Electronic Supplementary Material (ESI) for Nanoscale.

Supplementary Information

New DNA Sensor System for Specific and Quantitative Detection of Mycobacteria

Oskar Franch*[1], Xiao Han*[2], Lærke Bay Marcussen*[1], Asger Givskov[1], Marie Bech Andersen[1], Adwait Anand Godbole[3], Charlotte Harmsen[1], Niels Nørskov-Lauritsen[4], Jonas Thomsen[1], Finn Skou Pedersen[1,5], Yilong Wang[6,7], Donglu Shi[6,7], Christian Wejse[8], Lone Pødenphant[4], Valakunja Nagaraja[3], Johanna Bertl[8], Magnus Stougaard[9], Yi-Ping Ho[10], Marianne Smedegaard Hede[1], Rodrigo Labouriau[8] Birgitta Ruth Knudsen[#1,5]

[1] Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

[2] School of Materials Science and Engineering, Tongji University, Shanghai, PR China

[3] Indian Institute of Science, Bangalore, India

[4] Department of Clinical Microbiology, Aarhus University Hospital, Aarhus, Denmark

[5] Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Aarhus, Denmark

[6] The Institute for Translational Nanomedicine, Shanghai East Hospital, the Institute for Biomedical Engineering & Nano Science, Tongji University School of Medicine, Shanghai, P. R. China

[7] The Institute for Translational Nanomedicine, Shanghai East Hospital, the Institute for Biomedical Engineering & Nano Science, Tongji University School of Medicine, Shanghai, P. R. China

[8]Department of Mathematics, Aarhus University, Aarhus, Denmark

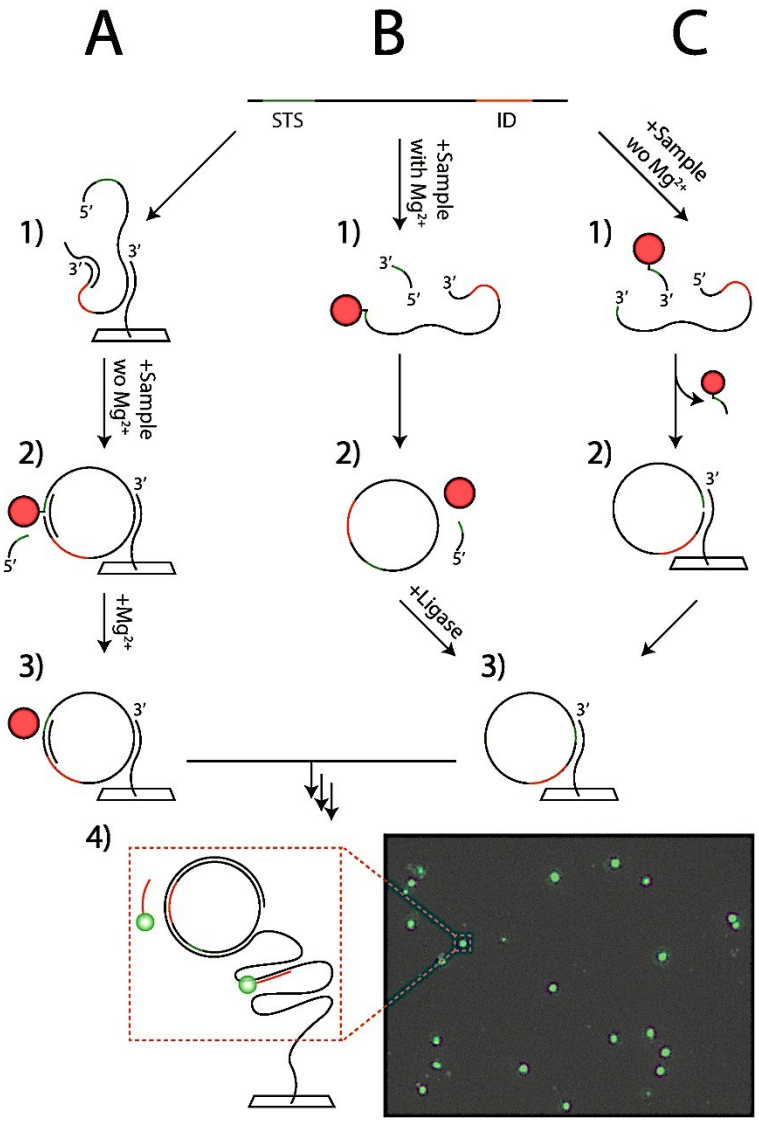
[9] Department of Clinical Medicine, Aarhus University Hospital, Aarhus, Denmark

[10] Department of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong SAR

* These authors contributed equally to this work

To whom correspondence should be addressed: brk@mbg.au.dk

Tested DNA substrate designs for mycobacteria TOP1A.



Supplementary Figure 1: An overview of three different designs that was tested in the attempt to develop a sensor system specific for TOP1A expressed in mycobacteria. In the first design (panel A), which is reminiscent to the functional design described in the main text, we used a solid support anchored DNA substrate composed of three DNA oligonucleotides (step 1). These were a solid support anchored "primer" DNA oligonucleotide, which also functioned as a primer for the subsequent RCA reaction (shown in steps 3-4), a scissile DNA oligonucleotide with a STS that could be cleaved by mycobacteria TOP1A and an additional DNA oligonucleotide annealed to the 3´-end of the scissile strand. This additional oligonucleotide

also had a sequence complementary to the 5′-DNA end generated after mycobacteria TOP1A cleavage and was believed to stimulate the ligation by mycobacteria TOP1A by bringing the 3'-OH end close to the cleavage complex (step 2). Hence, by using this substrate design we expected to increase the number of generated DNA circles and thereby the number or RCP signals. However, for unknown reasons the use of this substrate design resulted in the generation of only trance amounts of RCP signals even when using high concentrations of purified mycobacteria TOP1A (data not shown). Consequently, the design was discarded. Panel B, shows a schematic overview of an attempt to detect mycobacteria TOP1A activity without an isolation step to remove potential contaminants in the sample. In this setup, the substrate was composed of a single scissile strand with a STS that could be cleaved by mycobacteria TOP1A (step 1). Since the setup did not involve any isolation step Mg^{2+} was added together with the sample to support ligation that would generate a closed circle (step 2), which could subsequently be annealed to a surface attached primer and act as a template for RCA (step 3-4). This design was functional with purified mycobacteria TOP1A but no RCP signals above background could be observed when assaying mycobacteria TOP1A spiked in crude cell extracts (data not shown). The reason for this is most probably rapid degradation of the DNA substrate when incubated with crude cell extract in a Mg^{2+} containing buffer. Therefore, this design was given up on. The final assay design tested out is shown in panel C. The substrate in this design was composed of a single DNA oligonucleotide with a STS. Cleavage was suicidal and resulted in a short oligonucleotide covalently attached to mycobacteria TOP1A that diffused away from the rest of the substrate after cleavage. Hence, this assay setup did not benefit from the catalytic capability of mycobacteria TOP1A. Ligation of the cleaved substrate was accomplished by the addition of T4 ligase after hybridization to a surface attached primer (step 2-3). This procedure was expected to generate a circle that could act as a template for RCA (step 3-4). This setup generated large number of RCP signals even when no mycobacteria TOP1A was added (data not shown). The reason for the generation of false positive signals was most probably unspecific ligation. This problem could potentially have been solved by using a more specific ligase or by adding a 3'-blockage to the substrate. However, such modifications could not have circumvented the suicidal nature of the assay design and, consequently, the design was discarded.

In the figure, STS of the DNA substrates is shown by a green line, while the ID sequence that is identical to the sequence of the fluorescent labelled probe used for visualization of RCP signals is shown in red line. Mycobacteria TOP1A is shown by a red circle and the fluorescent labelled probe is illustrated by a red line attached to a green circle.

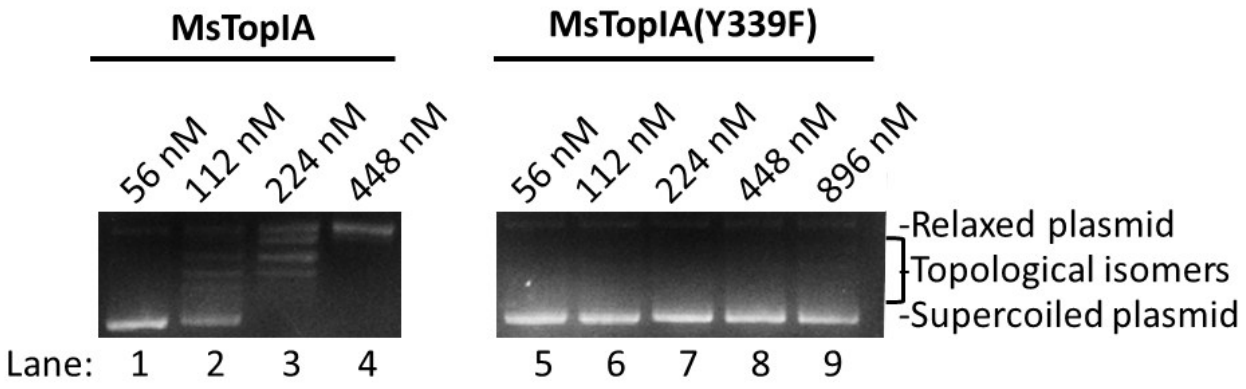
Relaxation activity of purified *M. smegmatis* TOP1A and *M. smegmatis* TOP1A(Y339F).

The activity of *M. smegmatis* TOP1A and the active site mutant of this enzyme, TOP1A(Y339F), was tested in a classical DNA relaxation assay.

This assay is based on the fact that a supercoiled DNA plasmid migrates faster in an agarose gel than the corresponding relaxed plasmid. Hence, it is possible to elucidate DNA relaxation activity of an enzyme by incubating it with a negatively supercoiled plasmid DNA and analyzing the product by electrophoresis in an agarose gel.

Supplementary Figure 2 shows an image obtained from gel electrophoretic analysis of 500 ng plasmid pUC18 DNA incubated with a titration of either *M. smegmatis* TOP1A (denoted MsTOP1A in the figure) (lane 1 to 4) or the active site mutant of this enzyme (denoted MsTOP1A(Y339F) in the figure) (lane 5 to 9) for 30 minutes at 37 °C in a buffer containing 40 mM Tris-HCl (pH 8.0), 20 mM NaCl, 5 mM $MgCl_2$, and 1 mM EDTA. The topological variants of the plasmid DNA were subsequently separated in a 1.2% agarose gel, which was then stained with ethidium bromide.

The result illustrates that incubation with *M. smegmatis* TOP1A (lane 1-4) resulted in the generation of relaxed plasmid DNA in a concentration dependent manner. As expected, the cleavage incompetent mutant, *M. smegmatis* TOP1A(Y339F), (lane 5 to 9) did not relax the plasmid. This result confirmed the catalytic activity of the utilized purified TOP1A and the catalytic inactivity of the utilized purified TOP1A(Y339F).



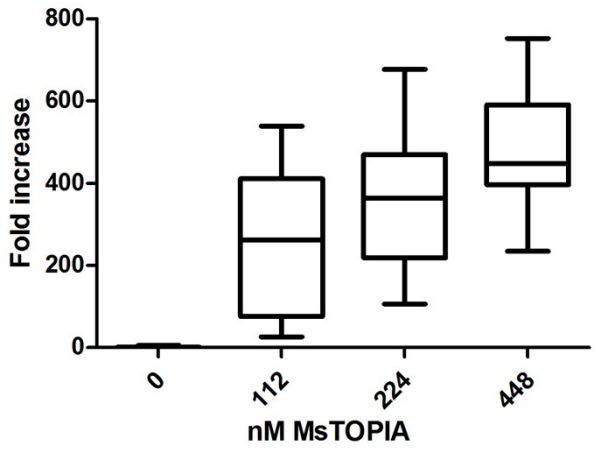
Supplementary Figure 2: Gel picture obtained from gel electrophoretic analysis of plasmid DNA incubated with a titration of either M. smegmatis TOP1A (denoted MsTOP1A in this figure) (lane 1 to 4) or the catalytic inactive mutant M. smegmatis TOP1A(Y339F) (denoted MsTOP1A(Y339) in this figure) (lane 5 to 9).

The mycobacteria sensor system detects mycobacterial TOP1A activity in a background of human cell lysate.

The ability of the mycobacteria sensor system to detect mycobacterial TOP1A activity in a background of human cell lysate was tested.

Human embryonic kidney (HEK293T) cells (kindly provided by Associate Professor Pia Møller Martensen, Department of Molecular Biology and Genetics, Aarhus University) were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum, 100 units/mL penicillin and 100 mg/mL streptomycin. Cells were incubated in a humidified incubator (5% CO_2) at 37 °C. At 70% confluency, the cells were harvested by trypsinization, washed in standard phosphate-buffered saline, and lysed in lysis buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT, and 1 mM PMSF) to give a final concentration of the extract corresponding to $6.6x10^6$ cells/mL. The lysate was spike-in with purified *M. smegmatis* TOP1A (concentrations indicated in the figure) and was subsequently analyzed using the mycobaceteria sensor system as described in Materials and Methods. The results were plotted as a function of the concentration of *M. smegmatis* TOP1A.

The results demonstrate the ability of the mycobacteria sensor system to detect mycobacterial TOP1A in a background of human cell lysate.

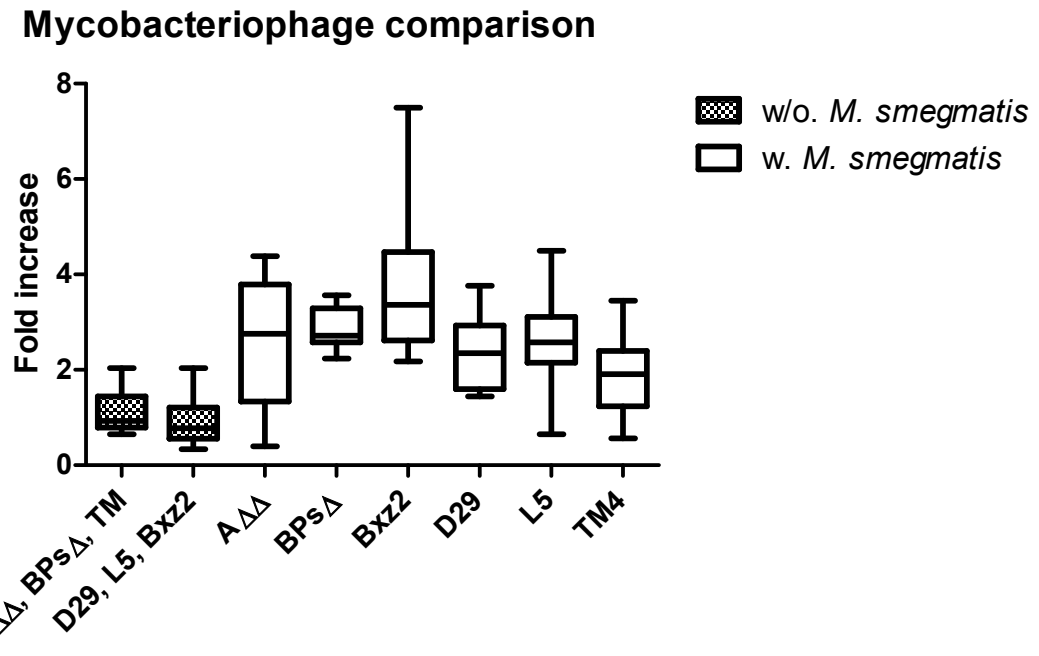


Supplementary Figure 3: Box plot showing the results of analysing the indicated concentrations of purified TOP1A from M. smegmatis in a background of human cell lysate. In order to allow comparison across slides the results are shown as fold increase over the average of the number of signals achieved from negative controls analysed on the same slide. This was done essentially as described previously and the equation below illustrates how the data points were processed.

$$Fold\ increase\ (Datapoint\ X) = \frac{Datapoint\ X}{Average(Negative\ controls)}$$

Comparison of various mycobacteriophages.

Growth media without Tween®80 (described in Materials and Methods) either without *M. smegmatis* (denoted w/o *M. smegmatis*, shown in checked box plots in Supplementary Figure 5) or containing 13000 CFU/µL of *M. smegmatis* (denoted w. *M. smegmatis,* shown in white box plots in Supplementary Figure 5) were supplemented with 1mM $CaCl_2$ (final concentration) before addition 17000 PFU/µL of either of the six mycobacteriophages AdephagiaΔ41,Δ43 (AΔΔ); BPsΔ33 (BPsΔ); Bxx2; D29; L5; TM4 as indicated in the figure. In the case of the samples without *M. smegmatis* the mycobacteriophages were grown individually and subsequently pooled as indicated in the figure and analysed in bulk. The infection and lysis was carried out for two hours at 37 °C. Mycobacterial TOP1A was released from the DNA by adding of NaCl to a final concentration of 560 mM followed by centrifugation at 20800 g for 5 minutes. Subsequently, the generated extracts were analysed using the mycobacteria sensor system and the results depicted in the box plot below. These results demonstrate that all the tested mycobacteriophages were able to lyse mycobacteria and allowed detection of the mycobacterial TOP1A by the mycobacteria sensor system.



Supplementary Figure 4: Box plot representing the results from analysing *M. smegmatis* lysed by infection with mycobacteriophages in the sensor system. The results of analysing mycobacteria lysed by either AdephagiaΔ41,Δ43 (AΔΔ); BPsΔ33 (BPsΔ); Bxx2; D29; L5; TM4 are shown by white boxes. The results of analysing of a pool of the mycobacteriophages without *M. smegmatis* are shown by checked boxes. In order to allow comparison across slides the results are shown as fold increase over the average of the number of signals achieved from negative controls analysed on the same slide. This was done essentially as described previously and the equation below illustrates how the data points were processed.

$$Fold\ increase\ (Datapoint\ X) = \frac{Datapoint\ X}{Average(Negative\ controls)}$$

Supplementary Table S1. Adjusted number signals (graphically depicted in Figure 2) for several concentrations of wild type *M. smegamatis* TOP1A (denotet wt in the table) and for the catalytic inactive mutant TOP1A(Y339F) (denoted mut in the table) that was used as a negative control. Column 1 indicates the identity and concentration of enzyme. Column 2 shows the estimated number of signals, including also the respective 95% confidence intervals in parenthesis and the statistical significance groups, at a 5% significance level, labelled by small letters. The p-values for pairwise comparisons (adjusted for multiple testing by the FDR mehtod) are shown in collumns 3-8.

		p-values					
conc.	Adj. n. signals	mut.	7	14	28	57	114
mut.	14.52(-31.68-60.71)a						
7 wt	18.45(-28.55-65.44)a	1					
14 wt	44.3(-8.6-97.19)a	0,981387	0,991386				
28 wt	50.14(-7.41-107.7)a	0,964709	0,981313	0,999999			
57 wt	83.63(28.22-139.04)a	0,49376	0,57506	0,952375	0,982708		
114 wt	304.88(246.54-363.21)b	1,27E-13	3,59E-13	9,05E-10	9,96E-09	6,06E-07	
227 wt	644.61(590.49-698.72)c	0	0	0	0	0	9,99E-16

Supplementary Table S2. Adjusted number of signals (shown in Figure 3) obtained when analysing active and heat inactivated (marked with "-I") extracts from different myco-bacteria species and four pools of non-mycobacteria species (control 1-4). Column 1 indicates the sample identity. Column 2 shows the estimated number of signals, including also the respective 95% confidence intervals in parenthesis and the statistical significance groups, at a 5% significance level, labelled by small letters. The p-values for pairwise comparisons are shown in collumns 3-11 (adjusted for multiple testing by the FDR mehtod).

		p-values								
Sample	Adj. n. signals	Control 1	Control 2	Control 3	Control 4	M. abscessus-I	M. abscessus	M. chelongae-I	M. chelongae	M. bolletii-I
Control 1	2.57(1.61-4.09)a									
Control 2	2.75(1.91-3.96)a	1								
Control 3	3.17(2.25-4.47)a	0,9973	0,999362							
Control 4	3.53(2.5-5)a	0,953582	0,961754	0,999907						
M. abscessus-I	2.79(2.08-3.74)a	0,999999	1	0,99939	0,950606					
M. abscessus	12.58(9.87-16.03)b	<0.000001	<0.000001	<0.000001	<0.000001	<0.000001				
M. chelongae-I	5.04(3.53-7.18)a	0,277308	0,196698	0,522465	0,84507	0,056927	0,000002			
M. chelongae	11.81(8.6-16.2)b	<0.000001	<0.000001	<0.000001	<0.000001	<0.000001	0,999993	0,000165		
M. bolletii-I	3.98(2.75-5.76)a	0,854682	0,860522	0,992137	0,999959	0,728289	<0.000001	0,978748	<0.000001	
M. bolletii	14.66(10.75-20)b	<0.000001	<0.000001	<0.000001	<0.000001	<0.000001	0,988806	0	0,938823	0

Supplementary Table S3. Adjusted number signals (graphically depicted in Figure 4) for increasing concentrations of *M. smegmatis* (CFU/μL). Column 1 displays the concentrations. Column 2 shows the adjusted number of signals including the respective 95% confidence intervals in parenthesis and the statistical significance groups, at a 5% significance level, labelled by small letters. The p-values for pairwise comparisons are shown in collumn 3-7 (adjusted for multiple testing by the FDR mehtod).

		p-values				
CFU/μL	Adj. n. signals	0	450	900	3600	7200
0	1.201(0.836-1.726)a					
450	1.792(1.314-2.443)ab	0,458506				
900	2.358(1.777-3.128)b	0,019304	0,605533			
3600	2.829(2.165-3.696)b	0,000643	0,069443	0,842406		
7200	5.029(3.997-6.329)c	<0.000001	<0.000001	0,000001	0,000166	
14400	7.654(6.304-9.294)d	<0.000001	<0.000001	<0.000001	<0.000001	0,000377

Supplementary Statistics

Detailed description of the Poisson mixed model used for modelling the number of signals: The number of signals per frame reported in each of the sequences reported (Figures 2, 3 and 4, Tables 1, 2, S1, S2 and S3) were modelled using a Poisson mixed model (see Jørgensen et al. 1996, Madsen et al. 2014) as described below. Denote by Y_{tsr} the random variable representing the number of signals per frame of the r^{th} repetition of the s^{th} slide/experiment subject to the t^{th} treatment (with variable index sets according to the experimental design employed). According to the Poisson mixed model used, those random variables are conditionally independent and Poisson distributed given two Gaussian distributed random components, U taking the same value for each observation arising from the same slide/experiment and V taking different value for each observation. Moreover, for a given treatment t, repetition r, and slide/experiment s the conditional expectation of Y_{tsr} is given by

$$E(Y_{tsr} \mid U = u_s, V = v_{tsr}) = \tau_t + u_s + v_{tsr}.$$

The specification of the model is completed by stating that $U \sim N(0, \sigma_s^2)$, $V \sim N(0, \sigma^2)$ and $Cov(U, V) = 0$. Using standard properties of conditional expectations (i.e. integrating with respect to the density of the distribution of U and the density of the distribution of V) yields that the expectation of Y_{tsr} is

$$E(Y_{tsr}) = \tau_t.$$

Therefore, the parameter τ_t is the expected value of the number of counts of the observations submitted to the t^{th} treatment, adjusted for possible effects of the slide/experiment (via the random effect U). Moreover, basic properties of generalised linear mixed models imply that the observations arising from the same slide/experiment are correlated due to the presence of the random component U. The random component V induces over-dispersion allowing the model to correctly describe data subject to a large amount of noise due to the presence of unknown uncontrolled factors affecting the responses. Note that the model described above corresponds to a generalised linear mixed model defined with the Poisson distribution the identity link function, one fixed effect indicating the treatment and two Gaussian random components: one taking the same value for each observation arising from the same slide/experiment (the random component U) and one taking a different value for each observation (the random component V).

References

B. Jørgensen, R. Labouriau, and S. Lundbye-Christensen. *Journal of the Royal Statistics Society*, 1996, B **58**, 573-592.

P. Madsen, J. Jensen, R. Labouriau, O. F. Christensen, G. Sahana. DMU - A Package for Analyzing Multivariate Mixed Models in quantitative Genetics and Genomics. 10th World Congress on Genetics Applied to Livestock Production (WCGALP), 2014, Vancouver, Canada.