PNAS PLUS

Intrinsic mutagenic properties of 5-chlorocytosine: A mechanistic connection between chronic inflammation and cancer

Bogdan I. Fedeles[a,b,c], Bret D. Freudenthal[d], Emily Yau[b,c], Vipender Singh[a,b,c], Shiou-chi Chang[b,c], Deyu Li[a,b,c,1], James C. Delaney[a,b,c,2], Samuel H. Wilson[d], and John M. Essigmann[a,b,c,3]

[a]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Department Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and [d]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709

Edited by Dennis A. Carson, University of California at San Diego, La Jolla, CA, and approved July 9, 2015 (received for review April 20, 2015)

During chronic inflammation, neutrophil-secreted hypochlorous acid can damage nearby cells inducing the genomic accumulation of 5-chlorocytosine (5ClC), a known inflammation biomarker. Although 5ClC has been shown to promote epigenetic changes, it has been unknown heretofore if 5ClC directly perpetrates a mutagenic outcome within the cell. The present work shows that 5ClC is intrinsically mutagenic, both in vitro and, at a level of a single molecule per cell, in vivo. Using biochemical and genetic approaches, we have quantified the mutagenic and toxic properties of 5ClC, showing that this lesion caused C→T transitions at frequencies ranging from 3–9% depending on the polymerase traversing the lesion. X-ray crystallographic studies provided a molecular basis for the mutagenicity of 5ClC; a snapshot of human polymerase β replicating across a primed 5ClC-containing template uncovered 5ClC engaged in a nascent base pair with an incoming dATP analog. Accommodation of the chlorine substituent in the template major groove enabled a unique interaction between 5ClC and the incoming dATP, which would facilitate mutagenic lesion bypass. The type of mutation induced by 5ClC, the C→T transition, has been previously shown to occur in substantial amounts both in tissues under inflammatory stress and in the genomes of many inflammation-associated cancers. In fact, many sequence-specific mutational signatures uncovered in sequenced cancer genomes feature C→T mutations. Therefore, the mutagenic ability of 5ClC documented in the present study may constitute a direct functional link between chronic inflammation and the genetic changes that enable and promote malignant transformation.

5-chloro-deoxycytidine | hypochlorite | myeloperoxidase | inflammatory bowel disease | carcinogenesis

Chronic inflammation is an established risk factor for many human cancers (1–6). By generating reactive chemical entities (e.g., reactive oxygen, nitrogen, and halogen species) intended to neutralize invading pathogens, inflammatory responses may also inflict chemical collateral damage on nucleic acids and proteins in surrounding healthy tissues (7–10). When the immune response aberrantly persists over months and years, the chemical damage may exceed the repair capabilities of the exposed cells, leading to cell death or to genetic and epigenetic modifications that, in principle, can set the stage for malignant transformation (7–10).

One type of chemical damage associated with inflammation is chlorination of cellular macromolecules by hypochlorous acid (HOCl). This type of damage is the result of the infiltration and activation of neutrophil granulocytes (the most abundant leukocyte), which produce myeloperoxidase, an enzyme capable of synthesizing HOCl from hydrogen peroxide and chloride anions (Fig. 1) (11–14). The main biomarkers found in cells exposed to HOCl are chlorinated tyrosines (in proteins) and chlorinated nucleobases (in nucleic acids and nucleotide pools) (15, 16). Among the chlorinated nucleobases, 5-chlorocytosine (5ClC) is the most abundant (16–19) and recently has been recognized as a biomarker associated with chronic inflammation (20, 21). In one study, the levels of 5ClC in both DNA and RNA were elevated in the colon of $Rag2^{-/-}$ mice infected with *Helicobacter hepaticus*, a model of chronic inflammation in the gut (20). In a follow-up study, the same biomarkers were found at similarly high levels in colonic tissue samples from human patients with inflammatory bowel disease (Crohn's disease or ulcerative colitis) (21). Surprisingly, these studies did not find significantly increased levels of classical DNA biomarkers of inflammation and oxidative stress, such as 8-oxo-7,8-dihydroguanine (8oxoG) and other oxidative DNA lesions (20, 21). These findings, together with the observed persistence and accumulation of genomic 5ClC in chronically inflamed tissues suggested that 5ClC is not efficiently repaired and could have a functional role in disease initiation or progression.

So far, research on 5ClC has focused on the lesion's ability to interact with DNA methyl transferases (22–24) when 5ClC is part of a CpG island; this interaction was shown to lead to changes in DNA methylation patterns (25). However, there are no reports affirming the ability of 5ClC to induce genetic changes directly during the replication or repair of genomic DNA. Knowledge of the mutagenicity of 5ClC would be especially

Significance

Chronic inflammation is a significant risk factor for cancer and other human diseases. During chronic inflammation, cells exposed to neutrophil-derived hypochlorous acid accumulate in their genomes the DNA lesion 5-chlorocytosine (5ClC). Using a battery of chemical, structural, and genetic tools, the present study demonstrates that 5ClC is a mutagenic lesion, suggesting that genomic 5ClC accumulation could have very serious biological consequences. 5ClC induces C→T transitions, a type of mutation commonly observed in tissues under inflammatory stress as well as in the genomes of inflammation-driven cancers. Thus, the mutagenic properties of 5ClC represent an appealing molecular mechanism by which chronic inflammation induces the genetic changes that potentially enable and stimulate carcinogenesis.

Author contributions: B.I.F., B.D.F., J.C.D., S.H.W., and J.M.E. designed research; B.I.F., B.D.F., E.Y., V.S., S.-c.C., and D.L. performed research; B.I.F., B.D.F., and S.-c.C. contributed new reagents/analytic tools; B.I.F., B.D.F., and E.Y. analyzed data; and B.I.F., B.D.F., S.H.W., and J.M.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The coordinates of the X-ray structures reported in this paper have been deposited at the Research Collaboratory for Structural Bioinformatics Protein Data Bank [RCSB PDB ID codes 5BOM (binary complex), 5BOL (ternary complex with dGTP analog), and 5BPC (ternary complex with dATP analog)].

[1]Present address: Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, RI 02881.

[2]Present address: Visterra, Inc., Cambridge, MA 02139.

[3]To whom correspondence should be addressed. Email: jessig@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1507709112/-/DCSupplemental.

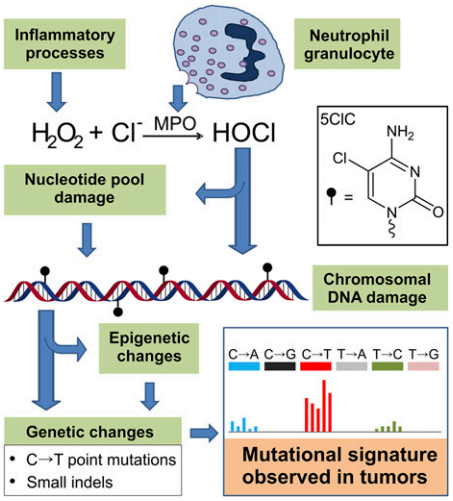

Fig. 1. A model of the 5ClC biomarker as a functional link between chronic inflammation and the genetic changes observed in tumors. During inflammation, activated neutrophil granulocytes produce the enzyme myeloperoxidase (MPO); other inflammatory processes produce hydrogen peroxide, among other reactive oxygen species. Using hydrogen peroxide and chloride, MPO catalyzes the formation of HOCl. Exposure of the surrounding cells to HOCl can generate 5ClC in their genomic DNA, a type of DNA damage occurring either by direct chlorination of cytosines or by the incorporation of damaged nucleotides from the pool. The present study showcases the mutagenic ability of 5ClC, which can directly induce C→T mutations. Additionally, 5ClC is known to induce epigenetic changes that could lead to the inactivation of tumor-suppressor genes such as *MLH1* by hypermethylation, in turn possibly leading to additional mutagenic outcomes such as small indels and microsatellite instability. Together, the biological effects of 5ClC generate a characteristic mutational signature (the relative distribution of all possible point mutations and indels over all possible sequence contexts), which could account for certain mutational signatures found in malignancies associated with chronic inflammation. The mutational signature shown is a cartoon representation of signature 6 from Alexandrov et al. (56), highlighting the preponderance of C→T transitions.

relevant for the cells surrounding the inflammation site, which have both an increased level of DNA repair (e.g., removing the aforementioned oxidative stress lesions that do not persist in inflamed tissues) and a higher propensity to replicate (to compensate for the inflammation-induced cell death). Both of these processes put the cells in the inflamed tissues at an increased risk for progressing toward cancer, particularly if the 5ClC they accumulate is a mutagenic lesion.

In the present work, we have used cellular, biochemical, and structural approaches to show that 5ClC is indeed mutagenic, inducing C→T transitions at a significant rate; when the levels of genomic 5ClC are taken into account, the mutagenesis of 5ClC translates into an anticipated 1–2 orders of magnitude increase in the cellular mutation rate in the inflamed tissues. The molecular basis for the mutagenic property of 5ClC was elucidated by crystallizing human polymerase β (pol β) as it incorporated dGTP or dATP analogs opposite the lesion. In conjunction with the previously known epigenetic effects of 5ClC, the mutagenic properties of 5ClC establish this biomarker as a promising functional link between chronic inflammation and the genetic and transcriptional changes observed in inflammation-associated tumors (Fig. 1).

Results

5ClC Is a Mutagenic Base That Is Easily Bypassed by Polymerases in Vivo. The mutagenic properties of 5ClC, as well as the lesion's ability to block replication, were evaluated by introducing the lesion site-specifically on a viral vector, replicating the vector in living cells, and analyzing the resulting viral progeny DNA using the established REAP (Restriction Endonuclease and Postlabeling) and CRAB (Competitive Replication of Adduct Bypass) assays (26–28). The 5ClC nucleoside was incorporated site-specifically into a 16mer oligonucleotide, which subsequently was ligated into an M13 single-stranded DNA vector (Fig. 2*A* and Fig. S1). Similarly, control vectors were constructed containing the negative controls cytosine (C) and 5-methylcytosine (m5C) and the positive control 3-methylcytosine (m3C). The m3C-containing vector was used primarily to validate the assays: in the absence of the Fe(II)/α-ketoglutarate-dependent dioxygenase AlkB, a direct reversal DNA repair enzyme, m3C is a strong block to replication and leads to a characteristic mutational pattern, whereas in wild-type cells m3C is repaired efficiently and completely, thus behaving like an unmodified cytosine (27, 29). Additionally, m3C was used to confirm the successful induction of the SOS response (see below).

Two strains of *Escherichia coli* AB1157 denoted $AlkB^+$ (wild type) and $AlkB^-$ (*AlkB* deficient) were used to replicate the constructed M13 vectors. Additionally, an SOS-induced version of the $AlkB^-$ strain (denoted $AlkB^-SOS^+$) was used. The induction of the SOS response, accomplished by briefly exposing the cells to UV radiation, increases the levels of cellular Y-class translesion polymerases such as Pol IV and Pol V. Following in vivo replication, the progeny phage DNA was extracted and analyzed to determine for each lesion (or controls) and cell condition the percentage bypass (a measure of how readily a lesion is bypassed during replication) and the percentage base at the lesion site, from which the type and amount of mutagenesis can be determined (26). A detailed graphical overview of the method is shown in Fig. S1.

The bypass experiments revealed that 5ClC is a mild block to replication; bypass relative to cytosine was ∼75% in both wild-type and $AlkB^-$ strains and increased to ∼100% in the SOS-induced cell line (Fig. 2*B* and Table S1). The negative control m5C showed a level of bypass indistinguishable from that of cytosine in all cell strains investigated. The behavior of the positive control m3C was consistent with previous studies (27, 29): in the $AlkB^-$ strain, m3C was a very strong block to replication (15% bypass relative to dC); the amount of bypass increased with SOS induction to 35%, whereas in the wild-type strain, because of its efficient repair by AlkB, m3C showed a bypass of ∼100% (Fig. 2*B* and Table S1).

The mutagenesis assay revealed that 5ClC is a mutagenic lesion, coding as a C ∼95% of the time and as a T ∼5% of the time, in all three cell strains investigated (Fig. 2*C* and Table S2). This result therefore suggests that 5ClC base-paired with guanine most of the time (∼95%), but it also base-paired with adenine (∼5%), an unexpected result. To put this value in perspective, 5ClC is approximately as mutagenic as 8oxoG, evaluated with a similar assay in a wild-type *E. coli* strain (i.e., a strain with normal levels of the 8oxoG repair enzymes MutM and MutY) (30, 31). The positive control m3C behaved as previously reported (27, 29). When fully repaired in the wild-type ($AlkB^+$) strain, m3C coded exclusively as C; however, in the $AlkB^-$ strains, m3C was highly mutagenic, coding primarily as a T or an A, with smaller amounts of C and G also observed. Additionally, SOS induction exacerbated the mutagenesis of m3C by increasing the percentage of non-C base detected at lesion site (Fig. 2*C* and Table S2). In contrast, the negative controls cytosine and m5C coded exclusively as C in all cell lines tested; the non-C levels shown in the inset graphs in Fig. 2*C* and in Table S2 were <0.5%, in good agreement with the expected sensitivity of the assay.

Establishing the Purity of 5ClC-Containing Oligonucleotides and Genomes. A series of validation experiments was performed to rule out any potential experimental artifacts (e.g., impurities in the starting materials) that could confound the in vivo experiment described above, which indicated that 5ClC was mutagenic. The 5ClC-containing oligonucleotide used to construct the M13 genome was thoroughly scrutinized using mass spectrometry and

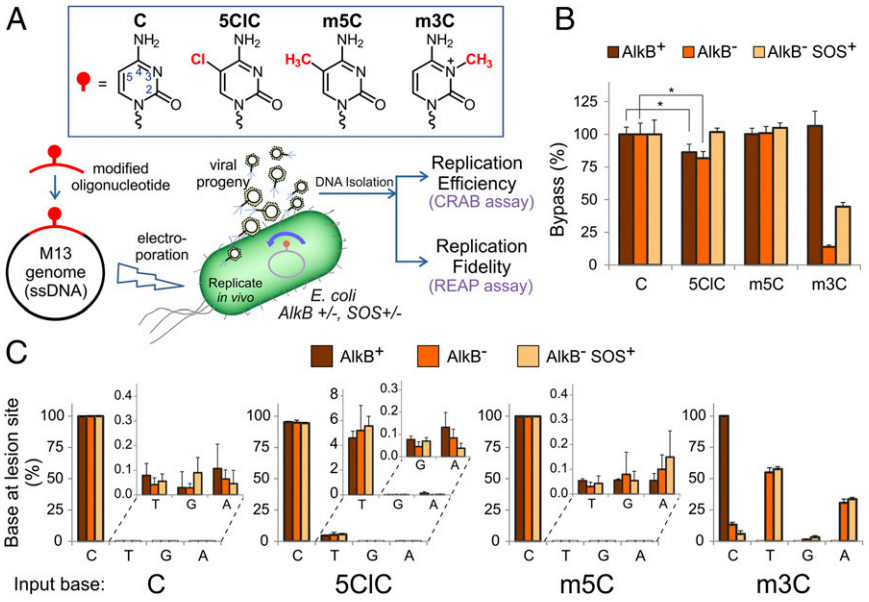

Fig. 2. The biological consequences of 5ClC in vivo. (*A*) General strategy for characterizing the in vivo properties of 5ClC. Oligodeoxynucleotides containing 5ClC, the negative controls cytosine (C) or 5-methylcytosine (m5C), or the positive control 3-methylcytosine (m3C) were ligated into M13 genomes, which were then introduced into *E. coli* cells of different genetic backgrounds and allowed to replicate. DNA from the resulting viral progeny was analyzed to determine the replication efficiency (a measure of how easily the lesion is bypassed) and replication fidelity (the ability of the lesion to induce mutations). A detailed graphical overview of the methods is shown in Fig. S1. (*B*) In vivo bypass efficiency of 5ClC and controls. The colors correspond to the three *E. coli* strains used, as indicated. Each graphed value corresponds to the mean of three independent biological replicates; error bars indicate one SD. $*P < 0.05$. Data are also tabulated in Table S1. (*C*) The mutagenicity of 5ClC in vivo. Each graph shows the base composition at the lesion site in the progeny phage coming from the constructed genomes containing the indicated input base at the lesion site. Each graphed value corresponds to the mean of three independent biological replicates; error bars indicate one SD. Data are also tabulated in Table S2.

was confirmed to be free of residual protecting groups and/or failure sequences (Fig. S2). As a further safeguard, a portion of the 5ClC-containing oligonucleotide was repurified with an independent HPLC method (*Materials and Methods* and Figs. S2 and S3) and was used to construct a genome that was analyzed side by side with the one prepared initially in the experiments outlined below. Finally, a lesion integrity assay was performed (described in *Materials and Methods* and Fig. S4*A*), which directly established the identity and purity of the 5ClC base present site-specifically in the constructed M13 genomes (Fig. S4*B*). Taken together, these analyses confirmed the high purity (>99%) of the 5ClC-containing oligonucleotides and M13 genomes, engendering confidence in the reported genetic results.

The Intrinsic Mutagenic Properties of 5ClC. Two possibilities could explain the in vivo mutagenic properties of 5ClC: either 5ClC is intrinsically mutagenic, or a cellular enzymatic process could act on 5ClC to convert it into a secondary, more mutagenic product. For example, 5ClC could be the substrate of a glycosylase, generating an abasic site, or it could be enzymatically deaminated to 5-chlorouracil (5ClU). Both scenarios, even if inefficient, could cause a portion of 5ClC to code as a T. To rule out a contribution of the intracellular milieu to the observed mutagenic properties of 5ClC, the modified M13 genomes were used as templates for an in vitro primer extension assay. The high-fidelity polymerase PfuTurbo was used to extend a primer and then generate a PCR product that can be analyzed directly with the REAP assay (Fig. 3*A*). The results of the in vitro primer extension were comparable to those obtained in the in vivo assays. Specifically, 5ClC was found to be mutagenic, coding both as a C (~94.2%) and as a T (~5.8%), whereas the controls (cytosine and m5C) coded, as expected, only as Cs (Fig. 3*B* and Table S3). Additionally, the results obtained with the repurified oligonucleotide were indistinguishable from the results obtained with the initial 5ClC oligonucleotide (Fig. 3*B* and Table S3). Thus, the qualitative and quantitative concordance of the in vivo and in vitro results strongly supports the view that 5ClC is intrinsically mutagenic and base-pairs with A at a rate of ~5%.

Next, we evaluated the mutagenic properties of 5ClC when traversed by different, but functionally well-understood, polymerases. A panel of eight different polymerases was chosen, covering all known classes of DNA polymerases (A, B, C, D, X, Y, and RT) and all domains of life (Table 1). The inclusion of three different human DNA polymerases (pol β, pol γ, and pol η) allowed exploration of the possible relevance of the 5ClC lesion in human cells and, by extension, to human disease.

Once again, the M13 genome site-specifically modified with 5ClC was used as a template in a modified primer extension assay denoted "iREAP" (in vitro REAP), which coupled the extension product of any test polymerase with the downstream quantitative analysis of the REAP assay (Fig. 3*C*). Here, the primer used (denoted "iPrimer") contained a unique 5′ overhang, which allowed a subsequent strand-specific PCR amplification of the newly synthesized strand (Fig. 3*C*). One key feature of this design was the low annealing temperature of the iPrimer (~40 °C), sufficient for a 37 °C primer extension reaction but inadequate for PCR. This feature prevented any unextended iPrimer from "jump-starting" the PCR and ensured that the outcome of the assay—the type and amount of the base present at the lesion site—reflected only the base-pairing preference and fidelity of the test polymerase used in the primer extension reaction and not that of the PCR polymerase. The strand-specific amplification approach permitted subsequent mutagenic analysis regardless of the yield of the primer extension reaction, and thus the assay could also provide data about the fidelity of low processivity polymerases. A set of control experiments was performed to demonstrate the primer design features. As shown in Fig. 3*D*, only when the iPrimer was successfully extended in the primer extension reaction did the subsequent PCR process yield the necessary amplicon for REAP analysis.

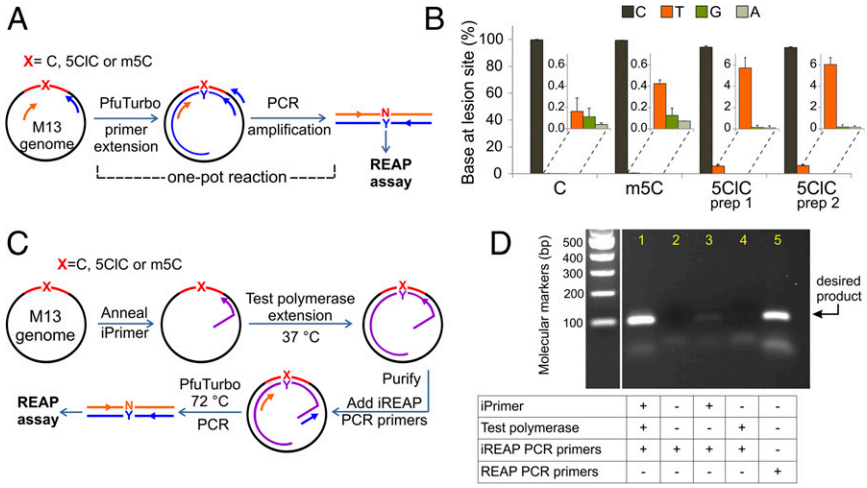

Fig. 3. The mutagenesis of 5ClC in vitro. (*A*) Strategy for analyzing the mutagenesis of 5ClC in vitro. The modified M13 genomes, used as templates, were incubated with the REAP PCR primers (blue and orange) and PfuTurbo polymerase. The first incubation extended only the blue primer, generating a double-stranded product, which was then PCR amplified with the aid of the orange primer. The identity and amount of base N is determined by the REAP assay. (*B*) In vitro mutagenesis of 5ClC and controls. The graphed values are averages of three replicates, with error bars representing one SD. Data are also tabulated in Table S3. (*C*) Schematic of the iREAP strategy. The M13 genome containing a lesion (5ClC) or a control base (cytosine or m5C) at a defined site was used as a template for a primer extension reaction. The primer (iPrimer) contained a unique 5′ overhang that did not anneal to the M13 genome. The primer extension was carried out with a chosen test polymerase at 37 °C. Subsequently, the extended product was purified to remove the polymerase and unextended iPrimer, and a PCR was performed so that only the newly synthesized strand was amplified. The PCR product was then analyzed using the REAP assay to identify the type and amount of base (denoted N) present at the lesion site. The results of this assay with eight different test polymerases are summarized in Table 1. (*D*) A control experiment demonstrating the specificity of the primers used in the iREAP assay. The PCR products obtained from control iREAP experiments 1–5 with a C genome have been separated on a 2% agarose gel and stained with ethidium bromide. In each experiment, the iPrimer, test polymerase (Klenow fragment), the iREAP PCR primers (or REAP PCR primers) were added (+) or left out (–), as indicated.

We found that 5ClC was mutagenic regardless of the polymerase used to replicate across the lesion, inducing a characteristic C→T transition at a frequency ranging from ∼3.5% (pol ζ and pol γ) to ∼9.1% (pol β) (Table 1). No clear correlation between the proofreading ability of the polymerase and the extent of miscoding was observed (Table 1). Together, these data strengthen the notion that 5ClC is an intrinsically mutagenic lesion being decoded as a base-pairing partner for both G and A by all classes of DNA polymerases found in nature.

Structural Analysis of 5ClC Replicative Bypass. To investigate the molecular basis for the mutagenic ability of 5ClC, a structural analysis was performed on a polymerase caught in the process of replicating across a templated 5ClC. To this end, we determined binary (pol/DNA) and ternary (pol/DNA/dNTP) X-ray crystal structures of human DNA polymerase β (pol β) with 5ClC in the templating position of a 1-nt–gapped DNA duplex (Table 2). Pol β is a model mammalian DNA polymerase involved in base excision repair and may encounter DNA lesions in the templating strand during repair of the genome (32–34). The binary DNA complex structure with 5ClC in the templating position diffracted to 2.0 Å. Fig. 4*A* shows pol β in an open conformation with 5ClC in an *anti*-conformation. The 5Cl atom is 3.9 and 4.3 Å away from the backbone phosphate oxygen atoms (Fig. 4*B*) and is accommodated within the major groove of the DNA, in contrast to other known mutagenic lesions, such as 8oxoG, which occupies both the *syn*- and *anti*-conformations in the open binary complex because of the steric clash of the O8 with the phosphate backbone (35).

Soaking pol β binary DNA complex crystals in a cryosolution with $MgCl_2$ and the nonhydrolyzable dGMP(CH_2)PP analog

Table 1. Mutagenic consequences of traversing 5ClC with several different DNA polymerases

Family	Polymerase	Domain	Species	Proofreading	% C→T induced by 5ClC
A	Pol I (Klenow fragment Exo$^+$)	Bacteria	*E. coli*	Yes	4.4 ± 0.3
A	Pol I (Klenow fragment Exo$^-$)	Bacteria	*E. coli*	No	4.4 ± 0.3
A	T4 DNA polymerase	Viruses	*T4 phage*	Yes	4.1 ± 0.2
A	Pol γ (gamma)	Eukaryota	*H. sapiens*	Yes	3.5 ± 0.3
B	Pol ζ (zeta)	Eukaryota	*S. cerevisiae*	No	3.4 ± 0.2
C	Pol III holoenzyme (in vivo)*	Bacteria	*E. coli*	Yes	4.6 ± 0.6
D	PfuTurbo†	Archaea	*P. furiosus*	Yes	5.8 ± 0.7
X	Pol β (beta)	Eukaryota	*H. sapiens*	No	9.1 ± 0.6
Y	Pol η (eta)	Eukaryota	*H. sapiens*	No	3.7 ± 0.7
RT	HIV RT	Viruses	*HIV*	No	4.9 ± 1.2

A panel of DNA polymerases spanning all known polymerase families (A, B, C, D, X, Y, RT) and all life domains was used to extend a primer annealed to an M13 ssDNA genome containing one 5ClC lesion. Subsequently, the newly synthesized strand was specifically PCR amplified (Fig. 3*C*), and the PCR product was analyzed with the REAP assay to determine the amount of C→T transitions induced by 5ClC.

*The result from Fig. 2*C*.

†The average of the two results from Fig. 3*B*.

Table 2. Data collection and refinement statistics of pol β binary and ternary complexes with 5ClC as the templating base

	Binary	Ternary incoming dGPCPP	Ternary incoming dAPCPP
Data collection			
Wavelength	1.54	1.54	1.54
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions			
a, *b*, *c*, Å	55.9, 79.5, 55.1	50.6, 79.1, 55.7	54.6, 78.6, 55.1
α, β, γ, °	90, 109.2, 90	90, 107.6, 90	90, 113.2, 90
Resolution, Å	50–2.0	50–1.98	50–2.0
R_{sym} or R_{merge}*, %	10.7 (48.1)	7.5 (50.3)	7.2 (51.1)
I/s*I*	14.1 (2.1)	16.3 (2.2)	19.8 (2.4)
Completeness, %	98.9 (90.8)	99.3 (95.9)	98.9 (90.9)
Redundancy	4.2 (2.5)	3.5 (2.2)	4.7 (2.8)
Refinement			
Resolution, Å	2.0	1.98	2.0
No. reflections	51,361	52,969	50,142
R_{work}/R_{free}	20.3/25.3	19.7/24.0	21.7/27.3
No. atoms			
Protein	2,593	2,674	2,670
DNA	644	644	644
Water	284	320	191
B-factors, Å			
Protein	37.1	28.1	41.4
DNA/5ClC/dNTP	33.6/45.8/–	35.4/24.5/19.4	40.8/54.1/33.4
Water	37.7	33.3	39.1
rms deviations			
Bond length, Å	0.01	0.01	0.01
Bond angles, °	1.1	1.1	1.07
Protein Data Bank ID	5BOM	5BOL	5BPC

*Highest-resolution shell is shown in parentheses.

promoted the open-to-closed conformational change without catalysis. Using this approach, we determined the nonmutagenic closed ternary preinsertion complex to 1.98 Å (Table 2). 5ClC remained in the *anti*-conformation and formed a good Watson–Crick base pair with the incoming dGTP analog (Fig. 4*C*). Mg^{2+} was bound both in the catalytic and nucleotide metal-binding sites, and O3′ of the primer terminus was in a catalytically competent position 3.5 Å to Pα of the bound nucleotide (Fig. 4*C*). Overlaying this structure with the nondamaged dC:dGTP ternary complex structure trapped with calcium showed they were nearly identical, with an rmsd of 0.17 Å; both structures formed identical Watson–Crick hydrogen-bonding interactions (Fig. 4*D*). The major groove can accommodate the 5Cl atom, with the closest active site residue being Asn37 (3.8 Å away). Additionally, the phosphate backbone did not have to alter its position and remained 3.6 Å away. In contrast to the major groove, the DNA minor groove restricts the geometry of the nascent base pair. Arg283 was within hydrogen bonding distance of O2 for both templating dC and 5ClC. These results demonstrate that the major groove accommodates the 5Cl atom, facilitating Watson–Crick base-pairing between 5ClC and dGTP within the active site. This accommodation explains the propensity for error-free lesion bypass (G being the major base-pairing partner for 5ClC) observed in the cellular studies.

The in vivo and in vitro results reported here suggest that 5ClC promotes C:G→T:A transition mutations by base-pairing with adenine in addition to the expected base-pairing partner guanine. To elucidate the mechanistic basis for the 5ClC mutagenesis, we obtained the closed ternary complex of a templating 5ClC with an incoming dAMP(CH_2)PP to 2.0 Å (Table 2). Because we were unable to obtain a closed ternary structure in the presence of $MgCl_2$, we instead used $MnCl_2$ during the soak step. The need for a more accommodating metal cation likely reflects the reduced binding affinity of dATP for the templating 5ClC. Fig. 5*A* shows the resulting closed ternary complex forming a mismatch intermediate conformation with Mn^{2+} in the catalytic and nucleotide-binding sites (36). This conformation resulted in the coding template base being shifted upstream by more than 3 Å and generating an abasic-like templating pocket within the active site that is occluded by Lys280 (Fig. 5 *A* and *B*). In this conformer, the only unique interaction of the templating 5ClC with the incoming dATP analog was between N4 of 5ClC and N6 of the incoming dATP analog. This interaction was facilitated by a well-ordered bridging water molecule near N4 of 5ClC, N6 of dATP, and Lys280 (Fig. 5*A*). In addition, the primer terminus sugar rotated away from the catalytic metal, generating a suboptimal O3′ position that did not coordinate the catalytic metal (Fig. 5 *A* and *B*). Accordingly, misinsertion of dATP would require a minor shift of O3′ to coordinate the catalytic metal and initiate catalysis, as has been proposed previously during mismatch insertion (37). These structural rearrangements are in contrast to the ordered active site and Watson–Crick base-pairing interactions observed between the templating 5ClC and dGTP analog (Fig. 5*C*).

During the mutagenic insertion of the dATP analog, 5ClC remained in the *anti*-conformation, and N4 of 5ClC was within hydrogen-bonding distance to N6 of the incoming dATP analog (Fig. 5*A*). An overlay of the pol β dAMP(CH_2)PP insertion opposite 5ClC or natural cytosine is shown in Fig. 5*D*. In contrast to 5ClC, the nondamaged templating cytosine adopted a *syn*-conformation with no hydrogen-bonding interactions between the templating base and the incoming dATP analog. This result indicates that the 5Cl substituent on 5ClC was promoting the *anti*-conformation of the base by favoring the major over the minor DNA groove position. Fig. 5*D* shows a surface representation highlighting the unrestrained space in the major groove that accommodates the 5Cl group. If 5ClC were to adopt a *syn*-conformation, similar to

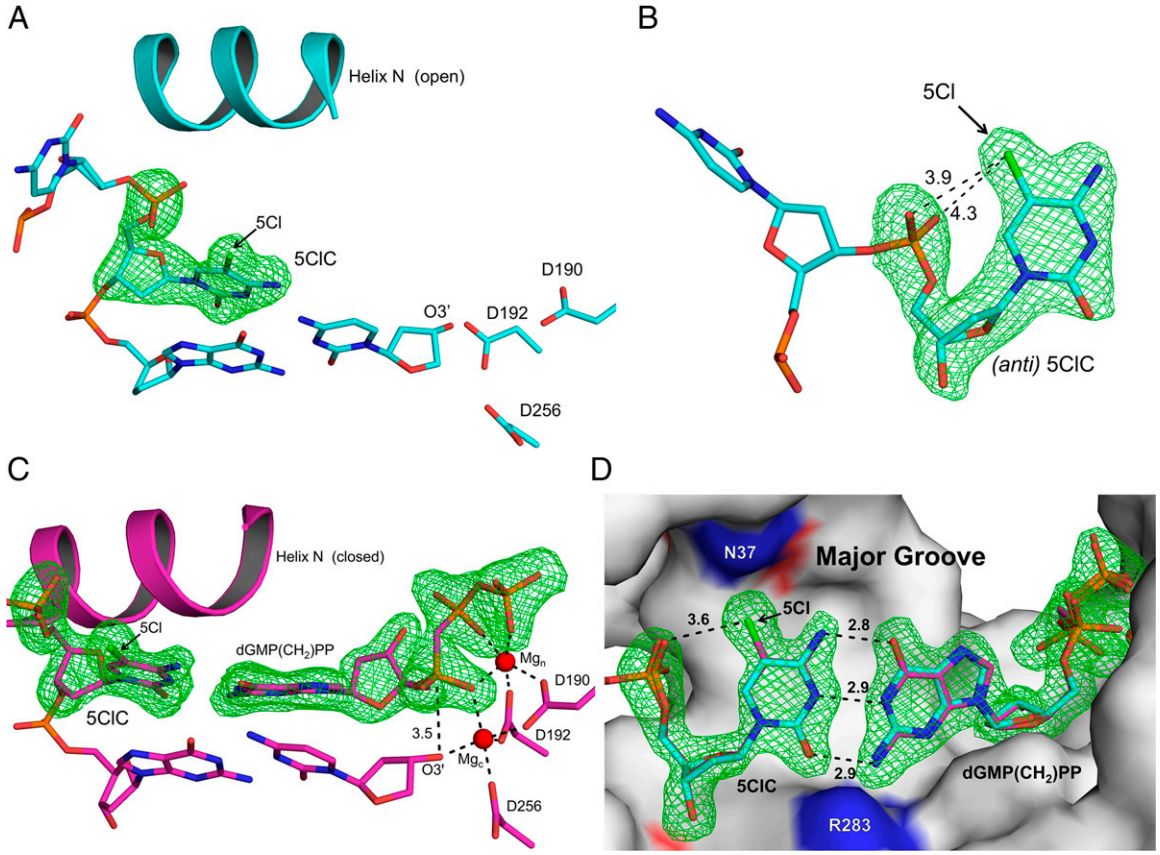

Fig. 4. Structure of pol β binary and nonmutagenic ternary complex with 5ClC-containing oligonucleotide. F_o-F_c omit maps (green) are contoured at 3σ. Helix N of pol β, primer terminus (O3′), 5Cl of 5ClC, and key active-site residues are indicated. (*A*) The binary pol β complex shown in the open conformation. (*B*) A different viewing angle of the structure highlights the *anti*-geometry of the 5ClC base and the distances (in Ångstroms) between the 5Cl atom and the oxygen atoms of the phosphate backbone. (*C*) The pol β ternary complex with an incoming dGMP(CH_2)PP base-pairing with 5ClC. Mg^{2+} ions are shown as red spheres in the catalytic and nucleotide-binding sites. The distance (in Ångstroms) between O3′ and Pα is shown. (*D*) The 5ClC and dGMP(CH_2)PP are shown in a stick representation (purple), and the protein is shown in surface representation (gray). The distances correspond to the Watson–Crick hydrogen bonding between 5ClC and the dGTP analog. The DNA major groove edge is indicated. For reference the dC:dGTP (PDB ID code 4UB4) ternary complex is superimposed (light blue).

the nondamaged cytosine, the 5Cl atom would sterically clash with Tyr271 and Glu295 (Fig. 5*D*). Therefore, the *anti*-conformation of the 5ClC base, adopted to accommodate the 5Cl bulky substituent in the DNA major groove, is a key factor that promotes a favorable hydrogen-bonding interaction between 5ClC and the incoming dATP, and likely constitutes the molecular basis for the mutagenic mispairing ability of 5ClC, given the constrained minor groove of most DNA polymerases.

Discussion

The Molecular Basis of 5ClC Mutagenesis. The present work unveils the mutagenic properties of 5ClC, a modified base previously shown to accumulate in cells and tissues exposed to chronic inflammation (20, 21). We found that when present in a template strand, 5ClC has an increased probability (compared with the replication fidelity of canonical bases) of mispairing with adenine and thus miscoding as a T. The structural evidence reported here confirms the existence of a direct interaction between 5ClC and incoming dATP in the active site of a polymerase. This unique interaction and the formation of an apparent abasic site templating pocket are sufficient to promote the mutagenic insertion of dATP opposite 5ClC by an "enhanced A-rule" mechanism whereby a polymerase incorporates an A opposite a structure resembling an abasic site (38).

The structural snapshots show that during the mutagenic bypass of the lesion, 5ClC adopts a conformation in which the 5Cl substituent is placed in the major groove of DNA. Because, for most DNA polymerases, the DNA major groove of the nascent base pair can generally accommodate bulky substituents, this phenomenon is likely universal, as supported by the mutagenesis data across all families of DNA polymerases (Table 1). In contrast, the minor groove edge often contacts protein side chains within the polymerase active site, thereby limiting base conformations that protrude in the DNA minor groove. By favoring the major groove position for the 5Cl group, the 5ClC base primarily adopts an *anti*-conformation, thereby presenting its Watson–Crick hydrogen-bonding edge to the incoming nucleotide (Figs. 4 and 5). In contrast, a nondamaged cytosine is known to occupy both the *anti*- and *syn*-conformations, and thus the probability of forming a stable base pair with incoming dATP is greatly reduced (Fig. 5*D*). Furthermore, our crystal structures reveal an unusual and rather unique interaction between 5ClC and the incoming adenine, with the N4 of 5ClC and N6 of adenine being within hydrogen-bonding distance (3.0 Å). Although the interaction has only one putative hydrogen bond (and/or a water bridge), it is sufficient to promote the conformational change in the polymerase from the open form to the prechemistry closed form. In fact, mutagenic mispairings of damaged bases that rely on only one hydrogen bond have been reported before [e.g., N^2,3-ethenoguanine mispairing with T (39)]. Moreover, the base-pairing we observe between 5ClC and A is strikingly different from the base-pairing between cytosine and adenine previously reported to occur across multiple polymerase families as either a wobble base pair or a base pair involving minor tautomeric forms of either base (40–42). Nevertheless, we cannot rule out the possibility that

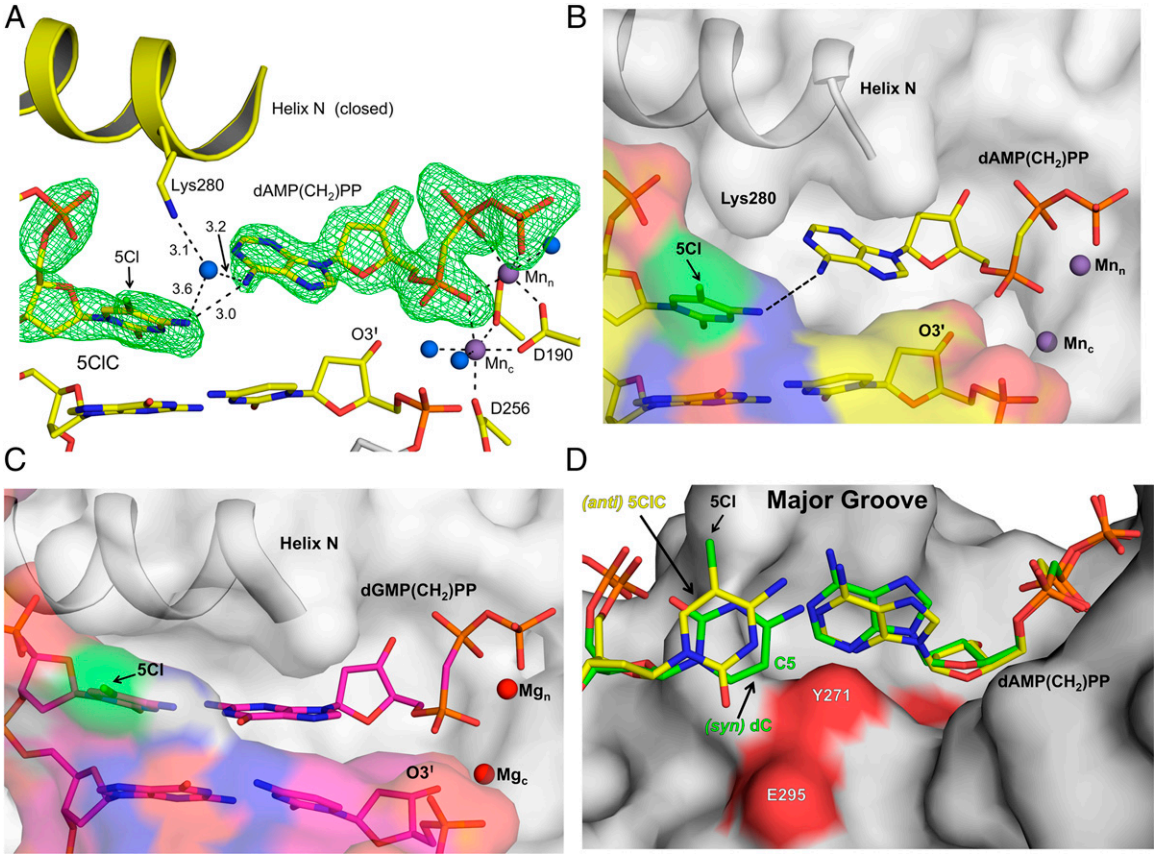

Fig. 5. Structures of the pol β closed ternary insertion complex during mutagenic 5ClC lesion bypass. F_o-F_c omit maps (green) are contoured at 3σ. Helix N, primer terminus (O3′), 5Cl of 5ClC, and key active-site residues are indicated. (*A*) The pol β mismatch ternary complex with an incoming dAMP(CH_2)PP opposite 5ClC with putative hydrogen-bonding interaction distances (in Ångstroms) shown. Mn^{2+} ions are shown as purple spheres in the catalytic and nucleotide-binding sites. Water molecules are shown in blue. (*B*) The view in *A* with the protein (gray) and DNA (yellow) shown in surface representation. Incoming dAMP(CH_2)PP is shown in stick representation. (*C*) The nonmutagenic ternary complex with an incoming dGMP(CH_2)PP base-pairing with 5ClC is shown with the protein (gray) and DNA (purple) in surface representation. The Mg^{2+} ions are shown in red, and the incoming dGMP(CH_2)PP is shown in stick representation. (*D*) The *anti*-5ClC and dAMP(CH_2)PP are shown in stick representation (yellow), and the protein is shown in surface representation (gray). The DNA major groove edge is indicated. For reference, the mismatch *syn*-dC:dATP (PDB ID code 3C2L) ternary complex is superimposed (green). The C5 position of the nondamaged cytosine is indicated.

during the nucleotidyl transfer reaction, the pairing between 5ClC and A transitions toward a wobble base pair or a tautomeric pair. Importantly, both of these possibilities will be stabilized by the presence of the 5Cl substituent, which promotes the *anti*-conformation of the 5ClC base, whereas an alternative *syn*-conformer would not form stable wobble base pairs.

Interestingly, by extending the steric argument, m5C, similarly to 5ClC, is expected to adopt a primarily *anti*-conformation. However, 5ClC is much more mutagenic than m5C, suggesting that factors other than sterics must be involved. For example, the presence of the chlorine atom on the base is known to affect the electron density and distribution on the aromatic ring; as a consequence, the pK_a of the N3 position of 5ClC (2.65) is significantly lower than the pK_a of cytosine at the same atom (4.30) (23). In comparison, m5C has a pK_a similar to that of cytosine (4.35) (23). The electronic effect of chlorine was shown to be strong enough to influence the energy of the hydrogen bonds involved in base-pairing; thus, the estimated base-pairing energy of 5ClC:G is slightly less than the energy of the C:G pair (43). Future studies probing the electronic configuration of 5ClC will be of interest to provide a more comprehensive understanding of the way that the chlorine substituent alters the base-pairing potential of 5ClC.

5ClC joins a substantial group of other substituted cytosines or cytosine analogs that are known to be mutagenic and induce C→T mutations. 5-Hydroxycytosine, an oxidative stress-induced lesion, is believed to mispair with adenine when it adopts an imine minor tautomeric form (44). Tautomerism is also the explanation for the mutagenic properties of 5-aza-5,6-dihydrocytosine, a cytosine analog used as an antiviral therapeutic (28). Cytosine analogs such as bicyclocytosine (45), N^4-hydroxycytosine (46), and N^4-aminocytosine (47) are mutagenic because of their ability to form Watson–Crick or wobble base pairs with both purines. Although 5-methylcytosine, an important epigenetic marker, is not a mutagenic base (as confirmed by the data in the present study), it has an increased potential to deaminate to thymine; if the resulting mismatch is not repaired, the deamination of 5-methylcytosine thus can lead to C→T mutations (48, 49). This mechanism is believed to account for the C→T mutations occurring in the CpG islands, where the cytosine is most likely to be methylated (48, 49). Additionally, 5-methylcytosine is also oxidatively processed by the TET enzymes to generate sequentially 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine; of these 5-substituted cytosines, only 5-formylcytosine can induce C→T mutations in vitro, at a frequency of ∼1% (50, 51).

The Mutagenic Impact of 5ClC in Mammalian Cells. Our data convincingly suggest that 5ClC is a mutagenic lesion, inducing about 3–9% C→T transition mutations depending on the polymerase at play. Although by itself this number seems relatively low, the impact of 5ClC and, by extension, the impact of chlorination damage on mammalian cell biology will depend on the total number of lesions present in the genome. The upper bound of the

extent of 5ClC formation in DNA was reported in a study in which physiologically achievable concentrations of HOCl transformed in vitro about 1–10 of every 1,000 cytosines in 5ClC (19). At this level of incorporation, the mutagenesis of 5ClC would be catastrophic for a mammalian-sized genome. More typically, the levels of genomic 5ClC in chronically inflamed tissues have been measured at ∼1–30 lesions per 10^8 bases (20, 21) or 66–1,980 lesions per diploid human genome. Given the inherent mutagenicity of 5ClC described here, when a cell having this lesion burden replicates, it would acquire about 3.3–99 new mutations. This number represents a notable increase over the spontaneous mutation rate, estimated for mammalian cells to be on average 1.1–1.5 mutations per cell division (52, 53). Given that no repair pathway specific for 5ClC has yet been described (perhaps explaining the accumulation and persistence of this lesion in the genome), it is plausible that 5ClC is an important contributor to the mutagenic burden previously observed in tissues subjected to chronic inflammation (54, 55) and thus increases the probability that such tissues progress toward malignancy.

The Mechanistic Link Between Chronic Inflammation and Genetic Modifications in Tumors. The patterns of genetic changes seen in the genomes of cancer cells are indicative of the mutagenic processes that contributed to the conversion of normal cells to tumors (48, 56, 57). Thirty such patterns, called "mutational signatures," have been identified; each consists of a histogram of relative frequencies of all possible point mutations occurring within all possible trinucleotide sequence contexts (48, 56). However, only a fraction of mutational signatures have been ascribed to specific mutagenic processes (56). Among all possible base substitution mutations, C→T (and correspondingly G→A) transitions are the most common, and for certain mutational signatures, they have been attributed to well-known mechanisms such as m5C deamination and AICDA (activation-induced cytidine deaminase)/APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like)-dependent deamination (48, 56). Nevertheless, other mutational signatures, including many of unknown etiology, include C→T mutations (48, 56). Given our present study, we suspect that 5ClC could also be a contributor to the C→T mutations observed in sequenced tumors as part of one or several different mutational signatures. The well-established infiltration of tumors with immune cells, including neutrophils (2, 9, 58), provides a rich opportunity for HOCl production and, hence, 5ClC formation.

We postulate that cells exposed to chronic inflammatory processes, more specifically those mediated by activated neutrophils and HOCl production, would bear in their DNA the mutagenic legacy of chlorination damage (Fig. 1). Such a mutagenic signature would feature C→T mutations characteristic of 5ClC formation and also other mutations corresponding to the variety of oxidative and chlorinated lesions induced by HOCl (18). Some features of the mutagenic spectrum generated by chronic inflammation already have been reported in mouse models, and they include increased C→T mutations as well as small DNA base insertions and deletions (indels) (54). Intriguingly, in tumor sequence data, both the mutational signature 6 of Alexandrov et al. (56), and the H signature of Jia et al. (57) feature extensive C→T transitions (but also smaller levels of other mutations such as T→C and C→A) and are strongly associated with colorectal cancers, malignancies that often develop in chronically inflamed tissues. It has been proposed that these mutational signatures reflect the microsatellite instability phenotype of many colorectal tumors (48, 56, 57), explaining well the indels that also have been reported to associate with these signatures (56). However, microsatellite instability does not fully explain the observed C→T-dominated point mutation pattern. Interestingly, one of the mechanisms that leads to microsatellite instability is inhibition of *MLH1* expression, a key mismatch repair protein, by hypermethylation (59). As mentioned earlier, one biochemical feature of 5ClC, when present in CpG islands, is the ability to induce DNA hypermethylation (25, 60). Therefore, we speculate that signature 6 (and/or signature H) may reflect both the mutagenic and the epigenetic features of genomic 5ClC accumulation and thus represent the mutational signature of neutrophil-driven chronic inflammation (Fig. 1). Further work is needed to investigate this hypothesis.

Materials and Methods

Enzymes. Polymerase pol β used for in vitro extension reactions was from Novoprotein, polymerases γ and ζ were from Enzymax LLC, HIV-RT was from Worthington Biochemical Corp., and PfuTurbo was from Agilent Technologies. Polymerase η was a generous gift from Wei Yang (National Institute of Diabetes and Digestive and Kidney Diseases/NIH, Bethesda). The other polymerases and restriction enzymes were from New England Biolabs.

5ClC Nucleoside. 5ClC deoxynucleoside was from Alchem Laboratories Corporation; its identity and purity was confirmed by NMR. The aqueous stability of 5ClC nucleoside was also evaluated. 5ClC was found to be stable to deamination at biologically relevant pHs. No appreciable deamination (evaluated on reverse-phase HPLC) was detected after incubation with 2 mM NaOH, 2 mM NaOMe, or 30% (wt/vol) ammonia for 24 h.

Oligonucleotides. All unmodified oligonucleotides were from Integrated DNA Technologies. 5ClC, m3C, and m5C modified bases were incorporated into 16mer oligonucleotides (5′-GAAGACCTXGGCGTCC-3′ and 5′CCGACXGCGCATCAGC-3′, where X is the modified nucleobase) by using phosphoramidite solid-phase methods described before (26, 29). The 5ClC nucleoside was protected at N4 with a benzoyl group and then was used to prepare the 5ClC phosphoramidite (at ChemGenes). Oligonucleotide syntheses containing the 5ClC phosphoramidite were done at ChemGenes or Boston Open Labs.

Oligonucleotide Purification and Characterization. The oligonucleotides containing 5ClC were deprotected with aqueous ammonium hydroxide and purified by reverse-phase HPLC on a Varian Microsorb-MV 100-5 250 × 4.6 mm C18 column (Varian) or on an Acclaim Polar Advantage 3-μm 250 × 3.0 mm C18 column (Thermo Scientific) using 10 mM ammonium acetate and acetonitrile as solvents. The 2× purified oligonucleotide was purified on both columns sequentially. Purity was evaluated on an Agilent 6510 electrospray ionization (ESI) quadrupole time-of-flight (Q-TOF) mass spectrometer (Fig. S2). The purification conditions allowed the separation of 16mer oligonucleotides that differ by only one mass unit (Fig. S3) and therefore were expected to separate the oligonucleotide containing 5ClC from one containing the putative contaminant 5ClU.

M13 Genomes Construction. Oligonucleotides containing lesions and controls were ligated into an M13mp7(L2) single-stranded viral genome by using reported methods (26, 29). The constructed genomes were purified using phenol-chloroform extraction and QIAquick columns (Qiagen).

Lesion Integrity Assay. A lesion integrity assay, adapted from ref. 29, was performed to establish directly the identity and purity of the site-specific base present in the constructed M13 genomes. Although the HPLC purification of the oligonucleotides should have removed any sequences containing 5ClU, detecting small levels of 5ClU in the MS analysis was challenging because 5ClU is only 1 Da heavier than 5ClC. It was important to detect and quantify the amount of 5ClU still remaining in the oligonucleotide sample or that may have formed during genome construction, because 5ClU would likely code as a T and contribute to the mutagenic signal. To that end, a scaffold was used to frame the lesion on the single-stranded M13 genome, after which the site of interest was analyzed using REAP. To account for any bias the scaffold might introduce, the experiment was performed independently with two scaffolds, one containing G and the other containing A opposite the lesion. To 150 fmol of M13 genome containing 5ClC prepared above, 250 fmol of a scaffolding oligonucleotide was added and incubated at 80 °C for 5 min, followed by a slow cooling (0.1 °C/s) for annealing. The scaffolds (5′-CCAGTGAATTGGACGCCYAGGTCTTCCACTGAATCA-3′, where Y is either G or A) spanned the lesion site, being complementary to 15 or 16 bases on either side of the lesion. This construct was subjected directly to the REAP assay, starting with the BbsI digest step (Fig. S4*A*) (26). To prepare 5CldCMP and 5CldUMP standards, an oligonucleotide containing 5ClC at the 5′ end was radiolabeled, isolated by PAGE, and digested to monophosphates with P1 nuclease. The mixture then was incubated in 0.1 M HCl for 30 min to promote partial deamination of 5ClC to 5ClU, and the spot for 5CldUMP was confirmed by a cold 5′ monophosphorylated standard. With either scaffold, the constructed

M13 genome was found to contain >99% 5ClC at the site of inquiry, whereas any putative 5ClU present (as judged from the TLC standard) was <1% (Fig. S4*B*).

In Vivo Lesion Bypass and Mutagenesis Assays. The constructed M13 viral genomes (mixed with a competitor genome) were transfected into *E. coli* strains by electroporation. The *E. coli* strains used were AlkB$^+$ (wild type) (AB1157, *nalA*) and AlKB$^-$ (AlkB deficient) (AB1157, *nalA*, *alkB22*). Lesion bypass was evaluated with the CRAB assay; lesion mutagenesis was measured with the REAP assay (26). Briefly, the resulting progeny phage were collected and reamplified; their DNA then was extracted using the QIAprep M13 kit (Qiagen). The region of interest (where the lesion used to reside) was PCR amplified with assay-specific primers. The CRAB primers amplified both input and competitor sequences and thus allowed the estimation of the relative bypass (Fig. S1). A genome normalization step (26) also was used to estimate the relative concentrations of the different M13 genomes accurately. The REAP primers amplified only the progeny coming from the input genomes (lesion-containing genomes) to maximize the signal obtained from the lesion site. Restriction digests and radiolabeling yielded an18mer fragment that contained the base at the lesion site at its 5′ end. PAGE isolation of this fragment and the subsequent P1 nuclease digest produced the radioactive monophosphate of the base at the lesion site, which was subsequently analyzed by TLC (Cellulose-PEI) and quantified by phosphorimagery (Fig. S1).

In Vitro Mutagenesis Assay. The M13 genomes constructed as described above were used as templates for the in vitro mutagenesis assay (iREAP). An extension primer (5′-CGTGACGATGCGCAGACTGATATCATGTGTAAAACGACGGC-3′) was added in equimolar amount and annealed by heating the mixture at 85 °C for 5 min, followed by a slow (0.1 °C/s) cooling to 4 °C. The primer was extended using 1 U of polymerase (as indicated) in NEBuffer1 (New England Biolabs) supplemented with 1 mM DTT and 200 μM of each of the four dNTPs (New England Biolabs) at 37 °C for 1 h. The polymerase was inactivated at 95 °C for 10 min; then the reaction was transferred quickly on ice to minimize reannealing of unextended primer. The extension product was purified from unextended primer and polymerase using QIAquick columns (Qiagen) and an additional wash with 0.5× PE buffer (Qiagen). The eluent was then PCR amplified (32 cycles of 95/70/72 °C for 30/15/60 s, respectively) with PfuTurbo polymerase, using the primers 5′-HGGAAACAGCTATGACCATGATTCAGTGGAAGAC-3′ (forward) and 5′-HCGTGACGA TGCGCAGACTGATATCATGT-3′ (reverse), where H is a phospho-hexylamino linker which prevents 5′ phosphorylation during subsequent steps. The PCR products were then purified and analyzed with the REAP assay as previously described (26). For the in vitro mutagenesis assay using PfuTurbo, the M13 genomic template containing a site-specific lesion was directly PCR amplified with PfuTurbo with the following primers: 5′-HCAGCTATGACCATGATTCAG-TGGAAGAC-3′ (forward) and 5′-HTGTAAAAC-GACGG CCAGTGAATTGGACG-3′ (reverse), because in this case the first PCR step is essentially a primer extension reaction using PfuTurbo. The resulting PCR products were subsequently purified and analyzed as above.

Structural Studies of 5ClC with DNA pol β. Human wild-type pol β was overexpressed in *E. coli* and purified as described previously (61). Binary complex crystals with a templating 5ClC in a 1-nt–gapped DNA were grown as previously described (62). The 16mer oligonucleotide containing 5ClC was synthesized and thoroughly purified on reverse-phase HPLC as described above. The binary complex crystals were transferred to a cryosolution containing 12% ethylene glycol, 50 mM imidazole (pH 7.5), 20% PEG3350, 90 mM sodium acetate, and 5 mM nonhydrolyzable dNTP analogs dGMP(CH_2)PP or dAMP(CH_2)PP with 200 mM $MgCl_2$ or $MnCl_2$, thus resulting in the ternary complex.

Data were collected at 100 K on an in-house Saturn92 CCD detector system mounted on a MiraMax-007HF rotating anode generator at a wavelength of 1.54 Å. Data were processed and scaled using the HKL2000 software package (63). Initial models were determined using molecular replacement with the open [Protein Data Bank (PDB) ID code 3ISB] or closed (PDB ID code 2FMS) structures of pol β as a reference. All R_{free} flags were taken from the starting model, and refinement was carried out using PHENIX (64) and model building using Coot (65). The figures were prepared using PyMOL and AxPyMOL Molecular Graphics plugin (v1.0, Schrodinger LLC), and all density maps were generated after simulated annealing was performed. Ramachandran analysis determined that 100% of nonglycine residues lie in the allowed regions.

ACKNOWLEDGMENTS. We thank Allison Simi, Maria Kulikova, and Nicole Zatorski for their assistance; Koli Taghizadeh for technical advice; and Charles Knutson, Steven Tannenbaum, and Gerald Wogan for valuable feedback on the manuscript. We also thank the MIT Center for Environmental Health Sciences (MIT CEHS) for providing access to the Instrumentation Core facility. This work was supported by NIH Grants P01 CA26731 and R37 CA080024 (both to J.M.E), NIH Center Grant P30 ES002109 (to the MIT CEHS), and by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences, Project Nos. Z01-ES050158 (S.H.W.).

1. Brenner DR, et al. (2014) A review of the application of inflammatory biomarkers in epidemiologic cancer research. *Cancer Epidemiol Biomarkers Prev* 23(9):1729–1751.
2. Elinav E, et al. (2013) Inflammation-induced cancer: Crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer* 13(11):759–771.
3. Kundu JK, Surh Y-J (2008) Inflammation: Gearing the journey to cancer. *Mutat Res* 659(1-2):15–30.
4. Hussain SP, Harris CC (2007) Inflammation and cancer: An ancient link with novel potentials. *Int J Cancer* 121(11):2373–2380.
5. Balkwill F, Charles KA, Mantovani A (2005) Smoldering and polarized inflammation in the initiation and promotion of malignant disease. *Cancer Cell* 7(3):211–217.
6. Coussens LM, Werb Z (2002) Inflammation and cancer. *Nature* 420(6917):860–867.
7. Mangerich A, Dedon PC, Fox JG, Tannenbaum SR, Wogan GN (2013) Chemistry meets biology in colitis-associated carcinogenesis. *Free Radic Res* 47(11):958–986.
8. Lonkar P, Dedon PC (2011) Reactive species and DNA damage in chronic inflammation: Reconciling chemical mechanisms and biological fates. *Int J Cancer* 128(9):1999–2009.
9. Grivennikov SI, Greten FR, Karin M (2010) Immunity, inflammation, and cancer. *Cell* 140(6):883–899.
10. Reuter S, Gupta SC, Chaturvedi MM, Aggarwal BB (2010) Oxidative stress, inflammation, and cancer: How are they linked? *Free Radic Biol Med* 49(11):1603–1616.
11. Mantovani A, Cassatella MA, Costantini C, Jaillon S (2011) Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat Rev Immunol* 11(8):519–531.
12. van der Veen BS, de Winther MPJ, Heeringa P (2009) Myeloperoxidase: Molecular mechanisms of action and their relevance to human health and disease. *Antioxid Redox Signal* 11(11):2899–2937.
13. Winterbourn CC, Hampton MB, Livesey JH, Kettle AJ (2006) Modeling the reactions of superoxide and myeloperoxidase in the neutrophil phagosome: Implications for microbial killing. *J Biol Chem* 281(52):39860–39869.
14. Henderson JP, Byun J, Heinecke JW (1999) Molecular chlorine generated by the myeloperoxidase-hydrogen peroxide-chloride system of phagocytes produces 5-chlorocytosine in bacterial RNA. *J Biol Chem* 274(47):33440–33448.
15. Winterbourn CC (2002) Biological reactivity and biomarkers of the neutrophil oxidant, hypochlorous acid. *Toxicology* 181-182:223–227.
16. Winterbourn CC, Kettle AJ (2000) Biomarkers of myeloperoxidase-derived hypochlorous acid. *Free Radic Biol Med* 29(5):403–409.
17. Kawai Y, et al. (2004) Endogenous formation of novel halogenated 2′-deoxycytidine. Hypohalous acid-mediated DNA modification at the site of inflammation. *J Biol Chem* 279(49):51241–51249.
18. Whiteman M, Jenner A, Halliwell B (1997) Hypochlorous acid-induced base modifications in isolated calf thymus DNA. *Chem Res Toxicol* 10(11):1240–1246.
19. Kang JI, Jr, Sowers LC (2008) Examination of hypochlorous acid-induced damage to cytosine residues in a CpG dinucleotide in DNA. *Chem Res Toxicol* 21(6):1211–1218.
20. Mangerich A, et al. (2012) Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc Natl Acad Sci USA* 109(27):E1820–E1829.
21. Knutson CG, et al. (2013) Chemical and cytokine features of innate immunity characterize serum and tissue profiles in inflammatory bowel disease. *Proc Natl Acad Sci USA* 110(26):E2332–E2341.
22. Valinluck V, Liu P, Kang JI, Jr, Burdzy A, Sowers LC (2005) 5-halogenated pyrimidine lesions within a CpG sequence context mimic 5-methylcytosine by enhancing the binding of the methyl-CpG-binding domain of methyl-CpG-binding protein 2 (MeCP2). *Nucleic Acids Res* 33(9):3057–3064.
23. Valinluck V, Wu W, Liu P, Neidigh JW, Sowers LC (2006) Impact of cytosine 5-halogens on the interaction of DNA with restriction endonucleases and methyltransferase. *Chem Res Toxicol* 19(4):556–562.
24. Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 67(3):946–950.
25. Lao VV, et al. (2009) Incorporation of 5-chlorocytosine into mammalian DNA results in heritable gene silencing and altered cytosine methylation patterns. *Carcinogenesis* 30(5):886–893.
26. Delaney JC, Essigmann JM (2006) Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions in vivo. *Methods Enzymol* 408:1–15.
27. Shrivastav N, et al. (2014) A chemical genetics analysis of the roles of bypass polymerase DinB and DNA repair protein AlkB in processing N2-alkylguanine lesions in vivo. *PLoS One* 9(4):e94716.
28. Li D, et al. (2014) Tautomerism provides a molecular explanation for the mutagenic properties of the anti-HIV nucleoside 5-aza-5,6-dihydro-2′-deoxycytidine. *Proc Natl Acad Sci USA* 111(32):E3252–E3259.
29. Delaney JC, Essigmann JM (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc Natl Acad Sci USA* 101(39):14051–14056.
30. Delaney S, Neeley WL, Delaney JC, Essigmann JM (2007) The substrate specificity of MutY for hyperoxidized guanine lesions in vivo. *Biochemistry* 46(5):1448–1455.

31. Wood ML, Dizdaroglu M, Gajewski E, Essigmann JM (1990) Mechanistic studies of ionizing radiation and oxidative mutagenesis: Genetic effects of a single 8-hydroxyguanine (7-hydro-8-oxoguanine) residue inserted at a unique site in a viral genome. *Biochemistry* 29(30):7024–7032.
32. Freudenthal BD, Beard WA, Wilson SH (2013) DNA polymerase minor groove interactions modulate mutagenic bypass of a templating 8-oxoguanine lesion. *Nucleic Acids Res* 41(3):1848–1858.
33. Kothandapani A, et al. (2011) Novel role of base excision repair in mediating cisplatin cytotoxicity. *J Biol Chem* 286(16):14564–14574.
34. Beard WA, Batra VK, Wilson SH (2010) DNA polymerase structure-based insight on the mutagenic properties of 8-oxoguanine. *Mutat Res* 703(1):18–23.
35. Batra VK, Shock DD, Beard WA, McKenna CE, Wilson SH (2012) Binary complex crystal structure of DNA polymerase β reveals multiple conformations of the templating 8-oxoguanine lesion. *Proc Natl Acad Sci USA* 109(1):113–118.
36. Batra VK, Beard WA, Shock DD, Pedersen LC, Wilson SH (2008) Structures of DNA polymerase beta with active-site mismatches suggest a transient abasic site intermediate during misincorporation. *Mol Cell* 30(3):315–324.
37. Freudenthal BD, Beard WA, Shock DD, Wilson SH (2013) Observing a DNA polymerase choose right from wrong. *Cell* 154(1):157–168.
38. Zahn KE, Averill A, Wallace SS, Doublié S (2011) The miscoding potential of 5-hydroxycytosine arises due to template instability in the replicative polymerase active site. *Biochemistry* 50(47):10350–10358.
39. Zhao L, et al. (2012) Basis of miscoding of the DNA adduct N2,3-ethenoguanine by human Y-family DNA polymerases. *J Biol Chem* 287(42):35516–35526.
40. Wang W, Hellinga HW, Beese LS (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc Natl Acad Sci USA* 108(43):17644–17648.
41. Xia S, Wang J, Konigsberg WH (2013) DNA mismatch synthesis complexes provide insights into base selectivity of a B family DNA polymerase. *J Am Chem Soc* 135(1): 193–202.
42. Zhao Y, et al. (2013) Mechanism of somatic hypermutation at the WA motif by human DNA polymerase η. *Proc Natl Acad Sci USA* 110(20):8146–8151.
43. Theruvathu JA, Yin YW, Pettitt BM, Sowers LC (2013) Comparison of the structural and dynamic effects of 5-methylcytosine and 5-chlorocytosine in a CpG dinucleotide sequence. *Biochemistry* 52(47):8590–8598.
44. Suen W, Spiro TG, Sowers LC, Fresco JR (1999) Identification by UV resonance Raman spectroscopy of an imino tautomer of 5-hydroxy-2′-deoxycytidine, a powerful base analog transition mutagen with a much higher unfavored tautomer frequency than that of the natural residue 2′-deoxycytidine. *Proc Natl Acad Sci USA* 96(8):4500–4505.
45. Negishi K, et al. (1997) The mechanism of mutation induction by a hydrogen bond ambivalent, bicyclic N4-oxy-2′-deoxycytidine in Escherichia coli. *Nucleic Acids Res* 25(8):1548–1552.
46. Sledziewska E, Janion C (1980) Mutagenic specificity of N4-hydroxycytidine. *Mutat Res* 70(1):11–16.
47. Nomura A, Negishi K, Hayatsu H, Kuroda Y (1987) Mutagenicity of N4-aminocytidine and its derivatives in Chinese hamster lung V79 cells. Incorporation of N4-aminocytosine into cellular DNA. *Mutat Res* 177(2):283–287.
48. Helleday T, Eshtad S, Nik-Zainal S (2014) Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 15(9):585–598.
49. Lutsenko E, Bhagwat AS (1999) Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. A model, its experimental support and implications. *Mutat Res* 437(1):11–20.
50. Münzel M, et al. (2011) Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chemistry* 17(49):13782–13788.
51. Xing X-W, et al. (2013) Mutagenic and cytotoxic properties of oxidation products of 5-methylcytosine revealed by next-generation sequencing. *PLoS One* 8(9):e72993.
52. Behjati S, et al. (2014) Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513(7518):422–425.
53. Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proc Natl Acad Sci USA* 99(2):803–808.
54. Sato Y, et al. (2006) IL-10 deficiency leads to somatic mutations in a model of IBD. *Carcinogenesis* 27(5):1068–1073.
55. Sheh A, et al. (2010) Mutagenic potency of Helicobacter pylori in the gastric mucosa of mice is determined by sex and duration of infection. *Proc Natl Acad Sci USA* 107(34):15217–15222.
56. Alexandrov LB, et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415–421.
57. Jia P, Pao W, Zhao Z (2014) Patterns and processes of somatic mutations in nine major cancers. *BMC Med Genomics* 7:11.
58. Lakritz JR, et al. (2015) Gut bacteria require neutrophils to promote mammary tumorigenesis. *Oncotarget* 6(11):9387–9396.
59. Weisenberger DJ, et al. (2006) CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* 38(7):787–793.
60. Valinluck V, Sowers LC (2007) Inflammation-mediated cytosine damage: A mechanistic link between inflammation and the epigenetic alterations in human cancers. *Cancer Res* 67(12):5583–5586.
61. Beard WA, Wilson SH (1995) Purification and domain-mapping of mammalian DNA polymerase beta. *Methods Enzymol* 262:98–107.
62. Batra VK, et al. (2006) Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* 14(4):757–766.
63. Otwinowski Z, Minor W (1997) Processsing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
64. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
65. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.