Molecular basis for DarT ADP-ribosylation of a DNA base

Marion Schuller[1], Rachel E. Butler[2], Antonio Ariza[1], Callum Tromans-Coia[1], Gytis Jankevicius[1,3], Tim D. W. Claridge[4], Sharon L. Kendall[5], Shan Goh[5] and Graham R. Stewart[2,*], Ivan Ahel[1,*]

[1]*Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom*

[2]*Department of Microbial Sciences, School of Biosciences and Medicine, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom*

[3]*Biozentrum, University of Basel, Basel CH - 4056, Switzerland*

[4]*Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford OX1 3TA, United Kingdom*

[5]*Centre for Emerging, Endemic and Exotic Disease, Pathology and Population Sciences, The Royal Veterinary College, Hawkshead Lane Hatfield Hertfordshire AL9 7TA, United Kingdom*

[*]Correspondence: ivan.ahel@path.ox.ac.uk; g.stewart@surrey.ac.uk

Key Words: ADP-ribosylation, ADP-ribosyltransferases, PARP, DNA modification, DNA repair, toxin-antitoxin system, *Mycobacterium tuberculosis*

ADP-ribosyltransferases (ARTs) utilise NAD^+ to catalyse substrate ADP-ribosylation[1], thereby regulating cellular pathways or contributing to toxin-mediated pathogenicity of bacteria[2–4]. Reversible ADP-ribosylation has traditionally been considered a protein-specific modification[5], but recent *in vitro* studies have suggested nucleic acids as targets[6–9]. Here, we present evidence that specific reversible DNA ADP-ribosylation on thymidine bases occurs *in cellulo* through the DarT/DarG toxin/antitoxin system which is found in a variety of bacteria including global pathogens such as *Mycobacterium tuberculosis,* EPEC and *Pseudomonas aeruginosa*[10]. We report the first DarT structure which identifies this protein as a diverged member of the PARP family. Moreover, a set of high-resolution structures in ligand-free, pre-and post-reaction states reveals a specialised mechanism of catalysis that includes a key active-site arginine, extending the canonical ART toolkit. Comparison with the well-established DNA-repair protein ADP-ribosylation complex, PARP/HPF1, offers insights into how the DarT class of ARTs evolved into specific DNA-modifying enzymes. Together, the structural and mechanistic data provide unprecedented detail for a PARP family member and contribute to fundamental understanding of nucleic acid ADP-ribosylation. We furthermore show that thymine-linked ADP-ribose DNA adducts reversed by DarG antitoxin, functioning as non-canonical DNA-repair factor, are utilised not only for targeted DNA damage to induce toxicity but also as a signalling strategy for cellular processes. Using *M. tuberculosis* as an exemplar we show that DarTG regulates growth by DNA ADP-ribosylation at the origin of chromosome replication.

DarT is a divergent member of the PARP family.

The DNA ADP-ribosyltransferase 'DarT' was discovered as bacterial ADP-ribosyltransferase ('ART') toxin encoded in the toxin/antitoxin ('TA') system DarTG, which catalyses the sequence-specific ADP-ribosylation of DNA[10]. *In vitro*, DarT has been shown to transfer ADP-ribose from NAD^+ onto thymidine bases present in single-stranded DNA (ssDNA) specifically at the four base motif TNTC, thereby showing no activity on RNA or protein targets[10]. The reaction is reversed by the antitoxin DNA ADP-ribosylglycohydrolase 'DarG' via the (ADP-ribosyl) hydrolase activity of its macrodomain (Figure 1A). Initial modelling and phylogenetic analysis suggested DarT as being distinct from other bacterial diphtheria toxin-like ADP-ribosyltransferases ('ARTDs')[10,11] and closer to eukaryotic ARTD members, referred to as poly(ADP-ribose)polymerases ('PARPs') (Figure 1B). For further clarification and insight into its function, we determined the first structures of DarT from *Thermus* sp. 2.9 in ligand-free and NAD^+-bound form, both to a resolution of 1.3 Å (Extended Data Figure 1, Extended Data Table 1). Secondary structure analysis confirmed the close similarity of DarT to ARTD family members, and PARPs in particular, with its fold-stabilising central 6-stranded β-sheet core and the ARTD-conserved helices between strand β1-2 and β2-3[11] (Extended Data Figure 1B), yet also highlighted structural differences. Most notably, the N-terminal extension of the β-sheet core found in PARPs, i.e. a strand-helix-strand arrangement next to β6, was found to be spatially replaced in DarT with a shorter C-terminal helix-strand extension (Extended Figure Data 1C). The NAD^+ substrate is bound by DarT as generally observed for ARTs[11,12] in a constrained conformation bent over the central split of the β-sheet core and is shielded from the solvent phase by a highly flexible loop-helix element (residues S35 to R53) (Extended Figure Data 1D), which structurally corresponds to the 'donor loop' described for other ART family members[1,13,14]

DarT links ADP-ribose to the thymidine base nitrogen.

In order to understand the mechanism of DNA ADP-ribosylation itself, we determined the structures of DarT of *Thermus* sp. 2.9 E160A in complex with the following ligands to 1.46-1.66 Å (Extended Data Table 1): (i) ADP-ribosylated ssDNA 5mer, 'ADPr-DNA', (ii) NAD^+ with ssDNA 5mer and (iii) carba-NAD^+ (a non-hydrolysable NAD^+ analogue) with ssDNA 5mer (Figure 1C, Extended Data Figure 2).

Overlaying the three structures showed that DNA binding stabilises the DarT fold for the ADP-ribosylation reaction, since previously disordered regions within the loop-helix element, including the NAD^+ 'donor loop', were now fully resolved. The DNA target is bound in a solvent-accessible groove orthogonally to the NAD^+ molecule while stabilised by the ADP-ribosylating turn-turn ('ARTT') loop which was suggested to contribute to substrate specificity in other ARTDs[15]. The ARTT loop is significantly longer in DarT than in other bacterial ARTDs but comparable to the ones in human ARTDs, i.e. PARP 1, 2 and 3, thereby forming a stable scaffold for the DNA target (Extended Data Figure 3A). Notably, as PARPs 1-3 have also been shown to catalyse DNA ADP-ribosylation[6,7,16,17], their extended ARTT loop may also be related to this activity.

The carba-NAD^+:DNA-bound DarT structure showed both ligands to be physically separate but positioned for linkage, thus representing a putative pre-reaction state (Figure 1C-left). Intriguingly, in the NAD^+ and DNA co-crystal structure, the ADPr-DNA product of DarT was identified despite the catalytic E160A mutation, thus its post-reaction state was captured with the nicotinamide ('NAM') reaction by-product still present in the binding site (Figure 1C-right). *In vitro* assays confirmed that DarT E160A still possesses weak ADP-ribosylation activity, rationalising the observed reactivity *in crystallo* (Extended Data Figure 2E). Since all complexed ligands were well-resolved in the high-resolution electron density map (Extended Figure Data 2C,D), this allowed us to unambiguously reveal the so far unknown atomic ADPr-linkage to the DarT-targeted thymidine. Thus, DarT catalyses DNA ADP-ribosylation by linking ADP-ribose at the NAM ribose C1" to the in-ring nitrogen N3 of the thymidine base. The stereocentre at the C1" atom has an α configuration suggesting anomeric inversion from the β-NAD^+-substrate as it is generally observed as a consequence of target ADP-

ribosylation by ARTs[18] (Figure 1C,D, Extended Data Figure 2B, Supplementary Figure 2, Supplementary Discussion).

Sequence-specific ADP-ribosylation by DarT

Mapping surface electrostatics and sequence conservation onto DarT showed that the DNA target is bound within a groove that is highly conserved among the DarT family and lined up with a positive electrostatic surface generated by several basic residues (Figure 2A). The five nucleotides of the co-crystallised DNA target span the entire groove, with the thymidine targeted for ADP-ribosylation pointing orthogonally to the DNA backbone deep into the active site of DarT (Figure 2B).

A series of interactions between the DNA fragment and DarT side-chains (Y44, H46, R50, R75, Y80, R154, Q158) and main-chains, in addition to structural waters, stabilise the phosphate-ribose backbone. The central R154 side-chain is thereby of particular importance through its coordination of three phosphate groups around the thymidine target site (Figure 2B). Furthermore, the DarT-specific four-base motif preference, i.e. TNTC for DarT of *Thermus* sp. 2.9 (and *Thermus aquaticus*[10]) can now be rationalised. The high specificity of DarT for the first thymine in the motif[10] is due to the recognition of all of its functional groups by forming hydrogen bonds with surrounding backbone amides and the R75 side-chain (Figure 2B-inset). Conversely, the second base does not show any specific interactions (Figure 2B), consistent with the flexibility of *Thermus* sp. 2.9 and *Thermus aquaticus* DarT with respect to the base in this position (Extended Data Figure 3B)[10]. The DNA-complex structures may also explain the preference of DarT for DNA over RNA modification[10] since the 2' hydroxyl groups present in RNA would likely lead to clashes within the DNA-binding site, while the methyl group of the active site thymine (absent in the corresponding uracil base) may also help in the steric orientation of the base for ADPr-linkage (Extended Data Figure 3C).

The high coordination of DNA binding by DarT was confirmed through mutagenesis studies. Most of *Thermus aquaticus* DarT mutants with single amino acids substitutions corresponding to *Thermus* sp. 2.9 DNA-binding residues (Extended Data Figure 4A) inhibited bacterial growth and thus still exhibited toxicity (Figure 2C), suggesting that individual mutations could not efficiently disrupt the extensive protein-DNA interface. This was consistent with the *in vitro* ADP-ribosylation activity of selected mutants (Figure 2D). Only mutagenesis of the central R154 to a tryptophan residue resulted in complete loss of ADP-ribosylation activity, predictably by sterically preventing DNA-binding. Finally, the numerous hydrogen bond interactions between DarT and ssDNA significantly increase the binding affinity for its ADPr-DNA product (K_D=961 nM) compared to the NAD^+ substrate (K_D=58 μM) (Figure 2E, Extended Data Figure 5D, Supplementary Figure 3), yet, the close to micromolar affinity also eases its release from DarT, enabling a higher substrate-product turnover.

The catalytic mechanism of DarT

Prior to catalysing DNA ADP-ribosylation, DarT binds its NAD^+ substrate with key interactions resulting in a constrained conformation (Extended Data Figure 5C, Supplementary Discussion). Analysis of reaction products after incubation of DarT in presence or absence of its DNA target, could not detect any NADase activity, auto-ADP-ribosylation activity or free ADP-ribose generation by DarT (Extended Data Figure 5E). This would suggest an efficient turnover of NAD^+-cleavage and ADPr-linkage in addition to NAD^+-cleavage happening only upon DNA binding. Structural analysis combined with DarT residue conservation and mutagenesis studies (Figure 3A,B,C,D, Extended Data Figure 4B) revealed a coordinated action of several residues in the active site that is required for DNA ADP-ribosylation catalysis. This includes (i) locking of the thymidine base in plane for reaction by H119, (ii) polarisation of the NAD^+ molecule for cleavage supported by Y71 and R51, (iii) stabilisation of the oxocarbenium ion resulting from NAD^+-cleavage by M78 and particularly (iv) proton abstraction from N3 of the thymidine base by R51 which finally enables the ADPr-linkage by a nucleophilic attack of the oxocarbenium ion in a S_N1-type reaction (Figure 3E). Thereby, the guanidinium group of R51 appears to be functioning as catalytic base, with the required prior

deprotonation and pK_a lowering being a consequence of a potential interaction with E160 (Extended Data Figure 3D), i.e. the ART-conserved catalytic glutamate. Interestingly, the traditional catalytic role of this glutamate in NAD^+ polarisation seems to be additionally taken over by R51 in DarT (Figure 3B). This is accomplished by R51 via the high flexibility of its side-chain, allowing it to adopt different and well-defined orientations in *apo* (for potential E160 interaction), NAD^+-bound (for C1"-ribose – N1 NAM bond polarisation) and DNA-bound (for thymine proton abstraction) state (Figure 3E, Extended Data Figure 3E, Supplementary Discussion).

DarT acts as a specific ADP-ribosyltransferase *in cellulo*.

Next, we visualised specific DarT ADP-ribosylation activity in cells on its physiological target, genomic DNA (gDNA). For this, we identified the first antibody capable of detecting ADP-ribose modifications on DNA (Extended Data Figure 6A, methods section). Expression of *Thermus aquaticus* and EPEC (a pathogenic *E. coli* strain) DarT in *E. coli* led to strong ADP-ribosylation of gDNA when compared to the characterised catalytic mutants[10,19] (Extended Data Figure 6B,C). DarT DNA ADP-ribosylation is perceived as severe DNA damage; leaving ADP-ribosylated sites unrepaired was shown to stall DNA replication and activates the DNA damage response[19] which is indicated by up-regulated RecA levels as in the DarT over-expressing cells (Extended Data Figure 6B). The phenotype could be rescued by co-expression with the cognate wild-type *Thermus aquaticus* DarG antitoxin and DarG homologues from EPEC and *M. tuberculosis* (Extended Data Figure 6D). In accompanying *in vitro* experiments, the ADP-ribose modifications on *Thermus aquaticus* DarT ADP-ribosylated gDNA could be removed by *Thermus aquaticus* DarG and DarG macrodomains from non-cognate species but not by the human (ADP-ribosyl) hydrolases ARH3 or the macrodomain-containing proteins MacroD1 and PARG (Extended Data Figure 6E). This positions DarG as a non-canonical DNA repair enzyme specific for thymine-linked DNA ADP-ribose-adducts. Together, this confirmed the targeted introduction of DNA damage by DarT as well as the specific removal of those DNA adducts and consequently repair by its DarG antitoxin partner.

DarT DNA ADP-ribosylation in *Mycobacterium tuberculosis*

We then characterised DNA ADP-ribosylation by DarT in an endogenous system using bacteria of the *M. tuberculosis* complex as an exemplar. By silencing DarG and thus deregulating DarT, the mycobacterial DarT was confirmed to perform gDNA ADP-ribosylation (Figure 4A) which led to profound induction of the DNA damage response including the prominent DNA damage markers RecA and DnaE2 amongst a regulon that incorporated all of the genes typically induced by the DNA-damaging agent mitomycin C (Figure 4B, Extended Data Figure 7A,B). It is notable that DarT and DarG are themselves intrinsic parts of the DNA damage response, being transcriptionally linked to the *dnaB* gene (Extended Data Figure 7C) in an operon that is upregulated following DNA damage (Figure 4B). Aside from its role in DNA repair, DnaB is the replicative helicase which interacts with ssDNA at the chromosome origin ('OriC') to initiate and then drive DNA branch migration during replication. We demonstrated that DarT preferentially ADP-ribosylates ssDNA with the motif TTTW which occurs densely in the AT-rich DnaB-loading region of the *M. tuberculosis* OriC (Figure 4C, Extended Data Figure 7D). The linked expression of DarT and DnaB combined with the potential for shared ssDNA substrates suggested that DarTG may be involved in control of replication by ADP-ribosylation of OriC. Indeed, *M. tuberculosis* DarT strongly ADP-ribosylates the OriC *in vitro* with preference at the TTTT and TTTA motifs in the lower strand (Figure 4C, Extended Data Figure 7E) and moreover, we confirmed that unregulated DarT activity targets the OriC for ADP-ribosylation *in cellulo* (Figure 4D). The physiological role of DarTG in growth control is further supported by experiments showing that unregulated DarT arrests the growth of mycobacteria (Figure 4E); knockout of *darTG* increases growth (Figure 4F) with a 5.2 fold (± 3.3, 95% CI) competitive advantage over wild-type *M. tuberculosis*; and disruption of *darT* by transposition confers increased growth in genome-wide TnSeq mutagenesis studies (Figure 4G)[20].

DISCUSSION

Our data provides the first evidence for the existence of specific reversible DNA ADP-ribosylation *in cellulo*. Effectively, the DarTG TA system serves as a DNA damage/repair module where DarG plays the role of a non-canonical DNA repair enzyme that specifically removes ADP-ribosylated DNA adducts, thereby counteracting DarT activity. An analogous, reversible DNA repair system has been seen for aprataxin DNA repair factors in eukaryotes. However, in this case the DNA nucleotidylation (adenylation) does not happen in a controlled, sequence-specific manner but instead is a consequence of abortive DNA ligation reactions[21,22].

Bacterial toxin/antitoxin systems often function as genetic stability factors, preserving the DNA molecule upon which they reside but their biochemical activities can be co-opted by 'host' bacteria for other cellular purposes[23]. In this study, we showed that in *M. tuberculosis, darTG* is co-expressed with a main replicative helicase (*dnaB*) that is under the control of DNA damage inducible promoters in mycobacteria[24]. We furthermore demonstrated that DarT is important for control of bacterial growth by ADP-ribosylation of ssDNA at the origin of chromosomal replication; carefully controlled, slow and non-replicating growth states are key for *M. tuberculosis* resulting in persistent, potentially life-long infection and antibiotic tolerance. The molecular structures presented here enable drug design and development for DarT inhibition which could potentially be a strategy to target persistent and phenotypically antibiotic tolerant tuberculosis.

Our DarT structures present the first example of a PARP-like ART captured in pre- and post-reaction states with a ssDNA target, revealing the molecular basis for specific DNA recognition and ADP-ribosylation and providing insights on how the DarT class of ARTs evolved into specific DNA-modifying enzymes (Extended Data Figure 8, Supplementary Discussion). Comparison with the PARP/HPF1 protein ADP-ribosylation complex[25] reveals conservation of spatial position and orientation of mechanistically relevant residues among the ART family and exposes the striking evolutionary adaption of ARTs for the specific recognition of different and unrelated macromolecular targets. Beyond DarTG, ADP-ribosylation of DNA/RNA has also been proposed for eukaryotic PARP family members[6–8] and the established method presented in our study which enables the visualisation of DNA ADP-ribosylation *in cellulo* may foster further studies of this modification. We predict that ADP-ribosylation of nucleic acids represents a common, but largely unknown aspect of ADP-ribosylation signalling and that it will become a new and exciting area in the fields of DNA damage response, epigenetics and beyond.

REFERENCES

1. Barkauskaite, E., Jankevicius, G. & Ahel, I. Structures and Mechanisms of Enzymes Employed in the Synthesis and Degradation of PARP-Dependent Protein ADP-Ribosylation. *Mol. Cell* **58**, 935–946 (2015).
2. Gibson, B. A. & Kraus, W. L. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat. Rev. Mol. Cell Biol.* **13**, 411–424 (2012).
3. Cohen, M. S. & Chang, P. Insights into the biogenesis, function, and regulation of ADP-ribosylation. *Nat. Chem. Biol.* **14**, 236–243 (2018).
4. Perina, D. *et al.* Distribution of protein poly(ADP-ribosyl)ation systems across all domains of life. *DNA Repair* **23**, 4–16 (2014).
5. Kraus, W. L. PARPs and ADP-Ribosylation: 50 Years . . . and Counting. *Mol. Cell* **58**, 902–910 (2015).
6. Talhaoui, I. *et al.* Poly(ADP-ribose) polymerases covalently modify strand break termini in DNA fragments in vitro. *Nucleic Acids Res.* **44**, 9279–9295 (2016).
7. Munnur, D. & Ahel, I. Reversible mono-ADP-ribosylation of DNA breaks. *FEBS J.* **284**, 4002–4016 (2017).
8. Munnur, D. *et al.* Reversible ADP-ribosylation of RNA. *Nucleic Acids Res.* **47**, 5658–5669 (2019).
9. Groslambert, J., Prokhorova, E. & Ahel, I. ADP-ribosylation of DNA and RNA. *DNA Repair* **105**, 103144 (2021).
10. Jankevicius, G., Ariza, A., Ahel, M. & Ahel, I. The Toxin-Antitoxin System DarTG Catalyzes Reversible ADP-Ribosylation of DNA. *Mol. Cell* **64**, 1109–1116 (2016).
11. Aravind, L., Zhang, D., de Souza, R. F., Anand, S. & Iyer, L. M. The Natural History of ADP-Ribosyltransferases and the ADP-Ribosylation System. in *Endogenous ADP-Ribosylation* (ed. Koch-Nolte, F.) 3–32 (Springer International Publishing, 2015). doi:10.1007/82_2014_414
12. Bell, C. E. & Eisenberg, D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Adv. Exp. Med. Biol.* **419**, 35–43 (1997).
13. Steffen, J. D., Brody, J. R., Armen, R. S. & Pascal, J. M. Structural implications for selective targeting of PARPs. *Front. Oncol.* **3**, 1–14 (2013).
14. Vyas, S. *et al.* Family-wide analysis of poly(ADP-ribose) polymerase activity. *Nat. Commun.* **5**, 1–13 (2014).
15. Hottiger, M. O., Hassa, P. O., Lüscher, B., Schüler, H. & Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* **35**, 208–219 (2010).
16. Belousova, E. A., Ishchenko, A. A. & Lavrik, O. I. DNA is a New Target of Parp3. *Sci. Rep.* **8**, 1–12 (2018).
17. Zarkovic, G. *et al.* Characterization of DNA ADP-ribosyltransferase activities of PARP2 and PARP3: New insights into DNA ADP-ribosylation. *Nucleic Acids Res.* **46**, 2417–2431 (2018).
18. Ueda, K. ADP-RIBOSYLATION. *Ann. Rev. Biochem.* **54**, 73–100 (1985).
19. Lawarée, E. *et al.* DNA ADP-Ribosylation Stalls Replication and Is Reversed by RecF-Mediated Homologous Recombination and Nucleotide Excision Repair. *Cell Rep.* **30**, 1373–1384 (2020).
20. Mendum, T. A. *et al.* Transposon libraries identify novel Mycobacterium bovis BCG genes involved in the dynamic interactions required for BCG to persist during in vivo passage in cattle. *BMC Genomics* **20**, 1–13 (2019).
21. Ahel, I. *et al.* The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature* **443**, 713–716 (2006).
22. Tumbale, P. *et al.* Structure of an aprataxin-DNA complex with insights into AOA1 neurodegenerative disease. *Nat. Struct. Mol. Biol.* **18**, 1189–1195 (2011).
23. Harms, A., Brodersen, D. E., Mitarai, N. & Gerdes, K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol. Cell* **70**, 768–784 (2018).
24. Gamulin, V., Cetkovic, H. & Ahel, I. Identification of a promoter motif regulating the major DNA damage response mechanism of Mycobacterium tuberculosis. *FEMS Microbiol. Lett.* **238**, 57–63 (2004).
25. Suskiewicz, M. J. *et al.* HPF1 completes the PARP active site for DNA damage-induced ADP-ribosylation. *Nature* **579**, 598–602 (2020).

LEGENDS FOR MAIN TEXT FIGURES

Figure 1. DarT – a PARP-like protein catalysing DNA ADP-ribosylation at the thymidine base nitrogen.

(A) Schematic representations of the DarTG Toxin-Antitoxin operon (**top**) and the DarT-catalysed reaction (**bottom**).

(B) Dendrogram of representative members of the ART superfamily. DarT clusters in the ARTD family, distinctly away from bacterial ARTDs and close to the human PARP family. (Ecto-ARTs: extracellular membrane-associated ADP-ribosyltransferases)

(C) Co-crystal structures of *Thermus* sp. 2.9 DarT E160A in substrate-bound state (**left**) and product-bound state after NAD^+-cleavage and reaction *in crystallo* (**right**). (**Middle**) Overlay focusing on the ADPr-thymine linkage. The substrate binding ('ARTT') loop is highlighted in green, the NAD^+-binding loop-helix element in purple (set for clarity in higher transparency). The catalytic glutamate E160 conserved in ARTs is modelled as red sticks.

(D) NMR ^{1}H-^{13}C HMBC spectrum (**left**) and schematic representation (**right**) showing key correlations establishing the connectivity between the NAM ribose C1" and N3 of the thymidine base.

Figure 2. DNA binding is highly coordinated by DarT for site-specific DNA ADP-ribosylation.

(A) Amino-acid residue conservation (**top**) and surface electrostatic potential (**bottom**) mapped onto the surface of *Thermus* sp. 2.9 DarT E160A in the carba-NAD^+:DNA-bound state.

(B) Cartoon-stick model showing DNA substrate coordination by *Thermus* sp. 2.9 DarT E160A with side- and main-chain interactions (dashed lines) including water (red spheres) contacts. The first thymidine base of DarT's DNA specificity motif is shown in the circle.

(C) Toxicity assay monitoring growth of DH5α under repression (glucose) and induction (arabinose) of expression of *Thermus aquaticus* DarT WT and DNA-binding mutants. *Thermus aquaticus* DarT residues are given in white labelling, corresponding residues of *Thermus* sp. 2.9 DarT in black. Representative of three biologically independent experiments.

(D) *In vitro* ADP-ribosylation activity of *Thermus aquaticus* DarT DNA-binding mutants compared to WT. Representative of three independent experiments.

(E) Integrated thermogram obtained by ITC giving ADPr-DNA binding parameters for *Thermus* sp. 2.9 DarT E160Q. Representative result from three independent experiments is shown, with the number of binding sites N and the dissociation constant K_D calculated from the repeats (mean±SD).

Figure 3. Mechanism of DNA ADP-ribosylation.

(A) Cartoon-stick model showing the active site of *Thermus* sp. 2.9 DarT E160A before (carba-NAD^+:DNA-bound, **left**) and after (ADPr-DNA:NAM-bound, **right**) ADPr-linkage. Active site residues relevant for the catalytic mechanism are highlighted in green with the catalytic glutamate E160 modelled in red.

(B) Comparison of the NAM ribose coordination of NAD^+ (cyan) and carba-NAD^+ (brown) relevant for NAD^+ substrate polarisation with interactions (dashed lines) shown in corresponding colours. Carba-NAD^+:DNA-bound structure (grey) is shown in overlay with the NAD^+-bound structure (cyan, interacting residues only).

(C) Toxicity assay monitoring growth of DH5α under repression (glucose) and induction (arabinose) of expression of *Thermus aquaticus* DarT active site mutants compared to WT. *Thermus aquaticus* DarT residues are given in white labelling, corresponding residues of *Thermus* sp. 2.9 DarT in black. Representative of three biologically independent experiments.

(D) *In vitro* ADP-ribosylation activity of *Thermus aquaticus* DarT active site mutants compared to WT. Representative of three independent experiments.

(E) Proposed molecular mechanism for catalysis of DNA ADP-ribosylation by DarT. The conformational dynamics of R51 of twists and flip is indicated by orange arrows.

Figure 4. gDNA ADP-ribosylation in *Mycobacterium tuberculosis.*

(A) *M. bovis* BCG DarT ADP-ribosylates gDNA *in cellulo* which is reversible by the *M. tuberculosis* DarG macrodomain. (ATC: anhydrotetracycline inducing CRISPRi knockdown of DarG). Representative of each of seven biologically independent repeats.

(B) Unregulated DarT activity (*darG* knockdown; **left**) in *M. bovis* BCG leads to profound induction of DNA damage response (DDR) while *darTG* is itself induced alongside *dnaB* as part of the DDR (mitomycin treatment; **right**). Volcano plots from RNA-Seq data show up (purple)- and down (blue)-regulated genes in respective conditions. The presence/ absence of ADPr-gDNA in those conditions is shown in the inset (**middle below**). Values are mean of three independent replicates.

(C) DarT ADP-ribosylates the OriC *in vitro* which is reversible by the *M. tuberculosis* DarG macrodomain. OriC oligo sequences are provided with potential DarT modification sites highlighted in purple. The preferred recognition motif of *M. tuberculosis* DarT is TTTW (underlined). Representative of three independent experiments.

(D) Unregulated DarT activity (*darG* knockdown) modifies the OriC *in cellulo*. ADP-ribosylated gDNA was immunoaffinity purified and quantified by qPCR. Data are presented as mean±SD, n = 3 biologically independent samples, ***p = 0.0009 by unpaired, two-tailed Student's *t* test. Representative of three independent experiments.

(E) Unregulated DarT activity (*darG* knockdown) leads to growth arrest in *M. bovis* BCG. Data are presented as mean±SD, n = 3 biologically independent samples. Representative of three independent experiments.

(F) Knockout of *darTG* in *M. tuberculosis* provides growth advantage. Data are presented as mean±SD, n = 3 of biologically independent samples, ***p = 0.0001 by unpaired, two-tailed Student's *t* test. Competitive advantage calculated at 14 days, N = 8 independent replicates (see main text).

(G) Tracking abundance of mutants in a *M. bovis* BCG transposon mutant library (TnSeq) demonstrates a growth advantage for *darT*-disrupted bacteria. Data from Mendum *et al.*[20].

METHODS

Materials, reagents and chemicals

The genes encoding *Thermosipho africanus* DarTG and *Thermus* sp. 2.9 DarT were synthesised by Thermo Scientific. *E. coli* bacterial strains were purchased from Merck Millipore. High-fidelity DNA polymerase Phusion and cloning reagents were obtained from New England Biolabs and Thermo Scientific. All DNA primers and ssDNA substrates (Supplementary Table 1 and 2) were synthesized by Thermo Scientific. Crystallisation screens were procured from Hampton Research. Carba-NAD^+ was synthesised by Hangzhou YiLu Biological technology Co., LTD. All remaining chemicals were purchased from Sigma unless stated otherwise. The antibody used for detecting the ADPr modification on gDNA was selected from screening all commercially available anti-ADP-ribose antibodies and the only one identified to be suitable and used in this study for this purpose is the mono-clonal "Poly/Mono-ADP Ribose (E6F6A)" rabbit antibody from Cell Signalling Technology (product #83732).

Constructs

The gene encoding full-length (FL)-DarTG E152A of *Thermosipho africanus* (residues 1-388) was cloned into a pET28a expression vector with His_6-N-terminal tag and TEV protease recognition site. The gene encoding FL-DarT E160A of *Thermus* sp. 2.9 (residues 1-209) was cloned into a pNIC28-Bsa4 expression vector, which adds an N-terminal His_6-tag and a TEV protease recognition site for removal of the tag. *Thermus aquaticus* DarT (FL) previously cloned[10] into a pBAD33 expression vector containing a ribosomal binding site and a N-terminal His_6-TEV cleavage site-V5 tag was used for biochemical studies. *Thermus* sp. 2.9 *darT* (FL) and *Mycobacterium tuberculosis darT* (FL) were cloned into a pBAD33 expression vector adding a N-terminal His_6-TEV cleavage site using the 'DH5α-macro' strain (as described below). Expression constructs encoding EPEC DarT G49D and E170A and EPEC DarG macrodomain as previously described[19] were provided by Christoph Tang Laboratory (University of Oxford). Mutations were introduced using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent). All plasmids were verified by Sanger sequencing. The constructs used in this study are summarised in the Supplementary Table 3.

Construction of sgRNA expression plasmids and strains

pRH2502 (an integrative plasmid expressing dCas9Spy from a Tet-regulated promotor), and pRH2521 (expressing the sgRNA scaffold from a Tet-regulated promoter), are described in Singh *et al*.[26]. sgRNAs targeting *darG* were designed as previously described[27]. A protospacer adjacent motif (PAM) site, "NGG", downstream of the start codon was identified, and 20 nucleotides downstream selected as genome specific sgRNA. Complementarity to other regions of the genome was assessed using the basic logical alignment tool (BLAST), demonstrating a full-length match specific to *darG*. A full length transcribed sgRNA including the terminators and dCas9 handle was designed, and M-fold was used to predict the secondary structure of the full length sgRNA transcript, confirming that the sgRNA was predicted to fold into the dCas9Spy and terminator hairpin loops. Complementary forward and reverse primers corresponding to the 20nt sequence (without the PAM) with appropriate ends for ligation into the pRH2521 vector were designed. Oligos were annealed and cloned into CRISPRi plasmids using BbsI (NEB) as previously described[26]. 1 µg of pRH2502 (dCas9Spy integrative vector) was electroporated into electrocompetent mycobacterial strains, which were selected in the presence of 25 µg/ml kanamycin, then further electroporated with 1 µg pRH2521 expressing *darG*-sgRNA and selected with hygromycin (50 µg/ml).

For generation of *M. tuberculosis* Δ*darTG*, regions of DNA flanking the *darT* (Rv0059) and *darG* genes (Rv0060) were PCR-amplified from genomic DNA using respective primer pairs (see Supplementary Table S4) for up- and downstream regions, respectively. The regions were cloned around the hygromycin-resistance gene (*hyg*) in the suicide delivery vector pG5. pG5 carries the *sacB* gene to provide counterselection for single-crossover integration of the gene-replacement vector. The resulting plasmid, pG5-RV59-60-KO, was introduced into *M. tuberculosis* via electroporation and gene replacement transformants were selected on 7H11 containing hygromycin at 50 µg/ml and 2% sucrose. Gene replacement was confirmed by PCR.

To combat the high toxicity of wild-type *M. tuberculosis* DarT for amenability to cloning and protein expression, the macrodomain sequence encoding residues 1-155 of *Thermus aquaticus* DarG was chromosomally integrated into *E. coli* DH5α (NEB) as described by St-Pierre *et al*. [28]. The DarG macrodomain DNA fragment was cloned into the 'One-Step Integration Plasmid' (pOSIP) encoding a kanamycin resistance marker by Gibson assembly

for integration at the phage P21 integration site. The integration module and antibiotic resistance marker, flanked by FRT sites, were removed using pE-FLP for FLP recombinase-mediated excision. Integration of *darG* macrodomain was verified by Sanger Sequencing, resulting in strain 'DH5α-macro'.

Mycobacterial strains and culture methods

M. bovis BCG and *M. tuberculosis* GC1237 were maintained on Middlebrook 7H11 solid medium containing 0.5% glycerol and 10% oleic acid-albumin-dextrose-catalase (OADC) supplement. Broth cultures were in Middlebrook 7H9 supplemented with 0.05% Tween80 and 10% OADC or ADC. Competitive growth of wild-type *M. tuberculosis* and Δ*darTG* was performed in shaken 7H9 broths at 37°C and cfu assessed at time intervals by plating on 7H11 medium with and without hygromycin at 50 µg/ml. Competitive index was calculated as (day 14 cfu Δ*darTG*/day14 cfu wild type)/(day 0 cfu Δ*darTG*/day 0 cfu wild type). Analysis of data shown in Figure 4F gives a 5.2 (±3.3, 95% CI, N=8) fold competitive index.

TnSeq of *M. bovis* BCG

TnSeq of a MycoMar transposon mutant library constructed in *M. bovis* BCG was performed as previously described[20]. Sequence files were deposited at the NCBI Sequence Read Archive, SRA accession number PRJNA532518 run SRR8886987. The frequency of transposon insertions per gene was plotted for the region surrounding the *dnaB-darT-darG* locus to illustrate essentiality and abundance of gene mutants in the library.

RNA-Seq

Triplicate cultures of BCG *darG*-sgRNA were seeded in tissue culture flasks at OD_{600nm}=0.05, and subjected to control treatment (untreated for 48h), *darG* silencing (200 ng/ml aTc for 48h), or induction of DNA damage with mitomycin C (20 ng/ml for the final 24 h). Cultures were mixed with iced PBS and centrifuged at 4°C, and stored in RNA Later reagent. RNA was extracted using Tri-reagent and Lysing Matrix B tubes (MP Biomedicals) according to the manufacturers' instructions. After two chloroform extractions had been performed, the aqueous phase was purified using RNA Clean and Concentrator columns (Zymo Research), with two on-column DNA digestions performed to remove genomic DNA. RNA concentration and integrity were assessed by Nanodrop and Agilent RNA Nano 6000 chips using an Agilent Bioanalyser, confirming a RIN ≥9.0. RNA was depleted of ribosomal RNA, fragmented and random primed for first and second strand cDNA synthesis. cDNA was end repaired, 5' phosphorylated and dA-tailed before adapter ligation, PCR enrichment and sequencing on Illumina HiSeq (GENEWIZ). Reads were trimmed and aligned to the *M. bovis* BCG Pasteur 1173P2 genome (GenBank: AM408590.1). After extraction of gene hit counts, DESeq2, was used to compare gene expression between BCG *darG*-sgRNA uninduced and aTc-induced, and mitomycin C treated bacilli. Log2 fold changes were calculated, and p-values generated with the Wald test with adjustment by Benjamini-Hochberg. Genes with an adjusted p-value < 0.1 and absolute log2 fold change > 1 were called as differentially expressed genes for each comparison. The data can be accessed under the GSE number GSE174526; data page link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174526.

RT-qPCR

2 µg of RNA (as described in RNA-Seq) was reverse-transcribed to cDNA using a High Capacity RNA to cDNA kit (ThermoFisher Scientific). An equivalent reaction was also performed omitting reverse transcriptase enzyme. 167 ng cDNA was amplified for 25 cycles of PCR using PowerTrack SYBR Green Master Mix (ThermoFisher Scientific) on a QuantStudio7 Real Time PCR machine, using cycling conditions and primer concentrations recommended by the manufacturer, and using the primers listed in Supplementary Table S4. Relative Quantification of mRNA expression was calculated using the 2^-ΔΔct method, using *sigA* as the endogenous control and the mean of the untreated group as the reference. Comparison of the ct values of RT+ vs RT- samples demonstrated <0.5% contamination with gDNA throughout.

Transcriptional linkage *darB-darTG*

Early log cultures of *M. bovis* BCG Pasteur were mixed with iced PBS, harvested by centrifugation at 4°C, and stored in Bacterial RNA Protect reagent (Qiagen) at -80°C. RNA was extracted using Tri-reagent and Lysing Matrix B tubes (MP Biomedicals) according to the manufacturers' instructions. After two chloroform extractions, the aqueous phase was purified using RNA Clean and Concentrator columns (Zymo Research), with

on-column DNA digestion performed to remove genomic DNA. RT-qPCR was performed as described above using the primers listed in Supplementary Table S4. Products were visualised by gel electrophoresis using a 2% TAE-agarose gel with SYBR Gold.

Recombinant DarT protein expression and purification

To enable the crystallographic studies, DarT proteins were expressed and purified with the earlier described catalytic-null glutamate substitution (corresponding to E160A in *Thermus aquaticus* DarT)[10] to counteract the inherent toxicity of DarT. *Thermosipho africanus* DarTG E152A and *Thermus* sp. 2.9 DarT E160A constructs were transformed into the *E. coli* Rosetta strain BL21(DE3). *Thermosipho africanus* DarTG E152A-expressing cells were grown at 37°C in LB medium (Miller) supplemented with 2 mM $MgSO_4$, 0.4% glucose (*w/w*), 4% ethanol (*v/v*), 50 µg/ml of kanamycin and 35 µg/ml of chloramphenicol. *Thermus* sp. 2.9 DarT E160A-expressing cells were grown at 37°C in Terrific Broth (Merck Millipore).
After reaching an OD_{600nm} of 0.5–0.6 (*Thermosipho africanus* DarTG E152A) and 1.2 (*Thermus* sp. 2.9 DarT E160A), respectively, the temperature was lowered to 18°C prior to induction of protein expression overnight (O/N) by adding 0.5 mM IPTG. Harvested cells were resuspended in lysis buffer (50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 20 mM imidazole, 0.5 mM TCEP, cOmplete EDTA-free protease inhibitors (Roche)) and stored at -20°C until purification.
For protein purification, pellets were gently thawed and lysed by high-pressure homogenisation. DNA was digested using benzonase and lysozyme was additionally added for purification of *Thermosipho africanus* DarTG. Proteins were purified by immobilised metal affinity chromatography (IMAC) using Ni-Sepharose resin (GE Healthcare) and eluted stepwise in binding buffer containing 40–500 mM imidazole. Typically, a high salt wash with 1 M NaCl was combined with the first elution step including 40 mM imidazole. Removal of the hexahistidine tag was carried out by addition of recombinant TEV protease during O/N dialysis into buffer without imidazole, followed by purification on a second IMAC column and finally by size-exclusion chromatography (SEC) (Superdex 75, GE Healthcare) in a buffer consisting of 50 mM HEPES (pH 7.5), 300 mM NaCl, 5% glycerol and 0.5 mM TCEP.

Thermus aquaticus DarT was chosen as a model toxin for biochemical assays due its lower toxicity to the *E. coli* expression strain when compared to *Thermus* sp. 2.9 DarT. For expression and purification of wild-type and mutant *Thermus aquaticus* DarT proteins, pBAD33 plasmids were transformed into *E. coli* BL21 cells (NEB). For expression and purification of wild-type *M. tuberculosis* DarT and *Thermus* sp. 2.9 DarT, the pBAD33 construct was transformed into *E. coli* 'DH5α-macro' cells. Cells were grown at 37°C in LB medium (Miller) supplemented with 25 µg/mL chloramphenicol and 0.8% (*w/w*) glucose to an OD_{600nm} of 0.8-1.0. Cells were then pelleted by centrifugation at 4000 x g for 15 min at RT and resuspended in fresh LB media containing 25 µg/ml chloramphenicol and 0.8% (*w/w*) arabinose to induce protein expression. After 2.0 h at 37°C, cells were harvested by centrifugation (4000 x g, 15 min) and resuspended in lysis buffer (50 mM TRIS-Cl (pH 8.0), 500 mM NaCl, 5% glycerol, 20 mM imidazole, 0.5 mM TCEP) and stored at -20°C until purification. Cells were lysed using BugBuster (Novagen) following the manufacturer's instructions after adding cOmplete EDTA-free protease inhibitors (Roche) and benzonase (Novagen). The DarT proteins were purified by IMAC using Ni-Sepharose resin (GE Healthcare). DarT proteins prepared for mutagenesis studies were further purified with an additional IMAC purification using TALON affinity resin (Clontech) after dialysis of the protein in lysis buffer. Finally, purified *Thermus aquaticus* DarT proteins were dialysed against protein storage buffer containing 50 mM TRIS-Cl (pH 8.0), 300 mM NaCl, 5% glycerol, 0.5 mM TCEP while *M. tuberculosis* DarT protein was buffer-exchanged (directly after the Ni-IMAC purification step) by repeated filter concentration in storage buffer containing 50 mM HEPES (pH 7.4), 300 mM NaCl, 5% glycerol, 0.5 mM TCEP.
All proteins were characterised by SDS-PAGE, then flash frozen in liquid nitrogen and stored at −80°C until required. Protein concentrations were determined by measuring absorption of the sample at 280 nm with the NanoDrop1000 (Thermo Scientific).

Isothermal titration calorimetry (ITC)

Binding experiments were carried out on a VP-ITC microcalorimeter (MicroCal). The protein was dialysed O/N at room temperature in 50 mM HEPES (pH 7.4), 300 mM NaCl, 5% glycerol, 0.5 mM TCEP using D-tube™ Dialysis Midi MWCO 3.5 kDa (Novagen) dialysis tubes prior to the experiment. Titrations experiments were performed at 25°C, a reference power of 12 µCal/sec and a stirring speed of 307 rpm with an initial injection of

2 µL followed by 28 identical injections of 10 µL (duration of 4 s/injection and spacing of 240 s between injections). Data were analysed using the MicroCal PEAQ-ITC analysis software (Malvern).

Toxicity assays

DH5α cells transformed with *Thermus aquaticus* DarT pBAD33 expression plasmids were grown in the presence of 0.8% (*w/w*) glucose and 25 µg/mL chloramphenicol O/N and streaked onto LB agar plates containing 25 µg/mL chloramphenicol for selection and 0.8% (*w/w*) glucose or 0.8% (*w/w*) arabinose for repression or induction of protein expression, respectively. The bacteriostatic effects were assessed after incubating the plates at 37°C O/N.

Thin layer chromatography (TLC)

To analyse reaction products and assess NADase activity of DarT, 1 µM *Thermus* sp. 2.9 DarT wild-type and E160A were incubated for 3 h at 37°C with 5 µM β-NAD^+ (supplemented with ^{32}P-NAD^+ at 10000 Bq/reaction) with (500 nM or 10 µM) or without DNA (DNA-5mer) in ADP-ribosylation buffer (50 mM TRIS pH 8.0, 50 mM NaCl). NADase from porcine brain (≥0.007 unit/mg, Sigma) was used as control. 1 µl of the reaction was spotted on a PEI cellulose plate (Macherey-Nagel) which was allowed to air dry and developed in 0.25 M LiCl and 0.25 M formic acid. The plate was dried and exposed to autoradiography films.

ADP-ribosylation activity assays

ADP-ribosylation reactions were performed in ADP-ribosylation buffer (50 mM Tris-Cl pH 8.0, 50 mM NaCl) supplemented with 5 mM ETDA at 37°C for 30 min unless otherwise indicated. In general, 1 µM DarT protein was incubated with oligonucleotides at a concentration of 3 µM and β-NAD^+ in excess (500 µM). Reaction products were analysed by separation on denaturing polyacrylamide gels run in TBE buffer, loading 0.02 nmol oligonucleotide with urea loading dye (10 mM TRIS pH 8.0, 10 mM EDTA, 4 M urea), and by following visualisation under UV light (340 nm) after ethidium bromide-staining.

ADPr-IP of ADP-ribosylated gDNA

BCG sgRNA-*darG* cultures were grown to late log phase, either untreated or treated with 200 ng/ml anhydrotetracycline (ATC) to induce *darG* silencing for 24 h. The cultures were then diluted to OD_{600nm} 0.15 to stimulate division, and fresh aTc was added as appropriate. After 3 days, the bacteria were harvested by centrifugation, resuspended in TE buffer pH 8.0, and incubated for 5 min rocking with an equal volume of methanol:chloroform 2:1. Bacilli were centrifuged and the pellet dried, before re-suspending in TE adding phenol:chloroform:isoamyl alcohol 25:24:1. Cells were disrupted with the Fastprep homogeniser (MP Biomedicals) and lysing matrix B which was followed by centrifugation to separate the aqueous and organic phases. The upper aqueous phase was re-extracted with an equal volume of phenol:chloroform:isoamyl alcohol 25:24:1, followed by two further chloroform extractions. gDNA was precipitated with 0.1 volume 3 M sodium acetate pH 5.2 and 1 volume propan-2-ol, pelleted by centrifugation, washed with 70% ethanol, and resuspended in TE buffer pH 8.0. gDNA was then digested with AluI (NEB) in the presence of RNAse (Roche), purified using MinElute columns (Qiagen), and the concentration determined using the QuantiFluor dsDNA system (Promega).

ADPr-IP was performed by adaptation of a methylated DNA immunoprecipitation protocol[29]. 200 ng of AluI-digested gDNA was retained as the "Input" sample. Digested gDNA was subjected to ADPr-IP using either rabbit anti-poly/mono-ADP ribose antibody (E6F6A, Cell Signaling Technology), or rabbit IgG (isotype control). 1.5 µg of Alu-1 digested gDNA was added to 500 µl IP buffer (10 mM $NaPO_4$ pH 7.0, 140 mM NaCl, 0.05% Triton X-100) containing 1.5 µg antibody, and rotated overnight at 4°C. 30 µl of equilibrated DynabeadsTM Protein A (ThermoFisher Scientific) were added, and rotated for a further 4 h at 4°C. The DynabeadsTM bound to gDNA-IgG complexes were then washed three times with ice cold IP buffer, using a DynaMag magnet device. The beads were resuspended in 400 µl Digestion Buffer (10 mM Tris pH 8.0, 100 mM EDTA, 0.5% SDS, 50 mM NaCl, 100 µg proteinase K per reaction) and digested for 2 h at 55°C. Beads were removed by magnetic separation, and gDNA was recovered from the supernatant by MinElute columns (Qiagen). qPCR was performed using PowerTrack SYBR Green Mastermix (ThermoFisher Scientific) using primers provided in Supplementary Table S4, at concentrations and cycling conditions specified by the manufacturer. The efficiency of PCR for both primer sets was assessed using gDNA; both sets were >95%

efficient, with efficiencies within 5% of each other. Relative quantification of the abundance of DNA fragments containing the origin of replication ('OriC') was compared to the abundance of DNA fragments containing Rv2129c (a probable oxidoreductase with 65.3% GC content), using the 2^-ΔΔct method. Data were normalised to the mean of untreated input samples.

Large scale preparation of ADPr-DNA

3000 nmole oligonucleotide (5 mM assay concentration) was ADP-ribosylated by incubation at 37°C O/N with *Thermus aquaticus* DarT and 15 mM β-NAD^+ in ADP-ribosylation buffer (50 mM Tris-Cl pH 8.0, 50 mM NaCl) supplemented with 5 mM ETDA. The ADP-ribosylated oligonucleotide was gel-purified on denaturing polyacrylamide gel run in TBE buffer, and recovered by excising from the gel and extracting from the gel pieces by diffusion with several elution steps in TE buffer. Gel residuals were removed by centrifugation of the eluates through 0.22 μm filter units (GE Healthcare). Finally, eluates were desalted using Sephadex G-25 in PD-10 Desalting Columns (GE Healthcare) equilibrated in nuclease-free water and concentrated to the desired concentrations. ADPr-DNA for NMR analysis was entirely concentrated and then dissolved in 20 mM potassium phosphate buffer, pH 7.4. The average yield of purified ADPr-DNA obtained was 45% of the original unmodified DNA input.

Detection of ADP-ribosylated genomic DNA

E. coli BL21 (DE3) were grown to OD_{600nm} of 0.2-0.3 in LB containing 0.8% (*w/w*) glucose before protein expression was induced with 0.8% (*w/w*) arabinose for DarT and 50 μM for DarG for 2 h. Cells were harvested by centrifugation (4000 x g, 3 min), washed with PBS, re-suspended in boiling lysis buffer (1.0% SDS, 10 mM Tris, 1 mM EDTA, pH 8.0) and lysed by heating to 95°C for 5 min. Cell lysates were subjected to proteinase K treatment for 1 h, 50°C. gDNA was then extracted by phenol:chloroform:isoamyl alcohol extraction and recovered by ammonium acetate/ethanol precipitation. The DNA pellets were washed twice with 70% ethanol before re-suspending in TE buffer and concentration determination using a DeNovix DS-11 FX nanodrop. Mycobacterial ADP-ribosylated genomic DNA were prepared from *darG*-silenced BCG (sgRNA-*darG*) cultures as described above (method section: 'ADPr-IP of ADP-ribosylated gDNA'). ~1 μg of gDNA was dotted onto a nitrocellulose membrane (Amersham Protran 0.45 NC nitrocellulose) and crosslinked with 1200 J using a Stratalinker UV crosslinker. Crosslinked DNA was then immunoblotted for gDNA (autoanti-dsDNA, DSHB, 1:200) or ADPr-gDNA (Poly/Mono-ADP ribose, E6F6A, Cell Signalling Technology, 1:1000) for 1 h at RT in 5% (*w/v*) powdered milk in PBS-T. Of note, the antibody used for detecting the ADPr modification on gDNA was selected from screening all commercially available anti-ADP-ribose antibodies and was identified as the only suitable one. Secondary peroxidase-couple antibodies (Dako) were incubated at RT for 1 h. ECL-based chemiluminesence was detected using Hyperfilms (GE). Autoanti-dsDNA was deposited to the DSHB by Voss, E.W. (DSHB Hybridoma Product autoanti-dsDNA).
For gDNA de-modification, gDNA was incubated with 1 μM of the indicated hydrolase at 37°C for 30 min and detection of ADPr-gDNA was performed by dot blot as described above.

Western blot procedure

For *Thermus aquaticus* DarT *in cellulo* studies, Western blot analysis was performed on samples generated for gDNA analysis. Cells were lysed in lysis buffer (1.0% SDS, 10 mM Tris, 1 mM EDTA, pH 8.0) by heating to 95°C for 5 min. Samples were treated with benzonase (0.5U/μL) and protein concentration determined for normalisation using Bradford reagent. 5 μg of protein lysate were resolved by SDS-PAGE and electrotransferred to 0.2 μm nitrocellulose membranes. Membranes were blocked in 5 % *(w/v)* milk-PBS with 0.05% *(v/v)* Tween 20 (PBS-T) for 1 h before detection of RecA levels with rabbit anti-RecA polyclonal antibody (1:10 000, ab63797, Abcam), DarT/DarG with mouse anti-His monoclonal antibody (1:1000, 631212, Takara), DarT with rabbit anti-V5 Tag polyclonal antibody (1:2500, A190-120A, Bethyl Laboratories) and GroEL with mouse anti-GroEL monoclonal antibody [9A1/2] (1:15000, ab82592, Abcam). IgG HRP conjugate secondary antibodies, i.e. goat anti-mouse (1:2000, P0447, Agilent) and goat anti-rabbit (1:2000, P0399, Agilent), were used with ECL western blotting detection kit (Pierce) for visualisation.
For *M. tuberculosis* studies, *M. tuberculosis* (GC1237) *darG* sgRNA or *M. tuberculosis* control sgRNA (non-targeting) were grown in shaking culture in 7H9 broth containing hygromycin (50 μg/ml) and kanamycin (20 μg/ml) to an OD_{600nm} of ~0.3. Cultures were then treated with 200 ng/ml ATC for 48 h; grown on for 24 h and

then treated with mitomycin C (MMC) at 20 ng/ml for 24 h or untreated 48 h. Cells were harvested by centrifugation and lysed by bead-beating in PBS containing protease inhibitors (Roche) before centrifugation. Proteins in the supernatant were separated by SDS-PAGE, blotted to nitrocellulose and probed with rabbit anti-RecA (1:2000, ab63797, Abcam) with secondary goat anti-rabbit IgG-HRP (1:4000, A16096, Life technologies) before detection with Clarity Max ECL substrate (Bio-rad). Blots were stripped and re-probed with monoclonal mouse anti-Hsp70 (1:4000, clone CosII, gift of Douglas Young) and secondary goat anti-mouse IgG-HRP (1:500, 31430, ThermoFisher).

Crystallisation, data collection, structure solution, refinement

Purified *Thermosipho africanus* DarTG E152A protein was concentrated to 10.7 mg/ml in 10 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT and 2 mM ADPr was added for at least 1 h prior to setting up crystallisation drops. Purified *Thermus* sp. 2.9 DarT E160A protein was concentrated to 21.5 mg/ml and incubated for co-crystallisation with different substrates for 30 min at RT after adding either 4 mM β-NAD^+, 4 mM carba-NAD^+ or 1.3 mM (1.5x) ADPr-DNA to the sample. For co-crystallisation with unmodified DNA, proteins were first pre-incubated with β-NAD^+/4 mM carba-NAD^+ for 30 min which was followed by incubation with 1.3 mM (1.5x) DNA-5mer for another 30 min. Crystallisation trials were performed at 20°C using the sitting-drop vapour-diffusion method. Crystallisation drops were set-up in MRC two-well crystallization microplates (Swissci) using the Mosquito Crystal robot (TTP Labtech) with protein to reservoir ratios of 1:1 and 1:2 in 300 nl total volume equilibrated against 75 µl of reservoir solution.

Crystals of *Thermosipho africanus* DarTG E152A protein grew in 0.2 M potassium thiocyanate, 0.1 M TRIS pH 7.5, 8% (*w/v*) PEG 20,000, 8% (*v/v*) PEG 500 MME. Crystals of *Thermus* sp. 2.9 DarT E160A *apo* were obtained in 100 mM potassium thiocyanate and 30% (*w/v*) PEG2000MME and *Thermus* sp. 2.9 DarT E160A co-crystals in crystallisation solutions of slight concentrations variations of these components, i.e. 50-200 mM potassium thiocyanate and 10-20% (*w/v*) PEG2000MME. Crystals were harvested using reservoir solution supplemented with 20% ethylene glycol (*v/v*) or 18% glycerol (*v/v*) as a cryo-protectant prior to flash freezing in liquid nitrogen.

X-ray data were collected at beamlines I24 (*Thermus* sp. 2.9 DarT:ADPr-DNA:NAM, *Thermus* sp. 2.9 DarT:ADPr-DNA), I03 (*Thermosipho africanus* DarTG:ADPr), I04 (*Thermus* sp. 2.9 DarT *apo, Thermus* sp. 2.9 DarT:NAD^+, *Thermus* sp. 2.9 DarT:Carba-NAD^+) and I04-1 (*Thermus* sp. 2.9 DarT:Carba-NAD^+**:**ssDNA) at the Diamond Light Source (Rutherford Appleton Laboratory, Harwell, UK) with the following X-ray wavelengths: 0.9795 Å (*Thermus* sp. 2.9 DarT *apo*, *Thermus* sp. 2.9 DarT:NAD^+, *Thermus* sp. 2.9 DarT:Carba-NAD^+), 0.9763 Å (*Thermosipho africanus* DarTG:ADPr), 0.8998 Å (*Thermus* sp. 2.9 DarT:ADPr-DNA:NAM, *Thermus* sp. 2.9 DarT:ADPr-DNA), and 0.9159 Å (*Thermus* sp. 2.9 DarT:Carba-NAD^+**:**ssDNA). Data collection statistics are provided in Extended Data Table 1.

X-ray data were processed using the XIA2 platform[30] and the ccp4i suite (v7.1.014). Phase information was obtained using the molecular replacement method with PHASER (v2.8.2)[31]. Density modification was implemented with PARROT (v1.0.5)[32] and initial models were build using the automated model building programme BUCCANEER (v1.16.9)[33]. Atomic models were improved following consecutive cycles of manual building in COOT (v0.9.4)[34] and structure refinement in REFMAC (v5.8.0267)[35]. *Thermosipho africanus* DarTG E152A was solved by molecular replacement using the previously published *Thermus aquaticus* DarG macrodomain structure (PDB ID: 5M31) as a search model. Initial phases for *Thermus* sp. 2.9 DarT *apo* were obtained by molecular replacement using the toxin of *Thermosipho africanus* DarTG (residues F46-V230) as search template. *Thermus* sp. 2.9 DarT *apo* was used as reference model for subsequently solving the ligand-bound structures of *Thermus* sp. 2.9 DarT. The structures were refined to good Ramachandran statistics without outliers except for the *Thermosipho africanus* DarTG:ADPr structure, which contains 0.4% outliers. MolProbity[36] was used to validate the models prior to deposition in the PDB. Processing and refinement statistics are given in Table S1. The PDB IDs for the atomic coordinates and structure factors reported in this manuscript are 7OMV, 7OMZ, 7ON0, 7OMY, 7OMW, 7OMX and 7OMU.

NMR analysis of *Thermus aquaticus* DarT ADP-ribosylated DNA

To confirm in solution the α stereospecificity of the reaction and the atomic NAM ribose C1"–N3 thymine linkage established by DarT, an ssDNA oligonucleotide with sequence GATGTCAG was modified by DarT *in vitro* and subjected to 1D and 2D NMR analyses. The ssDNA octamers (unmodified and ADP-ribosylated) were prepared as solutions in 17 mM phosphate buffer pH 7.4 in 90% H_2O/10% D_2O at final concentrations of a 2.8

mM (reference ssDNA) and 5.3 mM (modified ssDNA) in volumes of 180 µl. Samples were placed in 3 mm NMR tubes and NMR analyses were performed with a Bruker AVIII 700 spectrometer equipped with a TXI H/C/N room temperature probe regulated at 298 K. 1D ^{1}H spectra were collected using a 1D NOESY-presat scheme for solvent suppression with a 2 s pre-saturation period and a 50 Hz rf field. 2D multiplicity edited ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC spectra were acquired using windows of 8 ppm (^{1}H) and 180 ppm (^{13}C) centred at 4.7 and 90.0 ppm respectively, and employed solvent pre-saturation (50 Hz rf) during the recovery delays of 1.5 s (HSQC) or 2 s (HMBC). J_{CH}-coupling evolution delays were optimised for 145 Hz (HSQC) and 8 Hz (HMBC). 2D TOCSY spectra were acquired with ^{1}H windows of 8 ppm centred at 4.7 ppm and employed ^{1}H pre-saturation (50 Hz rf) during the 2 s recovery delay. Isotropic mixing was achieved using the DIPSI-2 mixing sequence for 100 ms. 1D selective TOCSY experiments employed similar pre-saturation and the DIPSI-2 mixing scheme. Additional suppression of zero-quantum interference was achieved using 60 kHz adiabatic CHIRP pulses of 20 ms and 15 ms before and after the mixing time, each combined with gradient amplitudes of 11% and 13% of maximum, respectively. Mixing times were varied from 40 to 150 ms to deconvolute proton spin systems. Selective excitation was achieved using 180^{o} Gaussian pulse of 80 ms within a single gradient spin-echo. 1D selective ROESY spectra used a similar selective excitation scheme with solvent pre-saturation during a 3 s recovery delay and a continuous wave spin-lock of 200 to 400 ms for ROE detection. NOESY spectra of the ssDNAs produced rather weak negative NOEs (consistent with their molecular masses of 2992 Da) and were not pursued for structural studies. NMR data analysis was performed with Bruker Topspin 3.2.

Data analysis and presentation

Structural alignments and analyses, as well as figure preparation, were carried out using PyMol (Molecular Graphics System, Version 2.3.3 Schrödinger, LLC). For multiple-sequence alignments, JalView v2[37] and MAFFT7[38] was used. The phylogenetic tree for the catalytic ART domains was generated with SplitsTree4 (v4.15.1) using the Neighbour-Joining (NJ) method[39] and confidence levels estimated using 1000 cycles of the bootstrap method. For DarT, the sequence of *Thermus* sp. 2.9 was used. Sequence conservation mapping was performed using ConSurf 2016[40]. Prism (v9.0.1) was used for statistical analysis and graph representation. ChemDraw (v15.0.0.106) was used for presentation of chemical structures. Inkscape (v0.91) was used for final figure preparation.

ADDITIONAL REFERENCES FOR METHOD SECTION

26. Singh, A. K. *et al.* Investigating essential gene function in Mycobacterium tuberculosis using an efficient CRISPR interference system. *Nucleic Acids Res.* **44**, e143 (2016).

27. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).

28. St-Pierre, F. *et al.* One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* **2**, 537–541 (2013).

29. Thu, K. L. *et al.* Methylated DNA immunoprecipitation. *J. Vis. Exp.* e935 (2009). doi:10.3791/935

30. Winter, G. *xia2*: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).

31. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**, 432–438 (2004).

32. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **66**, 470–478 (2010).

33. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **62**, 1002–1011 (2006).

34. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **60**, 2126–2132 (2004).

35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **53**, 240–255 (1997).

36. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallogr.* **D66**, 12–21 (2010).

37. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

739 38. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive
740 sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2018).

741 39. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol.*
742 *Evol.* **4**, 406–425 (1987).

743 40. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation
744 in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

745 41. Tsurumura, T. *et al.* Arginine ADP-ribosylation mechanism based on structural snapshots of iota-toxin and actin
746 complex. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7524–7528 (2013).

747

748

DATA AVAILABILITY

Crystallography atomic coordinates and structure factors are deposited in the Protein Data Bank (PDB) (www.rcsb.org) under the following accession codes: 7OMV, 7OMW, 7OMX, 7OMY, 7OMU, 7OMZ, 7ON0. RNA-Seq sequence files are deposited at the NCBI Gene Expression Omnibus GEO under the accession code GSE174526. TnSeq sequence files are deposited at the NCBI Sequence Read Archive, SRA accession number PRJNA532518 run SRR8886987. All data supporting the findings of this study are available within the paper and any further information will be provided upon request.

ACKNOWLEDGMENT

We thank Christoph Tang for guidance and discussion of this study and sharing reagents; Johannes Rack and Marcin Suskiewicz for discussions and critically revising the manuscript; Marcin Suskiewicz and Edward Lowe for help with X-ray diffraction data collection and processing; Svenja Meyer for help with biochemical experiments; Jon Elkins for discussions; Tom Mendum, Huihai Wu and Alex Smith for help with transposon mutagenesis; Suzie Hingley-Wilson for discussions on TA systems; Diamond Light Source for access to and assistance at beamlines I03, I04, I04-1, and I24 throughout the project [proposal numbers mx18069]. Work in the Ivan Ahel Laboratory is supported by the Wellcome Trust (101794 and 210634), Biotechnology and Biological Sciences Research Council (BB/R007195/1) and Cancer Research United Kingdom (C35050/A22284). The Stewart and Kendall labs were supported by Biotechnology and Biological Sciences Research Council grants BB/R006393/1 and BB/N004590/1, respectively.

AUTHOR CONTRIBUTIONS

IA and GRS conceived the project and conceptualised experiments with input from MS, REB and GJ. MS conducted biochemical and crystallographic studies including structure and data analysis/interpretation with assistance of other authors; AA solved *Thermosipho africanus* DarTG structure and refined structural data; CTC and GJ established method for detection of ADPr-DNA and supported strain construction; TDWC conducted NMR experiments and analysis; REB and GRS performed mycobacteria experiments with assistance from SLK and SG for DarG knockdowns. MS, IA, GRS and REB wrote the manuscript with support of all other authors.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information: The online version contains supplementary material.

Correspondence and requests for materials and resources should be addressed to Dr. Ivan Ahel (ivan.ahel@path.ox.ac.uk) and Professor Graham R. Stewart (g.stewart@surrey.ac.uk).

Peer reviewer reports are available.

FIGURE LEGENDS OF THE EXTENDED DATA

Extended Data Figure 1 [related to Figure 1]. DarT structure reveals a PARP-like ADP-ribosyltransferase. (A) Crystal structure of *Thermosipho africanus* DarTG E152A fusion protein served as model for solving DarT of *Thermus* sp. 2.9 by MR. The fused DarG macrodomain is coloured in light orange with the bound ADP-ribose molecule shown as an atom-coloured stick model in black. **(B, C)** Comparison of DarT with eukaryotic ARTD, i.e. PARP, and bacterial ARTD fold. **(B)** Secondary structure analysis shows the close similarity of DarT to ARTD family members, and PARPs in particular, with its fold-stabilising central 6-stranded β-sheet core and the ARTD-conserved helices between strand β1-2 and β2-3. The crystal structures of *Thermus* sp. 2.9 DarT E160A in ligand-free state (apo) was overlayed with diphtheria toxin (PDB ID: 1tox), PARP1 (PDB ID: 6bhv) and PARP13 (PDB ID: 2x5y). For clarity, only central secondary structure elements showing the similarity between the folds are depicted from diphtheria toxin, PARP1 and PARP13. **(C)** The N-terminal extension of the β-sheet core found in PARPs, i.e. a strand-helix-strand arrangement next to β6, is spatially replaced in DarT with a shorter C-terminal helix-strand extension. The crystal structure of *Thermus* sp. 2.9 DarT E160A in ADPr-DNA-bound state was overlayed with the crystal structure of PARP1 (PDB ID: 6bhv, left) and PARP13 (PDB ID: 2x5y, right). For clarity, only the secondary structure elements showing this difference between the folds are depicted from PARP1 and PARP13. **(D)** *Thermus* sp. 2.9 DarT E160A in complex with NAD^+ and carba-NAD^+. Overlay of the crystal structures with *Thermus* sp. 2.9 DarT E160A in ligand-free (apo) state is shown on the right. Unresolved regions of the NAD^+-binding loop-helical element including the ART 'donor loop' (purple) in the DarT:NAD^+ and DarT:carba-NAD^+ structure are marked with asterisks.

Extended Data Figure 2 [related to Figure 1C, D]. Structural details of ADPr-DNA bound to DarT. (A) Crystal structure of *Thermus* sp. 2.9 DarT E160A in complex with ADPr-DNA. Co-crystallisation with ADP-ribosylated DNA 5mer after *in vitro* modification by *Thermus aquaticus* DarT and purification. The substrate binding ('ARTT') loop is highlighted in green, the NAD^+-binding loop-helix element in purple (set for clarity in higher transparency). The catalytic glutamate E160, which is conserved in ARTs, is modelled as red sticks. **(B)** Overlay of the ADP-ribosylated DNA products of the ADPr-DNA:NAM-bound and the ADPr-DNA-bound DarT structures shows their perfect overlap apart from a slight tilting of the bonds which connect the NAM-ribose with the beta-phosphate. **(C, D)** The ADPr-DNA ligands are highly resolved in the *Thermus* sp. 2.9 DarT E160A co-crystal structures revealing the ADPr linkage to the thymidine base nitrogen N3 in α-conformation. The $2F_o$-F_c electron density maps contoured at 1.0 σ around the ligands is shown in blue. (C) The ADPr-DNA ligand in the ADPr-DNA co-crystal structure of 1.46 Å resolution. (D) The ADPr-DNA ligand in the ADPr-DNA:NAM-bound co-crystal structure at 1.66 Å resolution. The nicotinamide ('NAM') ligand left in the protein after ADP-ribosylation of the DNA is also clearly resolved. **(E)** ADP-ribosylation activity of the *Thermus aquaticus* DarT E160A mutant can also be observed in *in vitro* assays at low DNA (50 nM) and high protein concentrations under long incubation times. Modification of the Cy3-labelled oligo ('DarT-ADPr-27mer-Cy3') was visualised after separation of the reaction products on denaturing polyacrylamide gel. Representative of two independent experiments.

Extended Data Figure 3 [related to Figure 1,2,3]. Structural features of DarT for ssDNA binding and catalysis. (A) The ADP-ribosylating turn-turn ('ARTT') loop of *Thermus* sp. 2.9 in the ADPr-DNA:NAM-bound structure is shown in green with its stabilised DNA substrate in magenta. Several loops form together with the few short α-helices a stable scaffold which is held in position by a network of over 100 interactions between main chains, side chains and water molecules. Cartoon representation is shown on the **left**, the atom-coloured stick model of the 'ARTT' loop in the **middle**. Interactions are indicated with grey dashes and water molecules as red spheres. A table comparing the 'ARTT' loop length of DarT with other human and bacterial ARTDs[15] is provided on the **right**. **(B)** *Thermus* sp. 2.9 DarT preferentially modifies a TNTC motif in ssDNA which was verified by testing permutations of the motif. *In vitro* ADP-ribosylation activity of *Thermus* sp. 2.9 DarT was assessed by visualising the modification of the oligos under UV light after separation and ethidium bromide-staining of the reaction products on denaturing polyacrylamide gel. Representative of three independent experiments. **(C)** Close views on the nucleotide recognition of DarT rationalising its preferred modification of DNA over RNA. Cartoon-stick models of the *Thermus* sp. 2.9 DarT E160A structure in the ADPr-DNA:NAM-bound state are shown. **(Left, middle)** Additional 2' hydroxyl groups as in RNA strands may lead to clashes with parts of the proteins, i.e. W147 (1st nucleotide) and the α-helix between β2 and β3 (2nd nucleotide). **(Right)** As shown in previous studies the methyl group on the modified thymine base (circle) increases thymidine base modification, probably by locking the base in optimal conformation for the ADP-ribosylation reaction. Interactions are indicated with grey dashes and water molecules as red spheres. **(D)** Modelling of possible

rotamers of glutamate E160 into the *Thermus* sp. 2.9 DarT *apo* structure. Several conformations of the glutamate would allow a proton transfer from arginine R51 (green) to glutamate E160 (red). Possible interactions are shown with dashes in magenta. **(E)** Arginine R51 flexibility observed among different *Thermus* sp. 2.9 DarT *apo*, substrate- and product-bound states. The NAD^+:DNA-bound and as ADP-ribose unlinked state (fourth imagine from left) is modelled by superimposing the NAD^+ molecule with the NAD^+ co-crystal structure onto the carba-NAD^+:DNA-bound structure. R51 and ligands are shown as atom-coloured stick models, with R51 in green, NAD^+ in cyan, carba-NAD^+ in brown and higher transparency and DNA (thymine only) in magenta. Interactions are indicated with grey dashes.

Extended Data Figure 4 [related to Figure 2C,D, 3]. DarT sequence alignments. (A) Sequence alignment of *Thermus* sp. 2.9 DarT with DarT of *Thermus aquaticus*. Numbers on top of the alignments refer to *Thermus* sp. 2.9 DarT. Table provides a residue ID comparison for functional relevant residues. **(B)** Multiple sequence alignment of DarT sequences representing five main phylogenetically diverging branches. Numbers on top of the residues refer to *Thermus* sp. 2.9 DarT. Active site residues are highlighted in green, DNA-binding residues in magenta, with functionally similar residues as the reference in lower opacity. Shared sequence identities compared to *Thermus* sp. 2.9 DarT: group 1: 60%, group 2: 40-45%, group 3/4: 31-38%, group 5: 20-27%.

Extended Data Figure 5 [related to Figure 3B, Result Section, Supplementary Discussion]. NAD^+ coordination in the active site of DarT. (A) Comparison of the NAD^+-binding sites in the carba-NAD^+-bound structure and the NAD^+-bound structure of *Thermus* sp. 2.9 DarT E160A. **(top)** Overlay of the carba-NAD^+:DNA-bound structure (grey) with the carba-NAD^+-bound structure (brown) of which just the ligand and the side chains are shown as atom-coloured stick model. The carba-NAD^+ ligands of both structures perfectly overlay and DarT-interacting side chains show same positioning. DNA binding does not induce conformational changes upon the NAD^+ ligand. **(bottom)** Overlay of the carba-NAD^+:DNA-bound structure (grey) with the NAD^+-bound structure (cyan) shows slight differences in the ligand and the side chains positioning around the pyrophosphate-ribose moiety of the NAD^+ molecule which needs to be considered for analysis of NAD^+ polarisation. **(B)** Molecular structures of β-NAD^+ and carba-NAD^+. **(C)** Cartoon-stick model showing the coordination of the nicotinamide side ('NAM', left) and the adenine side (right) of the carba-NAD^+-ligand in the *Thermus* sp. 2.9 DarT E160A structure with side- and main-chain interactions (dashed lines) including water (red spheres) contacts. **(D)** Integrated thermogram obtained by ITC giving NAD^+-binding parameters for *Thermus* sp. 2.9 DarT E160Q. A representative result from three independent experiments is shown, with the number of binding sites N and the dissociation constant K_D calculated from the repeats with mean ± SD. **(E)** Autoradiography of TLC plate analysing the reaction products after incubation of *Thermus aquaticus* DarT WT and E160A mutant with NAD^+ and DNA. NADase from porcine brain was used as control for monitoring NADase activity. Representative of three independent experiments.

Extended Data Figure 6 [related to Figure 4, Result Section]. Visualisation of DarT's ADP-ribosylation activity in cells. (A) Validation of the antibody identified for detection of ADP-ribosylated DNA. ADP-ribosylation of the oligonucleotide by *Thermus aquaticus* DarT was verified by analysis of the reaction product on denaturing polyacrylamide gel (top panel) and visualised by immunoblotting using the Poly/Mono-ADP ribose antibody, E6F6A (Cell Signalling Technology) (bottom panel). Immunodetection of ssDNA served as loading control (middle panel). Representative result of four independent experiments with three individually purified *Thermus aquaticus* DarT-ADP-ribosylated oligonucleotides. **(B)** Dot blot showing DNA ADP-ribosylation activity by *Thermus aquaticus* and EPEC DarT WT and mutants on gDNA, its physiological target, (row 1 and 2 from top) which consequently induces DNA damage (RecA marker) in cells. EPEC DarT G49D is a characterised DarT mutant that retains ssDNA ADP-ribosylation activity, albeit to a lesser extent than the wild-type protein, while EPEC DarT E170 is its respective catalytically inactive mutant[19]. See also Extended Data Figure 6C. **(C)** G49D mutation in EPEC DarT reduces ssDNA ADP-ribosylation activity. Overlay of a homology model of EPEC DarT with the structure of *Thermus* sp. 2.9 DarT E160A in complex with ADPr-DNA indicates that the EPEC DarT mutation G49D reduces DarT ssDNA activity due to an aspartate side-chain pointing into the NAD^+-binding site towards the second phosphate group. This may sterically but also due to its negative charge impair NAD^+-binding, resulting in a less efficient ADP-ribosylation reaction. **(D)** DNA ADP-ribosylation by *Thermus aquaticus* DarT (dot blot, row 1 and 2 from top) and induction of DNA damage (RecA marker) is suppressed by *Thermus aquaticus* DarG with its macrodomain (MD) including by DarG macrodomains from non-cognate species (EPEC, *M. tuberculosis*). **(E)** Dot blot showing ADP-ribose removal from *Thermus aquaticus* DarT ADP-ribosylated genomic DNA by *Thermus aquaticus* DarG antitoxin with its

macrodomain (MD) and macrodomains from non-cognate species (EPEC, *M. tuberculosis*) in contrast to human hydrolases MarcoD1, PARG and ARH3.
(B), (D), (E): Cell lysates were prepared and genomic DNA was purified from samples before (+ glucose) and after (+ arabinose/IPTG) induction of protein expression and subjected to immunodetection. EV: empty vector. The N22A-K80A double mutation in *Thermus aquaticus* DarG results in loss of catalytic activity of the macrodomain. For gel source data, see Supplementary Figure 1. Results are representative for three biologically independent experiments.

Extended Data Figure 7 [related to Figure 4]. Characterisation of DarT gDNA ADP-ribosylation in *Mycobacterium tuberculosis.* (A) Unregulated DarT activity (*darG* silencing) and induction of DNA damage (mitomycin C, i.e. MMC, treatment) led to profound DNA damage response and induces expression of *dnaB-darT*. Gene transcription was compared by RT-qPCR of *M. bovis* BCG *darG*-sgRNA uninduced, ATC-induced, and mitomycin C-treated samples. Data are mean±SD of three biologically independent replicates. **(B)** Knockdown of *darG* expression in *M. tuberculosis* induces expression of RecA. *M. tuberculosis* were treated with 200 ng/ml ATC to induce dCas9 and *darG* sgRNA or non-targeting control sgRNA for 48 hours or with mitomycin C (MMC) for 24 hours. Cell-free bacterial lysates were probed by Western blotting with an anti-RecA antiserum or anti-Hsp70 (DnaK) antibodies as loading control. Representative of two biologically independent experiments. **(C)** *darTG* is transcriptionally linked to *dnaB*. PCR products were generated with the indicated set of primers (see Supplementary Table 2 for details) and visualised by gel electrophoresis. The presence of PCR products across the *dnaB-darT* and *darT-darG* junctions demonstrates the transcriptional linkage of *dnaB*, *darT* and *darG* as a polycistronic mRNA. Representative of three independent experiments. **(D)** *M. tuberculosis* DarT preferentially modifies a TTTW motif in ssDNA. Screening of 40 ssDNA oligo sequences with potential four-base motifs for ADP-ribosylation by DarT (data not shown) identified TTTW as targeted sequence which was verified by testing permutations of the TTTT motif. *In vitro* ADP-ribosylation activity of *M. tuberculosis* DarT was assessed by visualising the modification of the oligos under UV light after separation and ethidium bromide-staining of the reaction products on denaturing polyacrylamide gel. Representative of three independent experiments. **(E)** *M. tuberculosis* DarT ADP-ribosylates the OriC *in vitro* with preference for the lower strand at the TTTW motifs. ADP-ribosylation activity was assessed by visualising the modification of the oligos under UV light after separation and ethidium bromide-staining of the reaction products on denaturing polyacrylamide gel. Representative of three independent experiments. For gel source data, see Supplementary Figure 1.

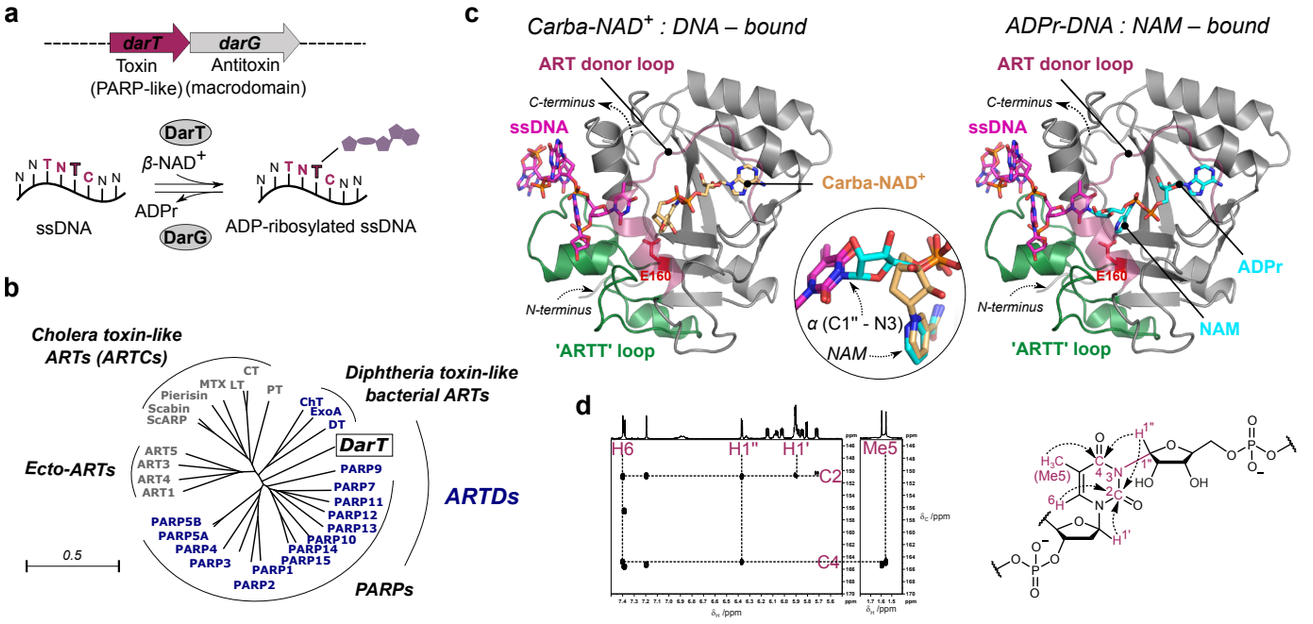
Extended Data Figure 8 [related to Discussion Section]. DarT is a PARP-like protein that evolved novel features that allow its specialised function as a DNA ADP-ribosyltransferase. (A) Schematic representations of the interactions between the NAD^+ substrate and the residues of the class-defining [H-Y-E] motif in ARTD members including PARPs compared to DarT. Conserved motif residues (purple) and additional active site residues (green) essential for catalysis with their relative position to the NAD^+ substrate are compared. **(B)** ARTs seem to share the spatial position and orientation of mechanistically relevant residues. Overlay of crystal structures of *Thermus* sp. 2.9 DarT E160A in ADPr-DNA:NAM-bound state with *Clostridium perfringens* iota-toxin (Ia)–actin complex (**left**, PDB ID: 4h0t) and PARP2 in the PARP2–HPF1 complex (**right**, PDB ID: 6tx3). H119 in DarT takes spatially the same position as Y375 in the iota-toxin (Ia) which was suggested to have a role in target protein, i.e. actin, recognition[41]. Both, Y375 and H119, are accommodated in the 'ARTT' loops, which do not show any similarity in either residue length or structural makeup. The approximate position of DarT H119 is occupied by E284 of HPF1 in the HPF1-PARP complex, whereby HPF1 sits on the 'ARTT' loop of PARP2. This leads to the formation of a composite active site with the catalytic glutamate residues E284 and E545 for catalysing serine ADP-ribosylation[25]. Enlarged views of the active sites are below the respective cartoon models. For clarity in the enlarged views, only the ADP-ribosylating turn-turn 'ARTT' loop from Ia-toxin and PARP2 and only a fragment of the respective binding partner, i.e. actin and HPF1 are shown as cartoon model. The substrate-coordinating and catalytic residues as well as the ADP-ribose products and complex-bound ligands are shown as sticks model. (rmsd (DarT-iota-toxin overlay): 2.71 Å; rmsd (DarT-PARP2 overlay): 2.58 Å).

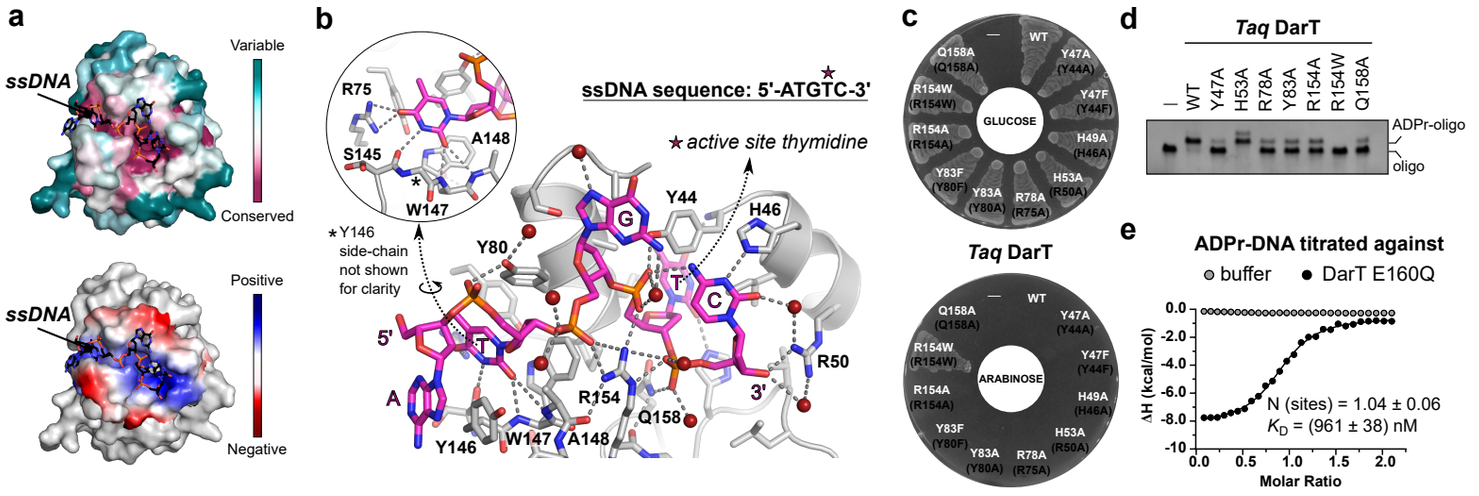
Extended Data Table 1. Data collection and refinement statistics for crystal structures of *Thermosipho africanus* DarTG and *Thermus* sp. 2.9 DarT described in this study.

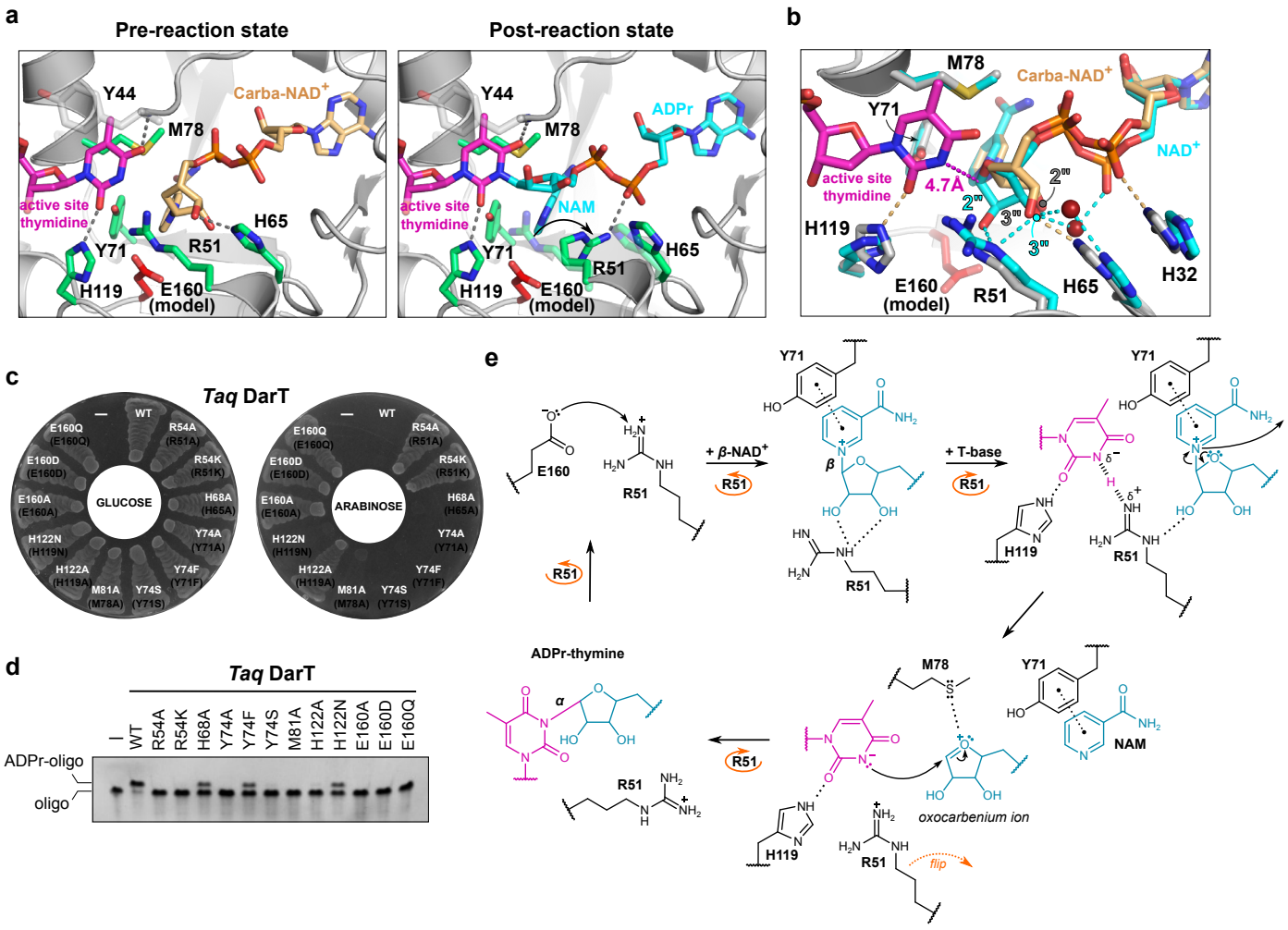
(a) Data for the highest resolution shell are given in parentheses.
(b) $R_{sym} = \Sigma|/-</>|/\Sigma/$, where / is measured density for reflections with indices *hkl*.

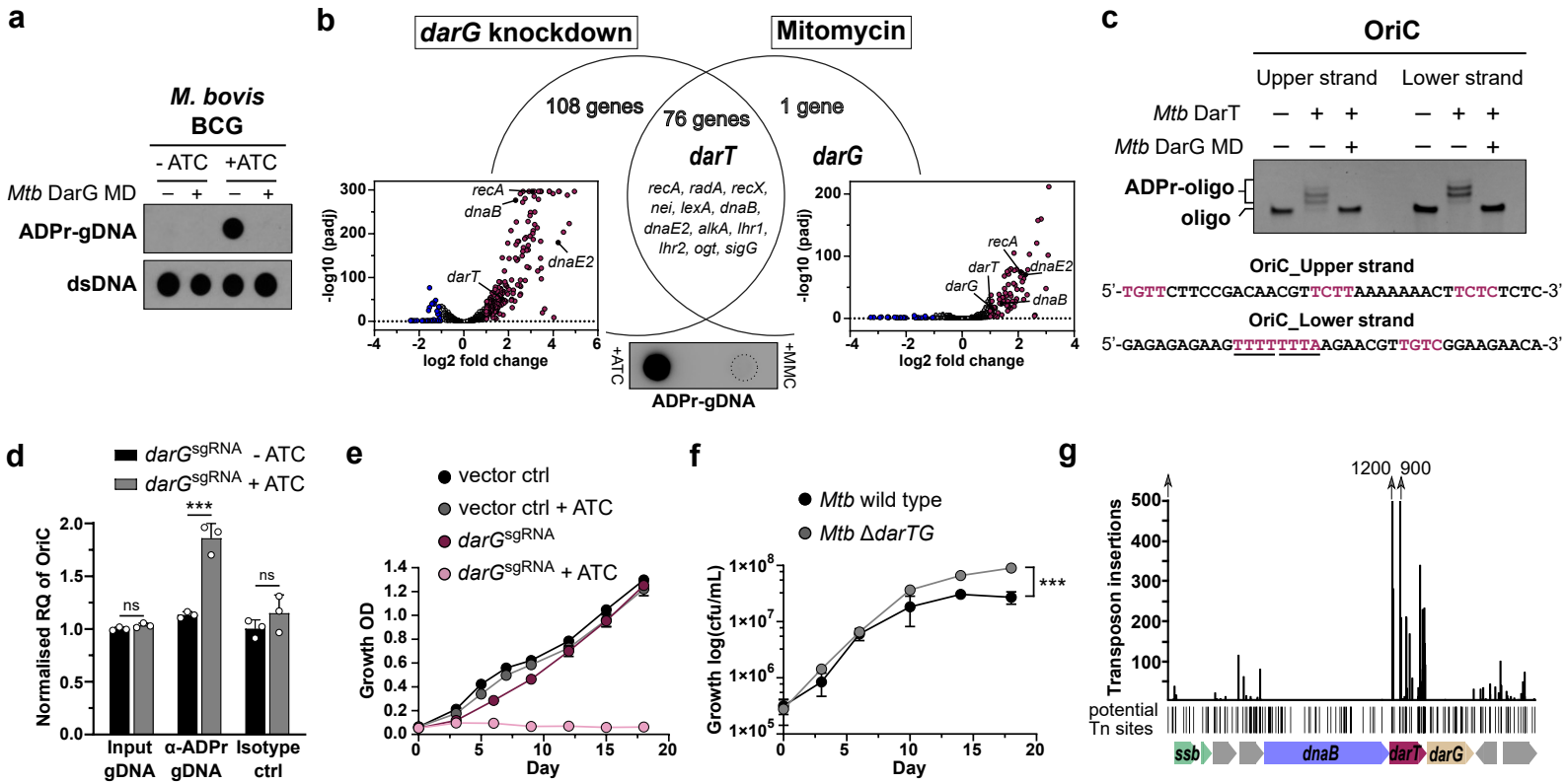
Extended Data Table 2. **Data collection and refinement statistics for DNA co-crystal structures of *Thermus* sp. 2.9 DarT described in this study.**

(a) Data for the highest resolution shell are given in parentheses.
(b) R_{sym} = Σ|/-</>|/Σ/, where / is measured density for reflections with indices *hkl*.


a
darT
darG
Toxin
(PARP-like)
Antitoxin
(macrodomain)
DarT
β-NAD+
ADPr
DarG
ssDNA
ADP-ribosylated ssDNA
b
Cholera toxin-like
ARTs (ARTCs)
Diphtheria toxin-like
bacterial ARTs
Ecto-ARTs
ARTDs
PARPs
CT
MTX LT
PT
Pierisin
Scabin
ScARP
ChT
ExoA
DT
DarT
ART5
ART3
ART4
ART1
PARP9
PARP7
PARP11
PARP12
PARP13
PARP10
PARP14
PARP15
PARP1
PARP2
PARP3
PARP4
PARP5A
PARP5B
0.5
c
Carba-NAD+ : DNA – bound
ADPr-DNA : NAM – bound
ART donor loop
C-terminus
ssDNA
Carba-NAD+
E160
N-terminus
'ARTT' loop
α (C1" - N3)
NAM
ADPr
d
H6
H1"
H1'
Me5
C2
C4


a
ssDNA
Variable
Conserved
ssDNA
Positive
Negative
b
R75
A148
S145
W147
★Y146
side-chain
not shown
for clarity
ssDNA sequence: 5'-ATGTC-3'★
★active site thymidine
Y44
H46
Y80
G
T
C
5'
T
A
R50
3'
R154
Q158
Y146
W147
A148
c
Taq DarT
GLUCOSE
ARABINOSE
WT
Y47A (Y44A)
Y47F (Y44F)
H49A (H46A)
H53A (R50A)
R78A (R75A)
Y83A (Y80A)
Y83F (Y80F)
R154A (R154A)
R154W (R154W)
Q158A (Q158A)
d
Taq DarT
—
WT
Y47A
H53A
R78A
Y83A
R154A
R154W
Q158A
ADPr-oligo
oligo
e
ADPr-DNA titrated against
buffer
DarT E160Q
ΔH (kcal/mol)
0.0
-2.0
-4.0
-6.0
-8.0
-10
0.0
0.5
1.0
1.5
2.0
Molar Ratio
N (sites) = 1.04 ± 0.06
K_D = (961 ± 38) nM


a
Pre-reaction state
Post-reaction state
Y44
M78
Carba-NAD+
ADPr
NAM
active site thymidine
Y71
R51
H65
H119
E160 (model)
b
M78
Carba-NAD+
Y71
NAD+
active site thymidine
4.7Å
H119
E160 (model)
R51
H65
H32
c
Taq DarT
GLUCOSE
ARABINOSE
WT
E160Q (E160Q)
R54A (R51A)
R54K (R51K)
H68A (H65A)
Y74A (Y71A)
Y74F (Y71F)
Y74S (Y71S)
M81A (M78A)
H122A (H119A)
H122N (H119N)
E160A (E160A)
E160D (E160D)
d
Taq DarT
—
WT
R54A
R54K
H68A
Y74A
Y74F
Y74S
M81A
H122A
H122N
E160A
E160D
E160Q
ADPr-oligo
oligo
e
E160
R51
+ β-NAD+
+ T-base
Y71
H119
ADPr-thymine
α
M78
NAM
oxocarbenium ion
flip


a
M. bovis BCG
- ATC
+ATC
Mtb DarG MD
– + – +
ADPr-gDNA
dsDNA
b
darG knockdown
Mitomycin
108 genes
76 genes
1 gene
darT
recA, radA, recX, nei, lexA, dnaB, dnaE2, alkA, lhr1, lhr2, ogt, sigG
darG
-log10 (padj)
log2 fold change
recA
dnaB
dnaE2
darT
darG
+ATC
+MMC
ADPr-gDNA
c
OriC
Upper strand
Lower strand
Mtb DarT
Mtb DarG MD
ADPr-oligo
oligo
OriC_Upper strand
5'-TGTTCTTCCGACAACGTTCTTAAAAAAACTTCTCTCTC-3'
OriC_Lower strand
5'-GAGAGAGAAGTTTTTTTAAGAACGTTGTCGGAAGAACA-3'
d
darGsgRNA - ATC
darGsgRNA + ATC
Normalised RQ of OriC
***
ns
ns
Input gDNA
α-ADPr gDNA
Isotype ctrl
e
vector ctrl
vector ctrl + ATC
darGsgRNA
darGsgRNA + ATC
Growth OD
Day
f
Mtb wild type
Mtb ΔdarTG
Growth log(cfu/mL)
***
Day
g
1200
900
Transposon insertions
potential Tn sites
ssb
dnaB
darT
darG