

# **Structural evolution of a DNA repair self-resistance mechanism targeting genotoxic secondary metabolites**

Elwood A. Mullins<sup>1,‡</sup>, Jonathan Dorival<sup>1,‡</sup>, Gong-Li Tang<sup>2</sup>, Dale L. Boger<sup>3</sup>, and Brandt F. Eichman<sup>1,4,\*</sup>

<sup>1</sup>Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA

<sup>2</sup>State Key Laboratory of Bio-organic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, China

<sup>3</sup>Department of Chemistry, The Scripps Research Institute, La Jolla, California, USA

<sup>4</sup>Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

<sup>‡</sup>Contributed equally to this work

\*Corresponding author: brandt.eichman@vanderbilt.edu

Microbes produce a broad spectrum of antibiotic natural products, including many DNA damaging genotoxins. Among the most potent of these are DNA alkylating agents in the spirocyclopropylcyclohexadienone (SCPCHD) family, which includes the duocarmycins, CC-1065, gilvusmycin, and yatakemycin. The yatakemycin biosynthesis cluster in *Streptomyces* sp. TP-A0356 contains an AlkD-related DNA glycosylase, YtkR2, that serves as a self-resistance mechanism against yatakemycin toxicity. We previously reported that AlkD, which is not present in an SCPCHD producer, provides only limited resistance against yatakemycin. We now show that YtkR2 and C10R5, a previously uncharacterized homolog found in the CC-1065 biosynthetic gene cluster of *Streptomyces zelensis*, confer far greater resistance against their respective SCPCHD natural products. We identify a structural basis for substrate specificity across gene clusters and show a correlation between *in vivo* resistance and *in vitro* enzymatic activity indicating that reduced product affinity—not enhanced substrate recognition—is the evolutionary outcome of selective pressure to provide self-resistance against yatakemycin and CC-1065.

## INTRODUCTION

Bacteria, fungi, and plants often produce toxic secondary metabolites as defense mechanisms against invading organisms. Streptomyces produce a staggering number and variety of natural products, including genotoxic DNA damaging antibiotics<sup>1</sup>. These agents may covalently alkylate or non-covalently intercalate the DNA nucleobases, forming mutagenic and/or toxic adducts that inhibit replication, transcription, or chromosome maturation, leading to chromosome rearrangement and instability, cell death, ageing, and disease<sup>2-7</sup>. Among the most potent genotoxic secondary metabolites are those of the spirocyclopropylcyclohexadienone (SCPCHD) family, which includes yatakemycin (YTM), gilvusmycin, CC-1065, and duocarmycin A and SA (DSA) (Fig. 1a)—all of which exhibit high cytotoxicity against tumor cells and pathogenic fungi<sup>8-16</sup>. These molecules bind in the minor groove of AT-rich regions of DNA and undergo binding-induced activation to modify N3 of adenine<sup>17</sup> (Fig. 1b). Despite forming a covalent link with only one strand of DNA, SCPCHD adducts create an extended network of non-covalent CH- $\pi$  interactions with both strands that greatly stabilizes the DNA duplex and inhibits nucleic acid metabolism<sup>18,19</sup>.

Antibiotic-producing microbes have resistance mechanisms against their own toxins that are often genetically clustered with the antibiotic synthesis operon<sup>20-22</sup>. In addition to toxin inactivation and efflux, DNA repair of the toxin-DNA adduct has been identified as a self-resistance mechanism against genotoxic natural products<sup>14,23-32</sup>. Nucleotide excision repair (NER) and base excision repair (BER) pathways serve to eliminate the majority of alkyl-DNA lesions. The NER pathway removes bulky, DNA distorting/destabilizing lesions, and is initiated by an excinuclease complex that isolates the damaged segment of DNA via dual incisions flanking the lesion<sup>33,34</sup>. By contrast, the BER pathway removes small nucleobase adducts, and is initiated by lesion-specific DNA glycosylases that hydrolyze the N-glycosidic bond to liberate the aberrant nucleobase from the DNA backbone<sup>35,36</sup>. The resulting apurinic/apyrimidinic (AP) site is incised by an AP endonuclease, generating a terminal 3'-hydroxyl group needed for polymerase-dependent synthesis of new DNA.

Owing to their exquisite helix-stabilizing properties, SCPCHD adducts are poorly repaired by the NER pathway, making these compounds highly toxic with strong antibiotic and antitumor potential<sup>37-40</sup>. To counter this toxicity, at least two of the Streptomyces that produce SCPCHD natural products encode DNA glycosylases within the corresponding biosynthetic gene clusters<sup>26,29</sup> (Fig. 1c). One of these homologs, YtkR2, has been shown to excise YTM-adenine (YTM-Ade) adducts *in vitro* and to confer self-resistance against YTM *in vivo*<sup>26</sup> (Fig. 1b). The other

homolog, C10R5, has not been characterized, but seems likely to provide similar self-resistance against CC-1065<sup>29</sup>. YtkR2 and C10R5 are homologs of *Bacillus cereus* AlkD, a unique DNA glycosylase able to excise bulky lesions by virtue of not using the base-flipping mechanism employed by other glycosylases<sup>41-45</sup>. Unlike YtkR2 and CC-1065, AlkD is present in non-antibiotic-producing species of *Bacillus* and provides only limited cellular resistance against YTM<sup>18</sup>.

Although YtkR2 has been shown to provide resistance against YTM, the corresponding resistance provided by C10R5 against CC-1065 is unknown. Moreover, the utility of the non-base-flipping mechanism employed by these enzymes to recognize SCPCHD adducts has not been explored outside YTM and AlkD. To elucidate how YtkR2 and C10R5 evolved to efficiently eliminate lesions generated by diverse SCPCHD natural products, we performed a comparative biochemical, biophysical, and structural analysis to identify differences between AlkD and these specialized homologs that are responsible for conferring enhanced cellular resistance against YTM, CC-1065, and DSA. Our data indicate that reduced product affinity, not enhanced substrate recognition, is the key factor associated with improved cellular resistance.

## RESULTS

### SCPCHD adducts stabilize the DNA duplex through a network of CH- $\pi$ interactions

The five known SCPCHD natural products can be divided into three sub-types based on the number and arrangement of the indole and pyrroloindole subunits (Fig. 1a). To better understand how these differences affect toxicity and repair, we selected one compound from each group for biochemical and biophysical characterization (Fig. 2). We previously demonstrated that a single YTM adduct in a short oligonucleotide retains a B-DNA conformation and dramatically increases the stability of the duplex, providing an explanation for why these lesions are poorly repaired by the NER pathway<sup>18,46</sup>. Similarly, earlier work showed that multiple CC-1065 adducts greatly increase the melting temperature of genomic DNA<sup>19</sup>. To directly compare the effects of the different SCPCHD sub-types on DNA stability and structure, we introduced a single YTM, DSA, or CC-1065 adduct onto a central adenine within the same GC-rich DNA dodecamer. The circular dichroism spectra of each construct showed negative and positive bands at ~240 and 270 nm, respectively, indicative of B-form DNA<sup>47</sup> (Fig. 2a). Additional features in the spectra likely result from either minor structural perturbation of the DNA or dichroism of the SCPCHD adducts themselves, which produce broad peaks above 300 nm in the absorbance spectra (Fig. 2b). To assess the effect of each compound on duplex stability, thermal denaturation was monitored by an increase in absorbance at 260 nm (Fig. 2c). Relative to unmodified DNA, the presence of a

single DSA, YTM, or CC-1065 adduct increased the melting temperature by 23, 36, or  $\geq 41^{\circ}\text{C}$ , respectively. The increase in duplex stability correlated with the stability of the adduct. After 28 days at  $30^{\circ}\text{C}$ , only 16%, 11%, and 8%, respectively, of DSA, YTM, and CC-1065 adducts had spontaneously depurinated (Fig. 2d and Supplementary Fig. 1). For comparison, the half-life of 3-methyl-2'-deoxyadenosine in genomic DNA is less than four days under similar conditions (interpolated from previously published rates at  $22$  and  $39^{\circ}\text{C}$ )<sup>48</sup>.

To elucidate the structural and chemical origins of these trends, we modeled each of the constructs by manually docking ideal B-form DNA and SCPCHD adducts taken from crystallographic AlkD product complexes (described below) (Fig. 2e). Despite being maintained as rigid bodies, all adducts fit in the minor groove without steric clashes. In these hypothetical models, the hydrophilic edges of the subunits are exposed to solvent, while the hydrophobic faces are sandwiched between the backbones of the two DNA strands. As we previously described with YTM adducts<sup>18</sup>, this arrangement creates an extended network of CH- $\pi$  interactions. While these interactions are individually often weak ( $\sim 1$  kcal/mol)<sup>49</sup>, the combined strength of many such interactions generates a strong energy barrier to duplex denaturation and lesion depurination. Indeed, DNA stability and lesion half-life correlate with the number of nucleotides that form CH- $\pi$  interactions (Fig. 2f). Thus, this network of weak interactions explains the trends in duplex stability and nucleotide half-life that we observe with each adduct. Importantly, our models and conclusions are consistent with a previously published NMR structure of DSA-DNA (PDB accession 1DSA [<http://doi.org/10.2210/pdb1DSA/pdb>])<sup>50</sup> (Supplementary Fig. 2), which further highlights the importance of CH- $\pi$  interactions on the stability of SCPCHD-modified DNA duplexes. Relative to ideal B-form DNA, the DSA adduct in the experimental structure induces a slight narrowing of the minor groove and twisting of the DNA backbone, which maximizes the number and strength of the CH- $\pi$  interactions formed between the adduct and the DNA. Thus, through a network of CH- $\pi$  interactions, SCPCHD adducts create an induced fit in the minor groove that tethers the opposing DNA strands and prevents gross deformation to the B-DNA conformation, explaining how these adducts pose a significant challenge to repair by the UvrABC NER excinuclease<sup>18</sup>.

### **YtkR2 and C10R5 evolved to provide self-resistance against YTM and CC-1065**

Historically, bulky DNA lesions, such as SCPCHD adducts, have been regarded as substrates for the NER pathway<sup>51-54</sup>. However, we previously showed that NER provides *Bacillus anthracis* with only limited resistance against YTM, while AlkD-mediated BER provides a modest degree of additional resistance<sup>18</sup>. At the time, we speculated the AlkD homologs that evolved in

antibiotic-producing bacteria would provide greater resistance against their SCPCHD natural products. To test this theory, we transformed *Escherichia coli* with either an empty plasmid or a plasmid encoding AlkD, YtkR2, or C10R5, and incubated the cultures with various concentrations of YTM, DSA, or CC-1065 (Supplementary Fig. 3). Sensitivity to the compounds was scored as the lag time, or delay prior to exponential cell growth<sup>55,56</sup> (Fig. 3). In the absence of any drug, cell growth was unchanged regardless of the protein being expressed. Conversely, cells transformed with only an empty vector showed different degrees of resistance against the three SCPCHD compounds. As previously observed with eukaryotic cells<sup>13</sup>, YTM was the most toxic, inhibiting growth at the lowest drug concentration, followed by CC-1065 and then DSA. In cultures containing YTM, only YtkR2 provided clear and significant resistance, shortening the lag phase before exponential growth. Similarly, in cultures containing CC-1065, C10R5 provided the greatest resistance, although YtkR2, and to a much lesser extent AlkD, also provided additional resistance. The same trend was observed in cultures containing DSA; C10R5 provided the greatest resistance, followed by YtkR2 and then AlkD. These data are consistent with the notion that YtkR2 and C10R5 evolved to provide self-resistance against YTM and CC-1065, respectively. However, both proteins also confer substantial resistance against DSA, and YtkR2 provides moderate resistance against CC-1065. This lack of specificity suggests that the features of YtkR2 and C10R5 responsible for drug resistance are largely independent of the differences between the SCPCHD compounds. The only exception to this generalization appears to be the lack of resistance against YTM provided by C10R5.

### **Product affinity inversely correlates with cellular resistance**

To understand how YtkR2 and C10R5 confer greater resistance than AlkD to SCPCHD natural products, we measured *in vitro* excision activity using purified proteins and defined oligonucleotide substrates (Fig. 4a,b and Supplementary Fig. 4). Under single-turnover conditions, in which the enzymes are in molar excess of the DNA substrates, both AlkD and YtkR2 rapidly excised all three adducts, approaching the endpoints of the reactions within 15 s (Fig. 4c, Supplementary Fig. 5, and Supplementary Table 1). C10R5 rapidly excised DSA and CC-1065 lesions, albeit more slowly than AlkD and YtkR2, but removed YTM adducts at a substantially reduced rate. Under multiple-turnover conditions, in which the DNA substrates are in molar excess of the enzymes, the results were strikingly different (Fig. 4d, Supplementary Fig. 6, and Supplementary Table 1). While AlkD rapidly excised a molar equivalent of each adduct (i.e. the first turnover in which ~10% product is formed), subsequent turnovers occurred several orders of magnitude more slowly. Excision of adducts by YtkR2 was similarly slowed, although to a lesser

extent. Surprisingly, the multiple-turnover rates of removal of DSA and CC-1065 lesions by C10R5 were effectively unchanged; only the rate of excision of YTM adducts was substantially reduced. Thus, under single-turnover conditions, all homologs have the capacity for rapid excision of diverse SCPCHD lesions, although this capacity appears to be reduced in C10R5, and inefficient substrate recognition by C10R5 seemingly further slows removal of YTM lesions. For all homologs, there is no correlation between rates of single-turnover excision and cellular resistance (Figs. 3 and 4c and Supplementary Table 1). Conversely, the multiple-turnover rates of the three enzymes are strongly correlated with their cellular resistance against each SCPCHD compound (Figs. 3 and 4d and Supplementary Table 1).

We previously demonstrated that slow multiple turnover by AlkD was correlated with inhibition of AP endonuclease activity *in vitro*, which we attributed to tight binding of the AP product generated upon base excision<sup>18,57</sup>. The substantially faster multiple turnover that we observed with YtkR2 and C10R5 therefore suggests that product affinity is weaker for these homologs. To test this possibility, we performed AP-DNA incision assays using the AP endonuclease EndoIV and either AP-DNA generated from each of the SCPCHD adducts or tetrahydrofuran (THF)-DNA containing an AP analog (Fig. 4e and Supplementary Fig. 7). Reaction mixtures with THF-DNA were supplemented with free 3-methyladenine (3mAde) nucleobase to approximate the nucleobase adducts excised from the SCPCHD lesions, while also providing a basis to assess the effects of the excised SCPCHD adducts on product affinity. Relative to AP-DNA produced by AlkD, AP-DNA generated by YtkR2 or C10R5 was incised by EndoIV to a substantially greater extent, regardless of which SCPCHD adduct was excised to produce the abasic site, consistent with relatively weak product binding by YtkR2 and C10R5. For each of the homologs, AP-DNA generated from DSA-DNA was cleaved to the greatest extent, which could be explained by DSA-Ade being bound by each of the homologs with weaker affinity, producing a less stable ternary product complex, and/or by DSA-Ade inhibiting AP-DNA incision by EndoIV to a lesser degree. The former would be consistent with the fewer observed or predicted protein-adduct binding interactions formed with DSA-Ade (described below), while the latter would be consistent with the fewer predicted CH- $\pi$  interactions formed with DSA (Fig. 2e). Unlike AP-DNA, the extent to which THF-DNA was incised by EndoIV was independent of which DNA glycosylase homolog was present, and no additional THF-DNA was incised in reactions mixtures in which AlkD, YtkR2, and C10R5 were omitted (Fig. 4e and Supplementary Fig. 7), indicating product affinity is greater for abasic DNA in the presence of the excised SCPCHD adducts. Taken together, the base excision and AP-DNA incision data suggest that YtkR2 and C10R5 evolved to possess lower product affinities than AlkD for abasic DNA derived from SCPCHD lesions. Moreover, this apparently

reduced product affinity correlates with the increased cellular resistance provided by these specialized homologs (Figs. 3 and 4d,e).

### **Diverse SCPCHD adducts are recognized primarily through non-specific interactions**

To understand AlkD's lack of specificity for the three SCPCHD sub-types and comparatively high product affinity, we determined crystal structures of AlkD bound to DNA containing each of the three adducts. We previously reported the structure of AlkD bound to AP-DNA and YTM-Ade (PDB accession 5UUG [<http://doi.org/10.2210/pdb5UUG/pdb>])<sup>18</sup>, and now report structures of AlkD bound to AP-DNA and either DSA-Ade or CC1065-Ade (Fig. 5 and Supplementary Table 2). Product complexes were generated by incubating AlkD with the corresponding DNA substrates, allowing for enzymatic hydrolysis of the N-glycosidic bond and formation of an AP site and a free nucleobase adduct (Fig. 5a). The conformations of AlkD and the abasic DNA are effectively identical in the three structures (Fig. 5b and Supplementary Fig. 8). Relative to the substrate models (Fig. 2e), binding by AlkD bends the B-DNA helix by ~30° and widens the minor groove by 4–5 Å, disrupting the CH- $\pi$  interactions between the adducts and the modified strand and enabling the three catalytic residues—Trp109, Asp113, and Trp187—to contact the modified nucleotide<sup>18,43,58</sup>. Despite this remodeling, the excised nucleobases remain stacked in the DNA duplex, with the bulky SCPCHD adducts bound within an extended cleft between the concave surface of the protein and the minor groove of the DNA.

Within the SCPCHD binding cleft, 14 protein residues contact one or more of the adducts (Fig. 5c). Only three—Gln38, Lys156, and Lys194—form adduct-specific hydrogen-bonding interactions. Gln38 interacts with the hydroxyl and methoxy substituents of the S<sup>-1</sup> subunit unique to YTM, while Lys194 interacts with the methoxy substituent of the S<sup>+2</sup> subunit unique to CC-1065. Otherwise, Gln38 and Lys194 function as DNA binding residues. Lys156 contacts all three adducts but only forms a hydrogen-bonding interaction with the hydroxyl substituent specific to the S<sup>+1</sup> subunit of CC-1065. In YTM and DSA, this position is occupied by a methoxy group, forcing Lys156 to adopt a different rotamer that hydrogen bonds with the DNA backbone. All other hydrogen bonds between the adducts and the protein are mediated by water. Additional recognition of the adducts occurs through hydrophobic contacts and CH- $\pi$  interactions. The majority of these binding interactions involve the S<sup>0</sup> and S<sup>+1</sup> subunits present in all of the SCPCHD adducts. Of seemingly principal importance are Tyr27, Met28, Trp109, Asp110, Leu155, and Trp187. Together these residues form a shelf below the adducts and provide an alternative network of CH- $\pi$  interactions, replacing the network of CH- $\pi$  interactions that is disrupted upon binding of the DNA by AlkD and widening of the minor groove. Thus, AlkD's substrate binding



cleft can accommodate each SCPCHD compound with strikingly few adduct-specific interactions. This lack of specificity is consistent with AlkD's similar rates of excision for the three substrates.

To explain the structural basis for cellular resistance against SCPCHD compounds provided by YtkR2 and C10R5, the AlkD product complexes were used as templates to create homology models of both homologs bound to abasic DNA and each of the nucleobase adducts (Supplementary Fig. 9). Despite the low sequence identity (19–25%) among the three proteins, the only putative structural difference of note is an extended loop between helices  $\alpha$ I and  $\alpha$ J in both YtkR2 and C10R5. The three homologs appear to share the same three catalytic residues and a high conservation among DNA binding residues, although YtkR2 and C10R5 possess fewer apparent DNA binding residues (15 and 14, respectively) than the 18 present in AlkD. Within the substrate binding cleft, the largest variation predicted by the homology models is in the residues that contact the SCPCHD adducts. However, despite this variation, YtkR2 and C10R5 seemingly form few specific interactions with any of the adducts, including the adduct that each evolved to excise. The only putative hydrogen-bonding interactions between YtkR2 and YTM are formed between Arg42 and the hydroxyl and methoxy substituents of the  $S^{-1}$  subunit. Similarly, the only likely hydrogen-bonding interactions between C10R5 and CC-1065 are formed between Lys159, Arg205, and the hydroxyl and methoxy substituents of the  $S^{+1}$  and  $S^{+2}$  subunits. Surprisingly, all of the adduct binding residues in YtkR2 and C10R5 that appear to be specific for YTM and CC-1065 are equivalent to the three DNA binding residues in AlkD that form similar interactions with YTM-Ade and CC1065-Ade (Fig. 5c and Supplementary Fig. 9). As with AlkD, the remainder of the binding interactions are composed of water-mediated hydrogen bonds, hydrophobic contacts, and CH- $\pi$  interactions. This apparent lack of structural specificity, combined with the nearly complete lack of specificity observed in the excision assays, suggests the enhanced cellular resistance provided by YtkR2 and C10R5 does not result from enhanced recognition of a particular SCPCHD adduct. Instead, the additional cellular resistance provided by YtkR2 and C10R5 seems more likely to result from decreased product affinity, which may be explained, at least in part, by the reduced number of predicted DNA binding residues in these homologs (Supplementary Fig. 9). It is also possible that other, more subtle, structural factors that are not apparent from the homology models may play a role.

### **C10R5 poorly accommodates the unique arrangement of subunits in YTM**

The only significant indication of substrate specificity by any of the three homologs was C10R5's relatively slow excision of YTM-Ade adducts *in vitro* and corresponding lack of YTM resistance in cells. Given the structural differences between the SCPCHD natural products, we

postulated that unfavorable interactions between C10R5 and the S<sup>-1</sup> subunit unique to YTM were likely responsible for this reduced activity. Comparison of the YtkR2 and C10R5 homology models indicated three residues (Phe22, Ser25, and Ala40) in C10R5 that were likely to contact the S<sup>-1</sup> subunit and that were not conserved in either YtkR2 or AlkD (Fig. 6a–c and Supplementary Fig. 9). To determine the influence of these residues on substrate specificity, we mutated them to the corresponding residues in YtkR2 and measured YTM-Ade excision under modified single- and multiple-turnover conditions (Fig. 6d,e, Supplementary Figs. 10 and 11, and Supplementary Table 3). Under single-turnover conditions, both the F22T and S25G single mutants exhibited a modest increase in YTM-Ade excision activity relative to wild-type C10R5, and this effect was further enhanced in the F22T-S25G double mutant. Moreover, the increased activity of the C10R5 F22T-S25G double mutant toward YTM-Ade was within five-fold of wild-type C10R5's activity toward CC1065-Ade (Supplementary Tables 1 and 3). In contrast, an A40R single mutant and an F22T-S25G-A40R triple mutant displayed decreased YTM-Ade excision activity relative to the wild-type enzyme and the F22T-S25G double mutant, respectively. Under multiple-turnover conditions, all C10R5 single mutants, including A40R, exhibited increased YTM-Ade excision activity, which was further increased in the F22T-S25G double mutant. As observed under single-turnover conditions, the F22T-S25G-A40R triple mutant displayed reduced YTM-Ade excision activity relative to the double mutant. Despite this reduction, the YTM-Ade excision activity of the triple mutant was comparable to that of YtkR2 under multiple-turnover conditions, while the YTM-Ade excision activities of the F22T single mutant and the F22T-S25G double mutant were slightly greater than that of YtkR2 (Supplementary Tables 1 and 3). These comparatively rapid multiple-turnover rates, however, are likely more indicative of C10R5's apparently low product affinity than vastly improved YTM recognition, as the rates at which these mutants removed YTM-Ade is more than 11-fold slower than the rates at which wild-type C10R5 removed CC1065-Ade and DSA-Ade. Moreover, under single-turnover conditions, the YTM-Ade excision activity of YtkR2 vastly exceeded that of all C10R5 mutants. Thus, while the substitutions (F22T and S25G) that increased the volume of the adduct binding cleft also increased the rate of YTM-Ade excision, suggesting that C10R5's low activity for YTM-Ade is dictated by a sterically constrained binding cleft at the position of the S<sup>-1</sup> subunit, additional unfavorable interactions may also be hindering recognition of YTM lesions (Supplementary Figs. 8 and 9), preventing more rapid excision of YTM-Ade adducts.

To determine the biological relevance of the gain-of-activity mutations, we transformed *E. coli* with either an empty plasmid or a plasmid encoding YtkR2, C10R5, C10R5 F22T S25G, or C10R5 F22T S25G A40R, and incubated cells on solid medium containing various concentrations

of YTM (Fig. 6f and Supplementary Fig. 12). In the absence of YTM, cell growth was unaffected by the protein construct being expressed. In the presence of YTM, YtkR2 provided the greatest resistance, while cells containing an empty plasmid were the most sensitive. As in our growth experiments in liquid medium (Fig. 3 and Supplementary Fig. 3), wild-type C10R5 provided little to no resistance to YTM (Fig. 6f and Supplementary Fig. 12). The C10R5 double mutant, however, provided clear resistance, albeit less than YtkR2. Consistent with our YTM-Ade excision assays (Fig. 6d,e), the C10R5 triple mutant provided less resistance than the double mutant, and only arguably provided more resistance than wild-type C10R5 (Fig. 6f and Supplementary Fig. 12). These relatively modest increases in YTM resistance may be limited by a number of factors, including potentially reduced expression and/or solubility of the mutant proteins. Regardless, these cellular results validate our structural and biochemical conclusions, showing homology-based mutations that likely only eliminate steric clashes between C10R5 and YTM—without introducing specific adduct binding interactions—are able to increase the YTM resistance conferred by C10R5.

## DISCUSSION

In our experiments, YtkR2 and C10R5 provided the greatest cellular resistance to YTM and CC-1065 toxicity, respectively, consistent with evolution of these proteins within the context of the biosynthetic gene cluster for each natural product. Both enzymes also conferred substantial resistance against DSA, even though DSA differs from YTM and CC-1065 in the substituents present on the  $S^{+1}$  subunit. YtkR2 also provided significant resistance against CC-1065, which was somewhat unexpected because of the presence of the unique  $S^{+2}$  subunit, and suggests that interactions with this subunit do not play an important role in determining substrate specificity. Indeed, the outward face of the  $S^{+2}$  subunit is largely solvent exposed and has the fewest protein contacts of any subunit. In contrast, the  $S^{-1}$  subunit unique to YTM makes several direct protein contacts, which we show largely account for the weak YTM-Ade excision activity of C10R5. Interestingly, the  $S^{-1}$  binding surface is formed predominantly from helix  $\alpha B$ , which is a critical DNA binding element within the AlkD family of enzymes<sup>57</sup> and whose structural variability among HLR glycosylases enables binding to a diverse array of ligands<sup>44,59</sup>. However, steric accommodation of the  $S^{-1}$  subunit is the only determinant of substrate specificity that we were able to identify through our comparative analysis. Within the largely water-filled adduct binding cleft, most hydrogen bonds are mediated by a network of water molecules, which creates a seemingly malleable hydrogen-bonding interface able to accommodate different substituents on

the S<sup>0</sup> and S<sup>+1</sup> subunits. Similarly, most of the protein residues that directly contact the adducts form a network of CH- $\pi$  interactions that is only selective for aromatic groups, which are present in all of the indole and pyrroloindole subunits in all of the SCPCHD adducts. This mode of substrate recognition seems largely incompatible with a high degree of substrate specificity, which is consistent with our experimental observations.

Despite selective pressure for YtkR2 and C10R5 to provide self-resistance against YTM and CC-1065, neither homolog evolved to remove YTM-Ade or CC1065-Ade adducts at a substantially increased rate, as observed by the strong correlation between multiple-turnover excision and cellular resistance. Instead, selective pressure seemingly induced both enzymes to develop reduced product affinity. We previously showed that AlkD's low level of protection against YTM results from product inhibition in which the enzyme remains bound to the newly formed abasic site in the DNA, slowing multiple-turnover excision, as well as impeding subsequent repair steps in the BER pathway<sup>18</sup>. Both YtkR2 and C10R5 have evolved to circumvent this inhibition, while also minimizing their impediment to later repair steps. Reduced product inhibition may be explained by the fewer DNA binding residues predicted in these proteins, which are also the only contacts predicted to form "specific" hydrogen bonds with the adducts. It is interesting to speculate that the evolution of YtkR2 and C10R5 to enable multiple turnover occurred as a means to handle the high concentration of YTM or CC-1065 in the producing strains, whereas the rapid single-turnover activity of AlkD is sufficient to provide resistance to the presumably much lower intracellular quantities of these compounds in *Bacillus* and other non-producing bacteria.

Until recently, all DNA glycosylases were believed to recognize and excise substrates using a base-flipping mechanism in which the target adduct is sequestered from the DNA helix and into a nucleobase binding pocket on the surface of the enzyme. The AlkD/YtkR2/C10R5 enzymes belong to one of two known bacterial DNA glycosylase superfamilies that have evolved a non-base-flipping mechanism as a means to remove structurally diverse duplex-stabilizing adducts<sup>18,43,45</sup>. While their SCPCHD substrates form a covalent bond with only one DNA strand, a network of CH- $\pi$  interactions non-covalently tethers the two DNA strands together. This greatly stabilizes the duplex, while also hindering the base-flipping activity necessary for recognition and excision by other repair proteins. An unrelated DNA glycosylase, AlkZ, from *Streptomyces sahachiroi* was recently found from its association with the biosynthetic gene cluster for azinomycin B, a highly toxic interstrand crosslinking agent, and shown to provide self-resistance through unhooking of the corresponding DNA crosslinks<sup>27,60</sup>. Enzymes in the AlkZ superfamily have since been shown to remove a variety of alkylated DNA adducts and interstrand crosslinks<sup>27,61,62</sup>. Despite their different polypeptide folds, AlkZ and AlkD family enzymes adopt a

similar C-shaped DNA binding architecture that enables access to the target N-glycosidic bond while the aberrant nucleotide remains stacked in the DNA. This mode of recognition and excision not only eliminates the problem of flipping a duplex-stabilizing lesion out of the DNA, but also avoids the steric limitations imposed by a nucleobase binding pocket. Thus, the expanded substrate spectrum of non-base-flipping DNA glycosylases broadens the biological utility of the BER pathway, providing a mechanism for resistance against the genotoxic natural products generated by many bacteria.

## METHODS

**Protein purification.** *Bacillus cereus* AlkD and *Streptomyces* sp. TP-A0356 YtkR2 were purified as previously described<sup>43,63</sup>. The gene encoding *Streptomyces zelensis* C10R5 (GenBank accession KY379149 [<https://www.ncbi.nlm.nih.gov/nucore/KY379149>]) was synthesized by Genscript and ligated into a modified pET27 expression vector encoding a Rhinovirus 3C-cleavable hexahistidine-SUMO fusion tag. Overexpression of C10R5 was performed in *Escherichia coli* BL21(DE3) cells grown in LB medium supplemented with 30 mg/L kanamycin. Upon reaching mid-log phase, cultures were cooled from 37 to 18°C, and IPTG and glucose were added to final concentrations of 0.5 mM and 0.4 g/L, respectively, to initiate slow protein production. After overnight incubation at 18°C, cells were harvested by centrifugation, resuspended in lysis buffer [50 mM sodium phosphate pH 6.8, 250 mM NaCl, 5% (v/v) glycerol, and 5% (w/v) mannitol], and lysed by passage through an Avestin Emulsifier C3 homogenizer operating at ~15,000 psi. Lysate was cleared by centrifugation and then injected onto a Ni-NTA column. The column was washed with lysis buffer supplemented with 40 mM imidazole, and the protein was then eluted using the same buffer but containing 300 mM imidazole. Fractions were pooled and the hexahistidine-SUMO tag was removed by overnight cleavage at 4°C. C10R5 was diluted three-fold in buffer H [25 mM sodium phosphate pH 6.8, 10% (v/v) glycerol, and 1 mM DTT] before being applied to a heparin Sepharose column equilibrated in buffer H supplemented with 100 mM NaCl. The protein was eluted by linearly increasing the concentration of NaCl to 1 M. Pooled fractions were reappplied to the Ni-NTA column, and C10R5 was eluted by washing the column with lysis buffer containing 50 mM imidazole. The protein was concentrated by diafiltration, injected onto a Superdex 200 column equilibrated in buffer S [20 mM MOPS pH 6.8, 300 mM NaCl, and 5% (v/v) glycerol], and eluted in the same buffer. Fractions containing C10R5 were pooled and concentrated by diafiltration. Aliquots were frozen in liquid nitrogen and stored at -80°C. C10R5 mutants were generated using the Q5 site-directed mutagenesis kit (New

England Biolabs) and purified in the same manner as the wild-type protein. *E. coli* EndoIV was purchased from New England Biolabs.

**Adduct preparation.** Oligodeoxynucleotides were purchased from Integrated DNA Technologies and used without further purification. YTM and CC-1065 were purified and DSA was synthesized as previously described<sup>14,29,64</sup>. DNA adducts were generated as previously described<sup>18</sup>, except reaction mixtures contained 10 mM MES pH 6.5, 40 mM NaCl, 10% (v/v) DMSO, 10  $\mu$ M DNA, and 150  $\mu$ M YTM or 100  $\mu$ M DSA or CC-1065. After 24 hr at 22°C, unreacted DSA, CC-1065, and YTM were removed by passing the reaction mixtures through a Sephadex G25 column equilibrated in annealing buffer (10 mM MES pH 6.5 and 40 mM NaCl). DNA intended for crystallographic or biophysical experiments was subsequently concentrated *in vacuo* and then exchanged into annealing buffer or melting buffer (10 mM sodium phosphate pH 7.0 and 10 mM NaCl) by diafiltration (3,000 NMWL).

**Adduct characterization.** Absorbance and circular dichroism measurements were collected from modified oligodeoxynucleotide duplexes [d(CCCCAAAGCCCG)/d(CGGGCTTTGGG)]; the underline denotes the modified nucleotide] as previously described<sup>18</sup>. Thermal melting profiles were also collected as previously described<sup>18</sup>. Melting temperatures were calculated by fitting the data to a polynomial function and determining the temperature at which the second-order derivative was equal to zero.

**AlkD/AP-DNA/DSA-Ade and AlkD/AP-DNA/CC1065-Ade crystallization.** Ternary product complexes were prepared by mixing equal volumes of 0.45 mM AlkD and 0.54 mM DSA-DNA or CC1065-DNA [d(AGCAAAGGC)/d(TGCCTTTGC); the underline denotes the modified nucleotide] and incubating the mixtures at 4°C for 30 min. Complexes were crystallized using the sitting-drop vapor-diffusion method. Drops containing DSA-DNA were prepared from 1.5  $\mu$ L of protein-DNA solution [0.22 mM AlkD and 0.27 mM DNA], 0.75  $\mu$ L of reservoir solution [24% (w/v) PEG 8,000, 50 mM HEPES pH 7.0, and 50 mM CaCl<sub>2</sub>], 0.75  $\mu$ L of seed solution [microscopic crystals of AlkD and DNA containing an A•C mismatch], and 0.75  $\mu$ L of additive solution [5% (w/v) benzamidine•HCl]. Drops containing CC1065-DNA were prepared from 1.5  $\mu$ L of protein-DNA solution [0.22 mM AlkD and 0.54 mM DNA], 1.5  $\mu$ L of reservoir solution [21% (w/v) PEG 8,000, 50 mM HEPES pH 7.0, and 50 mM CaCl<sub>2</sub>], and 0.75  $\mu$ L of additive solution [5% (w/v) benzamidine•HCl]. All drops were equilibrated at 21°C against 500  $\mu$ L of reservoir solution. After several days, crystals were harvested, briefly soaked in reservoir solution supplemented with 15% (v/v) glycerol, and flash-cooled in liquid nitrogen.

**X-ray data collection and refinement.** X-ray diffraction data were collected on beamlines 21-ID-F ( $\lambda = 0.97872$  Å) and 21-ID-G ( $\lambda = 0.97857$  Å) at the Advanced Photon Source. Each dataset was collected from a single cryopreserved crystal at  $-173^{\circ}\text{C}$  and processed using HKL2000<sup>65</sup>. Data collection statistics are provided in Supplementary Table 2. Initial phases were determined by molecular replacement, using a model of AlkD (PDB accession 3BVS [<http://doi.org/10.2210/pdb3BVS/pdb>])<sup>63</sup> placed with Phaser<sup>66</sup>. DNA was then manually built in Coot<sup>67</sup>. The entireties of the oligodeoxynucleotide duplexes were readily apparent in both complexes, as were the excised nucleobases. AlkD residues 1–225 were also clearly defined in both complexes. However, the last 12 residues (226–237) at the C-terminus could not be reliably modeled. Atomic coordinates, anisotropic temperature factors (TLS), and fractional occupancies were refined for all non-hydrogen atoms using PHENIX<sup>68</sup>. Hydrogen atoms were placed in riding positions and were not refined against the X-ray data. The final models were validated using MolProbity<sup>69</sup> and contained 96.9% of residues in the favored regions of the Ramachandran plot, 3.1% of residues in the allowed regions, and no residues in the disallowed regions. Additional refinement and validation statistics are included in Supplementary Table 2.  $mF_o - DF_c$  omit maps were generated with PHENIX after removing the AP site and the modified nucleobase and performing simulated annealing on the remaining AlkD/DNA complex. Figures were prepared in PyMOL (<https://www.pymol.org>).

**Homology modeling.** AlkD templates were extracted from each of the three crystallographic SCPCHD product complexes. For each of the homologs, a single homology model was generated from each of the three templates using SWISS-MODEL<sup>70</sup>. Global quality estimates (QMEANDisCo global) for all models reached or exceeded 0.60 (0–1 scale). Local quality estimates (QMEANDisCo local) were highest for residues constituting helices in the substrate binding cleft and lowest for residues constituting the extended loop between helices  $\alpha\text{I}$  and  $\alpha\text{J}$ . Relative to AlkD, the root-mean-square deviation of all  $\text{C}\alpha$  positions (excluding those in the loop between helices  $\alpha\text{I}$  and  $\alpha\text{J}$ ) was equal to or less than 0.3 and 0.8 Å for YtkR2 and C10R5, respectively. Hypothetical product complexes were generated from the three crystallographic product complexes by replacing AlkD with the corresponding YtkR2 or C10R5 homology models. As necessary, alternate rotamers were manually selected in Coot to eliminate steric clashes or to create favorable binding interactions.

**Adduct modeling.** DSA, CC-1065, and YTM from the AlkD product complexes and ideal B-form DNA [d(CCCCAAAGCCCG)/d(CGCGCTTTGGGG)] generated with Coot were manually docked. Both the DNA and the adducts were maintained as rigid bodies. Placement was dictated by the

geometric requirements of a covalent bond between the modification and the deoxyadenosine nucleotide and the avoidance of steric clashes with the remainder of the DNA duplex.

**Adduct excision.** Base excision was monitored by alkaline cleavage of the abasic products produced from alkylated and fluorescein (FAM)-labeled DNA [FAM-d(CGGGCGGCGGCAAAGGGCGCGGGCC)/d(GGCCC GCGCCCTTTGCCGCGCCCG)]; the underline denotes the modified nucleotides]. Reaction mixtures were prepared with 100 nM DNA, 20 mM Tris•HCl pH 8.0, 100 mM NaCl, 0.1 mg/mL BSA, and 1 mM DTT. To monitor excision under multiple- or single-turnover conditions, either 10 nM or 1  $\mu$ M, respectively, of AlkD, C10R5, or YtkR2 was added. For single-turnover base excision by C10R5 mutants, which are less stable than the wild-type enzyme, reaction mixtures contained 100 nM DNA, 20 mM MOPS pH 6.8, 150 mM NaCl, and 5% (v/v) glycerol. Similarly, multiple-turnover excision by C10R5 mutants was performed in reaction mixtures containing 100 nM DNA, 20 mM MOPS pH 6.8, 200 mM NaCl, 5% (v/v) glycerol, and 0.1 mg/mL BSA. All reactions were incubated at 30°C and quenched at different times by mixing 8- $\mu$ L aliquots with 2  $\mu$ L of 1 M NaOH and heating at 70°C for 20 min. An equal volume of loading buffer [80% (w/v) formamide, 10 mM EDTA pH 8.0, 1 mg/mL bromophenol blue, and 1 mg/mL xylene cyanol] was then added before the samples were heated at 70°C for an additional 2 min. The cleaved product was separated from the full-length substrate by denaturing urea-PAGE as previously described<sup>43</sup>, and the fluorescence intensity of the FAM-labeled DNA was measured on a Typhoon Trio variable mode imager (GE Healthcare). Rate constants were calculated by fitting the data to a single-exponential model. For multiple-turnover experiments with AlkD, the rapid first turnover that occurred during the initial 15 s of the reaction was omitted from the fit. Experiments were performed in triplicate.

**AP-DNA incision.** AP-DNA substrate [FAM-d(CGGGCGGCGGCAAXGGGCGCGGGCC)/d(GGCCC GCGCCCTTTGCCGCGCCCG)]; X = A or AP] was generated by incubating 500 nM YTM-DNA, DSA-DNA, or CC1065-DNA with 1  $\mu$ M AlkD, YtkR2, or C10R5 in 20 mM MOPS pH 6.8, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, 5% (v/v) glycerol, 0.1 mg/mL BSA, and 1 mM DTT at 30°C for 3 hr. THF-DNA substrate [FAM-d(CGGGCGGCGGCAXAGGCGCGGGCC)/d(GGCCC GCGCCCTTTGCCGCGCCCG)]; X = THF] was prepared by mixing THF-DNA and 3mAde nucleobase in an equal molar ratio and incubating with or without AlkD, YtkR2, or C10R5. AP-DNA incision reactions were initiated with aliquots from the base excision reaction mixtures and contained 0 or 2 nM EndoIV; 100 nM AP-DNA or THF-DNA; 200 nM AlkD, YtkR2, or C10R5; 20 mM MOPS pH 6.8; 150 mM NaCl; 10 mM MgCl<sub>2</sub>; 5% (v/v) glycerol; 0.1 mg/mL BSA; and 1 mM DTT. Reaction mixtures were incubated at



30°C for 1 hr. Aliquots were then removed and quenched by adding an equal volume of loading buffer supplemented with 0.1 M EDTA pH 8.0 and incubating at 70°C for 5 min. Alternatively, to determine the total fraction of AP-DNA present in the reaction mixtures, aliquots were quenched by adding 1 M NaOH to a final concentration of 0.2 M and heating at 70°C for 20 min before adding an equal volume of loading buffer and heating at 70°C for an additional 2 min. Samples were analyzed as described above. Experiments were performed in triplicate.

**Drug resistance.** *E. coli* BL21(DE3) cells were transformed with a pET27-derived plasmid encoding a fusion protein between a hexahistidine-SUMO tag and either AlkD, C10R5, or YtkR2, or encoding only the tag as a control. For growth experiments in liquid medium, precultures were grown overnight in LB medium supplemented with 0.1 mM IPTG and then diluted 100-fold in 200  $\mu$ L of LB medium without IPTG but containing different concentrations of YTM, DSA, or CC-1065. Cultures were incubated in 96-well, flat-bottom plates with shaking for 18 hr using a Synergy 2 multi-detector microplate reader (BioTek). Cell density was measured at 600 nm ( $OD_{600}$ ). Experiments were performed in quadruplicate. For growth experiments on solid medium, precultures were grown overnight in LB medium supplemented with 0.5 mM IPTG. Aliquots were then diluted to 0.1  $OD_{600}$  in LB medium containing 0.1 mM IPTG and grown to an  $OD_{600}$  of 1.0 to ensure that cells were in the exponential phase. New aliquots were diluted 10–1,000-fold in LB medium without IPTG and immediately spotted on LB agar plates supplemented with various concentrations of YTM. Plates were incubated for 18 hr before being imaged. Experiments were performed in duplicate. At all stages of both growth experiments, cells were grown at 37°C and LB medium was supplemented with 30 mg/L kanamycin.

**Immunoblotting.** Portions of the precultures used for the drug resistance assays in liquid medium were diluted to an  $OD_{600}$  of 0.8 in LB medium and immediately mixed with an equal volume of SDS loading buffer [100 mM Tris•HCl pH 6.8, 4% (w/v) SDS, 4 M urea, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, and 200 mM DTT]. Aliquots (10  $\mu$ L) corresponding to half of the quantity of cells added to inoculate the drug resistance cultures were then loaded on a NuPAGE 4–12% Bis-Tris SDS-acrylamide gel (Invitrogen). Proteins in the lysate were separated by electrophoresis at 50 V for 30 min and then 100 V for 2 hr before being transferred to a Trans-Blot Turbo Mini 0.2  $\mu$ m nitrocellulose membrane (Bio-Rad) using a Trans-Blot Turbo transfer system (Bio-Rad). The membrane was subsequently blocked overnight with 5% (w/v) BSA in TBS buffer (20 mM Tris•HCl pH 7.6 and 150 mM NaCl) supplemented with 0.1% (v/v) Triton X-100 (TBST buffer). Excess BSA was removed by washing twice for 5 min each in TBST buffer and again for 5 min in TBS buffer. Hexahistidine-tagged protein was detected by probing the membrane at room temperature for 2

hr with a monoclonal mouse anti-hexahistidine IgG1 antibody (diluted 1:5,000; Abgent). The membrane was then washed again as described above to remove non-specifically bound primary antibody before being probed at room temperature for 1 hr with a goat anti-mouse IgG antibody conjugated to calf intestinal alkaline phosphatase (diluted 1:3,000; Cell Signaling Technology). Non-specifically bound secondary antibody was removed by washing four times for 5 min each in TBST buffer. A SIGMAFAST BCIP/NBT tablet (Millipore Sigma) dissolved in 10 mL of water was then incubated with the membrane for 10 min to generate a colorimetric precipitate. The membrane was rinsed six times with water before being imaged using a GelDoc Go imaging system (Bio-Rad). Known amounts of pure hexahistidine-SUMO tag were processed in parallel with the lysate to enable determination of absolute quantities of AlkD, YtkR2, and C10R5. Experiments were performed in triplicate.

### Data availability

Atomic coordinates and structure factors generated in this study were deposited in the Protein Data Bank under accession codes 7LXJ [<http://doi.org/10.2210/pdb7LXJ/pdb>] (AlkD/AP-DNA/DSA-Ade) and 7LXH [<http://doi.org/10.2210/pdb7LXH/pdb>] (AlkD/AP-DNA/CC1065-Ade). Previously published structures of AlkD, AlkD/AP-DNA/YTM-Ade, and DSA-DNA are available in the Protein Data Bank under accession codes 3BVS [<http://doi.org/10.2210/pdb3BVS/pdb>], 5UUG [<http://doi.org/10.2210/pdb5UUG/pdb>], and 1DSA [<http://doi.org/10.2210/pdb1DSA/pdb>], respectively. Source data are provided with this paper.

### REFERENCES

1. Jacob, C. & Weissman, K.J. Unpackaging the roles of *Streptomyces* natural products. *Cell Chem. Biol.* **24**, 1194-1195 (2017).
2. Jackson, S.P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078 (2009).
3. Wilson, M.R. et al. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, eaar7785 (2019).
4. Frederick, C.A. et al. Structural comparison of anticancer drug-DNA complexes: adriamycin and daunomycin. *Biochemistry* **29**, 2538-2549 (1990).
5. Hecht, S.M. Bleomycin: new perspectives on the mechanism of action. *J. Nat. Prod.* **63**, 158-168 (2000).
6. Pommier, Y., Leo, E., Zhang, H. & Marchand, C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.* **17**, 421-433 (2010).
7. Tomasz, M. Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* **2**, 575-579

- (1995).
8. Hanka, L.J., Dietz, A., Gerpheide, S.A., Kuentzel, S.L. & Martin, D.G. CC-1065 (NSC-298223), a new antitumor antibiotic. Production, in vitro biological activity, microbiological assays and taxonomy of the producing microorganism. *J. Antibiot. (Tokyo)* **31**, 1211-1217 (1978).
  9. Martin, D.G. et al. CC-1065 (NSC 298223), a potent new antitumor agent. Improved production and isolation, characterization and antitumor activity. *J. Antibiot. (Tokyo)* **34**, 1119-1125 (1981).
  10. Takahashi, I. et al. Duocarmycin A, a new antitumor antibiotic from *Streptomyces*. *J. Antibiot. (Tokyo)* **41**, 1915-1917 (1988).
  11. Ichimura, M. et al. Duocarmycin SA, a new antitumor antibiotic from *Streptomyces* sp. *J. Antibiot. (Tokyo)* **43**, 1037-1038 (1990).
  12. Tokoro, Y., Isoe, T. & Shindo, K. Gilvumycin, a new antitumor antibiotic related to CC-1065. *J. Antibiot. (Tokyo)* **52**, 263-268 (1999).
  13. Igarashi, Y. et al. Yatakemycin, a novel antifungal antibiotic produced by *Streptomyces* sp. TP-A0356. *J. Antibiot. (Tokyo)* **56**, 107-113 (2003).
  14. Huang, W. et al. Characterization of yatakemycin gene cluster revealing a radical S-adenosylmethionine dependent methyltransferase and highlighting spirocyclopropane biosynthesis. *J. Am. Chem. Soc.* **134**, 8831-8840 (2012).
  15. Tichenor, M.S. et al. Systematic exploration of the structural features of yatakemycin impacting DNA alkylation and biological activity. *J. Am. Chem. Soc.* **129**, 10858-10869 (2007).
  16. Tichenor, M.S. & Boger, D.L. Yatakemycin: total synthesis, DNA alkylation, and biological properties. *Nat. Prod. Rep.* **25**, 220-226 (2008).
  17. MacMillan, K.S. & Boger, D.L. Fundamental relationships between structure, reactivity, and biological activity for the duocarmycins and CC-1065. *J. Med. Chem.* **52**, 5771-5780 (2009).
  18. Mullins, E.A., Shi, R. & Eichman, B.F. Toxicity and repair of DNA adducts produced by the natural product yatakemycin. *Nat. Chem. Biol.* **13**, 1002-1008 (2017).
  19. Swenson, D.H. et al. Mechanism of interaction of CC-1065 (NSC 298223) with DNA. *Cancer Res.* **42**, 2821-2828 (1982).
  20. Cundliffe, E. & Demain, A.L. Avoidance of suicide in antibiotic-producing microbes. *J. Ind. Microbiol. Biotechnol.* **37**, 643-672 (2010).
  21. D'Costa, V.M., McGrann, K.M., Hughes, D.W. & Wright, G.D. Sampling the antibiotic resistome. *Science* **311**, 374-377 (2006).
  22. Hubbard, B.K. & Walsh, C.T. Vancomycin assembly: nature's way. *Angew. Chem. Int. Ed. Engl.* **42**, 730-765 (2003).
  23. Lomovskaya, N. et al. The *Streptomyces peucetius drrC* gene encodes a UvrA-like protein involved in daunorubicin resistance and production. *J. Bacteriol.* **178**, 3238-3245 (1996).
  24. Garcia-Bernardo, J., Brana, A.F., Mendez, C. & Salas, J.A. Insertional inactivation of *mtrX* and *mtrY* genes from the mithramycin gene cluster affects production and growth of the producer organism *Streptomyces argillaceus*. *FEMS Microbiol. Lett.* **186**, 61-65 (2000).
  25. Galm, U. et al. Antitumor antibiotics: bleomycin, enediynes, and mitomycin. *Chem. Rev.* **105**, 739-758 (2005).

26. Xu, H. et al. Self-resistance to an antitumor antibiotic: a DNA glycosylase triggers the base-excision repair system in yatakemycin biosynthesis. *Angew. Chem. Int. Ed. Engl.* **51**, 10532-10536 (2012).
27. Wang, S. et al. Characterization of a novel DNA glycosylase from *S. sahachiroi* involved in the reduction and repair of azinomycin B induced DNA damage. *Nucleic Acids Res.* **44**, 187-197 (2016).
28. Yuan, H. et al. GyrI-like proteins catalyze cyclopropanoid hydrolysis to confer cellular protection. *Nat. Commun.* **8**, 1485 (2017).
29. Wu, S. et al. Unified biosynthetic origin of the benzodipyrrole subunits in CC-1065. *ACS Chem. Biol.* **12**, 1603-1610 (2017).
30. Tenconi, E. & Rigali, S. Self-resistance mechanisms to DNA-damaging antitumor antibiotics in actinobacteria. *Curr. Opin. Microbiol.* **45**, 100-108 (2018).
31. Ng, T.L., Rohac, R., Mitchell, A.J., Boal, A.K. & Balskus, E.P. An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **566**, 94-99 (2019).
32. Ma, L. et al. Three putative DNA replication/repair elements encoding genes confer self-resistance to distamycin in *Streptomyces netropsis*. *Acta Biochim. Biophys. Sin. (Shanghai)* **52**, 91-96 (2020).
33. Batty, D., Rapic'Otrin, V., Levine, A.S. & Wood, R.D. Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. *J. Mol. Biol.* **300**, 275-290 (2000).
34. Lin, J.J. & Sancar, A. (A)BC excinuclease: the *Escherichia coli* nucleotide excision repair enzyme. *Mol. Microbiol.* **6**, 2219-2224 (1992).
35. Hitomi, K., Iwai, S. & Tainer, J.A. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst.)* **6**, 410-428 (2007).
36. Krokan, H.E. & Bjørås, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012583 (2013).
37. Kisker, C., Kuper, J. & Van Houten, B. Prokaryotic nucleotide excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012591 (2013).
38. Baraldi, P.G. et al. DNA minor groove binders as potential antitumor and antimicrobial agents. *Med. Res. Rev.* **24**, 475-528 (2004).
39. Ghosh, N., Sheldrake, H.M., Searcey, M. & Pors, K. Chemical and biological explorations of the family of CC-1065 and the duocarmycin natural products. *Curr. Top. Med. Chem.* **9**, 1494-1524 (2009).
40. Hartley, J.A. Antibody-drug conjugates delivering DNA cytotoxics. in *Cancer Drug Design and Discovery* (Elsevier, 2013).
41. Alseth, I. et al. A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol. Microbiol.* **59**, 1602-1609 (2006).
42. Robinson, E.H., Gowda, A.S., Spratt, T.E., Gold, B. & Eichman, B.F. An unprecedented nucleic acid capture mechanism for excision of DNA damage. *Nature* **468**, 406-411 (2010).
43. Mullins, E.A. et al. The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature* **527**, 254-258 (2015).

44. Shi, R., Shen, X.X., Rokas, A. & Eichman, B.F. Structural biology of the HEAT-like repeat family of DNA glycosylases. *Bioessays* **40**, e1800133 (2018).
45. Mullins, E.A., Rodriguez, A.A., Bradley, N.P. & Eichman, B.F. Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.* **44**, 765-781 (2019).
46. Gunz, D., Hess, M.T. & Naegeli, H. Recognition of DNA adducts by human nucleotide excision repair. Evidence for a thermodynamic probing mechanism. *J. Biol. Chem.* **271**, 25089-25098 (1996).
47. Gray, D.M., Ratliff, R.L. & Vaughan, M.R. Circular dichroism spectroscopy of DNA. *Methods Enzymol.* **211**, 389-406 (1992).
48. Osborne, M.R. & Phillips, D.H. Preparation of a methylated DNA standard, and its stability on storage. *Chem. Res. Toxicol.* **13**, 257-261 (2000).
49. Shibasaki, K., Fujii, A., Mikami, N. & Tsuzuki, S. Magnitude of the CH/ $\pi$  interaction in the gas phase: experimental and theoretical determination of the accurate interaction energy in benzene-methane. *J. Phys. Chem. A* **110**, 4397-4404 (2006).
50. Eis, P.S. et al. High resolution solution structure of a DNA duplex alkylated by the antitumor agent duocarmycin SA. *J. Mol. Biol.* **272**, 237-252 (1997).
51. Friedberg, E.C. et al. *DNA Repair and Mutagenesis*, (ASM Press, 2006).
52. Selby, C.P. & Sancar, A. ABC excinuclease incises both 5' and 3' to the CC-1065-DNA adduct and its incision activity is stimulated by DNA helicase II and DNA polymerase I. *Biochemistry* **27**, 7184-7188 (1988).
53. Jin, S.G. et al. Excision repair of adozelesin-N3 adenine adduct by 3-methyladenine-DNA glycosylases and UvrABC nuclease. *Mol. Cells* **11**, 41-47 (2001).
54. Kiakos, K. et al. DNA sequence selective adenine alkylation, mechanism of adduct repair, and *in vivo* antitumor activity of the novel achiral *seco*-amino-cyclopropylbenz[e]indolone analogue of duocarmycin AS-I-145. *Mol. Cancer. Ther.* **6**, 2708-2718 (2007).
55. Fridman, O., Goldberg, A., Ronin, I., Shores, N. & Balaban, N.Q. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513**, 418-421 (2014).
56. Li, B., Qiu, Y., Shi, H. & Yin, H. The importance of lag time extension in determining bacterial resistance to antibiotics. *Analyst* **141**, 3059-3067 (2016).
57. Mullins, E.A., Shi, R., Kotsch, L.A. & Eichman, B.F. A new family of HEAT-like repeat proteins lacking a critical substrate recognition motif present in related DNA glycosylases. *PLoS One* **10**, e0127733 (2015).
58. Parsons, Z.D., Bland, J.M., Mullins, E.A. & Eichman, B.F. A catalytic role for C-H/ $\pi$  interactions in base excision repair by *Bacillus cereus* DNA glycosylase AlkD. *J. Am. Chem. Soc.* **138**, 11485-11488 (2016).
59. Shi, R. et al. Selective base excision repair of DNA damage by the non-base-flipping DNA glycosylase AlkC. *EMBO J.* **37**, 63-74 (2018).
60. Zhao, Q. et al. Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. *Chem. Biol.* **15**, 693-705 (2008).
61. Mullins, E.A., Warren, G.M., Bradley, N.P. & Eichman, B.F. Structure of a DNA glycosylase that unhooks interstrand cross-links. *Proc. Natl. Acad. Sci. USA* **114**, 4400-4405 (2017).

62. Bradley, N.P., Washburn, L.A., Christov, P.P., Watanabe, C.M.H. & Eichman, B.F. *Escherichia coli* YcaQ is a DNA glycosylase that unhooks DNA interstrand crosslinks. *Nucleic Acids Res.* **48**, 7005-7017 (2020).
63. Robinson, E.H., Metz, A.H., O'Quin, J. & Eichman, B.F. A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. *J. Mol. Biol.* **381**, 13-23 (2008).
64. Tichenor, M.S. et al. Asymmetric total synthesis of (+)- and ent(-)-yatakemycin and duocarmycin SA: evaluation of yatakemycin key partial structures and its unnatural enantiomer. *J. Am. Chem. Soc.* **128**, 15683-15696 (2006).
65. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326 (1997).
66. McCoy, A.J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).
67. Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
68. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
69. Davis, I.W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375-W383 (2007).
70. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303 (2018).

## Acknowledgments

This work was funded by the National Science Foundation (MCB-1928918 to B.F.E.) and the National Institutes of Health (R01 CA208669 to D.L.B.). Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Argonne National Laboratory, was supported by the U.S. Department of Energy (DE-AC02-06CH11357). Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (085P1000817).

## Author contributions

E.A.M. and B.F.E. conceived the project; E.A.M., J.D., and B.F.E. designed experiments; E.A.M. performed biophysical and structural experiments; J.D. performed cellular experiments; E.A.M. and J.D. performed biochemical experiments; G.L.T. and D.L.B. provided reagents; E.A.M., J.D., and B.F.E. analyzed data and wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## FIGURE LEGENDS

**Figure 1. DNA modification by the spirocyclopropylcyclohexadienone (SCPCHD) family of natural products.** (a) Chemical structures of yatakemycin (YTM), duocarmycin SA (DSA), and CC-1065. All three compounds share the  $S^0$  subunit (dark cyan), which includes the reactive SCPCHD moiety, and the  $S^{+1}$  subunit (pale cyan). YTM and CC-1065 also possess the additional subunits  $S^{-1}$  (magenta) and  $S^{+2}$  (green), respectively, on either side of the  $S^0$  and  $S^{+1}$  subunits. (b) Generalized mechanisms of DNA modification by SCPCHD compounds and nucleobase excision by associated DNA glycosylases. (c) *ytk* and *c10* gene clusters responsible for biosynthesis of YTM and CC-1065, respectively. Genes encoding the DNA glycosylases YtkR2 and C10R5 are colored red.

**Figure 2. Characterization of DNA adducts produced by YTM, DSA, and CC-1065.** (a,b) Circular dichroism (a) and absorbance (b) spectra of unmodified and alkylated DNA collected at 15°C. (c) Thermal melting profiles of unmodified and alkylated DNA. Melting temperatures ( $T_{m,DNA} = 49^\circ\text{C}$ ,  $T_{m,DSA-DNA} = 72^\circ\text{C}$ ,  $T_{m,YTM-DNA} = 85^\circ\text{C}$ , and  $T_{m,CC1065-DNA} \geq 90^\circ\text{C}$ ) were determined by fitting the data to a polynomial function and calculating the inflexion point of the curve. Spectra and melting profiles for unmodified DNA and YTM-DNA were previously described by Mullins *et al.*<sup>18</sup>. (d) Spontaneous depurination of alkylated DNA at 30°C. The data were fit to a linear model and half-lives ( $t_{1/2,DSA-DNA} = 142 \pm 11$  days,  $t_{1/2,YTM-DNA} = 224 \pm 56$  days, and  $t_{1/2,CC1065-DNA} = 288 \pm 36$  days) were calculated assuming first-order kinetics. Data are presented as the mean  $\pm$  SD from three replicate experiments. (e) Hypothetical models of DNA modified by YTM, DSA, and CC-1065. The dashed boxes indicate nucleotides that are predicted to form CH- $\pi$  interactions with the SCPCHD adducts. (f) Correlation between melting temperature, rate of spontaneous depurination (plotted as nucleotide half-life), and number of nucleotides forming CH- $\pi$  interactions, derived from c, d, and e, respectively. Horizontal error bars represent uncertainty in the melting temperature of CC1065-DNA caused by the lack of an endpoint in the melting profile. As such, the estimated melting temperature of CC1065-DNA was plotted as the average value  $\pm$  the range of values calculated assuming endpoints between those of unmodified DNA (minimum) and DSA-DNA (maximum). Nucleotide half-lives are presented as the mean  $\pm$  SD from three replicate experiments. Source data for a, b, c, d, and f are provided as a Source Data file.

**Figure 3. Determination of drug resistance.** *E. coli* BL21(DE3) cells transformed with either an empty plasmid (blank) or a plasmid encoding AlkD, YtkR2, or C10R5 were grown in the presence of various concentrations of YTM, DSA, or CC-1065. A shorter lag phase (lag time) before the onset of exponential growth indicates increased drug resistance. A complete lack of cell growth

was plotted as a lag time of 18 hr, the total incubation period for the cultures, as indicated by the dashed line. Data are presented as the mean  $\pm$  SD from four replicate experiments. Source data are provided as a Source Data file.

**Figure 4. Measurement of adduct excision activity.** (a) Sequence specificity of SCPCHD natural products. Modified nucleotides are indicated with arrows. The central 2'-deoxyadenosine nucleotide is modified by DSA with relatively low frequency. (b) Separation of full-length DNA substrate (25-mer) from alkali-cleaved product (12- or 13-mer) by denaturing gel electrophoresis. Oligonucleotide substrates were incubated with or without (mock) enzyme for 30 s before reactions were quenched with sodium hydroxide and heat to denature both the enzyme and the DNA and to cleave the DNA at the AP site. The high melting temperature of the GC-rich substrate prevented complete denaturation of the DNA duplex, producing a smeared substrate band. The unique sequence specificities of YTM, CC-1065, and DSA for different 2'-deoxyadenosine nucleotides result in generation of alkali-cleaved products of different lengths. The lower specificity of DSA allows for modification of one of two nucleotides, each corresponding to a different product band. Experiments were performed in triplicate. (c) Single-turnover excision of SCPCHD adducts. Reactions contained 1  $\mu$ M enzyme and 100 nM DNA. With all substrates, both AlkD and YtkR2 approached the endpoints of the reactions by the first time point, potentially reducing or masking differences in excision rates between the adducts and between the enzymes. (d) Multiple-turnover excision of SCPCHD adducts. Reactions contained 10 nM enzyme and 100 nM DNA. (e) Incision of AP-DNA. SCPCHD adducts were pre-incubated with AlkD, YtkR2, or C10R5 to generate AP-DNA for subsequent incision by EndoIV. THF-DNA and 3mAde nucleobase were also pre-incubated with each homolog or without enzyme (mock). Incomplete excision of YTM-Ade by C10R5 during the pre-incubation reaction reduced the amount of AP-DNA (Supplementary Fig. 7), limiting the possible fraction product. The affected quantity is indicated with an asterisk. Of the AP-DNA produced, 98% was subsequently incised by EndoIV. Data in c, d, and e are presented as the mean  $\pm$  SD from three replicate experiments. Source data are provided as a Source Data file.

**Figure 5. Comparison of AlkD product complexes.** (a) Excised SCPCHD adducts from the crystallographic AlkD product complexes. Annealed omit  $mF_o - DF_c$  electron density maps are contoured to  $3\sigma$  and carved around the omitted atoms with a 2-Å radius. (b) Orthogonal views of AlkD in complex with AP-DNA and either YTM-Ade (magenta), DSA-Ade (cyan), or CC1065-Ade (green). (c) Close-up views of the AlkD product complexes showing all residues that interact with



the SCPCHD adducts. Hydrogen-bonding interactions are indicated with dashed lines. The structure of AlkD bound to AP-DNA and YTM-Ade was previously described by Mullins *et al.*<sup>18</sup>.

**Figure 6. Molecular basis of C10R5 substrate specificity.** (a,b) Close-up views of YtkR2 (a) and C10R5 (b) homology models bound to AP-DNA and YTM-Ade. Hydrogen-bonding interactions are indicated with dashed lines. (c) Truncated sequence alignment of AlkD, YtkR2, and C10R5. Residues that interact with AP-DNA and YTM-Ade in the AlkD product complex (PDB accession 5UUG [<http://doi.org/10.2210/pdb5UUG/pdb>]) are indicated with orange triangles and magenta circles, respectively. (d) Single-turnover excision of YTM-Ade by wild-type and mutant C10R5. Reactions contained 1  $\mu$ M enzyme and 100 nM DNA. (e) Multiple-turnover excision of YTM-Ade by wild-type and mutant C10R5. Reactions contained 10 nM enzyme and 100 nM DNA. Mutated residues are indicated with bold font in a and b and with black circles in c. Data in d and e are presented as the mean  $\pm$  SD from three replicate experiments. Source data are provided as a Source Data file. (f) YTM resistance conferred by C10R5 mutants. *E. coli* BL21(DE3) cells were transformed with either an empty plasmid (blank) or a plasmid encoding YtkR2, C10R5, or one of two C10R5 mutants, and grown in the absence or presence of YTM. Experiments were performed in duplicate.

Fig. 1

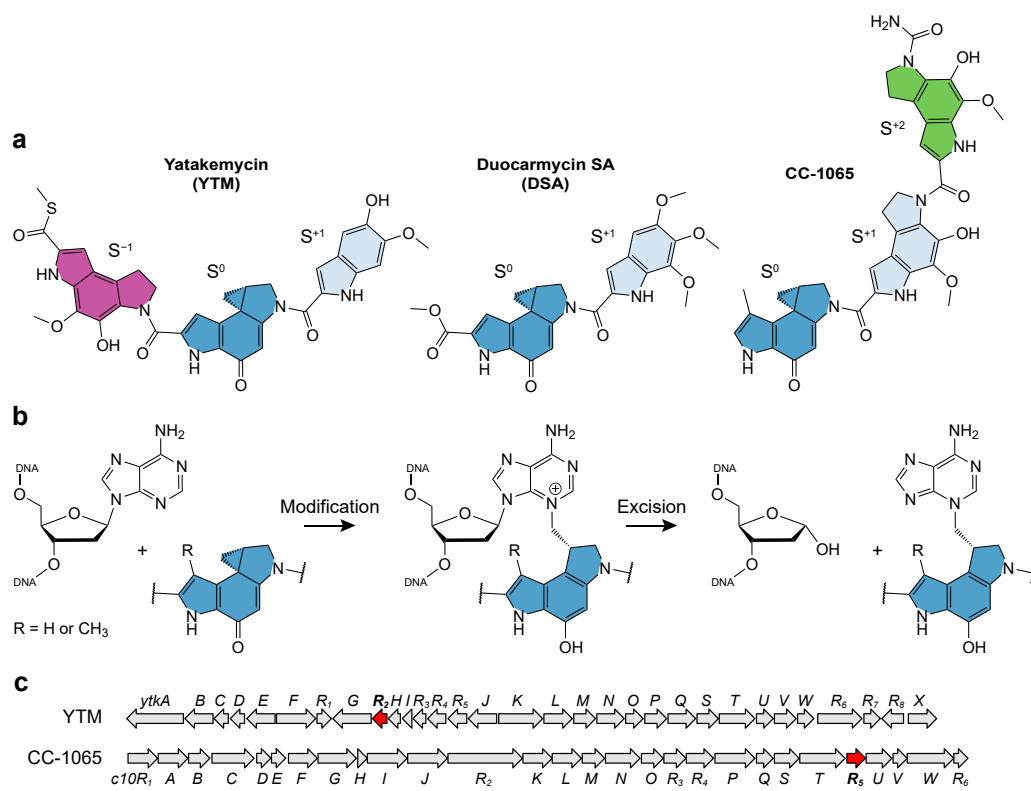


Fig. 2

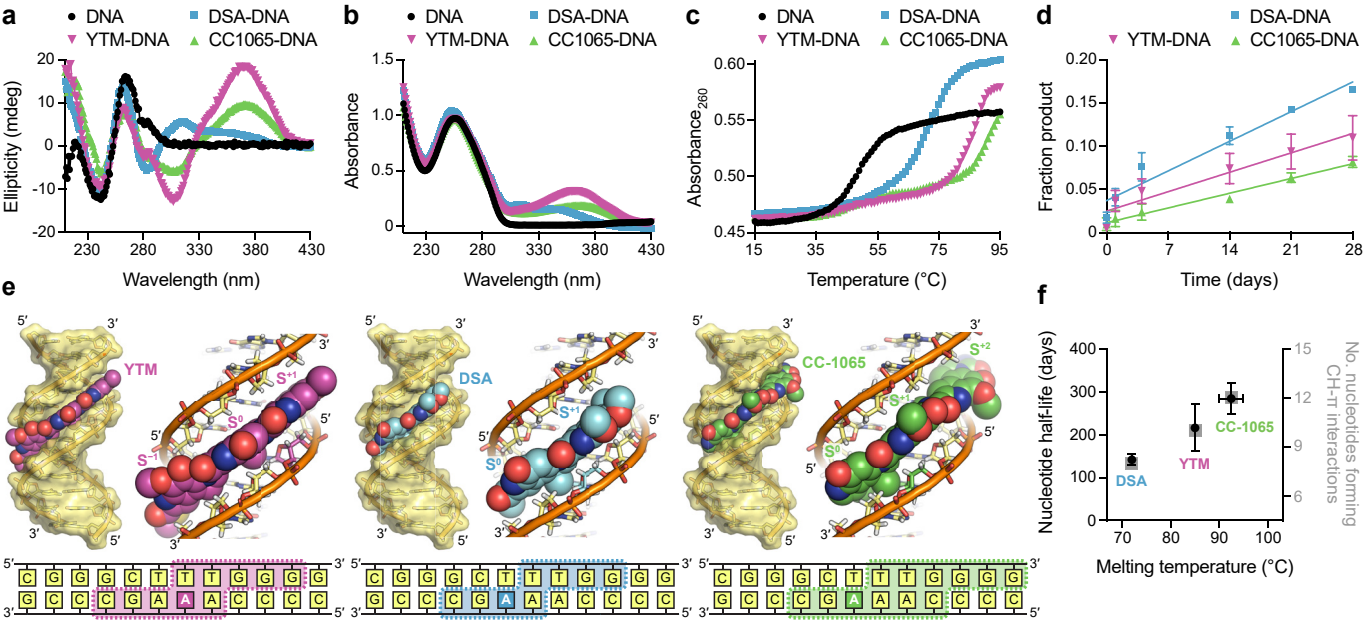


Fig. 3

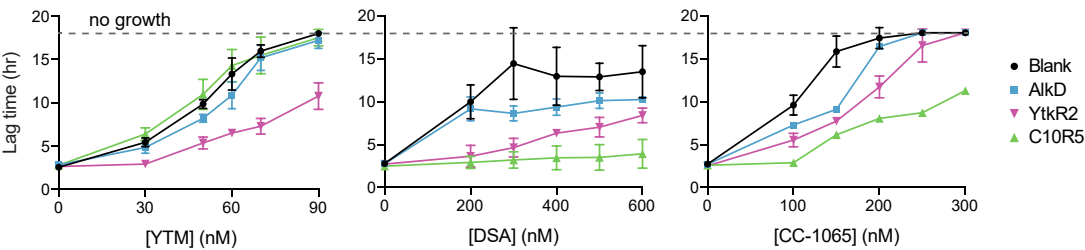


Fig. 4

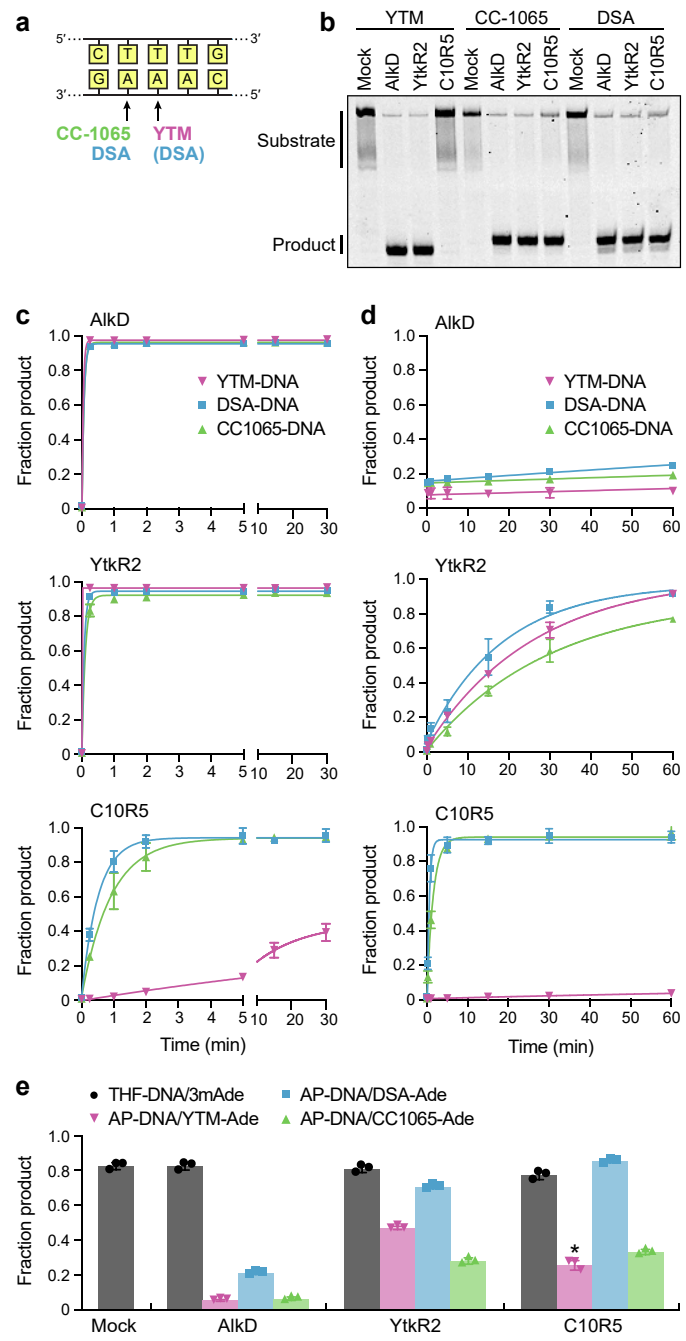


Fig. 5

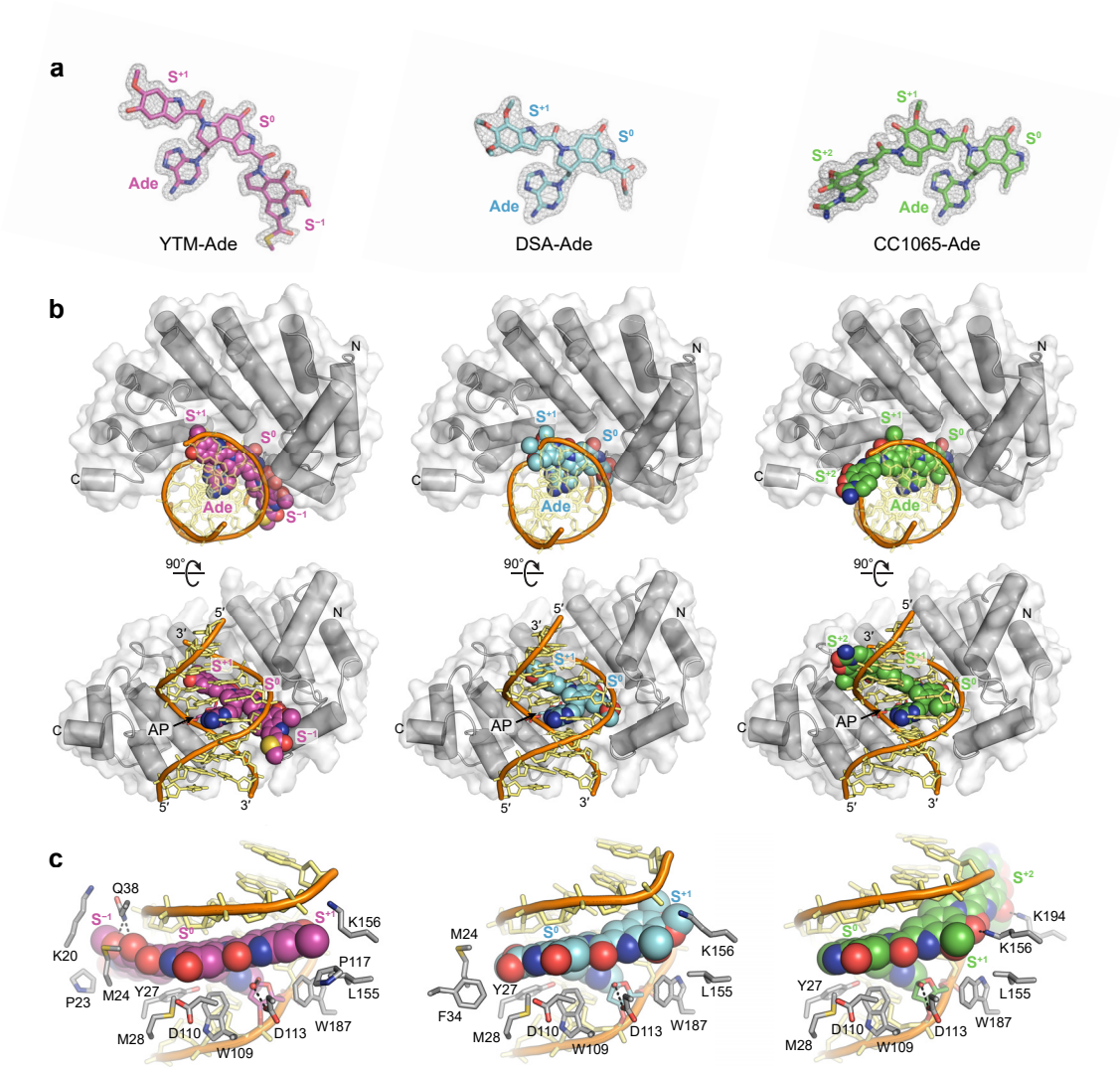


Fig. 6

