Semi-quantitative Biosensor for Naked-eye Detection of Arsenic Contamination

Duy Pham Minh Nguyen, Chae Hyon Lee, Yewei Sun, Vikesh Shah-Nesarajah, Xizi Wang, Agust Siregar, Jialu Li, Chethan Jaya Sai Nandamuri, Shaocong Fang, Paul Goodman, Michalis Michailidis

Department of Biochemical Engineering, University College London, London, United Kingdom

Abstract

Millions of people worldwide are exposed to arsenic contaminated environment, which poses serious health risks. In this proposal, we introduced a novel approach for arsenic detection through a genetically engineered *Escherichia coli* biosensor. This synthetic gene network utilises the expression of two coloured compounds, violacein and eforRed, to enable detection and semi-quantitative measurement of arsenic contamination levels. Using mathematical modelling, we validated the theoretical feasibility of this biosensor in distinguishing low from high arsenic levels. We also outlined a detailed plan for the development and optimisation of this biosensor, aiming at providing a cost-effective, accessible, and reliable alternative to current arsenic testing methods. By emphasising its potential for significant impact in environmental monitoring, particularly in resource-limited settings, our system represents a promising advancement in biotechnological applications for environmental safety.

1. Introduction

Arsenic is a metalloid element typically found in the Earth's crust and groundwater. It is released into the environment from natural sources such as volcanoes and hot springs, as well as through artificial means such as being a byproduct of ore smelting (1, 2). Historically, arsenic was used in pesticides, traditional medicines, and alcohol distillation (3). Today, its compound gallium arsenide is an important semiconductor used in electronic circuits manufacture (4).

In biology, arsenic typically appears in their two oxidation states of arsenic are As(III) and As(V) (5). At neutral pH, As(III) forms arsenious acid $As(OH)_3$ and is taken up by aquaglyceroporin channels (6). As(V) forms arsenate oxyanion and, owing to its resemblance to phosphate, is taken up by inorganic phosphate transporters (PiT).

Despite the wide range of applications throughout human history, arsenic is perhaps most known for its potent toxicity. In fact, while arsenic uptake is generally unintended, arsenic detoxification systems are found in nearly every organism. (7, 8). In human, arsenic poisoning is associated with several life-threatening diseases, such as nervous system disorders, cardiovascular diseases, diabetes, and cancer (9). Unfortunately, a report in 2007 found that hundreds of millions of people worldwide were consuming arsenic-contaminated water above the safety level proposed by the World Health Organization (10, 11). While tests for arsenic are available, they often require either an established laboratory facility, or expensive test kits that use toxic chemicals such as hydrochloric acid or mercury bromide. Such requirements mean detection of arsenic contamination is likely inaccessible and unfavourable at where it is needed the most, such as in rural areas or in household settings.

In this report, we proposed a novel synthetic gene network that transforms *Escherichia coli* into a biosensor of arsenic. In response to different arsenic levels, our network tunes the expression balance between two coloured compounds violacein and eforRed, allowing for semi-quantitative detection of arsenic presence. Furthermore, we used mathematical modelling to show the conceptual feasibility of our system and determine potential handles for optimisation. Lastly, we argued that this arsenic biosensor tackles the affordability and safety issues of the current arsenic testing method, therefore possessing high social value and commercial applicability.

2. Mechanism of action

In the absence of arsenic, the protein ArsR is expressed by the basal activity of the promoter P_{ars}. ArsR binds to the arsenic binding site (ABS) upstream of its own promoter, inhibiting further P_{ars}-dependent transcription in negative feedback loop. At low arsenic concentration, an arsenic influx causes ArsR to dissociate from ABS and activates the transcription of *arsR*, *arsB*, and the gene cluster *vioABCDE*. *arsB* encodes arsenic exporter protein ArsB, while *vioABCDE* encodes the enzymes responsible for the biosynthesis of a natural blue pigment, violacein (**Figure 1**).

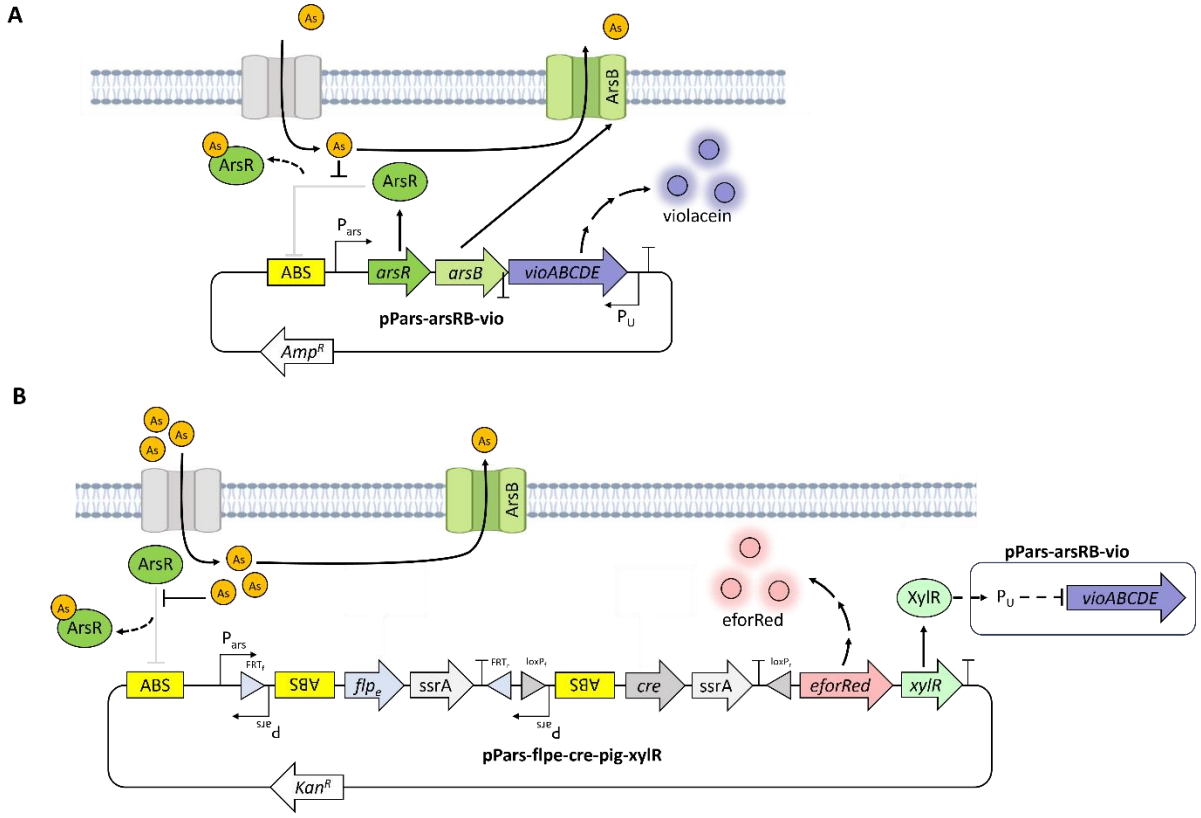

Figure 1. Mechanism of action of the arsenic biosensor (A) Schematic of the low arsenic reporter plasmid. Influx of arsenic removes ArsR suppression of P_{ars}, which drives expression of the arsenic exporter protein ArsB and the enzymes for biosynthesis of violacein. (B) Schematic of the high arsenic reporter plasmid. This plasmid functions on the premise that, under high arsenic level, arsenic influx exceeds efflux and therefore accumulate in the cell over time. The initial presence of intracellular arsenic drives Flpe expression by the first P_{ars}. Flpe inverts the sequence flanked by FRT sites and orient the second P_{ars} upstream of *cre*, which is activated by the persisting intracellular arsenic. Through similar mechanism, Cre in turn permits expression of eforRed and XylR by the third P_{ars}. XylR in turn activates P_u on the low arsenic reporter plasmid and supresses further violacein biosynthesis.

This system also employs a second plasmid for reporting high arsenic concentration (**Figure 2).** First, the presence of arsenic activates the first P_{ars} promoter, driving the expression of the recombinase Flpe, which inverse the double strand DNA in between its recognition sites FRF_f and FRT_r, thereby bringing the second P_{ars} promoter onto the other DNA strand. Because of the high arsenic concentration, this second P_{ars} promoter is also activated and drives the expression of the second recombinase Cre. Through similar mechanism to Flpe, Cre also brings the third P_{ars} promoter onto the other DNA strand. Again, the presence of a high amounts of arsenic activates this third P_{ars} promoter and drives the expression of the chromoprotein eforRed and the protein XylR.

Upon expression, XylR binds to and activates the promoter Pu, which drives expression of the anti-mRNA of the gene cluster vioABCDE (**Figure 1**). These anti-mRNAs can suppress further

violacein production, thereby enhancing the vibrancy of the red colour from eforRed at high arsenic level.

3. Materials and methods

3.1. Plasmid construction

3.1.1. Low arsenic reporter plasmid

A plasmid is designed to report low arsenic concentration in solution by expression of the blue pigment violacein (**Figure 2**). Using pUC57-Amp as the backbone, the arsenic-responsive promoter P_{ars} is placed upstream of *arsR*, a gene encoding the transcriptional repressor ArsR. ABS is positioned upstream of P_{ars}, allowing ArsR to bind to and supress P_{ars}-depedent transcription. *arsB,* encoding an arsenic efflux membrane protein ArsB, and *vioABCDE,* a gene cluster required for biosynthesis of the pigment violacein, are placed downstream of *arsR.* The promoter P_u is positioned downstream of *vioABCDE* in the reverse direction. Individual genes are cloned into the backbone plasmid via Gibson Assembly. See **Table S1** for description of the parts used and their sources.

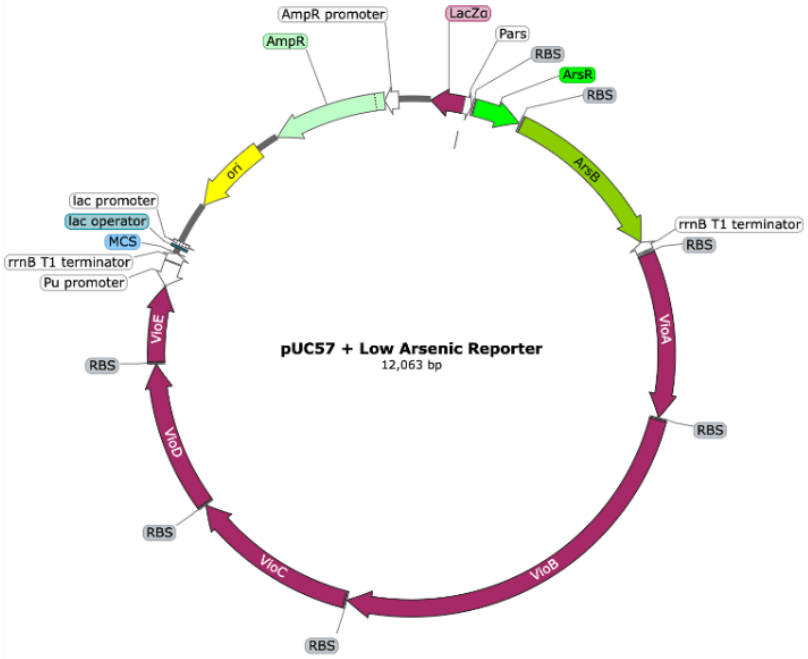

Figure 2. Design of the low arsenic reporter plasmid.

3.1.2. High arsenic reporter plasmid

A plasmid is designed to report high arsenic concentration in solution by expression of the red chromoprotein eforRed (**Figure 3**). Using pUC57-Kan as the backbone, a P_{ars} promoter is placed upstream of exchangeable elements formed by the FLP-FRT system. The exchangeable elements are formed by a reverse-directed P_{ars} promoter and *flp*, a gene encoding the Flpe recombinase, enclosed between forward and reverse flippase recombination target (FRT) recognition sites. Likewise, another set of exchangeable elements are formed by a reverse-directed P_{ars} and *Cre*, a gene encoding the Cre recombinase, enclosed between forward and reverse loxP recognition sites. Further downstream, genes encoding eforRed chromoprotein and XylR are inserted. Individual genes are cloned into the backbone plasmid via Gibson Assembly. See **Table S2** for description of the parts used and their sources.

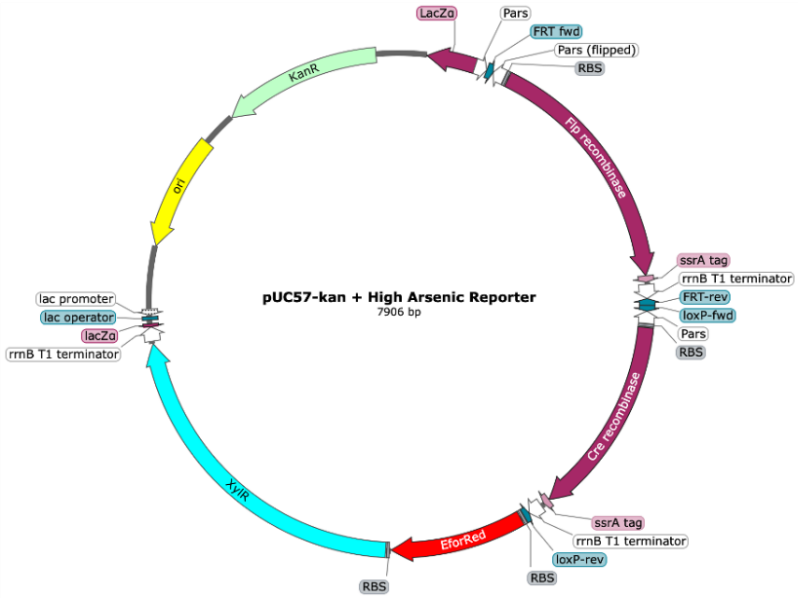

Figure 3. Design of the high arsenic reporter plasmid

3.2. Plasmid transformation

The recombinant plasmids are transformed into *E. coli* DH5α chemically competent cells using heat shock transformation. Post-transformation, 100 µL of the transformed DH5α cells are incubated in 1 mL of ampicillin-free LB liquid medium at 37°C with shaking at 225 rpm for 1 hour. This recovery phase allows for the expression of the ampicillin resistance gene present in the introduced plasmids. The cells are then plated on 100 mm LB agar plates containing 100 µg/mL ampicillin and incubated at 37°C overnight to select for successful transformants. Colony formation is assessed the following day.

Selected colonies of transformed DH5α cells are cultured in 5 mL LB medium with 100 µg/mL ampicillin at 37°C with shaking at 225 rpm for 6-8 hours, allowing bacterial growth and amplification of the transformed plasmids. To confirm correct insertion of promoters and genes the plasmids are then extracted and subjected to Sanger sequencing. Additionally, BLAST algorithm is used to verify that the insertions are correctly assembled and free from additional mutations.

3.3. Biosensor validation

3.3.1. Calibration curve

In this experiment, a series of arsenic solutions with known concentrations is prepared and incubated with the biosensor for a specific duration. The resulting colour changes and their intensities are measured and correlated with the corresponding arsenic concentrations. The calibration curve aims to establish a quantitative relationship between the biosensor's response and the actual arsenic concentration. This curve serves as a reference for accurately determining the concentration of arsenic in unknown samples based on the observed colour change.

3.3.2. Sensitivity and detection limit

The sensitivity of the biosensor is assessed by determining the lowest concentration of arsenic that elicits a noticeable colour change. This experiment aims to establish the lower limit of detection for the biosensor, providing crucial information about its ability to reliably detect low concentrations of arsenic. Gradually decreasing concentrations will be tested until the biosensor no longer shows a colour change visible to naked eyes, establishing a baseline for its sensitivity.

3.3.3. Specificity

To ensure specificity for arsenic, the biosensor is exposed to other common water contaminants. This experiment aims to verify that the biosensor only responses to arsenic and is not influenced by the presence of other pollutants. Specificity is crucial for the accuracy of the biosensor in real-world cases where water samples may contain various contaminants.

3.3.4. Reproducibility

Replicate experiments with the same concentration of arsenic are conducted to assess the reproducibility of the biosensor outputs. This experiment provides essential insights into its reliability and repeatability. Statistical analyses are employed to quantify the variation between replicate experiments.

3.3.5. Time course experiment

The time course experiment is designed to investigate the kinetics of the colour change over a range of arsenic exposure times. The objective is to determine how the biosensor's response to arsenic concentration evolves over time and to identify the optimal exposure time, which allow for reliable and timely detection of arsenic in a given sample.

3.3.6. Environmental sample testing

Actual water samples from polluted areas with varying arsenic concentrations are tested to assess how the biosensor performs under real environmental conditions. This experiment bridges the gap between controlled laboratory conditions and the complexities of real-world samples. It provides a practical evaluation of the biosensor's utility and reliability in diverse environmental samples.

3.3.7. Cross-validation with standard methods

The biosensor outputs are compared with those obtained from established arsenic detection methods, such as atomic absorption spectroscopy or inductively coupled plasma mass spectrometry (ICP-MS). This cross-validation experiment aims to assess the accuracy and reliability of the biosensor by benchmarking its performance against widely accepted analytical techniques, thereby validating its capabilities.

3.3.8. Temperature and pH sensitivity

The biosensor's responses to arsenic are tested at different temperatures and pH levels. This experiment aims to understand the biosensor's performance under varying environmental conditions. Temperature and pH sensitivity testing ensures that the biosensor remains effective and reliable across various conditions encountered in real-world water samples.

3.3.9. Long-term storage testing

The biosensor's performance is evaluated after prolonged storage to assess its stability over an extended period. This experiment is crucial for determining the shelf life and reliability of this biosensor over time. Insights gained from long-term storage testing will inform recommendations for storage conditions and provide an understanding of the biosensor's practical lifespan.

3.4. Mathematical model

We developed a mathematical model to simulate how the presence of arsenic drives pigment production in our synthetic gene network (**Figure 4A**). Components of the low-arsenic plasmid are described by ordinary differential equations (ODEs) in **Box 1** and **Box 2**, respectively. Simulations were performed using MATLAB.

We considered extracellular arsenic (As_E) to be passively imported into the cells and intracellular arsenic (As_I) to be exported in an ArsB-dependent manner. Intracellular arsenic activates the promoters P^n_{ars}, with $n \in \{1,2,3,4\}$ being the promoter index. P^1_{ars} drives the expression of ArsB and violacein. P^2_{ars} drives the expression of Flpe recombinase. Flpe in turn

reduces G_{Flpe}, representing the orientation of the sequence flanked by two FRT sites, from 1 down to 0. G_{Flpe} equals 1 reflects the *flp*$_e$ gene being directly downstream of. G_{Flpe} equals 0 reflects the sequence being inverted, at which point Flpe expression is abrogated and P^3_{ars} is positioned upstream of the *cre* gene. Through a similar mechanism, Cre expression positions P^4_{ars} upstream of the genes coding for eforRed and XylR. XylR then inhibits violacein production, enhancing the difference between eforRed and violacein.

See **Table S3** for detailed descriptions, values and units of all parameters used.

Box 1. Ordinary differential equations for components of the low arsenic reporter plasmid.

$$\frac{d[As_E]}{dt} = -k_{influx} \cdot [As_E] + k_{efflux} \cdot [ArsB] \cdot \frac{[As_I]}{[As_I] + T_{efflux}} \quad (1)$$

$$\frac{d[As_I]}{dt} = k_{influx} \cdot [As_E] - k_{efflux} \cdot [ArsB] \cdot \frac{[As_I]}{[As_I] + T_{efflux}} \quad (2)$$

$$\frac{d[P^1_{ars}]}{dt} = k_{on} \cdot \frac{[As_I]}{[As_I] + T_{on}} - k_{off} \cdot [P^1_{ars}] \quad (3)$$

$$\frac{d[ArsB]}{dt} = \alpha_{ArsB} \cdot [P^1_{ars}] - \beta_{ArsB} \cdot [ArsB] \quad (4)$$

$$\frac{d[violacein]}{dt} = \alpha_{violacein} \cdot \frac{[P^1_{ars}]}{1 + \gamma_{XylR} \cdot [XylR]} - \beta_{violacein} \cdot [violacein] \quad (5)$$

Box 2 Ordinary differential equations for components of the high arsenic reporter plasmid.

$$\frac{d[P^2_{ars}]}{dt} = k_{on} \cdot \frac{[As_I]}{[As_I] + T_P} - k_{off} \cdot [P^2_{ars}] \quad (6)$$

$$\frac{d[Flpe]}{dt} = \alpha_{Flpe} \cdot [P^2_{ars}] \cdot [G_F] - \beta_{Flpe} \cdot [Flpe] \quad (7)$$

$$\frac{d[G_{Flpe}]}{dt} = -\frac{[Flpe]}{[Flpe] + T_{Flpe}} \quad (8)$$

$$\frac{d[P^3_{ars}]}{dt} = k_{on} \cdot \frac{[As_I]}{[As_I] + T_P} \cdot (1 - [G_F]) - k_{off} \cdot [P^3_{ars}] \quad (9)$$

$$\frac{d[Cre]}{dt} = \alpha_{Cre} \cdot [P^3_{ars}] \cdot [G_{Cre}] - \beta_{Cre} \cdot [Cre] \quad (10)$$

$$\frac{d[G_{Cre}]}{dt} = -\frac{[Cre]}{[Cre] + T_{Cre}} \quad (11)$$

$$\frac{d[P^4_{ars}]}{dt} = k_{on} \cdot \frac{[As_I]}{[As_I] + T_P} \cdot (1 - [G_{Cre}]) - k_{off} \cdot [P^4_{ars}] \quad (12)$$

$$\frac{d[eforRed]}{dt} = \alpha_{eforRed} \cdot [P^4_{ars}] - \beta_{eforRed} \cdot [eforRed] \quad (13)$$

$$\frac{d[XylR]}{dt} = \alpha_{XylR} \cdot [P^4_{ars}] - \beta_{XylR} \cdot [XylR] \quad (14)$$

3.5. Optimisation

To optimise our whole cell-based arsenic detection system, we aim to improve two critical features: the low/high arsenic threshold and the exposure time needed for reliable output. Based on the results from dose response simulation analysis (**Figure 4D**), we found that several parameters significantly affect these two features. The most notable parameters the of arsenic influx and efflux rates, as well as the ArsB production and degradation rate.

3.5.1. Low/high arsenic threshold

The low/high arsenic threshold is mostly influenced by four parameters: arsenic influx, arsenic efflux, ArsB production rate, and ArsB degradation (**Figure 4D**, left panel). As arsenic enters *E. coli* cells through phosphate transporters such as PiT, which are likely important for survival (12), here we instead prioritise optimising arsenic efflux. Since the *arsB* gene encodes the arsenic exporter protein ArsB transporter for arsenic removal from the cell, its expression level will influence the rate arsenic efflux (6). The interdependence of ArsB level with both ArsB production rate and ArsB degradation rate suggests a parallel approach. We plan to manipulate the ArsB production rate using different transcriptional regulatory elements. Furthermore, ArsB degradation rate can be tuned by C-terminal tagging of ArsB with an ssrA tag (13) or a c-myc tag (14), which increases or decreases its degradation rate, respectively. **Table 1** shows our design strategy to find the best construct to optimise the low/high arsenic threshold.

Table 1. Construct library to tune ArsB production and degradation rates for optimisation of the low/high arsenic threshold.

	Transcription enhancer	No transcriptional regulator	Transcription repressor
With degradation tag	Construct 1	Construct 2	Construct 3
Without tag	Construct 4	Construct 5	Construct 6
With stabilisation tag	Construct 7	Construct 8	Construct 9

3.5.2. Exposure time for reliable output

The simulation result shows that the exposure time for reliable output is most sensitive towards the rates of arsenic influx, arsenic efflux, and ArsB production (**Figure 4D**, right panel). To optimise this feature, we plan to manipulate arsenic efflux by tuning ArsB production rate using different transcriptional regulatory elements, as shown in **Table 2**.

Table 2. Construct library to tune ArsB production rate for optimisation of the exposure time for reliable output.

	Transcription enhancer	No transcriptional regulator	Transcription repressor
Construct	Construct 10	Construct 11	Construct 12

4. Preliminary results

4.1. Different arsenic levels generate distinct pigment production profiles *in silico*

To study the behaviour of the synthetic gene network in response to different arsenic levels, we employed an ODE-based mathematical model. The model inputs an initial extracellular arsenic level and simulates the production profile of pigments over time (**Figure 4A**, see also **Materials and methods**). We defined the ethanol extract colour as the level of violacein relative to the total pigment level, which ranges from completely red at 0 from completely blue at 1. Our simulation indeed shows that if the initial arsenic level is relatively low, then violacein be the major pigment after a certain time window and the ethanol extract would appear blue (**Figure 4B**, left panel). If the initial arsenic level is instead 10-fold higher, then eforRed would eventually dominate after the same time window and the ethanol extract would appear red (**Figure 4B**, right panel).

4.2. Output time and low-high arsenic threshold majorly depends on arsenic flux

We next investigated how the output of our system varies over a range of initial extracellular arsenic level and read-out time. Given the same read-out time, we found that the ethanol extract colour exhibits ultrasensitivity, meaning its numerical value quickly switches from very low to very high as input arsenic level passes a threshold. With further read-out time, this threshold is reduced (**Figure 4C**). Such switch-like behaviour supports the ability of our system to provide clear distinction between low from high arsenic level, by reducing the likelihood of encountering intermediate output values.

To optimise the system for practical use, it is critical to determine the factors that affect the low-high arsenic threshold, as well as the time at which a reliable output could be obtained given a threshold. We therefore examined how these features respond to perturbations of each parameter in the model. Sensitivity analysis shows that both features are most sensitive to changes in arsenic influx and efflux rates. They are also highly sensitive to changes in ArsB production and degradation rates, which constitutes ArsB expression that drives arsenic efflux (**Figure 4D**). These results suggest that the processes governing arsenic flux could be important handles to adjust the low-high arsenic threshold and to optimise read-out time.

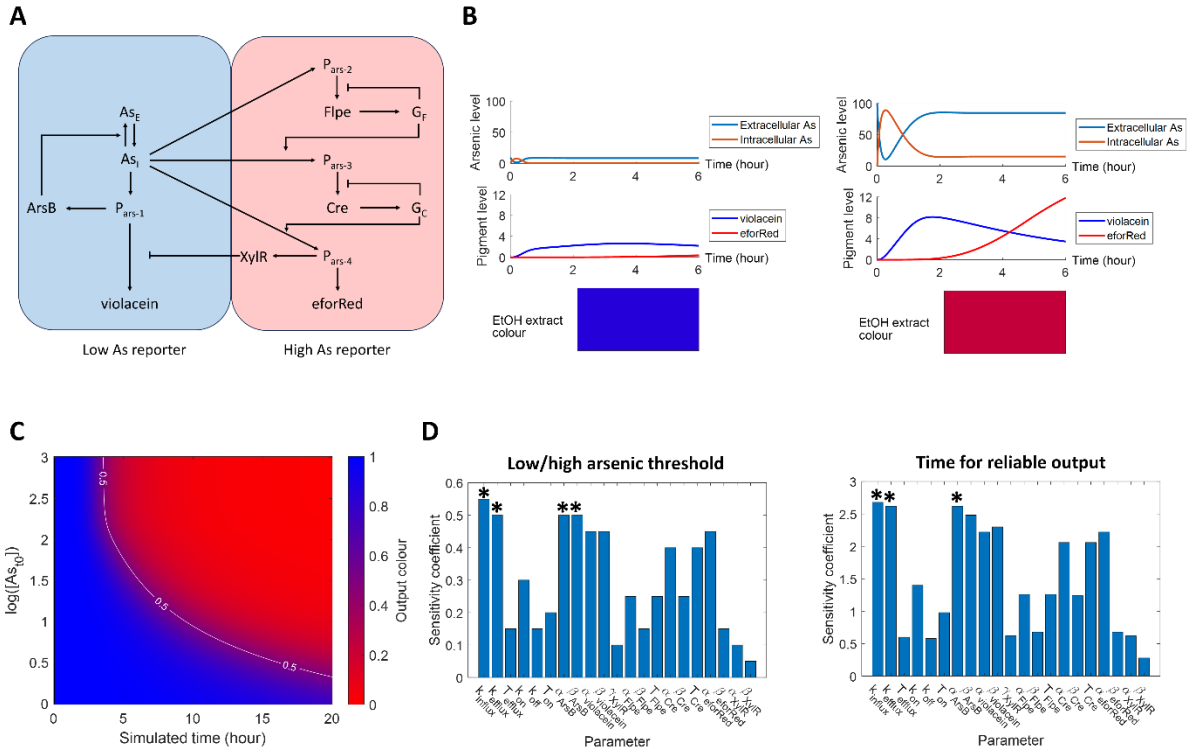

Figure 4. Validation of the arsenic biosensor system with mathematical modelling. (A) Topology of the model, representing the arsenic-dependent pigment production network. (B) Simulation of arsenic levels, pigment levels and ethanol extract colour over time in response to low (left panel) or high (right panel) initial extracellular arsenic level. (C) Output colour of the as a function of initial extracellular arsenic level and simulated time. White contour annotates the condition at which output colour equals 0.5, which indicates the low-high arsenic threshold. (D) Sensitivity of the low/high low-high arsenic threshold (left panel) and reliable read-out time (right panel) against individual parameter changes. Asterisks denote the top 3 parameters with the highest sensitivity coefficients.

5. Conclusion

In this proposal, we introduced the design for a novel *E. coli*-based biosensor for arsenic contamination. Our system leverages the expression of violacein and eforRed, providing a cost-effective, accessible, and semi-quantitative method for arsenic testing. The proposal outlined the potential for significant advancements in environmental arsenic monitoring, especially in areas lacking conventional testing resources. The focus on future optimisation and real-world application highlights the aim of our project in revolutionising arsenic detection, providing significant values in both scientific and practical domains.

6. Biosafety

Our system has very little biosafety risks. While we employ a whole cell-based arsenic detection system, we do not release any engineered *E. coli* into the environment during its use. Instead, water samples are collected and introduced into test tubes containing arsenic-sensitive *E. coli* cells. All residual tests are then destructed using alcohol sterilisation or autoclaving. *E. coli* is a risk group 1 organism which complies with the biosecurity whitelist guideline from Gogec [https://www.gogecconference.org/biosecurity-whitelist]. We use two recombinases in our system, Flp and Cre. However, these enzymes only function in the presence of their specific recognition target sequence, which is very rare in nature. Their activities are further limited by incorporation of ssrA tags to enhance degradation, reducing the likelihood of these recombinase contaminating the ecosystem.

7. References

1. Carlin DJ, Naujokas MF, Bradham KD, Cowden J, Heacock M, Henry HF, et al. Arsenic and Environmental Health: State of the Science and Future Research Opportunities. Environmental Health Perspectives. 2016;124(7):890-9.

2. Inskeep WP, McDermott TR. Geothermal Biology and Geochemistry in Yellowstone National Park: Proceeding of the Thermal Biology Institute Workshop, Yellowstone National Park, WY, October 2003: Montana State University Publications; 2005.

3. Huang JH, Hu KN, Ilgen J, Ilgen G. Occurrence and stability of inorganic and organic arsenic species in wines, rice wines and beers from Central European market. Food Additives and Contaminants Part A-Chemistry Analysis Control Exposure & Risk Assessment. 2012;29(1):85-93.

4. Moss SJ, Ledwith A. Chemistry of the Semiconductor Industry: Springer Science & Business Media; 1989.

5. Yang HC, Fu HL, Lin YF, Rosen BP. Pathways of arsenic uptake and efflux. Current Topics in Membranes. 2012;69:325-58.

6. Meng YL, Liu Z, Rosen BP. As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*. Journal of Biological Chemistry. 2004;279(18):18334-41.

7. Bhattacharjee H, Rosen BP, Mukhopadhyay R. Aquaglyceroporins and Metalloid Transport: Implications in Human Diseases. In: Beitz E, editor. Aquaporins. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 309-25.

8. Rensing C, Rosen B. Heavy metals cycle (arsenic, mercury, selenium, others). Encyclopedia of microbiology. 2009;205:19.

9. Shahid M, Dumat C, Khan Niazi N, Khalid S. Global scale arsenic pollution: increase the scientific knowledge to reduce human exposure. VertigO-la revue électronique en sciences de l'environnement. 2018(Hors-série 31).

10. Ravenscroft P, editor Predicting the global distribution of arsenic pollution in groundwater. Royal Geographical Society Annual International Conference, Department of Geography, Cambridge University; 2007.

11. World Health Organization. Exposure to arsenic: a major public health concern. 2010. WHO Document Production Service: Geneva, Switzerland. 2020.

12. Rosenberg H, Gerdes R, Chegwidden K. Two systems for the uptake of phosphate in *Escherichia coli*. Journal of bacteriology. 1977;131(2):505-11.

13. Karzai AW, Roche ED, Sauer RT. The SsrA–SmpB system for protein tagging, directed degradation and ribosome rescue. Nature structural biology. 2000;7(6):449-55.

14. Mikalsen T, Johannessen M, Moens U. Sequence- and position-dependent tagging protects extracellular-regulated kinase 3 protein from 26S proteasome-mediated degradation. The International Journal of Biochemistry & Cell Biology. 2005;37(12):2513-20.