; Cooper, M. D.; Horton, P.; Harris, C. M.; Harris, T. M. Improved strategies for postoligomerization synthesis of oligodeoxynucleotides bearing structurally defined adducts at the N^2 position of deoxyguanosine. *Chem. Res. Toxicol.* **1996**, *9*, 630−637.
(59) Sambrook, J.; Russell, D. W. *Molecular cloning: A laboratory manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y, 2001.
(60) Sheriff, A.; Motea, E.; Lee, I.; Berdis, A. J. The mechanism and dynamics of translesion DNA synthesis catalyzed by the *Escherichia coli* klenow fragment. *Biochemistry* **2008**, *47*, 8527−8537.
(61) Nevin, P.; Engen, J. R.; Beuning, P. J. Steric gate residues of Y-family DNA polymerases DinB and pol kappa are crucial for dNTP-induced conformational change. *DNA Repair* **2015**, *29*, 65−73.
(62) Nevin, P.; Lu, X.; Zhang, K.; Engen, J. R.; Beuning, P. J. Noncognate DNA damage prevents the formation of the active conformation of the Y-family DNA polymerases DinB and DNA polymerase κ. *FEBS J.* **2015**, *282*, 2646−2660.
(63) Ericsson, U. B.; Hallberg, B. M.; DeTitta, G. T.; Dekker, N.; Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **2006**, *357*, 289−298.
(64) Liu, Z.; Yang, C.; Li, X.; Luo, W.; Roy, B.; Xiong, T.; Zhang, X.; Yang, H.; Wang, J.; Ye, Z.; Chen, Y.; Song, J.; Ma, S.; Zhou, Y.; Yang, M.; Fang, X.; Du, J. The landscape of somatic mutation in sporadic Chinese colorectal cancer. *Oncotarget* **2018**, *9*, 27412−27422.
(65) Mouradov, D.; Sloggett, C.; Jorissen, R. N.; Love, C. G.; Li, S.; Burgess, A. W.; Arango, D.; Strausberg, R. L.; Buchanan, D.; Wormald, S.; O'Connor, L.; Wilding, J. L.; Bicknell, D.; Tomlinson, I. P. M.; Bodmer, W. F.; Mariadason, J. M.; Sieber, O. M. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res.* **2014**, *74*, 3238−3247.
(66) Fadlullah, M. Z. H.; Chiang, I. K.-N.; Dionne, K. R.; Yee, P. S.; Gan, C. P.; Sam, K. K.; Tiong, K. H.; Ng, A. K. W.; Martin, D.; Lim, K. P.; Kallarakkal, T. G.; Mustafa, W. M. W.; Lau, S. H.; Abraham, M. T.; Zain, R. B.; Rahman, Z. A. A.; Molinolo, A.; Patel, V.; Gutkind, J. S.; Tan, A. C.; et al. Genetically-defined novel oral squamous cell carcinoma cell lines for the development of molecular therapies. *Oncotarget* **2016**, *7*, 27802−27818.
(67) Bala, P.; Singh, A. K.; Kavadipula, P.; Kotapalli, V.; Sabarinathan, R.; Bashyam, M. D. Exome sequencing identifies ARID2 as a novel tumor suppressor in early-onset sporadic rectal cancer. *Oncogene* **2021**, *40*, 863−874.
(68) Giannakis, M.; Mu, X. J.; Shukla, S. A.; Qian, Z. R.; Cohen, O.; Nishihara, R.; Bahl, S.; Cao, Y.; Amin-Mansour, A.; Yamauchi, M.; Sukawa, Y.; Stewart, C.; Rosenberg, M.; Mima, K.; Inamura, K.; Nosho, K.; Nowak, J. A.; Lawrence, M. S.; Giovannucci, E. L.; Chan, A. T.; et al. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Reports* **2016**, *15*, 857−865.
(69) Grasso, C. S.; Wu, Y.-M.; Robinson, D. R.; Cao, X.; Dhanasekaran, S. M.; Khan, A. P.; Quist, M. J.; Jing, X.; Lonigro, R. J.; Brenner, J. C.; Asangani, I. A.; Ateeq, B.; Chun, S. Y.; Siddiqui, J.; Sam, L.; Anstett, M.; Mehra, R.; Prensner, J. R.; Palanisamy, N.; Ryslik, G. A.; et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **2012**, *487*, 239−243.
(70) Gingras, M.-C.; Covington, K. R.; Chang, D. K.; Donehower, L. A.; Gill, A. J.; Ittmann, M. M.; Creighton, C. J.; Johns, A. L.; Shinbrot, E.; Dewal, N.; Fisher, W. E.; Pilarsky, C.; Grützmann, R.; Overman, M. J.; Jamieson, N. B.; Van Buren, G.; Drummond, J.; Walker, K.; Hampton, O. A.; Xi, L.; et al. Ampullary cancers harbor ELF3 tumor suppressor gene mutations and exhibit frequent WNT dysregulation. *Cell Reports* **2016**, *14*, 907−919.
(71) McMillan, E. A.; Ryu, M.-J.; Diep, C. H.; Mendiratta, S.; Clemenceau, J. R.; Vaden, R. M.; Kim, J.-H.; Motoyaji, T.; Covington, K. R.; Peyton, M.; Huffman, K.; Wu, X.; Girard, L.; Sung, Y.; Chen, P.-H.; Mallipeddi, P. L.; Lee, J. Y.; Hanson, J.; Voruganti, S.; Yu, Y.; et al. Chemistry-first approach for nomination of personalized treatment in lung cancer. *Cell* **2018**, *173*, 864−878.e29.
(72) Sato, Y.; Yoshizato, T.; Shiraishi, Y.; Maekawa, S.; Okuno, Y.; Kamura, T.; Shimamura, T.; Sato-Otsubo, A.; Nagae, G.; Suzuki, H.; Nagata, Y.; Yoshida, K.; Kon, A.; Suzuki, Y.; Chiba, K.; Tanaka, H.; Niida, A.; Fujimoto, A.; Tsunoda, T.; Morikawa, T.; Maeda, D.; Kume, H.; Sugano, S.; Fukayama, M.; Aburatani, H.; Sanada, M.; Miyano, S.; Homma, Y.; Ogawa, S. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **2013**, *45*, 860−867.
(73) The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **2012**, *487*, 330−337.
(74) Pickering, C. R.; Zhou, J. H.; Lee, J. J.; Drummond, J. A.; Peng, S. A.; Saade, R. E.; Tsai, K. Y.; Curry, J. L.; Tetzlaff, M. T.; Lai, S. Y.; Yu, J.; Muzny, D. M.; Doddapaneni, H.; Shinbrot, E.; Covington, K. R.; Zhang, J.; Seth, S.; Caulin, C.; Clayman, G. L.; El-Naggar, A. K.; Gibbs, R. A.; Weber, R. S.; Myers, J. N.; Wheeler, D. A.; Frederick, M. J. Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin. Cancer. Res.* **2014**, *20*, 6582−6592.

(75) Bonilla, X.; Parmentier, L.; King, B.; Bezrukov, F.; Kaya, G.; Zoete, V.; Seplyarskiy, V. B.; Sharpe, H. J.; Mckee, T.; Letourneau, A.; Ribaux, P. G.; Popadin, K.; Basset-Seguin, N.; Chaabene, R. B.; Santoni, F. A.; Andrianova, M. A.; Guipponi, M.; Garieri, M.; Verdan, C.; Grosdemange, K.; Sumara, O.; Eilers, M.; Aifantis, I.; Michielin, O.; De Sauvage, F. J.; Antonarakis, S. E.; Nikolaev, S. I. Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma. *Nat. Genet.* **2016**, *48*, 398−406.
(76) Hayward, N. K.; Wilmott, J. S.; Waddell, N.; Johansson, P. A.; Field, M. A.; Nones, K.; Patch, A.-M.; Kakavand, H.; Alexandrov, L. B.; Burke, H.; Jakrot, V.; Kazakoff, S.; Holmes, O.; Leonard, C.; Sabarinathan, R.; Mularoni, L.; Wood, S.; Xu, Q.; Waddell, N.; Tembe, V.; Pupo, G. M.; De Paoli-Iseppi, R.; Vilain, R. E.; Shang, P.; Lau, L. M. S.; Dagg, R. A.; Schramm, S.-J.; Pritchard, A.; Dutton-Regester, K.; Newell, F.; Fitzgerald, A.; Shang, C. A.; Grimmond, S. M.; Pickett, H. A.; Yang, J. Y.; Stretch, J. R.; Behren, A.; Kefford, R. F.; Hersey, P.; Long, G. V.; Cebon, J.; Shackleton, M.; Spillane, A. J.; Saw, R. P. M.; López-Bigas, N.; Pearson, J. V.; Thompson, J. F.; Scolyer, R. A.; Mann, G. J. Whole-genome landscapes of major melanoma subtypes. *Nature* **2017**, *545*, 175−180.
(77) Van Allen, E. M.; Wagle, N.; Sucker, A.; Treacy, D. J.; Johannessen, C. M.; Goetz, E. M.; Place, C. S.; Taylor-Weiner, A.; Whittaker, S.; Kryukov, G. V.; Hodis, E.; Rosenberg, M.; Mckenna, A.; Cibulskis, K.; Farlow, D.; Zimmer, L.; Hillen, U.; Gutzmer, R.; Goldinger, S. M.; Ugurel, S.; Gogas, H. J.; Egberts, F.; Berking, C.; Trefzer, U.; Loquai, C.; Weide, B.; Hassel, J. C.; Gabriel, S. B.; Carter, S. L.; Getz, G.; Garraway, L. A.; Schadendorf, D. The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discovery* **2014**, *4*, 94−109.
(78) Flanagan, S. E.; Patch, A.-M.; Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic Testing and Molecular Biomarkers.* **2010**, *14*, 533−537.
(79) Hassan, M. S.; Shaalan, A. A.; Dessouky, M. I.; Abdelnaiem, A. E.; ElHefnawi, M. Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. *Genomics* **2019**, *111*, 869−882.
(80) Ng, P. C. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812−3814.
(81) Sunyaev, S. Prediction of deleterious human alleles. *Hum. Mol. Genet.* **2001**, *10*, 591−597.
(82) Shihab, H. A.; Rogers, M. F.; Gough, J.; Mort, M.; Cooper, D. N.; Day, I. N. M.; Gaunt, T. R.; Campbell, C. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **2015**, *31*, 1536−1543.
(83) Wolfle, W. T.; Washington, M. T.; Prakash, L.; Prakash, S. Human DNA polymerase κ uses template-primer misalignment as a novel means for extending mispaired termini and for generating single-base deletions. *Genes Dev.* **2003**, *17*, 2191−2199.
(84) Washington, M. T.; Johnson, R. E.; Prakash, L.; Prakash, S. Human *DINB1* encoded DNA polymerase κ is a promiscuous extender of mispaired primer termini. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1910−1914.
(85) Haracska, L.; Prakash, L.; Prakash, S. Role of human DNA polymerase κ as an extender in translesion synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16000−16005.
(86) Antczak, N. M.; Packer, M. R.; Lu, X.; Zhang, K.; Beuning, P. J. Human Y-family DNA polymerase κ is more tolerant to changes in its active site loop than its ortholog *Escherichia coli* DinB. *Chem. Res. Toxicol.* **2017**, *30*, 2002−2012.
(87) Ogi, T.; Limsirichaikul, S.; Overmeer, R. M.; Volker, M.; Takenaka, K.; Cloney, R.; Nakazawa, Y.; Niimi, A.; Miki, Y.; Jaspers, N. G.; Mullenders, L. H. F.; Yamashita, S.; Fousteri, M. I.; Lehmann, A. R. Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol. Cell* **2010**, *37*, 714−727.
(88) Choi, J.-Y.; Chowdhury, G.; Zang, H.; Angel, K. C.; Vu, C. C.; Peterson, L. A.; Guengerich, F. P. Translesion synthesis across O^6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem.* **2006**, *281*, 38244−38256.
(89) Song, I.; Kim, E.-J.; Kim, I.-H.; Park, E.-M.; Lee, K. E.; Shin, J.-H.; Guengerich, F. P.; Choi, J.-Y. Biochemical characterization of eight genetic variants of human DNA polymerase κ involved in error-free bypass across bulky N^2-guanyl DNA adducts. *Chem. Res. Toxicol.* **2014**, *27*, 919−930.
(90) Lone, S.; Townson, S. A.; Uljon, S. N.; Johnson, R. E.; Brahma, A.; Nair, D. T.; Prakash, S.; Prakash, L.; Aggarwal, A. K. Human DNA polymerase κ encircles DNA: Implications for mismatch extension and lesion bypass. *Mol. Cell* **2007**, *25*, 601−614.
(91) Iyengar, S. M.; Barnsley, K. K.; Xu, R.; Prystupa, A.; Ondrechen, M. J. Electrostatic fingerprints of catalytically active amino acids in enzymes. *Protein Sci.* **2022**, *31*, e4291.
(92) Kim, J.-K.; Yeom, M.; Hong, J.-K.; Song, I.; Lee, Y.-S.; Guengerich, F. P.; Choi, J.-Y. Six germline genetic variations impair the translesion synthesis activity of human DNA polymerase κ. *Chem. Res. Toxicol.* **2016**, *29*, 1741−1754.
(93) Fragoza, R.; Das, J.; Wierbowski, S. D.; Liang, J.; Tran, T. N.; Liang, S.; Beltran, J. F.; Rivera-Erick, C. A.; Ye, K.; Wang, T.-Y.; Yao, L.; Mort, M.; Stenson, P. D.; Cooper, D. N.; Wei, X.; Keinan, A.; Schimenti, J. C.; Clark, A. G.; Yu, H. Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations. *Nat. Commun.* **2019**, *10*, 4141.
(94) Ng, P. C.; Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res.* **2001**, *11*, 863−874.
(95) Ng, P. C.; Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* **2002**, *12*, 436−446.
(96) Rogers, M. F.; Shihab, H. A.; Mort, M.; Cooper, D. N.; Gaunt, T. R.; Campbell, C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **2018**, *34*, 511−513.

https://doi.org/10.1021/acs.chemrestox.3c00233
Chem. Res. Toxicol. 2023, 36, 1789−1803