

Cite this: *RSC Chem. Biol.*, 2025, 6, 772

# Development of a fluorescence-based assay for RecBCD activity using functional data analysis and design of experiments†

Adam Winnifrith,<sup>a</sup> Steven R. Brown,<sup>b</sup> Piotr Jedryszek,<sup>c</sup> C. Grant,<sup>b</sup> Philip E. Kay,<sup>d</sup> Adam M. Thomas,<sup>a</sup> Jacob D. Bradbury<sup>a,e</sup> and Thomas Lanyon-Hogg<sup>ib</sup>\*<sup>a</sup>

Biochemical assays are essential tools in biological research and drug discovery, but optimisation of these assays is often a challenging and lengthy process due to the wide range of input variables and the complex effects of these variables on one another. Traditional 'one-factor-at-a-time' optimisation is both time-consuming and fails to explore the full range of input combinations. In contrast, the modern 'design of experiments' (DoE) approach enables simultaneous investigation of multiple input variables and their interactions, leading to more information-rich and efficient experimentation. We therefore sought to apply DoE to the optimisation of a new fluorescence-based assay for the enzyme RecBCD, a helicase–nuclease–ATPase complex involved in bacterial stress responses. A novel 'functional data analysis' (FDA) approach was used to predict the shape of RecBCD reaction curves in response to different combinations of input variables, which successfully identified assay conditions suitable for drug screening. Collectively, this work delivers a new assay for the antibiotic target RecBCD and demonstrates the potential of DoE and FDA to accelerate biochemical assay development.

Received 29th November 2024,  
Accepted 12th March 2025

DOI: 10.1039/d4cb00291a

rsc.li/rsc-chembio

## Introduction

Analysis of complex biological systems is crucial for both fundamental research and drug discovery. Biochemical assays to study the function of individual proteins or protein complexes are essential tools in both contexts. However, the inherent complexity of these biochemical systems necessitates extensive effort in the design, analysis, and optimisation of assays due to the wide range of potential input variable combinations. The traditional approach to assay optimisation is by changing one variable (*e.g.*, temperature, pH, reagent concentration) and selecting the value that gives the optimal response, before changing the next variable (Fig. 1A). However, this 'one-factor-at-a-time' (OFAT) approach is both time-consuming and inefficient as only a limited proportion of the potential variable combinations is explored. Additionally, OFAT

overlooks possible interactions between different experimental variables in the assay response.

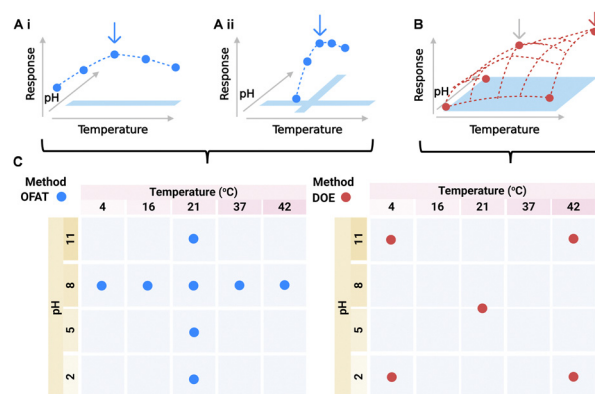


Fig. 1 Assay optimisation strategies. (A) Schematic representation of a traditional 'one-factor-at-a-time' (OFAT) optimisation. In this example, (i) the first input factor (temperature) is varied with a fixed value of the second input factor (pH), and the optimal temperature found; (ii) the second input factor, pH, is varied at the selected temperature to identify the apparent optimal combination of conditions. (B) Design of experiments (DoE) explores the full range of design space for both factors simultaneously. In this example, DoE identifies the global optimum at both high pH and high temperature, which is overlooked by OFAT, and achieves this with fewer experiments and fewer conditions tested. (C) Exploration of input factor design space using OFAT and DoE strategies.

<sup>a</sup> Department of Pharmacology, University of Oxford, OX1 3QT, UK.  
E-mail: thomas.lanyon-hogg@pharm.ox.ac.uk

<sup>b</sup> Synthace, 4th Floor The Westworks, 195 Wood Lane, London, W12 7FQ, UK

<sup>c</sup> Department of Biology, University of Oxford, OX1 3SZ, UK

<sup>d</sup> SAS Institute, Wittington House, Henley Road, Medmenham Marlow, Buckinghamshire, SL7 2EB, UK

<sup>e</sup> Ineos Oxford Institute for Antimicrobial Research, University of Oxford, OX1 3RE, UK

† Electronic supplementary information (ESI) available: Supplementary data and biological methods. See DOI: <https://doi.org/10.1039/d4cb00291a>

'Design of experiments' (DoE) was developed in the mid-20th century for agriculture field trials and chemical engineering, and has since expanded into a versatile tool for process and product optimisation.<sup>1,2</sup> DoE provides a mathematical framework to investigate the effect of multiple variables and their interactions by systematically changing all input variables simultaneously, and using statistical models to make predictions about how each individual variable affects the output response.<sup>3,4</sup> By using statistical experimental design to determine the smallest set of conditions required to examine the effect of multiple factors with sufficient statistical power, DoE overcomes the challenge of needing to test all possible factor combinations.<sup>5</sup> As a result, in contrast to OFAT exploration that follows one-dimensional lines in a multidimensional design space (Fig. 1A), DoE explores the overall response surface (Fig. 1B). DoE can enhance the information gathered from experiments and allow identification of optimal variable combinations potentially overlooked by OFAT, using fewer experimental conditions (Fig. 1C).

In DoE, independent experimental input variables whose values are controlled are termed 'factors', the values of factors selected for testing are termed 'levels', and the output dependent variables are termed 'responses'. To suit different experimental needs and constraints, pre-calculated design strategies have been developed, which include but are not limited to: full factorial, fractional factorial, response surface, mixture, Taguchi array, and split plot designs.

DoE has been applied in industrial drug discovery<sup>6</sup> to optimise cell based assays,<sup>7,8</sup> enzyme linked immunosorbent assays,<sup>9</sup> and recombinant protein production and purification,<sup>4,10</sup> as well as in design of hybrid synthetic promoters in yeast.<sup>11</sup> Yet despite successful implementation in industry, DoE lacks widespread adoption in academic research, where resource constraints mean efficient experimental design may yield most benefits.

The development of new medicines for bacterial infections is a field that would particularly benefit from more efficient academic research. Antimicrobial resistance (AMR) is one of the most serious threats to modern medicine and yet the current antibiotic pipeline is insufficient to meet this growing challenge due to both technical and economic challenges.<sup>12</sup> Use of any new antibiotic is limited to mitigate resistance emergence; however, this disincentivises commercial antibiotic development, as low unit sales diminish the appeal for investors. Consequently, much early-stage antibiotic research is now conducted in academia, where resources and funding are limited.<sup>13</sup> New methodologies that generate information-rich datasets rapidly, and with reduced resources, are therefore important to accelerate drug discovery efforts in this setting.

We therefore sought to use DoE to optimise a new biochemical assay for the emerging antibiotic target RecBCD, a DNA-damage repair enzyme that initiates the bacterial 'SOS response' and drives resistance evolution.<sup>14</sup> D-optimal and space-filling DoE designs were used to rapidly identify factors affecting RecBCD activity and conditions that lead to a suitable assay signal for high-throughput screening.<sup>2</sup> A novel 'functional data analysis' (FDA) approach was used to model the effect of different input factor combinations on the shape of the RecBCD

assay response curve. The presented DoE workflow is also compatible with automation, digital experimental design, and machine learning, to conduct assay optimisation in days or weeks rather than months.

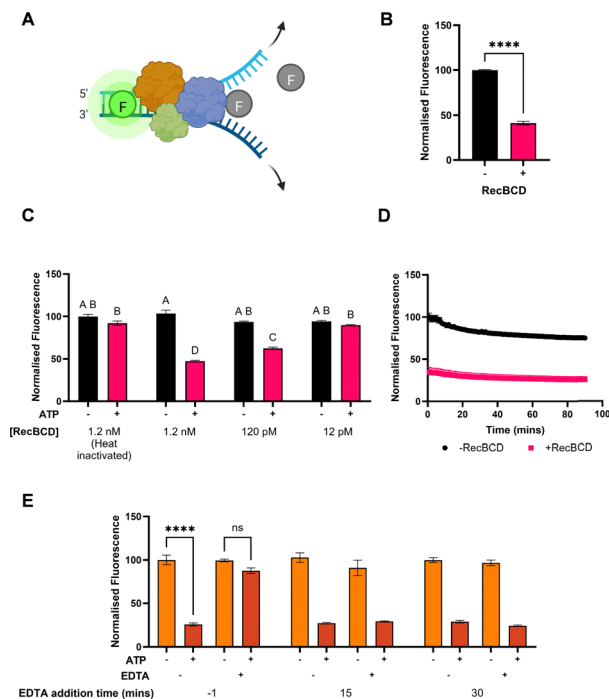
## Results and discussion

### Fluorescence-based assay for RecBCD activity

RecBCD is helicase–nuclease enzyme complex that initiates repair of double-stranded DNA (dsDNA) breaks in bacteria and activates the mutagenic 'SOS response', making this enzyme an attractive therapeutic target.<sup>14</sup> RecBCD separates dsDNA into single strands and cuts these DNA strands in an ATP-dependent manner.<sup>15</sup> We therefore sought to establish a biochemical assay for RecBCD using the fluorogenic dsDNA dye QuantiFluor (Promega, US), where enzymatic processing of dsDNA would result in decreased fluorescence (Fig. 2A). A reaction mixture was prepared containing RecBCD (4.8 nM), ATP (1 mM), NEBuffer 4 (1×) and single stranded DNA binding protein (SSB, 1.2 μM) to prevent reannealing of DNA strands if reaction conditions promoted helicase activity over nuclease cutting. Following addition of QuantiFluor dye and Lambda DNA (5 ng μL<sup>-1</sup>) to initiate the reaction, a significant decrease in fluorescence was observed in the presence of RecBCD following 3 h incubation (Fig. 2B). The inability of RecBCD to fully remove all double-stranded Lambda DNA signal under these assay conditions may result from the complementary 12 base-pair overhangs on the ends of double-stranded Lambda DNA, which would result in a certain proportion of cyclic dsDNA that cannot be processed by RecBCD. To confirm assay signal was dependent on enzyme activity, RecBCD was heat-inactivated, resulting in abrogation of fluorescence decrease (Fig. 2C). RecBCD contains two ATP-dependent helicase subunits, and correspondingly no change in signal was observed in the absence of ATP; assay signal in the presence of ATP was also dependant on RecBCD concentration (Fig. 2C).

Real-time measurements of fluorescence for 90 min following initiation of the RecBCD reaction revealed little change in signal over time, suggesting the enzymatic activity had reached completion before the first measurement (Fig. 2D). Given the known rapid processivity of RecBCD,<sup>16,17</sup> stop conditions were therefore sought to confirm the time dependence of the dsDNA fluorescent assay signal. EDTA inhibits RecBCD through chelation of Mg<sup>2+</sup> ions essential for helicase<sup>18,19</sup> and nuclease activity.<sup>20</sup> Addition of EDTA (0.1 M) before the enzyme reaction was initiated resulted in no decrease in fluorescence signal compared to no-ATP controls; EDTA addition 15 or 30 min after initiation of the enzyme reaction resulted in decreased fluorescence only in the presence of ATP (Fig. 2E). Inhibition of RecBCD activity by EDTA was dose-dependent with IC<sub>50</sub> = 7 mM (Fig. S1, ESI†). Alternative DNA substrates were also tested in the assay; Lambda DNA provided the best signal window compared to other tested DNA substrates (Fig. S2, ESI†) and was therefore used in subsequent experiments. The QuantiFluor dsDNA signal was also stable to dimethyl sulfoxide (DMSO) concentrations <10% (v/v) typically used in drug discovery assays (Fig. S3, ESI†).

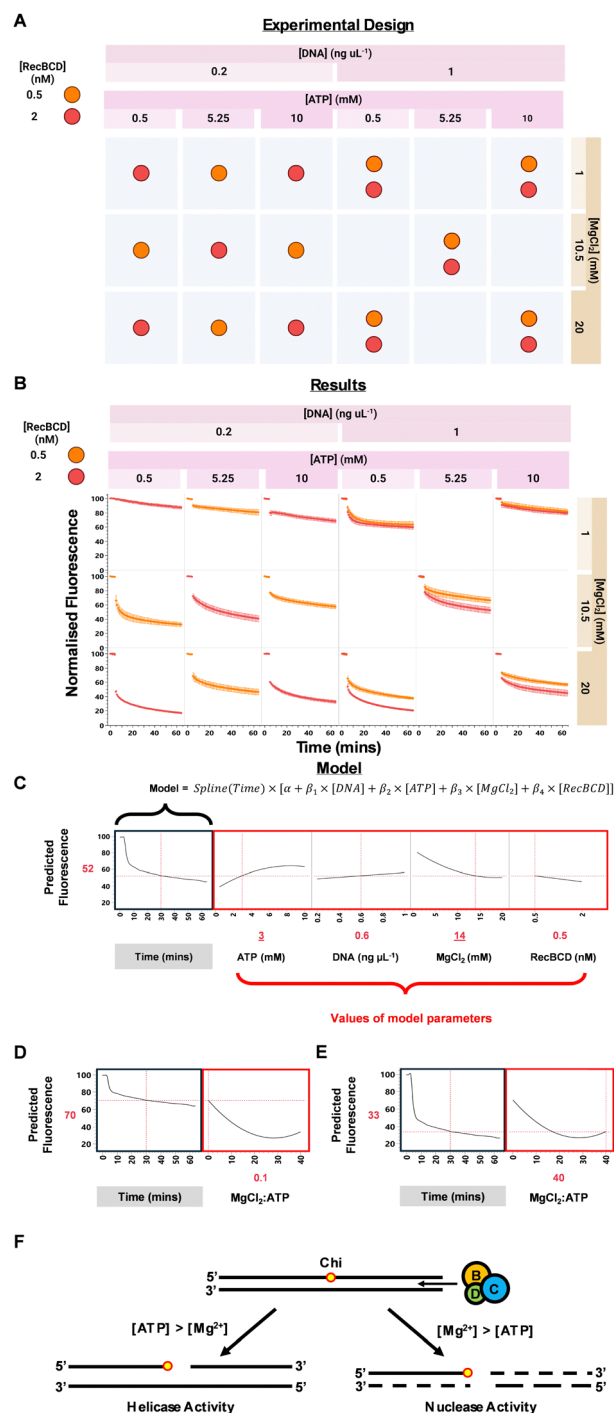




**Fig. 2** Fluorescence-based assay for RecBCD activity. (A) Schematic of the RecBCD assay. QuantiFluor dsDNA dye fluoresces when bound to dsDNA; RecBCD processing of dsDNA is detected as a decrease in fluorescence. (B) RecBCD activity measured as described in methods with RecBCD (4.8 nM), ATP (1 mM), Lambda DNA (5 ng  $\mu\text{L}^{-1}$ ), SSB (1.2  $\mu\text{M}$ ), and NEB Buffer 4 (1  $\times$ ). QuantiFluor (1 : 400 dilution, 50  $\mu\text{L}$ ) was added after 3 h. *t*-Test, \*\*\*\**P* < 0.05. (C) RecBCD activity after 2 min was measured following heat inactivation at 95 °C for 30 min or in dose–response. The assay was performed as described in methods with varying concentrations of RecBCD with or without ATP (1 mM). Two-way ANOVA main effects (RecBCD and ATP) and interaction effect, multiple pairwise comparisons using Turkey's test; conditions sharing a common letter (A–D) were not statistically significantly different. (D) Assay signal measured kinetically over 90 min. The assay was performed as in B, but QuantiFluor (50  $\mu\text{L}$ ) was added immediately after initiating the reaction. (E) Inhibition of RecBCD activity by EDTA. Reactions were prepared as described in methods, with or without ATP (1 mM). EDTA was added to a final concentration of 0.1 M 1 min before, or 15 and 30 min after, addition of ATP. QuantiFluor (20  $\mu\text{L}$ ) was added after the final EDTA stop. Three-way ANOVA main effects (EDTA, ATP, time) and interaction effect, multiple pairwise comparisons using Turkey's test, \*\*\*\**P* < 0.0001, ns, nonsignificant. Data represents means  $\pm$  SEM, *n*  $\geq$  3.

### DoE exploration of kinetic data

Having demonstrated that the QuantiFluor dsDNA fluorescence-based assay can assess RecBCD activity, but that the enzyme activity was too rapid for inhibitor screening, we employed a constrained D-optimal design of experiments approach to analyse the effect of different input factors and optimise the assay response. A D-optimal design algorithm selects an optimal set of conditions by maximising the determinant of the information matrix, thereby minimising parameter estimate uncertainty. Designs were generated within input factor constraints using either JMP or Synthace software platforms, ensuring the design was both statistically efficient and feasible to execute.



**Fig. 3** Design of experiments (DoE) and functional data analysis (FDA) exploration of RecBCD activity. (A) Experimental design matrix showing combinations of RecBCD, DNA, ATP, and MgCl<sub>2</sub> concentrations tested. (B) Time-course fluorescence results for each condition. (C) FDA model predicting fluorescence based on input factors. (D) and (E) FDA model predictions showing the effect of MgCl<sub>2</sub>:ATP ratio on fluorescence decrease: (D) low MgCl<sub>2</sub>:ATP ratio (1:10), (E) high MgCl<sub>2</sub>:ATP ratio (20:1). (F) Schematic illustrating how MgCl<sub>2</sub>:ATP ratio affects RecBCD activity, favouring helicase or nuclease activity, adapted from ref. 15.

Initially, a four-parameter experimental design was generated with RecBCD (0.5–2 nM), DNA (0.2–1 ng  $\mu\text{L}^{-1}$ ), MgCl<sub>2</sub> (1–20 mM),



and ATP (0.5–10 mM) as input factors, resulting in 19 unique conditions covering the four-dimensional experimental design space (Fig. 3A and Table S1, ESI†). Each condition was tested with 3–6 replicates, with fluorescence recorded for 5 min prior to ATP addition and then for a subsequent 60 min after addition (Fig. 3B).

Functional data analysis (FDA)<sup>21,22</sup> was used to model how the input factor combinations affected the shape of the assay response curve. Unlike kinetic analyses that reduce time-course data to single summary statistics (*e.g.* initial reaction rates or area under the curve), FDA models the entire enzyme reaction curve shape in response to different input factor levels. FDA therefore allows creation of interactive plots that predict the complete assay response curve for any combination of input factors (Fig. 3C). The JMP interactive Prediction Profiler was used to visualise the effect of different combinations of input factors on the response curve shape (Video S1, ESI†).

The FDA model generated insights consistent with previous validation experiments (Fig. 2), including that increased RecBCD generated a larger signal window between conditions with and without RecBCD (Fig. 3C). The model also showed that higher DNA concentrations increased raw fluorescence, while higher MgCl<sub>2</sub> concentrations decreased fluorescence across the entire curve, possibly due to Mg<sup>2+</sup> interference with dye fluorescence (Fig. S4, ESI†).

Further, FDA was capable of identifying two-factor interactions in the ratio of MgCl<sub>2</sub>:ATP concentration that affected the response curve shape. Increasing the MgCl<sub>2</sub>:ATP ratio led to a greater decrease in fluorescence, consistent with enhanced dsDNA degradation. When MgCl<sub>2</sub> = 1 mM and ATP = 10 mM (ratio 0.1), the FDA model predicted a 30% decrease in normalised fluorescence after 30 min (Fig. 3D). In contrast, when MgCl<sub>2</sub> = 20 mM and ATP = 1 mM (ratio 20), a 67% decrease in normalised fluorescence after 30 min was predicted (Fig. 3E). This MgCl<sub>2</sub>:ATP ratio effect is consistent with previous literature on RecBCD function.<sup>15,23–25</sup> When ATP concentration exceeds Mg<sup>2+</sup>, RecBCD primarily exhibits helicase activity; conversely, when Mg<sup>2+</sup> concentration is higher, nuclease activity is favoured (Fig. 3F). This modulation is thought to result from ATP coordinating Mg<sup>2+</sup> at low Mg<sup>2+</sup>:ATP ratios, with Mg<sup>2+</sup> being essential for nuclease activity.<sup>15</sup>

### High-dimensional experimentation DoE

To explore a broader range of variables on the assay signal, ten input factors which potentially impacted enzyme activity were investigated (Fig. 4A and Table S2, ESI†).<sup>26</sup> The design criteria covered input factors RecBCD (0–2 nM), DNA (2–5 ng μL<sup>-1</sup>), MgCl<sub>2</sub> (1–20 mM), ATP (1–100 μM), pH (6–8), BSA (0–30 g L<sup>-1</sup>), DTT (0–1 mM), NaCl *vs.* KCl, monovalent cation chloride salt concentration (1–100 mM), and reaction volume (40–80 μL, Fig. 4A).

120 combinations of input factors were tested in triplicate, with positive (+RecBCD) and negative (–RecBCD) controls. A space-filling design was used to evenly cover the whole experimental design space, ensuring that all possible combinations of factors were adequately represented (Fig. S5, ESI†).

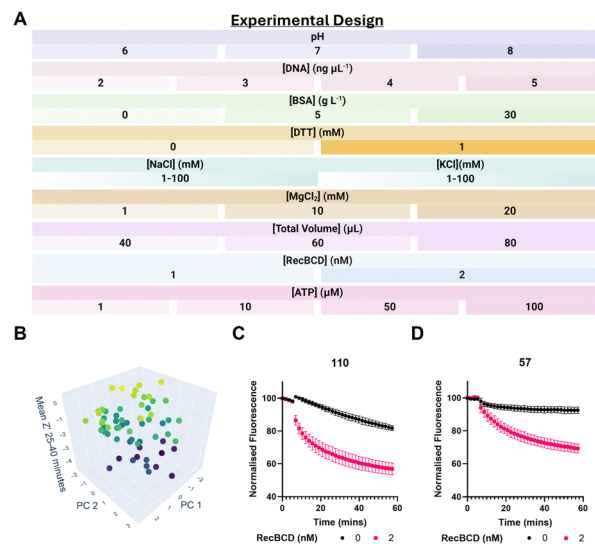


Fig. 4 High-dimensional experimentation exploration of RecBCD assay. (A) Conditions explored within the experimental design space. Assay were performed as described in the methods. (B) Principal component analysis (PCA) of 10 factors from A plotted against mean  $Z'$  (25–40 min) for each condition. Data points are coloured based on mean  $Z'$  from blue (low  $Z'$ ) to yellow (high  $Z'$ ). Conditions with  $Z' < -5$  excluded. (C) and (D) Highest  $Z'$  conditions. Data represents means  $\pm$  SD,  $n \geq 3$ .

This method is particularly useful when exploring a high-dimensional space, as it maximises the diversity of conditions tested, helping to identify interactions and nonlinear effects. Experimental conditions were also randomised across the assay plate to minimise edge effects.

Most conditions showed no change in fluorescence (Fig. S6, ESI†); however, FDA identified trends such as  $> 10$  μM ATP being required for assay signal, in line with previous reports.<sup>27</sup>  $Z'$  is a statistical measure of assay quality for high-throughput screening, with  $Z' \geq 0.5$  indicating an excellent assay.<sup>28</sup> To identify assay conditions compatible with small molecule screening, the  $Z'$  was calculated for each condition and time-point (see ESI†), with mean  $Z'$  (25–40 min) used for ranking.

Principal component analysis (PCA) of the ten input factors against mean  $Z'$  revealed that whilst most conditions had  $Z' < 0$ , there were 8 conditions with  $Z' > 0$  spread across the design space (Fig. 4B). ‘Hit’ conditions 110 ( $Z' = 0.32$ ) and 57 ( $Z' = 0.25$ ) were selected for further optimisation (Fig. 4C and D, see Table S2 for details of conditions, ESI†). In condition 110, the decrease in no RecBCD fluorescent signal over time may result from quenching of dye fluorescence under the assay conditions.

### DoE analysis of local areas of design space

Conditions 110 and 57 had different concentrations of DTT (1 vs. 0 mM), BSA (0 vs. 5 mg mL<sup>-1</sup>), NaCl (37 vs. 11 mM), DNA (5 vs. 3 ng μL<sup>-1</sup>), and pH (8 vs. 6), respectively. As the number of factors was reduced and their levels more well-defined, a constrained D-optimal design was used to systematically explore the local design space around these two conditions (Fig. S7, ESI†). 95 factor combinations were tested with input factors DNA (3, 5 ng μL<sup>-1</sup>), pH (6, 8), BSA (0, 2.5, 5 g L<sup>-1</sup>), DTT



(0, 1 mM), and NaCl (10, 20, 40 mM), with four replicates each (Fig. 5A and Table S3, ESI<sup>†</sup>). The MgCl<sub>2</sub> : ATP ratio was fixed at 200 : 1 (10 mM MgCl<sub>2</sub>, 50 μM ATP) to promote nuclease activity, as the nuclease activity of RecBCD requires both helicase and ATPase functions and can therefore be used as a readout when screening for inhibitors of any of these RecBCD functions.

The focused local design space yielded 72% of conditions with  $Z' > 0$  (Fig. 5A), compared to 8% in the high-dimensional screen of conditions (Fig. S6, ESI<sup>†</sup>). FDA indicated BSA, pH, and DNA concentration had the largest effects on assay signal

(Fig. 5B): pH 6 reduced RecBCD activity, BSA stabilised fluorescence at low DNA concentrations in the absence of RecBCD, and high BSA concentrations decreased the assay signal window between positive (+RecBCD) and negative (−RecBCD) controls (Video S2, ESI<sup>†</sup>). DTT and NaCl concentration had minimal effects on RecBCD activity. To check consistency between FDA and other statistical methods, a random forest model was constructed to predict fluorescence at 30 min based on RecBCD and the DoE factors. This model indicated RecBCD, BSA, and pH had feature importances of 53%, 23%, and 16%,

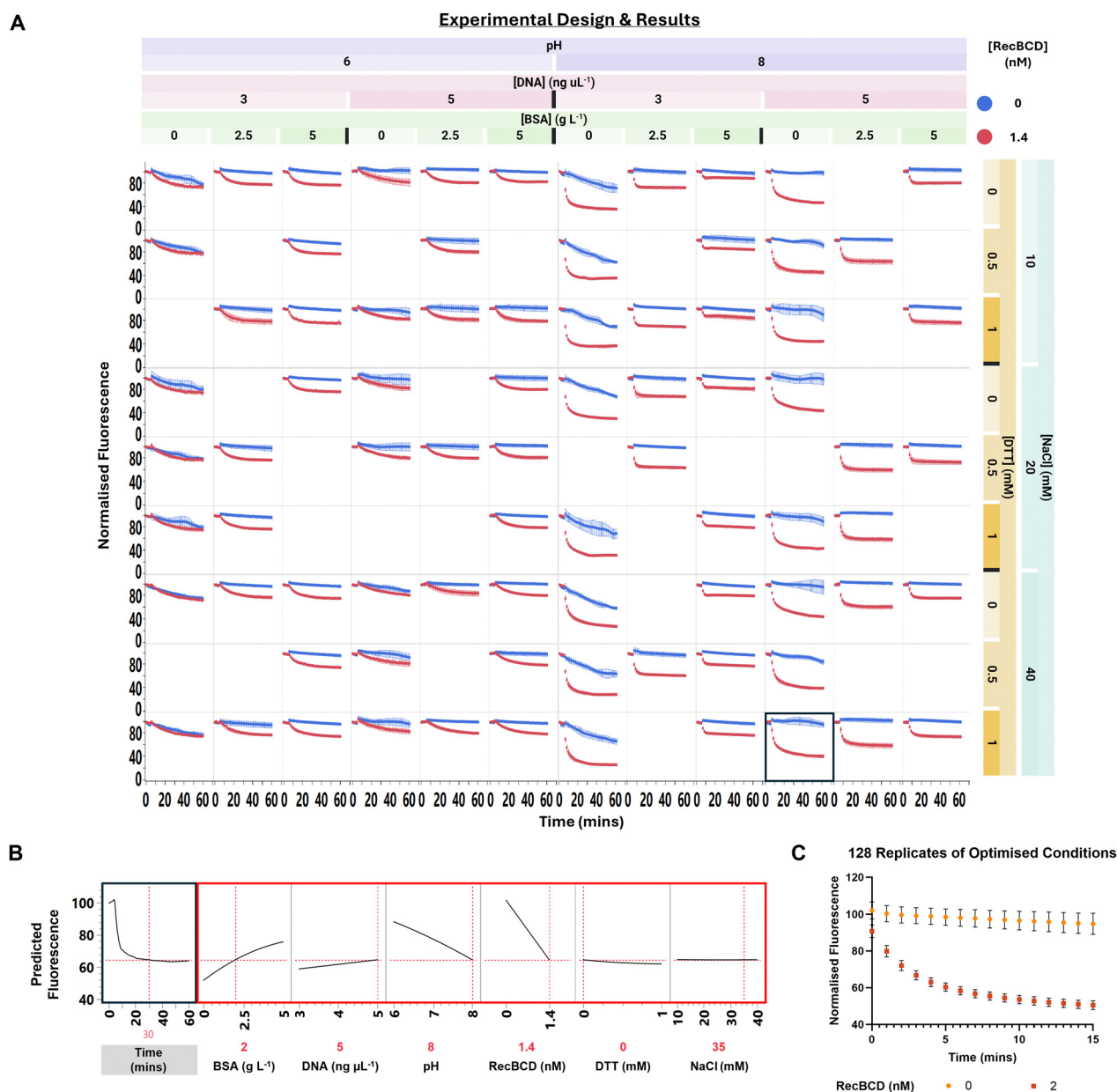


Fig. 5 Optimisation and validation of RecBCD assay conditions. (A) Experimental design and results of localized DoE experiment. Normalised fluorescence plotted against time for 95 variations of conditions 110 and 57, focusing on DNA (3, 5 ng μL<sup>-1</sup>), pH (6, 8), BSA (0, 2.5, 5 g L<sup>-1</sup>), DTT (0, 0.5, 1 mM), and NaCl (10, 20, 40 mM). Results show the effects of these variables on RecBCD assay signal, with best performing condition 82 boxed in bold. Data represents means ± SD,  $n \geq 4$ . (B) Functional data analysis model of raw data from (A), showing effects of BSA, pH, and DNA on assay signal. (C) Validation of best performing condition 82 (DNA = 5 ng μL<sup>-1</sup>, BSA = 0 mg mL<sup>-1</sup>, pH = 8, NaCl = 40 mM, DTT = 1 mM) with 128 replicates of positive and negative controls. Data represents means ± SD.



respectively ( $R^2 = 0.92$ , root mean squared error {RMSE} = 5.5, Fig. S8, ESI†). A linear model of the same prediction task showed significant interactions between pH and RecBCD ( $p = 0.0058$ ), BSA and RecBCD ( $p = 0.0023$ ), and pH and BSA-RecBCD ( $p = 0.00077$ ), suggesting RecBCD activity was sensitive to pH and BSA levels.<sup>20,29</sup> To check the robustness of the optimised assay condition, the best performing condition was repeated with 128 replicates, yielding a  $Z'$  after 10 min of 0.59 (Fig. 5C), highlighting the reproducibility of the assay.

## Conclusions

Biochemical assays present challenges in their analysis and optimisation due to the scale and interactions of potential input factors. This complexity makes assay optimisation both time consuming and expensive. To help address this challenge, we sought to employ a combination of DoE, automation, and digital experiment design to explore complex design spaces in a biochemical assay. As a test-case, we developed a new fluorescence-based assay for the helicase–nuclease enzyme RecBCD, an emerging target in antibiotic discovery.<sup>30–33</sup>

A single four-parameter DoE was capable of identifying conditions that promoted RecBCD nuclease activity at increased  $Mg^{2+}$ :ATP ratios (Fig. 3), consistent with prior literature using OFAT analysis.<sup>15,20,23,29</sup> Under reaction conditions where RecBCD was active, the reaction plateaued with 40% of the starting fluorescence signal remaining. This may suggest the presence of remaining dsDNA, potentially as a result of circular Lambda DNA formed from complimentary base pairs. Boiling of Lambda DNA to produce fully linear dsDNA prior to initiation of the RecBCD reaction may therefore provide a larger signal window and also increase the enzyme reaction kinetics. Expanding DoE analysis to ten input factors allowed screening of multiple conditions simultaneously and identified factors with no impact on assay readout as well as assay conditions yielding  $Z' > 0$ . A focused exploration of ‘hit’ conditions identified a reaction set up with robust assay performance ( $Z' = 0.59$  with 128 replicates).

We build on previous literature applying DoE to assay optimisation in two ways. Firstly, we introduce the application of functional data analysis (FDA) to modelling and optimising kinetic responses. FDA allows prediction of the effect of different input factor combinations on the response curve shape, offering insights into assay behaviour and enabling optimisation towards specific reaction profiles, such as linear reaction rates. Secondly, we combine D-optimal designs, automation, and digital experiment platforms to rapidly explore large areas of assay design space, identifying complex, multi-dimensional assay response surfaces with distinct optima. A key advantage of this approach was the ability to design and implement high-dimensional DoE experiments within one week using Synthace’s digital experiment platforms, which integrate DoE designs, liquid handling instructions, experiment simulations, and metadata tracking. This platform is also compatible with incorporation of reagent costs, such that an assay can be

optimised to maximise signal within the available financial budget for reagents.

The presented experimental approach therefore provides advantages in efficiency of both data gathering and data interpretation compared to conventional OFAT and DoE optimisation. Widespread adoption could substantially impact biochemical assay optimisation and drug discovery, particularly in resource-limited settings such as antibiotic development.

## Author contributions

A. W., S. B., C. G., and T. L.-H. designed experiments. A. W., P. J., and P. K. analysed data. A. W. and S. B. performed biochemical testing. T. L.-H. was responsible for funding acquisition. S. B., C. G., A. M. T., J. D. B. and T. L.-H. supervised the research. A. W., J. D. B. and T. L.-H. wrote the manuscript with input from all authors.

## Data availability

The data supporting this article have been included as part of the ESI.†

## Conflicts of interest

S. R. M. and C. G. are employees of Synthace LTD, UK. P. E. K. is an employee of JMP Statistical Discovery LLC, US. A. W. and P. J. are founders of and hold shares in Evolvere Biosciences LTD, UK.

## Acknowledgements

JMP Pro software used in the design, visualisation and analysis of DoE experiments was generously provided through a free license from JMP Statistical Discovery LLC, US. A. W. and J. D. B. are supported by the Wellcome Trust (Grant Number: 218514/Z/19/Z), Merck Sharp and Dohme Corp. and Janssen Pharmaceutica NV. J. D. B. is supported by the Ineos Oxford Institute for Antimicrobial Research. A. M. T. and T. L.-H. were supported by a Springboard Award from the Academy of Medical Sciences (AMS), the Government Department of Business, Energy and Industrial Strategy (BEIS), the British Heart Foundation and Diabetes UK [SBF007\100164]. T. L.-H. is supported by a Career Development Fellowship from the Department of Pharmacology, University of Oxford. The authors thank Prof. Gerald Smith and Dr Susan Amundsen (Fred Hutchinson Cancer Center) for technical advice, plasmids and feedback on the manuscript, and Dr Nuno Leitão and Dr Markus Gershater (Synthace LTD, UK) for assistance with assays and enabling the collaboration. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.



## References

- G. W. Snedecor, W. G. Cochran and W. G. Cochran, *Statistical methods*, Iowa State Univ. Press, Ames, Iowa, 8th edn, 7. print., 1996.
- D. C. Montgomery, *Design and Analysis of Experiments*, Wiley, Hoboken, NJ, 10th edn, 2020.
- G. E. P. Box and N. R. Draper, *Response Surfaces, Mixtures, and Ridge Analyses*, John Wiley, Hoboken, N. J., 2nd edn, 2007.
- C. Papanephytou, Design of Experiments As a Tool for Optimization in Recombinant Protein Biotechnology: From Constructs to Crystals, *Mol. Biotechnol.*, 2019, **61**, 873–891.
- S. A. Weissman and N. G. Anderson, Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications, *Org. Process Res. Dev.*, 2015, **19**, 1605–1633.
- H. Tye, Application of statistical ‘design of experiments’ methods in drug discovery, *Drug Discovery Today*, 2004, **9**, 485–491.
- X. C. Chen, L. Zhou, S. Gupta and F. Civoli, Implementation of design of experiments (DOE) in the development and validation of a cell-based bioassay for the detection of anti-drug neutralizing antibodies in human serum, *J. Immunol. Methods*, 2012, **376**, 32–45.
- C. A. Ray, V. Patel, J. Shih, C. Macaraeg, Y. Wu, T. Thway, M. Ma, J. W. Lee and B. Desilva, Application of multi-factorial design of experiments to successfully optimize immunoassays for robust measurements of therapeutic proteins, *J. Pharm. Biomed. Anal.*, 2009, **49**, 311–318.
- D. Joelsson, P. Moravec, M. Troutman, J. Pigeon and P. DePhillips, Optimizing ELISAs for precision and robustness using laboratory automation and statistical design of experiments, *J. Immunol. Methods*, 2008, **337**, 35–41.
- R. S. Islam, D. Tisi, M. S. Levy and G. J. Lye, Framework for the Rapid Optimization of Soluble Protein Expression in Escherichia coli Combining Microscale Experiments and Statistical Experimental Design, *Biotechnol. Prog.*, 2007, **23**, 785–793.
- S. R. Brown, M. Staff, R. Lee, J. Love, D. A. Parker, S. J. Aves and T. P. Howard, Design of Experiments Methodology to Build a Multifactorial Statistical Model Describing the Metabolic Interactions of Alcohol Dehydrogenase Isozymes in the Ethanol Biosynthetic Pathway of the Yeast *Saccharomyces cerevisiae*, *ACS Synth. Biol.*, 2018, **7**, 1676–1684.
- 2020 antibacterial agents in clinical and preclinical development, <https://www.who.int/publications-detail-redirect/9789240021303>, (accessed 19 May 2023).
- M. Miethke, M. Pieroni, T. Weber, M. Brönstrup, P. Hammann, L. Halby, P. B. Arimondo, P. Glaser, B. Aigle, H. B. Bode, R. Moreira, Y. Li, A. Luzhetskyy, M. H. Medema, J.-L. Pernodet, M. Stadler, J. R. Tormo, O. Genilloud, A. W. Truman, K. J. Weissman, E. Takano, S. Sabatini, E. Stegmann, H. Brötz-Oesterhelt, W. Wohlleben, M. Seemann, M. Empting, A. K. H. Hirsch, B. Loretz, C.-M. Lehr, A. Titz, J. Herrmann, T. Jaeger, S. Alt, T. Hesterkamp, M. Winterhalter, A. Schiefer, K. Pfarr, A. Hoerauf, H. Graz, M. Graz, M. Lindvall, S. Ramurthy, A. Karlén, M. van Dongen, H. Petkovic, A. Keller, F. Peyrane, S. Donadio, L. Fraisse, L. J. V. Piddock, I. H. Gilbert, H. E. Moser and R. Müller, Towards the sustainable discovery and development of new antibiotics, *Nat. Rev. Chem.*, 2021, **5**, 726–749.
- T. Lanyon-Hogg, Targeting the bacterial SOS response for new antimicrobial agents: drug targets, molecular mechanisms and inhibitors, *Future Med. Chem.*, 2021, **13**, 143–155.
- G. R. Smith, How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination: A Molecular Biologist’s View, *Microbiol. Mol. Biol. Rev.*, 2012, **76**, 217–228.
- A. Taylor and G. R. Smith, Unwinding and rewinding of DNA by the RecBC enzyme, *Cell*, 1980, **22**, 447–457.
- J. Wiktor, M. van der Does, L. Büller, D. J. Sherratt and C. Dekker, Direct observation of end resection by RecBCD during double-stranded DNA break repair in vivo, *Nucleic Acids Res.*, 2018, **46**, 1821–1833.
- V. Mackay and S. Linn, Selective inhibition of the dnase activity of the recBC enzyme by the DNA binding protein from Escherichia coli., *J. Biol. Chem.*, 1976, **251**, 3716–3719.
- L. J. Roman and S. C. Kowalczykowski, Characterization of the helicase activity of the Escherichia coli recBCD enzyme using a novel helicase assay, *Biochemistry*, 1989, **28**, 2863–2873.
- M. Wright, G. Buttin and J. Hurwitz, The Isolation and Characterization from Escherichia coli of an Adenosine Triphosphate-dependent Deoxyribonuclease Directed by rec B, C Genes, *J. Biol. Chem.*, 1971, **246**, 6543–6555.
- J. Gertheiss, D. Rügamer, B. X. W. Liew and S. Greven, Functional Data Analysis: An Introduction and Recent Developments, *Biomet. J.*, 2024, **66**(7), e202300363.
- Y. Li, Y. Qiu and Y. Xu, From multivariate to functional data analysis: Fundamentals, recent developments, and emerging areas, *J. Multivar. Anal.*, 2022, **188**, 104806.
- T. L. Pavankumar, A. K. Sinha and M. K. Ray, Biochemical characterization of RecBCD enzyme from an Antarctic Pseudomonas species and identification of its cognate Chi ( $\chi$ ) sequence, *PLoS One*, 2018, **13**, e0197476.
- S. K. Amundsen and G. R. Smith, RecBCD enzyme: mechanistic insights from mutants of a complex helicase–nuclease, *Microbiol. Mol. Biol. Rev.*, 2023, **87**, e00041-23.
- D. B. Wigley, Bacterial DNA Repair: Recent Insights into the Mechanism of RecBCD, AddAB and AdnAB, *Nat. Rev. Microbiol.*, 2013, **11**, 9–13.
- P. J. Goldmark and S. Linn, Purification and properties of the recBC DNase of Escherichia coli K-12, *J. Biol. Chem.*, 1972, **247**, 1849–1860.
- R. Zananiri, S. Mangapuram Venkata, V. Gaydar, D. Yahalom, O. Malik, S. Rudnizky, O. Kleifeld, A. Kaplan and A. Henn, Auxiliary ATP Binding Sites Support DNA Unwinding by RecBCD, *Nat. Commun.*, 2022, **13**, 1806.
- J.-H. Zhang, T. D. Y. Chung and K. R. Oldenburg, A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays, *SLAS Discovery*, 1999, **4**, 67–73.



- 29 D. C. Eichler and I. R. Lehman, On the role of ATP in phosphodiester bond hydrolysis catalyzed by the recBC deoxyribonuclease of *Escherichia coli*, *J. Biol. Chem.*, 1977, **252**, 499–503.
- 30 J. D. Bradbury, T. Hodgkinson, A. M. Thomas, O. Tanwar, G. L. Monica, V. V. Rogga, L. J. Mackay, E. K. Taylor, K. Gilbert, Y. Zhu, A. Y. Sefton, A. M. Edwards, C. J. Gray-Hammerton, G. R. Smith, P. M. Roberts, T. R. Walsh and T. Lanyon-Hogg, Development of an inhibitor of the mutagenic SOS response that suppresses the evolution of quinolone antibiotic resistance, *Chem. Sci.*, 2024, **15**, 9620–9629.
- 31 C. S. Q. Lim, K. P. Ha, R. S. Clarke, L.-A. Gavin, D. T. Cook, J. A. Hutton, C. L. Sutherland, A. M. Edwards, L. E. Evans, E. W. Tate and T. Lanyon-Hogg, Identification of a Potent Small-Molecule Inhibitor of Bacterial DNA Repair That Potentiates Quinolone Antibiotic Activity in Methicillin-Resistant *Staphylococcus Aureus*, *Bioorg. Med. Chem.*, 2019, **27**, 114962.
- 32 S. K. Amundsen, T. Spicer, A. C. Karabulut, L. M. Londoño, C. Eberhardt, V. F. Vega, T. D. Bannister, P. Hodder and G. R. Smith, Small-Molecule Inhibitors of Bacterial AddAB and RecBCD Helicase–nuclease DNA Repair Enzymes, *ACS Chem. Biol.*, 2012, **7**, 879–891.
- 33 A. C. Karabulut, R. T. Cirz, A. F. Taylor and G. R. Smith, Small-molecule sensitization of RecBCD helicase–nuclease to a Chi hotspot-activated state, *Nucleic Acids Res.*, 2020, **48**, 7973–7980.

