Yatakemycin biosynthesis requires two deoxyribonucleases for toxin self-resistance

Jonathan Dorival[1,†], Hua Yuan[2,3,†], Allison S. Walker[1,4], Gong-Li Tang[2,5,*], Brandt F. Eichman[1,6,*]

[1] Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA

[2] State Key Laboratory of Chemical Biology, Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200032, China

[3] Present address: College of Life Sciences, Shanghai Normal University, Shanghai 201418, China

[4] Department of Chemistry, Vanderbilt University, Nashville, Tennessee, USA

[5] School of Chemistry and Materials Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, 1 Sub-lane Xiangshan, Hangzhou 310024, China

[6] Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

[†] These authors contributed equally to this work

*Corresponding authors: 345 Lingling Road, Shanghai, 200032 / email: gltang@sioc.ac.cn; Box 351634 Station B, Nashville, TN 37235-1634 / email: brandt.eichman@vanderbilt.edu

The highly active natural product yatakemycin (YTM) from *Streptomyces* sp. TP-A0356 is a potent DNA damaging agent with antimicrobial and antitumor properties. The YTM biosynthesis gene cluster (*ytk*) contains several toxin self-resistance genes. Of these, *ytkR2* encodes a DNA glycosylase that is important for YTM production and host survival by excising lethal YTM-adenine lesions from the genome, presumably initiating a base excision repair (BER) pathway. However, the genes involved in repair of the resulting apurinic/apyrimidinic (AP) site as the second BER step have not been identified. Here, we show that *ytkR4* and *ytkR5* are essential for YTM production and encode deoxyribonucleases related to other known DNA repair nucleases. Purified YtkR4 and YtkR5 exhibit AP endonuclease activity specific for YtkR2-generated AP sites, providing a basis for BER of the toxic AP intermediate produced from YTM-adenine excision and consistent with co-evolution of *ytkR2*, *ytkR4*, and *ytkR5*. YtkR4 and YtkR5 also exhibit 3′-5′ exonuclease activity with differing substrate specificities. The YtkR5 exonuclease is capable of digesting through a YTM-DNA lesion and may represent an alternative repair mechanism to BER. We also show that *ytkR4* and *ytkR5* homologs are often clustered together in putative biosynthetic gene clusters, consistent with non-redundant roles in repair of other DNA adducts derived from genotoxic natural products.

Introduction

Some plant and microbial natural products are toxic by virtue of their ability to chemically modify DNA. Genotoxic compounds are diverse in chemical structure and generate an array of covalent and non-covalent DNA adducts that inhibit DNA processing and lead to cell death or disease [1-7]. The high cytotoxicity of DNA damaging natural products can be harnessed to develop antimicrobial and anticancer drugs such as actinomycin D, daunomycin, mitomycin C, and calicheamicins [1]. Cells have evolved several conserved pathways to detect and repair different types of DNA damage as a means of survival [2, 8-10]. For example, bulky, helix-distorting adducts are typically removed by nucleotide excision repair (NER), whereas smaller lesions are repaired by direct reversal or base excision repair (BER) pathways.

Yatakemycin (YTM, **1**) is produced by *Streptomyces* sp. TP-A0356 and belongs to the spirocyclopropylcyclohexadienone (SCPCHD) family of genotoxic natural products that includes duocarmycin A, and duocarmycin SA, CC-1065, and gilvusmycin (**Fig. 1A**) [11-15]. These compounds preferentially bind the minor groove of AT-rich sequences and alkylate the N3-position of adenine through ring opening of their cyclopropyl groups [16-19]. In addition to covalent adducts, SCPCHD compounds also form a network of non-covalent interactions with both DNA strands that stabilizes the duplex and effectively creates a non-covalent interstrand DNA crosslink that poses a challenge to excision repair pathways [20-24]. Consequently, SCPCHD-DNA adducts are potent blocks to DNA replication and exhibit antibiotic, antifungal, and antitumor properties [11-15, 25-28]. YTM is unique in that its cyclopropyl ring resides in the middle subunit, which forms a “sandwiched” structure with an enhanced rate of DNA alkylation, making YTM is the most potent SCPCHD member with an IC_{50} of 3-5 pM against the L1210 cell line [17].

Toxin-producing microorganisms have self-resistance mechanisms for survival and, in the case of genotoxins, to protect the integrity of the genome [29-37]. These self-resistance mechanisms are often embedded within and co-evolve with the biosynthesis gene cluster (BGC). Biosynthetic studies revealed that the YTM producer encodes multiple resistance/regulation genes in its BGC, including *ytkR1-ytkR8* (**Fig. 1B**) [26]. The *ytkR6* gene is homologous to drug-resistance transporters and was proposed to function as an efflux pump [26]. Three genes—*ytkR1*, *ytkR7*, and *ytkR8*—encode GyrI-like small molecule binding proteins; YtkR7 inactivates YTM by catalyzing the hydrolysis of the cyclopropyl warhead to form products **2** and **3** (**Fig. 1C**), and YtkR1 and YtkR8 have been proposed to sequester YTM or its intermediates and to regulate expression of the cluster, respectively [38].

YtkR2 confers self-resistance against YTM by functioning as a DNA glycosylase that excises YTM-adenine adducts [39]. DNA glycosylases initiate the BER pathway by hydrolyzing the N-glycosidic bond of the aberrant nucleotide to generate an apurinic/apyrimidinic (AP, or abasic) site. The AP site is a reactive intermediate that is removed in the subsequent steps of BER, whereby an AP endonuclease hydrolyzes the phosphodiester bond on the 5′ side of the AP site to generate a 3′-hydroxyl for gap filling synthesis by a DNA polymerase, followed by nick sealing by DNA ligase [40]. YtkR2 is a member of the AlkD family of DNA glycosylases that are unique in their ability to excise bulky DNA adducts [20-22, 39, 41-46]. Consistent with its co-

evolution within the YTM cluster, YtkR2 confers a greater cellular resistance against YTM than does *Bacillus cereus* AlkD [22]. This enhanced resistance is not the result of substrate specificity, but rather from a low affinity of YtkR2 for its AP site product, which presumably allows for enzymatic removal of the AP site [20, 22]. How the AP site is repaired in *Streptomyces* sp. TP-A0356, however, has been unclear.

Here, we show genetically and biochemically that *ytkR4* and *ytkR5* encoded in the YTM BGC are resistance genes (**Fig. 1C**) that encode multifunctional deoxyribonucleases (DNases) capable of processing YTM-DNA adducts. YtkR4 is homologous to the TatD family of DNases, and YtkR5 is a putative member of the xylose isomerase-like TIM barrel domain family found in the bacterial AP endonuclease, Endonuclease (Endo) IV [26, 39]. Consistent with these annotations, we found that YtkR4 and YtkR5 are 3′-5′ exonucleases with a preference for single- (ss) and double-stranded (ds) DNA substrates, respectively, and that YtkR5 can degrade DNA containing a YTM-adenosine (YTMA) lesion. In addition, both enzymes exhibit AP endonuclease activity toward the toxic AP product of YTM-adenosine hydrolysis by YtkR2. We also found through bioinformatic analysis that *ytkR4* and *ytkR5* homologs in other bacteria are often located together within putative BGCs, suggesting they play multiple, non-overlapping roles in repair of genotoxin DNA adducts.

Results and discussion

YtkR4 and YtkR5 are essential for yatakemycin biosynthesis

To investigate the effects of *ytkR4* and *ytkR5* on YTM biosynthesis, we generated Δ*ytkR4* and Δ*ytkR5* mutants of the YTM producing strain *Streptomyces* sp. TP-A0356 by constructing an in-frame deletion for each gene through homologous recombination (**Supplementary Fig. S1**). We then carried out fermentation analyses and monitored YTM production by HPLC. Both mutants lost the ability to produce YTM and its two hydrolyzed products, **2** and **3** (**Fig. 1D**). To verify the loss of YTM production was related to YtkR4 and YtkR5 activities, we introduced a plasmid expressing the native gene into the corresponding mutant. Complementation partially restored production of YTM and/or its two hydrolyzed products in each mutant (**Fig. 1D**). Thus, these genetic results indicate that the *ytkR4* and *ytkR5* genes each are essential for YTM biosynthesis.

YtkR4 and YtkR5 exhibit exonuclease and AP endonuclease activities

YtkR4 and YtkR5 are distant homologs of TatD and EndoIV deoxyribonucleases, respectively (**Supplemental Figs. S2 and S3**) [26, 39]. Both TatD and EndoIV exhibit AP endonuclease and 3′-5′ exonuclease activities [47-52]. We therefore tested nuclease activities of purified YtkR4 and YtkR5 proteins on 5′-FAM labeled ssDNA and dsDNA substrates containing either no modification or a centrally located tetrahydrofuran (THF) abasic site analog. In the presence of Mg^{2+} cofactor, YtkR4 and YtkR5 generated a ladder of products on both ssDNA and dsDNA, indicative of 3′-5′ exonuclease activity (**Fig. 2A**). YtkR4 showed greater exonuclease activity on ssDNA than on dsDNA, similar to the substrate specificity of

human and *E. coli* TatDs [47, 49]. In contrast, YtkR5 exhibited a preference for dsDNA, consistent with its homology to EndoIV, which also has greater activity for dsDNA [52]. In addition to exonuclease activity, both enzymes displayed AP endonuclease activity on the dsDNA THF substrate, as evidenced by the accumulation of a band corresponding to specific cleavage at the THF (**Fig. 2A**). YtkR5, but not YtkR4, showed a robust exonuclease degradation of the product of the AP endonuclease reaction, consistent with its preference for exonucleolytic activity in dsDNA. Neither enzyme exhibited AP endonuclease activity in ssDNA. THF inhibited the ssDNA exonuclease activity of YtkR4, as we observed an accumulation of a band corresponding to an oligonucleotide one nucleotide longer than the AP endo cleavage product on the THF-ssDNA but not the unmodified ssDNA substrate, indicating that YtkR4 pauses after cleaving the nucleotide immediately 3′ to the THF. We previously observed the same pausing behavior in response to THF from the human and *E. coli* TatD enzymes [47].

Since TatD and EndoIV utilize metal cofactors, we examined the metal dependence of YtkR4 and YtkR5 (**Fig. 2B**). We tested the activity of each protein against seven different divalent cations using the THF-containing dsDNA substrate so that we could monitor both AP endo and exonuclease activities. YktR4 exhibited the most exonuclease activity in the presence of Mg^{2+}, Mn^{2+}, and Co^{2+}, weak activity with Ca^{2+}, Zn^{2+}, and Ni^{2+}, and no activity with Cu^{2+}. AP endonuclease activity of YtkR4 was greatest in presence of Mg^{2+}, and a low level of activity was observed with Ca^{2+}. YktR5 exhibited both AP endonuclease and exonuclease activity in presence of Mg^{2+}, Mn^{2+}, and to a lesser extent Co^{2+}. In the presence of Ca^{2+}, YtkR5 displayed robust AP endonuclease activity and reduced exonuclease activity. Thus, whereas YtkR4 exhibited the same metal dependence as the TatD enzymes [47], the metal dependence of YtkR5 was markedly different than EndoIV. While EndoIV uses Zn^{2+} as a preferred metal for activity and remains active in presence of EDTA [52], YtkR5 exhibited the most nuclease activity in the presence of Mg^{2+}, Mn^{2+} and Ca^{2+} and very low activity in the presence of Zn^{2+} and EDTA. This indicates that YtkR5 may have evolved differently compared to EndoIV.

The metal preferences can be explained by similarities and differences in metal binding residues in the active sites of these enzymes (**Supplementary Fig. S2 and S3**). YtkR4's substrate specificity and metal dependence are reminiscent of the TatD enzymes [47, 49]. Despite only 20% sequence identity overall, the TatD metal binding residues are largely conserved in YtkR4, with only two out of six exceptions (**Supplementary Fig. S2**). YtkR4 Arg168 and Gly241 align with conserved TatD His and Asp residues, respectively. Although these substitutions would not coordinate the metals, they are not expected to alter the metal preference because they are each one of four that bind a separate metal [47]. YtkR5, on the other hand, has a markedly different metal dependence than EndoIV despite the similarity in preference for dsDNA [52]. Several EndoIV residues known to bind metals are different or absent in YtkR5, which likely affect the nature or number of metals bound in the active site (**Supplementary Fig. S3**). For example, Zn_A^{2+} and Zn_B^{2+} binding residues Asp229 and His216 in EndoIV align with His218 and Gln209 in YtkR5. EndoIV His69 and His109, which bind the third Zn^{2+} ion, align with Asp77 and Ala117 in YtkR5, such that only Glu155 would be involved in the binding of the third metal.

YtkR4 and YtkR5 AP endonucleases have a modest preference for the YtkR2 product

YtkR4 and YtkR5 presumably evolved with the *ytk* gene cluster, and thus we tested the hypothesis that these nucleases would show a preference for the products of YtkR2 cleavage of YTMA lesions (i.e., AP-DNA + YTM-adenine). We examined their AP endonuclease activities on substrates containing natural abasic sites generated either by uracil DNA glycosylase (UDG) excision of deoxyuracil or by YtkR2 excision of YTMA, as compared to the THF-containing substrate (**Fig. 3A**). The total (AP endo + exo) nuclease activity of YtkR4 was similar among the three substrates, as judged by the rates of product accumulation, whereas YtkR5 exhibited slightly higher overall nuclease activity for THF or UDG-derived AP sites (**Fig. 3B**). However, both enzymes showed slightly higher AP endo activity toward the YTMA-derived AP site (**Fig. 3C**). In the case of YtkR4, the incision product of the YTMA-derived AP site accumulated faster and to a greater extent than those of the UDG-derived or THF sites. In contrast, the products of YtkR5 AP endo activity from YTMA-derived AP sites accumulated to a greater extent, albeit slower, than those of the UDG-derived or THF AP sites (**Fig. 3C**). Together, these results indicate that the AP endonuclease activities of both YtkR4 and YtkR5 are specific for AP sites derived from YtkR2 excision of YTMA. Because the AP sites produced from YtkR2 and UDG are identical, the specificity is likely the result of the presence of the excised YTM-adenine lesion remaining non-covalently bound in the DNA, but may also be attributed to a potential interaction with YtkR2. The YTM moiety makes intimate interactions with both DNA strands [20] and thus we would not expect YTM-adenine to dissociate as readily as uracil.

Exonuclease activity of YtkR5 is not inhibited by YTMA lesions

Because of the possibility that YtkR4 and YtkR5 would encounter a YTMA lesion in DNA prior to its excision by YtkR2, we examined whether this adduct inhibits the exonuclease activities of YtkR4 and YtkR5. We incubated each enzyme with a dsDNA substrate containing a centrally located YTMA lesion and compared the kinetics of exonuclease activity against the same substrate containing no modification. Compared to its activity on unmodified DNA, YtkR4 did not produce excision products 5′ to the lesion, indicating that it is unable to digest DNA beyond YTMA (**Fig. 4A**). Interestingly, YtkR4 experienced a slight burst of exonuclease activity on the YTMA substrate, although the subsequent rates of excision were the same between the two substrates (**Fig. 4B**). In contrast, YktR5 was able to completely digest the YTMA substrate at the same rate as the unmodified substrate. Given that the YTMA lesion stabilizes the DNA duplex [20], this is consistent with YtkR5's preference for a dsDNA substrate and suggests that exonuclease digestion by YtkR5 constitutes an alternate YTMA-DNA repair pathway to BER by YtkR2.

YtkR4 and YtkR5 are found together in biosynthetic gene clusters

The specificity of YtkR4 and YtkR5 for the product of the YtkR2 reaction is consistent with co-evolution of these proteins. To investigate whether this is unique to the *ytk* cluster, we searched for co-existence of these proteins in other genomes. We performed a BLASTp

search to identify genomes that contain a YtkR4 homolog and obtained 566 hits. We then performed two additional BLASTp searches for YtkR2 and YtkR5 homologs in these genomes, of which 311 had a YtkR2 homolog and 526 had a YtkR5 homolog. Therefore, YtkR4 homologs are only very rarely present in genomes without YtkR5 homologs, while the same is not true for YtkR2. There were incomplete genomes present in our analysis so it is possible that the fraction of YtkR4 homolog-containing genomes that also contain YtkR5 or YtkR2 homologs are underestimated but we expect that the difference between YtkR5 and YtkR2 would hold for an analysis of only complete genomes.

YtkR5 is a member of the xylose isomerase-like superfamily that contains both sugar isomerases and DNA endonucleases [53], so we cannot be certain that all our hits have YtkR5-like activity rather than sugar isomerase activity. To account for this, we clustered enzymes by function [54] using a sequence similarity network (SSN). We found that there were four major sequence clusters of YtkR5 homologs as well as several outlier sequences in clusters with fewer than three sequences (**Fig. 5A**). We found that YtkR2 is generally distant from YtkR4 in the genome, with only a small fraction of the homologs being within 100 kpb of each other, meaning that most YtkR2 homologs are not in the same BGC as the YtkR4 homolog (**Fig. 5B**). Conversely, nearly all YtkR5 homologs from the four major SSN clusters are within 10 kbp of the YtkR4 homolog, making it highly likely that they are in the same BGC. This also suggests that all the major SSN clusters have YtkR5-like activity because it is unlikely that a sugar isomerase would be so tightly associated with YtkR4 homologs. The YtkR5 homologs that were in SSN clusters with three or fewer members were roughly evenly split between being within 10 kbp and being further than 100 kbp from YtkR4, which could indicate that only some of them have YtkR5-like functions (**Fig. 5**). This analysis agrees with the gene co-occurrence patterns described above in which YtkR4 and YtkR5 homologs co-occur more often than the YtkR4 and YtkR2 homologs.

Together, these results suggest that the YtkR4 and YtkR5 homologs are both important for resistance to other DNA damaging natural products and therefore often co-occur in the same genome or BGC. To confirm that *ytkR4* and *ytkR5* occur in BGCs, we used antiSMASH to determine if there was a predicted BGC within 40 kbp of YtkR4. BGCs were predicted for 113 out of the 566 YtkR4 hits. It is possible that the remaining YtkR4s are also in BGCs that antiSMASH is not capable of identifying. This co-existence is consistent with their different nuclease activities, which would provide non-redundant mechanisms for lesion repair (**Fig. 6**). The weak co-existence and genomic distance between YtkR2 and YtkR4 homologs in other genomes suggests that those YtkR2 homologs do not rely on a YtkR4 or YtkR5 nuclease to process the AP site product. Indeed, homologous *ytkR4* and *ytkR5* genes are not present in the CC-1065 BGC [22, 55]. Similarly, there is a class of BGCs that produce DNA crosslinking and intercalating natural products and that contain the AlkZ-like glycosylases, unrelated to YtkR2, that provide self-resistance by excising the resulting DNA adducts [37, 56, 57], although there are no apparent AP endonucleases or other DNA repair genes to remove the AP lesion in those clusters. Our genomic analysis suggests that the specificity of *Streptomyces* sp. TP-A0356 YtkR4 and YtkR5 for the product of the YtkR2 reaction, and thus the existence of at least a partial BER system for this lesion (**Fig. 6A**), is unique to the *ytk* cluster. If YtkR4 and YtkR5 homologs in other genomes have AP

endonuclease activity, it likely plays a more minor role than their exonuclease activities in lesion repair (**Fig. 6B**) or it is acting on the product of a different glycosylase. Thus, we speculate that the exonuclease activity of YtkR4 and YtkR5 is the major function of these enzymes that enables their repair of a broad range of DNA adducts and toxic DNA repair intermediates.

Experimental

Construction and complementation of ytkR4 and ytkR5 mutants

Bacterial strains and plasmids used in this study are listed in **Supplementary Table S1**. The *ytkR4* and *ytkR5* gene in-frame deletion mutants were constructed as previously described [55]. Briefly, two homologous DNA arms were cloned into the *Hin*dIII and *Eco*RI sites of the thermosensitive plasmid pKC1139 using primers listed in **Supplementary Table S2**. Then, the recombinant plasmid was introduced into *Streptomyces* sp. TP-A0356 from *E. coli* S17-1 to obtain apramycin-resistant exconjugants at 30°C. The exconjugant was then grown at 37°C to obtain apramycin-resistant single-crossover mutant culture, which was further used for screening apramycin-sensitive clones without apramycin selection for generations. The desired double-crossover mutants were verified by PCR analysis with primers listed in **Supplementary Table S2**.

For complementation experiments, each coding sequence was cloned into the *Nde*I and *Eco*RI sites of the integrative plasmid pIB139, and the recombinant plasmid was introduced into *ytkR4* or *ytkR5* mutant from *E. coli* S17-1 to obtain apramycin-resistant exconjugants at 30°C. The apramycin-resistant exconjugants were then used for further fermentation analysis.

Fermentation and analysis of YTM metabolites

For *Streptomyces* sp. TP-A0356 and its derivative strains, fermentation and analyses of YTM and other relative metabolites were carried out as previously described [38]. Briefly, *Streptomyces* sp. TP-A0356 and its derived mutant strains were first inoculated in the liquid seed medium (Tryptic Soy Broth) in a 250 mL flask, and then the culture was used for transfer onto the solid fermentation medium (International Streptomyces Project medium 2) for growth at 30°C for ~5 days. The culture was used for extracting YTM and other metabolites, which were for further HPLC analysis.

The HPLC analysis was performed on the Agilent 1200 HPLC system (Agilent Technologies Inc., USA) using a reverse-phase Alltima C18 column (5 μm, 4.6 × 250 mm). Solvent A was H_2O and Solvent B was CH_3CN, and the flow rate was 1 mL min−1 with DAD detector. The analytic HPLC conditions are as follows: the gradient program was 0–3 min 15% B, 3–6 min 15–40% B, 6–12 min 40% B, 12–19 min 40–55% B, 19–22 min 55–85% B, 22–28 min 85% B, and 28–29 min 15% B.

Protein expression and purification

The coding sequences of YtkR4 and YtkR5 were synthesized by GenScript without codon optimization and ligated into pBG102, a modified pET-27 expression vector encoding an N-terminal Rhinovirus 3C-cleavable hexahistidine-SUMO fusion tag. For YtkR4 expression, the plasmid was transformed into C41 cells along with the pG-Tf2 vector (Takara Bio) encoding Trigger Factor, GroES, and GroEL chaperones. Cells were grown at 37°C in Luria Bertani (LB) medium supplemented with 5 µg/L tetracycline. When the cultures reached an A_{600} of 0.6, they were incubated for 1 hr at 18°C before protein expression was induced by addition of 0.1 mM IPTG. Induced cells were grown overnight at 18°C. The plasmid encoding YtkR5 was co-transformed into BL21 cells along with the pGro7 vector (Takara Bio) encoding GroES and GroEL chaperones. Cells were grown at 37°C in LB medium supplemented with 5 g/L arabinose, as well as 2 mM betaine and 50 mM sorbitol to increase protein solubility. When the A_{600} reached 1.0, the cells were incubated for an hour at 18°C and 0.09 mM IPTG was added to induce the protein expression. Cultures were then incubated at 18°C overnight.

For both YtkR4 and YtkR5, cells were harvested by centrifugation and the resulting cell pellet was resuspended in lysis buffer (30 mM MOPS pH 7.5, 300 mM NaCl, 10% glycerol). Benzonase (25 U/L culture, Sigma-Aldrich) was added to the lysis buffer at a concentration of 25 U/L culture, as well as 6 mM $MgCl_2$, 10 mM KCl, and 4 mM ATP to detach the chaperones from the proteins during cell lysis. Cells were lysed using an Avestin C3 Emulsiflex operating at 15,000 psi. Cell debris was removed by centrifugation at 45,000 x g for 30 min. The supernatant was supplemented with 40 mM imidazole and applied to a Ni-NTA column. The column was washed with 20 column volumes of lysis buffer supplemented with 40 mM imidazole and the protein was eluted using buffer B (lysis buffer supplemented with 500 mM imidazole). The hexahistidine-SUMO tag was removed by overnight cleavage at 4°C while dialyzing against the lysis buffer to remove imidazole. The sample was then reinjected onto the Ni-NTA column and the flow-through collected. For YtkR4 only, an additional purification step was performed, whereby the protein was then diluted four times in buffer Q (20 mM HEPES pH 7.5, 5% glycerol) and injected onto a Q-sepharose column. The protein was eluted using a 0-1 M NaCl gradient. YtkR4 and YtkR5 were concentrated using an Amicon Ultracel-10 (Merck Millipore), incubated for 15 min in 10 mM EDTA, and injected over onto a HiLoad Superdex 200 16/60 pg column (Cytiva) equilibrated in buffer S (30 mM HEPES pH 7.7, 300 mM NaCl, 5% glycerol). The proteins were concentrated to 100 µM and aliquots flash frozen in liquid nitrogen and stored at -80°C.

DNA substrate preparation

Oligodeoxynucleotides were purchased from Integrated DNA Technologies. The sequences used are provided in **Supplementary Table S3**. Oligonucleotides containing a centrally located THF or deoxyuridine residue were labeled with 6-carboxyfluorescein (FAM) on the 5′-end and HPLC purified. Double-stranded substrates were formed by annealing FAM-oligonucleotides to an unlabeled complementary strand in annealing buffer (20 mM EPPS pH 8.0, 50 mM NaCl). Sequences containing natural AP sites were generated by reacting 50 µL of 50 µM dsDNA uracil-containing oligonucleotide with 5 U uracil DNA glycosylase (New England Biolabs) for 60 min at 37°C. YTM was purified as previously

described [26] and YTM-DNA substrates were generated as previously described [20], except reaction mixtures contained 20 mM EPPS pH 8.0, 50 mM NaCl, 10% (v/v) dimethylsulfoxide, 10 µM DNA, and 150 µM YTM. After 18 hr at 22°C, excess YTM was removed by passing the reaction mixture through a Microspin G25 column (Cytiva) equilibrated in annealing buffer.

Nuclease assays

Nuclease reactions were carried out at 37°C with 10 µM protein, 100 nM DNA, and buffer containing 20 mM EPPS pH 8.0, 50 mM NaCl, and 10 mM $MgCl_2$. Experiments to test activity in the presence of different divalent metals were carried out under the same conditions but supplemented with either 5 mM $CaCl_2$, 0.05 mM $ZnCl_2$, 3 mM $MnCl_2$, 1 mM $NiCl_2$, 1 mM $CuCl_2$, or 10 mM EDTA. Reactions were quenched by addition of an equivalent volume of loading buffer (80% formamide, 25 mM EDTA pH 8.0, 2 mg/mL orange G, and 1 mg/mL xylene cyanol, 10 U proteinase K), heated at 70°C for 10 min, and subjected to denaturing polyacrylamide gel electrophoresis (PAGE) on 20% acrylamide/8M urea sequencing gels. Fluorescence from the FAM-labeled DNA was detected using a Typhoon Trio variable mode imager (GE Healthcare). For excision of the YTM adducts, 10 µM YtkR2 was added to the reactions prior to YtkR4 or YtkR5. To verify excision of the YTM-adduct by YtkR2, reactions were quenched by addition of 100 mM NaOH and heated at 70°C for 10 min before addition of loading buffer.

Identification of YtkR4, YtkR2, and YtkR5 homologs

BLAST version 2.8.1 command line software was used to perform a blastp search against the refseq_protein database with YtkR4 as the query sequence. The cutoff E-value was set at 10^{-20} and the maximum target sequences was set at 100,000. The output format was set with the following options: “7 qacc sacc sgi evalue qstart qend sstart send”. Genomes containing hits were obtained by first downloading the record for the protein from NCBI, extracting the genome accession, and then downloading the genome. Downloads were performed using the NCBI datasets command line tool using a custom script, download_blast_hit_genomes.py. A full list of genomes downloaded and used in subsequent analysis is available in **Supplementary Table S4**. We used tblastn to confirm presence of a YtkR4 homolog and search for YtkR2 and YtkR5 homologs in the downloaded genome. One BLAST search was run for each gene-genome pair, again using the command line with an E-value of 10^{-5} and output format options: “7 qacc sacc sgi evalue pident qstart qend sstart send”. This process was automated with a custom script, search_for_other_genes.py. The SSN was created by uploading the fasta file containing YtkR5 blastp hits to the EFI-Enzyme Similarity Tool server [58] and using a score cutoff of 80. The SSN was colored by cluster membership and visualized using Cytoscape.

Analysis of genomic distance between YtkR4 and YtkR2/YtkR5

Complete genomes were determined to be circular or linear from the genbank file record using a custom script (genomes_circular_list.py). To determine the distance for a linear genome, we identified the minimum distance between the end of one gene and the

start of the other. To determine the distance for a circular genome we measured the minimum number of nucleotides between genes looking in both directions in the chromosome. For genomes that were not complete, we were unable to analyze the distance in both directions. For these genomes, we assumed that if the two genes were observed on the same contig that the shorter distance between them would be their distance on the contig rather than the other possible distance if the genome is circular. Distances were not determined for any pairs that did not fall in the same contig and these pairs are not represented in the plot in **Fig. 5**. Distances were determined using the custom script search_for_other_genes.py. These distances were then plotted as a strip plot using the python modules seaborn and matplotlib using a custom script (make_stripplots.py). To analyze if YtkR4 was in a putative BGC, the 40 kbp on either side of YtkR4 was extracted and used as input for antiSMASH version 6 [59] (**Supplementary Table S4**).

Conclusions

We have defined the previously uncharacterized genes *ytkR4* and *ytkR5* within the YTM BGC as resistance genes important for YTM production. YtkR4 and YtkR5 are AP endonucleases acting on the product of the YtkR2 glycosylase. Both enzymes contain 3′-5′ exonuclease activity, which in YtkR5 is most pronounced on dsDNA substrates and capable of excising a YTM-adenosine lesion. Genes encoding YtkR4 and YtkR5 homologs in other bacteria are often located together within putative BGCs, but without a YtkR2 homolog, suggesting YtkR4 and YtkR5 play multiple, non-overlapping roles in repair of genotoxin DNA adducts.

Author contributions

Jonathan Dorival: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. Hua Yuan: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. Allison S. Walker: Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft, Visualization, Writing – review & editing. Gong-Li Tang: Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. Brandt F. Eichman – Conceptualization, Funding acquisition, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article, including custom scripts for genomic analysis and fasta files containing homologous sequences are available at GitHub at https://github.com/aswalker-lab/YtkR4_bioinformatics.git. Additional data supporting this article have been included as part of the Supplementary Information.

Acknowledgements

We thank Prof. Yasuhiro Igarashi from Toyama Prefectural University, Japan for kindly providing *Streptomyces* sp. TP-A0356. This work was supported by grants from the National Science Foundation (MCB-1928918 and MCB-2341288) to B.F.E. and the National Natural Science Foundation of China (31930002) to G.-L.T.

Notes and References

1. W. C. Tse and D. L. Boger, *Chem. Biol.*, 2004, **11**, 1607-1617.
2. S. P. Jackson and J. Bartek, *Nature*, 2009, **461**, 1071-1078.
3. M. R. Wilson, Y. Jiang, P. W. Villalta, A. Stornetta, P. D. Boudreau, A. Carrá, C. A. Brennan, E. Chun, L. Ngo, L. D. Samson, B. P. Engelward, W. S. Garrett, S. Balbo and E. P. Balskus, *Science*, 2019, **363**, eaar7785.
4. C. A. Frederick, L. D. Williams, G. Ughetto, G. A. van der Marel, J. H. van Boom, A. Rich and A. H. Wang, *Biochemistry*, 1990, **29**, 2538-2549.
5. S. M. Hecht, *J. Nat. Prod.*, 2000, **63**, 158-168.
6. Y. Pommier, E. Leo, H. Zhang and C. Marchand, *Chem. Biol.*, 2010, **17**, 421-433.
7. M. Tomasz, *Chem. Biol.*, 1995, **2**, 575-579.
8. J. H. Hoeijmakers, *Nature*, 2001, **411**, 366-374.
9. C. Yi and C. He, *Cold Spring Harb Perspect Biol*, 2013, **5**, a012575.
10. S. N. Andres, M. J. Schellenberg, B. D. Wallace, P. Tumbale and R. S. Williams, *Environ. Mol. Mutagen.*, 2015, **56**, 1-21.
11. Y. Igarashi, K. Futamata, T. Fujita, A. Sekine, H. Senda, H. Naoki and T. Furumai, *J. Antibiot. (Tokyo)*, 2003, **56**, 107-113.
12. L. J. Hanka, A. Dietz, S. A. Gerpheide, S. L. Kuentzel and D. G. Martin, *J. Antibiot. (Tokyo)*, 1978, **31**, 1211-1217.
13. Y. Tokoro, T. Isoe and K. Shindo, *J. Antibiot. (Tokyo)*, 1999, **52**, 263-268.
14. I. Takahashi, K. Takahashi, M. Ichimura, M. Morimoto, K. Asano, I. Kawamoto, F. Tomita and H. Nakano, *J. Antibiot. (Tokyo)*, 1988, **41**, 1915-1917.
15. M. Ichimura, T. Ogawa, K. Takahashi, E. Kobayashi, I. Kawamoto, T. Yasuzawa, I. Takahashi and H. Nakano, *J. Antibiot. (Tokyo)*, 1990, **43**, 1037-1038.
16. D. H. Swenson, L. H. Li, L. H. Hurley, J. S. Rokem, G. L. Petzold, B. D. Dayton, T. L. Wallace, A. H. Lin and W. C. Krueger, *Cancer Res.*, 1982, **42**, 2821-2828.
17. J. P. Parrish, D. B. Kastrinsky, S. E. Wolkenberg, Y. Igarashi and D. L. Boger, *J. Am. Chem. Soc.*, 2003, **125**, 10971-10976.
18. S. E. Wolkenberg and D. L. Boger, *Chem. Rev.*, 2002, **102**, 2477-2495.
19. K. S. MacMillan and D. L. Boger, *J. Med. Chem.*, 2009, **52**, 5771-5780.
20. E. A. Mullins, R. Shi and B. F. Eichman, *Nat. Chem. Biol.*, 2017, **13**, 1002-1008.
21. E. A. Mullins, R. Shi, Z. D. Parsons, P. K. Yuen, S. S. David, Y. Igarashi and B. F. Eichman, *Nature*, 2015, **527**, 254-258.
22. E. A. Mullins, J. Dorival, G. L. Tang, D. L. Boger and B. F. Eichman, *Nat. Commun.*, 2021, **12**, 6942.
23. P. G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani and R. Romagnoli, *Med. Res. Rev.*, 2004, **24**, 475-528.
24. N. Ghosh, H. M. Sheldrake, M. Searcey and K. Pors, *Current topics in medicinal chemistry*, 2009, **9**, 1494-1524.

25. D. G. Martin, C. Biles, S. A. Gerpheide, L. J. Hanka, W. C. Krueger, J. P. McGovren, S. A. Mizsak, G. L. Neil, J. C. Stewart and J. Visser, *J. Antibiot. (Tokyo)*, 1981, **34**, 1119-1125.
26. W. Huang, H. Xu, Y. Li, F. Zhang, X. Y. Chen, Q. L. He, Y. Igarashi and G. L. Tang, *J. Am. Chem. Soc.*, 2012, **134**, 8831-8840.
27. M. S. Tichenor, K. S. MacMillan, J. D. Trzupek, T. J. Rayl, I. Hwang and D. L. Boger, *J. Am. Chem. Soc.*, 2007, **129**, 10858-10869.
28. M. S. Tichenor and D. L. Boger, *Natural product reports*, 2008, **25**, 220-226.
29. T. Matsumoto, Y. Okuno and Y. Sugiura, *Biochem Biophys Res Commun*, 1993, **195**, 659-666.
30. N. Lomovskaya, S. K. Hong, S. U. Kim, L. Fonstein, K. Furuya and R. C. Hutchinson, *J. Bacteriol.*, 1996, **178**, 3238-3245.
31. A. Gatignol, H. Durand and G. Tiraby, *FEBS Lett*, 1988, **230**, 171-175.
32. T. W. Martin, Z. Dauter, Y. Devedjiev, P. Sheffield, F. Jelen, M. He, D. H. Sherman, J. Otlewski, Z. S. Derewenda and U. Derewenda, *Structure*, 2002, **10**, 933-942.
33. J. B. Biggins, K. C. Onwueme and J. S. Thorson, *Science*, 2003, **301**, 1537-1541.
34. S. Malla, N. P. Niraula, K. Liou and J. K. Sohng, *Microbiol. Res.*, 2010, **165**, 259-267.
35. J. Foulke-Abel, G. T. Kelly, H. Zhang and C. M. Watanabe, *Mol. Biosyst.*, 2011, **7**, 2563-2570.
36. J. M. Coughlin, J. D. Rudolf, E. Wendt-Pienkowski, L. Wang, C. Unsin, U. Galm, D. Yang, M. Tao and B. Shen, *Biochemistry*, 2014, **53**, 6901-6909.
37. X. Chen, N. P. Bradley, W. Lu, K. L. Wahl, M. Zhang, H. Yuan, X. F. Hou, B. F. Eichman and G. L. Tang, *Nucleic Acids Res.*, 2022, **50**, 2417-2430.
38. H. Yuan, J. Zhang, Y. Cai, S. Wu, K. Yang, H. C. S. Chan, W. Huang, W. B. Jin, Y. Li, Y. Yin, Y. Igarashi, S. Yuan, J. Zhou and G. L. Tang, *Nat. Commun.*, 2017, **8**, 1485.
39. H. Xu, W. Huang, Q. L. He, Z. X. Zhao, F. Zhang, R. Wang, J. Kang and G. L. Tang, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 10532-10536.
40. H. E. Krokan and M. Bjørås, *Cold Spring Harb. Perspect. Biol.*, 2013, **5**, a012583.
41. I. Alseth, T. Rognes, T. Lindback, I. Solberg, K. Robertsen, K. I. Kristiansen, D. Mainieri, L. Lillehagen, A. B. Kolsto and M. Bjoras, *Mol. Microbiol.*, 2006, **59**, 1602-1609.
42. E. H. Rubinson, A. S. Gowda, T. E. Spratt, B. Gold and B. F. Eichman, *Nature*, 2010, **468**, 406-411.
43. R. Shi, E. A. Mullins, X. X. Shen, K. T. Lay, P. K. Yuen, S. S. David, A. Rokas and B. F. Eichman, *EMBO J.*, 2018, **37**, 63-74.
44. R. Shi, X. X. Shen, A. Rokas and B. F. Eichman, *Bioessays*, 2018, **40**, e1800133.
45. E. A. Mullins, A. A. Rodriguez, N. P. Bradley and B. F. Eichman, *Trends Biochem. Sci.*, 2019, **44**, 765-781.
46. S. C. Brooks, S. Adhikary, E. H. Rubinson and B. F. Eichman, *Biochim. Biophys. Acta.*, 2013, **1834**, 247-271.
47. J. Dorival and B. F. Eichman, *Nucleic Acids Res.*, 2023, **51**, 2838-2849.
48. M. Wexler, F. Sargent, R. L. Jack, N. R. Stanley, E. G. Bogsch, C. Robinson, B. C. Berks and T. Palmer, *J. Biol. Chem.*, 2000, **275**, 16717-16722.
49. Y. C. Chen, C. L. Li, Y. Y. Hsiao, Y. Duh and H. S. Yuan, *Nucleic Acids Res.*, 2014, **42**, 10776-10785.
50. S. Ljungquist, T. Lindahl and P. Howard-Flanders, *J. Bacteriol.*, 1976, **126**, 646-653.
51. R. P. Cunningham, S. M. Saporito, S. G. Spitzer and B. Weiss, *J. Bacteriol.*, 1986, **168**, 1120-1127.
52. S. M. Kerins, R. Collins and T. V. McCarthy, *J. Biol. Chem.*, 2003, **278**, 3048-3054.

53. L. Aravind, D. R. Walker and E. V. Koonin, *Nucleic Acids Res.*, 1999, **27**, 1223-1242.
54. H. J. Atkinson, J. H. Morris, T. E. Ferrin and P. C. Babbitt, *PLoS One*, 2009, **4**, e4345.
55. S. Wu, X. H. Jian, H. Yuan, W. B. Jin, Y. Yin, L. Y. Wang, J. Zhao and G. L. Tang, *ACS Chem. Biol.*, 2017, **12**, 1603-1610.
56. S. Wang, K. Liu, L. Xiao, L. Yang, H. Li, F. Zhang, L. Lei, S. Li, X. Feng, A. Li and J. He, *Nucleic Acids Res.*, 2016, **44**, 187-197.
57. N. P. Bradley, K. L. Wahl, J. L. Steenwyk, A. Rokas and B. F. Eichman, *mBio*, 2022, **13**, e0329721.
58. R. Zallot, N. Oberg and J. A. Gerlt, *Biochemistry*, 2019, **58**, 4169-4182.
59. K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2021, **49**, W29-w35.

FIGURE LEGENDS

Figure 1. Effects of ytkR4 and ytkR5 on yatakemycin biosynthesis. (A) Structures of yatakemycin (YTM, **1**), CC-1065, gilvusmycin, and duocarmycins. (B) The biosynthetic gene cluster of YTM, showing only the outermost (*ytkA*, *ytkX*) and the resistance (*ytkR*) genes. Wavy lines represent biosynthesis genes not shown. (C) YTM hydrolysis products, **2** and **3**. (D) HPLC traces of fermentation products extracted from the wild-type TP-A0356 strain and Δ*ytkR4* and Δ*ytkR5* mutants. Each mutant was analyzed alone and complemented with the corresponding native *ytkR4* or *ytkR5* gene (pIB139 vector). The injected concentration of the extracted fermentation products from TP-A0356 is 1/20 that of the other strains. The detection wavelength used was 383 nm.

Figure 2. YtkR4 and YtkR5 exhibit 3′-5′ exonuclease and AP endonuclease activity. (A) Denaturing PAGE of time courses of YtkR4 and YtkR5 incubated with various 5′-FAM-labeled substrates. The location of the FAM label is marked with an asterisk (*). Black triangles designate the bands resulting from AP endonuclease activity. THF, tetrahydrofuran. Duplex DNA persists on the gel because of the high GC content (Supplementary Table S3). (B) Metal dependence for nuclease activity on a 5′-FAM-labeled double-stranded THF-DNA substrate. Reactions were carried out for 2 hr (YtkR4) or 90 min (YtkR5) and contained 10 µM protein, 100 nM DNA, and either 10 mM $MgCl_2$, 5 mM $CaCl_2$, 0.05 mM $ZnCl_2$, 3 mM $MnCl_2$, 1 mM $NiCl_2$, 1 mM $CuCl_2$, or 10 mM EDTA.

Figure 3. YtkR4 and YtkR5 have greater AP endonuclease activities on abasic sites generated by YtkR2 excision of YTM-adenine. (A) Nuclease activities of YtkR4 and YtkR5 on DNA substrates containing either THF (THF-DNA), an AP site generated by UDG excision of deoxyuracil (AP-DNA), or an AP site generated by YtkR2 excision of YTMA (YTM-DNA). Bands resulting from AP site cleavage are indicated by black triangles. The far-left panel shows non-enzymatic (NaOH) cleavage of the AP site generated from the YtkR2/YTMA-DNA glycosylase reaction. (B) Quantification of both AP endo and exonuclease products. Data are represented as the mean ± SD (n = 2). (C) Quantification of AP endonuclease products. Mean ± SD (n = 2).

Figure 4. YTMA lesions do not inhibit exonuclease activities of YtkR4 or YtkR5. (A) Denaturing PAGE of YtkR4 and YtkR5 nuclease products on unmodified and YTM-modified dsDNA in the presence of 10 mM $MgCl_2$. Mock, no enzyme control. The two lanes at the far right are controls to quantify the percentage of spontaneous depurination of YTMA (Mock + NaOH) and the percentage of DNA alkylated with YTM (YtkR2 + NaOH). (B) Quantification of the gel shown in panel a. Data are represented as the mean ± SD (n = 2).

Figure 5. YtkR4 and YtkR5 are found together in clusters. YtkR2 and YtkR5 subfamily distances to YtkR4. (A) Sequence similarity network (SSN) of BLASTp hits for YtkR5. Nodes are clustered by sequence similarity and colored based on cluster membership, with all clusters with three or fewer members colored light blue. The YtkR5 query sequence is

colored yellow. (B) Distance in base pairs in the genome between YtkR4 and YtkR2 or YtkR5 BLASTp hits. YtkR5 hits are further divided into subclasses based on the SSN, with colors corresponding to the SSN in panel a. The “YtkR5 other” category includes all proteins that were in a cluster with three or fewer members.

Figure 6. Two possible mechanisms for repair of a YTMA-DNA adduct. (A) The YtkR2 glycosylase initiates a BER pathway by excising YTM-adenine. The AP site product is incised by YtkR4 or YtkR5 to generate a 3′-OH, which is a substrate for 5′-deoxyribosephosphate displacement synthesis by a DNA polymerase. (B) The 3′-exonuclease activity of YtkR5, acting at a downstream nick in the DNA, is capable of removing the YTMA nucleotide.

Figure 1

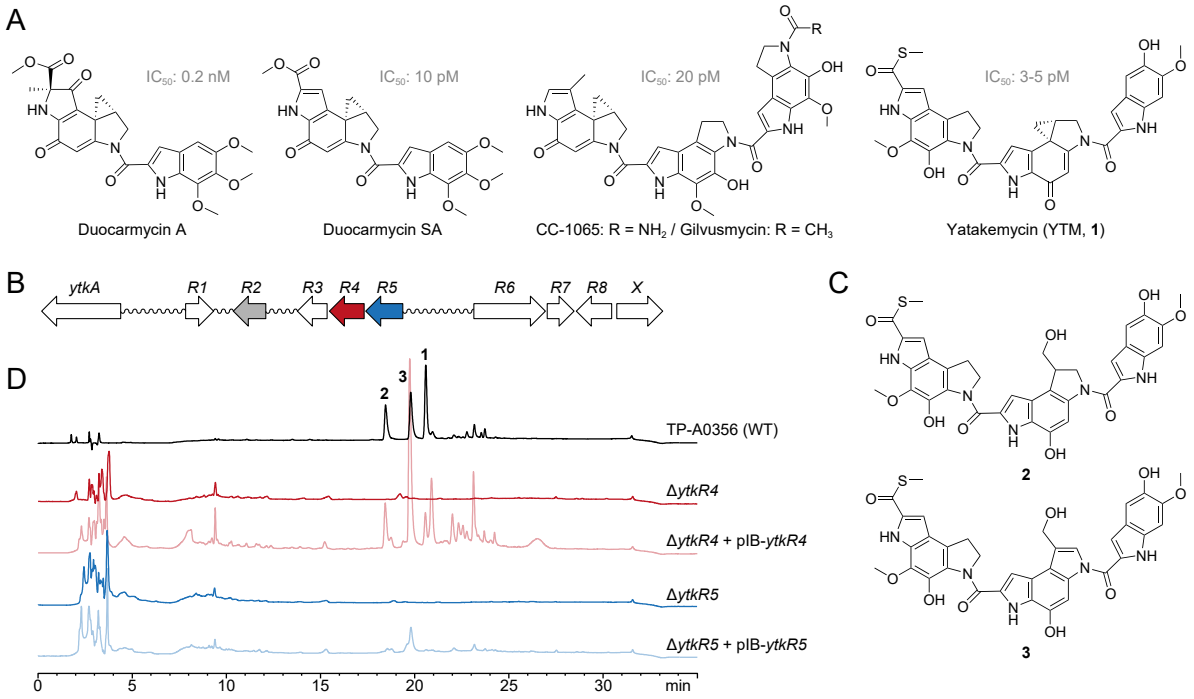

A
IC50: 0.2 nM
IC50: 10 pM
IC50: 20 pM
IC50: 3-5 pM
Duocarmycin A
Duocarmycin SA
CC-1065: R = NH2 / Gilvusmycin: R = CH3
Yatakemycin (YTM, 1)
B
ytkA
R1
R2
R3
R4
R5
R6
R7
R8
X
C
2
3
D
TP-A0356 (WT)
ΔytkR4
ΔytkR4 + pIB-ytkR4
ΔytkR5
ΔytkR5 + pIB-ytkR5
0
5
10
15
20
25
30
min

Figure 2

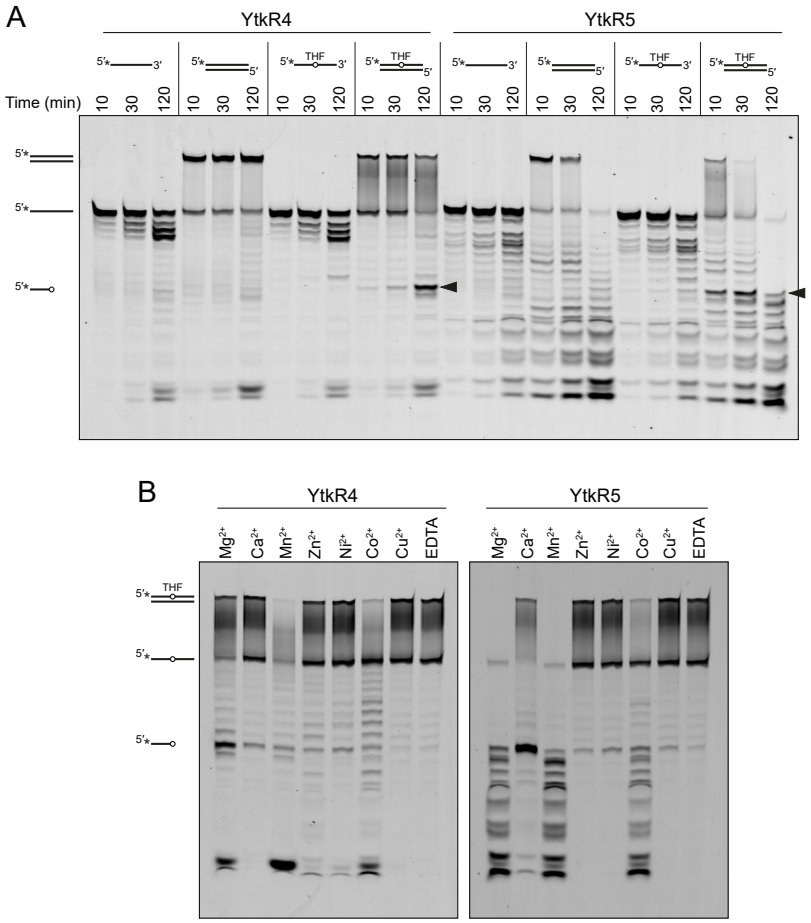

A
YtkR4
YtkR5
Time (min)
10
30
120
THF
B
Mg2+
Ca2+
Mn2+
Zn2+
Ni2+
Co2+
Cu2+
EDTA

Figure 3

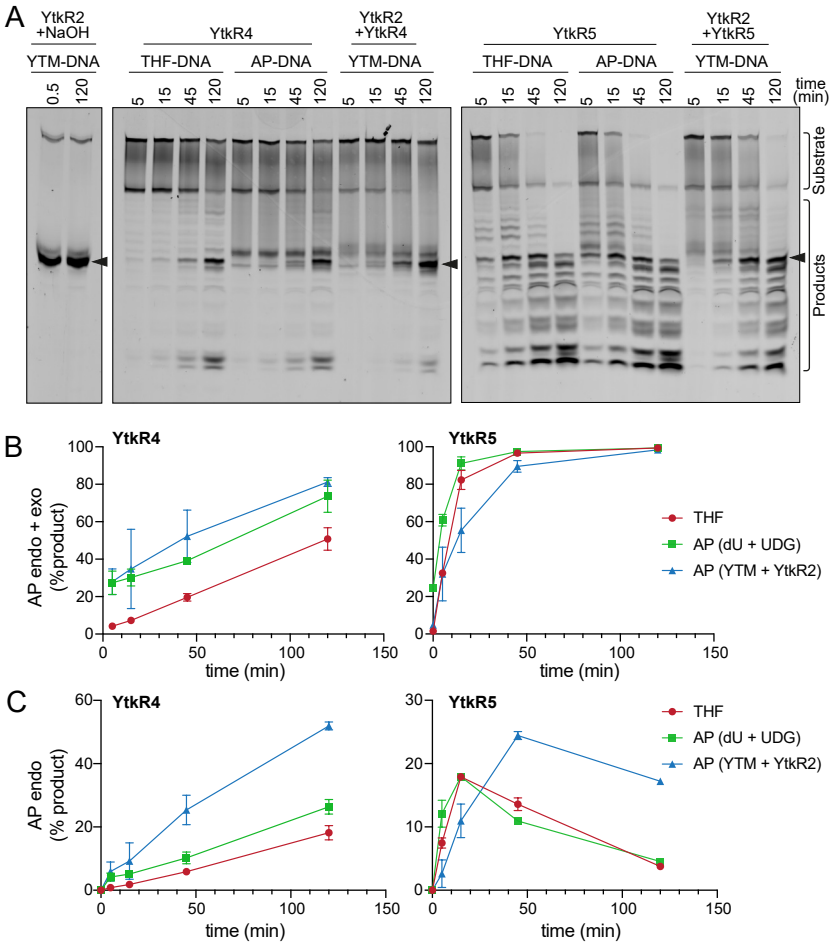

A
YtkR2 +NaOH
YtkR4
YtkR2 +YtkR4
YtkR5
YtkR2 +YtkR5
YTM-DNA
THF-DNA
AP-DNA
YTM-DNA
THF-DNA
AP-DNA
YTM-DNA
0.5
120
5
15
45
120
time (min)
Substrate
Products
B
YtkR4
YtkR5
AP endo + exo (%product)
time (min)
THF
AP (dU + UDG)
AP (YTM + YtkR2)
C
YtkR4
YtkR5
AP endo (% product)
time (min)
THF
AP (dU + UDG)
AP (YTM + YtkR2)

Figure 4

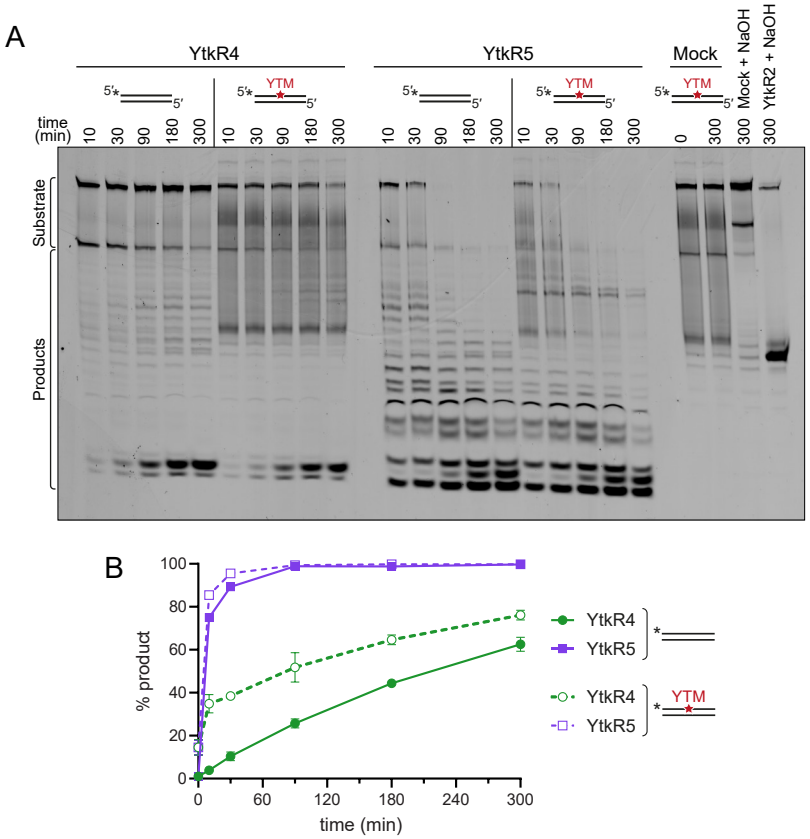

A
YtkR4
YtkR5
Mock
5'*
5'
YTM
time (min)
10
30
90
180
300
0
Mock + NaOH
YtkR2 + NaOH
Substrate
Products
B
% product
time (min)
YtkR4
YtkR5
*
YTM

Figure 5

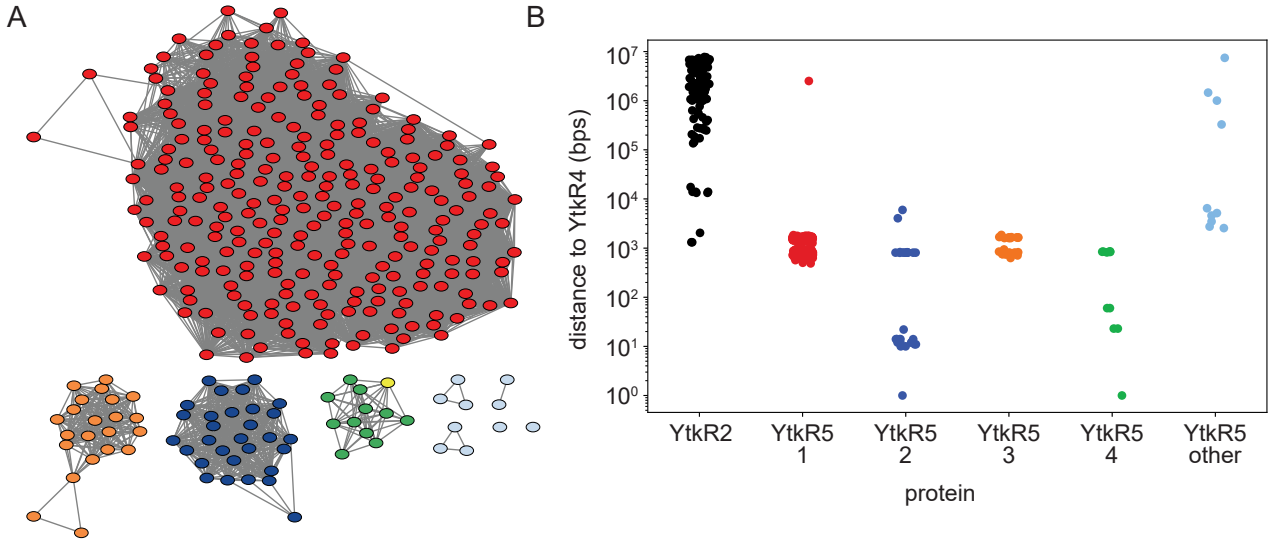
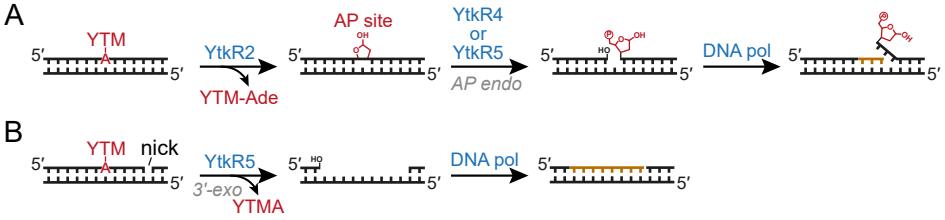

A
B
distance to YtkR4 (bps)
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
YtkR2
YtkR5 1
YtkR5 2
YtkR5 3
YtkR5 4
YtkR5 other
protein

Figure 6


A
YTM
5′
5′
YtkR2
YTM-Ade
AP site
5′
5′
YtkR4
or
YtkR5
AP endo
5′
5′
DNA pol
5′
5′
B
YTM
nick
5′
5′
YtkR5
3′-exo
YTMA
5′
5′
DNA pol
5′
5′