Modeling Whole-Cell Bioluminescent Biosensor

Dr. Lingchong You
BME574: Engineering & Modeling Gene Circuits
Bora Keresteci, Phong Le, Minxue Lin
December 8th, 2025

ABSTRACT

Whole-cell biosensors have the potential to provide inexpensive, field-deployable detection of environmental contaminants, but their practical utility depends on accurately converting a complex cellular response into a quantitative estimate of analyte concentration. Here, we present a mechanistic forward-inverse modeling framework for an *Escherichia coli* recA::*luxCDABE* sensor engineered to report nalidixic acid (NA)-induced DNA damage via the SOS pathway (Daniel et al., 2008). Unlike earlier simplified models, this report develops a nonlinear ODE model to include the full sensing cascade: NA-induced DNA damage, RecA activation, SOS-driven transcription, gene expression, and Michaelis-Menten substrate cycling. This model provides a mechanistic mapping from NA concentration to a time-resolved luminescence trajectory.

However, this mapping cannot be inverted analytically, and the original study did not report discrete photon-count time series. To address this, a subsequent inverse modeling framework is constructed. Synthetic ground-truth trajectories with 8% parameter perturbations and measurement noise were then generated to evaluate identifiability. A hybrid global-local solver (Differential Evolution + nonlinear least squares) consistently recovered unique NA values across a broad mid-range of concentrations (10^{-6} to 10^{-2} μM). In this mid-range, fitted trajectories closely matched ground-truth data, and bootstrap uncertainty analysis produced tight confidence intervals (e.g., median 1.96×10^{-3} μM, 95% CI 1.80-2.12$\times10^{-3}$ μM fit for a true concentration of 2.00×10^{-3} μM). Across all tested concentrations, the inverse model deviates substantially from ground truth at the extremes of the NA domain, revealing failure modes that delineate the sensor's practical analytical limits. This framework establishes how a mechanistic SOS-based reporter can be quantitatively inverted to recover analyte concentrations and provides a template for evaluating identifiability and analytical limits in other whole-cell biosensor pathways.

INTRODUCTION

Whole-cell biosensors leverage genetically engineered organisms to detect environmental stimuli and transduce them into measurable outputs, such as fluorescence, bioluminescence, or electrochemical signals. They are highly versatile, since they can be constructed from diverse regulatory pathways present in bacteria. Compared to traditional chemical sensors, biosensors offer advantages including low cost, self-replication, and enhanced specificity using natural regulatory pathways (Bazhenov et al., 2023).

As a result, these whole-cell biosensors have become increasingly attractive for field deployment, especially environmental monitoring of contaminants. In particular, biosensors allow real-time monitoring of toxicological pathways, making them useful in identifying DNA damage, oxidative stress, heat shock, or other cellular responses to environmental hazards (Bazhenov et al., 2023). Daniel et al. (2008) engineered an *Escherichia coli* strain carrying a

*recA::*luxCDABE transcriptional fusion. It demonstrated that nalidixic acid (NA), a DNA gyrase inhibitor that creates single-stranded DNA, induces a measurable & predictable bioluminescent response through the SOS response pathway.

Their work established the viability of SOS-driven lux reporters and introduced a simplified kinetic model of light production (Figure 1), providing initial insight into system dynamics. However, the model focused primarily on post-promoter kinetics and did not capture the regulatory layers of the SOS pathway or provide numerical photon-count time-series data. As a result, it did not address the inverse problem of estimating NA concentration from a measured light trajectory (Figure 2).

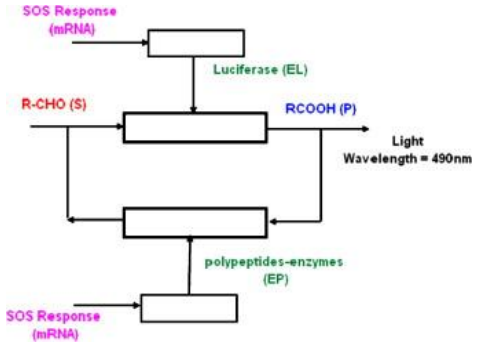

Figure 1. Daniel et al. (2008) simplified kinetic model

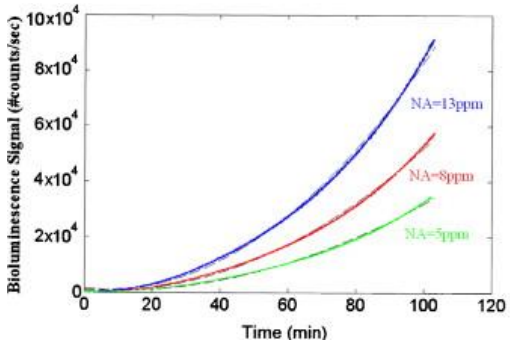

Figure 2. Modeled light trajectory vs time from different initial [NA]

This gap is significant because the utility of whole-cell biosensors depends not only on producing a measurable signal, but also on being able to quantitatively infer the analyte concentration. These systems often exhibit nonlinear dynamics, with delayed dynamics from transcription & translation, saturation of enzymatic capacity, and substrate cycling, complicating direct inversion without advanced modeling. Addressing this gap requires a more rigorous forward model of the cellular pathways. The light signal is not an instantaneous reporter of NA concentration but instead emerges from a cascade that includes NA-induced DNA damage, RecA activation, promoter-driven transcription, mRNA turnover, protein expression, and the enzymatic substrate cycling of the luciferase-reductase complex. To address this need, we develop a mechanistic forward model and a corresponding inverse model to estimate NA concentrations from luminescence trajectories. The inverse framework combines nonlinear optimization with residual bootstrapping to estimate NA and quantify estimation uncertainty, enabling the identification of a practical operating range. Together, these components provide a framework to generalize and quantitatively analyze SOS-driven lux reporters, and more broadly, pathway-driven whole-cell biosensors.

RESULTS

Biosensor Model

We developed an ordinary differential equation (ODE) model to simulate the dynamic response of the E. coli recA::*luxCDABE* biosensor to nalidixic acid (Fig. 3). NA first induces single-stranded DNA (ssDNA) breaks, leading to the accumulation of ssDNA. The accumulation of ssDNA triggers the activation of the RecA protein (RecA*). This active form promotes the expression of *lux* genes, producing luciferase (EL) and reductase complex (EP). The final module represents the bioluminescent reaction as a closed enzymatic loop where the aldehyde (S) is converted to fatty acid (P) by luciferase, and then recycled back to aldehyde by reductase complex. In the module, the photon emission rate is directly proportional to the reaction rate of the aldehyde to fatty acid conversion, which we use as the main readout of biosensor activity. In all simulations, the plotted light signal corresponds to the modelled reaction rate of aldehyde to fatty acid, used as a proxy for photon emission.

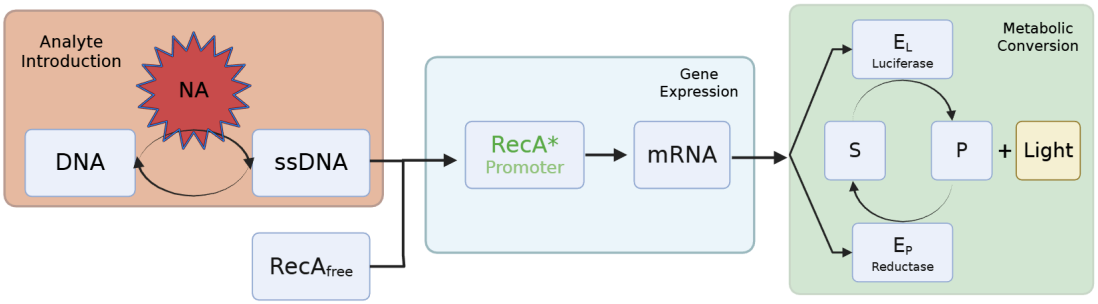

Figure 3. Key mechanisms of the biosensor system.

With no toxin input, the system maintains a low and stable basal expression. Before introducing NA, the concentrations of ssDNA, RecA*, and lux enzymes settle at a low, basal steady state, resulting in a negligible background luminescence (Fig. 4). When NA is added, the model predicts a distinct dynamic response. The slight time lag between toxin introduction and the rise in light output reflects the biological time required for the transcription and translation to occur (Fig. 4D). Following the delay, the light intensity increases rapidly as the enzymatic loop amplifies the signal, and it eventually saturates due to limited cellular resources. The biosensor response to a range of input NA concentrations, higher NA leads to a faster rise and a higher final stable state (Fig. 5).

The fitted parameters control both the sensitivity and the timing of the biosensor response. The DNA damage parameters determine how quickly ssDNA is created and repaired. The RecA activation rates, together with the RecA production and degradation terms, control the duration of the RecA* burst. Transcription and translation parameters control the speed and

amplitude of reporter expression. The enzyme kinetic parameters set the upper limit of light signal that the biosensor can produce.

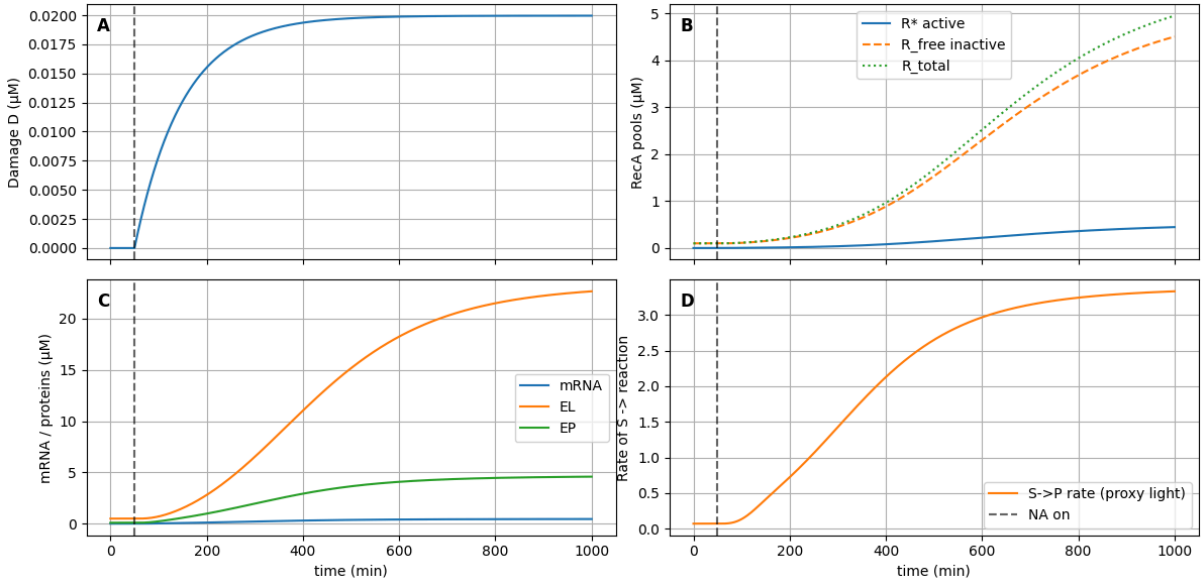

Figure 4. Dynamics of the biosensor system with input NA = 2 nM. NA is added at t = 50 min (grey dashed line). *(A)* Time course of accumulated ssDNA damage. *(B)* Time course of active RecA* (R*), inactive free RecA (R_free), and total RecA. *(C)* Time courses of mRNA, luciferase, and reductase concentrations. *(D)* Reaction rate of S-to-P conversion, as a proxy for light output.

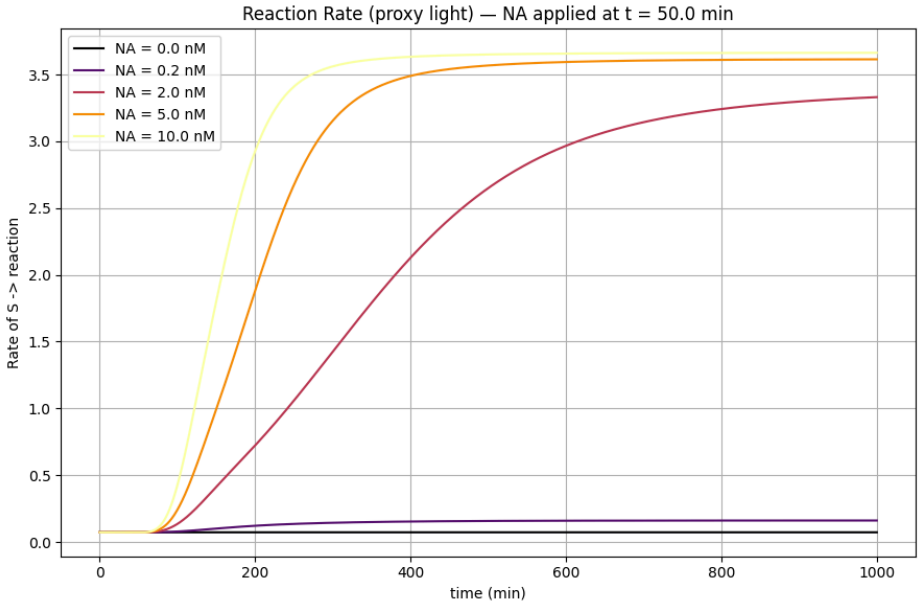

Figure 5. Proxy light signal trajectories for different input NA concentrations (NA = 0, 0.2, 2, 5, 10 nM). NA is added at t = 50 min.

Pseudo-Experiment

To test whether the model could predict NA from realistic measurements, we generated synthetic data because no experimental data were available. We started from a base parameter set, representing the biosensor response of an idealized cell. A perturbed parameter set was then created by perturbing each rate constant in the base set by ±8%. This avoids the inverse crime, as it simulates a slightly different real cell. Then, Gaussian noise with a standard deviation equal to 8% of the signal range was added to the perturbed model. The noisy datapoints were treated as pseudo-experimental measurements.

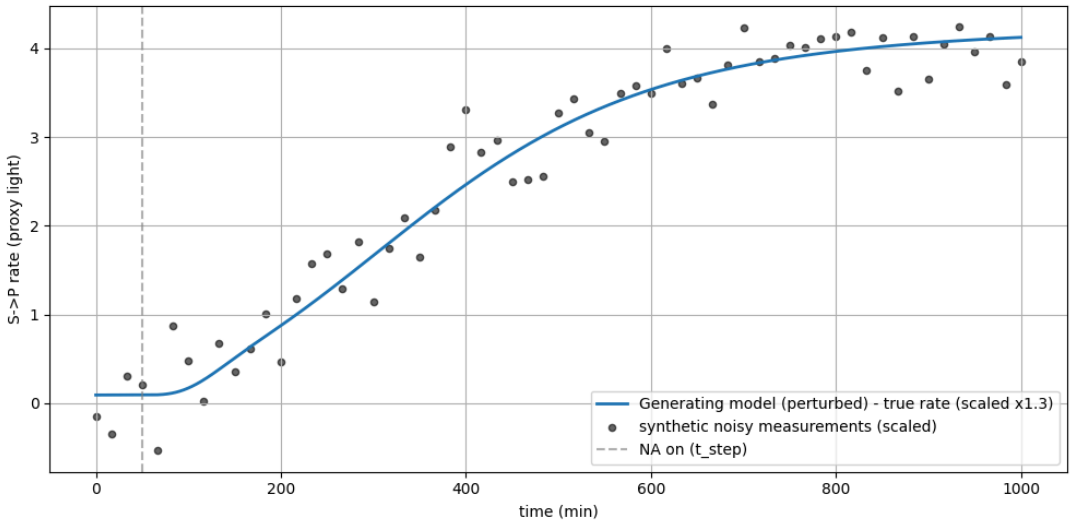

Figure 6. Synthetic time course and data. NA is added at t = 50 min (grey dashed line). The blue line is simulated with a perturbed parameter set, serving as a ground truth model. The grey dots are the noisy measurements.

Inverse Problem

Optimization

We estimated NA and the photon detection scale factor by adjusting them until the simulated light curve matched the measured data with the smallest error. This was done with a two-stage optimizer. First, a global Differential Evolution search explored the bounded NA-scale space, identifying regions of local minima. Then, a local non-linear least-squares step refined the best candidate. NA and the scale factor were restricted to lie within preset bounds and to remain non-negative. For a single light signal input with scaling, the optimization method recovered a unique best-fit NA concentration and scale factor (Fig. 7 and 8). In this example, the true NA was 0.002 µM and the fitted NA was also 0.002 µM, while the true and fitted scale factors were 1.30 and 1.22 respectively. The error curve has a clear single minimum near the true NA value. The corresponding best-fit trajectory closely overlaps the true signal.

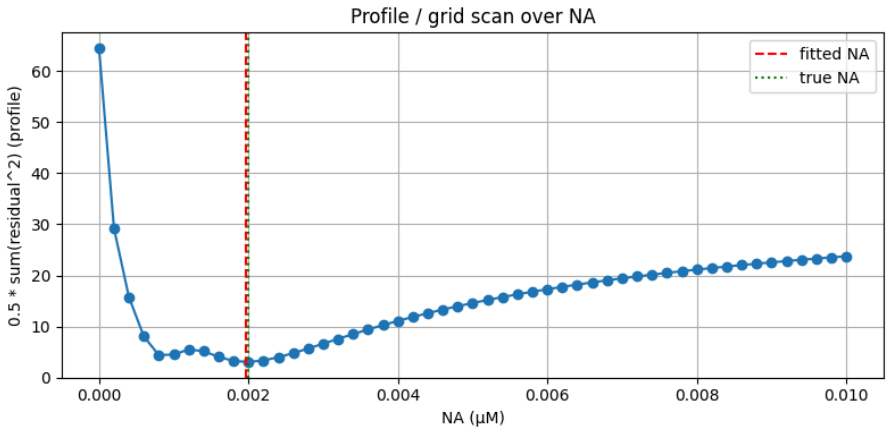

Figure 7. Error profile over NA when fitting both NA and the scale factor.

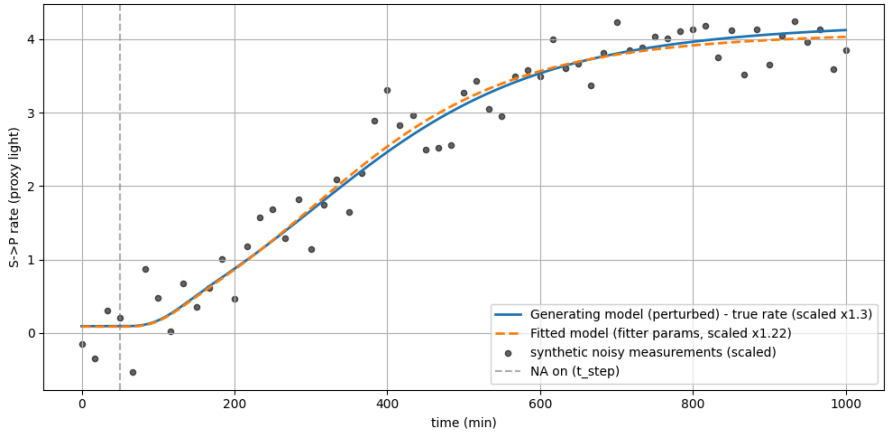

Figure 8. Synthetic data and the best-fit model. NA is added at t = 50 min (grey dashed line). The blue line is simulated with a perturbed parameter set, serving as a ground truth model. The grey dots are the noisy measurements. The orange line is the best-fit trajectory.

Uncertainty Quantification

We used a residual bootstrap to assess the reliability of the estimates. It was conducted by reshuffling residuals to generate new noisy datasets and refit NA for each one, resulting in a distribution of fitted NA and scale values (Fig. 9). The bootstrap distribution for NA had a median of 1.96×10^{-3} μM, with a 95% interval of 1.80×10^{-3} to 2.12×10^{-3} μM. The bootstrap distribution for the scale factor had a median of 1.21, with a 95% interval of 1.15 to 1.29.

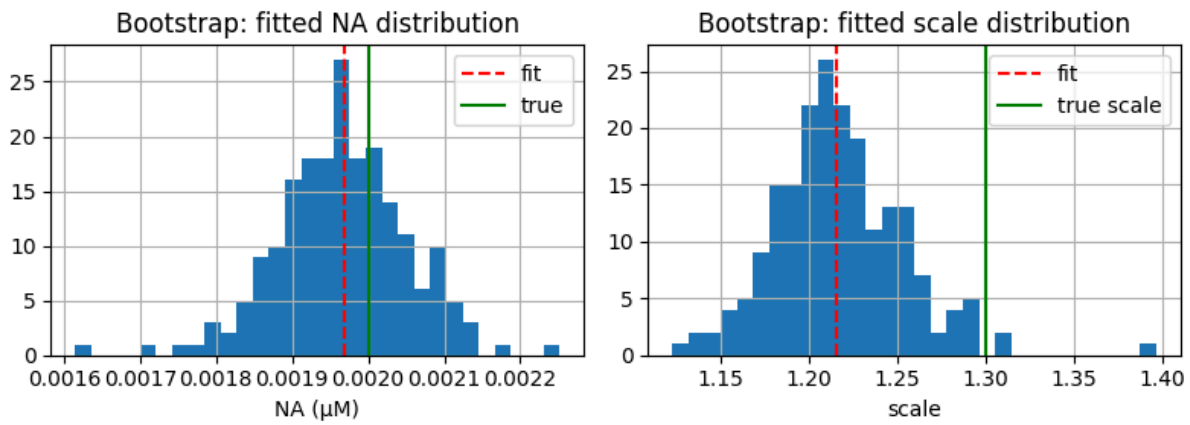

Figure 9. Distribution of fitted NA values (left) and scale values (right).

Analyte estimation range

After testing the prediction on a single NA value, we fitted NA for a series of simulated concentrations to assess the usable analyte range. Fig. 10 shows true NA versus fitted NA, both on a log10 scale. Over a broad mid-range of NA, the points lie close to the identity line. After taking logs, the fitted NA matches the true NA over several orders of magnitude with very little bias in this region. At very low NA, the points flatten and cluster near a lower limit, suggesting a practical limit of detection. The biosensor signal may be too weak compared with noise. At high NA, the estimates remain centered around the identity line but show increased scatter. Prediction at high NA may be less precise.

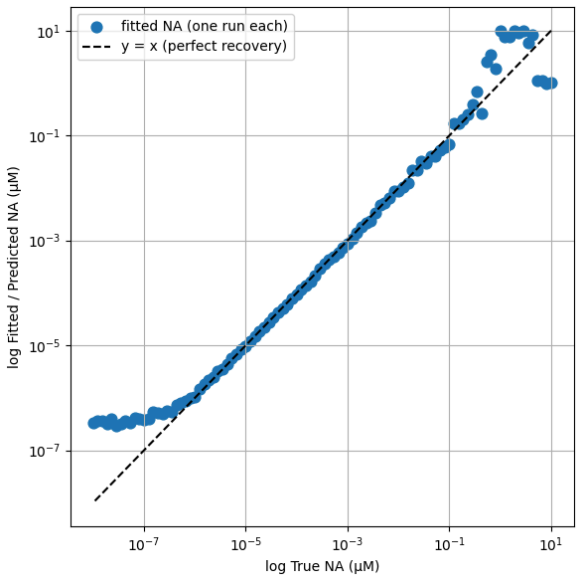

Figure 10. log10(true NA) versus log10(fitted NA). The dashed line corresponds to perfect recovery.

DISCUSSION

In this work, we developed and evaluated an inverse-modeling framework to quantitatively infer nalidixic acid (NA) concentrations from the time-resolved luminescence of a whole-cell *recA::lux* biosensor. Our results demonstrate that, despite the nonlinear, delayed, and saturating nature of the underlying biological pathway, a hybrid global-local optimization pipeline can recover NA robustly across a substantial operational range of 10^{-6}-10^{-1} µM.

A central finding of this project is that the mid-range NA values are both structurally and practically identifiable, with fitted NA closely matching the true concentrations across repeated synthetic experiments. In this regime, NA distinctly influences the rise time, curvature, and amplitude of the light trajectory, leading to a well-shaped cost landscape in parameter space. The hybrid solver, Differential Evolution (DE) for global exploration followed by nonlinear least-squares refinement, consistently converges to the correct solution (example in Figure 11).

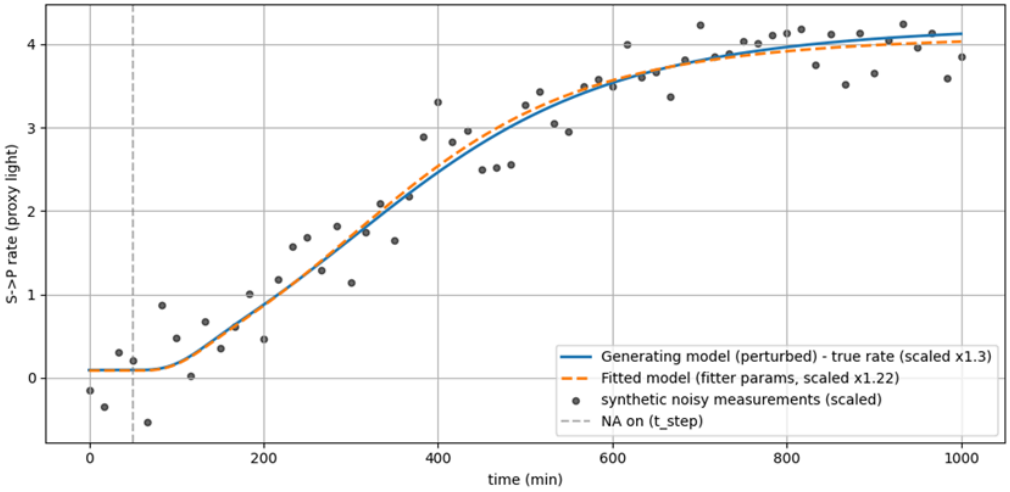

Figure 11. 0.002 µM NA synthetic data with 8% noise fitted with our solver.

An important aspect of the inverse problem concerns the multiplicative scale parameter, which represents instrument gain or unit conversion between reaction rate and measured photons. Including this parameter significantly improves the agreement between model predictions and observed data. However, we found that the fitted scale often deviated slightly from the true generating value, even though NA was accurately recovered. This behavior arises because the scale partially compensates for discrepancies between the generating model and the inference model, specifically, the deliberate kinetic parameter perturbations introduced to avoid inverse crime and the presence of measurement noise. Furthermore, scale and NA are partially confounded; increases in NA can produce changes in luminescence amplitude that resemble

multiplicative rescaling. As a result, the optimizer may accept a slightly biased scale so long as NA absorbs the dynamic features of the signal as seen in Figure 12.

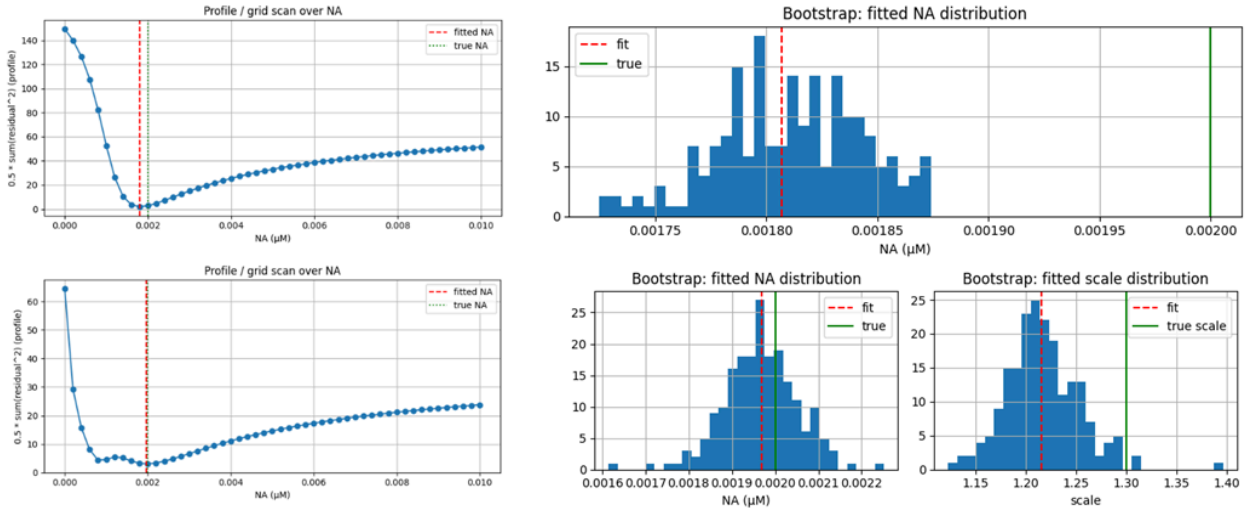

Figure 12. Profile and bootstrap histogram of 0.002 µM NA. Without scaling fitting (Top), with scaling fitting (Bottom), showing the improved fit when including scale fitting.

While the model performs well in the central concentration window, the limits of the NA domain reveal important constraints on identifiability. At very high NA concentrations, the SOS response saturates; DNA damage approaches a maximum, RecA activation plateaus, and downstream luminescence dynamics become insensitive to further increases in NA. The result is a flat profile likelihood, a broad bootstrap confidence interval, and fitted NA values that are unpredictable. Conversely, at very low NA concentrations, the emitted signal approaches the noise floor, making the measured trajectories indistinguishable across NA values (Figure 13). These behaviors reflect intrinsic biophysical information limits rather than shortcomings of the optimization routine.

The model itself embodies several simplifying assumptions that affect interpretability and generalizability. It treats the cell population as well-mixed and non-growing, ignores stochastic gene expression, assumes constant cofactors such as ATP and $FMNH_2$, and models photon detection through additive Gaussian noise rather than Poisson statistics. These abstractions enable tractable inference but may underrepresent noise levels and parameter variability observed in real biological systems. The inclusion of parameter perturbations in synthetic data generation was a strength, as it avoided inverse crime, but it also means scale and NA are estimated under mild model mismatch.

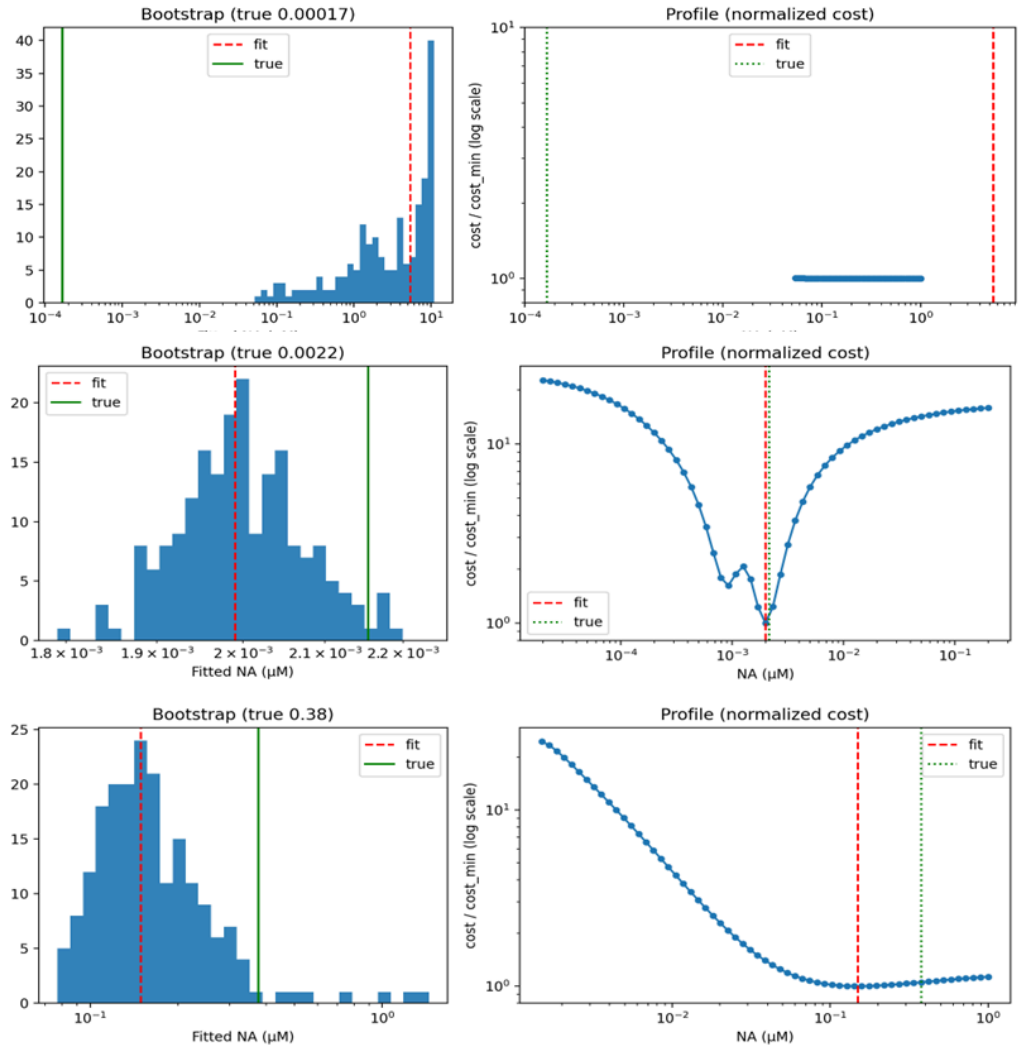

Figure 13. Limits of the model, showing the bootstrap histogram (Left), and the cost profile (Right). Comparing the extremes (Top= Lower Limit, Bottom = Upper Limit) to a good result (Middle).

Despite these limitations, the methodology for solving the inverse problem is justified. The forward model is highly nonlinear and cannot be analytically inverted; therefore, the inverse must be cast as an optimization problem over a cost function comparing predicted and observed light trajectories. DE provides a global search resilient to local minima, while least-squares polishing exploits curvature near the optimum to achieve accuracy and computational efficiency. Bootstrapping provides empirical uncertainty estimates that quantify practical identifiability, and profile likelihoods allow structural identifiability to be visualized directly.

Future work can strengthen the model in several ways. Sensitivity analysis, not fully explored here, would reveal which kinetic parameters most influence NA identifiability and help prioritize which biological measurements to obtain experimentally. Incorporating Poisson noise

models or stochastic gene-expression models would better reflect the statistics of photon detection and biological variability. Extending the inference pipeline to simultaneously estimate biological parameters and NA could allow real-time calibration in changing experimental conditions. Finally, combining luminescence with orthogonal readouts (fluorescence reporters or optical density measurements) could reduce NA-scale confounding and expand the quantifiable range at both low and high concentrations.

Overall, this project shows that a mechanistic RecA-based biosensor can serve as a quantitative chemical detector when paired with a carefully designed inverse-modeling framework. While the sensing range is ultimately constrained by biophysical saturation and noise limitations, the hybrid optimization paired with an uncertainty-quantification approach provides a principled way to define that range and to evaluate the confidence of inferred concentrations. This integrated methodology enables the transition from qualitative reporter behavior to robust, quantitative biosensing.

APPENDIX

Full ODE Model Equations

The mechanistic model captures NA-induced DNA damage, RecA activation, transcriptional induction, translation of luciferase components, and luminescent enzyme-substrate kinetics. The state vector is:

$$y(t) = \left[D(t), R^*(t), R_{free}(t), m(t), EL(t), S(t)\right]$$

Where:

- D: DNA damage level
- R^* : active RecA
- R_{free} : inactive RecA
- m: lux promoter mRNA
- EL, EP: luciferase enzyme and substrate-processing protein
- S: luminescent substrate (with product $P = S_{tot} - S$)

Table 1: Model ODEs

$[NA] + [DNA] \xrightarrow{k_{dmg}} [ssDNA]$	DNA damage and creation of ssDNA (NA is not consumed)
$[ssDNA] \xrightarrow{k_{rep}} [DNA]$	Repair of DNA
$dD = k_{dmg} * \frac{NA}{K_{NA} + NA} - k_{rep} * D$	DNA -> ssDNA rate (D = ssDNA)
$[\mathbf{ssDNA}] + \mathbf{RecA} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathbf{RecA^*}$	Activation of RecA
$dRecA^* = k_{on} * RecA_{free} * D - k_{off} * RecA^*$	RecA activation (D = ssDNA)
Model as Hill equation (h=1) See equation below table; too long to fit.	Free RecA kinetics due to promoter activation from RecA*, off rate of ssDNA and RecA*, degradation of RecA, including basal activation.
n=1 $\frac{d[mRNA]}{dt} = k_{0+k_{tx}\frac{[RecA]^n}{K_d^n+[RecA]^n}-\delta_{mRNA}[mRNA]}$	Transcription kinetics (activation of RecA* is treated as promoter activation)
$\frac{dE_L}{dt} = k_{TL}[mRNA] - \delta_L[E_L]$	Translation kinetics of luciferase enzyme

$\frac{dE_P}{dt} = k_{TP}[mRNA] - \delta_P[E_P]$	Translation kinetics of reductase enzyme complex
$\mathrm{S} \xrightarrow{E_L} \mathrm{P} + \mathrm{Light}$	Enzyme reaction of Product Formation
$\mathrm{P} \xrightarrow{E_P} \mathrm{S}$	Enzyme reaction of Substrate Formation
$dS = -vL + vP$	Rate of change in substrate
$vP = \frac{k_{catP} * EP * P}{K_{mP} + P}$	Rate of production of substrate (P->S)
$vL = \frac{k_{catL} * EL * S}{K_{mL} + S}$	Reaction Rate of S -> P (proportional to the rate of light production)
$P = S_{tot} - S$	Total S and P = S tot - S

$$dRecA_{free} = k_{RecA0} + k_{RecAind} * \frac{RecA^{*h}}{K_{RecA}^{h} + RecA^{*h}} - delta_{RecA} * RecA_{free} - (k_{on} * RecA_{free} * D) + (k_{off} * Rstar)$$

Derivation of Steady-State ($NA = 0$)

Prior to NA exposure, the model is initialized at NA = 0 steady state to avoid artificial transients.

- No damage or RecA activation:

$$D_0 = 0, R_0^* = 0$$

- Inactive RecA balances basal production and degradation:

$$R_{\text{free},0} = \frac{k_{\text{recA0}}}{\delta_{\text{recA}}}$$

- Basal transcription and translation give:

$$m_0 = \frac{k_0}{\delta_m}, EL_0 = \frac{k_{\text{tlL}} m_0}{\delta_L}, EP_0 = \frac{k_{\text{tlP}} m_0}{\delta_P}$$

- Substrate steady state is defined implicitly by:

$$v_L(S_0) = v_P(S_{\text{tot}} - S_0)$$

which we solve numerically.

Table 2: Basal steady-state values.

Variable	Expression	Value (approximate)
D_0	-	0
R_0^*	-	0
$R_{free,0}$	k_{recA0}/δ_{recA}	0.1
m_0	k_0/δ_m	0.01
EL_0	$k_{tIL} m_0/\delta_L$	0.5
EP_0	$k_{tIP} m_0/\delta_P$	0.1
S_0	$Root\ of\ v_L = v_P$	~100

Noise Model

Synthetic measurements incorporate both *model mismatch* and *measurement noise*:

1. Model mismatch:
   Selected parameters in the generating model are perturbed by 8% Gaussian noise to avoid an inverse crime.

2. Measurement noise:

$$y_{\text{meas}}(t_i) = y_{\text{clean}}(t_i) \cdot \text{TRUE\_SCALE} + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

where

$$\sigma = \text{noise\_level} \times (\max\ y_{\text{clean}} - \min\ y_{\text{clean}})$$

This yields approximately 8% variation relative to dynamic range.

Solver Settings

Differential Evolution (DE) was used for global exploration of parameter space. It is a population-based stochastic method that is resilient to local minima, so it is suitable for highly nonlinear ODE-driven cost landscapes.

Table 3: Differential Evolution solver settings.

Setting	Value	Notes
Maxiter	20	More -> better global search but slower
Polish	true	After DE, refine with LS
Seed	fixed	Makes it replicable
Bounds	[10E-12, 11]	Wide enough to contain expected values

Least Squares was used for refinement. It performs gradient-based minimization and uses bounds matching the DEs to find a local optimum with precision.

Table 4: Settings for Least Squares solver.

Setting	Value
Xtol	10^{-8}
Ftol	10^{-8}

bounds	Same as DE

Bootstrap Algorithm

Bootstrapping assesses uncertainty by sampling residuals, generating new synthetic datasets, and refitting NA. For each bootstrap replicate b:

1. Compute residuals

$$r_i = y_{\text{fit}}(t_i) - y_{\text{meas}}(t_i)$$

2. Resample residuals with replacement: $r_i^{*(b)}$
3. Generate new data:

$$y_i^{*(b)} = y_{\text{fit}}(t_i) + r_i^{*(b)}$$

4. Refit NA (and scale).
5. Store estimated NA for confidence interval estimation.

This yields empirical 95% CIs that reflect both measurement noise and model mismatch.

Additional Figures are included in the additional figures folder to not increase the size of this report. Parameters and their definitions are also included in an excel file.

AUTHOR CONTRIBUTIONS

Table 5. Contributions per Member

Team Member	Contribution
Bora Keresteci	Model: Updated forward model code, Built Inverse Model, figures, model validation/tests. Written Report: Appendix, Discussion Presentation: Intro (Midterm), Results, discussion & conclusion (Final)
Phong Le	Model: Assumptions & Parameters, Diagrams & Visualization, Discussion Written Report: Abstract, Introduction Presentation Parts: Assumptions & Params (Midterm), Inverse Model & Methods (Final)
Minxue Lin	Model: Build and update the forward model, checked the code Written Report: Result Presentation Parts: Forward model (Midterm), intro (Final)

BIBLIOGRAPHY

Bazhenov, S., Novoyatlova, U., Scheglova, E., Prazdnova, E., Mazanko, M., Kessenikh, A., Kononchuk, O., Gnuchikh, E., Liu, Y., Al Ebrahim, R., Zavilgelsky, G., Chistyakov, V., & Manukhov, I. (2023). Bacterial lux-biosensors: Constructing, applications, and prospects. *Biosensors and Bioelectronics: X*, *13*, 100323. https://doi.org/10.1016/j.biosx.2023.100323

Daniel, R., Almog, R., Ron, A., Belkin, S., & Diamand, Y. S. (2008). Modeling and measurement of a whole-cell bioluminescent biosensor based on a single photon avalanche diode. *Biosensors and Bioelectronics*, *24*(4), 882-887. https://doi.org/10.1016/j.bios.2008.07.026

Ma, H. R., Xu, H. Z., Kim, K., Anderson, D. J., & You, L. (2024). Private benefit of β-lactamase dictates selection dynamics of combination antibiotic treatment. *Nature Communications* , *15*(1). https://doi.org/10.1038/s41467-024-52711-w