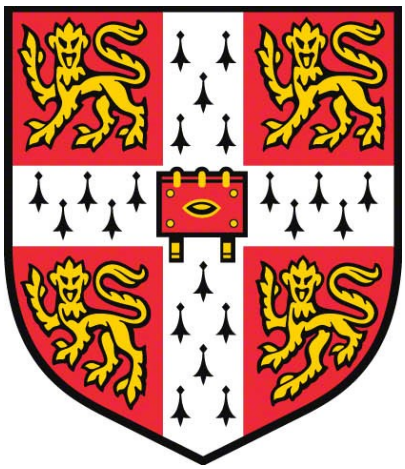
Towards smart molecular biosensors

Modulation of DNA strand displacement reactions with transcription factors for robust and tunable arsenic quantification



James Coxon

Jesus College

Department of Pathology

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

June 2021

Declaration

I hereby declare that my dissertation entitled

Towards smart molecular biosensors: Modulation of DNA strand displacement reactions with transcription factors for robust and tunable arsenic quantification

- Is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.
- Is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.
- Does not exceed the prescribed word limit for the Biology Degree Committee.

James Coxon

June 2021

Towards smart molecular biosensors: Modulation of DNA strand displacement reactions with transcription factors for robust and tunable arsenic quantification

Abstract **James Coxon**

Biosensors are increasingly important for gathering information in modern life. Advances in synthetic biology tools, such as isothermal DNA amplification, CRISPR-Cas and DNA strand displacement (DSD), have accelerated this transition and the ongoing COVID-19 pandemic has made biological testing ubiquitous. Nucleic acid diagnostics have been at the forefront of this trend, particularly for viral genomes, but increasingly other analytes are being addressed. Arsenic is a pervasive contaminant of drinking water in many countries, especially Bangladesh where the WHO has described the situation as the 'largest poisoning of a population in history'. Reliable, sensitive, affordable and practical sensors are urgently required to reduce prevalence of chronic exposure.

Here, a novel *in vitro* biosensory system is presented, using allosteric transcription factors (aTF) to modulate reversible DSD reactions. By designing fluorescent probes containing an appropriate aTF operator sequence, repressor concentration can influence the underlying DSD reaction equilibrium. Once a stable state between these components is established, addition of the allosteric ligand, in this case arsenite, perturbs the system, and the resulting equilibrium shift and intermediate dynamics can be correlated with ligand concentration, enabling quantification.

In this project, the *Corynebacterium glutamicum* arsenic repressor (CgArsR) was expressed, purified and characterised, and its suitability for DSD integration established. A mutant operator sequence was designed to provide a structureless, high-affinity DNA probe with an architecture enabling comparison of alternative sequences while maintaining a consistent fluorophore environment. This was combined with a bespoke normalisation methodology for meaningful, reliable signal conversion to concentration or reaction balance metrics, allowing optimisation of both speed and sensitivity of arsenic quantification. Finally, integration of a parallel arsenite-insensitive mutant repressor and a differential rate analysis enabled not only robust selectivity but detection below the 10ppb WHO threshold in under an hour.

The resulting system has great potential for further enhancement. Addition of chelating and cryopreservative agents are likely to improve ability to withstand environmental variation and lyophilisation, while an improved mechanistic model could permit optimisation of component concentrations and properties. Combined with suitable hardware, a practical arsenic biosensor may be possible. Perhaps more importantly, however, this project provides clear demonstration that allosteric transcription factors can modulate DSD reactions, creating potential for many other applications, whether directly as biosensors for alternative analytes or as bespoke signal control elements for DSD computation.

Acknowledgements

I would like to thank the following people for their guidance and support over the course of this project:

- Dr Jim Ajioka, my primary supervisor, for his advice and mentoring over the past four years. I am particularly thankful for his creation of the initial project, his aid in ensuring the work was properly resourced and the independence he allowed me to express in choosing project directions.
- Dr Peter Davenport (University of New Mexico), my post-doctoral advisor, who provided frequent detailed supervision despite moving abroad early in the project. His continued engagement was highly appreciated and our fortnightly discussions were vital to the progress made.
- Dr Boyan Yordanov (formerly Microsoft Research, Cambridge), my supervisor on project computational aspects, who guided the modelling as well as contributed to discussions in all other areas. His presence and suggestions at the fortnightly meetings were also highly valued.
- Dr Andrew Phillips (Microsoft Research, Cambridge) for taking over Microsoft supervision of this project in the last nine months.
- Dr Charlie Gilbert and Akash Das for their help and good company in the Ajioka lab. I am also grateful for the camaraderie provided by many other friends in the Molteno building.
- Professor Kevin Land (University of Pretoria) for the wisdom he provided in our weekly lab meetings.
- Ben Baker, my summer student in 2018, who helped with the early repressor characterisation work including EMSA images in Chapter 4.

This work was supported by Microsoft Research through its PhD Scholarship Programme. I am grateful for their funding and expertise during this project.

Contents

List of Figures

List of Tables

List of Abbreviations

AFS	Atomic fluorescence spectrometry
ArsR	Arsenic repressor
As	Arsenic
As(III)	Arsenite
As(V)	Arsenate
aTF	Allosteric transcription factor
BGS	British Geological Survey
bp	Base pair(s)
BSA	Bovine serum albumin
BsArsR	*Bacillus subtilis* arsenic repressor
CgArsR	*Corynebacterium glutamicum* arsenic repressor
CRISPR	Clustered regularly interspaced short palindromic repeats
CRN	Chemical reaction network
DBD	DNA-binding domain
DSD	DNA-strand displacement
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EMSA	Electrophoretic mobility shift assay
DTT	Dithiothreitol
FEOC	Free energy of ordered complex
FT	Flow through
GMO	Genetically modified organism
GOx	Glucose oxidase
gRNA	Guide RNA
HCR	Hybridisation chain reaction
HG	Hydride generation
hGC	human chorionic gonadotrophin
HPLC	High-performance liquid chromatography

ICP-AES	Inductively-coupled-plasma atomic-emission spectrometry
ICP-MS	Inductively-coupled-plasma mass spectrometry
IVT	*in vitro* transcription
IPTG	Isopropyl β-d-1-thiogalactopyranoside
IR	Inverted repeat
LAMP	Loop-mediated isothermal amplification
LFD	Lateral flow device
MCS	Multiple cloning site
MWCO	Molecular weight cut-off
MFE	Minimum free energy
MM	Master mix
NASBA	Nucleic acid sequence based amplification
nt	Nucleotide(s)
PAGE	Polyacrylamide gel electophoresis
PAT	Potassium antimony tartrate
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
POC	Point-of-care
ppb	Parts per billion
ppm	Parts per million
PSB	Protein storage buffer
RBS	Ribosome-binding sequence
RFU	Relative fluorescence units
RLS	Resource-limited setting
ROC	Repressor-operator complex
RPA	Recombinase-polymerase amplification
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electophoresis
ssDNA	Single-stranded DNA
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
TF	Transcription factor
TSP	Total soluble protein
TXTL	Transcription-translation
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Environmental sensing

All organisms require information about their environment in order to survive. These can be positive stimuli, such as identification of useful resources, beneficial conditions and potential mates, or negative, such as presence of predators and toxins. Gathering data on the current situation, and processing these to inform decisions that improve chances of survival or reproduction, is key to evolutionary success. Methods of detecting positive and negative stimuli are therefore present in all forms of life.

Our basal senses enable us to see, hear, smell, touch and taste, but each of these is limited in both the range and sensitivity of targets commonly found throughout human evolution. Rapid changes to technology and society mean that modern life is filled with additional sensors constantly sampling our environment. This wealth of information allows us to make better and more complex decisions than ever before, as well as respond to circumstances brought about by these changes themselves. Carbon monoxide alarms are a classic example of a technology that has allowed us to sense the presence of a deadly gas that we naturally find odourless, as our evolution has not adapted us to life in confirmed spaces.

For a long time detection of new analytes relied on the use of physical or chemical methods, however recently biology has provided the tools to sense targets in different ways. Any device based on a biological component for the purposes of detection is known as a biosensor [1]. Household examples of these

include pregnancy and glucose tests, which harness either antibodies or enzymes for detection or quantification of particular analytes. The use of biological components can improve the speed, sensitivity, affordability or practicality compared to traditional methods [2, 3].

The recent COVID-19 pandemic has highlighted the utility of widespread, accurate data collection to anticipate viral outbreaks and respond quickly, and how the features of a particular detection method can help inform life-or-death decision making when faced with deadly agents. While RT-qPCR remains the gold standard for SARS-CoV-2 viral detection due to its high accuracy, less reliable lateral flow tests have been crucial to recent efforts to release lockdown measures, due to their ease-of-use, rapid results presentation and relative low cost [4]. Even in wealthy countries such as the UK, expense and speed are crucial factors in mass testing efforts, often outweighing precision. The resulting rapid progress of biotechnological development in this area is sure to continue increasing the prevalence of biosensors in our lives.

1.2 Synthetic biology

The term synthetic biology is wide-ranging, generally defined as the 'rational and systematic engineering of complex biological systems with novel functions' [5, 6]. It can be split into two facets - the editing of living cells or organisms to alter their behaviour, and the creation of new *ex vivo* biological parts, devices or systems to perform desired functions. Both have been driven primarily by rapid technological progress removing the boundaries originally thought to constrain molecular biology. Since the advent of PCR the costs of DNA sequencing and synthesis have dramatically reduced, and combined with new methods to assemble (e.g. GoldenGate, Gibson) and edit (e.g. CRISPR) genetic parts, bioengineering has reached this truly synthetic phase [7, 8].

Many of the early success stories were based on the first facet, through the engineering of microorganisms to produce proteins or molecules of value. A common example is the synthetic production of artemisinin, an important antimalarial drug naturally produced by plants. Growing and harvesting *Artemisia annua* through conventional means is very inefficient, producing low yields of the desired terpenoids, and so used to be very costly. Through the engineering of heterologous mevalonate synthesis pathways into first *E. coli*

bacteria and later *S. cerevisiae* yeast, artemisinin production has become much more efficient and as a result much cheaper, profoundly impacting treatment of malaria worldwide [9, 10]. Similar work has achieved large scale biosynthesis of numerous antibiotics, antivirals, anticancer agents, pesticides and vaccines [11]. Indeed, synthetic biology can be credited with many of the techniques successful SARS-CoV-2 vaccines have used in their development pipelines [12].

Apart from improving production of pharmaceuticals, synthetic biology has also enabled the endowment of engineered microorganisms with new or heightened abilities. Demonstrations, such as the creation of *E. coli* capable of producing a high-resolution image in response to light stimulus, have shown how transforming cells with new genetic circuits can enhance their capabilities [13]. An important area of development is cell/gene therapies, whereby heterologous circuits enable the identification and removal of undesirable cells, a growing tool in combating cancers and other metabolic conditions as well as antibiotic-resistant microorganisms [7, 14, 15].

The ambition for what synthetic biology can achieve is only growing. Further to altering specified cell populations, work is also being attempted engineering entire microbial communities [16], while concerted efforts are also being made to expand the genetic code [17] and make new genetic polymers [18]. If these succeed, radically new organisms and ecologies will be designed capable of ever greater possibilities.

In addition to the development of directly valuable products and technologies, synthetic biology is also enabling a greater understanding of natural systems. Richard Feynman's 'What I cannot create I do not understand' is often quoted to justify how the application of engineering principles to biology forces the full comprehension of complex systems, which in turn eventually enables better designed devices [19]. Examples of this include the reconstitution of chemotaxis [20] and phagocytosis [21], through which their most important components were identified and characterised.

Synthetic biology is not without its issues, however. Concern over the ethics and safety of such work is rightly voiced, but given the technological possibilities open but well-regulated science is vastly preferred to limiting research to illicit groups [19]. Innovation can provide solutions to these problems too, with work on synthetic biocontainment circuits quelling fears of escape of genetically-modified organisms (GMOs) [22, 23].

1.2.1 Cell-free biology

The creation of *ex vivo* devices has taken longer to reach fruition but these have become increasingly common over the last ten years. These non-living systems address many of the safety concerns with GMOs and do not suffer the burden of having to maintain crucial cell processes, but are complex in their own ways. Within this area of synthetic biology two routes have distinguished themselves - 'cell-free' and '*in vitro*'. Cell-free systems harness extracts of lysed cells to power desired synthesis without the constraints of maintaining living organisms. By removing all membranes and native nucleic acids, a barrier-free homogenous cytosol can be used to produce DNA, RNA or proteins of interest simply through addition of engineered plasmids, supplemented with an energy source and requisite monomers. If productive conditions are maintained then the transcription and translation machinery of the lysed microorganisms will instead act on the recombinant DNA to synthesise encoded products [24].

Cell-free, or transcription-translation (TXTL), systems, have therefore been used primarily for protein synthesis [24]. The ability to focus all energy and resources for production of desired polypeptides makes this theoretically much more efficient than conventional methods, and enables the generation of proteins that would usually be toxic to growing cells. Without barriers, however, solutions are more sensitive to contamination, with any extraneous genetic material able to compete for expression and mRNA stability reduced [25]. Cell extracts are inherently highly nutritious too, making them effective growth media for other microorganisms, which are harder to control with antibiotics. Extracts are notoriously inconsistent too, with minute differences in culture conditions altering the metabolite and protein profile of a solution to limit reproducibility of TXTL experiments [26].

While *in vitro* is technically an umbrella term for all biology taken out of natural context, including cell-free systems, within synthetic biology it is often used to describe systems built up through the addition of individual purified components, rather than primarily consisting of relatively crude cell extracts. As a result *in vitro* systems tend to be simpler, but through the avoidance of complex transcription and translation tend to be cheaper and more reliable.

The use of *in vitro* formats for experimentation has therefore existed since the inception of molecular biology, but recently their use in stand-alone devices has

grown dramatically, largely due to the expansion of characterised parts and reporters. Their lack of translation ability or metabolism pathways mean they cannot be used for direct synthesis, but instead focus on another synthetic biology application - diagnostics.

1.3 Biosensors

As defined earlier, a biosensor is any detection device containing at least one biological component [1]. The inclusion of such elements is key to providing either specialised activity or structural specificity that physical or chemical methods would struggle to achieve, particularly at low costs. The *in vitro* format has been highly successful for such devices - combining a few purified components and arranging them to behave differently depending on the presence or absence of their analyte results in very controlled, reliable outputs.

Many *in vitro* biosensors make use of antibodies as their biological component [27]. These can be highly specific for their target and can be conjugated to reporters to allow a detectable output. Enzyme-linked immunosorbent assays (ELISAs) harness this principle, using either primary or secondary antibodies conjugated to enzymes such as horseradish peroxidase to catalyse a chemical colour change, which will only occur when antigen is present to stop the antibodies washing away.

Pregnancy tests use an ELISA derivative known as a 'sandwich' assay, requiring two different antibodies and a conjugated, coloured nanoparticle instead of an enzyme [3]. The labelled antibody is able to bind any human chorionic gonadotrophin (hCG) hormone present in the urine of a pregnant woman and as the solution wicks up the lateral flow device (LFD) another antibody, fixed to the test line, traps the bound complex by binding to another hCG epitope. This results in a solid coloured line forming in the presence of hCG, while no line forms without hCG as the labelled antibody passes unperturbed. The use of two antibodies makes false positives very rare, while the specificity of hCG for pregnancy means false negatives occur only in limited circumstances. Together, many hCG lateral flow tests claim to be over 99% accurate, outperforming ultrasound analysis and able to identify pregnancies even before a first missed period. Being cheap, discreet, able to use in the privacy of your own home and producing increasingly clear results (some with digital outputs) very quickly,

these devices have been revolutionary in improving family planning and women's health and rights worldwide [3].

Aside from antibodies, enzymes are another class of biomolecule commonly used in *in vitro* biosensors. These proteins catalyse specific changes to targets through the lowering of reaction activation energy. Blood glucose monitors, or 'glucometers', make use of the enzyme glucose oxidase (GOx), which catalyses the conversion of glucose to gluconolactone. As this oxidation reaction indirectly reduces an electrode, the charge through the electrochemical circuit is proportional to the concentration of glucose in the solution and can be measured very accurately [2]. Leland Clark, who developed this system in the 1950s, was subsequently known as the 'father of biosensors' for this pioneering work [28]. The precise, quantitative reading of glucometers is often seen as the gold-standard for biosensors, although the accuracy of these readings can be perturbed by a multitude of factors [2]. Once again, however, the ability to self-monitor blood glucose content has transformed diabetic care, vastly improving the management of the disease for millions in the UK alone.

1.3.1 Whole cell biosensors

While such application-specific *in vitro* biosensors have appeared over the years, the advent of genetic engineering and synthetic biology opened up an entirely different approach. Instead of identifying or manipulating proteins to be useful in completely unnatural contexts, existing sensory systems could be hijacked instead [29]. Non-human life has evolved to detect and utilise countless ions, compounds and molecules that humans cannot naturally sense, and so exploiting these properties has been a constant area of interest for all types of synthetic biology.

Initially, genetically engineering reporters into living microorganisms was employed. Bacteria such as *E. coli* have hundreds of operons in their genome to regulate expression of various effector proteins, preventing their expensive synthesis until they are required. The *mer* operon, for example, responds to presence of mercury in the cytosol, and upon detection increases the expression of enzymes needed to detoxify and remove the element. Simply cloning a reporter gene such as luciferase into this operon, in frame with the other effector genes, provides a visual output whenever the *E. coli* strain detects mercury in

the presence of luciferin [30].

This approach has a number of advantages. First, generating this recombinant strain is technically straightforward. Second, manufacturing costs for a potential product are low - maintaining a culture of *E. coli* only requires provision of a nutrient broth in stable conditions. Third, detection is relatively robust. Cells naturally filter their environment through their membranes and their operons have evolved to respond to biologically relevant concentrations of target solutes in spite of potentially confounding external factors, such as presence of other compounds or fluctuations in temperature. Lastly, an enormous wealth of natural operons exist that can be tapped into, making this simple approach easy to replicate.

However, several drawbacks exist too. While countless useful operons naturally exist, only a small selection are in easy-to-culture bacteria like *E. coli*. While it is possible to characterise and clone others, not only is this technically much harder but they also may not function as well as in its original strain, or even at all. Second, whole cell biosensors tend to be very slow as bacteria usually make transcriptional changes on hour-long timescales rather than seconds. For a reporter gene to be expressed in sufficient quantities to be detected, the analyte must enter the cells and interact with the appropriate TFs before transcription and then translation changes. Typically humans desire information in seconds or minutes and not longer. The above example of a mercury biosensor was impressively sensitive, but response time took one hour of incubation after the bacteria were prepared [30]. Third, natural operons have a limited dynamic range. Many microorganisms have evolved to detect presence or absence of a particular substance, rather than specific quantities. The various sensitivities built into the operon, therefore, tend to reflect thresholds biologically relevant to the host, which may differ to humans. Altering these sensitivities and expanding detection ranges is possible through careful genetic manipulations, but has limits within a living cell.

Fourth, despite the robust characteristics of bacteria, whole cell biosensors are usually unreliable. Microorganisms constantly evolve and so maintaining strain identity can be difficult, particularly if a quantitative output is desired. Environmental fluctuations may not alter long-term responses to particular concentrations of particular targets, but greatly affect the speed or strength of the response. Finally, practical issues plague whole cell biosensors. Housing live

bacteria in a device that enables them to respond to analytes is difficult while simultaneously providing nutrients and removing waste. There is a lot of concern over escape of GMOs, even if they are demonstrated to be benign, so employing measures to limit their access or chances of external survival are important, but these can interfere with device capabilities. Overall, while the potential of whole cell biosensors is great, these final hurdles are usually too high to overcome.

1.3.2 Cell-free TXTL biosensors

Designing TXTL versions of these regulatory networks solves many of these problems, but introduces others. As cell-free systems do not need to support a fully living organism they are not prone to regulatory or evolutionary issues, and focussing all resources on detection is much more efficient than having to support numerous other processes. They can be much more controlled than living cells, although extracts will inevitably display batch variability. They are, however, much more expensive to produce and more sensitive to environmental variation and contamination.

Despite these concerns cell-free systems have been successfully adapted for biosensors, particularly in the detection of nucleic acid sequences. Innovative work by Pardee *et al.* designed synthetic riboswitch RNA species that altered their structure if bound by a complementary trigger, freeing up the ribosome-binding sequence (RBS) and enabling the translation of a reporter protein [31]. Figure 1.1 shows how their design enabled them to remove constraints usually present in these species to target almost any RNA sequence, such as the mRNA of expressed genes or the genomes of RNA viruses. Once added to cell-free extracts and freeze-dried onto paper disks, these riboswitch biosensors were able to detect different regions of the Ebola virus genome as well as bacterial antibiotic resistance genes, either with a fluorescent or colorimetric output. The ability to embed these sensors and extracts into paper, lyophilise the prepared solutions and store them at room temperature made them highly convenient for use as a point-of-care (POC) diagnostic. These innovations would certainly help drive down costs of cell-free systems, making similar biosensors more affordable as well as practical, important for uptake in resource-limited settings (RLS) [32].

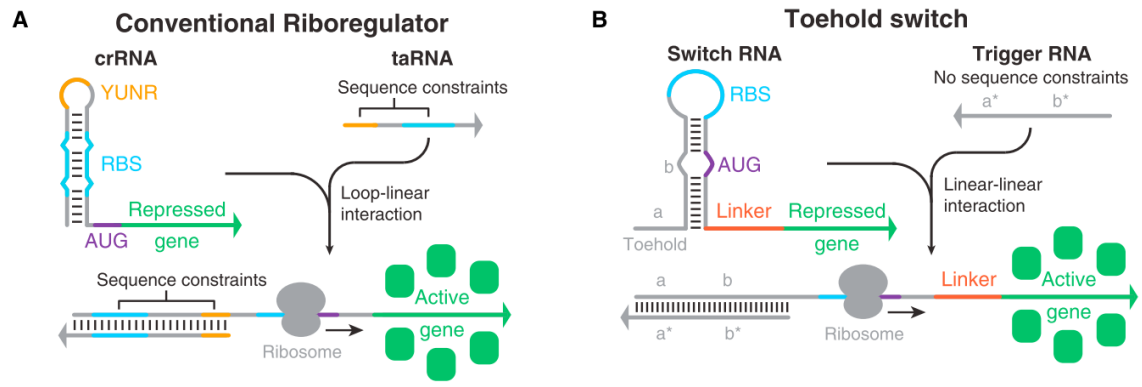


Figure 1.1: **Synthetic riboswitch design by Pardee *et al.*** (a) Structure of naturally occurring riboswitch regulators, where expression is induced by a trans-acting RNA. (b) Synthetic riboswitch designed to remove sequence constraints, allowing the trigger RNA to be any identified mRNA sequence or an artificial oligo. Figure from Green *et al.* [33].

Since this work, similar systems have been developed for numerous other targets including Zika virus, plant pathogens, gut microbiota and malaria [34–37]. In order to boost sensitivity to meet detection targets, various amplification methods were integrated to increase the concentration of the triggering RNA molecules. To prevent denaturation of the extract elements conventional thermocycling (as in PCR) cannot be used, so low temperature isothermal amplification methods were required, such as nucleic acid sequence based amplification (NASBA), recombinase-polymerase amplification (RPA) or loop-mediated isothermal amplification (LAMP). As these required use of intermediate DNA templates, *in vitro* transcription (IVT) was a crucial step in this process to resynthesise the RNA triggers, converting these devices from translation-only to full TXTL systems.

The dependency of these systems on specialised isothermal amplification methods, requiring supplementary enzymes not found within the original cell extracts, shifted the focus away from use of translation machinery. Following the development of faster and more reliable reporting methods from nucleic acid intermediates, the cell extracts themselves became obsolete, resulting in a new range of *in vitro* biosensors, albeit containing complex mixtures of enzymes and their substrates. For example, Hu *et al.* were able to combine RPA with pH sensors to monitor the rate of hydrolysis caused by nucleotide incorporation into amplifying DNA, to create a fast and sensitive biosensor for detecting presence of antimicrobial resistance genes in bacterial isolates [38].

Recently *in vitro* biosensors have been further revolutionised by the introduction

of CRISPR. The ability to design guide RNAs (gRNAs) to target sequences with single-nucleotide precision has substantially increased the specificity these biosensory systems can achieve. Pardee *et al.* included a supplementary CRISPR-Cas9 modality in their Zika virus diagnostic, whereby particular sequences were cut by Cas9, preventing them from triggering a secondary riboswitch. This allowed them to discriminate between different Zika strains once the presence of the virus had been identified by their initial assay. This level of detail is very impressive and if successfully developed would enable tracking of variants within an outbreak with much greater speed than current sequencing allows.

Use of other Cas enzymes has further removed the need for a translation system. The SHERLOCK platform, building on the work by Pardee *et al.*, replaced the riboswitch sensor with direct detection by Cas13a, which is stably activated upon recognition of a complementary sequence by its gRNA and following this carries out enzymatic collateral cleavage of fluorophore-quencher ssDNA oligos [39]. The combination of RPA, IVT and CRISPR-Cas provides three stages of signal amplification and two layers of specificity checks resulted in claimed attomolar sensitivity within an hour. An update to this showed how combination with Cas12a enabled dsDNA recognition and how the system could be adapted for multiplex analysis and LFDs [40]. SHERLOCK has since been demonstrated on different Plasmodium species [41], while a related amplification-free, tandem CRISPR system has claimed detection of clinical SARS-CoV-2 samples within 5 minutes [42].

Notably, this astounding recent progress has focussed on the detection of nucleic acid sequences. While this is understandable given the rapid expansion of CRISPR technology and the ongoing COVID-19 pandemic, sensing of other analytes has lagged behind, with far lower sensitivity and speeds reported. In many cases this is the result of having to design and optimise individual systems for specific targets, whereas the gains in genomic diagnostics are largely sequence-independent so immediately adaptable for countless applications. Attempts have been repeatedly made to develop aptamers - oligonucleotides that change structure upon binding a ligand - to bridge this gap, but these are plagued by specificity issues.

Other analytes have therefore depended on proteins for biosensor signal transduction method, which are much slower to develop. While this is a

limitation, proteins are the macromolecule through which the vast majority of natural sensing occurs, usually in the form of transcription factors regulating expression of effectors. Harnessing these evolved systems instils much greater confidence in detection specificity, as any off-target responses would likely reduce fitness, but the process for doing so is more laborious than for nucleic acids.

For example, an *in vitro* mercury biosensor was recently created through the design of a MerR-EYFP chimera protein [43]. Özyurt *et al.* fused the Hg^{2+}-responsive MerR transcription factor to the N-terminus of the enhanced yellow fluorescent protein, sensitising the reporter to presence of mercury. The resulting biosensor was rapid, sensitive and specific, but the engineering process was not guaranteed to work and may struggle to be replicated for other targets. Gaining a deeper understanding of how cells and organisms naturally use transcription factors to quantify and compute analyte presence may enable more reproducible biosensor designs.

1.4 Biological computation

All cells make decisions based on external stimuli interacting with internal components; even in the most simple systems this aggregation of information and response generation is a form of computation. This is largely carried out through the control of transcription and translation by regulatory proteins responding to small molecule effectors or other proteins. Understanding how hundreds of these function simultaneously in three-dimensional cytoplasm to produce beneficial responses is incredibly difficult, but if successful could lead to the development of powerful cell models as well as improved biosensors.

The other half of biological computation is concerned not with characterising existing decision-making processes, but the creation of new computational methods with a biological basis. One commonly cited advantage of such an aim is the suitability of nucleic acids for information storage. DNA, with its quaternary base system and minimalistic molecular structure, was already shown in 2012 to be capable of storing data at greater densities than conventional non-biological drives, and since then even greater gains have been made [44, 45]. DNA's stability - capable of persisting for thousands of years without significant degradation - also makes it well-suited for storage, while

numerous methods to read and write also exist.

While information storage is useful, operations must be performed on these data to carry out computation, and this is fundamentally more noisy in solutions. Conventional electronic computing organises ever increasing numbers of binary transistors into complex two-dimensional networks, but only desired interactions are ever permitted. Cells are incapable of isolating connections in this way; instead species constantly move throughout a three-dimensional solution with their structures determining the strength of various specific or non-specific interactions. While living systems are inherently more noisy as a result, potential advantages exist over silicon-based logic, with this stochasticity probably harnessed at population-level behaviours such as differentiation, bet hedging and evolution [46].

Despite these emergent properties, the complexity in engineering these systems from scratch is currently too great, so focus has been on digital computation, which revolves around the use of logic gates. Synthetic genetic circuits have been designed to carry out these operations, such as the use of orthogonal activator-chaperone pairs to create layers of AND gates [47]. Other work has used repressors to make NOT gates, or co-opt particular bacterial sub-systems or CRISPR to create ever more functionality [48, 49].

An alternative technique has built on the riboswitch work by Pardee *et al.*, through the realisation that independently controlling the transcription of trigger and switch RNA created an AND gate. As the riboregulated output could also be transcription factors activating synthesis of further triggers or switches, these RNA gates can be used for layered logic too [33]. By co-localising multiple sensing modules onto a single riboswitch, OR gates could be created, while complementary sequences resulted in NOT gates [50]. More bespoke designs enabled creation of NAND and NOR logic gates, critical for more complex functionality [51]. These could be layered to process up to twelve inputs in *E. coli*, but required hours-long timescales and flow cytometry for output quantification. While these are notable developments in synthetic biological computation, their applicability to *in vitro* biosensing is currently limited.

1.4.1 DNA strand displacement

The ability of RNA to form functional structures is well-known. In Figure 1.1, it was shown how both natural and synthetic riboregulators could sequester an RBS within a hairpin structure, preventing translation, until a trigger or trans-acting RNA promoted the formation of a linear molecule that a ribosome could bind. The synthetic riboswitch utilised an exposed single-stranded domain, known as a toehold, to initiate binding of the second RNA and cause the overall reaction to be energetically favourable through the net gain of base-pair formation. This is known as toehold-mediated strand displacement (TMSD), although in this case the displaced sequence is in fact another part of the same switch RNA as the toehold.

The same principle can be used with DNA, although less structure is possible. RNA's continual problem is its lack of stability, whereby its ease of degradation makes it risky to use *in vitro*, unprotected by cells. In the case of the Pardee *et al.* system, reporter expression could be initiated not only by presence of the trigger RNA, but any degradation that released the repressive riboswitch structure. Alternatively, hydrolysis elsewhere could prevent translation occurring even once the trigger was present. DNA is much less prone to hydrolysis and so is more reliable to use *in vitro*.

DNA strand displacement (DSD) is TMSD occurring exclusively between DNA molecules. As DNA does not readily form tight structures such as the riboswitch hairpins, DSD reactions are usually designed to occur between structureless, linear species, resulting in the full release of a displaced strand [52]. Figure 1.2 illustrates how an ssDNA oligo can react with a partially single-stranded duplex to cause a displacement reaction, given the design of appropriately complementary domain sequences. By modifying the gate duplex G with a fluorophore at the 5' end of the top strand and a quencher at the 3' end of the bottom, reaction progress can be visualised and quantified by the increase in fluorescence resulting from the separation of these proximity-dependant modifications. The extent of the reaction can also be altered by changing the ratio of G:X.

This basic bimolecular displacement effectively takes the form of an arbitrary chemical reaction $G + X \rightarrow H + Q$. By carefully expanding the number of orthogonal domains and designing DNA species with more elaborate structures,

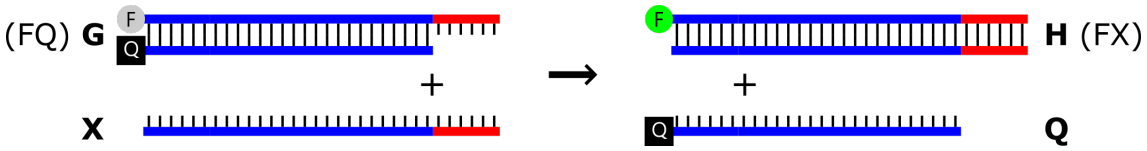


Figure 1.2: **Basic DSD reaction.** Schematic showing the duplex G, comprising of a dsDNA recognition domain (blue) and an ssDNA toehold domain (red) and modified by a fluorophore on one strand and a quencher on the other, reacting with fully complementary displacing oligo X. X can bind G through its toehold and once bound progressive branch migration causes the full release of oligo Q and formation of the fully dsDNA fluorescent duplex H.

more complex chemistry can be achieved, such as the creation of DSD-only logic gates and amplification methods [53, 54]. Signal processing functions such as thresholding, oscillations and complex logic circuits can also be created [55, 56].

Unlike other systems, DSD reactions are completely enzyme-free, relying purely on DNA molecules interacting according to mass-action kinetics and the thermodynamics of their base-pairing. Reliance on these simple primitives makes them not only much more reproducible systems but also relatively easy to model, given estimations of domain binding energies and their resulting association and dissociation rate constants [57]. As a result the Visual DSD programming language was created to enable abstracted modelling of these systems at different levels of mechanistic detail, using established syntax for defining DNA structure and providing methods for both deterministic and stochastic simulations as well as parameter inference [58–60].

Since its inception, Visual DSD has demonstrated the ability to model complex behaviours such as analogue signal processing and feedback control circuits [61, 62]. The ability to simulate these complex systems *in silico* not only requires a deep understanding of the underlying reaction mechanics, but speaks for the reproducibility of these circuits and the precision with which specific behaviours can be engineered.

1.4.2 DSD biosensing

The highly controlled nature of DSD makes this framework well-suited for biosensory applications. When precise, reliable quantification is desired, a system built and simulated with robust mathematics is greatly preferable to overly complex biological soups much more prone to stochastic responses.

The aim of this project was to develop methods of limiting effects of environmental variability within *in vitro* biosensors, such as those based on DNA strand displacement. Like other cell-free systems, exposure to unknown analyte solutions containing differing quantities of contaminants can greatly affect their output, limiting target quantification. A system that could report the same concentration regardless of variable backgrounds, without extensive sample preparation methods, would massively increase the range of biosensor targets. Commercially successful *in vitro* biosensors rely on sampling relative stable biological solutions such as blood or urine for this reason.

The most direct application of DSD biosensing is for nucleic acid detection. Designing signal transduction elements for these is very straightforward, as shown by Pardee *et al.* [31]. However, despite the advantages that enzyme-free, DNA-only systems provide, DSD is currently outperformed in a couple of key metrics. First, response speed lags orders of magnitude behind contemporary CRISPR-based systems. Initially sensitivity was the primary issue, however new methods such as branched hybridisation chain reaction (HCR) claim to bring detection limits down to attomolar levels [63]. This is similar to that suggested by the SHERLOCK or FIND-IT systems, but takes hours rather than minutes to achieve [42, 64]

Another problem, surprisingly, is specificity. While an appropriately designed DSD gate or riboswitch would seem to be highly specific for its complementary sequence, basic DSD reactions are able to tolerate a small number of nucleotide mismatches. While in many circumstances this may not be a problem, mismatched targets would lead to incorrect results as they would displace duplexes with lower but perceptible rates. Alternative probe designs enable better discrimination of single-nucleotide variants, but this is still less than CRISPR-Cas12/13 systems, which are highly sensitive to bulges caused by mismatches and so offer unmatched specificity [65].

As a result, while DSD may be capable of more precise quantification, for

nucleic acid detection (particularly in the context of field testing) this is currently overshadowed by the speed and precision of CRISPR systems. DSD biosensors have been developed for other targets, although these predominantly rely on aptamers, which are notoriously non-specific [66]. To achieve the aim of developing a viable DSD-based biosensor, which could not only match top performing sensitivity and selectivity metrics but also enable testing of methods to control environmental variability, another target and another method of signal transduction was necessary.

1.5 Arsenic

Arsenic is well known for its acute toxicity. In popular culture it is used by assassins to murder foes without detection, through its addition to the food or drink of a victim. This is possible through our inability to see, smell or taste the metalloid, as well as the non-specific nature of its lethality. However, the full properties and real-world consequences of arsenic are much less appreciated.

Modern exposure to arsenic is commonly associated with bad industrial practices such as in mining, but the element is naturally present in the Earth's crust all over the world. Due to the geology that influences its distribution, arsenic is most abundant in South and East Asia, as well as South America, but it is found in every continent to varying degrees [67]. Figure 1.3 highlights the regions most affected by arsenic-contaminated aquifers. These areas are associated with river deltas, due to groundwater concentrating the movement of arsenic from surrounding deposits into these low-lying regions. As the original deposits are associated with other minerals, certain mining practices tend to enrich for it, hence the human sources of environmental pollution.

Arsenic can exist in four oxidation states, but two are most prevalent - arsenite (As(III), commonly as AsO_3^{3-}) and arsenate (As(V), commonly as AsO_4^{3-}). The latter is more common in aerobic conditions while the former is more mobile in water [69]. Industrial uses for arsenic include pesticides, chemotherapy and high-performance semiconductors.

While toxic to most life, arsenic has some important biological roles. Arsenate is used by some microorganisms in anaerobic respiration, while others oxidise arsenite during carbon fixation [70]. A by-product of this arsenic metabolism

Figure 1.3: **Worldwide distribution of concentrated arsenic.** Illustration highlighting the regions where arsenic-contaminated aquifers are found at the highest concentrations. Taken from Chowdhury [68].

is the occasional formation of organic arsenic species (organoarsenicals) such as methylarsonic acid, which importantly can accumulate in food like rice. While all these different species play slightly different roles, they are broadly all poisonous and hard to detect.

1.5.1 Arsenic poisoning

Arsenic's toxicity is a direct consequence of its chemistry. The arsenate ion is structurally extremely similar to phosphate, as arsenic is in the same periodic group as phosphorous, and so competitively inhibits oxidative phosphorylation, key to respiration in all aerobic organisms. Arsenite, on the other hand, reacts strongly to thiol groups and as these are crucial to the structure and function of a huge range of proteins even anaerobic microorganisms are not safe from arsenic's effects. Together these explain how arsenic can be dangerous to all life forms [69].

Single-celled organisms must avoid arsenic or resist its effects to survive. Multicellular organisms, on the other hand, can risk exposure if poisoned cells are eventually replaced, and so are not as evolutionarily pressured into requiring direct detection and resistance mechanisms. Thankfully most humans do not encounter arsenic regularly and so this vulnerability is inconsequential. However, in the regions highlighted in Figure 1.3 where arsenic deposits are

picked up by groundwater and leach into aquifers, locals can be exposed to contaminated water and food. Colourless, odourless and tasteless, it is then used and consumed unknowingly.

The effects of arsenic poisoning are strongly dependant on exposed concentration, which can vary greatly. Ingestion of high quantities results in vomiting, diarrhoea, nerve issues and, if high enough, death. However, the amounts required for these acute effects are rarely encountered naturally and the more widespread issues derive from long term, chronic exposure from lower quantities. Changes to skin pigmentation, the formation of lesions (arsenical keratosis) and 'blackfoot' disease (from extremity blood vessel damage) are common visible symptoms if dose is relatively high, shown in Figure 1.4, but most symptoms are internal and less identifiable as resulting from arsenic [71].

Figure 1.4: **Visible symptoms of arsenicosis.** Examples of superficial lesions caused by high arsenic exposure. Taken from Barkat *et al.* [72] and Smith *et al.* [73].

Arsenic poisoning is known to contribute to diabetes, renal failure, peripheral neuropathy and a host of cardiovascular and developmental diseases, including to unborn children from exposure during pregnancy. Arsenic is also classified by many agencies as a carcinogen, significantly contributing to skin, bladder and lung cancers among others. These wide-ranging, largely non-specific, chronic symptoms make it hard to conclusively identify arsenic as the major cause of eventual death in casualties, and therefore quantify the full prevalence of arsenic related disease and death [70].

While the level of arsenic 'safe' to drink will vary person to person, the World Health Organisation (WHO) have established a limit of 10 µg/l (10ppb), revised down from previous limits of 50ppb [74]. Drinking water below this threshold is supposedly unlikely to result in adverse effects, while prolonged exposure greater

than this could. In reality a continuous spectrum exists, where the greater the consumption the greater the risk.

1.5.2 Bangladesh groundwater contamination

Out of all the countries affected by arsenic contamination of water sources, Bangladesh is widely agreed to be worst-hit. The WHO have even described the situation here as the 'largest poisoning of a population in history', surpassing the likes of Chernobyl [73]. Several factors contribute to this circumstance. First, the natural geology in the surrounding area contains many arsenic-rich minerals. Second, Bangladesh is dominated by low-lying countryside surrounding river deltas, with water originating from an extensive basin combining the Ganges, Brahmaputra and Meghna rivers. Third, the country is the 6th most densely populated in the world, below only much smaller countries, with roughly 170 million people living in an area similar to England, largely in rural communities with relatively high rates of poverty [67, 74].

As a comparatively poor country with challenging geography, nationwide infrastructure is limited, notably the provision of clean, piped drinking water. Historically this led to a dependency on surface water (e.g. rivers and ponds) for both drinking and bathing, but this lack of separation between consumption and waste resulted in high spread of bacterial gastrointestinal diseases. To combat this, a public drive backed by UNICEF from the 1970s installed millions of shallow (<150m) tubewells to instead tap into the plentiful, and supposedly safe, groundwater [68].

Initially this was seen as a success, with rates of pathogenic diseases substantially reduced. However, towards the end of the 1990s it became clear that rates of skin lesions and cancers had risen dramatically, eventually traced back to arsenic-contaminated groundwater. Several surveys have since been conducted to estimate the full scale of this problem, notably by the WHO and the British Geological Survey (BGS) [67, 73–75].

Due to the lack of local testing capability, samples were predominantly collected and sent abroad for physical spectrometry analysis, revealing several important findings. First, there was large country-wide variation in arsenic content, ranging from regions with minimal quantities to areas with extremely high

amounts, shown in Figure 1.5 (a). Second, beneath this broad variation there were great differences in arsenic from different tubewells even within a single village. Figure 1.5 (b), shows data for the village of Mandari, with wells coloured by arsenic content to highlight the large differences present across this small area. Some of this variation can be attributed to differences in tubewell length, with the medium depth pipes tending to have the most arsenic, while the most shallow and the deepest seemingly safer. However, great spatial heterogeneity clearly existed even with similar depth tubewells [67].

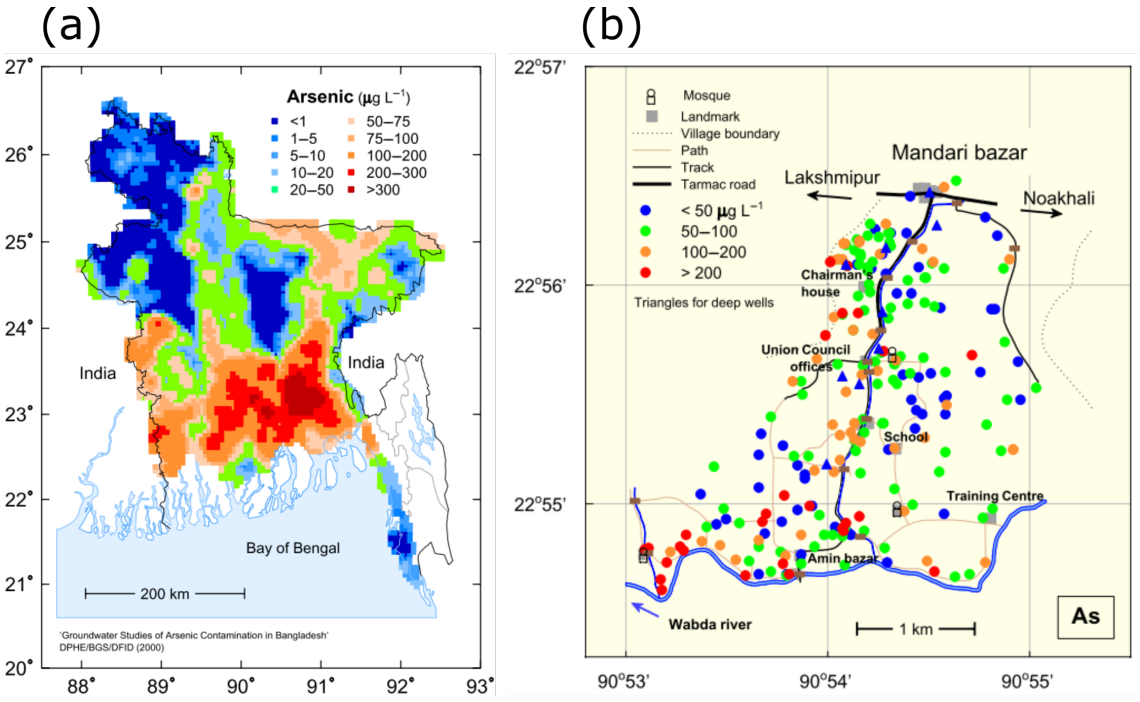


Figure 1.5: **Bangladesh tubewells show great variation in extent of arsenic contamination.** (a) Country map coloured according to estimated average arsenic content of shallow tubewells, smoothed over large areas. (b) Variation in arsenic concentration from individual tubewells within Mandari, an example village, coloured by different thresholds. Taken from Kinniburgh & Smedley [67].

Third, variation exists over time as well as space. This is largely associated with periodic leaching events caused by rainfall, particularly in monsoon season, but other factors may play a role. Together, these highlighted the importance of regularly testing wells, enabling locals to identify the safest water sources in real time [67].

It is estimated that over 150 million people worldwide are regularly exposed to drinking water with greater than 10ppb arsenic, 50 million of which live in Bangladesh. 5 million people in this country are thought to drink water with greater than 200ppb arsenic, a staggering 20x above this limit. At least 40,000

people a year in Bangladesh are believed to die directly from drinking this water, but this number is likely to be far greater due to the difficulties in weighting causes of multifactorial diseases. Counting deaths alone also grossly simplifies the adverse effects this element causes, with many disability-adjusted life years (DALYs) lost as well as billions of annual dollars of GDP in the resulting wasted productivity and healthcare costs [74].

Long term, the solution to this problem is an overhaul of public health infrastructure with piped clean water available to all homes, however the cost of this is enormous and will require decades of investment. Efforts to filter water exist but these are also difficult and costly, even with progress in bioremediation technology [76]. In the short term therefore, regular testing of tubewells, enabling well-switching if contamination is high, is seen as the most practical method to best reduce the burden of arsenic-contaminated drinking water [71].

1.5.3 Conventional methods of arsenic detection

The large-scale surveys referenced above relied on physical methods to interrogate arsenic concentrations of water samples. These included atomic fluorescence spectrometry (AFS), inductively-coupled-plasma atomic-emission spectrometry (ICP-AES) and inductively-coupled-plasma mass spectrometry (ICP-MS), occasionally preceded by hydride generation (HG). All these methods require large, expensive equipment operated by trained technicians after carrying out extensive sample preparation. As a result, the vast majority of the data were produced by storing acidified samples at low temperature and flying them to laboratories in other countries such as the UK [67].

These efforts were necessary to produce accurate and reliable arsenic quantification. With HG, some of these methods had detection limits under 1ppb at the time, reproducible within around 10% variation, far more sensitive than other techniques. To identify safe water sources (<10ppb) and quantify how dangerous others were, there was no alternative. However, for the purpose of informing locals of real-time toxicity, rather than producing a report, these methods are clearly unusable. Modern-day spectrometry is even more sensitive but transport time, cost of equipment and testing, need for technicians and sample preparation are all prohibitive [1].

POC testing requires reliable, cheap, quick and practical field kits. A number of chemical kits have been produced and used over the years to address this, but they too have issues. The 'Arsenator' is one of the most widely marketed examples and was compared to the laboratory methods in Kinniburgh & Smedley's BGS report. They found the Arsenator data did correlate fairly well with the spectrometry results, although only at high arsenic concentrations and even then with some notable outliers, despite claims of detection limits under 10ppb. And while the kit was able to be used *in situ* and takes only 20 minutes, the device costs over $1000 and the consumed reagents are both difficult to use and inherently dangerous themselves.

Many of these chemical methods, including the Arsenator, rely on the production of arsine gas from inorganic arsenic to then react with mercuric bromide to give faint yellow colour, known as the Gutzeit method [77, 78]. If released this gas is very toxic, much more than dissolved ions, and the protocol to perform measurements is too complex to guarantee safety if undertaken by untrained users. After accounting for operator skill, reagent quality, contaminant presence and sample preparation the resulting data were very unreliable [1, 76].

For these reasons there is still demand for cheaper and more reliable portable sensing devices, and this is where biotechnology has made an impact.

1.6 Arsenic biosensors

Arsenic biosensors have the potential to solve many of the problems posed by current chemical and physical methods. Appropriately selected biological components should enable greater selectivity than chemical alternatives, while the range of signal amplification methods should allow sensitivity below 10ppb. The different designs employed, however, can greatly affect assay speed, cost, usability and level of sample preparation required.

As with the non-arsenic biosensors described above, whole-cell and *in vitro* attempts have been made. The former tend to rely on natural *ars* operons, while the latter tend to harness purified proteins or nucleic acids. Over 50 arsenic biosensors have been reported, with detailed reviews comparing their properties [1, 79, 80]. A few of the more successful attempts are highlighted below.

1.6.1 Whole cell arsenic biosensors

In order to avoid the consequences of arsenic poisoning, many microorganisms encode *ars* operons to detect and respond to its ions [81, 82]. These typically include an arsenic repressor (ArsR) controlling the expression of an arsenate reductase (ArsC, reducing arsenate to arsenite for easier removal) and an efflux pump (ArsB, actively exporting arsenite from cells) [83]. Additional regulators (ArsD) and ATPases (ArsA) can also be found in more extensive operons that provide higher levels of resistance [84]. Arsenic resistance operons are found all over prokaryotes and are highly related to each other and to operons for resistance of other metals, often in mobile elements but chromosomally integrated into many genomes, highlighting both shared ancestry and the common need for these detoxification methods [85].

Most arsenic biosensors harness these operons, particularly *E. coli* homologs, for the creation of whole cell reporters. These tend to use recombinant fluorescent protein genes engineered into the operon, to be co-expressed with the other effectors once arsenic is detected by ArsR. Despite the similarity to previous work, new attempts at this are constantly made with minor tweaks to try to improve performance. For example, Jia *et al.* used the LuxR transcriptional activator, regulated by ArsR and itself, to amplify the expression of mCherry through a positive feedback loop [86]. While this improved arsenic sensitivity to just below 10ppb, incubation still took 6-10 hours.

Luciferase is an alternative bioreporter, not requiring light excitation and generally offering greater sensitivity compared to fluorescent proteins [1]. Siegfried *et al.* transformed *E. coli* with a plasmid in which ArsR controlled the expression of a bacterial *lux* operon, thereby synthesising the enzyme in response to arsenite [78]. Culture aliquots were lyophilised in glass vials, to be reactivated by the injection of water samples. After a 2 hour incubation at 30 °C, luminescence was analysed and compared to standards for inference of arsenic concentration, with a claimed 5ppb detection limit. Whereas many other biosensors have not been field tested, this 'ARSOlux' system was trialled in Bangladesh and although quantitative accuracy was limited, the system was fairly good at identifying samples above or below their 50ppb threshold.

Siegfried *et al.* noted their assay was safer and simpler to use than chemical systems such as the Arsenator, while providing similar results. The ARSOlux

took longer, but many samples could be prepared and tested in parallel, so all the wells within a village could be assayed in a day. However, the luciferin substrate and the luminometer equipment are relatively expensive and the prepared vials had to be stored at 4 °C. While simpler than chemical methods, the assay, including sample and standard preparation, would still require training to perform. Also, despite the vials being 'closed' GMO systems they still required specialist permission for testing, and could not be legally commercialised [79, 87].

Another method to improve whole cell bioreporting sensitivity utilised modular synthetic biology parts to tune the response of genetic circuits. Wang *et al.* altered RBS and promoter strength controlling both ArsR and its GFP reporter to boost arsenic sensitivity to under 10ppb, including an innovative Buffer gate to further amplify signal, another example of successful implementation of biological signal processing [88–90]. They also designed strains layering logic gates to improve selectivity for a range of metal ions, including increased discrimination of arsenic from mixed arsenic/mercury solutions.

Further work has brought this system closer to use. To reduce background fluorescence a protein degradation tag was fused to GFP, separated by a TEV protease recognition site. Tightly controlling TEV protease expression by ArsR meant that high arsenic stabilised the reporter, while GFP was degraded faster at low concentrations [91]. In addition, a device was created trapping differently-tuned strains in a hydrogel grid to produce a visible bar scale output depending on detected arsenic concentrations. Images could then be taken by a fluorometer or a phone to interpret the results.

This design shows real promise, with good sensitivity, selectivity, dynamic range and output thresholding. The attempt to replace requirement for a fluorometer with a phone is admirable, although supplementary diodes and filters are required for this in practice, and specialist software is needed to analyse the images. Despite this progress, the main limitations to the use of whole cell biosensors still exist. The Wan *et al.* system required 24 hours of incubation at 37 °C - the trade-off for the layered circuits to boost sensitivity and reliability - and as the cells were not lyophilised they required extensive preparation immediately prior to use. While these hurdles may be reduced in time, better reporting methods and a change in legislation will be required before whole cell biosensors are widely adopted [92].

1.6.2 *In vitro* arsenic biosensors

The shortcomings of whole cell biosensors have driven the push for *in vitro* solutions. The use of purified biomolecules circumvents the need for live organisms and enables integration with more practical reporting methods. However, without the natural filtration cells provide, these systems are highly vulnerable to contamination and interference from other solutes, often rendering them unreliable outside the lab.

Numerous aptamers have been designed for this purpose, claiming high sensitivities and speeds and integration with outputs such as electrodes or nanoparticle aggregation [93, 94]. Despite these advantages, however, their specificity has been called into question many times. The best known arsenite/arsenate aptamer, Ars-3, was originally selected from *in vitro* evolution experiments for potential bioremediation, but has since been the key element to at least 24 aptamer biosensor papers [1, 95]. Recent work, however, claimed that Ars-3 was not capable of binding arsenic ions stronger than random control oligos, hypothesising instead that the gold nanoparticles used in the original work provided the non-specific affinity [96].

This highlights the difficulty in using key signal transduction elements derived from artificial, not natural, selection. Specificity is not proven unless fully controlled and tested against all contaminants likely to be found in samples, at relevant concentrations and combinations. Reliability is much more likely if the biological element is one used by organisms for natural detection - the difficulty is harnessing these for biosensing.

One highly developed method harnessing a naturally evolved bioelement relies on the use of arsenite oxidase (Aio) enzymes. Work by the Santini lab at UCL identified these periplasmic proteins from *Rhizobium* bacteria, isolated from an Australian gold mine known to be high in arsenic [97]. This chemolithoautotrophic prokaryote could not only tolerate these conditions but used arsenite oxidation to drive carbon fixation. Despite the difficulty in working with novel bacteria, they were able to express, purify and characterise the proteins involved in this mechanism, and identified their potential for use in a biosensor, by replacing GOx in a glucometer with Aio [98, 99].

Together with Professor Tony Cass, who was involved in the creation of the first electrochemical glucometers, an 'AquAffirm' device was prototyped and

patented [100]. This claimed rapid sensitivity down to 1ppb arsenic with a disposable cartridge, to which the water sample is applied, inserted into a handheld reader. This format is highly convenient, however the device is yet to hit the market despite the original patent being published in 2013. Limited data are available on this system, but the original publication demonstrated the ability to detect 20ppb arsenite spiked into a few different real water samples, although it is not clear how reliable this is or whether dose linearity was preserved [101]. The device did seem to experience interference from river water and electroactive solutes like humic acid and a dual mediator approach was proposed to combat this (using an Aio-free control electrode), but no data was provided for the performance of the resulting system [102]. This device would also inherently only be able to detect arsenite, not arsenate, although identification of parallel arsenate reductases (Arr) may solve this issue [103].

An alternative signal transduction element is ArsR itself. This is the protein through which most known microorganisms sense arsenite, with a ligand-induced conformational change causing dissociation of the repressor from its operator sequence, enabling expression of the effector proteins. While this is the key mechanism behind all whole cell and TXTL arsenic biosensors, few genuinely *in vitro* biosensors utilise this, largely due to the difficulty in finding a suitable transcription-free reporting system.

One innovative system simply expressed an ArsR-GFP fusion protein, with arsenic samples added directly to cell lysates [104]. The incubated mixture was then added to microplate wells with copies of the ArsR operator dsDNA immobilised to the bottom; after a further incubation the solution was washed off, leaving only ArsR-GFP proteins that had not been bound by arsenic. Fluorescence in the washed wells then negatively correlated with sample arsenic content. This simple system was capable of sensing down to 5ppb, and could detect arsenate too through the use of sodium thiosulfate reduction to arsenite. A selection of other metal ions did not significantly alter fluorescence, although antimony seemed to cross-react and fluorescence was also affected by NaCl content. While correlation seemed relatively robust within experiments, readings were variable between different backgrounds, suggesting difficulty in actually inferring concentrations from unknown samples.

A lyophilised version of this system reduced reaction time to 30 minutes and was supposedly capable of 10ppb arsenic detection in milk and yoghurt, while a

further update using thermoresponsive magnetic beads dropped these to 1ppb in 5 minutes [105–107]. While they were able to develop parallel systems for cadmium and mercury, they all still lacked consistent quantification across different backgrounds [108]. This reflects the key limiting factor for *in vitro* biosensors - tolerance of environmental variation.

Aside from the work by Santini *et al.*, the vast majority of both whole cell and *in vitro* systems utilise genes, proteins or operons from *E. coli*. While this is often the easiest approach, it is highly likely alternative microorganisms may have components enabling greater sensitivity, selectivity, speed, or even different dynamics. Arruda *et al.* found that the *Chromobacterium violaceum* ArsR, while less sensitive to arsenite than *E. coli*, produced a steeper response curve indicative of thresholding dynamics [109]. Without the need to culture cells, *in vitro* biosensors should be primed to exploit the vast array of natural variation and be more ambitious when designing new systems.

1.7 Design of a repressor-DSD biosensor

The idea for this project stemmed from the desire to expand the range of targets offered by DSD biosensors and to explore new methods of biological signal processing. Whole cell arsenic biosensors had been previously investigated by members of this lab and while their limitations were identified, the unique circumstances surrounding arsenic detection presented an inviting challenge. Reviewing the literature on current *in vitro* methods above identified the importance of ArsR proteins to many of these and led to the hypothesis that these repressors could be used to modulate DSD reactions directly. If possible, integration with existing DSD designs and creation of new methods of signal processing could overcome existing challenges.

The schematic presented in Figure 1.6 demonstrates how this could work. If a DSD probe was designed such that the operator sequence of a repressor was contained within the double-stranded region, a reaction that naturally progressed (a) could be prevented through pre-incubation with this repressor (b). Addition of the allosteric ligand, in this case arsenic, should cause repressor dissociation and thereby concentration-dependant signal increase.

Such a minimal system could have a number of advantages. First, the design has

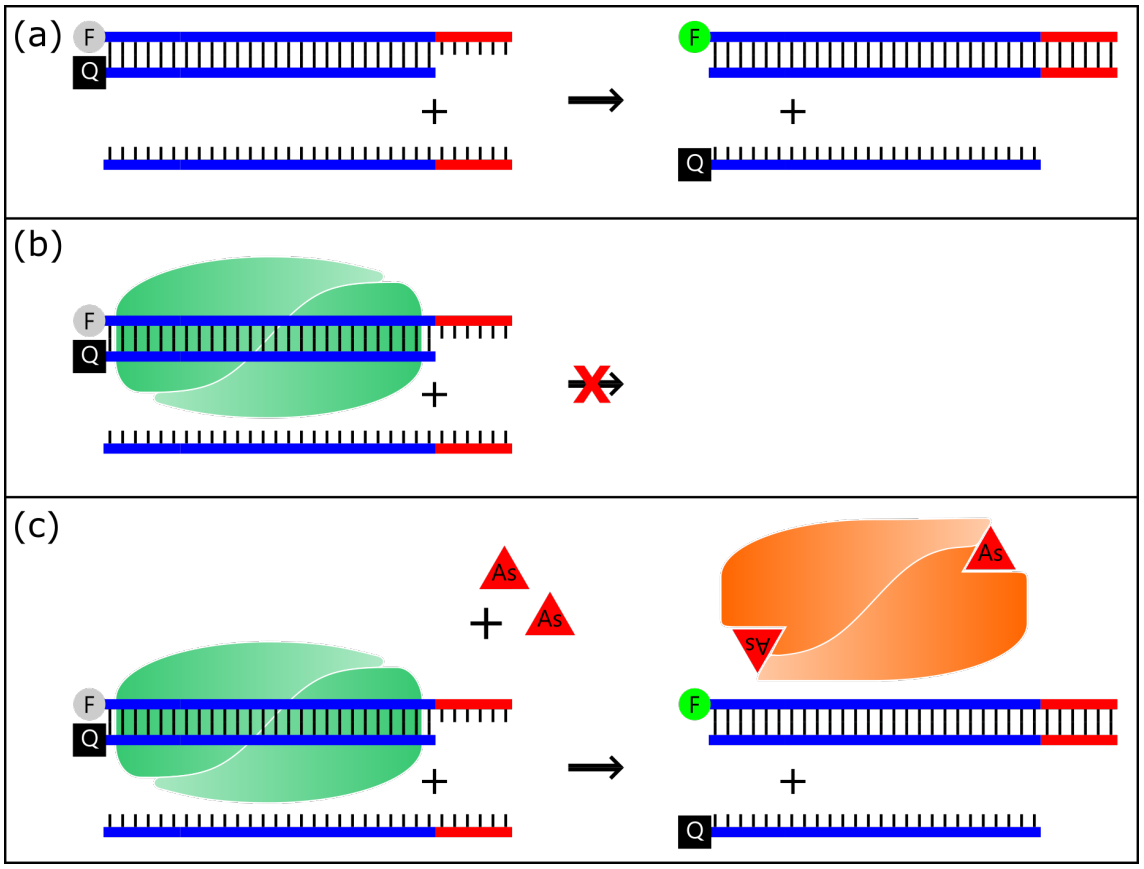


Figure 1.6: **Repressor-mediated control of DSD reaction.** Schematic illustrating how a repressor protein could modulate a basic DSD reaction: (a) Basic irreversible DSD reaction. (b) Repressor binds operator sequence (blue), preventing DSD from occurring. (c) The ligand (in this case arsenic) induces dissociation of the repressor from its operator sequence, thereby enabling the DSD reaction to progress and fluorescence to accumulate.

a good chance of being robust to the environment. Without containing any RNA and using only a single protein, possibilities for degradation or interference are limited. While the repressor could be vulnerable to both of these, several could be compared and the most robust selected. Second, the assay should be very cheap. Unmodified DNA and culture-expressed proteins are inexpensive, while costs of modified oligos greatly reduce with scale. Third, fluorophore-derived fluorescence is a stable output, much more reliable than fluorescent proteins or enzymes. This should be a useful interrogation tool, although reporting method could be changed to a colorimetric/electrochemical output if eventually desirable.

Finally, this minimal system is suitable for mechanistic modelling. DSD-only systems are well-characterised and if the behaviour of the repressor can be captured and the complete assay simulated, this could lead to possibilities including component optimisation and integration of amplification or

thresholding methods. This is a good way to better understand the parameters affecting biosensor performance, compared to simple trial-and-error. The effect of contaminating solutes could then be investigated and techniques developed to control for these.

Several challenges are sure to present themselves, however. First, appropriate repressors will need to be selected. Both *in vivo* and *in vitro* behaviour will need to be scrutinised, including the suitability of its operator sequence. So far, DSD reactions predominantly deal with non-biological, structureless oligos and ensuring the sequences used conform may be difficult. Second, candidate repressors will need to be expressed, purified and characterised for ability to function in a DSD system. If any are sufficiently compatible then design of their probes, reactions and buffers will need to be optimised. If this can be achieved then basal biosensor performance would need to be established, before testing methods to normalise signal and control variability. Finally, efforts should be made to model the system, to improve understanding of interactions, optimise composition and test potential extensions.

Combining DNA strand displacement technology with repressor proteins is something never attempted before, but if successful could open up a range of similar biosensory systems. Transcription factors are pervasive natural signal transduction elements and if they can be harnessed in controlled *in vitro* systems, a huge diversity of analytes could be explored. As well as attempting to create a novel arsenic biosensor for proof of this principle, equally important will be the identification of design principles that may influence other systems, so that parallel biosensors can be rapidly developed in future.

Chapter 2

Materials and Methods

2.1 Expression plasmid preparation

Repressor coding sequences were obtained from the NCBI database using the accession numbers found in the primary literature. The destination vector chosen was the pET-28a plasmid (Invitrogen), previously used for protein expression by other members of the lab as well as a couple of the more recent repressor protein studies. The pET-28a plasmid contains a pBR322 origin of replication (15-25 copies per cell), a kanamycin resistance gene, a LacI coding sequence and a T7 promoter upstream of a multiple cloning site (MCS) containing options for both N-terminal and C-terminal His-tags depending on cloning. Most functional studies, however, have suggested the N-terminal sequences of the arsenic repressor proteins contained both binding and dimerisation domains.

While various ArsR proteins have been successfully purified with N- and C-terminal tags, the latter option seemed less likely to impede protein functionality. The crystal structure of the closest related protein, the *E. coli* cadmium repressor (CadC), on which the R773 *E. coli* ArsR protein structure was homology modelled, also suggested the C-terminus of this (and similarly structured proteins) was easily accessible and physically further away from the functional domains of the protein [110, 111]. A C-terminal His-tag was therefore encoded in each of the repressor gBlock constructs, and with the pre-existing N-terminal tags excised from the pET-28a plasmid.

Transformation protocol

E. coli DH5α competent cells (NEB) were thawed on ice for 15 mins and split into 2x25 µl aliquots. 1 µl of pET-28a stock (50 ng) was added to one (none to the other as a negative control) and samples were incubated on ice for 30 min. Samples were heat shocked for 40s at 42 °C and then placed back on ice for 5 min. SOC outgrowth medium was added to each tube (250 µl, NEB) before incubation (37 °C, 180 rpm) for 2 hr. 50 µl samples of each transformation (plus 1/10 dilutions in SOC) were spread on LB-agar plates (supplemented with kanamycin for 50 µg/ml final concentration) using glass rattler beads (Zymo Research). Plates were then incubated overnight at 37 °C.

Culture growth

Individual colonies were picked from the diluted transformation plate and grown in 10 ml LB Lennox media with kanamycin (50 µg/µl) in a 50 ml falcon tube. Cultures were grown overnight (37 °C, 200 rpm) and then centrifuged (5000 g, 5 min, 4 °C). Pellets were frozen at −20 °C before further usage.

DNA extraction

Plasmids were purified from 10 ml culture pellets using the QIAprep Spin Miniprep Kit (Qiagen). Buffer volumes were doubled and the lysis supernatant was added to the column in two stages in accordance with the manufacturer's instructions. DNA was eluted in 40 µl nuclease-free water (pre-heated to 50 °C and incubated on column membrane for 5 min) and the eluate recycled with one further centrifugation step through the column to maximise yield.

2.2 Repressor coding construct design and cloning

A hybrid restriction digest/Golden Gate approach was chosen to clone the synthesised ArsR genes into pET-28a. The pET-28a MCS can be excised with the NcoI and NotI restriction enzymes, leaving a linear fragment with a bottom strand 3'-GTAC-5' overhang (able to accept a 5'-CATG-3' overhang including ATG start codon) and top strand 5'-GGCC-3' overhang. A restriction-enzyme-only approach was unfeasible due to the need for a CCATGG recognition sequence in the gBlock, requiring the second codon to have a guanine first base. Not all the genes met this, so an ArsR-independent method of

creating these overhangs was used.

The gBlocks were therefore designed with asymmetric BsaI recognition sequences at either end, inwardly cutting to produce the requisite 5'-CATG-3' and 5'-GGCC-3' overhangs. The gBlocks were combined with the purified, digested pET-28a so gBlock digestion and ligation into the plasmid could occur simultaneously. The palindromic overhangs enabled the formation of vector-vector or insert-insert chains/circles, however the correct vector-insert conformation was expected to form with sufficient yield using this method, and the incorrect species were unlikely to transform/proliferate/resist kanamycin as well. The end result was an efficient hybrid cloning strategy independent of the insert sequence and without the need for any polymerase steps that could introduce point mutations.

Raw ArsR genes were supplemented with a C-terminal 6xHis tag, separated by a TEV protease recognition sequence and a spacer. This enabled the removal of the His-tag post-purification if found to interfere with functionality. The resulting codon sequences were codon-optimised for *E. coli* expression using the IDT tool, after which the BsaI recognition sequences were added either side with terminal spacers, to give the final gBlock sequences shown in Table A.3.

Restriction digest conditions

In each 50 µl tube – 19 µl pET-28a (110 ng/µl, 2 µg), 5 µl CutSmart buffer (10x), 1 µl NotI-HF (NEB, 20 U/µl), 1 µl NcoI-HF (NEB, 20 U/µl), 24 µl water. Two 50 µl tubes were incubated at 37 °C for 1 hr, then 80 °C for 20 min to inactivate the enzymes.

Gel electrophoresis conditions

Gels were made with 1% UltraPureTM Agarose (Invitrogen) with 10 µl SYBR Safe DNA Stain (Invitrogen) in a 100 ml 1x TBE solution. 10 µl of 6x loading dye was added to each 50 µl restriction digest sample, and four lanes of 30 µl were loaded. 10 µl of the GeneRuler DNA ladder (ThermoFisher) was loaded in adjacent lanes to check product sizes and gels were run for 80 V for 10 min then 100 V for a further hour in 1x TBE buffer.

Gel extraction

DNA was extracted from gel bands using the QIAquick Gel Extraction Kit

(Qiagen). The manufacturer's protocol was followed, with the DNA eluted in 40 µl nuclease-free water (pre-heated to 50 °C and incubated on column membrane for 5 min) and the eluate recycled with one further centrifugation step through the column to maximise yield. After determination of concentration with a NanoDrop spectrophotometer, samples were concentrated using the DNA Clean & Concentrator Kit (Zymo Research). The pooled samples contained 32 µl of digested plasmid at 49 ng/µl.

Golden gate reaction conditions

In each 20 µl tube – 3 µl digested plasmid (49 ng/µl), 3 µl insert DNA (40 fmol) or 3 µl water for control samples, 2 µl ligation buffer, 1 µl BsaI-HF (NEB, 20 U/µl), 1 µl T4 DNA ligase (NEB, 400 U/µl), 10 µl water. Samples were incubated at 37 °C for 2 hr, then 50 °C for 5 min (final digestion step) and 80 °C for 5 min to inactivate the enzymes. (Engler and Marillonnet – Golden Gate Cloning chapter of DNA Cloning and Assembly Methods [112]).

CgArsR-16S mutagenesis

To create the expression plasmid for the CgArsR-16S mutant repressor, the pET-28a-CgArsR plasmid underwent mutagenesis PCR. Primers were designed using http://nebasechanger.neb.com/ to give the sequences listed in Table A.2. In a 25 µl tube - 12.5 µl Q5 High-Fidelity 2x Master Mix (NEB), 1.25 µl each 10 µM primer, 0.25 µl of 100 µM pET-28a-CgArsR, 9.75 µl water. Cycling conditions - 30 s at 98 °C; 25 cycles of 10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C; 2 min at 72 °C.

The linear PCR product was gel-purified then a KLD reaction was performed to degrade any remaining template and circularise the mutant plasmid. In a 10 µl tube - 1 µl purified PCR product, 1 µl 10x T4 DNA ligase buffer (NEB), 1 µl T4 DNA ligase (NEB), 1 µl DpnI (NEB), 1 µl T4 PNK (ThermoFisher), 5 µl water. Samples were incubated at room temperature for 1 h before transformation.

Transformation protocol

E. coli DH5α competent cells (NEB) were transformed with the cloning products as for the original pET-28a plasmid. 10 µl of each golden gate reaction, or 5 µl of the KLD reaction, was added to each 25 µl aliquot of competent cells. Incubation with SOC outgrowth medium (37 °C, 180 rpm) was carried out for 75 min. LB-kanamycin plates spread with the incubated cells were then incubated overnight at 37 °C.

Sequence verification

Colonies were picked and grown as before, and plasmids purified with the same miniprep kit. DNA samples were diluted to 100 ng/µl with nuclease-free water and 5 µl aliquots were sent for Sanger sequencing (SourceBioscience). T7 forward and reverse stock primers were used to sequence the inserts and flanking regions.

2.3 Expression of repressor proteins

Preparation of competent cells

50 µl aliquots of T7 Express and T7 Express *lysY/I^q* competent cells (NEB) were thawed on ice for 15 min and used to inoculate 100 ml of SOB media. The cultures were incubated for 18 hours (20 °C, 200 rpm) until the ODs were 0.3, then were each split into two 50 ml falcons and centrifuged (3000 rpm, 4 °C, 10 min). The media was removed and cells resuspended in 16 ml ice cold CCMB80 buffer then left on ice for 20 min. Cells were pelleted again (3000 rpm, 4 °C, 10 min) before being resuspended in 1.75 ml ice cold CCMB80 (250 µl increments until OD was 1.0-1.5). Aliquots were then stored at −80 °C.

Transformation protocol

Thawed competent cell aliquots were transformed with assembled plasmids as for the original pET-28a plasmid. 1 µl (100 ng) of each plasmid was added to each 25 µl aliquot of competent cells and heat shocked at 42 °C for 30 s. Incubation with SOC outgrowth medium (37 °C, 180 rpm) was carried out for 1 h. LB-kanamycin plates spread with the incubated cells were then incubated overnight at 37 °C.

Growth of T7 express cultures

Several variations of culture growth conditions were used while optimising protein expression. All media used was LB Lennox with 50 µg/µl kanamycin. Colonies from fresh transformations or glycerols stock strains were picked to inoculate overnight 10 ml cultures (37 °C, 200 rpm). The following morning these were used to seed new cultures (50-500 ml) grown until an OD of 0.6-0.8.

Induction of protein expression

Induction was carried out as suggested in the T7 Express documentation. Once the cultures reached an OD of 0.6-0.8, IPTG was added until a final concentration of 0.4 mM and either incubated at 37 °C for 2 h (initial trials) or overnight at 15 °C (standard protocol). Later iterations of this standard protocol also included 30 min on ice before IPTG induction as a cold shock step.

2.4 Purification of repressor proteins

Small scale – 10-50 ml cultures

Induced 10 ml cultures were centrifuged in 1.5 ml aliquots (5000 xg, 2 min, 20 °C) and pellets were lysed using 400 µl CelLytic B 2x (Sigma) diluted from 10x in 100 mM HEPES-NaOH, pH 7.5. 50 ml cultures were centrifuged in falcon tubes (5000 xg, 10 min, 4 °C) and 10 ml CelLytic B 2x was added per gram of wet cell paste. Samples were vortexed and mixed for 10 min then centrifuged (10 000 xg, 5 min, 20 °C) with the soluble fraction stored at 4 °C before purification.

Small-scale purification was carried out using HIS-Select HF Nickel Affinity Gel (Sigma) according to the manufacturer's instructions. The native equilibration/wash (EQ/W) buffer used was 50 mM sodium phosphate buffer, pH 8.0, 300 mM sodium chloride, 10-20 mM imidazole while the elution buffer was identical except for containing 250 mM imidazole. Flow-through, wash and eluate samples were stored at 4 °C.

Large scale – 500 ml cultures

Induced 500 ml cultures were combined or made up to 1 l with MilliQ water and centrifuged (6000 xg, 10 min, 4 °C). The resulting wet cell paste was weighed and frozen at −20 °C. Cells were resuspended in 20 ml HisTrap binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4) and lysed using a Cell Disruptor (Constant Systems) at 30 kPa. Samples were then centrifuged (40 000 xg, 1 hr, 4 °C) and the soluble fraction stored at 4 °C before purification.

These larger samples were purified using a HisTrap HP 1 ml column (GE Healthcare) according to the manufacturer's instructions. Column purification was carried out at 4 °C using a peristaltic pump with a flow rate of 1 ml/min.

The binding buffer was as described above, while the elution buffer contained 500 mM imidazole. Elution was carried out using 3 ml of elution buffer in a syringe. Flow-through, wash and eluate samples were stored at 4 °C.

2.5 Protein validation and analysis

Buffer exchange and protein storage

Eluted protein samples were dialysed when altering buffer composition. 1-2 ml eluate was loaded into SnakeSkin Dialysis Tubing (3.5K MWCO, ThermoFisher) and placed into 1 l of the new buffer. A magnetic stirrer bar maintained buffer mixing and dialysis was carried out overnight at 4 °C. Dialysis was primarily used to remove imidazole from protein eluates, but also for transfer of samples into protein storage buffers (PSB):

- **PSB1** - 10 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM EDTA, 0.1 mM DTT, 0.01 mg/ml BSA, 5% v/v glycerol
- **PSB2** - 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5% v/v glycerol
- **PSB3** - 10 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM $MgCl_2$, 5% v/v glycerol
- **PSB4** - 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM $MgCl_2$, 2 mM TCEP
- **PSB5** - 10 mM Tris-HCl pH 7.5, 250 mM NaCl, 0.5 mM $MgCl_2$, 1 mM TCEP, 50% v/v glycerol

Occasionally faster buffer exchange was desired, in which case Zeba Spin Desalting Columns were used (10 ml, ThermoFisher). Up to 3 ml of protein sample was loaded per column, according to the manufacturer's instructions.

SDS-PAGE

Protein samples were analysed for distribution of sizes using denaturing gel electrophoresis. 10 µl protein samples were incubated (95 °C, 5 min) with 10 µl 2x Laemmli sample buffer (Bio-Rad) supplemented with 5% v/v β-mercaptoethanol. 15 µl was loaded onto lanes of Mini-PROTEAN TGX

Precast Gels (15-well, 15 µl well capacity, Bio-Rad) with 5 µl of the Amersham ECL Rainbow Low Range Marker (GE Healthcare) as a ladder. Tanks were filled with 1x TGS running buffer and run at 200 V for 30 min. For the western blot a 12% acrylamide gel was cast in MES buffer. Samples were loaded as above and the gel was run at 200 V for 40 min.

Gels were stored in Sterilin plastic containers (ThermoFisher) and stained overnight with 25 ml of Quick Coomassie stain (Generon). Gels were destained with MilliQ water for 1 h and visualised under white light (Azure 600, Azure Biosystems), or on a Light Pad (MiniSun) photographed with an iPhone SE camera for colour images. These were then processed using ImageJ software.

Protein quantification

Protein sample quantification was initially determined using the Pierce BCA Protein Assay Kit (ThermoFisher). As the maximum compatible concentration of imidazole was 50 mM, eluate samples purified by resin were diluted 1/5 with imidazole-free buffer. BSA protein standards were prepared with the same end solution (50 mM sodium phosphate, pH 8.0, 300 mM sodium chloride, 50 mM imidazole). The microplate procedure was used with three replicates of each standard and sample, and absorbance at 562 nm was measured using a FLUOstar Omega plate reader (BMG LABTECH). A standard linear curve was plotted using Microsoft Excel and used to determine sample protein concentration.

Once proteins were stored in buffers containing reducing agents, concentrations were instead determined using absorbance at 280 nm using a NanoDrop spectrophotometer. Absorbance relative to PSB blank was converted to concentration using extinction coefficients estimated for the full fusion protein sequence by Expasy (https://web.expasy.org/protparam/), both for monomer and dimer concentrations.

His tag cleavage

EZCut TEV Protease (BioVision) was used to remove the C-terminal His-tag from the purified proteins. Two units (0.2 µl) were added to 10 µl protein sample and 10 µl cleavage buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM DTT) and incubated for 1 h at 34 °C, before repurification as above this time collecting the unbound flow through.

Western blots

Proteins were transferred from unstained SDS-PAGE gels onto a nitrocellulose membrane in a transfer tank filled with 1x transfer buffer. The tank was cooled to 10 °C and run at 100 V for 1 h. The membrane was then blocked using 5% w/v skimmed milk in PBS-T for 1 h on an orbital shaker and washed three times for 5 min with PBS-T. The membrane was bound by primary antibody (monoclonal anti-polyHistidine produced in mouse, Sigma H1029, diluted 1/1,000 in 1% BSA in PBS) for 2 h before 3x 5 min washes with PBS-T. The membrane was then stained with secondary antibody (700 nm fluorophore-conjugated anti-mouse, diluted 1/5,000 in 1% BSA in PBS-T) for 1 h before another 3x 5 min washes with PBS-T.

The blot was visualised at 700 nm excitation wavelength (Azure 600, Azure Biosystems) and images were processed using ImageJ software.

2.6 Functional assays

Oligonucleotide annealing

Single-stranded oligos were ordered from IDT (except for Alexa488-Fprobe from ThermoFisher) and resuspended in a basic annealing buffer (AB - 10 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, pH 8.0) to a stock concentration of 100 µM. For EMSAs, 10 µM working dilutions were prepared and aliquots combined with complementary oligos for annealing to produce dsDNA at a 5 µM concentration. Probe components were prepared at 4/4.8/5.76/6.912 µM (lowQ) or 4/6/12/20 µM (highQ) concentrations (4/6/12 µM for Lprobes), to be combined for 1 µM probes.

Complexes were then annealed in a thermocycler with the following protocol: 95 °C for 5 min, temperature reduced at 1 °C/min until 25 °C reached, 5 min at 25 °C then hold at 10 °C. DNA was kept at 4 °C for short-term use and −20 °C for long-term storage.

Electrophoretic mobility shift assays (EMSAs)

Many EMSA iterations were attempted while optimising binding conditions. 2-3 µl dsDNA (200-300 ng for most oligos) was incubated with purified protein

samples (up to 10 µl) for 30 min at room temperature (20 °C). 1 µl loading dye (6x, ThermoFisher) was added to each tube to aid visualisation and samples were loaded onto gels. Initial gels were 5% agarose gels in 1x TBE or 6% acrylamide pre-cast DNA Retardation gels (Tnvitrogen, 0.5x TBE), while later gels were 20% acrylamide pre-cast Novex gels (Invitrogen) or self-made 20% acrylamide (29:1 acrylamide to bisacrylamide) TBE gels. The pre-cast gels were run at 100 V in 0.5x TBE buffer while the self-made gels were run at 200 V in 1x TBE buffer, both for 1 h.

Gels were then stained in 1x SYBR Safe (Invitrogen) for 30 min on a platform rocker, before visualisation under EpiBlue light (Azure 600, Azure Biosystems). Images were then processed using the ImageJ software.

2.7 Fluorescence time course assays

In Chapters 3 and 4 fluorescent DSD assays were highly individual so preparation details are specified in the figure caption for each. Following the methodology developments in Chapter 5, seven standardised protocols were established, outlined in Figure 2.1. The standard formulations included a 10x excess of CgArsR over dsDNA probe operators (2 µl CgArsR/PSB stock per 15 wells); if more was desired this volume was proportionally increase and compensated with less buffer.

Except for DSD6, where the reversible X oligo was include in the master mix (MM), X oligos were first loaded into the bottom of each well with a 100 µl combitip. In DSD1-3, 3 µl of 2/4/8 µM X oligo stock was added for 2/4/8x excesses over probe G strands, with 5 µl of 1.2/2.4/4.8 µM X added in DSD4-7 to load the same amount but allowing 45/195 µl MM to be added with a combitip. Control oligos were added at the 8x concentration for each.

Assays were then triggered by the addition of the specified volumes of each MM row by row. DSD1-5 were read immediately after MM addition, with 0/10/100 µM sodium arsenite was added to triplicate wells 30 min after the last master mix had been added, while for a total volume of 250 µl per well. For DSD7 MMs were added to wells and left to incubate for 1 h before arsenite addition or lyophilisation.

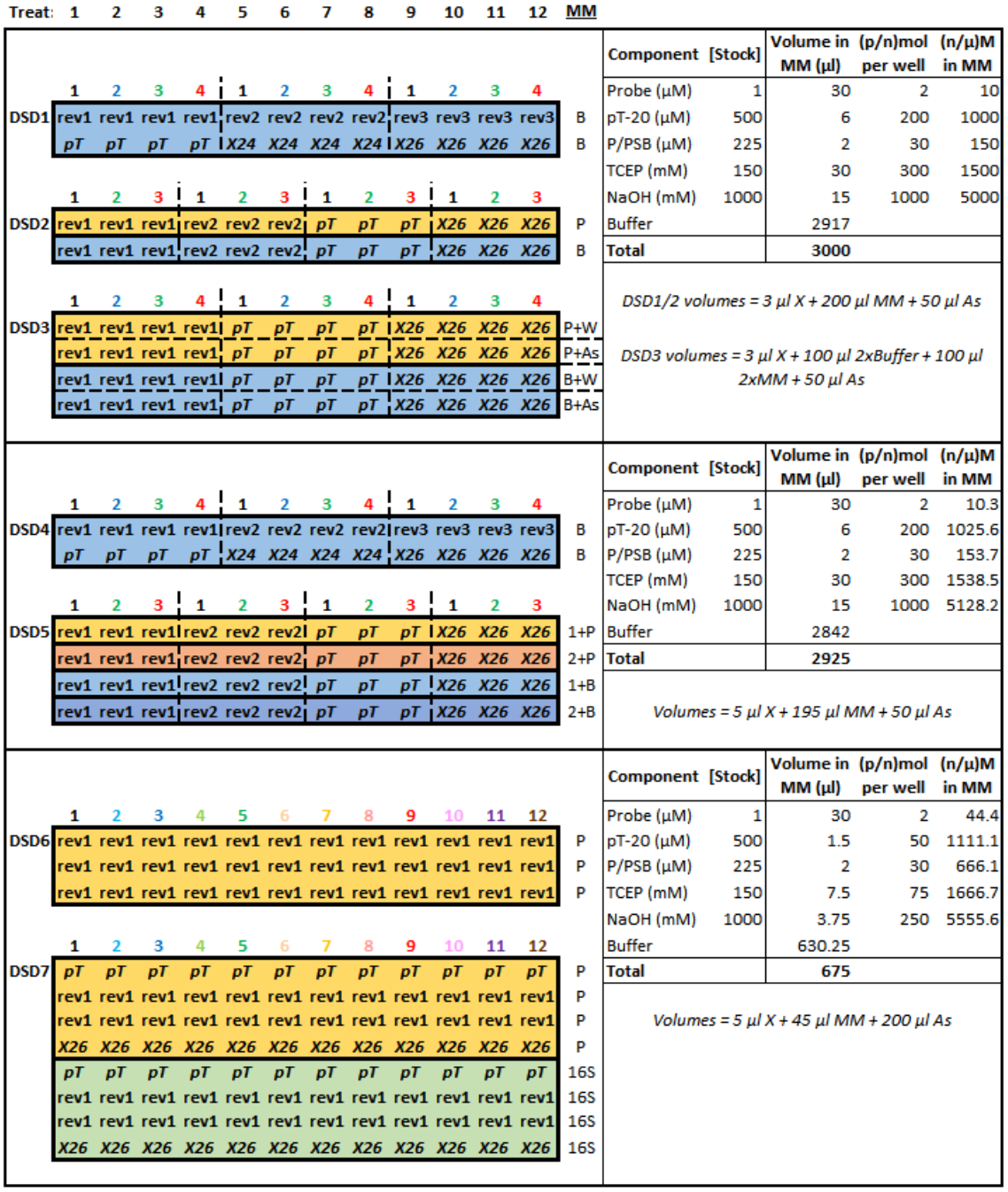


Figure 2.1: **DSD protocol layouts and master mix compositions.** Illustration of the seven standardised DSD protocols developed in Chapters 5 and 6. Component volumes are provided per 15-well MM, made up for each 12-well row with a 3-well excess. The units for each MM species depend on whether the stock was originally in µM or mM concentrations. TCEP was prepared as frozen aliquots of TCEP.HCl equimolar balanced with NaOH while probe concentration refers to fluorophore content, with other strands in excess. For the layouts, row colours specify different MMs, with protein (P) rows placed higher than buffer (B) rows to capture earlier dynamics in DSD2/3/5. DSD5 contained 2-4 rows, in the example containing two probes +/- ArsR, but elsewhere with multiple ArsR concentrations. Columns 1-12 specify different treatments, usually repeating sets of arsenite concentrations as indicated by the coloured numbers. Replicates for DSD6/7 were by row instead to ensure they derived from the same storage wells. DSD3 compared non-As treatments so water/arsenite was added by row too and MMs were double concentrated to enable variable buffer addition.

When preparing master mixes, NaOH and TCEP added first, followed by half buffer volume then pT-20, probe, protein/PSB then the remaining buffer. Tubes were inverted twice then gently vortexed briefly before incubation for 1 h at room temperature in covered LoBind tubes (Eppendorf). Master mix volumes under 2 ml were prepared in 2 ml tubes, those from 2-6 ml in 5 ml falcons, and greater than 6 ml in 15 ml falcons to minimise excess sticking to tube walls as well as ensure combitip access. MMs were recycled in combitips three times before addition to X oligos to limit differential sequestration.

Plates were purchased from Greiner Bio-One, initially 384-well F-bottom, HiBase non-binding (784900), later 96-well F-bottom, non-binding (655900). Assays were read on a FLUOstar Omega reader (BMG LABTECH), with 485-10 nm excitation and 520-10 nm emission filters, 20-flash orbital averaging and a gain of 2200 at 25 °C. Data was exported to .csv files for processing and plotting in R.

Displacement buffers

The following displacement buffers were used to provide different reaction conditions for the DSD assays:

- **DB1** - 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$
- **DB2** - 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 20 mM $MgCl_2$
- **DB3** - 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$
- **DB4** - 5 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$
- **DB5** - 100 mM NaCl, 10 mM $MgCl_2$
- **DB6** - 10 mM Tris-HCl pH 8.6, 50 mM NaCl, 5 mM $MgCl_2$
- **DB7** - 10 mM Tris-HCl pH 8.6, 50 mM NaCl, 5 mM $MgCl_2$, (2.6/5.3/10.6)% w/v PEG8k
- **DB8** - 100 mM Tris-HCl pH 8.6, 200 mM NaCl, 20 mM $MgCl_2$, 11.9% w/v PEG8k

2.8 Visual DSD modelling

DSD reaction models were created in Visual DSD using their online interface: https://classicdsd.azurewebsites.net/ Code was written as in Figure 5.11, with alternatives and syntax described in Section 5.6.

Further details can be found in the user manual: https://www.microsoft.com/en-us/research/project/programming-dna-circuits/

Inference was typically run for 10,000 iterations, half with randomly varied burn-in parameter sets and half to assess posterior distributions. More iterations were used if parameters were not judged to have converged, displaying either non-normal posterior distributions or notable increase in LogLikelihood over time.

Chapter 3

Initial repressor characterisation

3.1 Introduction

Following proposal of a repressor-DSD *in vitro* biosensor design, suitable components needed to be identified. This chapter outlines how these components were chosen, how their repressors were cloned and expressed, and how a subset were subsequently purified. Two arsenic repressors, from *Bacillus subtilis* and *Corynebacterium glutamicum*, were eventually selected for *in vitro* characterisation, where their interactions with operator oligonucleotides and potential ligands were analysed.

Further experiments tested their suitability for integration with DSD reactions and the *Bacillus subtilis* protein, despite some promising early results, ultimately proved too problematic to use long-term. The *Corynebacterium glutamicum* repressor, however, was similarly effective but much more stable, and so was selected to be utilised for the remainder of this work.

3.2 Selection of nine repressor proteins

An extensive literature search was conducted in order to identify suitable ArsR proteins. Hundreds of bacterial ArsRs have been identified or predicted within sequenced genomes, but relatively few have been studied in great detail. Previously unstudied repressors may be particularly sensitive or selective to

arsenic, but such discoveries would be laborious without guaranteed success. Therefore in order to develop a working repressor-DSD biosensor as soon as possible, the best characterised ArsRs were shortlisted for use. Several pieces of evidence were used to narrow down appropriate transcription factors likely to be expressed and purified easily, as well as behave as desired *in vitro* - these included:

- Proof of repressor expression *in vivo* in response to arsenic
- Previous repressor purification, ideally from recombinant plasmid expression
- Identification of operator sequence and level of evidence
- Demonstration of operator binding *in vitro* and size of fragment
- Demonstration of dissociation induced by arsenic *in vitro*
- Evidence of selectivity in favour of arsenic over other heavy metals or metalloids
- Possibility of binding to multiple arsenic species

Eventually eight arsenic repressors were identified for recombinant expression and purification. These ArsR proteins originated from the *Escherichia coli* chromosome (EcCArsR), the *Escherichia coli* R773 plasmid homolog (EcRArsR), the *Bacillus subtilis* skin element (BsArsR), the *Staphylococcus xylosus* pSX267 plasmid (SxArsR), the *Acidithiobacillus ferrooxidans* chromosome (AfArsR), the *Pseudomonas stutzeri* chromosome (PsArsR), the *Cupriavidus medallidurans* chromosome (CmArsR) and the *Corynebacterium glutamicum* chromosomal ArsR1 (CgArsR). This set of repressor proteins was chosen due to the likelihood of encountering expression difficulties with some of them, while the selections were made to cover the best-studied proteins as well as a range of potentially variable properties. It was expected that different ArsR proteins would vary in arsenic sensitivity and selectivity, stability in different solutions, and binding dynamics. Their respective operator sequences may also vary in suitability to DSD applications, and discrimination of oligonucleotides (oligos) may also be variable. These operator sequences were identified to variable degrees of certainty among the different repressors, from hypothetical presence within promoter regions hundreds of bases long to detailed footprinting assays. Table 3.1 summarises the available evidence for the chosen arsenic repressors.

Organism	ArsR Class	Induction *in vivo*	Purification method	Operator sequence evidence	DNA probe used	ArsR-operator kD	Arsenic-binding evidence	Selectivity	Mr /kDa	References
Escherichia coli chromosome (EcCArsR)	1	Yes	Ion-exchange and His-tagged (both termini)	DNase I footprinting	Radiolabelled 153bp and 208bp probes	1 µM (4x)	EMSA only	Dissociates with arsenite, antimonite and PAO; not arsenate	13.3	[113–118]
Escherichia coli R773 (EcRArsR)	1	Yes	Ion-exchange	DNase I and hydroxyl footprinting	Radiolabelled 153bp probe	0.33 µM (1.3x)	EMSA only	Dissociates with arsenite, antimonite and bismuth; not arsenate	13.2	[119–123]
Bacillus subtilis skin element (BsArsR)	1	Arsenite, arsenate, antimonite	N/A	Hypothesised inverted repeat only	N/A	N/A	N/A	N/A	12.3	[124]
Staphylococcus xylosus pSX267 (SxArsR)	1	Arsenite, arsenate, antimonite	Ion-exchange	DNase I and hydroxyl footprinting	Radiolabelled 75bp, 186bp and 312bp probes	0.67 µM (4.5x)	EMSA only	Dissosiates with arsenite only; not arsenate or antimonite	11.9	[125, 126]
Pseudomonas stutzeri (PsArsR)	1/3	Yes	Ion-exchange	N/A	N/A	N/A	N/A	N/A	13.5	[127]
Corynebacterium glutamicum ArsR1 (CgArsR)	2	Arsenite and antimonite	C-terminal His-tag	DNase I footprinting + fluorescent anisotrophy	Fluorescein-labelled 30bp probes	0.15 µM (5x)	Fluorescent anisotrophy and EMSA	Dissociation greater with antimony than arsenite (opposite *in vivo*). May bind at dimer interface	13.0	[128–130]
Cupriavidus metallidurans (CmArsR)	3	Arsenite, arsenate, bismuth, cadmium	N-terminal His-tag (affected binding), C-terminal intein affinity tag	EMSA only	Biotin-labelled EtBr-stained 142bp and 58bp probes	N/A	EMSA only	Dissociates with arsenite, bismuth, cobalt, copper, nickel, cadmium, lead and zinc; not arsenate	11.5	[131]
Acidithiobacillus ferrooxidans (AfArsR)	3	Arsenite, arsenate, antimonite	C-terminal His-tag	DNase I footprinting + fluorescent anisotrophy	199bp probe; post-stained with SYBR Green	0.22 µM (4.4x)	MS + X-ray absorption spectroscopy	Affinity 6x greater for antimonite than arsenite. May bind at dimer interface	13.0	[110, 132, 133]

Table 3.1: **Arsenic repressor proteins chosen for expression.** Comparison of the selected ArsRs, their claimed properties and the experimental methods used to purify and analyse them. Dissociation constants are listed as the concentration of ArsR required to bind half of the total operator DNA, alongside the relative excess of total ArsR over total operator DNA available

The first and best studied arsenic repressors belong to the *E. coli* chromosomal operon and the *E. coli* R773 plasmid [113, 119]. While these have slightly different protein and target operator sequences, they have been claimed to be able to bind each other's promoter and repress each other's transcription. However, the likely difference in affinities and the extensive *in vitro* work made both of these proteins worth investigating. The R773 plasmid is known to confer greater arsenic resistance, however this may be due to the other effector proteins contained within the operon and higher copy number of the plasmid compared to the chromosome, rather than differences in the repressor-operator complex (ROC) itself [116]. The demonstrated footprints of these proteins are unusually imperfect inverted repeat sequences, which is a potentially useful feature in DSD applications.

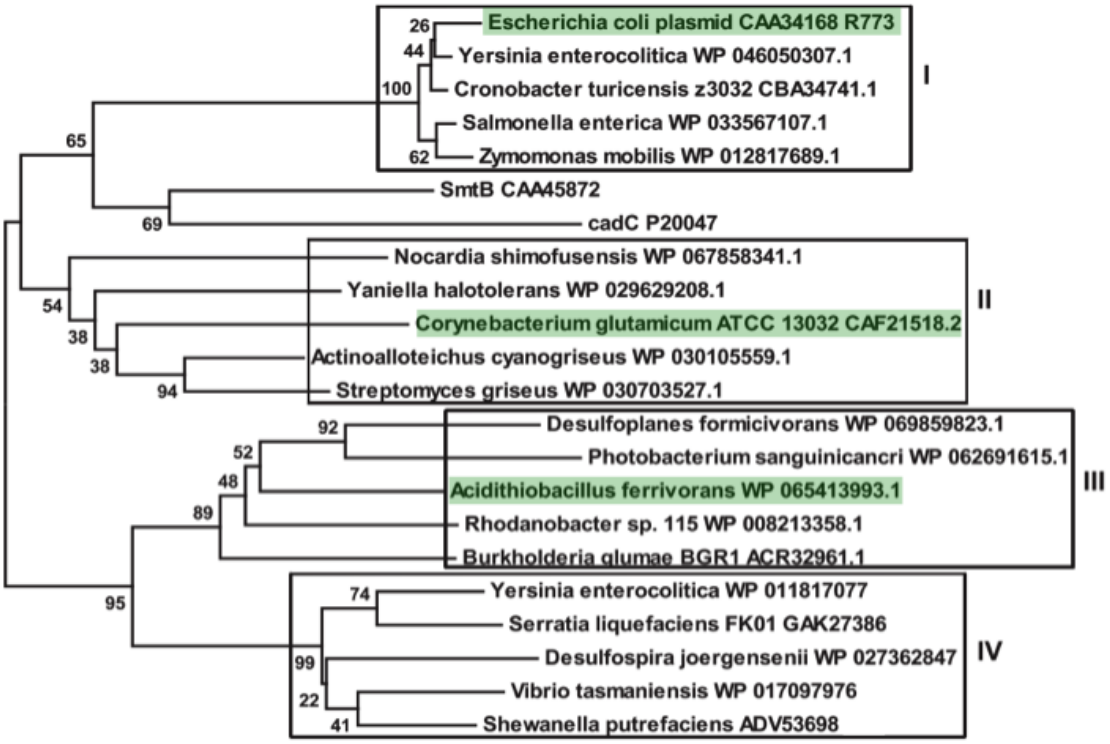


Figure 3.1: **Classification of ArsR proteins.** Chen *et al.* [134] used a neighbour-joining method to group a sample of arsenic repressor proteins into four classes. Three of these classes are represented by proteins selected for expression (highlighted in green), while the fourth class contains only poorly studied proteins or those shown to bind organic arsenic species.

B. subtilis lacks this *in vitro* evidence, but has strong resistance to arsenic and induced expression of ArsR, a hypothesised inverted repeat operator and high natural selectivity for arsenite [Yarkoni, pers. comm.]. The *S. xylosus* and *S. aureus* plasmid *ars* operons are both well studied and share high similarity, however the SxArsR protein has been investigated more thoroughly *in vitro*.

According to Chen *et al.*, however, all of these repressors belong to the same first class of ArsR proteins [134]. Figure 3.1 shows the classification of some identified ArsR protein sequences into four groups.

Each of the class I ArsR polypeptides contain an arsenic-binding domain dominated by a 'CVCDLC' box, or a sequence with high homology to this, as shown in Figure 3.2. The three cysteine residues have been demonstrated to be crucial for response to arsenic, triangulating one arsenite ion within this structure through each thiol group interacting with one of the three oxygen atoms (or hydroxyl groups) from the arsenite species, with at least two of these cysteines necessary for response to arsenic [123]. The arsenic-binding domain is closely followed by a helix-turn-helix DNA-binding domain roughly 22 amino acids long. Two ArsR polypeptides, each roughly 100-120 amino acids long, have been shown to combine to form a homodimeric functional protein with two arsenic-binding domains and two DNA-binding domains, hence recognition of an inverted repeat (or near-repeat) operator sequence within the promoter of the *ars* operon. It is unclear whether one or two bound arsenite ions are required to cause repressor dissociation from its operator sequence, or whether these class I ArsRs exhibit cooperative behaviour [117].

The remainder of the ArsR proteins were chosen to diversify the structures of the repressors selected. *C. glutamicum* was placed in the second class of arsenic repressors by Chen *et al.* - while its DNA-binding domain is related to that of the class I proteins, its arsenic-binding domain is noticeably different. Instead of a 'CVCDLC' box, a 'CC' motif near the N-terminus of the protein is combined with another cysteine residue around the middle of the protein, as shown in Figure 3.2. Homology modelling to the *Staphylococcus aureus* cadmium repressor (CadC) indicates arsenic binds at the interface of the CgArsR homodimer, stabilised by the 'CC' of one monomer and the third cysteine of the other. This has the potential to give CgArsR radically different dynamics and sensitivity in response to arsenic compared to class I repressors, as a single arsenite ion may destabilise the protein more than the class I orthologs, and perhaps CgArsR would also have a different selectivity profile [130]. Ordóñez *et al.* also found *C. glutamicum* to contain two functional chromosomal *ars* operons each with a separate repressor, sharing 66% sequence identity. These were shown to have very similar biochemical properties and so most experiments were subsequently performed using their CgArsR1 protein - this was therefore the paralog chosen for this work too, and will be referred to as simply CgArsR.

```
EcCArsR------------------------MSFLLPIQLFKILADETRLGIV-LL-LSELGELCVCDLCTALDQSQPKISRHLALLRESGLLLDRKQGKWVHYRLSPHIPAWAAKIIDE-----AWRCEQEKVQAIVRNLARQNCSGDSKNICS    Class 1    117aa
EcRArsR------------------------MLQLTPLQLFKNLSDETRLGIV-LL-LREMGELCVCDLCMALDQSQPKISRHLAMLRESGILLDRKQGKWVHYRLSPHIPSWAAQIIEQ-----AWLSQQDDVQVIARKLASVNCSGSSKAVCI    Class 1    117aa
BsArsR-------------------MDETKSELLRKYEQKFKALADQKRLEIM-YE-LCQRGKTCVCDLTEIFEVTQSKLSYHLKILLDANLITKETKGTWSYYDLNDEEVN---GLLSEELCC------------IFRKKGEGDCC--------    Class 1    105aa
SxArsR-------------------------MSYKELSTILKVLSDPSRLEILD---LLSCGELCACDLLEHFQFSQPTLSHHMKSLVDNELVTTRKNGNKHMYQLNHEFLDYINQNLDIIN--------TSDQGCACKNMKSGEC---------    Class 1    104aa
CgArsR-MTTLHTIQLANPTECCTLATGPLSSDESEHYADLFKVLGDPVRLRILSQLAAGGCGPVSVNELTDLMGLSQPTISHHLKKMTEAGFLDRVPEGRVVLHRVRPELFAELRTVLQIGSME----------------L--------------    Class 2    119aa
PsArsR-------------------------MSFLLLHGIVHQFLL-MKPGWASFY-CWANWESYPSAISALLDQSQPKKSRHLALLRESGLLLDRKKRYVGSL---PLITAYS-SIFGENYCCGLAMCHRKRFRRLSAHLARQNCSGDSKNICS    Class 1/3  118aa
CmArsR-------------------------METENALEALAAALAHGIRLAVFRLLVQAGPEGLPAGRIAELMEMPASSLSFHLKELHRAGLLASRQEGRSIIYMAQFETMNALLGYLTENCCGGAPCSPVSSCSV-ATES--------------    Class 3    109aa
AfArsR-------------------MEP--LQDPAQIVARLEALASPVRLEIFRLLVEQEPTGLVSGDIAEHLGQPHNGISFHLKNLQHAGLVTVQREGRYQRYRAAMPVVRALVAYLTENCCHGTRDCALSGETR-SPSVQEGNQ---------    Class 3    118aa
Conserved residues                      . :    :                 :   :      * *:  :    ::    :                 :
```

Figure 3.2: **Alignment of the chosen ArsR protein sequences.** A full ClustalO alignment of the eight selected ArsR proteins, displaying the conservation of certain residues underneath ('*' = fully conserved residue; ':' = strong residue group conservation; '.' = weak residue group conservation). The known/hypothesised arsenic-binding and DNA-binding domains have been manually highlighted in red and green, respectively. The DNA-binding domains of CmArsR and AfArsR are likely to also align with the other six. The classification of each ArsR according to Chen *et al.* and the length of each protein are also displayed.

```
EcCArsR 5'------GCACTTACACATTCGTTAAGTCATATATGTTTTTGACTTATCCGCTTCGAAGAGAGACACTACCTGCAAC----------------3'
EcRArsR 5'---GTGATTAATCATATGCGTTTTTGGTTATGTGTTGTTTGACTTAATATCAGAGCCGAGAGATACTTGTTTT--------------------3'
BsArsR  5'-----TTACAATTAATCAAAATAAATTGATTTATTTGCTTGCATTATTTTAAAAATCATGAGTATAATAAATACA----------------3'
SxArsR  5'-----TATATAGAATTAAAAACAGTAGATAAAAGAGAATAGACACAAATCTATATAGATGTTAATCTATTAACTG----------------3'
CgArsR  5'-----CATAAATCCACTATATATTGACGAATGTCGATATTGAAAGTATTTTGAATATCGACAGGTATCAATATACCGAAAGGTGTCGCATG-3'
PsArsR  5'---------GGGTTTGTACTGCGACACATGGAAATCCGTTGACATTTTTGATCCGTACGCTTCGAATAATCATAGTATG-------------3'
CmArsR  5'----------------AGCGACACTGCCAATCTTCCATTGCACGGTTTCCATGATTCGACTATATTCGAATCATGGAAACCGAAAA-----3'
AfArsR  5'-------ACAGAGCCAAGTCTTGACGAGCGCCACGGTCTTGTCAATACTGCAAGAAATATTCGTGGATAATACGATG---------------3'
EcTetR  5'-TTTCAGATCCTGTGTTAATTCCCTATCATTGATAGAGTTGGAATATATCCCTATCAATGATAGAGTGTCA--------------------3'
```

Figure 3.3: **Comparison of the promoter sequences for each of the chosen repressor proteins.** Promoter sequences have been manually aligned by their identified/hypothesised -35 box (underlined in middle). Any identified -10 boxes are also underlined further to the 3' end. Red highlighting identifies regions protected by DNase I footprinting, while blue colouring is for more stringent contacts identified by hydroxyl radical footprinting (or crystallography for EcTetR). Green highlighting is for inverted repeat (or nearly symmetrical) regions hypothesised to be operator sequences - some of these are obscured where actual footprinting has been carried out. Where transcription start sites have been identified these have been highlighted in orange, otherwise presumed start codons have been coloured purple.

AfArsR was placed in the third class of ArsR proteins. Its arsenic-binding domain containing the three cysteine residues is towards the C-terminus of the protein in a 'CC-XXXXX-C' conformation [133]. CmArsR has a similar structure, with four amino acids separating the three cysteine residues at the C-terminus instead of five. *C. metallidurans* is also well known for its adaptation to a variety of metal stresses, but its response selectivity is questionable [131]. *A. ferrooxidans* is another strain found in mineral deposits, and as a result is well adapted to detoxifying and removing metal ions [132]. *P. stutzeri* was the last strain chosen, as it does not seem to fall into one specific ArsR category. Its arsenic-binding domain is clearly similar to those in class 3, however its DNA-binding domain more closely resembles those in class 1 and 2, and a phylogenetic analysis placed its *ars* operon more closely related to the *E. coli* chromosomal operon than even the *E. coli* R773 plasmid operon, despite this difference in arsenic-binding domain [127]. While this has interesting implications for the evolutionary relationships between these operons, Patel *et al.* also suggested these C-terminal arsenic-binding domains may allow response to arsenate ions, instead of only arsenite with the class I proteins.

Most of the remaining studied ArsRs were other *E. coli* plasmid operons or similar class 1 proteins. One particular ArsR not chosen for investigation belonged to the *Acidiphilum multivorum* pKW301 plasmid (AmArsR). It has high similarity to the *E. coli* repressors in the first 33 amino acids, but the rest of the protein is very divergent and much shorter at 84 amino acids total [135]. The result is that AmArsR contains a very different DNA-binding domain, and a footprinting assay demonstrated it bound a very different operator sequence to the *E. coli* repressors. The repressor coding sequence reveals divergence from *E. coli* ancestors through a single base insertional mutation – removing a single adenosine nucleotide from the *A. multivorum* sequence causes a frame shift to produce a 116-amino acid long protein with 86% identity to the EcRArsR sequence, and crucially reverts to having an identical DNA-binding domain to the R773 repressor [136]. This does leave the possibility that a sequencing error is responsible for this otherwise interesting repressor, and it is in fact a classic class I ArsR, and so the *in vitro* evidence alone was not sufficiently strong to warrant an investigation.

Another notable exclusion was the *Shewanella putrefaciens* ArsR (SpArsR). This was the subject of work by Chen *et al.* and was placed in the elusive fourth class of arsenic repressor, in part due to its seemingly unique structure of arsenic-binding domain, with C-terminal cysteines likely involved in

coordination of arsenic species. Induction *in vivo* suggested that SpArsR responded much stronger to methylarsenite than arsenite itself and that the loss of an additional cysteine residue was responsible for this. While the prospect of a repressor more selective for organo-arsenicals was very enticing, once again the lack of *in vitro* assays and operator identification put this ArsR outside the scope of this work.

The ninth protein included for recombinant expression was the *E. coli* tetracycline repressor (EcTetR), intended to be used as a potential control system. This transcription factor has been studied much more extensively than any of the ArsRs, with extensive literature on purification methods, crystal structure and operator sequence [137–143]. EcTetR would serve as a useful test system for *in vitro* assays and DSD integration if the arsenic repressors proved difficult to handle, as well as serving as an expression control, being roughly twice the size of the ArsRs.

Looking towards making a functional arsenic biosensor, one potential design would be to produce a ratiometric output since any *in vitro* system based on binding and unbinding would be affected by physical parameters such as temperature, pH and salt concentrations. Having a parallel system similarly affected by these variables but unresponsive to arsenic would serve as a useful control. An arsenic-insensitive mutant of the main ArsR protein would have the most similar kinetics, however this would still bind the same operator sequence - so the ideal internal control (within a single reaction solution) would need to have similar kinetics but bind a different sequence. An alternative repressor such as TetR still responsive to a different substance may also be a useful point of calibration or system testing. An analysis by Rosinski *et al.* [144] showed the tetracycline repressor is the second closest related helix-turn-helix repressor to the ArsR family, except for the lysine repressors, but the LysR family can function as homotetramers so are likely to have different dynamics. TetR binds its operator exclusively as a homodimer, so its binding dynamics should be more similar ArsR. While other members of the ArsR family of repressors (e.g. CadC) may be most closely related, drinking water contaminated with arsenic may well contain noticeable levels of cadmium, zinc or other ions to which these respond, whereas the presence of tetracycline would be less likely as it is only naturally synthesised by soil bacteria [145].

3.3 Successful expression and purification of *B. subtilis* and *C. glutamicum* ArsRs

The nine selected repressor genes were designed into C-terminal His-tag fusion gBlocks, incorporating a TEV protease cleavage sequence between repressor and tag, and codon-optimised for expression in *E. coli* as specified in the Materials and Methods. Following golden gate cloning into a pET-28a vector, eight out of the nine plasmids sequenced as expected, with only pET-28a-AfArsR failing due to the presence of a strong C-terminal hairpin between His-tag and poly-glycine spacer following unfortunate codon-optimisation.

The remaining eight constructs underwent initial expression in small *E. coli* cultures and purification trials using nickel-conjugated resin. Six of these are shown in Figure 3.4, with BsArsR, CgArsR, CmArsR and EcTetR expressing proteins of their expected masses (14.3, 15.0, 13.5 and 26.4 kDa respectively), but the latter two only weakly. Figure A.1 shows a similar SDS-PAGE gel for EcCArsR and SxArsR, neither of which expressed. The use of the CelLytic lysis reagent resulted in total soluble protein (TSP) samples with high salt content, which distorted bands in the low molecular weight regions of these protein gels.

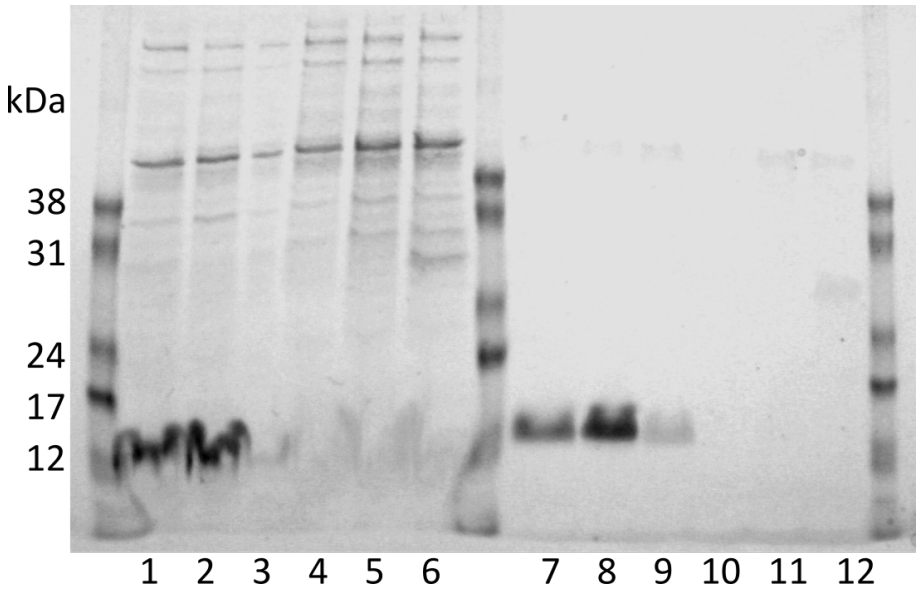


Figure 3.4: **Expression attempt of six repressor proteins.** Coomassie-stained SDS-PAGE gel of IPTG-induced 10 ml cultures, with lanes 1-6 containing the total soluble protein from the induced cultures, while lanes 7-12 contained the batch-purified eluate fractions. (1,7) BsArsR; (2,8) CgArsR; (3,9) CmArsR; (4,10) EcRArsR; (5,11) PsArsR; (6,12) EcTetR

The four successfully expressed proteins were further investigated at this batch purification stage, to determine whether washing the samples without imidazole improved purity without significantly reducing yield. As shown in Figure 3.5, washing with 10 mM imidazole did not noticeably reduce yield but aided sample purity, particularly removing contaminant high molecular weight species from CmArsR and EcTetR samples, so subsequent washing stages always included a low concentration of imidazole.

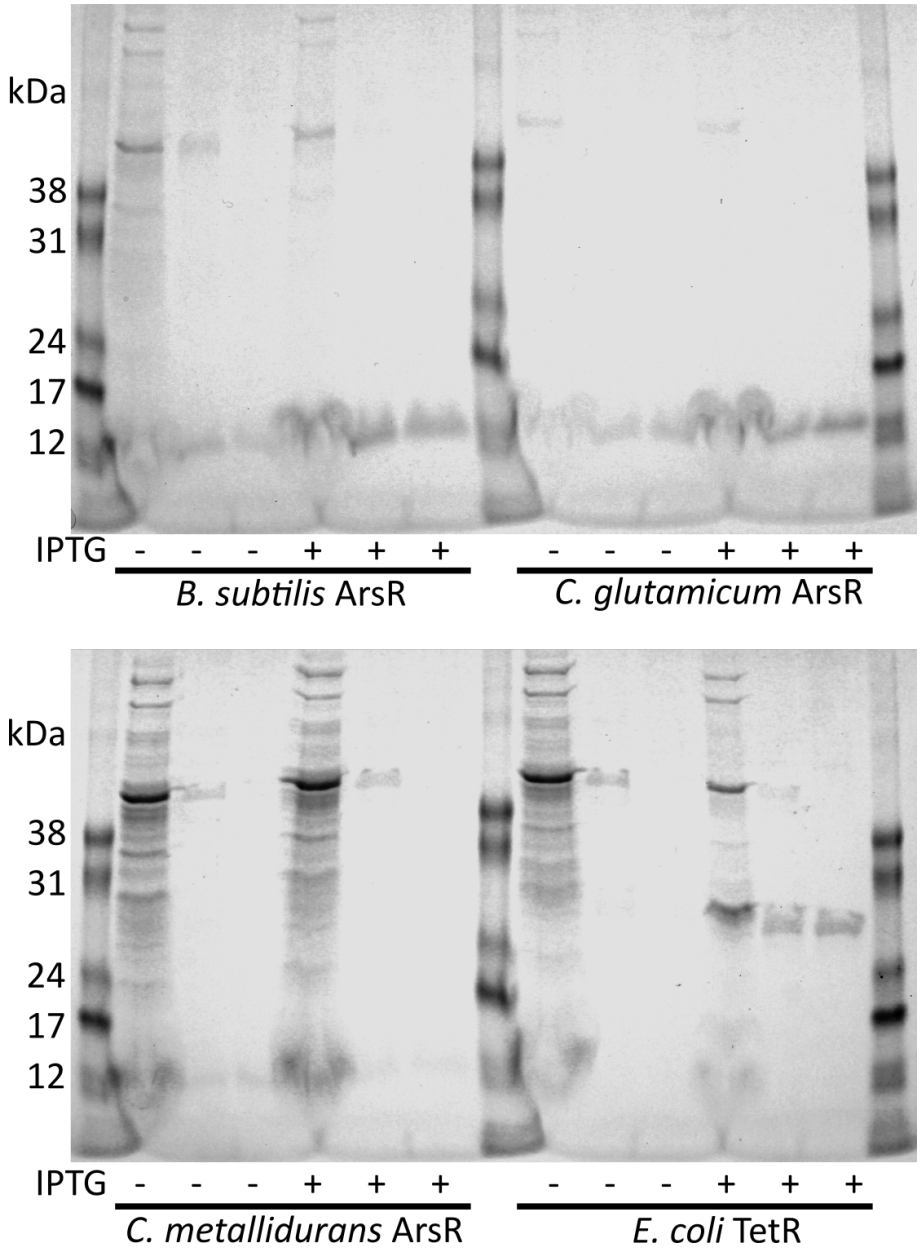


Figure 3.5: **Improving purification of the four expressing repressor proteins.** Coomassie-stained SDS-PAGE gel of uninduced and IPTG-induced 10 ml cultures, comparing their total soluble protein (left), the purified protein washed with no imidazole (middle), and the purified protein washed with 10 mM imidazole (right).

This second round of expression and purification did reveal, however, that the *C. metallidurans* ArsR was only very weakly expressing. Figure 3.4 suggested that CmArsR was expressing moderately, but in hindsight this may have been due to protein spill-over from the adjacent *C. glutamicum* lane. The independent subsequent expression shown in Figure 3.5 revealed only a very faint band for CmArsR, suggesting only very low levels of expression for this proteins and therefore expression attempts on this construct were not continued.

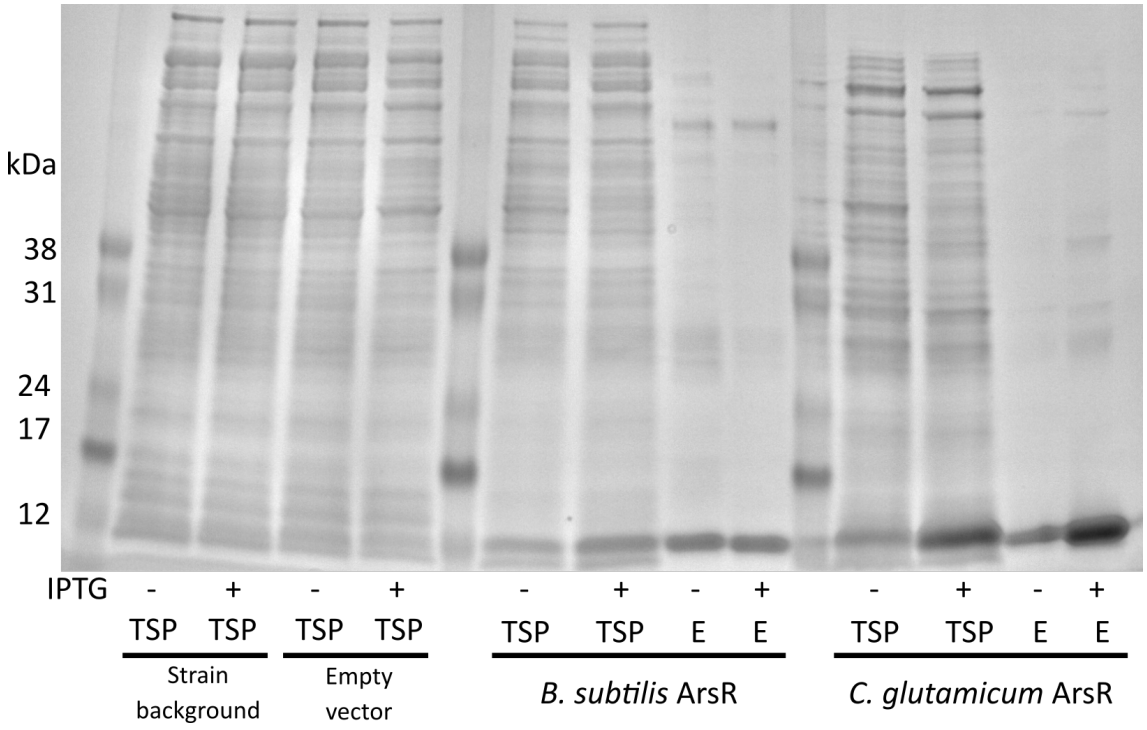


Figure 3.6: **Large scale expression and purification of BsArsR and CgArsR.** Coomassie-stained SDS-PAGE gel of uninduced and IPTG-induced 500 ml cultures, comparing the strain background (T7 Express *E.coli*), the same strain transformed with the original uncloned expression vector (pET-28a), and those containing plasmids cloned to express BsArsR or CgArsR. The total soluble protein (TSP) of each of these is also compared to the nickel column-purified eluates (E) of the ArsR-expressing strains.

The remaining three repressors were then taken for larger scale expression trials, using 500 ml cultures, a non-chemical cell lysis method and purification using a HisTrap nickel column. Unfortunately this new method reduced the yield of EcTetR production, as seen in Figure A.2, but expression and purification of BsArsR and CgArsR improved. Figure 3.6 shows how the expression of these two proteins compared to their control strains. Neither the vector-free strain (T7 Express *E.coli*) or the same strain transformed with the original uncloned pET-28a expression vector synthesised high quantities of a protein in the 14-15 kDa mass range, either with or without IPTG-induction.

The strains transformed with the pET-28a-BsArsR and pET-28a-CgArsR plasmids, however, did produce large quantities of protein within this range, strongly suggesting our repressors of interest had been correctly synthesised. Interestingly, BsArsR expression seemed similar with or without IPTG-induction of its cultures, suggesting weak control of T7 RNA polymerase expression by its *lac* repressor. CgArsR also displayed appreciable background expression, but this did increase upon IPTG stimulation. Later rounds of expression made use of the T7 Express *lysY/I^q E.coli* strain to increase control of background expression and thereby potentially improving protein yield.

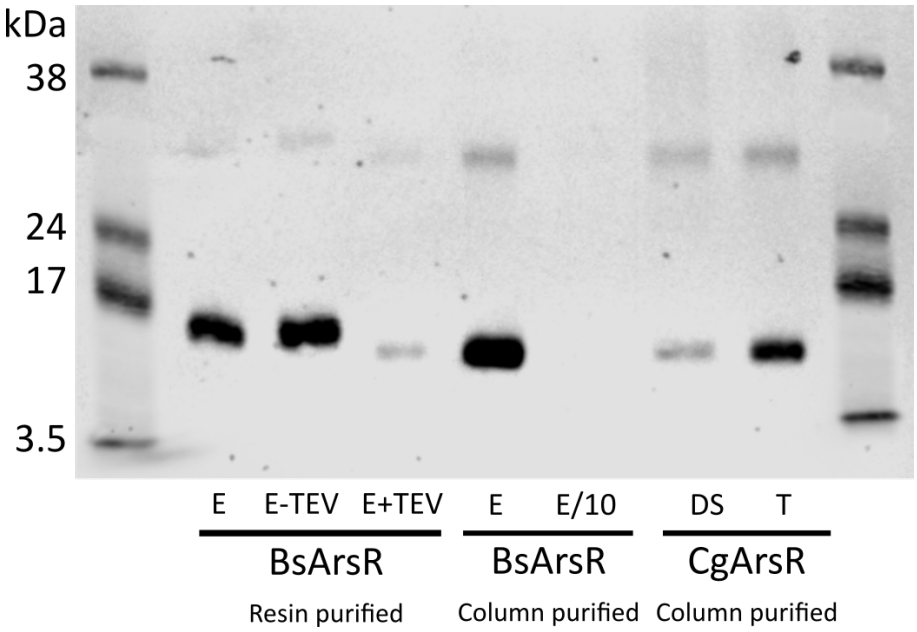


Figure 3.7: **Western blot of BsArsR and CgArsR proteins.** SDS-PAGE gel of BsArsR and CgArsR samples, blotted by an anti-His tag primary antibody and visualised by a fluorescent secondary antibody. Samples were either raw eluate (E), eluate diluted 1/10 with elution buffer (E/10), desalted eluate (DS) or eluate buffer exchanged into a Tris-based buffer (T). For the resin-purified BsArsR, the samples were either raw eluate (E), eluate in TEV buffer but without protease (E-TEV) or eluate in TEV buffer with TEV protease (E+TEV).

Another control for correct expression was a Western blot against the purified proteins using an anti-His antibody. All the repressors were constructed as His-tag fusion proteins, with the histidine residues separated from the C-terminus of the repressor by a TEV protease cleavage site. Figure 3.7 shows the successful binding of this antibody to two purifications of the BsArsR protein (small scale using resin and large scale using a column), as well as the CgArsR protein. TEV protease was also able to successfully cleave off the C-terminal His-tag from most of the BsArsR resin sample, with the anti-His antibody subsequently unable to bind to the trimmed protein. This is further proof that the expression products

were those intended by the cloning design - not only were the proteins the correct mass, but also contained an exposed 6xHis group that could be separated from the rest of the protein by cleavage of a TEV protease site as expected.

In a number of the repressor purifications (e.g. Figure 3.6) a faint band can be seen in the sample eluates around the 30 kDa mass. These bands are seen more clearly in Figure 3.7, indicating these proteins are also His-tagged species. Given the size of these products, and the propensity of these repressors to form homodimers, it seems likely these bands do in fact correspond to such dimers. Other possibilities include run-through translation products, but this would be very unlikely given the multiple stop codons within the terminator sequence following the His-tag. The unusual feature of such dimers, though, would be their ability to resist the denaturing conditions within an SDS-PAGE gel. The β-mercaptoethanol in these samples should reduce most internal disulfide bonds, but it is possible some can resist this treatment. Such an occurrence would not be problematic unless it reduced repressor functionality.

3.4 Characterisation of BsArsR *in vitro*

Following successful expression and purification of the BsArsR and CgArsR proteins, the next step was to assess their functionality. These repressors have two broad properties required to perform *in vivo* - to bind their operator sequence, thereby preventing operon transcription, and to dissociate from this DNA when allosterically bound by its ligand.

The most common method used to probe these interactions *in vitro* is electrophoretic mobility shift assays (EMSAs), more colloquially known as gel shift assays. These rely on the fact that most DNA-binding proteins will alter the ability of bound DNA to migrate through an electrophoresis gel and the resulting difference in migration distance can be visualised using intercalating DNA dyes or radiolabelled DNA probes.

In most cases a transcription factor will slow the progression of operator DNA through a gel matrix, a combination of the larger hydrodynamic cross-section of the ROC versus the operator alone and the likely positive shift in net complex charge, giving the ROC the appearance of a larger DNA species, but this is not always the case. Clear EMSAs rely on pre-gel incubation conditions suitable for

operator-protein binding, as well as electrophoresis conditions that allow these complexes to stay intact. These conditions and the structure of the gel matrix (pore size and distribution) also affect the strength of discrimination between bound and unbound DNA - pores too large may allow both species to migrate equally, while pores too small may prevent large complexes entering the gel at all [146]. It can therefore be necessary to vary buffer pH and salt/solute concentrations, as well as gel voltage and temperature, until a combination is found that sufficiently preserves ROCs.

As expected, these EMSAs took a number of iterations to optimise. BsArsR was one of only two of the initially selected repressors without proof of a particular operator sequence and so care had to be taken when selecting a dsDNA probe to trial with EMSAs. Sato *et al.* [124], who characterised the *B. subtilis ars* operon *in vivo*, identified the promoter region controlling this operon, including its transcription start site as determined by primer extension analysis. Upstream of this, putative -10 and -35 box sequences for binding by RNA polymerase were also suggested. Figure 3.3 highlighted these, alongside the inverted repeat sequence adjacent to this -35 box, hypothesised to be the operator sequence for the BsArsR. No footprinting work or EMSAs were carried out to prove this, but the structure (inverted repeat for binding by a homodimer with 180° rotational symmetry) and placement (close enough to RNA polymerase binding sites to interfere with transcription), made this a highly likely candidate.

3.4.1 Identification of conditions necessary for BsArsR to bind its operator *in vitro*

Initially a 44bp DNA sequence was selected for use in EMSA trials with BsArsR (B44: 5'-TTACAATTAATCAAAATAAATTGATTTATTTGCTTGCATTATTT-3'). This contained the 8bp inverted repeat (underlined) and 13bp either side in case the repressor required bases outside this most minimal sequence. Following the example EMSA protocol from Hellman *et al.* [146], the purified BsArsR was exchanged into protein storage buffer 1 (PSB1) and incubated with the B44 operator for 30 minutes at room temperature, before being run on a 6% acrylamide 'DNA Retardation gel' from Invitrogen. The first few attempts at this protocol were unsuccessful, unfortunately, so possible reasons for this failure were investigated.

One theory was that the C-terminal His-tag prevented binding through steric hindrance of the DNA-binding site. To check, BsArsR was incubated with TEV protease, which should cleave off this C-terminal extension to leave a protein very close to the wild-type repressor. This truncated BsArsR, alongside a parallel sample incubated with the TEV protease buffer but without the protease itself, was incubated with the B44 operator and run on a gel, shown in Figure 3.8. In this gel the B44 operator migrated as expected, but both protein samples show two bands - one at the same 44bp distance and another further up, suggesting some of the DNA in each well was retarded by the presence of the repressor. The second band in the well containing the truncated protein migrated slightly further than the second band in the well for the full length protein - the retardation effect therefore correlating with the mass of the repressor in each well.

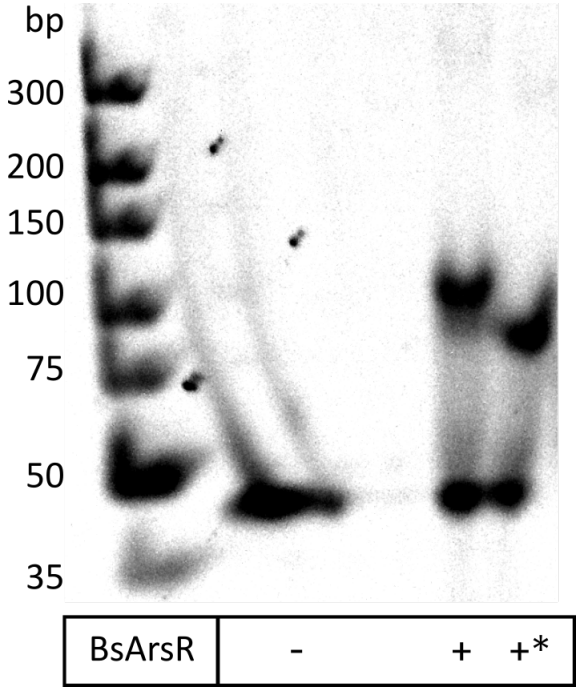


Figure 3.8: **B44 dsDNA can be bound by BsArsR *in vitro*.** The 44bp BsArsR operator (B44) on its own (-), incubated with His-tagged BsArsR in TEV protease buffer (+) or trimmed BsArsR with its His-tag cleaved off by TEV protease (+*).

This surprising result suggested three conclusions. First, that the B44 oligo did in fact contain the native BsArsR operator sequence. Second, that BsArsR was capable of binding this operator with or without the C-terminal His-tag. Third, that something within the TEV protease buffer or protocol, but not the action of the protease itself, enabled the formation of the ROC. The protease protocol included 1 h incubation at 34 °C, but this temperature was unlikely to enable binding as the mix would have cooled before loading onto the gel. The final

concentration of buffer components in each sample - a combination of protein storage buffer, oligo annealing buffer and TEV buffer - was only slightly different for most components compared to when the TEV buffer was not present. Salt, glycerol and Tris-HCl were minimally altered, however the pH would have been marginally elevated by the presence of the TEV buffer and the DTT concentration would have increased from a final 0.077 mM to 1.423 mM with the TEV buffer included. This 18.5-fold increase in reducing agent concentration was the most noticeable difference between the two formulations and so its effect was tested.

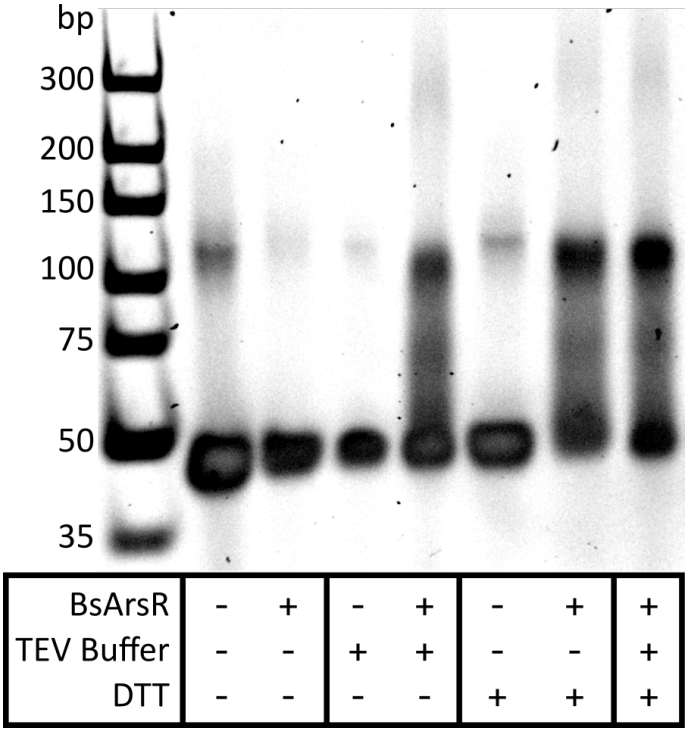


Figure 3.9: **High reducing agent concentration required for BsArsR operator binding.** 6% PAGE EMSA comparing B44 operator dsDNA (10 pmol) incubated with combinations of BsArsR (42 pmol), TEV protease buffer (3 µl) and/or DTT (5 mM).

Figure 3.9 shows the result of this experiment. Once again, BsArsR without the TEV buffer present was unable to bind and shift the B44 oligo, but the band ~100bp in lane 4 and the dark smear between this and the 44bp band indicates that incubation in the TEV buffer enables the formation of this complex. This effect is replicated in the last two lanes, where TEV buffer is either replaced or combined with supplementary DTT. Addition of more DTT alone therefore enabled the repressor-operator complex to form.

While this result is robust, the second band in the other lanes of Figure 3.9 does make this less clear. Its presence is unfortunately close to the protein-DNA band seen in lanes 4, 5 and 6, but does appear slightly higher. While this could be

the result of protein contaminating all the samples (and some ROC formation without high DTT/TEV buffer), this seems unlikely given further protein does not produce a stronger second band in lane 2 versus lane 1. The more likely explanation is that the B44 DNA sample is not comprised purely of one 44 base pair double-stranded DNA duplex, but also contains a higher molecular weight species formed by the concatemerisation of the sB44 or sB44c single strands. This is a constant risk when working with short oligos containing inverted repeats as single strands are prone to forming hairpins or polymeric structures if one strand exists in a slight excess of the other. Figure 3.10 confirms this theory as the B44 ROC appears as a much larger molecular weight species in the higher percentage acrylamide gel, while the secondary band in the DNA-only well is unchanged and a faint hairpin band can also be seen at a lower molecular weight.

3.4.2 The minimum BsArsR operator length is roughly 28bp

Following confirmation that the B44 sequence contained the BsArsR operator, the minimum sequence within this necessary for BsArsR binding was investigated. In order to do this, a range of smaller DNA duplexes were formed by trimming one base pair from each end of the B44 sequence, preserving the inverted repeat in the middle of the duplex. A selection of these oligos is shown in Figure 3.10, comparing each with and without the repressor added, all samples now incubated under the stronger reducing conditions identified above.

This experiment demonstrated that the minimal BsArsR operator sequence was about 28bp long, much shorter than the initially selected 44bp. Operators this size, or with additional nucleotides at each end were all capable of being bound, while those smaller did not produce a clear, shifted, ROC band when combined with BsArsR. The slight smearing present in the lanes containing repressor for these shorter oligos suggests that these sequences retain some affinity for the protein, but the interaction is not sufficiently strong to preserve an ROC in an EMSA. This suggested that while the inverted repeat within the operator is the most important sequence, some external DNA is necessary to stabilise the bound complex (B28: 5'-AATCAAAATAAATTGATTTATTTGCTTG-3').

This EMSA used a 20% acrylamide gel instead of the initial 6% gel and this protocol produced a much clearer distinction between operator and ROC bands. The smaller pore sizes within this denser matrix still allowed these complexes to

enter, but they were retarded much more strongly than in the lower percentage gels, appearing equivalent to 300bp or even larger species compared to roughly 100bp dsDNA. As explained above this clarifies the presence of the B44 concatemer. This gel also showed further bands in the B30 lanes, likely corresponding to multiple hairpin structures. This suggested that again one of the individual strands may have been annealed in a slight excess, and possibly that this hairpin species may have been subject to some slight degradation too.

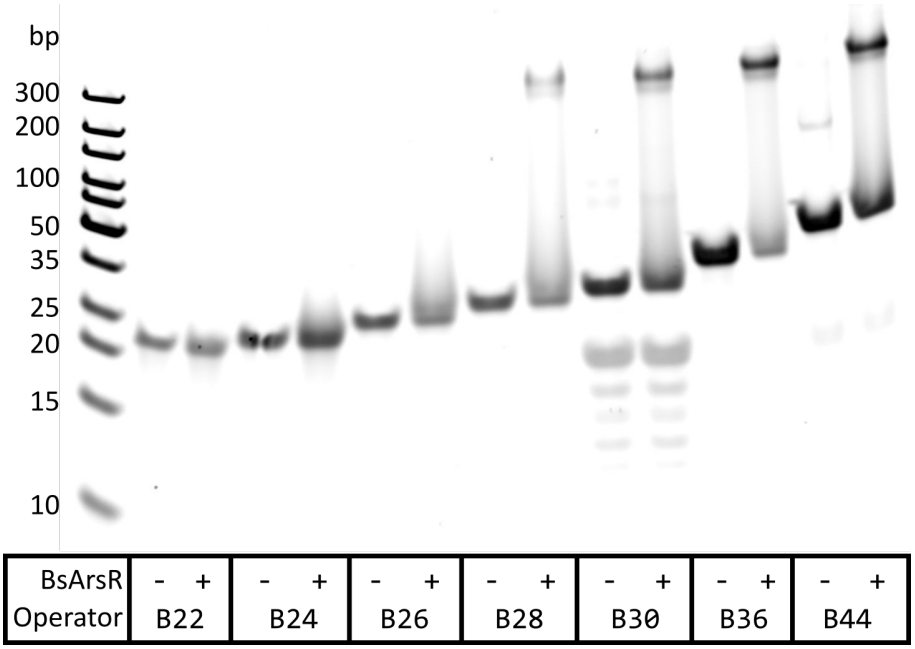


Figure 3.10: **Identifying the minimal BsArsR operator sequence.** 20% PAGE EMSA comparing 10 pmol BsArsR operator dsDNA of varying lengths (22-44bp) incubated with and without 175 pmol BsArsR.

3.4.3 Sodium arsenite causes the dissociation of BsArsR from its operator

Following identification of the minimal BsArsR operator sequence, the other important experiment required to validate the function of this repressor was demonstration of dissociation induced by its ligand, the arsenite ion. To do this a double titration EMSA was performed - first increasing the concentration of BsArsR incubated with a fixed amount of B30 operator until the DNA was saturated, then adding increasing quantities of sodium arsenite into parallel samples to probe if, and how much of, this arsenic species was able to cause dissociation.

Figure 3.11 shows the result of this experiment, with the clear conclusion that sodium arsenite is able to prevent the formation of the BsArsR-B30 repressor-operator complex. About 70 pmol BsArsR was required to bind roughly half of the 10 pmol B30 dsDNA, with the operator maximally bound by 175-210 pmol of BsArsR. This produces a dissociation constant of 7 µM for this 10 µl reaction, with the B30 operator at 1 µM concentration and the repressor at a 7x excess, although a much finer titration would be required to produce a more accurate figure.

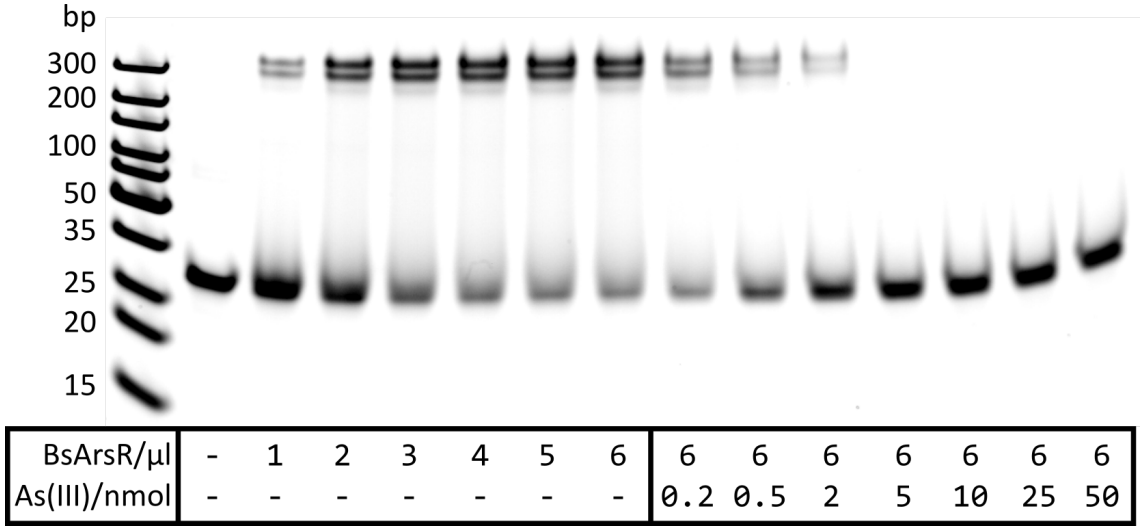


Figure 3.11: **Titration of BsArsR and sodium arsenite against the B30 operator.** 20% PAGE EMSA titrating BsArsR (at 35 pmol dimer/µl) and sodium arsenite against a constant 10 pmol B30 operator DNA.

Even the lowest quantity of added arsenite (200 pmol) produced noticeable ROC dissociation and 500 pmol, a 2.4x excess over the 210 pmol BsArsR dimer present (and 1.2x excess over the theoretical number of arsenite-binding sites), resulted in dissociation of roughly half the ROC. Addition of 5000 pmol or more arsenite caused full dissociation, leaving only the single band of unbound B30 operator on the gel. This was clear evidence that the purified recombinant repressor was able to fulfil its two basic functions *in vitro*.

The shifted ROC bands in this assay presented as a pair, unlike previous assays but consistent with other EMSAs using BsArsR from a second purification and expression cycle. The paired bands suggested the presence of two different ROCs, resulting either from multiple operator or repressor species, or from the possibility of a second complex stoichiometry other than one operator to one dimerised repressor. The uniformly single operator-only band reduced the likelihood of there being multiple DNA species present, and it seemed highly

unlikely that multiple repressors could bind a single B30 operator sequence, or vice versa, given the absence of this occurrence in previous EMSAs and the minimal nature of the operator sequence. The most likely explanation seemed to be two repressor products from the most recent purification - either a read-through translation or a short truncation of some of the protein at some point that did not affect expression, purification or function.

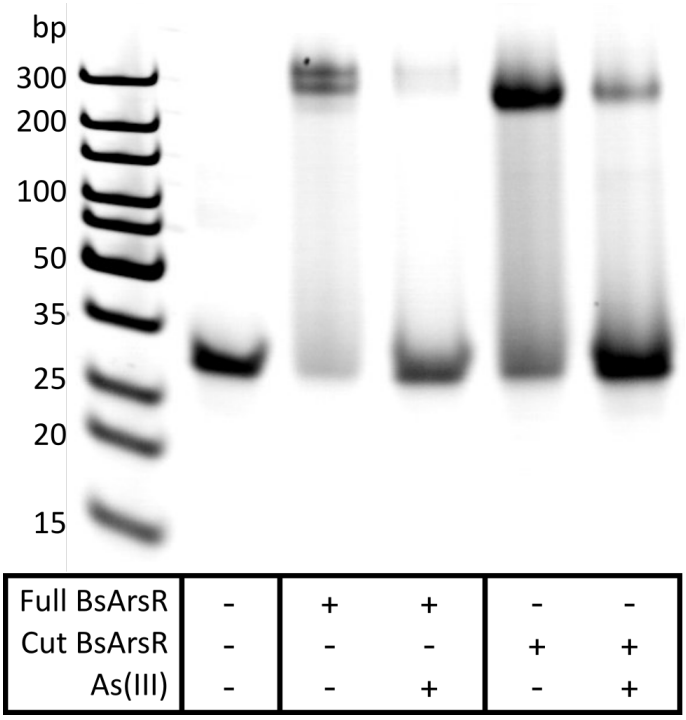


Figure 3.12: **Read-through translation explains multiple ROC species.** 20% PAGE EMSA comparing the appearance of 15 pmol B30 operator DNA bound by 200 pmol full-length BsArsR or BsArsR cut by TEV protease, with or without 2500 pmol sodium arsenite.

The possibility of multiple BsArsR proteins was investigated by comparing the ROC of a full-length BsArsR protein to that formed by BsArsR cut by TEV protease. Any difference in protein length of multiple BsArsR species, as a result of C-terminal read-through or degradation past the TEV protease recognition sequence, would be removed following cleavage. Figure 3.12 shows the result of this assay. While the full-length BsArsR forms two ROCs once more, there is only one ROC formed by the cut BsArsR. This demonstrates that either a read-through translation or a second small, discrete, truncated BsArsR was responsible for the double-banded ROC.

3.4.4 BsArsR does not bind single-stranded operator oligos

The last experiment required before testing a DSD system was to check that single-stranded operator DNA species could not be bound by the repressor. In the basic design for the repressor-modulated DSD reaction (Figure 1.6), the repressor must be able to bind the double-stranded full operator sequence, but not the single-stranded displacing oligo or the resulting duplex with the partially single-stranded operator in order to affect the reaction equilibrium. Although unlikely, it seemed appropriate to test whether the hairpin species formed by these single strands could be bound at all by the repressor, given they would form one half of the inverted repeat dsDNA.

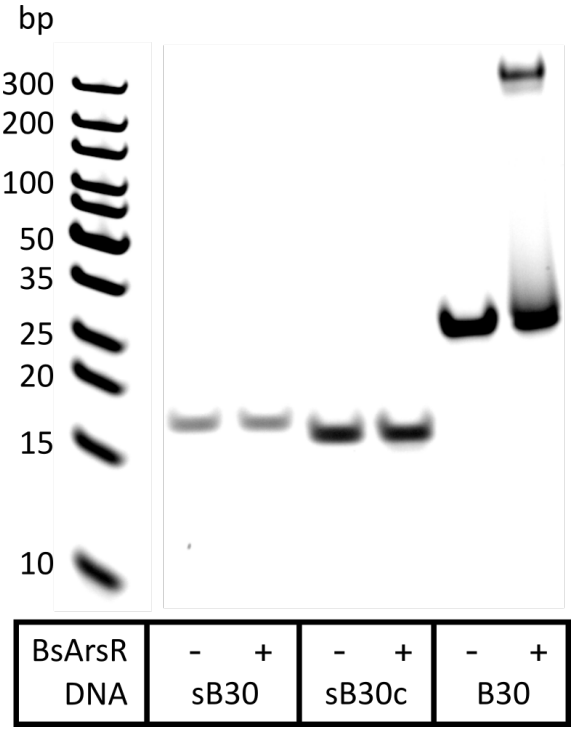


Figure 3.13: **BsArsR does not bind single-stranded B30 oligos.** 20% PAGE EMSA testing the ability of BsArsR (140 pmol) to bind the sB30 and sB30c single strands compared to the double stranded B30 (all DNA 15 pmol). Other lanes of the original gel image have been removed for clarity.

Figure 3.13 shows the result of this experiment, where each of the B30 operator single strands (sB30 - forward/top strand; sB30c - complementary strand) was incubated with and without the repressor and run on an EMSA. The protein-free hairpin DNA appear on the gel equivalent to slightly larger than 15bp dsDNA, owing to their non-simple structure, but are clearly unaffected by the presence of BsArsR, unlike the fully double-stranded B30 DNA.

3.4.5 BsArsR can modulate *in vitro* DSD reactions

Following successful demonstration that BsArsR can bind the double-stranded B30 operator, but not its single-stranded component oligos, an attempt was made at performing an *in vitro* DSD reaction. Figure 3.14 displays the desired DSD behaviour, an adaptation of Figure 1.6 from the introduction, with unmodified DNA used and each species labelled.

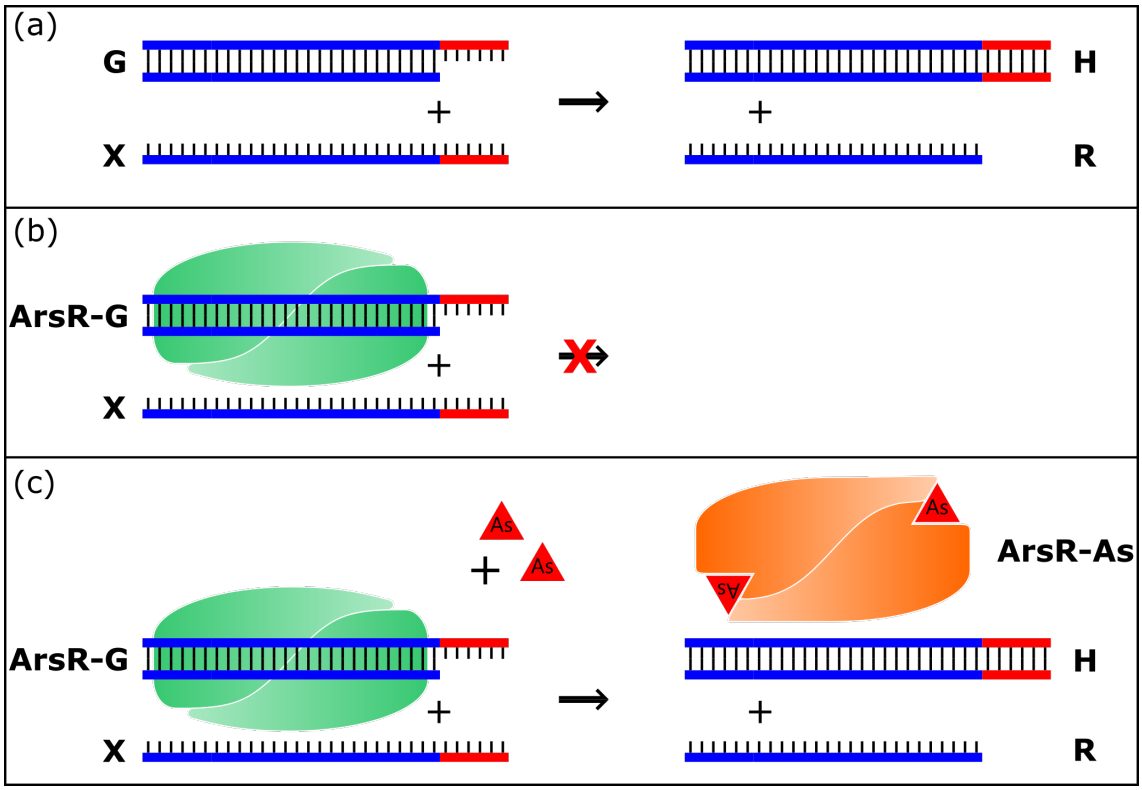


Figure 3.14: **ArsR-mediated control of basic DSD reaction.** Schematic illustrating how ArsR could modulate a DSD reaction between unmodified DNA species: (a) Basic irreversible DSD reaction. (b) ArsR binds operator sequence, preventing DSD from occurring. (c) Arsenic induces dissociation of ArsR from its operator sequence, thereby enabling the DSD reaction. The labelled species are: the 30/36nt B30-G6 gated duplex (G), the 36nt single-stranded X6-sB30c displacing oligo (X), the 36bp double-stranded DSD product (H), the 30nt single-stranded sB30c product oligo (R), BsArsR-G repressor-operator complex (ArsR-G) and the dissociated arsenic-bound BsArsR (ArsR-As).

In this case, the G duplex (B30-G6) comprised of the double-stranded B30 operator sequence, with a 6 nucleotide single-stranded toehold or gate sequence (5'-CACTGG-3') at the 3'end of its forward strand. The X displacing oligo (X6-sB30c) therefore consisted of the complementary toehold sequence followed by the reverse sequence from the B30 operator. The H duplex (B30-H6) is the 36bp product of their reaction, with R (sB30c) the displaced 30nt reverse

operator strand. ArsR-G is the ROC and ArsR-As the arsenic-bound repressor unable to bind its operator sequence.

While DSD reactions are usually carried out using fluorophore-modified oligos and investigated using plate readers, these are restricted to following the behaviour of only the fluorophore-modified DNA itself and not other species. It seemed possible to track more DSD species, albeit at a lower resolution and at a single time-point, by using an EMSA to separate out all the component DNA complexes, so a set of samples were prepared to test whether BsArsR performed as expected by preventing a DSD reaction from occurring until arsenite was added. Figure 3.15 shows the result of this experiment.

B30-G6	15	-	15	15	15	15	15	15	15	15	15	15	15	15
X6-sB30c	-	30	7.5	15	22.5	30	-	7.5	15	22.5	30	30	30	30
Cut BsArsR	-	-	-	-	-	-	200	200	200	200	200	200	200	200
As(III)	-	-	-	-	-	-	-	-	-	-	-	600	1500	3000
Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Figure 3.15: **BsArsR can control B30 DSD reaction.** 20% PAGE EMSA testing the ability of BsArsR to prevent a DSD reaction, and of arsenite to subsequently relieve this repression. The B30-G6 duplex, BsArsR cut with TEV protease and sodium arsenite were incubated together for 20 min before addition of the X6-sB30c displacing oligo and a further 20 min incubation in a total volume of 11 µl, in the indicated quantities given in pmol. The bands corresponding to the ArsR-G, H, G, X and R species are indicated on the right.

Lane 1 shows the band corresponding to the G duplex, which migrated roughly like 34bp dsDNA. The X displacing oligo is shown in lane 2, appearing like a 25bp dsDNA molecule. A small amount of G from lane 1, spilled over into lane 2, allowing some G and X to react and form H and R - the H can be seen in lane 2 as a 36bp band. Lanes 3-6 demonstrate the result when G is incubated with increasing amounts of X. As more X was added to the reaction, more G was consumed and more H and R were formed. In lane 6, where there was twice the amount of X as G, almost all G reacted, leaving a dominant H band and clear

smaller bands corresponding to excess X and R. The presence of significant quantities of these single strands also produces higher molecular weight concatemers above the H band, like those seen in Figures 3.9 and 3.10.

Lanes 7-11 show the scenario depicted in Figure 3.14 (b), where the repressor should prevent the DSD reaction from occurring. All these lanes show a clear band for the BsArsR-G repressor-operator complex and this persisted as an increasing amount of X was added. In lane 11 double the amount of X compared to G was added, but only a very weak H band is visible, there is no clear R band and the ArsR-G band is dominant, unlike in lane 6 when no repressor is present and R is seen, G has been consumed and H is dominant. In lanes 12-14 an increasing quantity of arsenite is added, with the result that the ArsR-G band weakens and the H and R bands become more visible.

This result clearly demonstrated that this DSD system behaved as expected, with the BsArsR repressor able to prevent a DSD reaction from occurring, thereby sensitising displacement to the presence and concentration of arsenite. This proof-of-concept enabled the subsequent design and testing of a fluorescence-based version of this assay.

3.4.6 BsArsR can be used in a fluorescence-based DSD system

Conversion to an initial fluorescence-based system was relatively straightforward. The sB30-G6 strand was replaced by one with a TYE563 fluorophore conjugated to its 5' end; the reverse sB30c strand was modified with a 3' Iowa Black Dark Quencher. Once annealed together they formed a quenched TYE563-G duplex, which should produce a fluorescent TYE563-H species and a free quencher oligo once reacted with the X displacing oligo, as depicted back in Figure 1.6 (a). Once again pre-incubation of this G duplex with BsArsR should prevent the reaction with X taking place, until the addition of arsenite. Figure 3.16 shows the result of an hour-long time course for this system, using three replicates for each of six different arsenite concentrations.

The relative fluorescence units (RFU) for each sample start at 4000 and then quickly increase correlating with the concentration of added arsenite. Within 10 minutes of addition the top four concentrations have clearly separated from the baseline signal and from each other, with the replicates very similar indicating

this differential signal is reliably due to an arsenite-specific response. These top four arsenite concentrations continue to increase at a steady rate over the hour, with the top two saturating towards the end of the assay. The lowest arsenite concentration, 10 µM, did not significantly separate from the arsenic-free control throughout the hour, but its trajectory suggested this would be likely if the time course was continued.

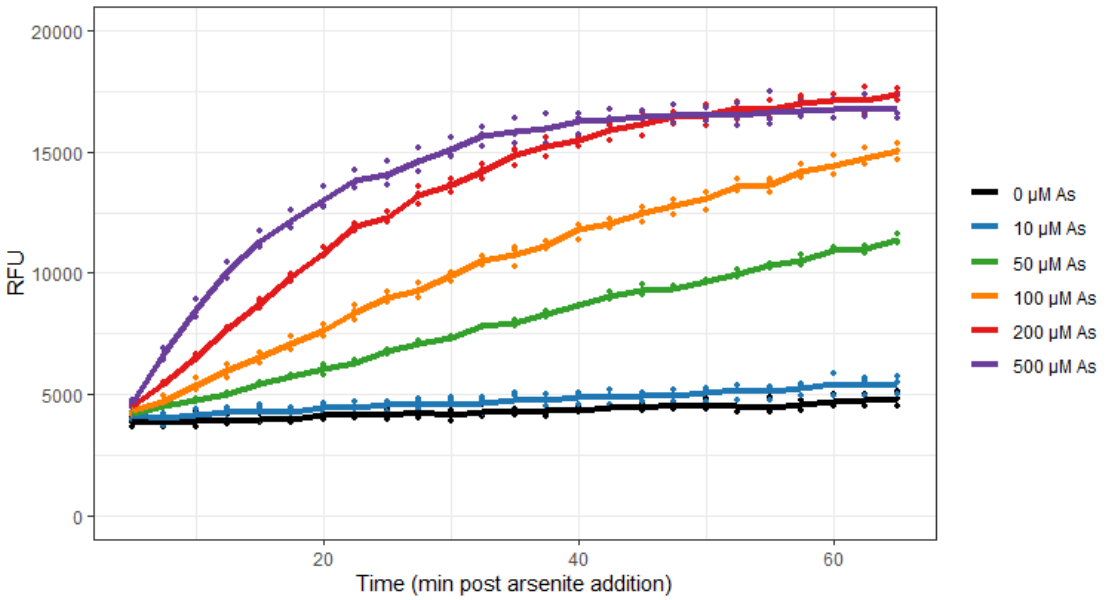


Figure 3.16: **Fluorescent B30-BsArsR DSD system responds in arsenite concentration-dependant manner.** Fluorescence time course displaying the RFU for samples containing different arsenite concentrations. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each sample contained 15 pmol TYE563-G and 240 pmol BsArsR co-incubated for 45 min before addition of 30 pmol X, a further 20 min incubation then addition of 3 µl sodium arsenite at the specified concentrations, for a total volume of 10 µl per well of 384-well plate.

This result conclusively demonstrated that the proposed repressor-operator DSD system was capable of quantitative detection of arsenic. While the arsenite concentrations tested in this first fluorescence-based assay were much higher than those likely to be found in contaminated drinking water (10 µM ~ 750ppb), this would be certain to improve through increasing the analyte volume (>3 µl) and optimising the relative concentrations of components.

3.4.7 BsArsR is less stable than CgArsR

Throughout this work with BsArsR, multiple problems presented themselves. One was the constant presence of operator DNA hairpin and concatemer species

resulting from its long inverted repeat sequence, while another was the presence of read-through translation products of the repressor itself. A further problem occurred several times while carrying out these assays - protein stocks had a high tendency to precipitate and could not be resolubilised, so further work required more rounds of protein expression and purification.

A common occurrence is illustrated by Figure 3.17, where an attempt to improve the sensitivity of the fluorescent B30-BsArsR DSD system was carried out by repeating the assay shown in Figure 3.16, but with half the BsArsR. This change was expected to improve arsenic sensitivity, perhaps at the cost of increased background noise or variability, however instead all the samples, including those with no added arsenic, immediately produced saturating signal with higher variability than seen in the previous assay. The protein sample used for this assay shortly crashed out of solution, following the trend whereby a BsArsR stock would stop being functional before precipitating soon after.

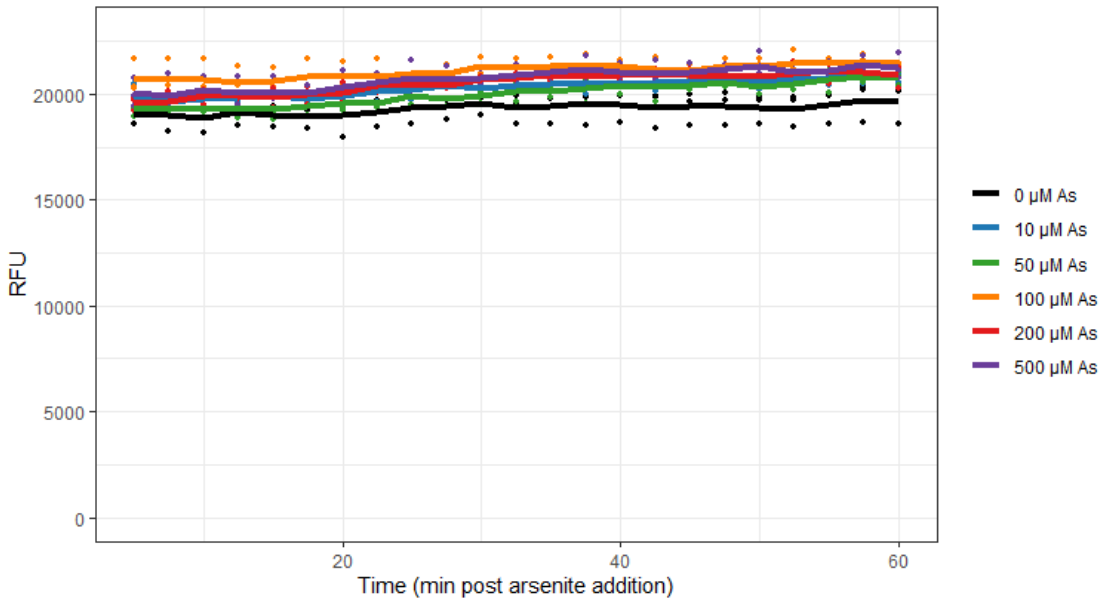


Figure 3.17: **BsArsR loses functionality before precipitating.** Fluorescence time course displaying the RFU for samples containing different arsenite concentrations. Graph shows readings from individual wells as points and averages across each set of three replicates as lines, with preparation method identical to Figure 3.16, except half the amount of BsArsR was added to each sample (120 pmol).

While subsequent rounds of BsArsR expression produced purified protein of seemingly similar mass, as visualised on coomassie-stained SDS-PAGE gels, their behaviour in EMSAs differed, despite deriving from the same sequence-verified glycerol stock of transformed *E. coli*. As noted earlier, the first

purification produced BsArsR that formed a single ROC, while the second produced two ROC species, as seen in Figure 3.12 and attributed to a read-through translation product. A third and fourth expression was compared to this second sample in Figure 3.18, demonstrating further disparate behaviour. The third sample seemed to contain an even larger BsArsR, but one that did not reduce to the smaller species following TEV protease digestion. The fourth sample seemed to contain a single, small protein, but one either with lower B30 operator affinity or much lower purity than previous samples given its reduced gel shift ability despite A280 concentration standardisation. The inconsistency of purity and behaviour between subsequent preparations of BsArsR made it very difficult to reproduce previous results precisely once a protein sample had precipitated, thereby limiting continuity of results.

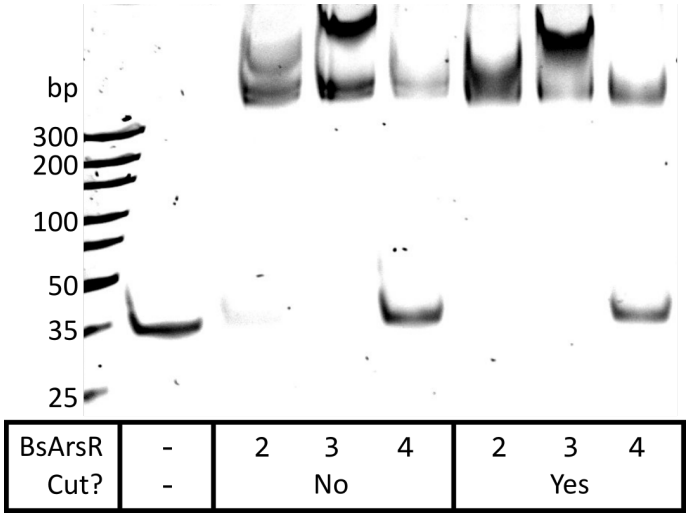
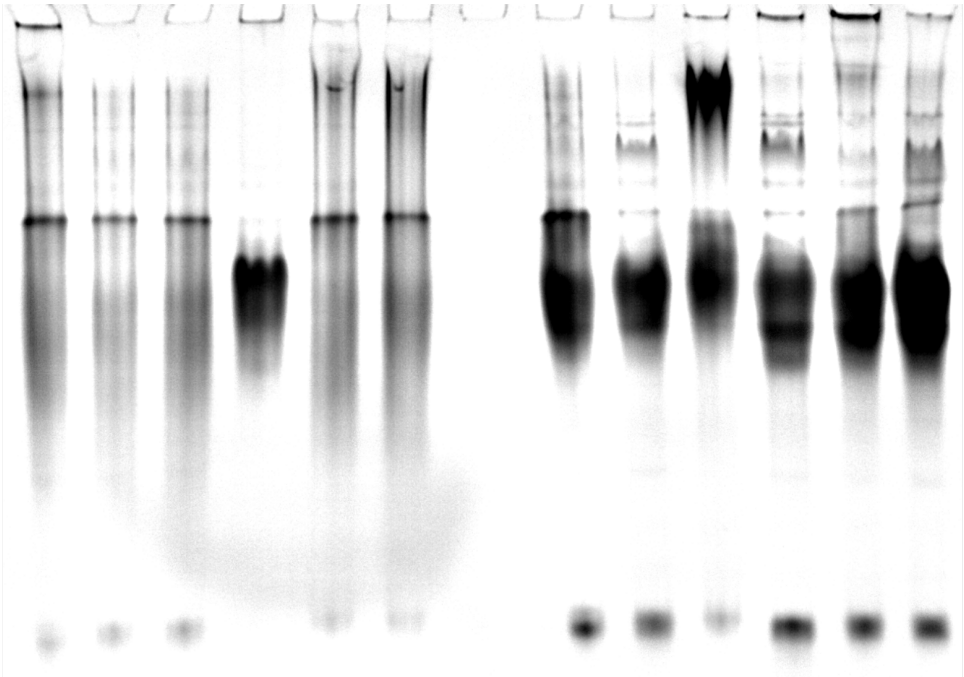


Figure 3.18: **BsArsR behaviour variable between independent purifications.** 20% PAGE EMSA comparing the migration of 15 pmol B30G6 duplex when incubated with estimated 200 pmol of different BsArsR samples. Repressors were from different rounds of expression (2-4), either as purified (uncut), or digested with TEV protease (cut). Lanes are skewed and the ladder partially lost due to a leak in the electrophoresis apparatus.

While it should be possible to find buffers that this small, globular, theoretically homodimeric protein should be stable in, initial attempts at altering pH, salt concentrations and reducing agents did not noticeably aid stability. The isoelectric point for all the theoretical wild-type, tagged and cut protein sequences was in the 4.96-5.78 range, but all buffers were at least one pH unit away from this range.

A native PAGE gel was performed to test whether different reducing conditions may aid sample stability, as this was the factor most responsible for determining BsArsR functionality. Figure 3.19 shows the result of this, comparing a number

of samples of both BsArsR and CgArsR. The BsArsR samples were either free of reducing agent and having stopped binding its operator but having not precipitated (lane 1), this sample incubated with the specified concentration of fresh reducing agent for one hour (lanes 2, 3, 5, 6) or having been incubated for a month with the specified 'old' reducing agent but still no longer functioning (lane 4).



RA	-	DTT	DTT	TCEP	TCEP	TCEP		-	DTT	TCEP	DTT	TCEP	DTT
[RA]/mM	-	2.5	5	5	1	2		-	5	5	5	2.5	5
Fresh/Old RA	-	F	F	O	F	F		-	O	O	O*	F	F
Lane	1	2	3	4	5	6		7	8	9	10	11	12
ArsR	BsArsR							CgArsR					

Figure 3.19: **CgArsR is more stable than BsArsR.** Coomassie-stained native PAGE gel of BsArsR and CgArsR protein samples, either free of reducing agents (RA), stored with the specified concentration of reducing agent (O), or incubated with fresh reducing agent for one hour (F). All samples were stored at 4 °C except for O* (lane 10), which was stored at room temperature for one month.

The single, clearly dominant band seen in lane four is suggestive of a single globular protein species, whereas the smears seen in the other BsArsR lanes suggest that these samples may have started to unfold, degrade or aggregate. This long term storage in TCEP may therefore have aided stability somewhat, but it was not able to prevent loss of functionality nor precipitation (as this sample did also precipitate soon after) nor could TCEP treatment rescue solubility/function from a defective sample. It may be possible that BsArsR was incredibly sensitive to irreversible oxidation, which even cold storage with TCEP

was unable to prevent for more than a couple of weeks and as a result became very hard to use.

CgArsR, on the other hand, proved a much easier protein to work with. It only precipitated when brought close to its isoelectric point of 5.44-6.09 (depending on exact isoform), but even this was reversible when pH was restored to 7.4-8.0. As shown in Figure 3.19, CgArsR maintained its globular structure even without the presence of reducing agents, although a high concentration of TCEP (lane 9) may have begun to denature the repressor. Notably CgArsR maintained its stability and functionality even when stored at room temperature for a month (lane 10) and such high stability made it preferable both for research but also for potential real-world biosensing applications. Most lanes in the native PAGE gel also contained a small, fast migrating band - this may be the less common monomer form of each repressor, but was not investigated.

3.5 Characterisation of CgArsR *in vitro*

The biochemical mechanism behind arsenic resistance in *Corynebacterium glutamicum* was first studied by Ordóñez *et al.* [128], who identified that this bacteria's genome contained two *ars* operons responsible for this characteristic. Each operon consisted of genes encoding a repressor, a permease and a reductase, with further permease and reductase proteins found elsewhere in the genome - together this large system enabled the bacteria to have very high resistance to both arsenite and arsenate.

A second study by the same authors investigated the arsenic-responsive transcription factors in more detail [130]. They found that both operons could be induced by arsenite addition, and that both repressors could control the expression of both operons and dissociate upon arsenite addition with similar efficacy. The two ArsRs shared 66% sequence identity and contained an unusual arsenic-binding domain, suspected to be formed at the interface between the monomers within its homodimer. Closer inspection of the *ars1* operon revealed that its promoter was in fact bidirectional, with the effector genes (permease and reductase) coded on one strand and the regulatory gene (ArsR) coded upstream from this region on the reverse strand.

Footprinting analysis using the CgArsR protein identified two 30bp sequences

in the intergenic promoter region that were bound by the repressor, as shown in Figure 3.20. Fluorescence anisotropy further confirmed that double-stranded DNA oligos of each sequence could be bound by CgArsR, and that the operator closer to the effector proteins (CR30) seemed to have greater affinity for the repressor than CL30. The authors noted that CgArsR may exhibit cooperative behaviour when binding these operators and that 10bp within each sequence formed an inverted repeat together. These sequences were therefore used as the starting point for *in vitro* characterisation.

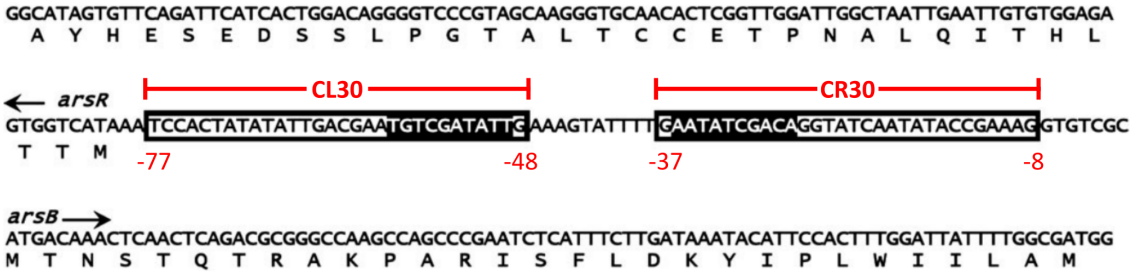


Figure 3.20: **CgArsR operator footprints.** Image adapted from Ordóñez *et al.* [130], highlighting the sequences within the native *C. glutamicum ars1* operon footprinted by CgArsR. The left-hand (CL30) and right-hand (CR30) sequences are labelled as shown, with their numbered positions relative to the *arsB* coding sequence identified below. The black highlighted sequences form a 10bp inverted repeat, suggested by Ordóñez *et al.* to be the key bases for repressor binding and possible cooperative behaviour.

3.5.1 CgArsR can bind both native 30bp operators *in vitro*

Following the successful demonstration of BsArsR binding its B30 operator *in vitro*, the same conditions were used to test for CgArsR's functionality. Figure 3.21 shows the result of an EMSA testing the ability of CgArsR to bind and shift different operators - both the CL30 and CR30 dsDNAs, the 70bp region covering both theoretical operators and the 10bp separating them (C70), and the B30 operator for BsArsR. As shown, CgArsR was capable of binding all three *C. glutamicum* sequences, but not the B30 dsDNA, confirming the footprinting results by Ordóñez *et al.* as well as demonstrating the specificity of CgArsR.

Both the CL30 and CR30 operators showed propensity to form concatemers rather than short hairpins (unlike B30 forming both), while the C70 promoter sequence presented as a slight smear, consistent with the difficulty synthesising and annealing these slightly longer oligos. The intensity of the CR30 ROC band suggests CgArsR may have slightly greater affinity for this sequence over the

CL30 operator as suggested by Ordóñez *et al.*, but the CR30-only lane was similarly more intense so this difference is not clear.

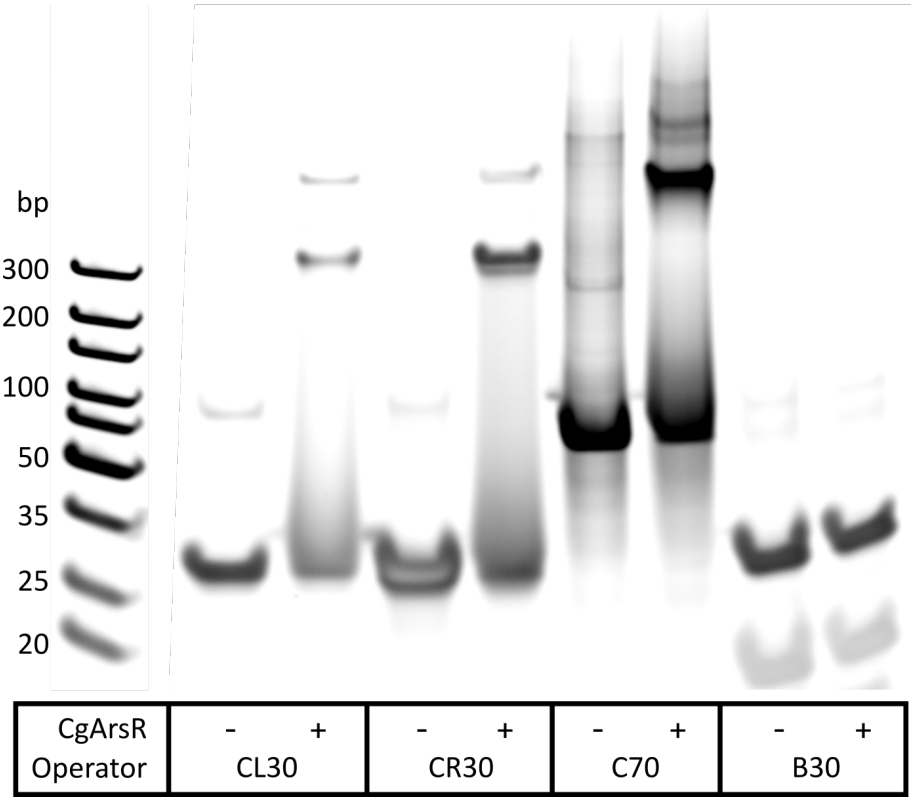


Figure 3.21: **CgArsR can specifically bind both native operators.** 20% PAGE EMSA comparing the migration of 10 pmol different dsDNA operator sequences when incubated with or without 80 pmol CgArsR. Other lanes of the original gel image have been removed for clarity.

3.5.2 Arsenite, but not arsenate, induces CgArsR dissociation

CgArsR was then assayed for response to arsenic. So far, no ArsR protein has been shown to bind to arsenate *in vitro*, with bacterial operons instead relying on reductase enzymes to convert this species into arsenite for detection, sequestration and removal. While Ordóñez *et al.* demonstrated CgArsR could bind to arsenite, they did not test arsenate, and it was possible the repressor could interact with the larger ion as well, due to its unusual arsenic-binding domain.

Figure 3.22 shows the result of the EMSA testing arsenic response. While arsenite clearly induced the dissociation of CgArsR from the CR30 operator, arsenate did not at any concentration tested. Even at the highest tested concentration, where arsenate would have been present at a 25-fold excess over

theoretical arsenic-binding domains, the ROC band intensity remained as strong as in the lanes without any arsenic, suggesting no arsenate-induced dissociation at all. Arsenite, on the other hand, caused complete CgArsR dissociation at the two highest concentrations (10-fold and 25-fold excess), while a strong response was seen at all the lower concentrations too. Roughly half of the ROC dissociated at the lowest concentration assayed, where the amount of arsenite would be equimolar to the number of potential binding sites (200 pmol). The affinity of CgArsR to arsenite was therefore roughly similar to that of BsArsR, although these estimates only derive from semi-quantitative EMSAs.

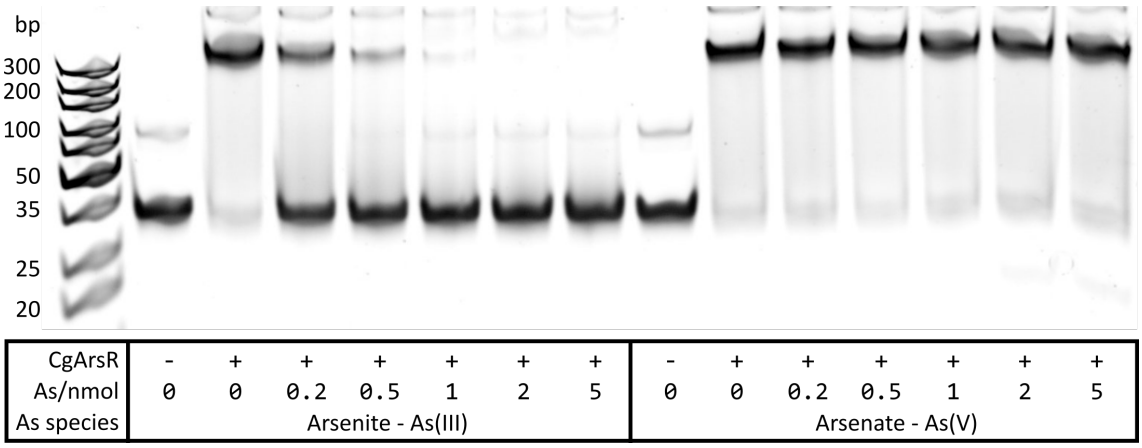


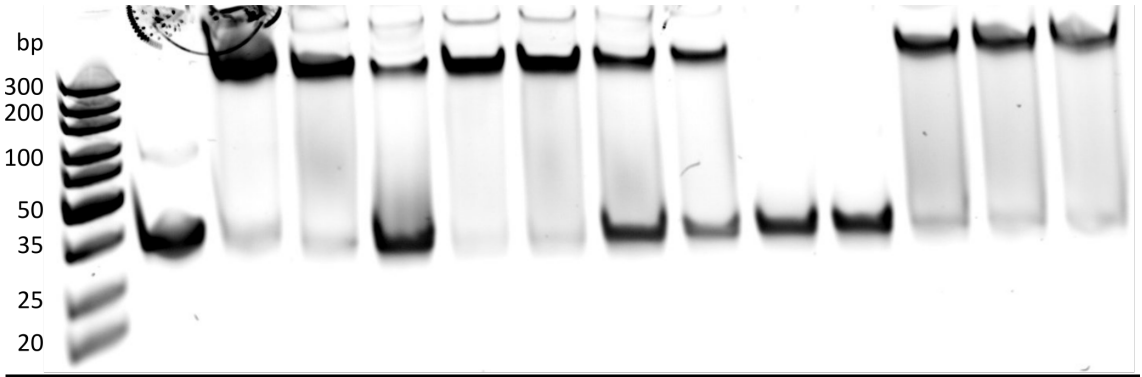
Figure 3.22: **CgArsR responds to arsenite but not arsenate.** 20% PAGE EMSA testing whether 100 pmol CgArsR, bound to 15 pmol CR30, dissociates if incubated with various concentrations of sodium arsenite or sodium arsenate. Protein and dsDNA were co-incubated for 30 minutes before another 30 minute incubation with the specified quantity of arsenic in a total volume of 10 µl.

3.5.3 CgArsR requires reducing conditions for response to arsenite, but not to bind operator

Previously BsArsR was shown to require reducing conditions to bind its operator DNA, so initial tests for CgArsR binding were also carried out under these conditions. The *in vitro* work carried out on CgArsR before, however, either used very weak reducing conditions (0.2 mM DTT for footprinting and fluorescence anisotrophy, as described in [110]) or none at all (EMSA protocol from [129]). Mutagenesis and chemical modification of the cysteine residues even led Ordóñez *et al.* to claim that 'none of the cysteine residues is required for binding to *ars1* o/p DNA'. As reducing conditions primarily help to prevent oxidation of thiol groups in proteins, this work suggested that reducing agents

may not be required at all for CgArsR to bind its operator.

Ordóñez *et al.* did, however, suggest that mutagenesis of the cysteines did reduce affinity of CgArsR for its operator, and that addition of glutathione, which acts as a reducing agent, increased the repressor's affinity for arsenite. Figure 3.23 shows the result of an EMSA comparing the effect of different reducing treatments on CgArsR's ability to bind its operator and respond to arsenite. All conditions, including without the presence of any reducing agent, allowed CgArsR to bind the CR30 operator. This difference in requirement compared to BsArsR is intriguing, and suggests a subtle difference in how changes within the arsenic-binding domain affect the structure of each repressor's DNA-binding domain.



CgArsR	-	+	+	+	+	+	+	+	+	+	+	+	+	
As/nmol	0	0	10	50	0	10	50	0	10	50	0	10	50	
RA	-	5 mM DTT			10 mM DTT			6 mM TCEP			None			

Figure 3.23: **CgArsR requires reducing conditions for response to arsenite but not to bind operator.** 20% PAGE EMSA testing whether 100 pmol CgArsR binds to 15 pmol CR30 and dissociates with high concentrations of sodium arsenite, when incubated with or without different reducing agents. Protein and dsDNA were co-incubated for 30 minutes before another 30 minute incubation with the specified quantity of arsenic in a total volume of 10 µl.

Without reducing agents, however, CgArsR did not dissociate from CR30 upon arsenite addition. This was true even at the highest concentration tested, with a 250-fold excess of arsenite over arsenic binding domains. Addition of DTT or TCEP, however, sensitised the repressor to arsenite, allowing dissociation to occur. At both concentrations of DTT tested, only partial dissociation was seen with the large arsenite addition, whereas maximal dissociation was seen with both arsenite concentration when TCEP was used. TCEP, however, may have caused some denaturation of CgArsR at its 6 mM concentration used here, as CR30 was not fully shifted when CgArsR was added under these conditions. When 1 mM TCEP was used instead, as seen in Figure 3.22, not only did the

same amount of CgArsR cause maximal retardation of CR30, but the response to arsenite was two orders of magnitude more sensitive than for 5 or 10 mM DTT.

The reason for this discrepancy between the effect of TCEP and DTT is likely due to the latter's instability. Although DTT is effective at reducing thiol groups in proteins, it is prone to rapid oxidation, especially if in solution with multivalent metal cations or if not stored at cold temperatures, in which case it could fully oxidise within a week. TCEP, however, is much more stable, oxidising only up to 10% under these same conditions as long as not in a phosphate-based buffer [147]. In addition, TCEP is a stronger reducing agent and therefore lower concentrations are required for use [148]. The DTT used in Figure 3.23 may well have at least partially oxidised, explaining its lower efficacy at sensitising CgArsR to arsenite response, and so in future CgArsR was both stored and assayed in 1 mM TCEP.

3.5.4 CgArsR can be used in a fluorescence-based DSD system

Following optimisation of the reducing conditions enabling CgArsR functionality *in vitro*, the repressor was tested in a fluorescence-based DSD system. A more complex probe design was used for these assays, explained in the following chapter, but the same core principles remained - the probe contained an operator dsDNA sequence and a fluorophore and quencher on different strands, to be separated by the toehold-mediated displacement reaction of a single-stranded oligo. The repressor impeded this DSD reaction until arsenic was added to the solution.

Figure 3.24 shows the results of one such system. Unlike with the B30-BsArsR assay, there is noticeable increase in background signal across the hour, but despite this a clear arsenic concentration-dependant increase in signal above this. The highest concentration of arsenite, 1 mM, is immediately distinguishable from the others, while the 100 µM samples rapidly increase in RFU over the first 20 minutes. Although the 10 µM (~ 750ppb) arsenite samples did not significantly differ from the arsenite-free background by the end of the hour, one third of the volume of arsenite was added relative to those used to initially test B30-BsArsR (to compensate for increased probe volume) so their overall sensitivities seemed very similar.

While there were some slightly more complex dynamics seen in this assay than the B30-BsArsR system shown in Figure 3.16, this result was at least reproducible due to the stability of the CgArsR protein. 750ppb is at the top end of arsenic concentrations expected to be found in contaminated groundwater samples in Bangladesh and so this result showed great promise for the repressor-DSD system and the use of CgArsR in particular, with plenty of optimisations expected to be able to increase sensitivity further.

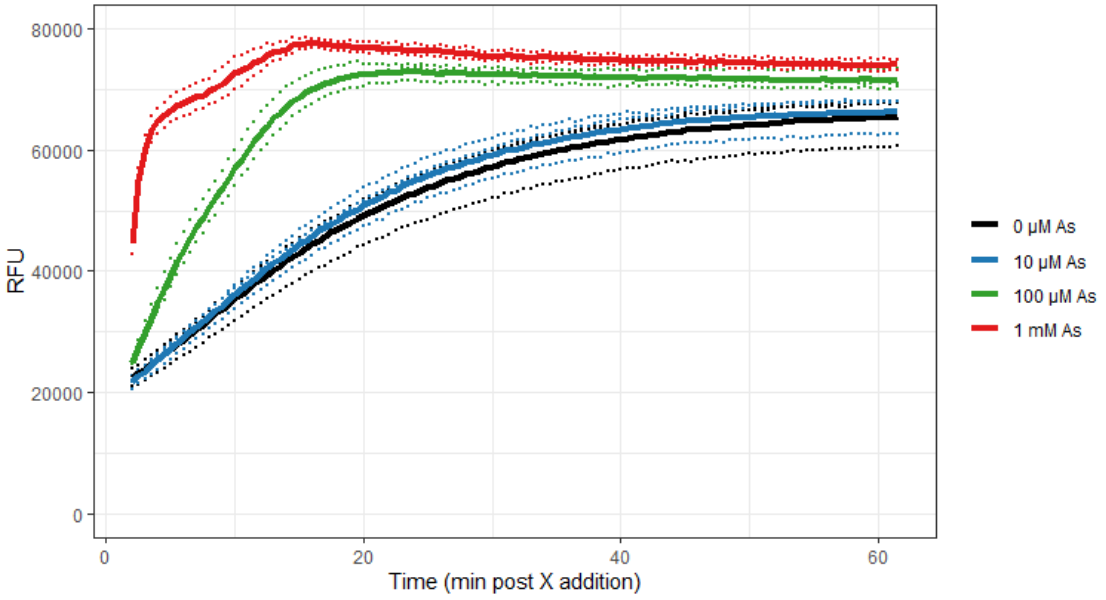


Figure 3.24: **TR30-G6a-CgArsR has high background DSD.** Fluorescence time course displaying the RFU for samples with or without displacing oligo or arsenite. Graph shows readings from individual wells as points and averages across each set of three replicates as lines (two replicates for 1 mM As sample). Each sample contained 10 pmol TR30-G6a and 200 pmol CgArsR co-incubated for 15 min before addition of 1 µl sodium arsenite at the specified concentrations, a further 15 min incubation then addition of 48 pmol X6a-sCR30c, for a total volume of 10 µl per well of 384-well plate.

3.6 Conclusion

The results presented in this chapter demonstrated the feasibility of building a sensory system based on the integration of repressor proteins with DNA strand displacement reactions. As with any new assay, several bottlenecks appeared in the research pipeline - there were many difficulties in selecting, expressing and testing these repressors. Thankfully all of these problems were eventually overcome and two purified proteins displayed clear promise for use in the intended system.

This work contained the first known expression and purification of the *Bacillus subtilis* ArsR protein. EMSAs were able to prove the identification of its operator sequence as well as demonstrate its response to arsenite, both previously only hypothesised. These characteristics were then successfully employed to control DSD reactions and thereby transduce the detection of arsenite. Unfortunately instability, problems with reproducible expression and an operator sequence with strong secondary structure all contributed to make BsArsR a difficult protein to work with. With more time or as a separate future project these challenges could all be overcome and its suitability for DSD integration could be better characterised, but researching this new area in sufficient depth required the selection of a single repressor.

The *Corynebacterium glutamicum* ArsR was initially only included to compare to the more classical arsenic repressors, with the possibility its atypical arsenic-binding domain may provide slightly different properties. While an in-depth comparison was not possible, this structure may well have contributed to its ease of expression as well as improved stability. CgArsR's ability to bind its operator without any reducing agents made it noticeably different to BsArsR, and perhaps to all the class I ArsRs, and is certainly a useful but unexpected characteristic. Although it did not respond to arsenate as hoped, CgArsR demonstrated good sensitivity to arsenite *in vitro* and strong binding to its native operators, both of which contributed to promising early results in a fluorescent DSD system. The remainder of this thesis builds on the work done on CgArsR in this chapter, using this repressor to develop a greater understanding of the considerations required to build functional repressor-DSD biosensors.

Chapter 4

Improving CgArsR operator DSD probe design

4.1 Introduction

The fluorescent DSD probes used in the previous chapter each consisted of four parts - an operator dsDNA region, an ssDNA toehold sequence and the fluorophore and quencher modifications enabling signal quantification. Each of these components have potential for optimisation. For example, different fluorophores have different extinction coefficients, different excitation and emission spectra and different costs. These must be well matched with the available reader, whether a lamp filter or laser, as well as the choice of quencher modification. Good combinations can vastly improve sensitivity and noise reduction over poor ones.

This chapter explores a number of changes to probe design that improved performance of the DSD system. These required the adoption of a 'Tprobe' instead of a linear configuration, to allow higher throughput and cheaper comparison of different designs. Subsequently, a reversible reaction format was adopted to aid signal stability and mutant operators were identified that removed secondary structure. This was combined with the introduction of a polyT spacer sequence adjacent to the fluorophore, creating a more consistent fluorescence environment and thereby enabling a more meaningful quantitative signal. The combination of all these advances resulted in a much faster DSD reaction, capable of more stable and more accurate arsenite detection.

4.1.1 Transcription factors are likely to have a wide sequence space with variable operator affinities

Operator sequences themselves are likely to be open to optimisation. The affinity between a repressor and its operator is a property that will be under selection pressure *in vivo*. Too weak and a cell would waste resources synthesising unnecessary effector proteins; too strong and it would lose sensitivity in responding to its ligand. This property would have evolved, however, for the scenario where only a few copies of a repressor protein and its operator sequence are present within a cell - it may be possible that for the unnatural *in vitro* environment of a potential biosensor, where concentrations of both components could be substantially enriched, a different affinity between repressor and operator could be optimal.

Historically, operator sequences have been identified through characterisation of genomic footprints. While this tells us what sequence, or sequences, a transcription factor is likely to have bound within the genome of one particular organism, it does not tell us the total range of DNA sequences that protein could possibly bind. A repressor is unlikely to have one single operator sequence that it can bind and be unable to tolerate any deviations from this - such a system would be very unlikely to evolve naturally. Instead, a repressor is likely to have a specific 'sequence space', in which every possible DNA sequence (up to the maximum length where affecting an interaction would be possible) has a particular affinity for this repressor in a particular environment - most would have none but many would have low affinity, some would have medium affinity, and a few may have high affinity.

The operator used in the previous chapter for assaying BsArsR notably contained an 8bp inverted repeat sequence, likely the bases making the most important contacts with the BsArsR DNA-binding domain. But EMSAs suggested that the minimum length for the operator was 28bp, thereby containing 12bp in addition to this core sequence. These may provide a structural role, but it's possible the exact sequence of these bases does also influence binding affinity. The inverted repeat itself may also tolerate mutations; perhaps its affinity is even improved by some. Successful EMSAs only prove that a repressor can bind a particular sequence with sufficient affinity to be visible on a gel, and, if a native sequence is used, suggest this plays a role *in vivo*.

Work by Ordóñez *et al.*, highlighted in Figure 3.20, demonstrated that the CgArsR promoter region contained two non-identical operator sequences, each capable of binding CgArsR independently [130]. Ordóñez *et al.* suggested that 10bp forming an inverted repeat across these two operators was important, but this is at odds with their ability to bind CgArsR separately. The similarity between the two 30bp sequences is highlighted in Figure 4.1, comparing the native top strand sequence of CR30 to the reverse complement of the CL30 operator, from the perspective of the *arsB* coding sequence.

This alignment reveals that bases are identical at 19 out of the 30 positions, within two blocks and a lone base highlighted in blue. The longest continuous sequence able to form an inverted repeat within these sequences is an 'ATAT' (in green and bold) - these would be speculative candidates for the most important bases within these operators, following the example of the BsArsR operator, but this may not necessarily be the case.

		1	2	**3**	**4**	**5**	**6**	7	8	9	10	11	12	13	14	15	16	17	18	**19**	**20**	**21**	**22**	23	24	25	26	27	28	29	30	
CL30	5'-	C	A	**A**	**T**	**A**	**T**	C	G	A	C	A	T	T	C	G	T	C	A	**A**	**T**	**A**	**T**	A	T	A	G	T	G	G	A	-3'
CR30	5'-	G	A	**A**	**T**	**A**	**T**	C	G	A	C	A	G	G	T	A	T	C	A	**A**	**T**	**A**	**T**	A	C	C	G	A	A	A	G	-3'

Figure 4.1: **Native C30 operators share 63% identity.** Alignment of the CR30 and CL30 native CgArsR operator sequences, with the reverse complement of the CL30 sequence shown in Figure 3.20 used. Identical bases are highlighted in blue, while an inverted repeat within these is highlighted in green with bases and positions in bold.

The fact that 11 out of the 30 bases in these operators differ is clear proof that CgArsR must have a large possible binding repertoire out of the 30bp dsDNA sequence space. Ordóñez *et al.* claimed that the CL30 operator had a much weaker interaction with CgArsR than CR30, possibly backed up by the EMSA in Figure 3.21 - this difference must clearly result from the alteration of at least one of the twelve differing bases, or more likely the combination of several. If CL30 is a native operator sequence that does not have the highest possible affinity of a 30bp dsDNA for CgArsR, then the same may be true for CR30, in which case a non-native sequence could be found that improved repressor-operator affinity.

4.1.2 Displacing X oligos using native CgArsR operator sequences are highly structured

While altering the operator sequence would clearly affect the affinity of repressor binding, it would also affect the behaviour of the displacing X oligo. As these are single-stranded DNA molecules they can form secondary structures both internally or with other single-stranded DNA molecules. The two parts of the displacing X oligo - its toehold and operator domains - determine both the specificity of what it reacts with, and the dynamics with which this occurs.

The toehold sequence of a DSD reaction greatly determines its speed and strength. Longer toeholds, and those with a higher GC content, tend to react faster, due to the larger free energy release following hybridisation with their complementary single strands. This correlation is true, however, only over a small range of toehold lengths and binding energies - an analysis by Zhang *et al.* suggested that hybridisation rate constants strongly plateaued once greater than six nucleotides long [57].

Choice of toehold sequence also affects potential secondary structure within a displacing or product oligo; presence of these can complicate and slow DSD reaction dynamics. Removing or being able to control the level of secondary structure within the operator may therefore be beneficial. The homodimeric nature of these repressors mean that operators are likely to contain inverted repeat sequences prone to hairpin and concatemer formation, but it may be possible to remove or alter these with targeted mutations that do not lose repressor binding.

Figure 4.2 compares the predicted structures of the B30, CL30 and CR30 operators when in the single-stranded displacing oligo form. Structures are given both for the 30nt sequences alone (with 'c' suffix to denote reverse complement of operator top strand) and for oligos with 6nt, 5' toehold sequences as used in some of the displacement reactions - either X6 (CCAGTG) or X6a (GCCCGC). The minimum free energy (MFE) of these structures has also been calculated for the specified conditions - the more negative the greater the stability of the ssDNA.

The operator-only structures in the top row demonstrate the propensity for the internal inverted repeats to form hairpins. The sCL30c oligo forms an even more

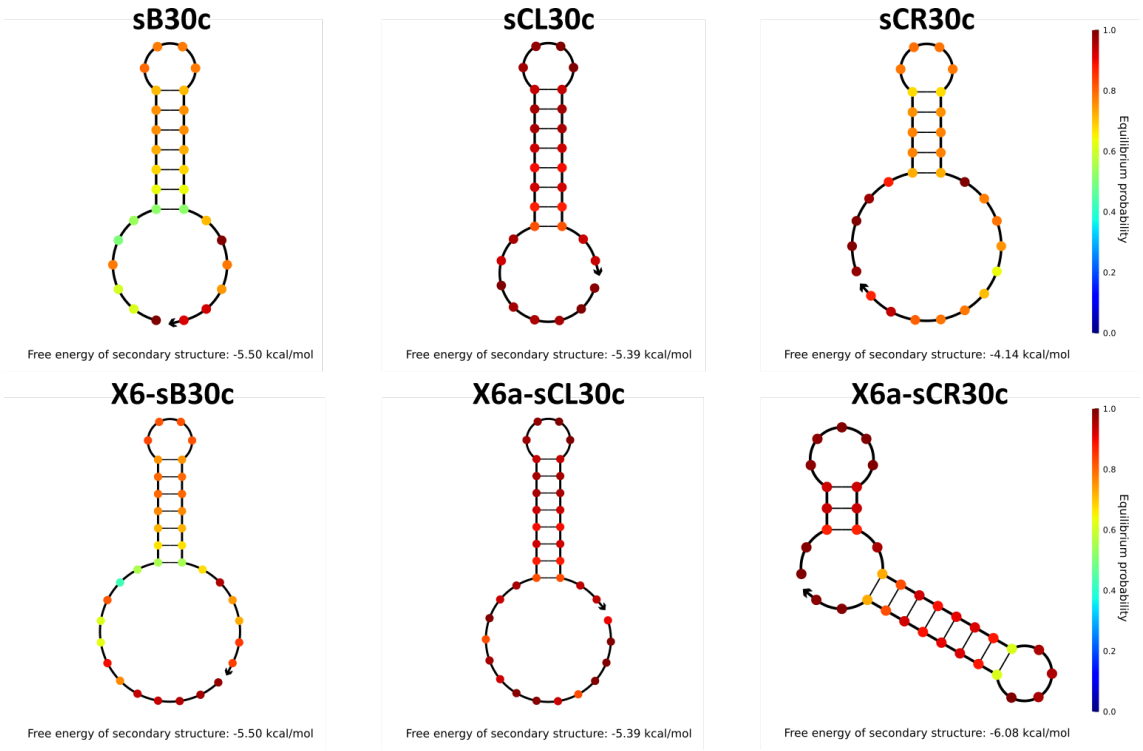


Figure 4.2: **Native operator ssDNA oligos naturally form hairpin structures.** Minimum free energy secondary structures formed by single-stranded oligos comprised either of the 30nt reverse complement sequence of the identified B30/CL30/CR30 operators, or these downstream of a 6nt toehold sequence (X6 or X6a). The structures and the associated minimum free energies were predicted using NUPACK at a concentration of 1 µM and a temperature of 25 °C, with individual bases coloured according to the probability of being in the suggested state using the scale on the right. The 3' end of each oligo is labelled with an outward arrow.

stable hairpin than the sB30c oligo, while the sCR30c hairpin is noticeably weaker. The first two of these are unaffected by the addition of the toehold sequence, maintaining the same hairpin and calculated free energy, but the X6a toehold causes the sCR30c oligo to form two different hairpins, demonstrating how toeholds and operator (or 'recognition') domains can interact to form new structures. The NUPACK software used to perform these predictions allowed for the formation of non-canonical G-T wobble base pairs, highlighted in blue in Figure 4.3, which compares the full sequences and structures of these six oligos.

As mentioned earlier, presence of secondary structure in these oligos is not inherently bad, but may lead to slow or complex dynamics. Reduction should lead to faster and stronger DSD reactions, and more simple, predictable behaviour. Testing this hypothesis, however, could be extremely costly. The B30-G6 duplex used in the fluorescent DSD reactions in Figures 3.16 and 3.17 comprised of two strands - an operator-toehold strand modified by a 5'

fluorophore and a complementary operator strand modified by a 3' quencher. Both these modifications are expensive and require the synthesised oligo to undergo HPLC purification, further increasing the cost of the component strands more than 10-fold greater than unmodified counterparts. Comparison and optimisation of many such probe designs would be prohibitively expensive, and also take much more time, if new modified oligos had to be ordered for each test. A new probe design was therefore necessary to enable higher throughput and cheaper probe development.

	X1	X2	X3	X4	X5	X6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
sB30c 5'-							T	A	A	T	C	A	**A**	**A**	**A**	**T**	**A**	**A**	**A**	**T**	T	G	**A**	**T**	**T**	**T**	**A**	**T**	**T**	**T**	G	C	T	T	G	C	-3'
X6-sB30c 5'-	C	C	A	G	T	G	T	A	A	T	C	A	**A**	**A**	**A**	**T**	**A**	**A**	**A**	**T**	T	G	**A**	**T**	**T**	**T**	**A**	**T**	**T**	**T**	G	C	T	T	G	C	-3'
sCL30c 5'-							T	C	C	A	C	T	A	T	**A**	**T**	**A**	**T**	T	G	A	C	G	A	A	T	G	T	C	G	**A**	**T**	**A**	**T**	T	G	-3'
X6a-sCL30c 5'-	G	C	C	C	G	C	T	C	C	A	C	T	A	T	**A**	**T**	**A**	**T**	T	G	A	C	G	A	A	T	G	T	C	G	**A**	**T**	**A**	**T**	T	G	-3'
sCR30c 5'-							C	T	T	T	C	G	G	T	**A**	**T**	**A**	**T**	T	G	A	T	A	C	C	T	G	T	C	G	**A**	**T**	**A**	**T**	T	C	-3'
X6a-sCR30c 5'-	G	C	C	C	G	C	C	T	T	T	C	G	G	T	**A**	**T**	**A**	**T**	T	G	A	T	A	C	C	T	G	T	C	G	**A**	**T**	**A**	**T**	T	C	-3'

Figure 4.3: **X strand toehold sequence can affect overall oligo structure.** Alignment of the sB30c, sCL30c and sCR30c operator sequences, alongside their X strand counterparts, with highlighting to demonstrate predicted secondary structure from Figure 4.2. Green colouring corresponds to the regions likely to form hairpins; within these blue colouring highlights G-T wobble base pairing. Bold lettering indicates the inverted repeats identified either by Sato *et al.* [124] or by Figure 4.1 to possibly be most important for repressor binding.

4.1.3 Tprobe design enables fast and cheap DSD testing

One such system was pioneered in 2015 by Wang *et al.* [65]. In order to improve the sensitivity of DSD methods testing for single-nucleotide allele variants, they designed an 'X-Probe' - a four-strand DNA complex comprising of two 'universal' fluorophore- and quencher-modified oligos and two variable oligos - displayed in Figure 4.4 (a). The universal strands were formed from separate 21nt sequences enabling hybridisation to the variable strands, and a 7nt sequence allowing the universal strands to anneal to each other and form a double-stranded domain with a terminal fluorophore and quencher adjacent to each other, on anti-parallel strands as in a basic linear DSD probe. The variable strands comprised of (blue) adapter sequences, enabling annealing to the universal strands, and a (green) sequence that formed a double-stranded recognition domain with a single-stranded toehold. This design enabled a large quantity of each modified oligo to be synthesised and purified, and different allele targets to be probed by designing new, unmodified variable strands. Aliquots of the universal and variable strands could then be annealed together

to form functional X-Probes, as long as each of the four double-stranded sequences were sufficiently distinct to prevent other structures from forming.

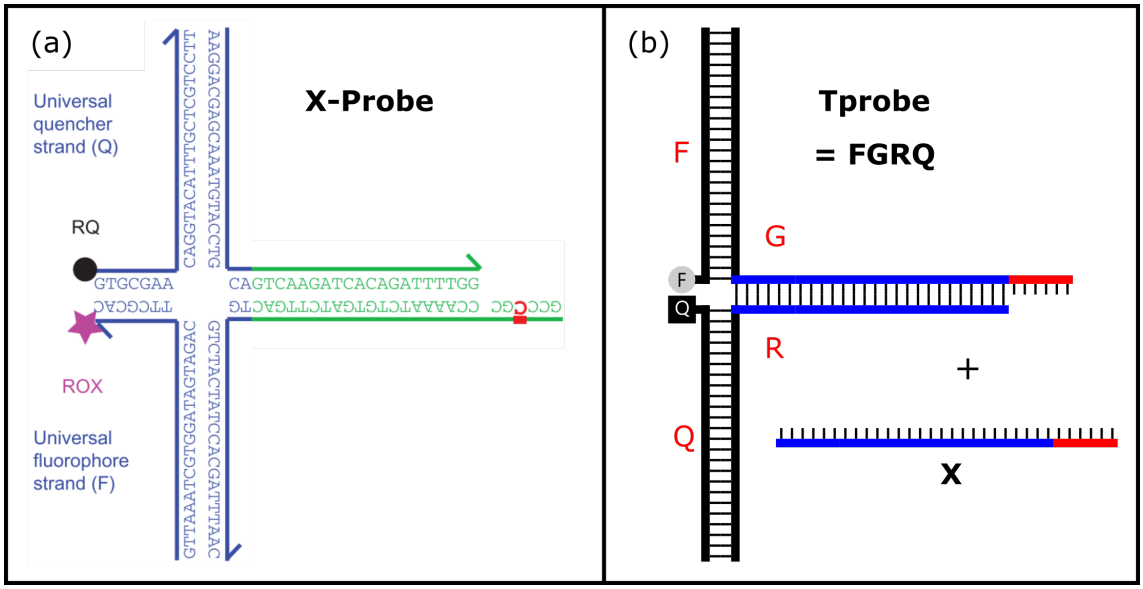


Figure 4.4: **Tprobe design is based on published X-Probe.** (a) X-Probe from Wang *et al.* [65], with universal sequences in blue, variable sequences in green and single-nucleotide variant being targeted in red. (b) Tprobe design, with the universal sequences in black, variable recognition domain in blue and variable toehold sequence in red. The F, G, R and Q strands are labelled and the displacing X oligo is shown alongside.

One downside to this design is that for the fluorophore and quencher to separate, not only must there be displacement of the green recognition domain, but this must cause the 7bp domain adjacent to the modifications to separate too. A subsequent update to the X-Probe design, used for estimating DNA hybridisation rates in collaboration with Microsoft Research, removed this short arm of the probe, thereby placing the modifications at the probe junction itself, opposing one another rather than being adjacent [149]. This formed the basis for the 'Tprobe' design used in this project, shown in Figure 4.4 (b), with full reaction mechanism displayed in Figure 4.5.

The Tprobe, named after the rough shape of the complex compared to the four-arm X-Probe, also contains four strands. Each of these is labelled with a letter related to the strand's function - F (fluorophore), G (gate), R (recognition) and Q (quencher), while the X displacing oligo is labelled as such from notation used by Soloveichik *et al.* [55]. Annealed together they form the FGRQ complex (Tprobe), with three double-stranded domains (F, Q and R) and a gate (G) toehold. The products of the Tprobe reacting with X are shown in Figure 4.5 (a) - named FGX and QR after their component strands. FGX is fully double-stranded, but with

a nick between the F and X strands where the fluorophore modification resides, while QR is only partially double-stranded.

The schematics in Figure 4.5 show that the new probe design should be able to take part in DSD reactions, with or without binding to a repressor, in the same manner as the original linear probe. The use of a four-strand complex does, however, introduce some new considerations. To minimise background signal it is important to prevent the occurrence of free F oligo, or any probe complex containing the F strand but not the Q strand, which would otherwise be fluorescent regardless of DSD state. It is therefore important to anneal the component oligos together with clockwise excess, using the structure in Figure 4.4 (b) for reference, meaning the following hierarchy of concentrations: F<G<R<Q. Two ratios of F:G:R:Q were used in this project - 1:1.2:1.44:1.78 (lowQ: a constant 20% excess at each strand) or 1:1.5:3:5 (highQ: as used in Wang *et al.* [65]). These ensure (to different degrees of certainty), that each F strand is in an FGRQ complex, while likely forming some GRQ and RQ complexes and free Q.

The cost of this insurance is a slight reduction of sensitivity - any G not in a Tprobe (free G, GR or GRQ) can undergo DSD without associated increase in fluorescence. Also in the case of this project, any GR or GRQ complexes would contain functional operator sequences able to be bound by the repressor, whose arsenic-induced dissociation would also not result in fluorescence increase post-DSD. These slight reductions in sensitivity were deemed offset, however, by the time and cost savings not re-ordering modified oligos or attempting to purify and quantify stable complexes.

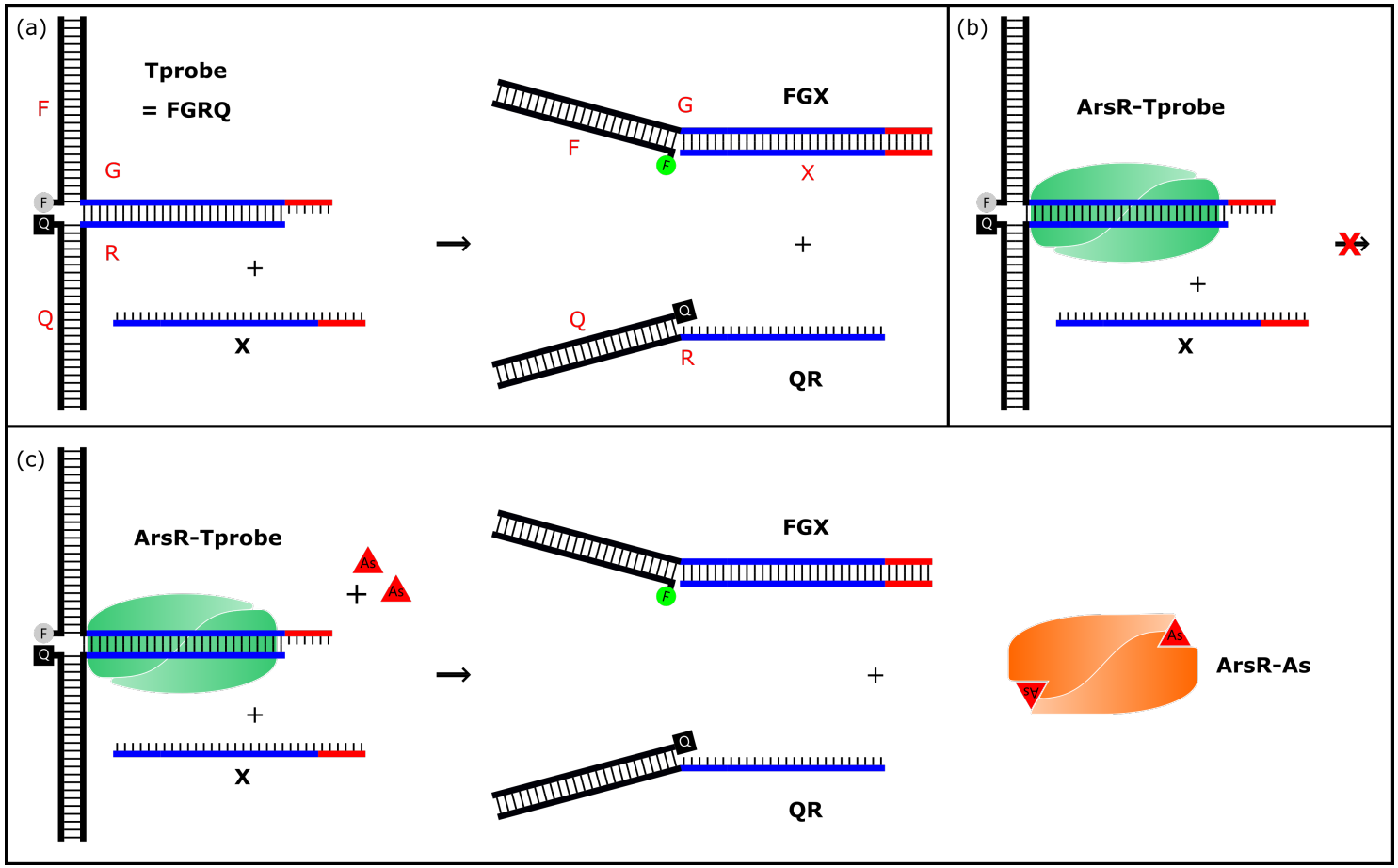


Figure 4.5: **Basic Tprobe DSD reaction.** Schematic illustrating the structure of a Tprobe and how it undergoes a basic irreversible DSD reaction. (a) The Tprobe is comprised of F, G, R and Q DNA strands (labelled in red), the first and last of which are modified by a fluorophore and quencher, respectively. When reacted with displacing oligo X, the products FGX and QR are formed. FGX is fluorescent as its fluorophore no longer in the proximity of a quencher modification. (b) ArsR binds its operator sequence within the Tprobe (blue dsDNA), preventing DSD from occurring. (c) Arsenic induces dissociation of ArsR from its operator sequence, thereby enabling the DSD reaction.

4.2 Establishing CgArsR DSD system

4.2.1 TR30-G6a probe has high background displacement

Once this Tprobe design was confirmed, an initial construct was annealed using the CR30 operator and 6a toehold. Figure 4.6 shows the fluorescence time course of this probe (TR30-G6a) under three conditions. The base signal, with just the probe incubated with repressor, is shown in green. The black and red traces show parallel samples with X6a-sCR30c displacing oligo added after 30 minutes, with 1 µl water or 1 mM sodium arsenite added 15 minutes later.

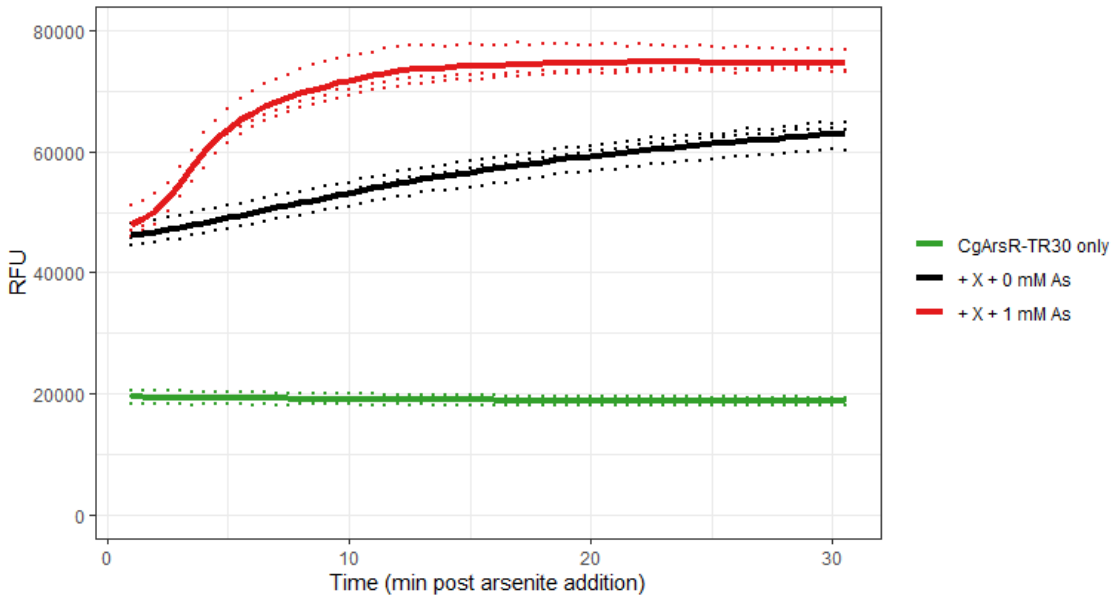


Figure 4.6: **Fluorescent CgArsR-TR30-G6a system has high background DSD.** Fluorescence time course displaying the RFU for samples with or without displacing oligo or arsenite. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each sample contained 10 pmol TR30-G6a and 200 pmol CgArsR co-incubated for 30 min before addition of 48 pmol X6a-sCR30c, a further 15 min incubation then addition of 1 µl sodium arsenite at the specified concentrations, for a total volume of 10 µl per well of 384-well plate.

This result indicated the presence of a large background effect, with the black traces starting at a far higher RFU than the DSD-free green samples. This background continued to increase over the half hour assay. The red samples, with arsenite added one minute before the first read, started at a similar RFU before increasing in signal much faster and plateauing at around 75k RFU.

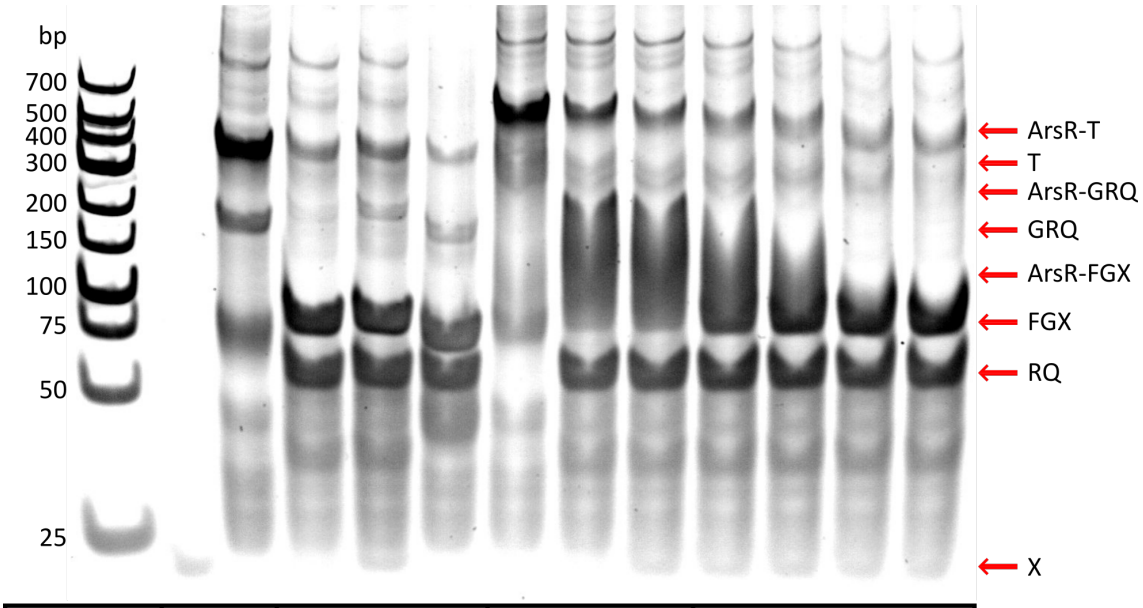
These data suggested that while arsenite addition was able to cause an increased rate in DNA strand displacement, arsenite was not necessary for DSD to occur

and the repressor in these samples could only slow rather than prevent DSD from occurring as intended. In the 15 minutes prior to arsenite addition, the X6a-sCR30c oligo seemed to have caused DSD of the Tprobe despite the presence of CgArsR, but this displacement was subsequently accelerated by arsenite addition, suggesting non-functional repressor was not the cause of this background.

An EMSA was run to confirm the performance of CgArsR and understand the behaviour of components in more detail. As Tprobes form more complex three-dimensional structures than usual double-stranded DNA, and would have a variable cross-section unlike the 'rod' model of a simple double helix, the Tprobes migrated very slowly through 20% polyacrylamide gels. For EMSAs on these species, therefore, 10% polyacrylamide gels were used - these were found to allow good Tprobe migration as well as discrimination between structures bound or unbound by CgArsR.

Figure 4.7 shows the result of this EMSA. As expected, many different DNA species were present in all the lanes containing the Tprobe, due to the excesses of the four strands present and the many possible interactions these could have other than formation of the desired complex. Lane two shows the annealed products by themselves - the three dominant species are the Tprobe, migrating roughly as 400bp dsDNA, the GRQ complex, migrating as just under 200bp, and the RQ duplex, migrating as a roughly 75bp smear. Upon addition of the X6a-sCR30c displacing oligo (lanes 3 and 4), most of the Tprobe and GRQ complexes reacted to form the dominant FGX and RQ product bands as expected. Slightly different band migrations were seen for these reactions compared to the pre-annealed products (lane 5), but as the reactions may not have reached a full thermodynamic equilibrium small differences in the exact distribution of DNA species would be expected. Also, the RQ band in the unreacted Tprobe lane 2 seemed to migrate slightly slower than those in others, but as much of this complex was single-stranded its exact conformation could be influenced by the presence of the X oligo.

Lane 6 shows the Tprobe solution incubated with CgArsR, resulting in both the Tprobe and GRQ complexes displaying shifted migration, indicative of being bound by the repressor through their complete CR30 operator domains. The RQ duplex did not shift as it only contained a single-stranded operator sequence. Upon addition of X6a-sCR30c (lanes 7 and 8) much of the Tprobe stayed shifted,



TR30-G6a	-	10	10	10	10*	10	10	10	10	10	10	10
X6a-sCR30c	12	-	12	24	24*	-	12	24	24	24	24	24
CgArsR	-	-	-	-	-	100	100	100	100	100	100	100
As(III)	-	-	-	-	-	-	-	-	100	200	500	1000
Lane	1	2	3	4	5	6	7	8	9	10	11	12

Figure 4.7: **Testing a Tprobe CR30-CgArsR DSD reaction.** 10% PAGE EMSA testing the ability of CgArsR to prevent a Tprobe DSD reaction, and of arsenite to subsequently relieve this repression. The TR30-G6a probe and CgArsR were incubated together for 20 min before addition of the X6a-sCR30c displacing oligo, then incubated for a further 20 min before sodium arsenite addition for a total volume of 10 µl, in the indicated quantities given in pmol, before a final 20 min incubation. The asterix for the components in lane 5 denotes they were pre-annealed together instead of added separately. The bands corresponding to the ArsR-T, T, ArsR-GRQ, GRQ, ArsR-FGX, FGX, RQ and X species are indicated on the right.

however a large fraction reacted to form FGX and RQ despite no arsenite addition. In these lanes, FGX presented as a smear as it could also be bound by CgArsR, in addition to Tprobe or GRQ. Presence of arsenite (lanes 9-12) removed this behaviour, converting the smear into a discrete band, while enabling the remainder of the Tprobe to undergo DSD.

These results confirmed that the repressor was functional in these assays - it could bind its operator and dissociate with arsenite - and this included binding the Tprobe as intended. However, CgArsR clearly could not prevent DSD initiated by X6a-sCR30c and so this reaction occurred despite the lack of arsenite, with the repressor subsequently free to bind the only species present with its full dsDNA operator - FGX.

One protocol alteration that could reduce the effect of this background

displacement would be to swap the order of X oligo and arsenite addition. If the CgArsR-Tprobe mix is pre-incubated with different concentrations of arsenite, different concentrations of unbound Tprobe would be expected to form. As Figure 4.6 demonstrated that bound Tprobe did have a noticeably slower DSD rate than unbound Tprobe, the exact rate of signal increase should be indicative of the concentration of added arsenite.

Figure 3.24 at the end of the previous chapter showed the result of such a rearranged assay. Here, all the samples except for the highest arsenite concentration (1 mM) initially have an RFU of around 20k, in line with the base signal of the bound probe seen in Figure 4.6. From here they all increase quickly, with the trace for the background arsenite-free sample for the 15 to 45 minute time-frame (from roughly 40k to 60k RFU) matching the increase seen in the 30 minutes of the Figure 4.6 assay, which started 15 minutes after arsenite addition.

While this reversal of addition order enabled greater clarity of the underlying DSD behaviour, it is not a practical arrangement for a potential biosensor. The true time for the assay would still start from the moment the unknown samples would be added, and requiring a subsequent addition of the displacing oligo (manual or automated) would increase the chance of misuse or at least introduce unnecessary sources of variation. It was therefore deemed necessary to remove, or at least greatly limit, this background DSD and to understand its underlying mechanism.

Several possible explanations were hypothesised for why this behaviour, with the DNA-strand displacement able to overcome the blocking effect of the repressor, was seen so clearly in this system but less for BsArsR, as seen in Figure 3.16. These centred around the possibility of toehold-mediated strand displacement while CgArsR was bound to the probe and reasons for a weakened repressor-operator interaction:

- The affinity of BsArsR for its operator is naturally much greater than that for CgArsR and its operator
- The CR30 operator for CgArsR was not ‘minimal’ enough, unlike the B30 operator, and so the X6a-sCR30c oligo was able to access the TR30-G6a toehold even when CgArsR was bound, while access to the B30-G6 toehold was more sterically hindered by BsArsR for displacement by X6-sB30c

- The Tprobe structure interferes with CgArsR binding
- The X6a-sCR30c toehold was stronger (higher GC content) than the X6-sB30c toehold
- The structure of the X6a-sCR30c oligo enabled higher background than the X6-sB30c oligo

The first of these potential explanations seemed unlikely - if anything the EMSA data, and the dissociation constant estimation by Ordóñez *et al.* [130], suggested the CgArsR-CR30 interaction was stronger than that for the BsArsR-B30 complex. Each of the remaining hypotheses was investigated in order to gain a deeper understanding of the important factors affecting behaviour of repressor-operator DSD systems, as well as to improve performance of the CgArsR arsenic biosensor.

4.2.2 The minimal CgArsR operator is roughly 26bp, but its use does not prevent background DSD

To determine if a smaller CR operator could be used, an EMSA was run testing CgArsR binding to truncated CR30 duplexes. Shortened sCR30 oligos (named sCRx-y depending on their 5' x and 3' y positions from sCR30) were annealed to sCR30c, creating partially single-stranded DNA species, and incubated with CgArsR before run on a gel. Figure 4.8 shows the result of this experiment. While CR1-28 and CR1-26 displayed similarly stronger shifts to CR30, both CR1-24 and CR3-30 seemed to have reduced affinity to CgArsR. CR5-30 and CR1-22 (run on a separate gel) were unable to be retarded by CgArsR to any degree, showing major loss of binding affinity.

This result was interesting because not only did it show that a more minimal 26bp operator could be used (CR26 is the main operator used from this point onwards - this is CR1-26 without the ssDNA overhang), but also that this was asymmetric within the original CR30 sequence. While the loss of base 2 did reduce CgArsR affinity it did not abolish it, unlike base 23, both of which were previously suggested as important due to their conservation with CL30.

These data are summarised in Figure 4.9, highlighting how the 4bp at the 3' end of CR30 seem unnecessary for function. Bases 3-24 therefore form the core

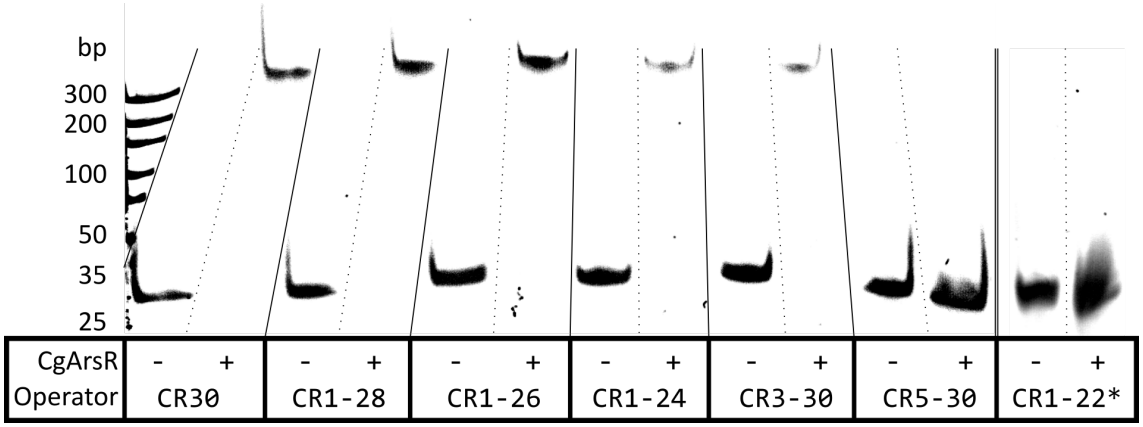


Figure 4.8: **Identifying the minimal CR operator length.** 20% PAGE EMSA testing the ability of CgArsR to bind truncated CR30 duplexes of different lengths. DNA species were formed by annealing the sCR30c oligo with the sCR1-28, sCR1-26, sCR1-24, sCR3-30, sCR5-30 or sCR1-22 oligos, respectively, with the numbers indicating the 5' and 3' bases of the sCR30 sequence from which it derived. The duplexes (other than the fully double-stranded CR30) therefore had a dsDNA core (22-28bp) with a single-stranded tail on one end (2-8nt). 15 pmol of each annealed duplex was incubated with 200 pmol CgArsR at room temperature for 30 min in a binding buffer containing 5 mM DTT before being run on the gel. A small leak within the tank apparatus caused the lanes to diverge and part of the ladder to be lost - rough lane boundaries are annotated onto the gel image. The CR1-22 samples (*) were run on a separate gel following the result of this EMSA, but are included alongside for completeness; both gels were run by B. Baker.

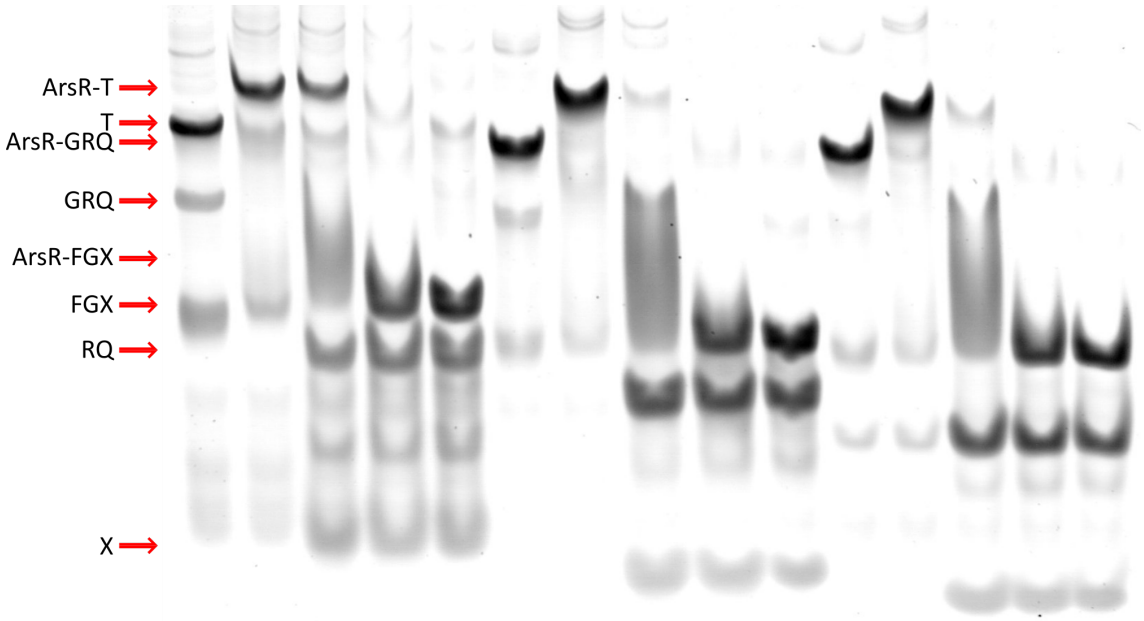
operator, while bases 1-2 and 25-26 increase the affinity for CgArsR, perhaps by providing structural support than important base-specific interactions. This prompted further experiments probing the sequence affinity space of the CR operator, described later in this chapter.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
sCR30	5'-	G	A	A	T	A	T	C	G	A	C	A	G	G	T	A	T	C	A	A	T	A	T	A	C	C	G	A	A	A	G	-3'
sCR30c	3'-	C	T	T	A	T	A	G	C	T	G	T	C	C	A	T	A	G	T	T	A	T	A	T	G	G	C	T	T	T	C	-5'

Figure 4.9: **Only the 5' 26bp of CR30 are necessary for normal CgArsR binding.** Full sequence of both strands of the CR30 dsDNA, with the light colouring showing conservation with CL30 as in Figure 4.1. Overlaid dark colouring for sCR30 sequence highlights the results from Figure 4.8, with bases that can be lost without noticeably affecting CgArsR affinity in green, bases that slightly reduce affinity when lost in yellow, and bases that abolish CgArsR affinity when lost in red.

Following this work, Tprobes were designed to investigate whether the use of a shorter operator reduced background DSD, through increased occlusion of the toehold. If, when using the TR30-G6a probe, the 6nt toehold is fully exposed and accessible by the X displacing oligo even when CgArsR is bound, DSD may be able to occur at a slower albeit appreciable rate if the CgArsR-CR30 affinity

was outweighed by the thermodynamics of the DSD reaction. Bringing the toehold closer to the core operator by removing unnecessary 3' dsDNA may therefore make DSD less likely when CgArsR is bound and so would reduce the background rate. Tprobes containing the CR26 and CR24 operators, removing 4/6bp from the dsDNA adjacent to the G6a toehold respectively, were compared to TR30-G6a via EMSA below.



X6a-sCR*nc*	-	-	48	48	48	-	-	48	48	48	-	-	48	48	48
CgArsR	-	200	200	200	-	-	200	200	200	-	-	200	200	200	-
As(III)	-	-	-	2000	-	-	-	-	2000	-	-	-	-	2000	-
Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Tprobe	TR30-G6a					TR26-G6a					TR24-G6a				

Figure 4.10: **Comparison of Tprobe DSD reactions with shorter operators.** 10% PAGE EMSA testing the ability of CgArsR to prevent DSD reactions of Tprobes with 26bp or 24bp operators compared to the 30bp operator. The Tprobes (all 10 pmol) and CgArsR were incubated together for 25 min before addition of X6a-sCRnc displacing oligo and sodium arsenite together for a total volume of 10 µl, in the indicated quantities given in pmol, before a final 15 min incubation. The displacing oligos used were the full length for each respective Tprobe, hence the use of variable *n*. The bands corresponding to the ArsR-T, T, ArsR-GRQ, GRQ, ArsR-FGX, FGX, RQ and X species are indicated on the left for the TR30-G6a reaction - the band identities for the other Tprobes follow the same pattern but are shifted slightly lower.

Figure 4.10 compares each of these three Tprobes under five conditions - on their own, with CgArsR, with CgArsR and their respective X oligo, with CgArsR, X and arsenite, and the uninhibited Tprobe + X reaction. First, lanes 2, 7 and 12 show all these Tprobes can be bound by CgArsR, while lanes 5, 10 and 15 demonstrate that without CgArsR all the probes would react fully with their X oligos within the preceding 15 minute incubation. Lanes 3, 8 and 13 show that this DSD is

reduced/slowed if each of the Tprobes has been pre-incubated with CgArsR, but the shorter operators probes do not appear to reduce this background as hoped. If anything these shorter operators have less CgArsR-Tprobe maintained when X added (lanes 8/13 versus 3) indicating a potentially higher background rate. The addition of arsenite (lanes 4, 9 and 14) does increase the DSD to replicate the CgArsR-free situation as expected.

Unfortunately this result showed that the background DSD rate was not reduced simply by trimming the 3' end of the operator. It is possible that with all the operator lengths assayed the toehold was so close to the dsDNA operator, and thereby the bound repressor, that it was fully protected from DSD until CgArsR dissociated. Alternatively the opposite scenario could be true and the toehold was not sufficiently hidden even with the TR24 probe to make a noticeable difference to the DSD rate. Without knowing the likelihood or prevalence of such a DSD mechanism, which may require detailed crystal structures to ascertain the exact degree of toehold access, it would be difficult to suggest one or the other, but reducing operator length did not help reduce background DSD in this case.

Figure 4.10 did in fact suggest the shorter operators had a higher background DSD rate than TR30-G6a. One explanation could be that they had lower affinity for CgArsR - for TR24-G6a this would be consistent with the weaker CR1-24 shift seen in Figure 4.8. Reduced repressor affinity could enable faster DSD, but the worse background may also be down to differences in X oligo structure. As shown in Figure 4.11, the X oligos for the shorter operators form less stable structures (MFE of −4.88 kcal/mol for X6a-sCR26c and −3.20 kcal/mol for X6a-sCR24c) than for the full operator (−6.08 kcal/mol for X6a-sCR30c) so may be expected to react faster with their respective Tprobes. To add a further complication, under the modelled conditions 7% of X6a-sCR24c was predicted to form a strong dimer (all other X oligos preferring internal hybridisation so far) so simply comparing the monomeric minimum free energy (MFE) values would be imperfect. Again, this highlights the difficulty in making conclusive comparisons between sequence alterations when this results in structural changes to their single-stranded counterparts.

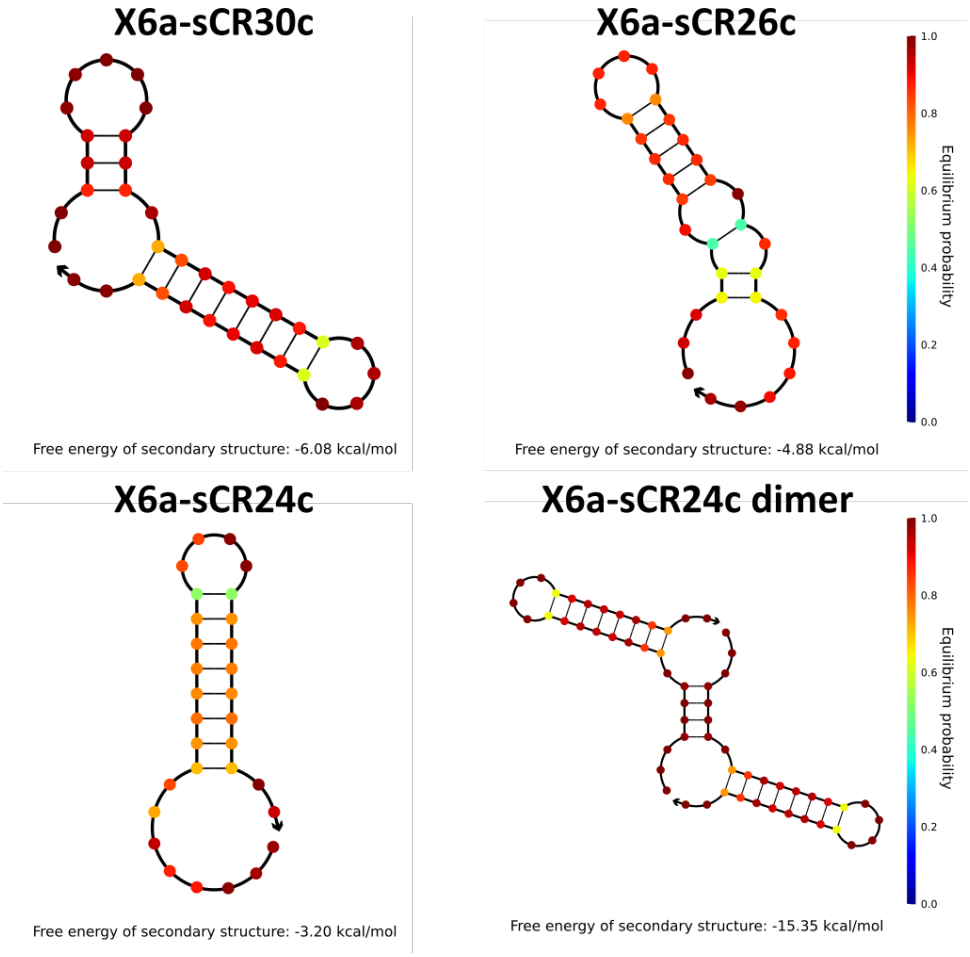


Figure 4.11: **Shortening native CR operator alters X oligo structures.** Minimum free energy secondary structures formed by single-stranded displacing oligos comprised of the sCR30c/sCR26c/sCR24c operators downstream of the X6a toehold. The structures and the associated MFEs were predicted using NUPACK at a concentration of 1 µM and a temperature of 25 °C, with individual bases coloured according to the probability of being in the suggested state using the scale on the right. The 3' end of each oligo is labelled with an outward arrow. Under the specified conditions 7% of the X6a-sCR24c oligo was predicted to form the dimer shown instead of its monomeric structure.

Finally, a fluorescence assay was carried out to make more accurate comparisons between the background rates of these variable length Tprobes. Figure 4.12 shows the result.

This suggested that despite the differences in X oligo structure, the background rates were even more similar than suggested in Figure 4.10. The combination of these assays demonstrated that shorter operators do not seem to offer greater DSD protection, reducing the likelihood that X oligo access to the Tprobe while CgArsR is bound is the reason for high background rate, but at least showing that this intervention was not beneficial.

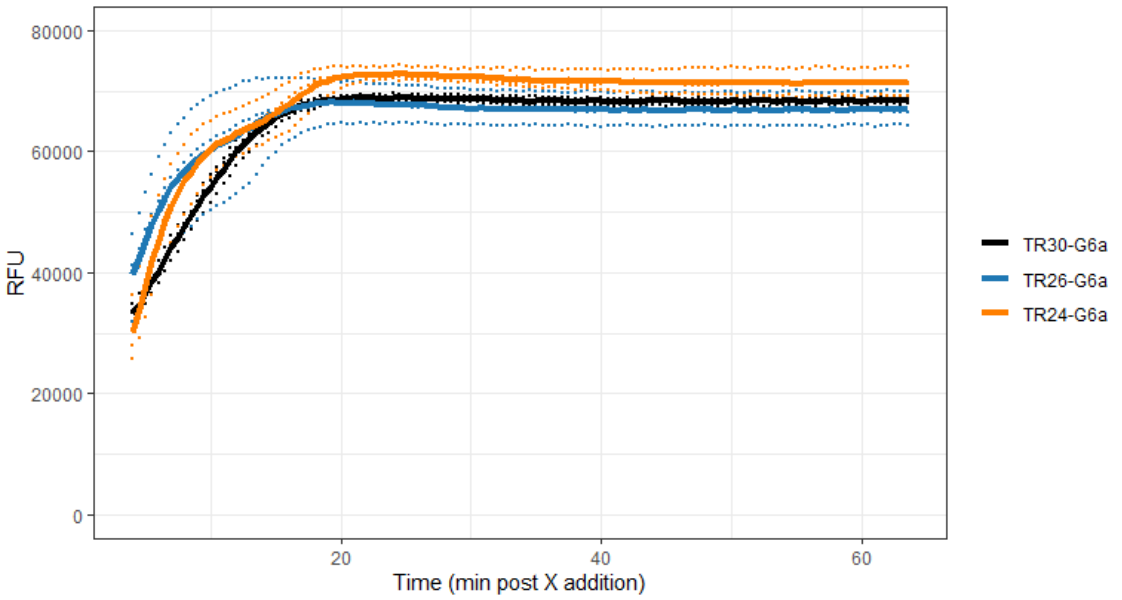


Figure 4.12: **Shortening native CR operator does not reduce DSD background rate.** Fluorescence time course displaying the RFU for TR30-G6a, TR26-G6a and TR24-G6a samples with CgArsR reacting with displacing oligo. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each sample contained 10 pmol Tprobe and 100 pmol CgArsR co-incubated for 1 h before addition of 48 pmol X6a-sCRnc (n denoting full length for each Tprobe), for a total volume of 10 µl per well of 384-well plate.

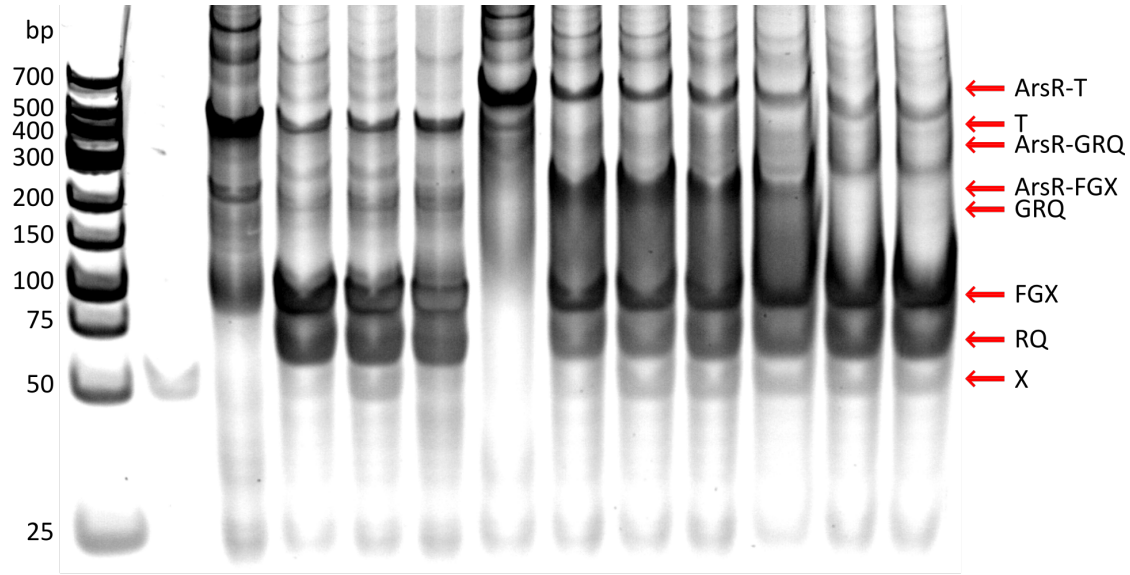
4.2.3 Tprobe structure does not impede CgArsR binding

The next hypothesis to test was if the general Tprobe structure interfered with CgArsR's ability to bind its operator, thereby reducing the repressor-operator affinity and explaining the high background rate of TR30-G6a compared to the linear B30-G6 probe. The theory was that with the 'vertical' fluorophore and quencher arms of the probe directly adjacent to the 30bp operator, this may sterically hinder CgArsR binding.

In order to test this, a Tprobe was designed with a 24bp spacer in between the modified arms and the operator. To prevent any chance of CgArsR binding this upstream region, the 24bp extension was comprised of the reverse complement of the first 18bp of the native CgArsR coding sequence and the 6bp directly upstream of CR30. This was effectively the 24bp preceding CR30 as shown in Figure 3.20, removing CL30 and a few bases either side - these chosen regions had direct evidence from Ordóñez *et al.* that CgArsR did not bind them.

An EMSA was run on this TR54-G6a probe, analogous to the TR30-G6a EMSA in Figure 4.7, with the result shown in Figure 4.13. Once again Tprobe formation

and DSD upon addition of X oligo without CgArsR is clear (lanes 1-5), although inevitably there is a greater prevalence of minor bands and background smears due to the larger number of possible non-specific interactions with this longer probe. The Tprobe is shown to shift with CgArsR (lane 6), but once again much of the ArsR-Tprobe complex reacts when X oligo is introduced, forming significant quantities of RG, FGX and ArsR-FGX (lanes 7 and 8). These bound species reduce upon arsenite addition (lanes 9-12), but sensitivity has already been limited by the high background displacement.



TR54-G6a	-	10	10	10	10*	10	10	10	10	10	10	10
X6a-sCR54c	12	-	12	24	24*	-	12	24	24	24	24	24
CgArsR	-	-	-	-	-	100	100	100	100	100	100	100
As(III)	-	-	-	-	-	-	-	-	100	200	500	1000
Lane	1	2	3	4	5	6	7	8	9	10	11	12

Figure 4.13: **Testing a CgArsR DSD reaction with TR54-G6a.** 10% PAGE EMSA testing the ability of CgArsR to prevent a TR54-G6a DSD reaction, and of arsenite to subsequently relieve this repression. The TR54-G6a probe and CgArsR were incubated together for 20 min before addition of the X6a-sCR54c displacing oligo, then incubated for a further 20 min before sodium arsenite addition for a total volume of 10 µl, in the indicated quantities given in pmol, before a final 20 min incubation. The asterix for the components in lane 5 denotes they were pre-annealed together instead of added separately. The bands corresponding to the ArsR-T, T, ArsR-GRQ, GRQ, ArsR-FGX, FGX, RQ and X species are indicated on the right.

This similarity in the behaviour of the TR54-G6a and TR30-G6a probes suggested that hindrance of CgArsR binding by the Tprobe arms was not a significant factor in the high background DSD rate seen in the reactions so far. Later assays therefore utilised the more minimal probe designs to limit the chances of non-specific interactions and formation of unintended structures.

4.2.4 Reducing toehold strength does reduce DSD background, but not completely

The final hypotheses to test involved assessing the effect of altering the displacing oligo structure and toehold strength. The X6 toehold used for B30 was moderate (CCAGTG), while the X6a toehold was very strong (GCCCGC) - it was very possible that this alteration affected the background DSD rate more than anticipated. New probes were designed to test the effect of altering toehold strength, utilising either the weak X6b toehold (ATTTAT) or the moderate (strong but short) X4a toehold (CCGC).

Figure 4.14 shows the result of DSD reactions of both TR26 and TR24 operator probes using these toeholds, incubated with CgArsR, as well as the X6a reactions seen in Figure 4.12. This clearly shows that the weaker X6b and X4a toeholds significantly reduced the background rate of displacement, with none of their reactions reaching saturation within the hour-long time course. This contrasted strongly with the reactions using the 6a toeholds, which saturated within 20 minutes.

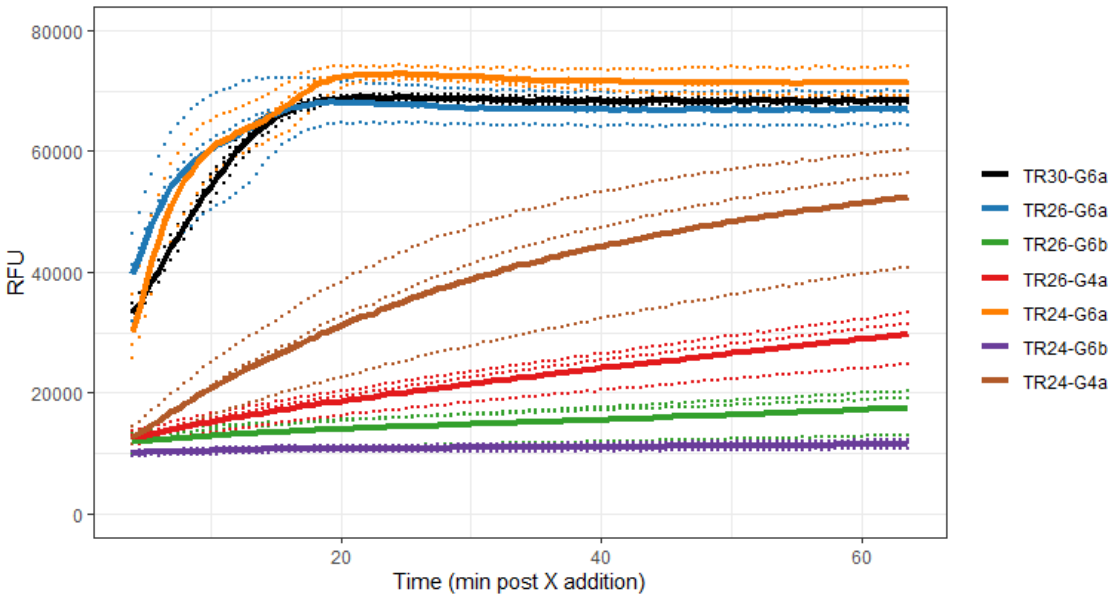


Figure 4.14: **Reducing toehold strength does reduce DSD background rate.** Fluorescence time course displaying the RFU for TR30-G6a, TR26-G6a, TR26-G6b, TR26-G4a, TR24-G6a, TR24-G6b and TR24-G4a probes with CgArsR reacting with their displacing oligos. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each sample contained 10 pmol Tprobe and 100 pmol CgArsR co-incubated for 1 h before addition of 48 pmol X-sCRnc (n denoting full length for each Tprobe; X denoting each respective toehold), for a total volume of 10 µl per well of 384-well plate.

The CgArsR-TR24-G6b probe, reacting with X6b-sCR24c, showed the lowest background rate, with almost no displacement occurring during the assay (<3% relative to the 6a equilibrium RFU). The CgArsR-TR26-G6b probe displayed roughly 12% average DSD, while the CgArsR-TR26-G4a and CgArsR-TR24-G4a probes reacted on average 33% and 71%, respectively. This suggested that the 6b toehold was weaker than the 4a toehold, despite its greater length. Interestingly, the operator length did substantially affect the DSD rate unlike with the 6a toeholds, but not in a consistent manner. For the 4a toehold, combining with the TR24 operator gave a much higher background rate than the TR26 operator, while the opposite was true when using the 6b toehold. The X6b-sCR26c oligo had a lower MFE and a structure obscuring the toehold more than the X6b-sCR24c oligo, both opposite to usual explanations for a faster DSD rate. This was also true of the X4a-sCR24c versus X4a-sCR26c comparison - again the more structured and obscured oligo reacted faster against expectation.

While the lower background DSD rates using these toeholds was promising, slowing displacement when CgArsR is bound was only desirable if not at the expense of speed when CgArsR was not bound. The ideal scenario envisaged for this system would be no displacement until CgArsR dissociated, after which DSD would be rapid - this would result in the fastest and most sensitive response to arsenite.

Figure 4.15 displays the result of repressor-free Tprobe reactions with displacing oligos for the alternative toeholds. While the TR30-G6a probe could be fully displaced in under 5 minutes, these other Tprobes were much slower. Both the TR24-G4a and TR26-G4a probes peaked after 15-20 minutes, while neither of the 6b probes were fully displaced after 45 minutes. By the end of this assay, TR26-G6b had reacted roughly 62% while TR24-G6b had only reacted an incredibly low 10%.

These results demonstrated that DSD rates of free Tprobe and bound Tprobe did largely correlate, and altering toehold strength alone seemed unlikely to prevent background displacement completely while enabling fast DSD when desired. The very slow rate of the 6b oligos reacting with free probe was clearly undesired, as it would severely limit speed of response to arsenite detection, even if background rates of bound probe DSD were low.

The speed of the 4a toeholds looked more promising, although still much slower than the stronger 6a toeholds. Interestingly, the TR26-G6a probe displayed a

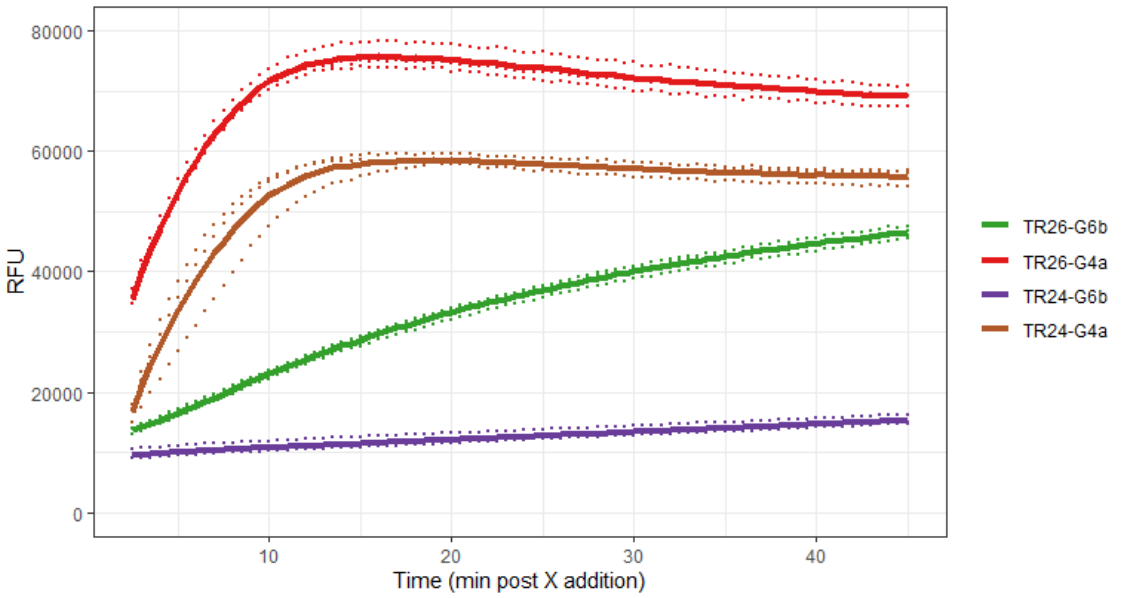


Figure 4.15: **Reducing toehold strength also slows maximum DSD rate.** Fluorescence time course displaying the RFU for TR26-G6b, TR26-G4a, TR24-G6b and TR24-G4a probes without CgArsR reacting with their displacing oligos. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each sample contained 10 pmol Tprobe reacting with 48 pmol X-sCRnc (n denoting full length for each Tprobe; X denoting each respective toehold), for a total volume of 10 µl per well of 384-well plate.

faster initial DSD rate than the shorter operator and while they both plateaued at roughly the same time, the latter settled at an unexpectedly low RFU. The relative initial speeds of these 4a toeholds was therefore the reverse of repressor-bound situation. Together, with free TR26-G4a reacting only slightly slower than the 6a probes but with a much lower bound probe background rate, these experiments suggested that using TR26-G4a instead of TR30-G6a may improve sensitivity to arsenite.

4.2.5 Increasing volume of arsenite sample addition increases sensitivity

A couple of other protocol alterations were trialled at this stage to test their effect on arsenite response speed and sensitivity as well as control of background DSD rate. In the fluorescence plate reader assays so far, total volume had been kept at 10 µl per well, which resulted in most Tprobe reactions containing only 1 µl of arsenite sample each. The wells within the 384-well plates used could accommodate up to 20 µl of solution and so larger volumes were trialled to assess whether this improve biosensor performance.

In theory, increasing the volume of arsenite sample added should elevate the ratio of arsenite to repressor and probe within the solution and thereby increase sensitivity. However, this would also have the trade-off of diluting all the other components, affecting salt concentrations, CgArsR binding, DSD speed and raw fluorescence values. Without a complete model of this system such compromises are hard to predict and so are best tested empirically.

Preliminary results had shown that this doubling in volume only had a minor effect on total RFU, so assays were carried out comparing the behaviour of TR30-G6a and TR26-G4a under these conditions. Figure 4.16 shows the result of this, with TR30-G6a in (a) and TR26-G4a in (b). These clearly show the effect of the diluted solutions and the differing behaviours between the probes. Previously, 10 pmol TR30-G6a probe incubated with 100 pmol CgArsR underwent full DSD within 20 minutes (Figure 4.12) but this background rate was greatly slowed under the diluted conditions, not quite saturating within the hour-long assay. The samples responding to arsenite show complex (and highly variable) dynamics, but the two highest concentrations were both distinguishable from the background early on - this detection of 10 µM sodium arsenite was greatly improved from the original assay in Figure 3.24. Towards the end of the assay even the 1 µM sodium arsenite samples seem to be separating from the background, albeit with high variation limiting the significance of this response.

The TR26-G4a probe also displayed a reduction in background DSD due to the dilution. In this case, almost no displacement of the arsenite-free samples occurred across the hour, starkly reduced from the 33% seen with the original volumes in Figure 4.14.

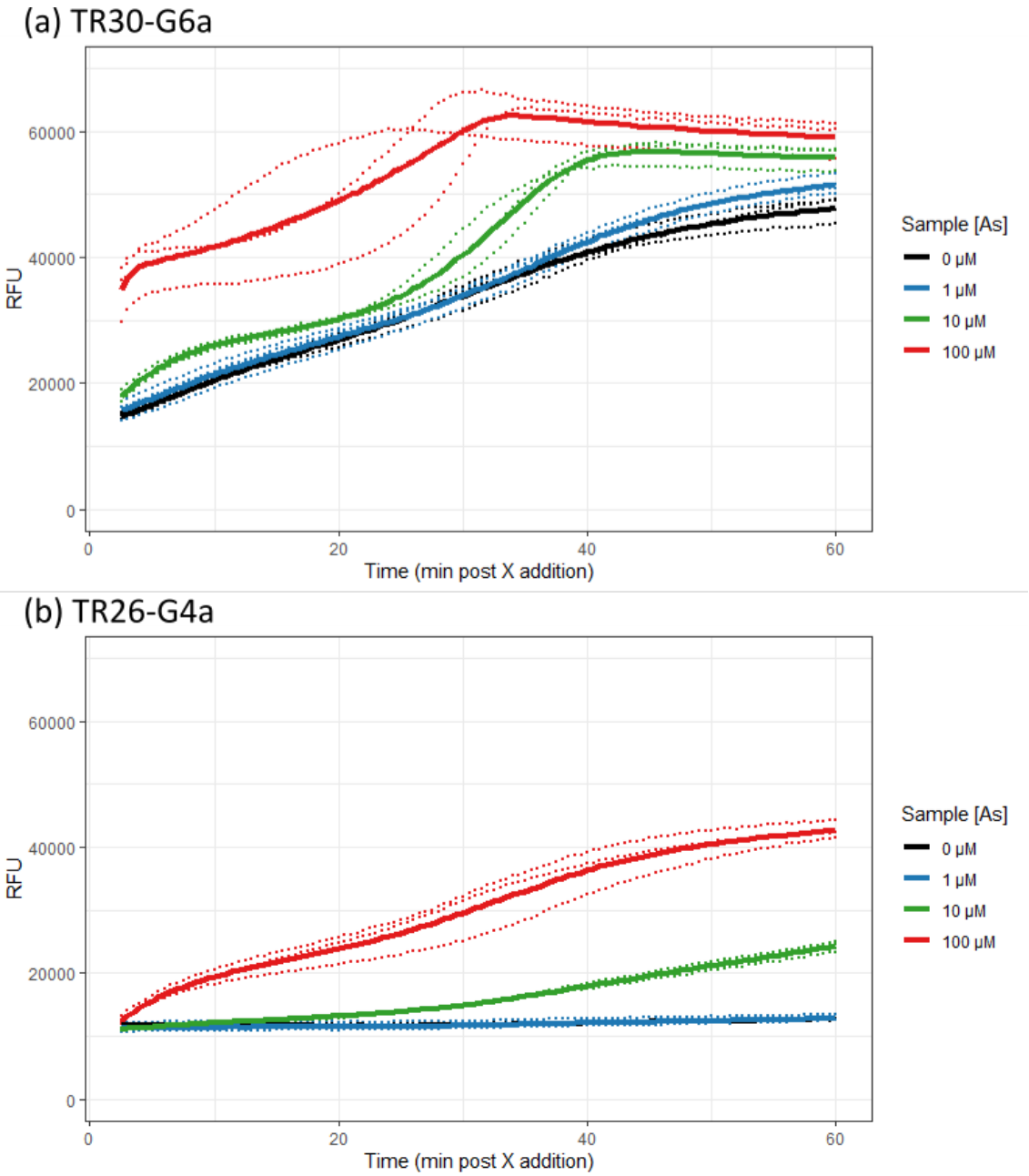


Figure 4.16: **Comparison of TR30-G6a and TR26-4a response to higher volume arsenite samples.** Fluorescence time courses displaying the RFU for CgArsR-bound Tprobes responding to varying concentrations of arsenite. The assay methods were identical except the probes used in each: (a) TR30-G6a and (b) TR26-G4a. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each well contained 10 pmol Tprobe and 100 pmol CgArsR co-incubated for 30 min before 10 µl sodium arsenite addition at the indicated concentrations, then another 30 min incubation before addition of 48 pmol X oligo (X6a-sCR30c or X4a-sCR26c), for a total volume of 20 µl per well of 384-well plate.

As expected, such a low rate of repressor-bound probe DSD did translate into a slower arsenite response compared to the TR30-G6a probe, but this notably improved reproducibility and dynamic range of the system. While this

formulation was not sensitive enough to respond to 1 µM sodium arsenite within the time-frame, this may have been possible if the assay was run for longer.

Figure 4.16 demonstrated that these two probes had large differences between their DSD rates both with and without CgArsR, in large part due to the presence or absence of two nucleotides in their toehold sequence. One way to fine-tune DSD rates instead of making such discrete jumps is to alter the concentration of displacing oligo, particularly the relative excess of this X oligo over its complementary G strand (within both Tprobe and GRQ species). The assay shown in Figure 4.16 (b) was repeated with twice the concentration of X4a-sCR26c, with the result in Figure 4.17 below.

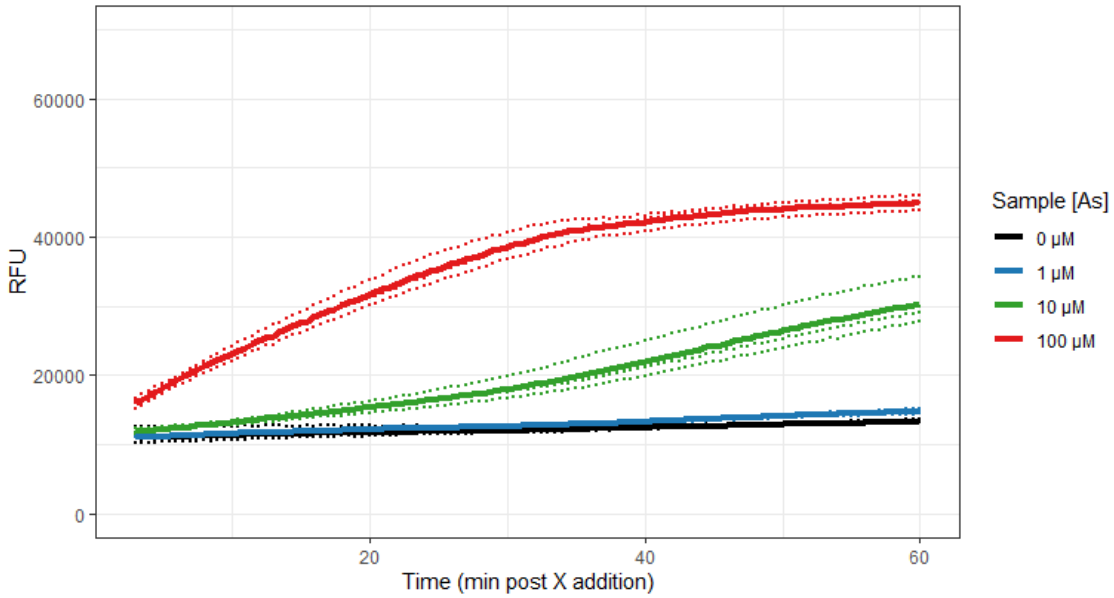


Figure 4.17: **Increasing X oligo concentration boosts TR26-G4a sensitivity and speed of response.** Fluorescence time courses displaying the RFU for CgArsR-bound TR26-G4a responding to varying concentrations of arsenite. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each well contained 10 pmol TR26-G4a and 100 pmol CgArsR co-incubated for 30 min before 10 µl sodium arsenite addition at the indicated concentrations, then another 30 min incubation before addition of 48 pmol X4a-sCR26c, for a total volume of 20 µl per well of 384-well plate.

This alteration resulted in a small rise in background DSD, but a notable change in the response to arsenite. The detection of the high concentrations was quicker, with the 10 µM samples distinguishable from the arsenic-free samples in under 10 minutes, compared to 20 minutes previously. Towards the end of the assay the 1 µM arsenite samples were all over 1k RFU higher than the arsenite-free average, suggesting the sensitivity had also increased.

4.2.6 Conclusions

In this section the factors affecting DSD rates with and without repressor were explored. None of the interventions were able to completely remove background DSD, but several of them provided useful insights into the relative importance of the factors determining reaction behaviour. These are summarised below:

- Tprobe structure does not appear to significantly impeded repressor-based DSD, but enables rapid prototyping of new designs
- Displacement of repressor-bound probe is either negligible or minimally affected by reducing the length of the operator in an attempt to improve toehold occlusion
- Increasing volume of arsenite added is a relatively easy method of improving sensitivity, but will inevitably slow reaction speed if components are significantly diluted
- Reducing toehold strength can substantially slow DSD - but rates both without repressor and the background rate with bound probe are affected in parallel. Speed and sensitivity can be improved by maximising the window between these rates, but doing so is not trivial
- Altering displacing oligo concentration can fine-tune these rates and so is a better intervention if small changes are desired
- X oligo structure can complicate dynamics and make sequence adjustments hard to predict
- It may be impossible to completely remove background DSD and attempting to do so may limit speed and sensitivity

4.3 Transitioning to a reversible DSD system

The inability to completely shut down background DSD, even with very weak toehold sequences and large repressor excesses, led to the formulation of a new hypothesis. While there had been some appreciation for the affinity the repressors have for their operator sequences, and the effect the strength of this

interaction has on the excess of repressor required to 'fully' bind a fixed quantity of operator dsDNA, little attention had been given to the possible kinetics of this situation. The evidence presented in the previous section failed to prove that DSD could occur to a repressor-bound probe, which eventually led to the suggestion that there may always be some repressor-free probe able to undergo DSD, even with large repressor-to-probe excesses.

Figure 4.18 investigated the effect of increasing CgArsR concentration on DSD of repressor-bound Tprobes. While even the lowest quantity (a five-fold excess over TR26-G6a and four-fold excess over available operator dsDNA) would be sufficient for a 'full' EMSA shift, this and larger excesses are unable to prevent full displacement and signal saturation after 40 minutes. While it was expected that it may be impossible to guarantee 100% operator binding even with these substantial excesses, it was assumed that, once bound, a probe would stay bound over the timescale of the assays conducted, unless forcibly displaced by an X oligo or dissociated by the addition of arsenite. A scenario that was initially missed was that there would be a high natural rate of unforced unbinding and rebinding.

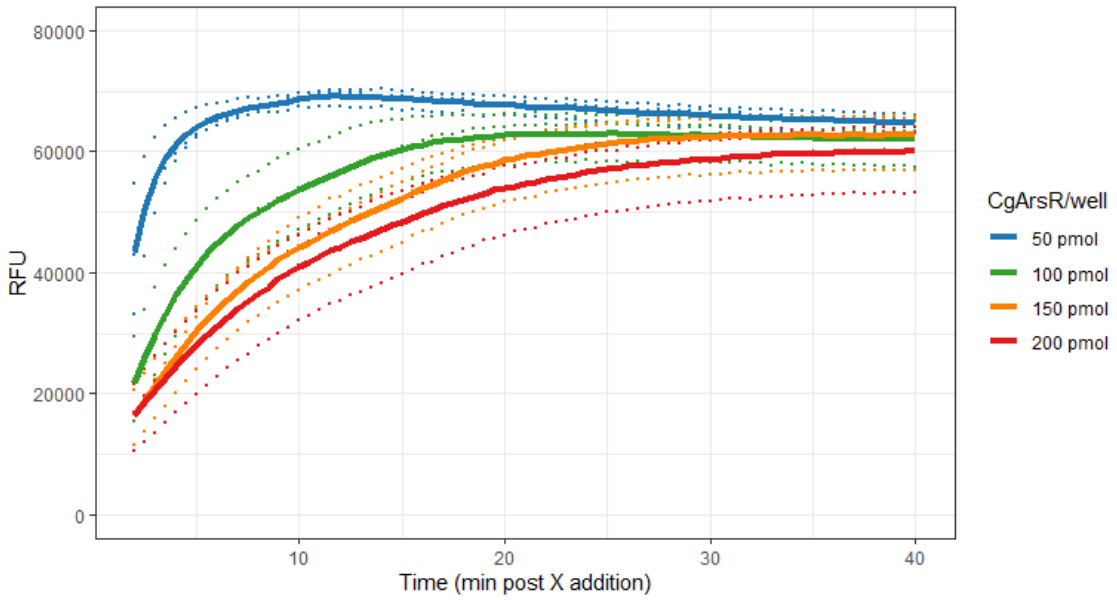


Figure 4.18: **Increasing CgArsR concentration slows, but does not prevent, background DSD.** Fluorescence time course displaying the RFU for samples with varying quantities of CgArsR. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each sample contained 10 pmol TR26-G6a and 50-200 pmol CgArsR co-incubated for 30 min before addition of 48 pmol X6a-sCR26c, for a total volume of 10 µl per well of 384-well plate.

If most repressor-bound probe complexes stayed bound for the during of this assay, unless forcibly displaced by X oligos, signal increase should only be

possible at two rates - fast displacement of unbound probe and slower displacement of bound probe. Small initial differences in DSD rate between the samples in Figure 4.18 would be expected due to the assumed minor differences in the small quantity of unbound probe initially present in each. Once the unbound portion had reacted, each would then be expected to increase in signal at a similar rate - the slower rate of bound probe displacement - leading to somewhat biphasic dynamics like those seen in Figure 3.24. The smooth kinetics for each of the CgArsR concentrations seen in Figure 4.18 would suggest these assumptions were incorrect.

Instead, the observation that each CgArsR concentration seemed to have its own, consistent, background DSD rate greatly suggested that there was a constant replenishment of unbound probe, the rate of which was dependant on CgArsR concentration. This makes sense in the context of a high rate of both binding and unbinding between repressor and operator - when co-incubating just these at any one time 95% could be bound but the exact species involved would constantly change. In the context of an irreversible DSD reaction any available unbound probe could react with a displacing oligo, and subsequently more repressor-free probe would be released simply through stochastic unbinding. The likelihood of an individual probe complex undergoing a DSD reaction would depend on both the prevalence and reactivity of the displacing oligo, and the length of time the probe was available until it was rebound by a repressor - determined by the concentration of free repressor.

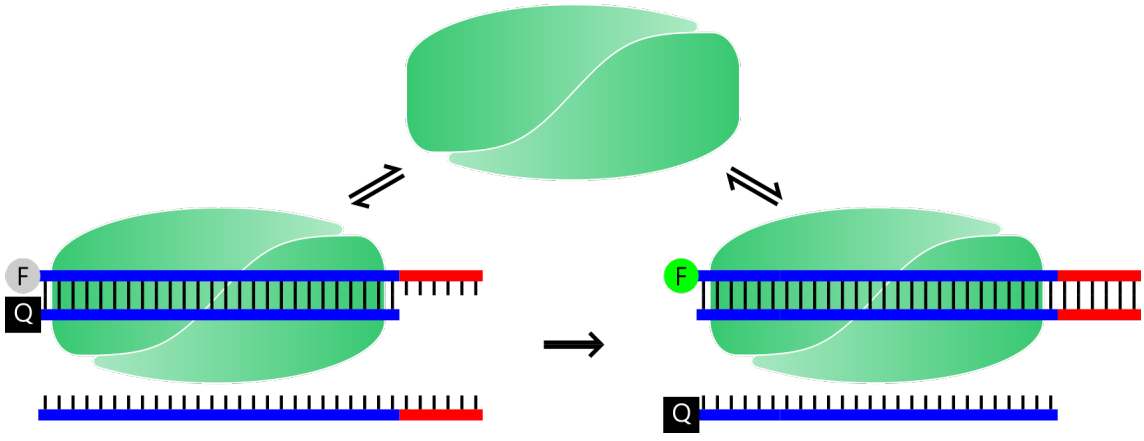


Figure 4.19: **Natural repressor unbinding allows background DSD reaction.** Schematic illustrating how a repressor may spontaneously unbind from its operator within a probe, allowing an irreversible DSD reaction to occur. The repressor may then bind another free probe or the DSD product which would also contain a full dsDNA operator.

As Figure 4.19 shows, this natural unbinding would enable the continual

accumulation of the DSD products, one of which could itself be bound by the repressor. This is the ArsR-FGX species seen in the EMSAs in Figures 4.7, 4.10 and 4.13. Reducing toehold strength and increasing repressor concentration are the two most effective ways to reduce the rate of this accumulation, but neither can be altered to completely remove this background. Importantly, partial alterations to these parameters would reduce the speed and sensitivity of the system, severely limiting the performance of an eventual biosensor. To ensure some reproducibility for such a diagnostic it would be necessary to either separate the displacing oligo from the probe and only combine them at the point of sample addition, or to freeze-dry reactions immediately after mixing together. Either would present substantial practical challenges.

Instead, a more elegant biochemical solution was proposed. The DSD reactions so far have involved full-length X oligos irreversibly displacing quencher strands - the permanence of this reaction enabled the problematic background signal accumulation. If the DSD reaction was itself reversible, then this background would not continually increase but plateau to reach an equilibrium. Adding more repressor to the system should lower this stable signal, while adding arsenite or more X oligo should increase it.

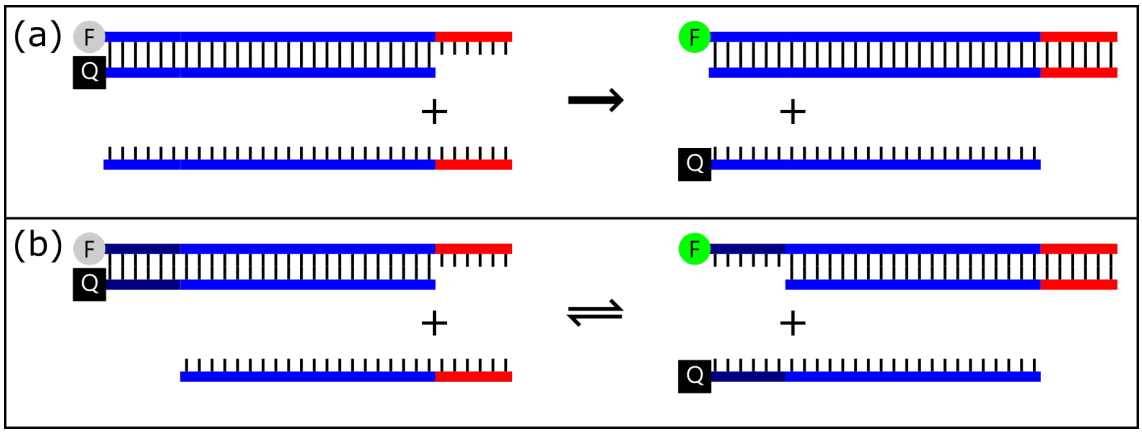


Figure 4.20: **Comparison of reversible and irreversible DSD reactions.** Schematic illustrating how an irreversible DSD reaction can be adapted into an irreversible one. (a) An X oligo comprised of the reverse toehold and the full-length operator reacts irreversibly with an FQ probe to allow fluorescence. (b) By shortening the length of the displacing oligo a reverse toehold sequence is created (dark blue). This allows free quencher strand to itself displace the X strand, reforming the original quenched probe. Across a large population of molecules both reactions will happen continuously to set up a dynamic equilibrium.

Figure 4.20 illustrates how an irreversible DSD reaction can be adapted to set up a reversible equilibrium. Simply by reducing the length of the displacing X oligo a reverse toehold is created following the initial reaction. If the strength

of this new toehold is roughly comparable to that of the forward reaction, the free quencher strand would be able to occasionally displace the X oligo off the fluorescent product to reform the quenched probe. Eventually the rate of the competing reactions would match to set up a dynamic equilibrium.

By combining the reversible repressor binding with a reversible DSD reaction there would not be constant signal accumulation limiting the window in which sample analysis could occur. Instead the system would shift between new stable points, determined by the concentrations of the DNA species, repressor and arsenite.

An example of this is displayed in Figure 4.21. Here, the TR26-G6a probe reacts with three X oligos differing by a single base - X6a-sCR17c, X6a-sCR18c and X6a-sCR19c (labelled X17/X18/X19). These each stabilise at an RFU determined by the dynamic equilibrium point of their respective forward and reverse reactions, correlating with the length of X oligo. The RFU for all the samples is reduced following addition of CgArsR, reflecting the shift towards the quenched probe in each, while addition of arsenite shifts the equilibrium forwards again.

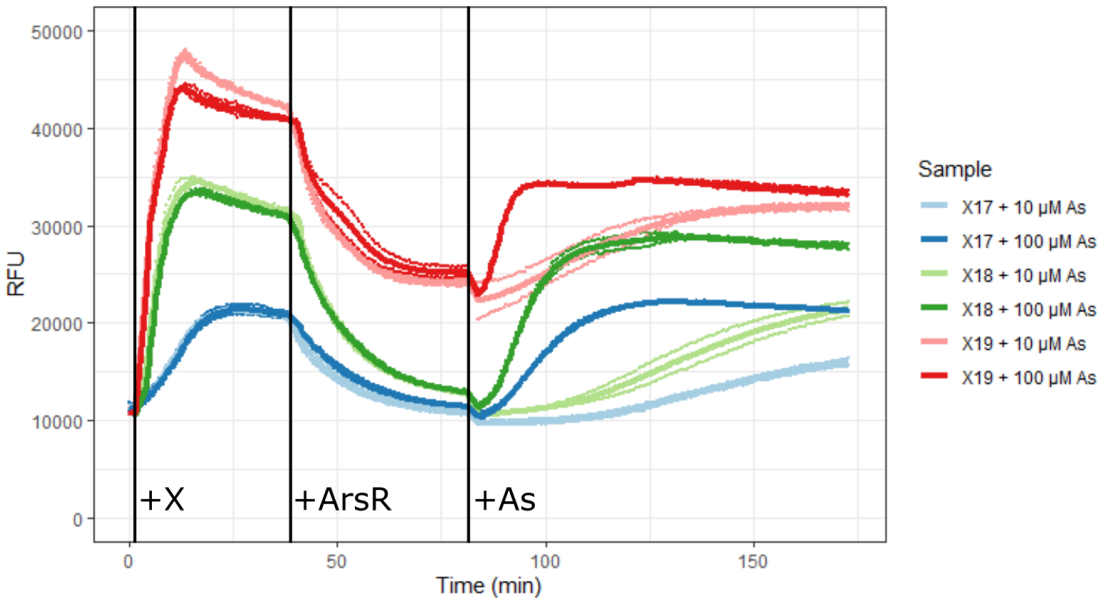


Figure 4.21: **CgArsR and arsenite can shift the equilibrium of a reversible DSD reaction.** Fluorescence time course displaying the RFU for probes reacting to different length X oligos and varying quantities of arsenite. Graph shows readings from individual wells as points and averages across each set of two replicates as lines. Each sample contained 10 pmol TR26-G6a before addition of 48 pmol X6a-sCR17c/X6a-sCR18c/X6a-sCR19c after 2 min (+X), 50 pmol CgArsR after 38 min (+ArsR) and 10 µl of 10 or 100 µM sodium arsenite after 81 min (+As) for a total volume of 20 µl per well of 384-well plate.

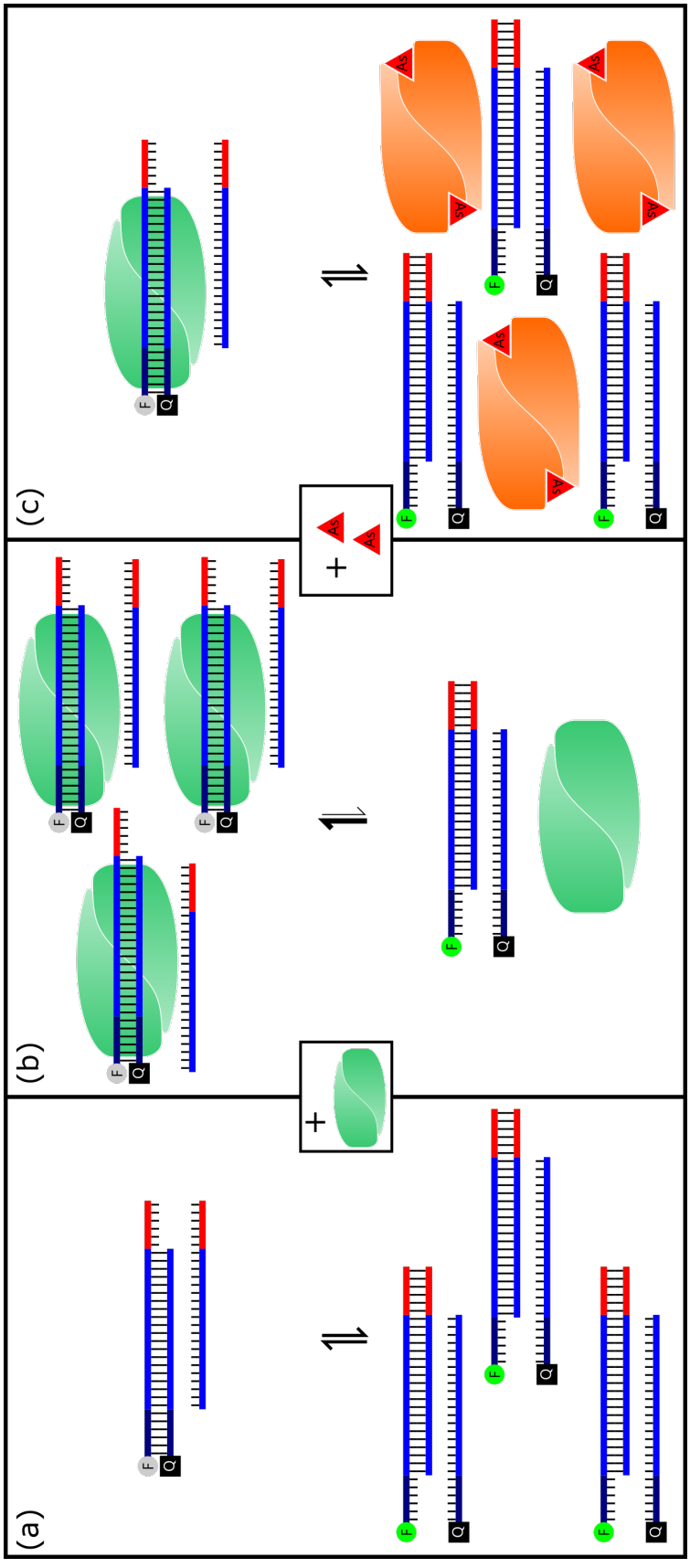


Figure 4.22: **Basic reversible repressor-DSD reaction.** Schematic illustrating a reversible DSD reaction and how its equilibrium is altered by the addition of repressor and its ligand. (a) The annealed FQ probe reacts with a partial-length displacing oligo, setting up an equilibrium between the forward and reverse reactions. The forward and reverse toehold domains are coloured in red and dark blue, respectively. The light and dark blue domains together form the operator sequence for the repressor. (b) ArsR reversibly binds its operator sequence within the FQ probe, but it cannot bind the FX product as the full operator dsDNA is not present. This stabilises the FQ probe and slows the rate of the forward reaction without stopping it completely, thereby shifting the overall reaction equilibrium towards the more quenched state. (c) Arsenic induces dissociation of ArsR from its operator sequence, removing repressor from the active system and therefore allowing the forward DSD rate to speed up and shift the reaction equilibrium closer to the repressor-free state. This schematic uses a basic FQ probe for clarity, but most of the subsequent assays used Tprobes reacting in a similar fashion.

Figure 4.22 illustrates the underlying reaction shifts using a basic linear probe for simplicity. The changes in equilibrium following addition of repressor and arsenite reflect the behaviour of the overall population of DSD probes, not every individual DNA species. The RFU reduction following CgArsR addition is therefore not down to the absolute baseline - it simply makes the reverse DSD reaction more favourable than it previously was, through the selective stabilisation of the quenched probe, which contains the complete dsDNA operator sequence, over the fluorescent product, which does not. Addition of arsenite increases RFU in a concentration-dependant manner through gradual removal of functional repressor from the system. The difference in behaviour between the X17, X18 and X19 reactions is largely due to the variable strength of the reverse toeholds they create following displacement of the quencher strand - GAATATCGA, GAATATCG or GAATATC, respectively. The longer (and stronger) reverse toehold created by the X17 reaction limits the maximal signal of the system, while the weak X19 reverse reaction cannot be aided enough by CgArsR addition to shift the equilibrium close to the baseline. As a result, the X18 oligo seemed to generate the largest dynamic range in which to detect arsenite concentrations for this particular probe setup.

4.3.1 Reversible DSD reactions enable greater signal control

In order to compare the performance of the reversible system to previous irreversible iterations, an assay similar to those in Figure 4.16 was set up. In this case, however, following pre-incubation of CgArsR with the TR26-G6a probe, the intermediate length displacing oligo X6a-sCR18c (X18) was added. After another 30 minute incubation sodium arsenite at different concentrations was added and fluorescence recorded in Figure 4.23. The stability of the reversible system allows time to be calculated from the moment of arsenite addition rather than X addition. Unlike in Figure 4.16 (a), where the 6a toehold is also used, the arsenite-free signal does not rapidly rise as soon as the displacing oligo is added as the X18 oligo had already pre-equilibrated with the probe. The background rose ever so slightly over the course of the assay, but a small equilibrium adjustment to the dilution effect of the added volume would be expected.

Over the course of two hours the assay showed clear, reproducible signal increases to 10 and 100 µM arsenite. These took longer to separate from the background than the irreversible systems in Figure 4.16, but these previous

assays should include an extra 30 minutes for the true time since arsenite addition. Towards the end of the assay even the 1 µM arsenite samples seem to be increasing above the background, showing very similar sensitivity to Figure 4.16 (b). This suggests that sensitivity is not substantially reduced by the switch from irreversible to reversible reaction, and while slightly slower in this form the much improved stability should allow it to be greatly optimised further.

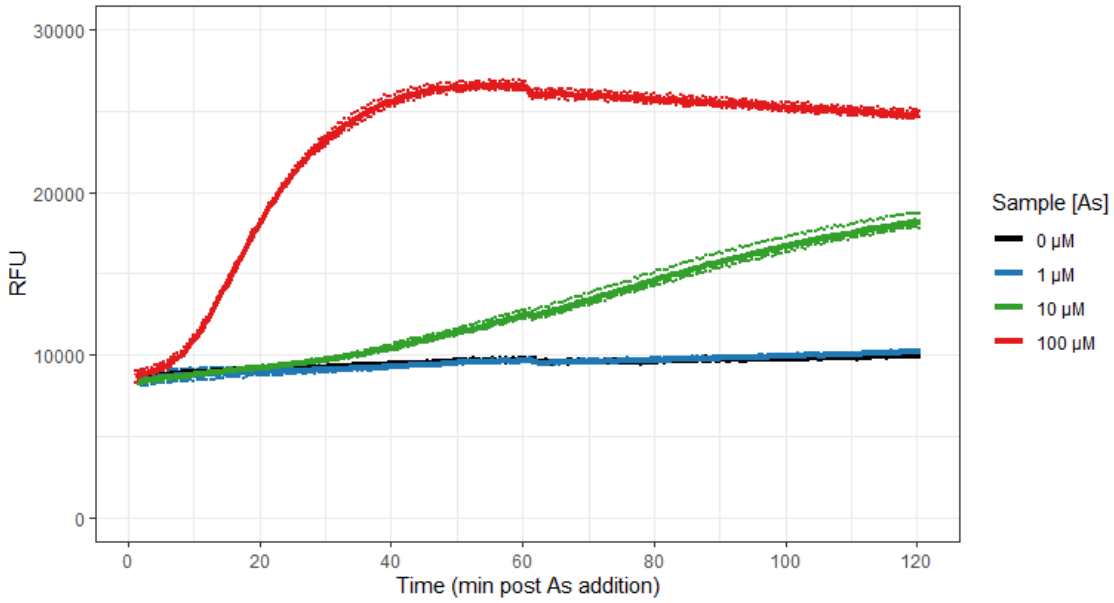


Figure 4.23: **Reversible DSD assay results in slower but more stable arsenite detection.** Fluorescence time courses displaying the RFU for CgArsR-bound TR26-G6a responding to varying concentrations of arsenite. Graph shows readings from individual wells as points and averages across each set of three replicates as lines (only two replicates for100 µM samples). Each well contained 10 pmol TR26-G6a and 150 pmol CgArsR co-incubated for 30 min before addition of 48 pmol X6a-sCR18c, then another 30 min incubation before 10 µl sodium arsenite addition at the indicated concentrations, for a total volume of 20 µl per well of 384-well plate. The assay was restarted after 1 h resulting in the plate moving, giving the slight signal adjustment at this time.

4.3.2 Displacing oligo length can significantly alter structure

Choosing the optimal length of displacing oligo for these reversible reactions can be difficult. As demonstrated in Figure 4.21, increasing length results in a higher (more fluorescent) equilibrium point, as a result of the reduced reverse toehold. In many cases the speed at which the reaction initially progresses correlates with the magnitude of this eventual stable point, but this is not necessarily the case.

An example of this scenario is shown in Figure 4.24, which compares the DSD dynamics of TL26-G6a, the Tprobe containing the second native operator,

reacting to X oligos of different lengths. Here, pT refers to a 30nt poly-T oligo, used as a control showing the baseline fluorescence of a probe that has not undergone any displacement. Concentrations of the X oligos are also given as fold changes relative to their G strand counterparts, as explained in the figure caption. Data are shown both for an initial two hour time course after combining the probe and displacing oligos, and a short series of reads the following morning when the reactions were assumed to have reached equilibrium.

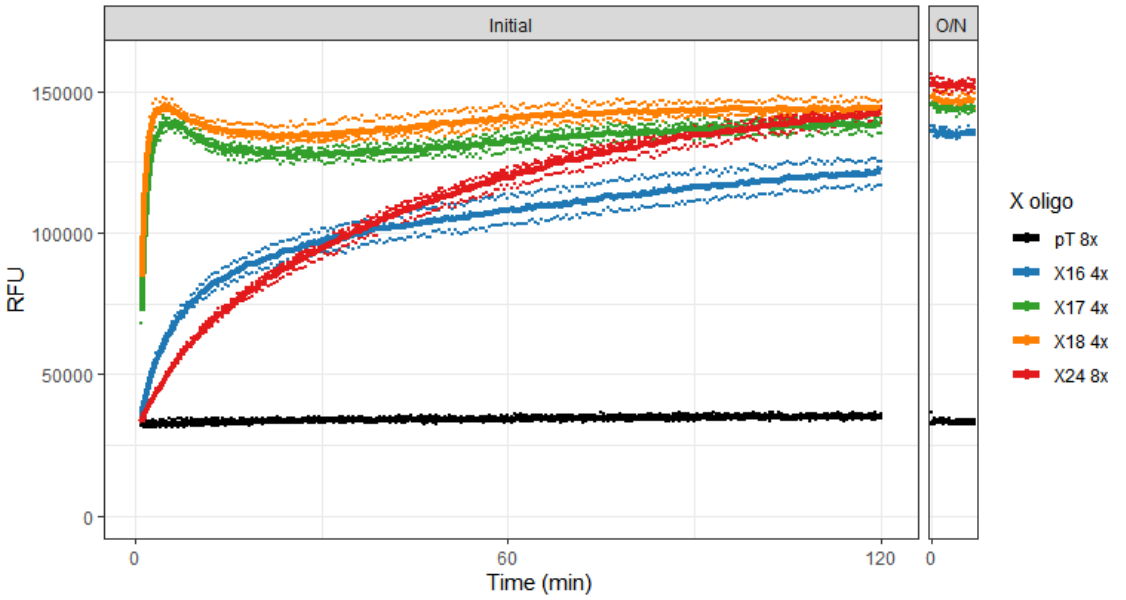


Figure 4.24: **Longer displacing oligos do not always give faster reactions.** Fluorescence time courses displaying the RFU for unbound TL26-G6a responding to corresponding displacing oligos of different lengths. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each well contained 1.5 µl of 8/16 µM X oligo stock (12/24 pmol) to which 200 µl of 10 nM TL26-G6a (2 pmol) was added to trigger the reactions, for a total volume of 201.5 µl per well of 96-well plate. The Tprobe was annealed using the highQ strand ratio described in section 4.1.3, resulting in a 50% G strand excess over full probe (for which the above concentration is given) and so the X oligos were added in a 4x or 8x excess over their complementary G strand toeholds. O/N data is from reading the same plate for 7 min the following morning.

Figure 4.24 shows that the short X16 oligo was slow to displace as expected, increasing in signal throughout the whole initial two hours. Its overnight equilibrium was higher still, suggesting perhaps another hour was needed before this mix stabilised. The reactions with X17 and X18 were much faster, peaking after only five minutes before displaying dampened oscillations around this point for the rest of the assay, but eventually settling at a higher RFU than X16. While X24 would be expected to react even quicker than these, with its

reverse reaction using only 2nt toehold, instead it was initially slower than even X16 and took over 30 minutes to surpass this supposedly weakest reaction. By the end of the assay the X24 DSD had steadily increased to a point similar to X17 and X18 and at the eventual equilibrium did stabilise higher as expected, confirming that the experiment was prepared correctly.

The explanation for this stark difference in initial DSD rate is shown in Figure 4.25. The X16, X17 and X18 oligos, without 8-10 bases of the CL26 operator, are almost structureless, with only a couple of weak internal bonds likely to form. X24, on the other hand, only has 2nt removed, enabling the strong extended hairpin to form from the inverted repeat sequence within the full native operator. So while the longer complementary sequence and the smaller reverse toehold meant the X24 reaction eventually stabilised at a higher equilibrium, the internal structure of this oligo greatly reduced the speed at which this was reached.

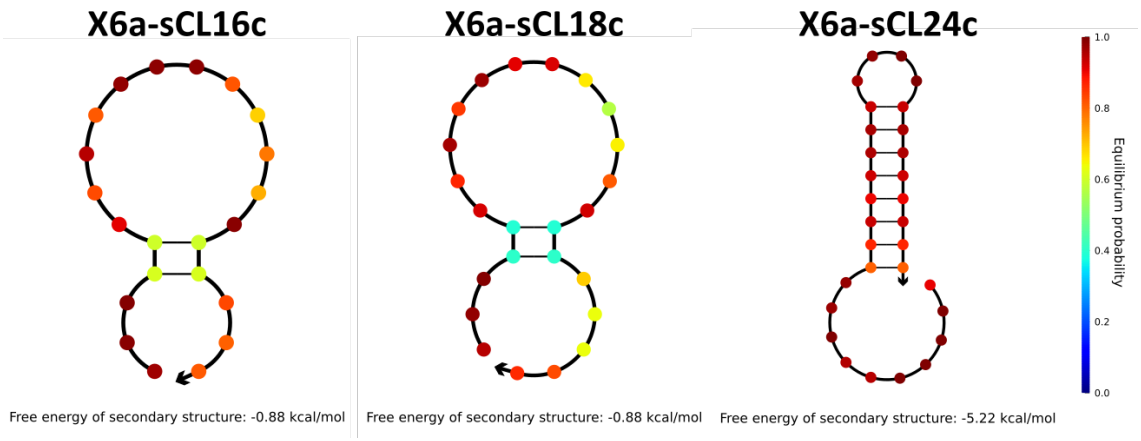


Figure 4.25: **Longer CL operator displacing oligos are more structured.** Minimum free energy secondary structures formed by single-stranded oligos comprised of the X6a toehold and different lengths of the CL26 operator reverse complement sequence. The structures and associated minimum free energies were predicted using NUPACK at 1 µM and 25 °C, with individual bases coloured according to the probability of being in the suggested state using the scale on the right. The 3' end of each oligo is labelled with an outward arrow.

This potential relationship between displacing oligo length and structure for these reversible reactions is yet another reason why altering operator sequences to prevent such formations is desired. Manipulating displacing oligo length is an easy and effective way to control the speed and strength of these reversible DSD reactions, but presence or variability in structure can limit speed as well as increase the difficulty in predicting the effect of such changes.

4.4 Removing X oligo secondary structure

Altering the probe operator sequence has been a repeated theme since starting work on CgArsR. Early in this chapter it was noted that the two native 30bp operators had numerous differences (Figure 4.1) but despite this displayed similar affinities for CgArsR (Figure 3.21). Two overarching reasons for investigating alternative operator sequences emerged since - to manipulate the CgArsR-operator affinity and to reduce the structure of the single-stranded X oligo. Following transition to a reversible DSD system the structure of different lengths of X and the released quencher species were also important to consider due to their roles in the competing DSD reaction.

Removing ssDNA structure itself has a number of advantages. First, it makes comparisons between probe alterations affecting speed and sensitivity easier - it takes away changes to oligo MFEs and toehold accessibility, like those seen in Figure 4.11 and Figure 4.25, as possible explanations for differences in behaviour. Understanding the effect of altering operator length, operator sequence and toehold strength would then become much easier. Second, it should allow faster DSD reactions. A displacing oligo that forms a stable hairpin or dimer will undergo a smaller free energy change upon binding its complement compared to a similar structureless oligo - removing such structures should therefore increase the speed at which shifts between stable states would occur.

Third, it should make the system as a whole simpler and therefore more predictable. Dynamics more complex than those expected for simple bimolecular reactions have been observed (e.g. Figure 4.16) and secondary structure is a likely factor in causing this. Simpler kinetics are beneficial both for building a biosensor whose output will need to be interpreted and to be able to predict the effect of design changes. The relatively minimal system presented here - with only DNA, repressor and arsenite in solution - has the potential to be accurately modelled if components interact through only a few mechanisms. Such a model would be invaluable in optimising the composition of an assay for speed or sensitivity, or for simulating the effect of design alterations. Removing DNA secondary structure as a variable would greatly simplify this process.

4.4.1 Targeted mutations identify scope for operator alterations

With the assistance of a summer student (B. Baker), CgArsR operator sequence dependency was initially investigated through gel shift assays. A number of 30bp constructs were designed to test small hypotheses derived from the preliminary comparison between CR30 and CL30. The EMSAs for these dsDNA species are displayed in Figures A.3 and 4.27, with their sequences and assessment of CgArsR binding ability summarised in Figure 4.26.

The first hypothesis was that the 'optimal' CgArsR operator was likely to contain an inverted repeat. Homodimeric transcription factors often bind each other with 180° rotational symmetry, and so they commonly have two anti-parallel DNA-binding domains, either one in each monomer or both at the interface between them. This structure means they tend to preferentially bind inverted repeat sequences - BsArsR had such an operator (B28: 5'-AATCAAAATAAATTGATTTATTTGCTTG-3'). CL30 and CR30 shared two regions of sequence identity, but only an 'ATAT' within this formed an inverted repeat. New 30bp constructs were therefore made which reversed each entire conserved subsequence in turn around the suspected 4bp 'spacer' to force longer perfect inverted repeats.

As each mutant sequence used the CR30 operator as its origin, these were named CR30-IR1 and CR30-IR2. The ultimate effect of these changes were that in each case six bases were changed from the native CR30 operator, as highlighted in yellow in Figure 4.26. The result of the EMSA on these constructs is shown in Figure A.3 (a), with neither of them shifted by CgArsR. Clearly, rather than creating a 'better' operator these changes abolished binding completely, also demonstrating that CgArsR did have a somewhat stringent sequence dependency.

Likewise, two constructs were made to alter the separation between the two identical native subsequences. As bases 12-15 were all different in CR30 and CL30, it seemed likely that these positions had lower sequence stringency than those either side. CR30-6Space introduced 'TC' between bases 13 and 14, shifting bases 14-28 in the 3' direction and removing bases 29 and 30 (we knew these 3' end bases were less important than those at the 5' end), while CR30-2Space removed 'GT' from bases 13 and 14, shifting bases 14-30 in the 5' direction and introducing 'TC' at the 3' end. The same figures showed that these

changes also lost CgArsR binding, suggesting either the repressor is not flexible enough to accommodate such a spacing difference or that the sequences in these positions do matter.

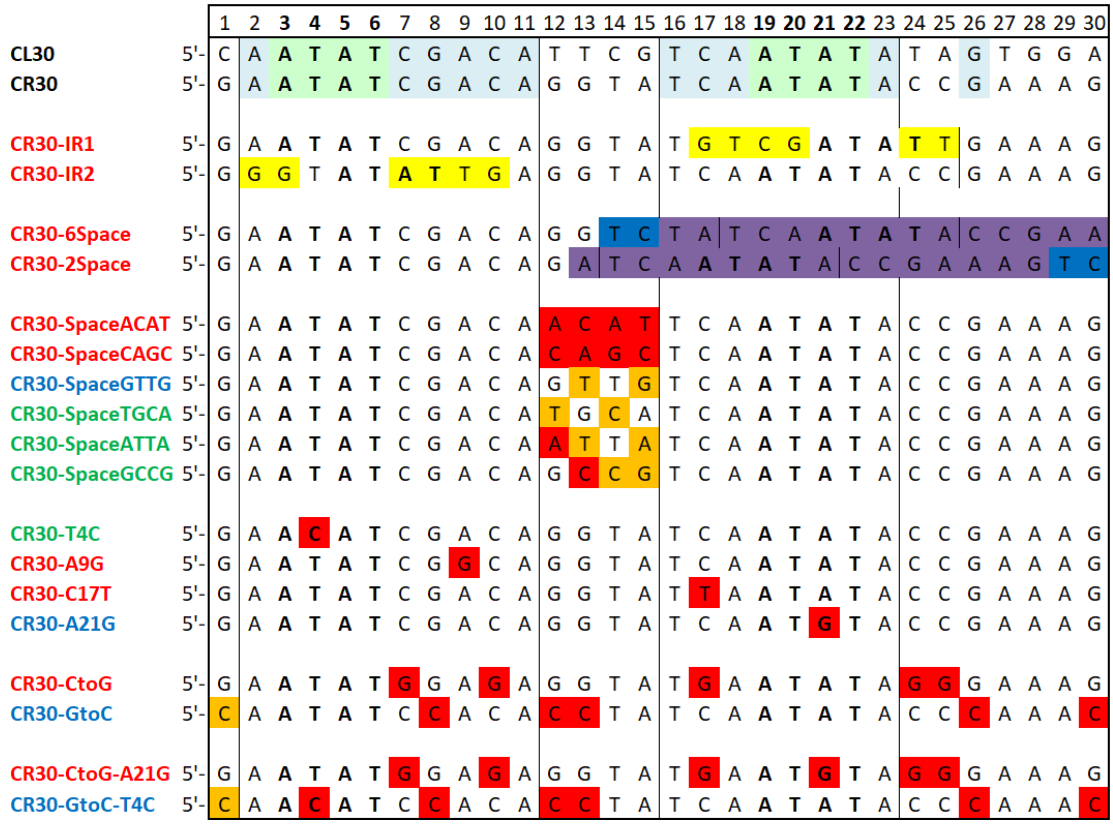


Figure 4.26: **EMSA summary for targeted 30bp mutant operators.** Comparison of the native and mutant operators assayed for CgArsR binding by this first round of EMSAs. The native operators are highlighted as in Figure 4.1, with identical bases in light blue and the inverted repeat within these highlighted in light green with bases and positions in bold. The mutant operators all derive from CR30 - substitutions to bases that are present in CL30 are highlighted in orange, or in yellow if duplicated to form an inverted repeat. Point mutations to bases not in either CR30 or CL30 are highlighted in red, with inserted bases in blue and shifts resulting from indels in purple. The names of each of the resulting theoretical operators are coloured according to how well they were shifted by CgArsR - green if similar to wild-type, blue if bound with reduced affinity, or red if binding was substantially reduced or completely lost.

A further six mutants were created to investigate sequence dependency in this spacer region. Two were designed so that none of the bases at these positions were present in the native operators (CR30-SpaceACAT and CR30-SpaceCAGC), while another two derived from new combinations of the native bases (CR30-SpaceGTTG and CR30-SpaceTGCA). The final two were intermediates containing two CL30 bases and one non-native base (CR30-SpaceATTA and CR30-SpaceGCCG). Figure A.3 (b) and (c) tested these for CgArsR binding and

found that both the fully non-native spacers (CR30-SpaceACAT and CR30-SpaceCAGC) displayed almost complete loss of binding while the others showed at least some affinity. CR30-SpaceATTA, CR30-SpaceGCCG and CR30-SpaceTGCA seemed to have close to wild-type affinity, while CR30-SpaceGTTG seemed slightly weaker. This showed that despite the lack of conservation between the native operators in this region, the exact sequence was still hugely important - changes to these four bases alone were able to either alter or abolish CgArsR binding.

Another four sequences were designed to test a selection of point mutations. CR30-T4C and CR30-A21G made alterations within the potentially important 'ATAT' inverted repeat, while CR30-A9G and CR30-C17T contained substitutions elsewhere in the maintained regions. In each case transition interchanges were selected rather than transversions to make the mutations as subtle as possible. Contrary to expectations, both of the ATAT mutants were still capable of binding CgArsR, although CR30-A21G did seem slightly weaker. Even more unexpectedly, both the other mutants displayed almost complete loss of binding. These results therefore disproved the hypothesis that the ATAT inverted repeat contained the most integral bases within the operator - these positions could tolerate at least some mutations while others clearly could not.

4.4.2 CR30-GtoC mutant operator maintains significant CgArsR affinity

One final set of mutants were also tested at this point. These created sequences containing only a 'three-letter alphabet' - removing either all Gs or Cs from each strand of the duplex. This approach is commonly used for DSD reactions as it prevents the formation of G-C base pairing within a single-stranded oligo, either internally or for self-dimerisation. Removing the possibility of these strong bonds massively reduces the likelihood of stable ssDNA structures, as only weaker A-T bonds are likely to form. Such sequences are trivial to design for theoretical *in vitro* reactions with no biological integration, but are far less likely to be acceptable in natural sequences as many independent mutations may be necessary. In this case, both CtoG and GtoC mutants were made, preserving the overall GC content of the sequences and limiting the potential for additional internal A-T base pairing. Two more constructs were designed which also

contained the T4C and A21G mutations to disrupt the ATAT repeat that was the final source of internal secondary structure.

As shown in Figure 4.27, the CR30-CtoG mutant suffered substantial loss of affinity for CgArsR, which seemed to be compounded further by the A21G mutation. Surprisingly, the CR30-GtoC duplex preserved significant binding, despite no less than six different mutations to the native CR30 operator. The additional T4C mutation did seem to reduce the resulting affinity, but did not abolish it despite over one quarter of bases having been altered. This discovery was hugely important and laid the groundwork for future probe designs.

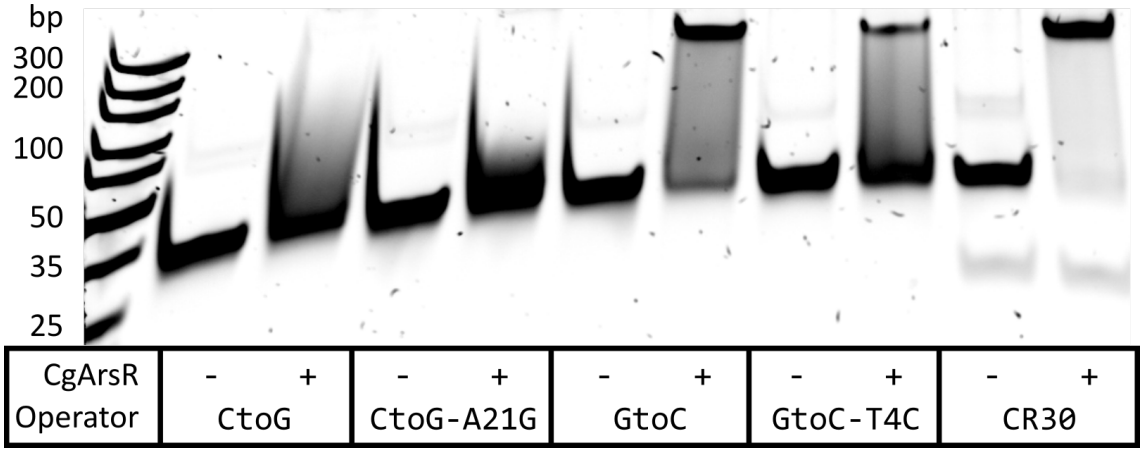


Figure 4.27: **Testing CgArsR binding of GC mutant operators.** 20% PAGE EMSAs assaying the ability of CgArsR to bind various three-letter 30bp mutant operators. The sequence of each dsDNA construct is given in Figure 4.26. 15 pmol of each annealed duplex was incubated with 100 pmol CgArsR at room temperature for 30 min in a binding buffer containing 5 mM DTT before being run on the gel.

4.4.3 Single mutant library screen reveals crucial operator inverted repeat

The surprising results from the EMSAs of targeted mutations made it clear that too much had been assumed from the initial comparison between the native CgArsR operators. The two identical subsequences did not need to be as conserved as first thought, while the sequence of the intervening spacer bases did matter. Bases A9 and C17, however, were clearly important as lone point mutations at each of these positions abolished binding, but these sensitive nucleotides were unlikely to be in the majority of places as a few constructs with several changes could still bind.

To investigate the operator sequence space more thoroughly, and without bias, a dsDNA library was created that tested every possible single mutation at every position in the CR26 operator. It seemed unnecessary to test mutants in bases 27-30 when removing these did not affect binding, and this still left 78 distinct mutants to assay. These constructs were annealed and run on EMSAs with and without CgArsR alongside CR26 across 13 gels, displayed below in Figure 4.28.

The samples below were deliberately run with enough CgArsR for good binding of the native CR26 but insufficient for maximal binding. This allowed increased affinity to be detected as well as reduced binding. Due to the batch nature of performing these EMSAs, conditions were not identical between all gels - most were run at room temperature but some were re-run refrigerated to increase clarity and contrast between differences in affinity. For this reason CR26 was included in every gel to serve as a control - degree of binding should only be observed and described relative to the CR26 shift seen within each individual gel and not between the independent experiments. A summary of these subjective, qualitative observations is provided afterwards in Figure 4.29.

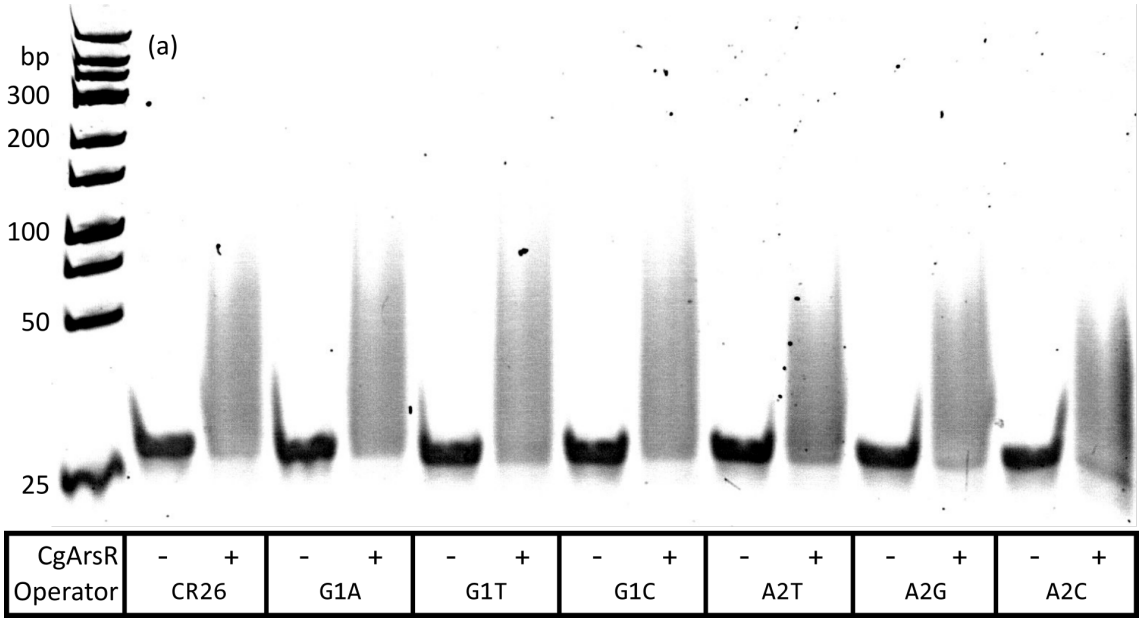


Figure 4.28: **Testing CgArsR binding of single mutant operator library.** 20% PAGE EMSAs assaying the ability of CgArsR to bind each 26bp single mutant operator. 10 pmol of each annealed duplex was incubated with 50 pmol CgArsR (stored in PSB4) at room temperature for 1 h in a solution made up to 10 µl with AB. Gels were subsequently run at 200 V for 1 h but some were run at room temperature and others were refrigerated.

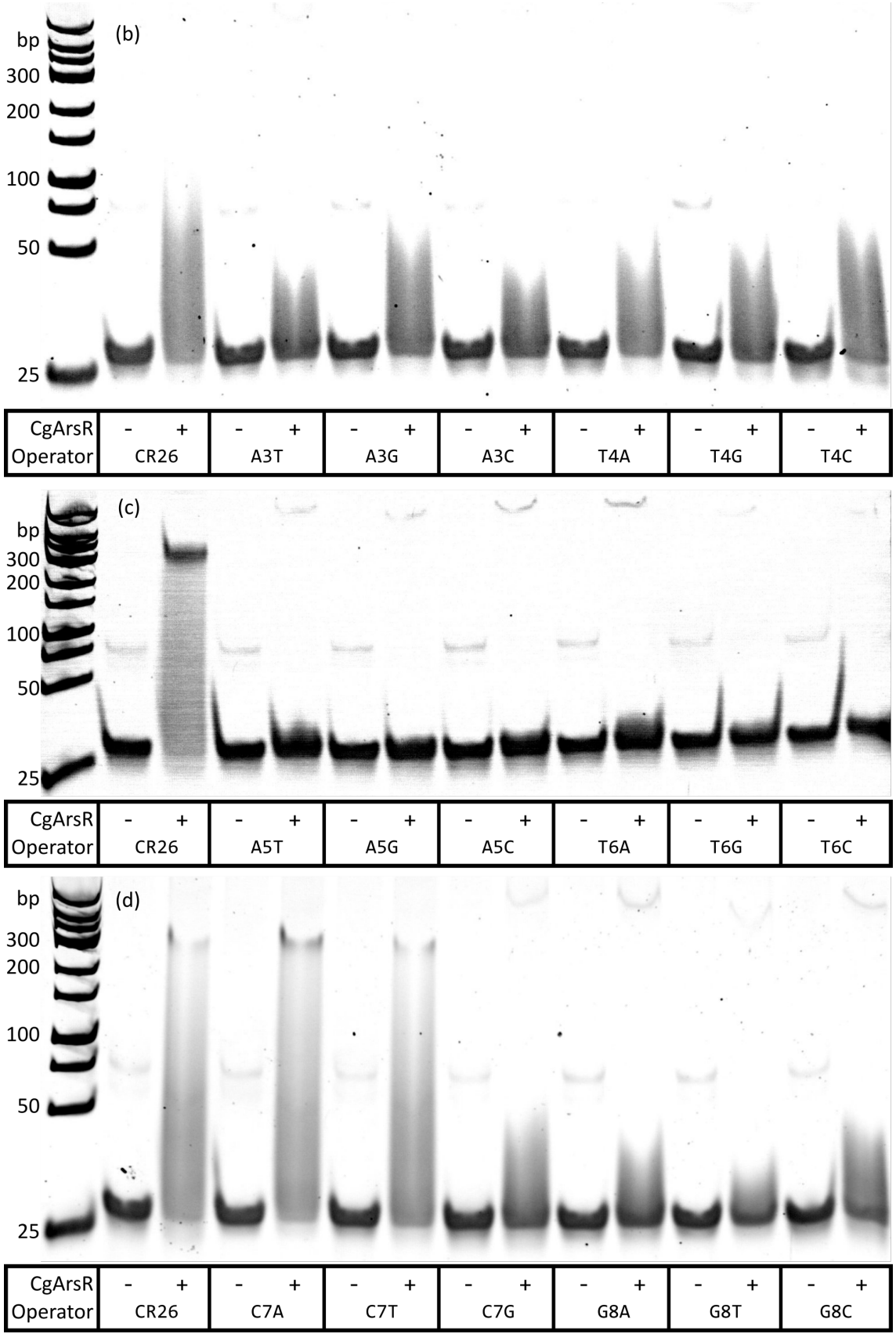


Figure 4.28: **Testing CgArsR binding of single mutant operator library.**

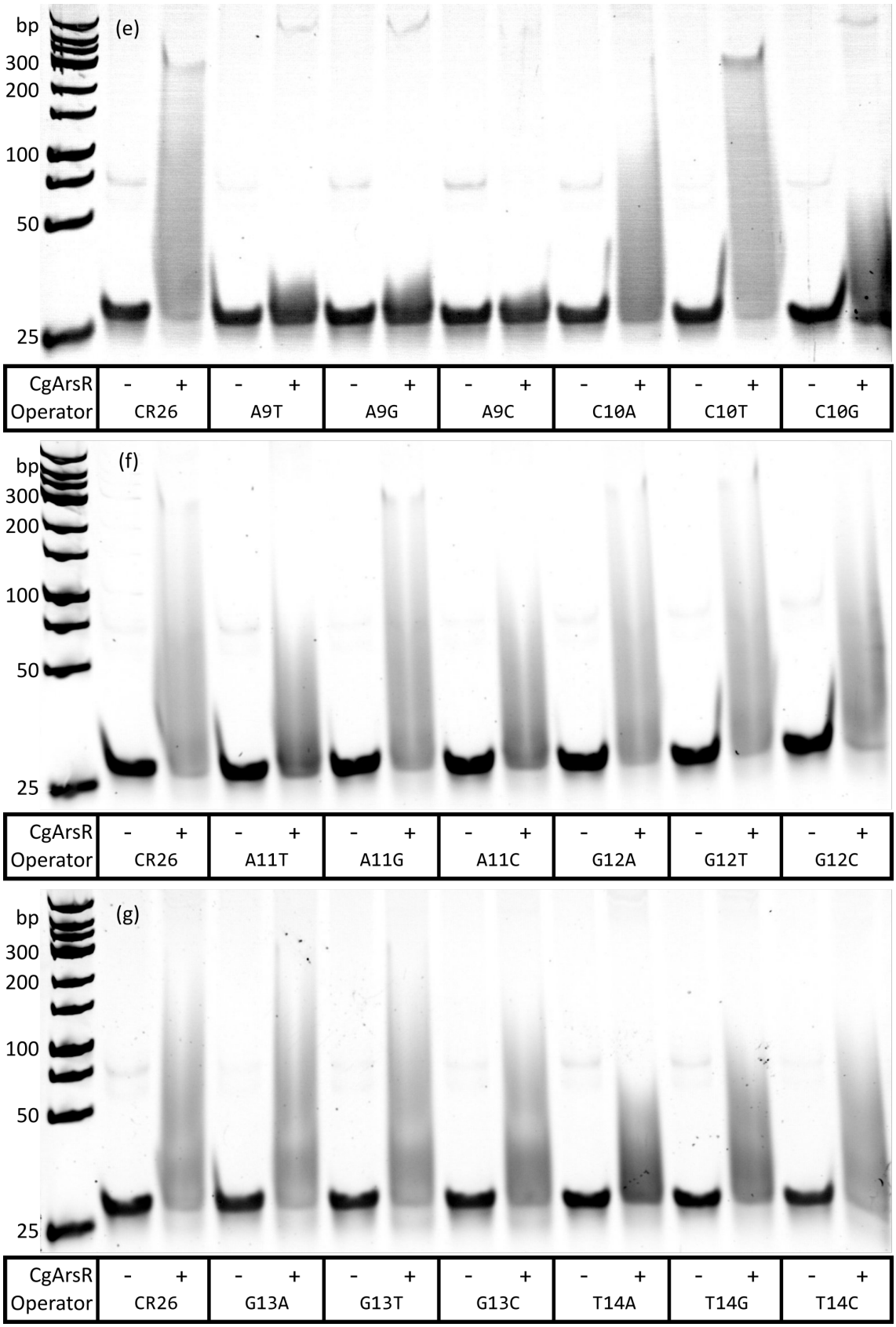


Figure 4.28: **Testing CgArsR binding of single mutant operator library.**

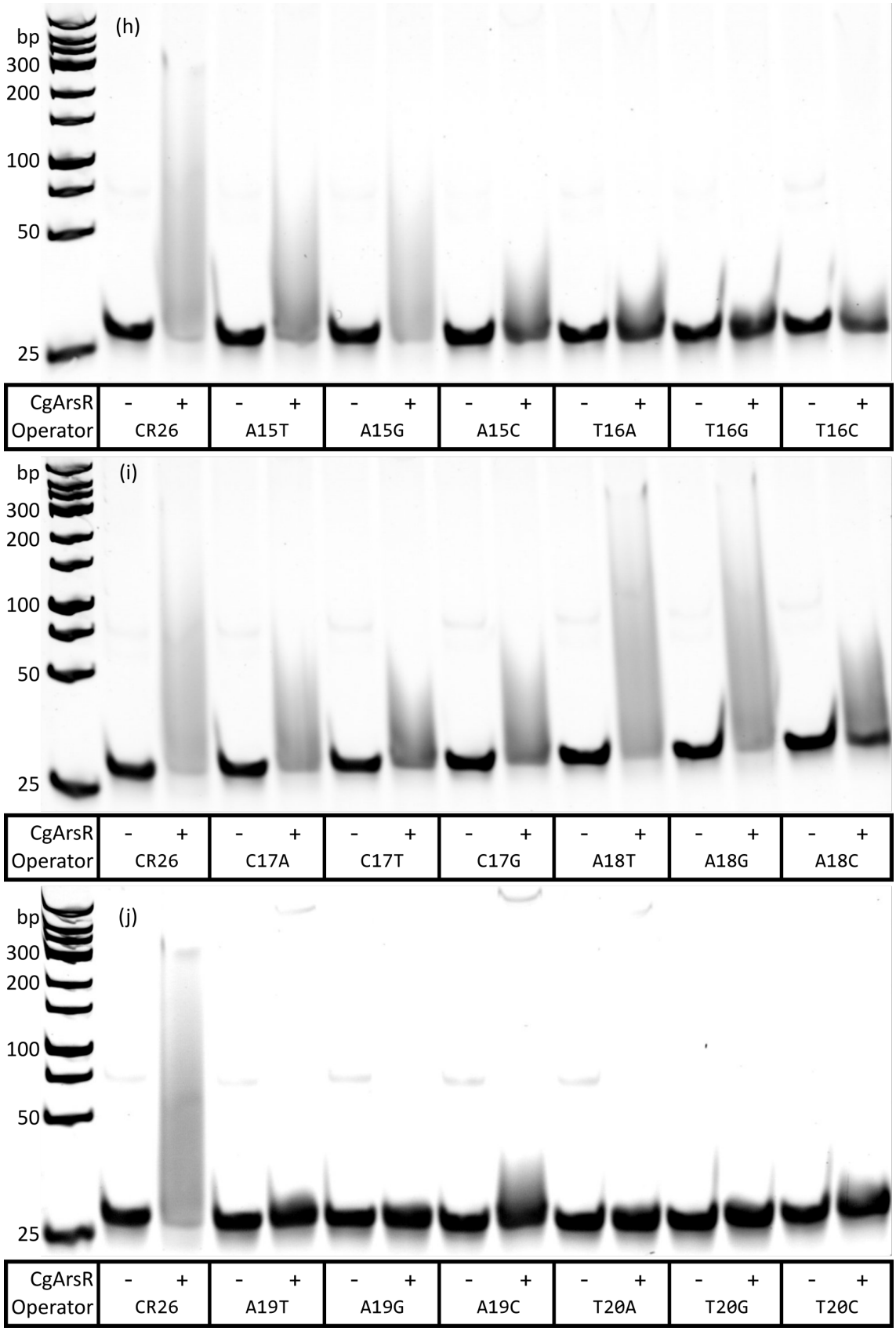


Figure 4.28: **Testing CgArsR binding of single mutant operator library.**

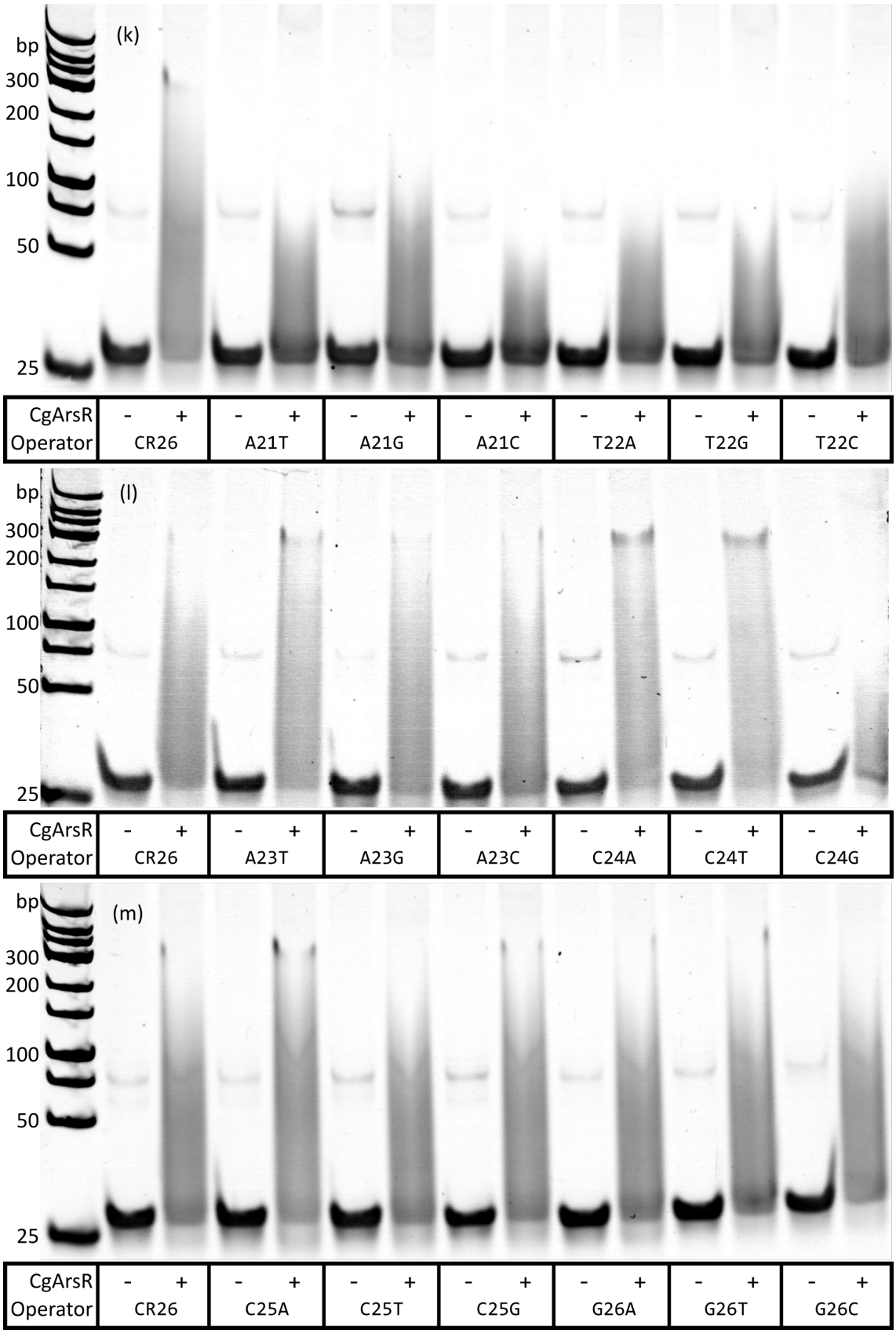


Figure 4.28: **Testing CgArsR binding of single mutant operator library.**

A wealth of information was provided by these experiments. Across the 26 positions only three were judged to be completely sequence-independent, with all four bases resulting in similar binding - positions 1, 13 and 26. A further few bases around the middle and towards the 3' end seemed to have relatively minor roles in determining CgArsR affinity - positions 11, 12, 14, 23 and 25. There were, however, six positions where any mutation away from the bases seen in CR26 resulted in total loss of binding - positions 5, 6, 9, 16, 19 and 20. Specific mutations in a few adjacent positions could also abolish binding, but alternative mutations in the same place would have a smaller effect on affinity. The bases around these still had a significant influence on binding, able to slightly or substantially change affinity depending on the precise change. Finally, in six positions (bases 7, 10, 18, 23, 24 and 25) particular alterations showed promise in possibly increasing affinity relative to CR26. Each shift is coloured in Figure 4.29 (b), while the overall behaviour of each position compared to the CR26 benchmark is given in Figure 4.29 (c).

The different bases in CL26 were subsequently highlighted according to these groupings, with the result in Figure 4.29 (a). This suggests that two bases provide slightly improved affinity, while one alteration reduces it - these could easily cancel each other out, or more likely result in only marginally divergent affinities that single EMSA lanes lack the resolution to differentiate. As all of the groupings were the result of subjective assessments, and not quantified changes, they could easily be slightly misjudged so only the major losses in affinity are reliable.

The crucial positions alone provided an interesting story. Bases A5, T6, A19 and T20 were the innermost pairs from the originally identified 'ATAT' inverted repeat - together with A9 and T16 and the intervening bases they did seem to make an inverted repeat, albeit imperfect, somewhat validating the initial hypothesis about the expected behaviour of homodimeric repressors. Indeed, carrying across the mutations with the strongest suggestion for improving affinity results in the sequence in Figure 4.29 (d), which contains a perfect 10bp inverted repeat with the most important bases in the middle 8bp of each. Figure 4.29 (e) takes this new reference sequence and reapplies the same colouring to demonstrate the likely effect of alterations at each position from this theoretically higher affinity operator - this reinforces the symmetry both of the sequence but also the relative importance of certain bases. Interestingly this suggests the 'middle' of the operator is between bases 12 and 13, not 13-14, but

the sequence isn't fully symmetrical around this point. The 3' end of the operator is slightly longer than the 5' end making it 26bp not 24bp and base 24 had an unexpectedly strong influence on affinity. While the C24A and C24T mutations seemed to notably increase CgArsR affinity, C24G almost completely abolished it, completely unlike the neutral behaviour of base 1. So while the EMSA screen reveal strong symmetry in sequence dependence, this was clearly not absolute.

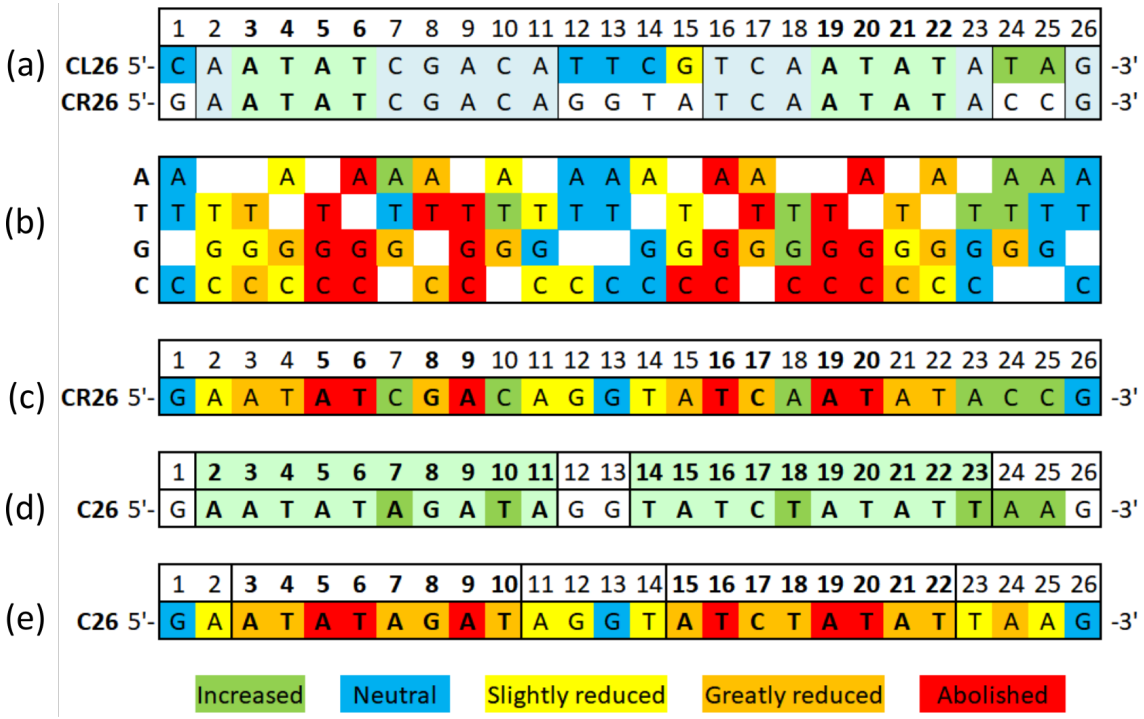


Figure 4.29: **EMSA summary for 26bp single mutant operator screen.** Assessment of CgArsR binding to single mutant operators compared to native CR26, as seen in Figure 4.28. Altered bases are highlighted according to the coloured legend above, depending on the effect of their introduction. (a) The native operators are highlighted as in Figure 4.1, with identical bases in light blue and the inverted repeat within these highlighted in light green with bases and positions in bold. The different bases in CL26 are further highlighted as in the legend. (b) Assessment of CgArsR binding to each single mutant operator compared to native CR26. (c) A condensation of the results relative to the CR26 operator - highlighted according to which colour grouping best fit each position. Any potential improvement was given green, while red was only chosen if all other bases lost binding. (d) A mutant operator sequence theoretically optimised to include all changes that could improve affinity (dark green). Bases forming an inverted repeat are coloured in light green underneath this. (e) The same optimised operator shown in (d), but with bases highlighted according to the likely effect of changes away from this new reference sequence at each position. The core of the inverted repeat, containing the least variable residues, is in bold and bordered.

Another important takeaway to note was the unknown effect of combining these individual mutations. Given small mutations are unlikely to change the lateral

position of CgArsR on its operator, combining mutations could be assumed to have a degree of independence, but to what extent and if this were to be additive or multiplicative is unknown. The EMSA screen did, however, back up the results of the first round of targeted mutations. And while two of the alterations in the CR30-GtoC operator should have negative impacted its affinity, none were the 'lethal' mutations at crucial bases. The CR30-CtoG operator, however, included four changes that each should have greatly reduced affinity - together they effectively abolished it. The CR30-SpaceCAGC operator included the lethal A15C mutation, explaining its failure to bind, but interestingly the non-binding CR30-SpaceACAT operator only included two slightly reducing mutations - T14A and A15T - the combination of which seemed to be greater than the sum of their individual effects.

The potential affinity increase from combining six small changes could, therefore, be meaningful. However, as has been noted previously, strong secondary structure is prohibitive to efficient DSD reactions and the 10bp perfect inverted repeat in this 'optimised' sequence consequently results in a very strong hairpin in corresponding single-stranded displacing oligos. A careful balance, therefore, needs to be struck between removing secondary structure and diminishing CgArsR affinity.

4.4.4 GtoC operator forms a very strong, unpredicted structure with 6d toehold

The primary reason behind the above single mutant screen was to enable the design of structureless sequences that still bound CgArsR with reasonable affinity. The 'three-letter' GtoC mutant, assayed in Figure 4.27, was previously shown to bind with only slightly reduced affinity, despite 5 changes within the core 26bp sequence. Its reverse complement, upon which a displacing oligo would be based, would only form a weak hairpin through an 'ATAT' inverse repeat as a result of preventing any G-C base-pairing and therefore this operator sequence provided a useful template from which various structureless operators could be tested.

The CR30-GtoC-T4C operator, further mutated to remove even the weak hairpin, bound to CgArsR noticeably weaker than the GtoC operator alone (Figure 4.27 again). The single mutant screen confirmed that any change to the ATAT repeat

(bases 3-6 and 19-22) would likely reduce affinity to a similar extent or even more, so the CR(GtoC)26 operator was selected for initial analysis. Its sequence is shown in Figure 4.30, with colouring to show mutations from CR26 - only two out of the five changes were suggested to reduce its affinity.

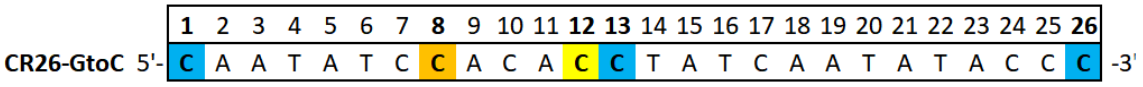


Figure 4.30: **GtoC operator only has two mutations reducing CgArsR affinity.** CR(GtoC)26 mutant operator sequence, with bold and highlighted bases having been mutated from G, in the native CR26 operator, to C. These altered bases are coloured according to whether their independent mutations didn't affect (blue), slightly reduced (yellow), or noticeably reduced (orange) CgArsR affinity, as in the Figure 4.29 (b) summary.

Combining the reverse complement of the CR(GtoC)26 operator with the X6a toehold used for native Tprobe reactions, however, would have been ill-fated. X6a-sCR(GtoC)26c (5'-GCCCGC-GGGTATATTGATAGGTGTGGATATTG-3'), with three consecutive cytosine residues in the toehold and three consecutive guanines in the mutant operator complement, would have formed an exceptionally strong dimer. Instead, a new 6d toehold - the reverse complement of 6a - was used to produce the X6d-sCR(GtoC)26c displacing oligo (5'-GCGGGC-GGGTATATTGATAGGTGTGGATATTG-3'). This maintained the 100% GC content of the toehold while minimising cytosine residues in the displacing oligo, and preventing runs of four or more consecutive guanines that could enable G-quadruplex formation. This X oligo only contained the weak hairpin seen in the toehold-free sequence - a comparison of these three structures can be found in Figure A.4.

A new Tprobe - TR(GtoC)26-G6d - was annealed using these components. Its reaction with different length displacing oligos is shown in Figure 4.31. Contrary to the expectation of a fast reaction, signal increase was incredibly slow across all oligo lengths. After three hours the X24 oligos had caused less than half of the expected maximal signal increase, while the shorter displacing oligos had barely induced any rise. The assay was briefly resumed two hours later, by which time only minimal further gains had been made. The following morning, roughly 23 h after the reactions had been triggered, the RFU for all samples were still far below those seen for TL26-G6a, suggesting that the samples had still not reached their equilibriums. Reordered oligos and repeated assays confirmed that this result was accurate for the intended probe and not a mix-up of components.

The most likely explanation for such a slow reaction would be a great deal of structure within all the displacing oligos. If a very strong hairpin or dimer were to form then there would only be a small thermodynamic gain from triggering displacement, so the final equilibrium state is reached only after a long time. However, such an occurrence would be polar opposite of both the design intention and the NUPACK simulations. It would also be unexpected for reactions using different length X oligos to be slowed largely proportionally and not in the length-dependant manner seen in Figure 4.24 - this suggests similar structure limited to the 5' end of the sequence maintained in all the oligos.

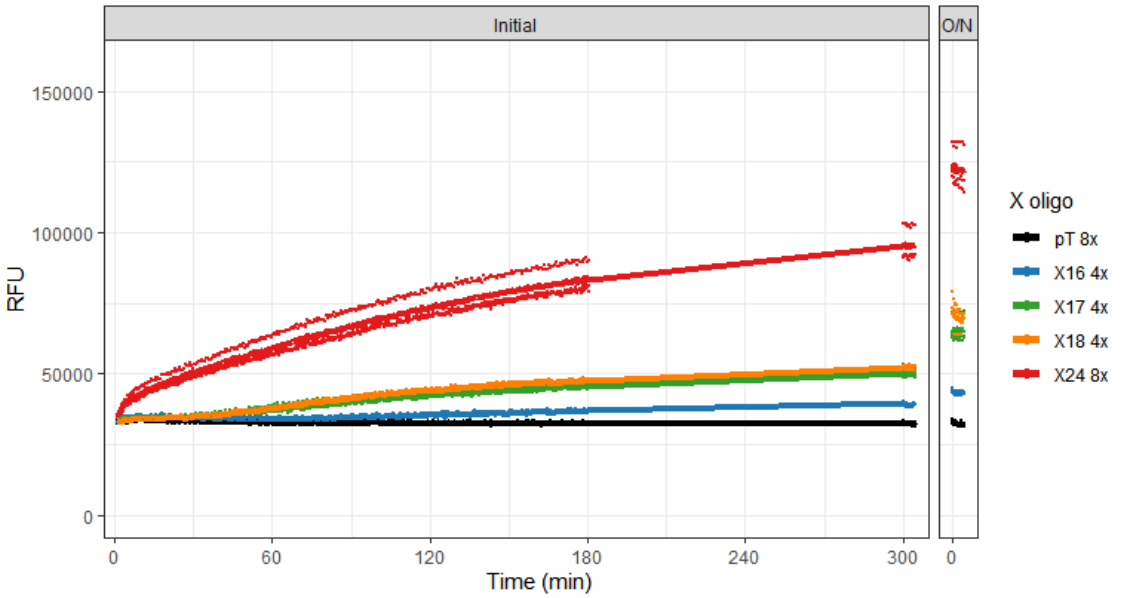


Figure 4.31: **GtoC mutant probe reacts unexpectedly slowly.** Fluorescence time courses displaying the RFU for unbound TR(GtoC)26-G6d responding to corresponding displacing oligos of different lengths. Graph shows readings from individual wells as points and averages across each set of three replicates as lines. Each well contained 1.5 µl of 8/16 µM X oligo stock (12/24 pmol) to which 200 µl of 10 nM TR(GtoC)26-G6d (2 pmol) was added to trigger the reactions, for a total volume of 201.5 µl per well of 96-well plate. The initial assay was run for 3 h, then for 10 min two hours later. O/N data is from reading the same plate for 10 min the following morning.

Native and denaturing acrylamide gels were run to test for the presence of unintended structures - these are shown in Figure 4.32. The denaturing gel in (b), which should prevent the formation of any structures, confirmed that each displacing oligo was their expected length (22-30nt containing a 6nt toehold) for both probes. However, the oligos appeared very different on the native gel. The CL oligos all appear as bands having run similarly to the 25nt ssDNA standard, despite their variation in length and structure.

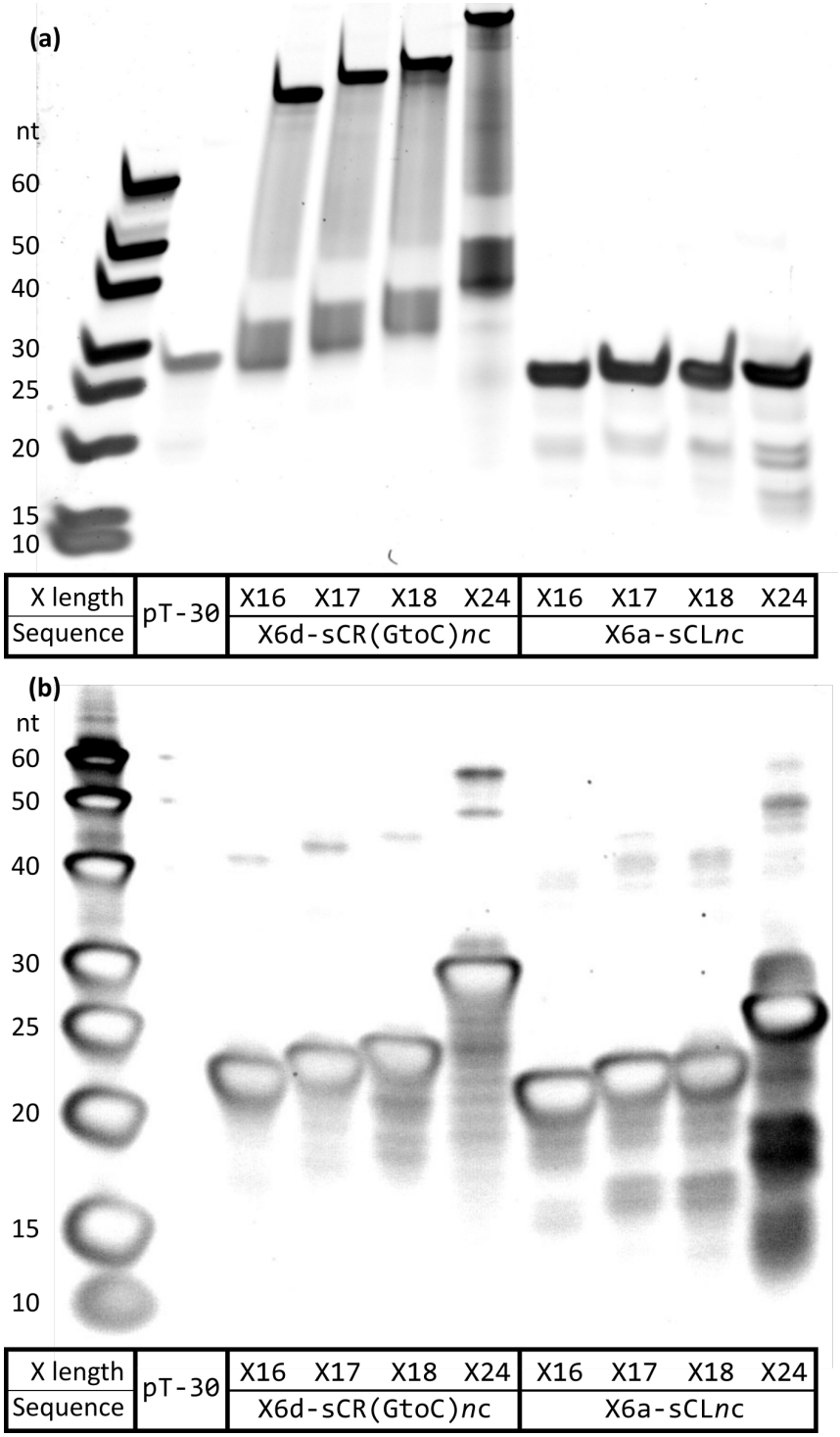


Figure 4.32: **X6d-GtoC displacing oligos are highly structured.** Two PAGE gels comparing the structure of two sets of displacing oligos - X6a-sCL*n*c and X6d-sCR(GtoC)*n*c, where *n* is 16/17/18/24. Actual oligo lengths are therefore 22/23/24/30nt, with pT-30 a 30nt poly-thymine oligo. Each lane contained 32 pmol of the specified oligo and a ssDNA ladder was used for reference. Gels were stained with 1x SYBR gold for 50 min before visualisation. (a) 20% native PAGE TBE gel, run for 50 min at 200 V at room temperature. (b) 15% denaturing TBU gel, run for 90 min at 180 V at room temperature. Samples were incubated in a denaturing buffer containing formamide for 5 min at 95 °C before loading onto the gel.

While ssDNA gel migration can be hard to predict and is not fully understood, structureless single-stranded oligos tend to run slower than their dsDNA counterparts while ssDNA with internal hairpin structures tend to run quicker than those without [150]. This is likely a result of dsDNA stiffness causing it to migrate in a rod conformation, while unstructured ssDNA is more open and this greater hydrodynamic radius slows its movement. In Figure 4.25 the structure of the X6a-sCLc oligos were compared, and it is possible that the tight hairpin structure of X6a-sCL24c compensated for its greater length, resulting in its similar migration to its shorter derivations.

Although the sequences of the ssDNA ladder are unknown, they do have balanced base content and were spaced at relatively even intervals. Comparing their migration to the CL oligos suggests they have intermediate structure - perhaps some weak hairpin formation but less than that seen in X6a-sCL24c. The behaviour of the GtoC oligos, however, does not fit within this paradigm as they all migrated remarkably slowly - all far less than the much larger 60nt control and the X6d-sCR(GtoC)24c oligo barely entered the gel at all. This suggested that they are not forming hairpins, which would speed up their migration, nor are they forming straightforward dimers as this would at most halve their migration speed, not almost prevent it. This result therefore indicated the GtoC oligos were all forming strong, unusually large structures, greatly slowing their migration through acrylamide gels and causing their low DSD rate.

Other than hairpins and dimers, the only other common ssDNA structure is the G-quadruplex. This phenomenon is not fully understood, but avoiding runs of four or more consecutive Gs was known to be good practice to avoid such structures, and was employed earlier when designing the 6d toehold. However, the creation of the GtoC operator still resulted in a 'G-rich' displacing oligo, for example 5'-GCGGGC-GGGTATATTGATAGGT-3' for X6d-sCR(GtoC)16c. In particular, three guanines (complementary to the three cytosines at the 3' end of the mutant operator) were adjacent to the toehold and this resulted in a 5' end with Gs at 7/9 positions.

To investigate the possibility that this sequence resulted in G-quadruplex formation, a selection of ten alternative 22nt oligos were designed that were likely to form either G-quadruplex or dimer structures, then run on a native TBE gel alongside X6d-sCR(GtoC)16c. These combined the X6d (GCGGGC),

X6dt (GTGGGT), X6g (GGGGGG) and X6gc (GCGCGC) 5' toeholds with the 'GtoC' (GGGTATATTGATAGGT), t16 (TTTTTTTTTTTTTTT) or g3t13 (GGGTTTTTTTTTTTT) 3' domains. Each was named according to their specific pairing, but of course without the context of a DSD reaction all were simply 22nt oligos without distinct domains.

Figure 4.33 shows the result of this gel. Once again X6d-GtoC appeared as a smear with a band larger than the 60nt reference. Adjacent, X6dt-GtoC (with Ts replacing the two Cs in the toehold) migrated quicker, as a thick band close to the bottom of the X6d-GtoC smear. Its appearance was close to the X6gc oligos, which all appeared similar regardless of 3' domain - a thick band around the 40nt mark suggestive of self-dimers formed through complementary toeholds as expected. The X6d-t16, X6dt-g3t13 and X6dt-t16 oligos appeared as if they too formed dimers but these may have been weaker. The appearance of X6d-g3t13 was similar to X6d-GtoC, but not exactly the same, suggesting that most of X6d-GtoC's structure derived from its first 9 bases, but that the latter 13nt did affect this, perhaps disrupting it slightly.

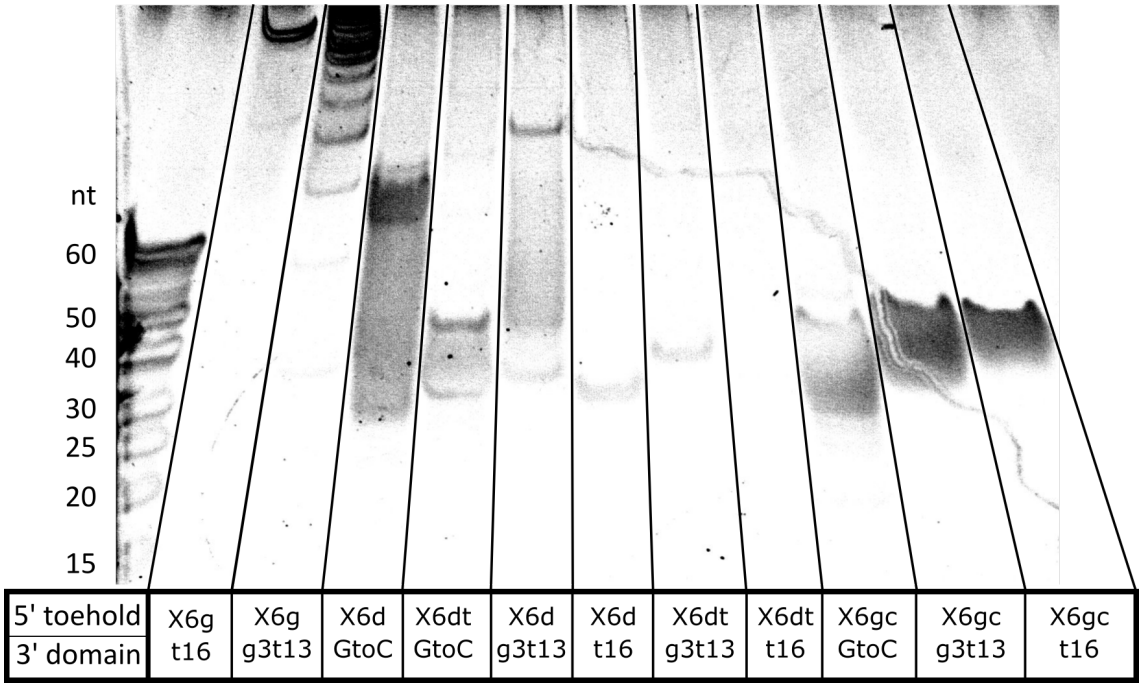


Figure 4.33: **X6d-GtoC displacing oligos may form G-quadruplex/dimer intermediates.** 20% PAGE TBE gel comparing the structure of X6d-sCR(GtoC)16c (abbreviated to X6d-GtoC) to other 22nt oligos. 5' toeholds were X6d (GCGGGC), X6dt (GTGGGT), X6g (GGGGGG) and X6gc (GCGCGC), while 3' domains were 'GtoC' (GGGTATATTGATAGGT), t16 (TTTTTTTTTTTTTTT) or g3t13 (GGGTTTTTTTTTTTT). Each lane contained 32 pmol of the specified oligo and a ssDNA ladder was used for reference. Gels were run for 1 h at 200 V while refrigerated and stained with 1x SYBR gold for 20 min before visualisation.

The two X6g oligos formed incredibly slow migrating structures - their 6/9 consecutive guanines were expected to form G-quadruplex structures but this effect was surprisingly strong. While X6g-t16 seemed to form a single dominant species, the longer run of Gs in X6g-g3t13 appeared to allow multiple conformations to form with slightly different migration speeds. The overall result of this analysis was therefore that several features of X6d-GtoC seemed to contribute to its final behaviour. The toehold alone may have allowed for some weak dimerisation but this was enhanced by the GtoC operator. For some reason separating the runs of Gs by a C instead of a T enabled G-quadruplex-like structures to form. More research into G-quadruplex topology revealed that multi-G repeats, even shorter than four in a row, can allow dimeric and tetrameric complexes to form parallel or anti-parallel quadruplex planes [151]. It seemed likely, therefore, that the X6d-GtoC oligos formed hybrid structures in-between conventional dimers through G-C base pairing and multimeric quadruplexes - indeed each oligo population may have contained a dynamic equilibrium of many such structures. These conformations are clearly missing from software like NUPACK, which predicted the X6d-GtoC to be almost structureless. Interestingly, a few groups have attempted to exploit the coordination of metal ions by these quadruplexes to use them for biosensor applications, although their sensitivity and selectivity are often weaker than other systems [152].

4.4.5 Reversed GtoC operator enables fast DSD reactions

The solution to this problem of structure formation from G-rich displacing oligos was to switch the operator strands within the Tprobe. By taking the reverse complement, the top strand in the probe became sCR(GtoC)26c (5'-GGGTATATTGATAGGTGTGGATATTG-3') with the gate toehold to follow the 3' end. As this confused the oligo/probe naming system, it was subsequently referred to as sCR0C26. The X oligos then contained different lengths of sCR0C26c (5'-CAATATCCACACCTATCAATATACCC-3', the operator sequence in Figure 4.30), with the toehold preceding their 5' ends. Crucially, this limited G-rich ssDNA to the gate toehold of the probe (which at 6nt is unlikely to form complex structures) while the displacing oligo is instead C-rich so unable to form any G-quadruplex conformations.

To test this design alteration, three Tprobes were annealed using different

toeholds - 6a, 6c and 6m. The 6a toehold (GCCCGC) was used previously, but the operator reversal meant that only one C was now at the 5' end of the X oligo's recognition domain, enabling the new 6c (CCCGCC) to use one fewer G and still prevent four consecutive Gs in the complementary gate toehold. The 6m toehold (TCTCCA) was also generated to test the effect of a medium strength reaction. The full length displacing oligos for all these variants were predicted to form only the weak ATAT hairpin, giving an MFE of −0.18 kcal/mol, while shorter oligos had none at all.

Figure 4.34 shows the result of this assay. All of the reactions reached their equilibrium within the first 5 minutes - far faster than any up until this point and a dramatic improvement over the day-long original GtoC probe. In the X24 reactions, the 6m probe was slowest as expected, but 6c was incredibly fast, peaking in time for its first read only one minute after addition. Its X18 reaction was also faster than the 6a probe, suggesting the removal of that second G was beneficial. The X18 length combined with the 6m toehold reduced the equilibrium to a similar level as the X16 oligos with the stronger toehold - just above the background.

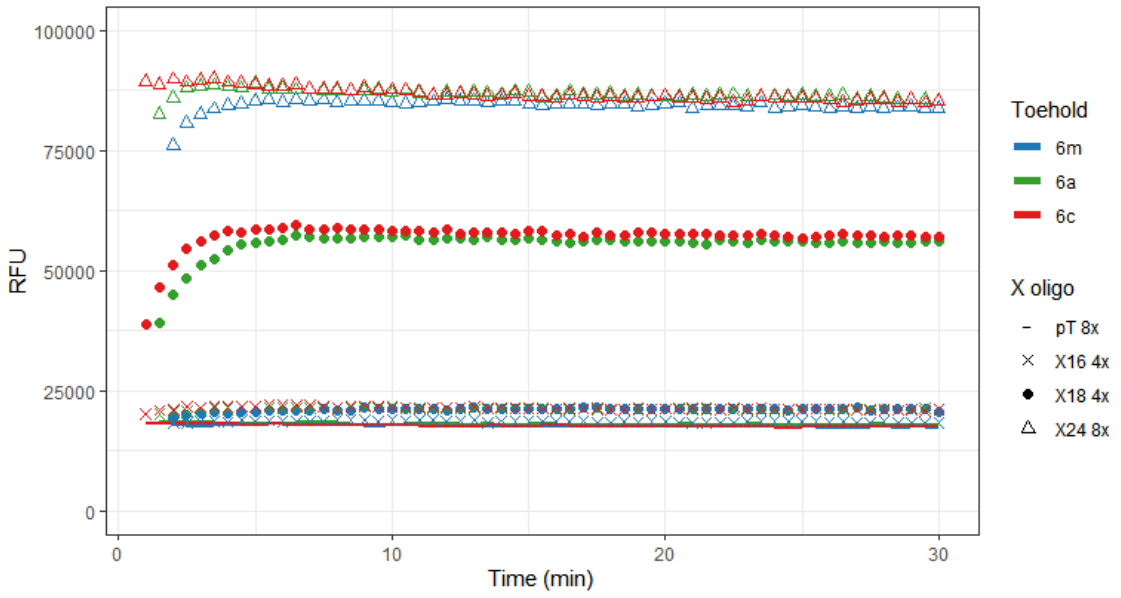


Figure 4.34: **Reversed GtoC operator probes react very quickly.** Fluorescence time courses displaying the RFU for unbound TR0C26 probes with G6m/G6a/G6c toeholds responding to corresponding displacing oligos of different lengths. Graph shows averages from three replicates as points, with minimal variability between the individual wells. Each well contained 1.5 µl of 8/16 µM X oligo stock (12/24 pmol) to which 200 µl of 10 nM Tprobe (2 pmol) was added to trigger the reactions, for a total volume of 201.5 µl per well of 96-well plate. Start times are staggered for each probe due to the time take to switch mixes.

This substantial improvement in both the speed and simplicity of the kinetics for these reactions demonstrated how useful it is to remove secondary structure from DSD probe design. These subtle changes reduced reaction time from hours to a few minutes - a feature very desirable for *in vitro* biosensors.

To test whether a probe using this mutant operator could actually function with CgArsR and respond to arsenite, a titration was set up with three concentrations of its X18 oligo. The data for this are shown in Figure 4.35. In this assay, CgArsR was added to one well per reaction after 41 minutes, then arsenite was added to all wells once the new equilibria had been reached. Again, the initial DSD is very quick with much cleaner dynamics than the similar assay with the native operator probe in Figure 4.21. It took roughly 20 minutes for the wells with added CgArsR to stabilise at their new RFU, half the time of the native probe, then a clear, steady signal increase was seen in these samples once arsenite was added, unlike the ArsR-free wells which quickly adjusted to the dilution effect of the arsenite solution then maintained rather than increased subsequent signal.

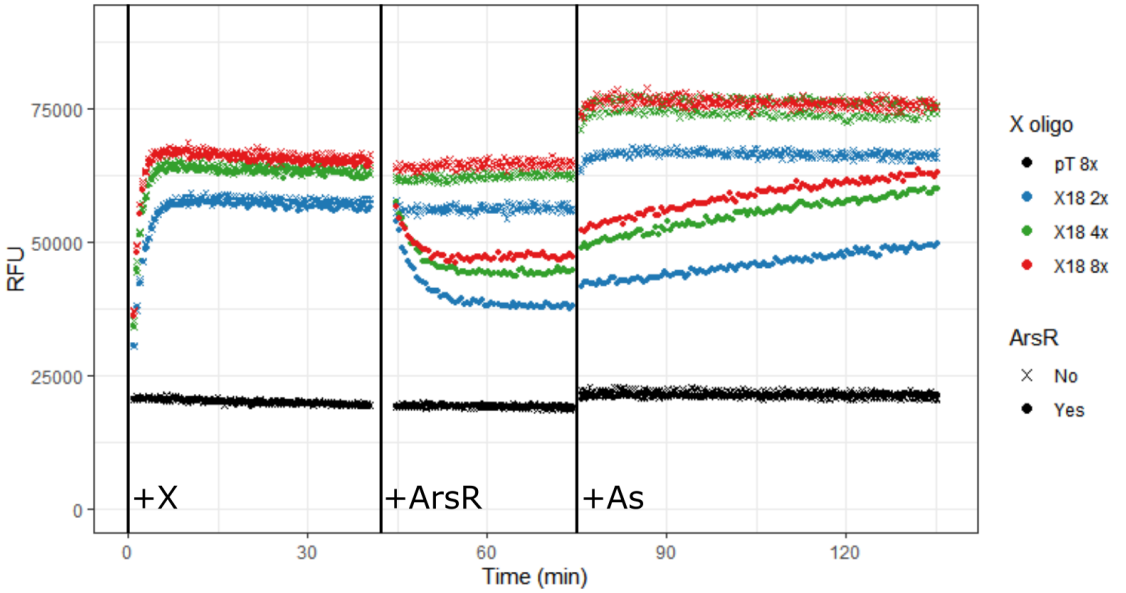


Figure 4.35: **CgArsR can modulate mutant operator DSD reaction.** Fluorescence time course displaying the RFU for probes reacting to different, and variable concentrations of, X oligos. Graph shows readings from individual wells as points. Each well contained 3 µl of 2/4/8 µM X oligo stock (6/12/24 pmol pT-30/X6c-sCR0C18c) to which 200 µl of 10 nM TR0C26-G6c (2 pmol) was added to trigger the reactions (+X). 200 pmol CgArsR (2.4 µl) was added to the 3rd well of each reaction after 41 min (+ArsR) and 50 µl of 10 µM sodium arsenite was added to all wells after 75 min (+As) for a total volume of 253/255.4 µl per well of 96-well plate.

This assay demonstrated that using mutant operators in a DSD system was not only possible, but created clear advantages over the native sequences. However,

one feature of these last assays was cause for concern. The maximum RFU for the X24 samples in Figure 4.34 was under 90k, far below the 130k seen in Figure 4.31 and 150k in Figure 4.24 for the other 96-well assays with the same fluorophore content. The reason for this sizeable difference and the deeper problem it posed is investigated in the following section.

4.5 Controlling the fluorophore environment

A potential disparity in the inherent fluorescence of particular species had been identified earlier in this project. An assay using the native TR26-G6a probe and comparing its reaction with different length displacing oligos is shown in Figure 4.36. The aim of this experiment was to identify the various signal equilibria that each may form, and the speed at which this would be reached. However, it instead revealed that reaction with its full-length X26 oligo caused substantially more fluorescence than the same reaction with X25 (219468.3 versus 161090.3 average RFU, a 36% increase), despite both being expected to reach close to 100% reaction.

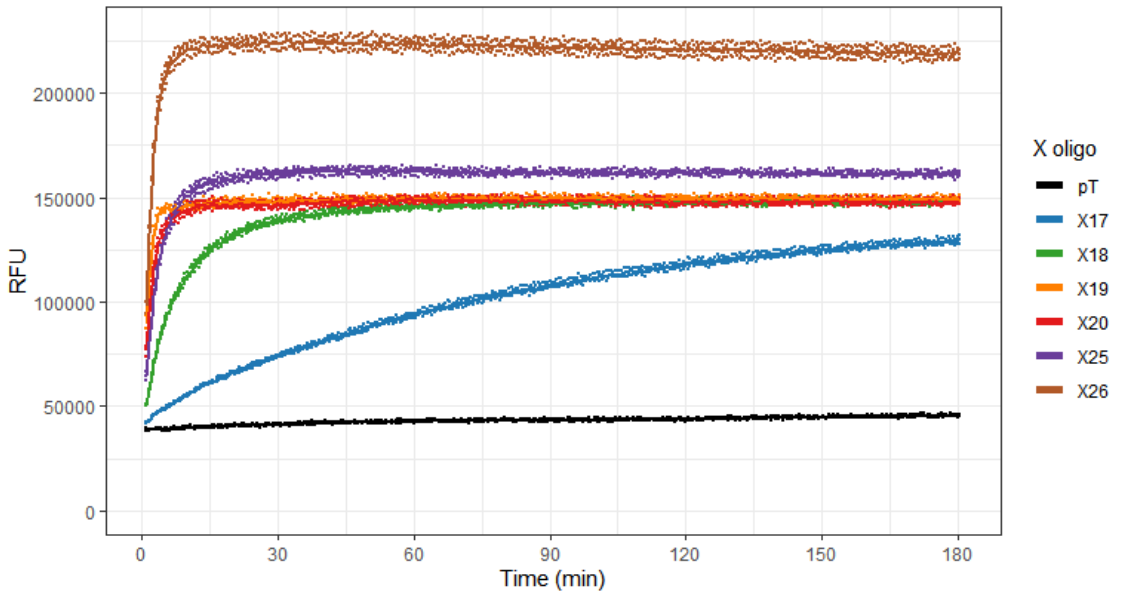


Figure 4.36: **Full length TR26-G6a DSD results in much greater fluorescence than X25.** Fluorescence time course displaying the RFU for TR26-G6a reacting to different lengths of its complementary X oligo. Graph shows readings from individual wells as points and average of triplicates as lines. Each well contained 1.5 µl of 8 µM X oligo stock (12 pmol pT-30/X6a-sCR*n*c, where *n*=17/18/19/20/25/26) to which 200 µl of 10 nM TR26-G6a (2 pmol) was added to trigger the reactions, for a total volume of 201.5 µl per well of 96-well plate.

Even X20 may have been expected to react fully, with its reverse toehold (GAATAT) much weaker than its forward toehold (GCGGGC), so it was suspected that the products from the longer X oligo reactions may have anomalously high fluorescence. Having a good grasp on the maximum possible RFU of each reaction is important for these reversible systems, as it enables you to infer the equilibrium point of the smaller X oligo reactions. If the output signal is in fact a combination of varying equilibria and varying product fluorescence then the true equilibrium point would be unknown.

For this reason the X24 oligo had been used for a while as proxy for a maximal reaction. Even using this, however, led to some questionable results - the X16/X17/X18 oligos for TL26-G6a (Figure 4.24) reacted almost fully, while the X18 oligos for TR0C26-G6c seemed to react only approximately 60% (Figure 4.34) despite similar strength of forward and reverse toeholds.

This observation, combined with much lower RFU values in the TR0C26 probes, prompted this investigation into the varying fluorescence of DSD products. All the Tprobe reactions generate the FGX complex as the fluorescent species (as shown in Figure 4.5 (a)), but the exact appearance of this species will vary with different displacing X oligos. One hypothesis, therefore, was that the longer X oligos produced FGX products where the fluorophore was sterically hindered, and this somehow enhanced their fluorescence. However, no mechanism could be found in the literature for such a phenomenon.

Another hypothesis was that instead of the longer X products having enhanced fluorescence, the shorter X products may have reduced fluorescence. A potential mechanism for this was local quenching from unpaired nucleotides - this would correlate with the observation that FG-X26, with no unpaired bases, had higher fluorescence than FG-X25 and other products, which all had unpaired bases. This phenomenon was investigated by Noble *et al.*, who tested the effect of different unpaired 4nt overhangs adjacent to FAM or Alexa-488 fluorophores, the latter of which was used in these probes [153]. They found that exposed purines, particularly guanine, exhibited notable quenching and that this effect was roughly additive, with more purines resulting in even lower signal. A single exposed guanine could therefore reduce the quantum yield of a 5' Alexa-488 fluorophore to 90-93% of a probe with no 3' overhang, while four consecutive Gs brought this down to 66%. An AAAA overhang decreased quantum yield to 87%, while CCCC or TTTT actually increased fluorescence by 5-7%.

These results seemed to correlate with the observations made across the various Tprobe reactions. The moderate signal reduction seen in the native probes could be derived from exposure to adjacent unpaired GAAT/CAAT bases (for an X22 oligo reaction or shorter, relative to full DSD), while the stronger signal reduction in the reversed GtoC operator could be because of exposure to its more quenching GGGT sequence. The native unpaired bases were not explicitly tested by Noble *et al.*, but would be expected to quench signal by 5-15% if effects are roughly additive, while GGGT reduced fluorescence by 27%.

4.5.1 Introduction of polyT spacer reduces fluorescence variation in FGX products

To test this possibility, unquenched FG duplexes were annealed using three different gate sequences. The first was Fa-sCR0C26-G6c, the gate oligo used to assemble the TR(GtoC)26c-G6c probe used in Figures 4.34 and 4.35. The other two introduced TTTT (4T) or TTTTTT (6T) spacers separating the fluorophore adapter from the operator. These designs should ensure that, regardless of X oligo length, the fluorophore would always be adjacent to the same unpaired polyT bases, which could even enhance the signal. While a 4T spacer should be enough to ensure removal of a significant quenching effect, Noble *et al.* suggested that interaction could be possible with the two bases further away, so a 6T design was also made to test if it further improved signal. The use of unquenched FG duplexes instead of Tprobes removed the complication of reversible equilibria that would affect the concentration of FGX products formed, while an FG + X reaction was chosen over annealing the different FGX products independently to limit the sources of variation in fluorophore addition to a single pipetting step from one communal master mix. The reactions and products for 6T and spacer-free duplexes are compared in Figure 4.37.

The result of this experiment is shown in Figure 4.38. After 30 minutes, the fluorescence of the unreacted (+ polyT) spacer-free FG duplex was similar to that for reactions up to the X22 oligo - in the 82500-85000 RFU range. The average fluorescence for the same reaction with the X23, X24, X25 and X26 oligos, however, was 85912, 88363, 95612 and 125391 RFU, reflecting a substantial increase in signal above baseline from the two longest displacing oligos in particular. As the same amount of fluorophore was added to all these

samples and no quencher modifications were present, this confirmed that the inherent fluorescence of these products was greater than their smaller X oligo counterparts, likely due to the presence of unpaired guanines quenching signal intensity in these latter species.

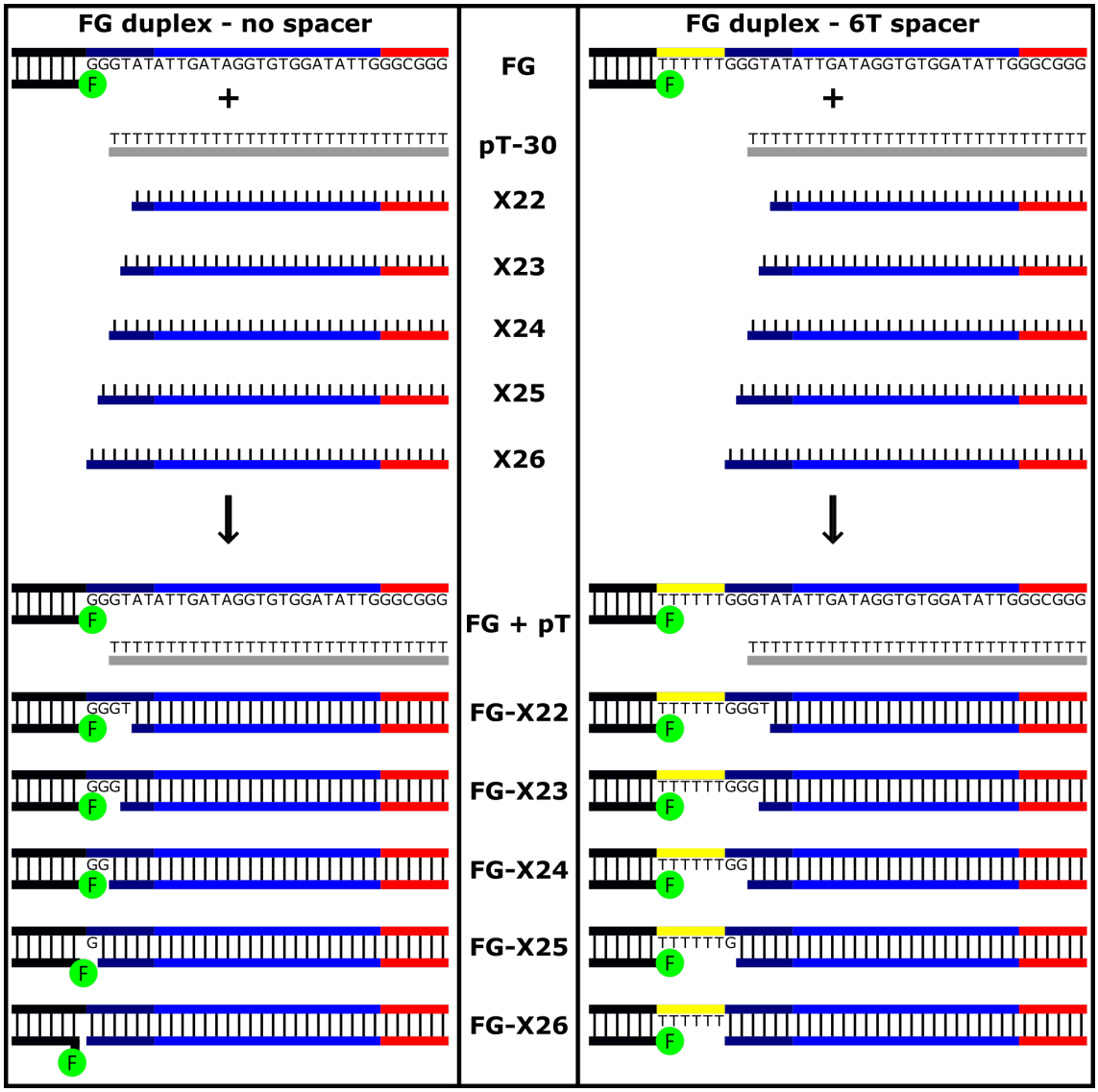


Figure 4.37: **polyT spacer improves consistency in FGX fluorophore environment.** Schematic displaying the various products formed when the F-CR0C26-G6c (or F-6T-CR0C26-G6c) duplex is combined with pT-30 or complementary displacing oligos of different lengths, as per the reactions seen in Figure 4.38. The sequence of any unpaired bases on the gate strand is written to reveal the immediate environment of the fluorophore in these different species. The FGX products are the same as those resulting from Tprobe + X oligo reactions, but the QR duplex is also produced allowing the reactions to be reversible. The universal FG arm is coloured black, with the forward toehold red and the polyT spacer yellow. The CR0C26 CgArsR operator is coloured blue, with the dark blue subset indicating the reverse toehold assuming an X20 reaction. The length of the universal FG arm has been artificially shortened from 24bp to 6bp to improve clarity.

The species with four or more unpaired bases, therefore, displayed 66-68% of the maximum signal for this construct, close to the 73% seen by Noble *et al.*. Using the probes containing a polyT spacer, however, massively reduced the variation in signal between the FGX species, with each resulting in at least 97% of the signal of their respective FG-X26 products. With both the 4T and 6T spacers, the species with the three highest RFU were still FG-X24, FG-X25 and FG-X26, suggesting the unpaired operator bases may still enable a minor quenching effect from further away, but this effect was much smaller than the range seen before.

Another observation was that the fluorescence of the FG-X26 species between the three probe designs differed, suggesting either variation in master mix preparation or the polyTs did enhance fluorescence relative to the fully double-stranded spacer-free conformation. It was noted that the maximum signal of all these systems was still substantially below the X26 reaction seen in Figure 4.36, suggesting that variation in local paired nucleotides may also affect RFU, but limiting the variation between FGX products from the same probe was more important than maximising the end signal - for this reason the 6T spacer design was incorporated for all future Tprobes.

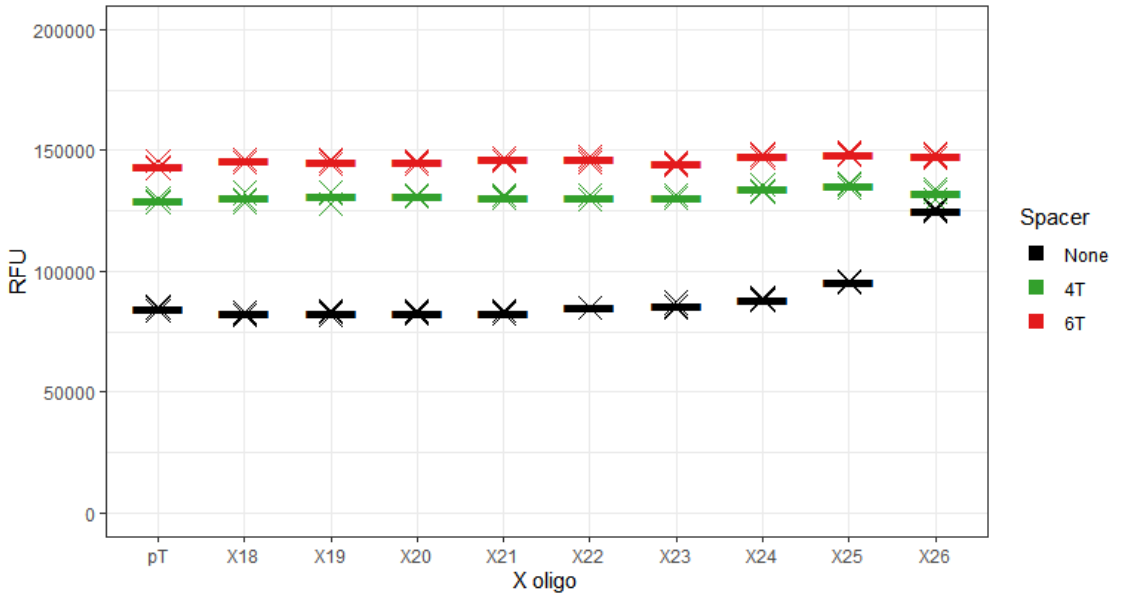


Figure 4.38: **polyT spacer increases signal and removes variation in FGX fluorescence.** Fluorescence end point reads displaying the RFU for F-CR0C26-G6c duplexes with no/4T/6T spacers annealing to different lengths of its complementary X oligo. Graph shows readings from individual wells as crosses and average of triplicates as lines. Each well contained 3 µl of 4 µM X oligo stock (12 pmol pT-30/X6c-sCR0C*n*c, where *n*=18-26) to which 200 µl of 10 nM FG duplex (2 pmol) was added to initiate annealing, for a total volume of 203 µl per well of 96-well plate. Readings were taken 30 min after probe addition.

4.6 Conclusions

Throughout this chapter, several probe design alterations have been explored and integrated. Testing these relied on the use of a Tprobe structure, which enabled fast and cheap comparison between different functional sequences. Transitioning from an irreversible to a reversible DSD system prevented the accumulation of background signal while allowing the natural on/off rate of repressor binding to guide design rather than limit it. This also enabled a shift towards developing design principles for any such repressor-operator DSD system, rather than simply trying to iteratively improve creation of one specific biosensor.

Removing secondary structure was a big focus in this chapter - both from the operator sequence, the toehold, and possible interactions between the two. The mutant library screen unearthed some interesting sequence dependencies within the CgArsR operator and guided the design of new operators containing multiple point mutations to remove structure while preserving function. After identifying an unexpectedly structured toehold sequence, the importance of a C-rich displacing oligo was established and the resulting design displayed impressively fast kinetics while maintaining integration with CgArsR.

Introducing a polyT spacer was the final substantial alteration to probe design. This corrected a subtle variation in signal between different DSD products, which misled identification of equilibria for reversible reactions. Standardising the fluorophore environment in this way allowed a more quantitative approach to be taken when probing other system variables and this is explored in the following chapter.

Chapter 5

Enabling system optimisation

5.1 Introduction

While the previous chapter centred on adjusting probe sequences to amend behaviour, this chapter focusses on how other aspects of system design can be altered to improve assay performance. First, changes to experimental protocols were explored in an attempt to reduce noise and increase precision. Second, a normalisation methodology was established in order to translate RFU signals into meaningful data about the state of the underlying system. This enabled more quantitative comparisons to be made between different probe designs and ratios. Third, alterations to buffer components were tested to identify their relationships with different aspects of assay behaviour, while trying to optimise these for overall performance.

Finally, attempts were made to model the fundamental system. While empirical experimentation enabled small improvements to the assay, the low-throughput nature of these experiments limits the number of variables that could ever be tested. A working model, however, would allow many more simulations to be run than could be feasible by hand, potentially generating predictions about how to optimise the overall speed and sensitivity of the system.

5.2 Improvements to assay protocol

Over the course of the DSD assays presented so far, numerous experimental alterations have been made in an attempt to improve both sensitivity and reliability of the system. Initially, reactions were performed in 10 µl volumes in 384-well plates, with components at micromolar concentrations. Usually a master mix of probe and repressor was prepared and split between wells, before addition of X oligo and arsenite solution, in the desired order, to each well individually. While these could be performed in triplicate without prohibitive noise, it was soon clear that adjusting these variables could lead to improvements.

5.2.1 Increasing analyte volume

In section 4.2.5, increasing the volume of arsenite added to reactions from 1/2 µl to 10 µl was found to improve sensitivity and reduce background DSD rate of irreversible reactions. In this case the 2-fold overall dilution from increasing the total solution volume to 20 µl, and the resulting slowing of reaction speeds, was outweighed by the 5 to 10-fold increase in arsenic content that could be added. For a set concentration of arsenic stock, the improved ratio of arsenite to CgArsR in the final solution inevitably boosted sensitivity.

While *in vitro* assays commonly deal with small volumes, primarily due to cost or scarcity of the solution to be sampled (e.g. blood/saliva), this does not have to be the case. Groundwater, the main analyte of an arsenic biosensor, is typically measured in units of litres or greater - scarcity or cost of this component is negligible from a sensing standpoint. However, this does not mean that adding more to a reaction guarantees better sensitivity. Excessive dilution of the DSD reaction presented here would inevitably reduce signal intensity, severely slow detection speed and, perhaps most acutely for this system, reduce the proportion of probe bound by the repressor.

This last point means that the greater the dilution of DSD components, the more CgArsR would be needed to maintain the same reversible equilibrium point. As this does not scale linearly, as you add more arsenic there will eventually be an inflexion point at which you start to lose rather than gain sensitivity. As with many of the adjustments referenced in this chapter, this optimum volume of

arsenic sample may be estimated empirically, but an accurate model might be able to predict this for many different conditions much quicker.

At the same time as transitioning to a reversible DSD system, assays were trialled in 96-well plates with 200 µl volumes and Tprobes at much lower 10 nM concentrations. Despite a net reduction in fluorophore content the solutions produced sufficiently high signals, however a much larger CgArsR excess was needed to generate a comparable equilibrium shift (100x in Figure 4.35; 5-20x previously).

One clear benefit of the change to larger, more dilute probe formulations was improved mixing. Unlike early reactions, where 2-3 µl X oligo was added to a larger volume of probe, instead a small volume (≤5 µl) of concentrated X stock could be pipetted onto the bottom of a well and a much larger volume (200 µl) of probe master mix could be added on top. This reordering and clear volume differential should ensure that X oligo is rapidly distributed evenly around the solution, removing any variation that might derive from heterogeneous oligo dispersion. Initially 50 µl arsenite solution was added to this; later these volumes were reversed - 200 µl arsenite was added to 50 µl of a more concentrated probe + X oligo reaction. The various standardised protocols used are detailed in the Methods chapter. These not only contributed to increased sensitivity, but ensured that at each solution addition a larger volume was added, minimising time until well homogeneity.

5.2.2 Preventing differential fluorophore sequestration

Manual pipetting can be a major source of variability within *in vitro* assays, particularly if other sources of noises are being tightly controlled. As a result, an electronic pipette repeater was employed to dispense master mix solutions to these larger wells but, as seen in Figure 4.31, substantial variation between replicates was still noticeable.

Eventually the reason for this noise was traced back to differential fluorophore sequestration. Oligos are known to stick to the surface of standard wells, thereby reducing the concentration of DNA species in solution over time. This behaviour had been understood early on, so plates with a treated surface reducing such interactions were used throughout the project. However, this

problem resurfaced after moving to larger volumes despite continued use of non-binding plates. In this case, the combitips used for the pipetted repeater seemed to sequester a substantial quantity of DNA itself, including the fluorophore-modified species. When dispensing a large volume of master mix requiring the combitip to be refilled, the wells filled from the first round were invariably lower signal than the others. Here, when filled the first time the combitip sequestered DNA until its surface was saturated, dispensing solutions which had fluorophore removed, while later rounds did not lose any DNA.

Introducing a 1 µM polyT-20 carrier DNA into all the low concentration probe master mixes reduced this problem, while making sure these solutions were recycled through their combitip at least three times before final dispensation into wells removed it fully. Together these protocol adaptations ensured that oligo sequestration primarily occurred to the non-functional carrier DNA over the probe complexes, and if any probe was removed then recycling the solution spread this effect across all the wells equally.

These improvements resulted in assays with minimal variation between replicates. Not only did this enable greater discrimination of arsenite concentrations and probe designs, but allowed a more quantitative approach to be taken with later comparisons. Variation between different experiments, rather than within a single one, is much harder to control, however. Protocol developments can aid to an extent, but a robust normalisation of raw data into standardised, meaningful concentration of underlying components is much more effective.

5.3 Normalisation methodology

In any fluorescence assay, a relationship exists between the chemistry of the target solution and the emission of light at the assayed wavelengths. This can be extremely complex and subject to many sources of noise, but the aim of a normalisation method is to translate the raw RFU data into measurements of the underlying chemistry as accurately as possible. There is a trade-off between accuracy/reliability of data and time/materials needed to produce these, influenced by what kind of data is needed. Fully accurate kinetic data are therefore likely to require more stringent controls than semi-quantitative endpoint measurements.

Adoption of a reversible DSD system using intermediate length displacing oligos enabled development of the normalisation protocol presented here. The methodology centres on comparison of sample wells with the RFU from control wells - the same probe master mix responding to either unreactive polyT oligos or fully displacing X oligos. As the equilibria of these control wells should be unaffected by repressor presence, these should accurately reflect the minimum and maximum signals possible by each system and as they derived from the same master mix as the sample wells, many of the variables known to affect fluorescence should be controlled.

5.3.1 Problems with conventional methods

Normalising to fixed backgrounds/maxima assumes these values are constant over the course of an assay – this is rarely true. The dynamics of these values could be modelled (e.g. using a linear approximation) but this assumes some knowledge of how these values change over time. These days, signal leak from poor DNA synthesis is unlikely to be significant while inefficient quenching, photobleaching, oligo sequestration and localised heating are tangible factors. While poor synthesis is relatively easy to model these other factors are much harder. If not fully understood it is better to constantly measure these values rather than attempt to model them.

It can be hard to identify appropriately accurate controls. When prepared independently from test samples these are susceptible to measurement and pipetting errors as well as variation in storage conditions and final buffers.

Previous work also highlighted that fluorophores are very sensitive to their immediate environment (e.g. sequence/structure). Free fluorophore-modified oligos, purposely annealed products and quencher-free probes all proved unreliable at replicating the behaviour of factors influencing the signal of DSD reactions.

5.3.2 Control DSD normalisation method

The normalisation aim is to determine the 'reaction balance' for each sample, which is equivalent to the fraction of quenched probe that has undergone a displacement reaction to become fluorescent. This is achieved through measuring parallel wells designed to have undergone 0% or 100% reaction, assuming that all wells have received the same amount of initial probe and that each sample represents an intermediate state between the minimum and maximum determined values.

This normalisation uses the average RFU of at least 3 independent wells for both the minimum and maximum values, determined by reacting the same volume of probe master mix used for the reversible oligos with either polyT-30 or full length X oligos (here usually X26), respectively. This is possible because the irreversible maximum DSD control is able to reach full displacement very quickly, and ensures the same final fluorophore, oligo and buffer composition in both the sample and control wells.

A small volume (up to 5 µl) of polyT-30, full length X oligo and sample length X oligo is therefore loaded into wells on a 96-well plate, before all are triggered in quick succession by addition of a larger set volume of homogenous master mix (containing pre-incubated DSD probe, repressor, carrier DNA and appropriate buffer) from a multi-dispensing pipette. This is carried out for each independent master mix for every fluorescence assay. If sample wells are subsequently altered (e.g. arsenic addition), these control wells should be similarly altered at the same time.

Variable definitions

For each master mix, at each time t:

F = amount of fluorophore initially added to each well (pmol)

$V(t)$ = volume of every well (µl)

$bRFU(t)$ = background RFU contribution present equally in all wells

$Sq(t)$ = specific RFU of quenched fluorophore (/nM)

$Sf(t)$ = specific RFU of fluorescent fluorophore (/nM)

Therefore:

$$\begin{aligned} minRFU(t) &= \text{RFU of solution of fully quenched probe (0\% rxn)} \\ &= bRFU(t) + \frac{F}{V(t)} \times Sq(t) \\ maxRFU(t) &= \text{RFU of solution of fully reacted, fluorescent product (100\% rxn)} \\ &= bRFU(t) + \frac{F}{V(t)} \times Sf(t) \end{aligned}$$

And for each sample well from this master mix:

$A(t)$ = amount of quenched fluorophore (pmol)

$B(t)$ = amount of fluorescent fluorophore (pmol)

$sRFU(t)$ = RFU of sample solution

Variable relationships

If the fluorophore can only be in two states - quenched or fluorescent - and the amount of probe/fluorophore initially added to each well is fixed and cannot vary over the course of the assay:

$$A(t) + B(t) = F \tag{5.1}$$

Assuming that sample RFU is additive from three independent sources - background signal present equally in all wells, fluorescence from quenched fluorophore and fluorescence from fluorescent fluorophore, then

$$sRFU(t) = bRFU(t) + \frac{A(t)}{V(t)} \times Sq(t) + \frac{B(t)}{V(t)} \times Sf(t) \tag{5.2}$$

Using the definitions of $minRFU(t)$ and $maxRFU(t)$, we can substitute out $Sq(t)$ and $Sf(t)$, leaving

$$sRFU(t) = bRFU(t) + \frac{minRFU(t) - bRFU(t)}{F} \times A(t) + \frac{maxRFU(t) - bRFU(t)}{F} \times B(t)$$

Which rearranges to

$$F \times sRFU(t) = bRFU(t) \times (F - A(t) - B(t)) + minRFU(t) \times A(t) + maxRFU(t) \times B(t)$$

This cancels the background, and substituting $A(t) = F - B(t)$ gives

$$F \times sRFU(t) = minRFU(t) \times (F - B(t)) + maxRFU(t) \times B(t)$$

Finally rearranging for B(t) gives:

$$B(t) = F \times \frac{sRFU(t) - minRFU(t)}{maxRFU(t) - minRFU(t)} \tag{5.3}$$

This gives us a way of calculating the amount of fluorescent fluorophore in a sample well from RFU data and the known quantity of added fluorophore. This can be adjusted to calculate the concentration of fluorescent fluorophore in a sample:

$$\frac{B(t)}{V(t)} = \frac{F}{V(t)} \times \frac{sRFU(t) - minRFU(t)}{maxRFU(t) - minRFU(t)} \tag{5.4}$$

Or the percentage of fluorophore that is fluorescent i.e. the reaction balance:

$$\frac{B(t)}{F} = \frac{sRFU(t) - minRFU(t)}{maxRFU(t) - minRFU(t)} \tag{5.5}$$

These final calculations are similar to those done with other fluorescent systems, but crucially the minimum and maximum RFU values of each solution are continuously monitored in parallel with the sample wells, enabling normalisation specific to the precise conditions at the time of measurement, rather than to average or estimated conditions. The design of the controls themselves are also believed to more accurately replicate the environment of the sample probes than conventional use of fluorophore standards.

5.3.3 Assumptions

The relationships and equations stated above make a number of assumptions about underlying components and their behaviour in the described system. These are stated below for clarity and to outline scenarios in which they did not, or may not, hold true. Some assumptions also may not be perfectly true, but this does not necessarily invalidate the normalisation method. Comparing only raw data or using an alternative standardisation protocol are likely to be less appropriate and make more incorrect assumptions, so this method may represent the best practical solution.

1. Fluorescent signal can be separated into three additive components - background signal present equally in all wells, fluorescence from quenched fluorophore and fluorescence from fluorescent (non-quenched) fluorophore

    (a) This would be untrue if fluorescence could not be independently partitioned into these three categories. In reality, fluorophore fluorescence is unlikely to be truly binary, but this may be a reasonable approximation

    (b) Fluorescent signal deriving from either quenched or non-quenched fluorophores is assumed to be directly proportional to the concentration of each species within each solution. This may not be always true, but if this relationship is close to linear for most of the dynamic range used by the assay then this should be a reasonable approximation

    (c) This assumption would be invalid if these sources of signal were not additive, but interacted in some way or saturated the detector

2. The RFU from the 'minimum signal' wells can be separated into two additive components - background signal present equally in all wells and fluorescence from quenched fluorophore

    (a) Most of the background signal likely derives from non-quenched complexes (fluorescent complexes lacking a quencher strand due to insufficient excesses during annealing or synthesis errors), any external light entering the reader, any other fluorescent components of the solutions (nucleic acids, buffers, proteins etc.) or particulate

contaminants that may alter light scattering or surface tension. This is assumed to be constant in all wells, as it should derive from the same volume of master mix or the displacing oligo buffer. This would be untrue if, for example, the different sequences between the sample and control oligos differentially absorbed/reflected/emitted light at the emission wavelength

(b) The fluorescence from quenched fluorophores encompasses any signal that might be generated from fluorophores still within the same complex as a quencher, but which has still produced light at the emission wavelength (i.e. due to imperfect quenching)

(c) This assumption would be incorrect if the polyT-30 oligos did in fact cause any degree of displacement reaction, increasing fluorescence (e.g. strand 'fraying')

(d) This would also be incorrect if for some reason the minimum wells in fact did not behave comparably to the true background in the other wells (e.g. if they did not heat up, evaporate, sequester oligos or photobleach at the same rate)

3. The RFU from the 'maximum signal' wells can be separated into two additive components - background signal present equally in all wells and fluorescence from fluorescent (non-quenched) fluorophore

(a) The fluorescence from fluorescent fluorophores encompasses any signal that might be generated from fluorophores once within the same complex as a quencher, but which has been separated due to a displacement reaction

(b) This assumption would be incorrect if any probe complexes in the maximum signal wells were not displaced and unquenched at any timepoint (<100% reaction)

(c) This assumption would also be incorrect if the individual product species in the maximum signal wells (usually FG-X26) did not have the same specific fluorescence as the individual fluorescent sample product species (e.g. FG-X19), or were differentially affected/quenched by other solution components (e.g. proteins)

4. Sample RFU can be separated into three additive components - background signal present equally in all wells, fluorescence from quenched fluorophore and fluorescence from fluorescent (non-quenched) fluorophore

   (a) As for the minimum and maximum RFUs, this would be untrue if fluorescence could not be independently partitioned into these three categories

5. Every well (sample, min or max) receives the exact same volume and composition of master mix, at the same time, after addition of a fixed volume of displacing oligo

   (a) Ideally the timing would in fact match the timing offset of the plate reader or could be calculated and compensated, but small deviations from this are likely to be insignificant

   (b) This relies on the master mix being perfectly homogenous (well-mixed) and no loss effects are experienced during transfer to wells

6. Each well within a plate is identical (e.g. no manufacturing defects leading to variable rates of binding and no particle contaminants) and is treated identically other than the addition of different oligos

7. Composition of the X or pT-30 oligo aliquots is identical in every way except for oligo sequence (e.g. same buffer, temperature, inherent fluorescence, possible degradation) and concentration, and that identical volumes of each have been added to sample and control wells

8. Solutions are assumed to be homogenous for every reading, with fluorophores (quenched or otherwise) evenly distributed throughout each well

5.3.4 Identified problems and limitations

The probe background assumption (2c) has failed when the polyT-30 aliquot has either degraded (leading to short oligos potentially able to encourage probe 'fraying'), or has been contaminated with X oligos thereby increasing the determined background and so reducing the calculated reaction balance of sample wells. This could lead to underestimating displacement or exaggerating repressor activity and so raw minimum control values should be inspected pre-normalisation. With standard fluorophore, buffer, concentrations and gain settings this was usually ~30k RFU.

While a significant excess of fully complementary displacing oligo (X26 8x) should rapidly react irreversibly with quenched probe to expose all the fluorophores in a solution, this reaction will not necessarily finish before the first fluorescence reading (invalidating 3b). The presence of repressor protein will often delay 100% displacement, as the bound probes are more stable so undergo DSD more slowly, but the final equilibrium should be unaffected. Likewise, high percentage crowding agents increase the time needed for the solutions to fully mix so it will take longer for all the quenched probes to encounter and react with the displacing oligos (8). If samples are subsequently normalised to a max value that derives from incomplete reactions, the reaction balance of these samples will be overestimated at this timepoint. The normalised sample data would then suggest an overly fast initial reaction until the first timepoint, followed by an overly slow rate until the max control completes its reaction, after which the data should be correct.

If this problem exists, it could be largely solved by triggering the max control wells with the master mix sufficiently in advance of the sample wells to ensure 100% reaction before any wells are read. However, these controls may not reflect the same conditions experienced by the sample wells as a result (5; any sequestering, heating or evaporation may start earlier). If these sources of variation are deemed to be less significant than read-induced sources (e.g. micro-heating, photobleaching) and less significant than the inaccuracy of the maximum control, this could be a worthwhile trade-off. One further inconvenience would be not knowing exactly how long a particular master mix would need for its max control to fully react – this could be anywhere from 30s to over 30 minutes so pilot experiments would be needed to estimate this.

This normalisation method, while improving the accuracy of the processed data relative to other methods, is likely to increase the variance in the resulting sample data. There is noise present in the control data, just as there is in the test samples, and so this normalisation method will compound this variability, whereas normalisation to fixed values or to a continuous model will not. It is therefore important the controls are sufficiently reliable so this increased variance is not problematic.

5.3.5 Methodology development

Independent controls were tested but there were inconsistencies in relative fluorescence and RFU behaviour during assays. Single stranded fluorophore oligos (F), annealed products (FGX), probes annealed with displacing oligo (T/FQ+X) and probes without quencher (T/F-Q) were all tested but none of these seemed to accurately capture probe behaviour. The largest factor likely was having to prepare these separately from sample mixes, enabling differences in pipetting, buffers and storage, but some species clearly had inherent differences in specific fluorescence or dynamics. Modelling to data normalised to such controls proved challenging and so suggested the underlying system was not being represented – this greatly improved following introduction of the new methodology.

Significant redesigns were made to finally ensure that DSD products all had the same fluorescence (assumption 3c). Several experiments highlighted that initial probe design resulted in FGX products where the fluorophore was adjacent to variable exposed bases depending on the length of the displacing oligo. For the FG-X26 control to have the same fluorescence as the sample products (e.g. FG-X19 or FG-X20), it was necessary to introduce a 6T spacer between the operator and fluorophore domains.

Some experiments also highlighted variance between replicates and this was concluded to result from sequestration within a combitip as the master mix was added to wells. Recycling each mix at least three times through the combitip before addition to wells was found to abolish this variation.

5.3.6 Conclusion

This bespoke normalisation method drastically improved the accuracy of processed data compared to previous attempts. The main drawback was having to prepare more master mix for each experiment (at least twice as much), which therefore consumed more DNA, fluorophore, protein and buffer and also took longer to prepare, as well as taking more space on a plate and limiting how many samples could be investigated at once. However, there were a number of occasions where this normalisation method managed to remove unexplained artefacts from the raw data, where a simpler method would have not. At the

moment, therefore, this is a worthwhile trade-off in order to have more confidence in the data and with further adjustments to the experimental design it may be possible to improve this.

5.4 Comparison of probe designs

The previous chapter presented many iterations of probe designs, but a continual theme was the difficulty in accurate and reliable comparisons between them, whether due to confounding variables such as secondary structure or the lack of appropriate controls. Design advances in combination with improvements to protocol robustness and the normalisation methodology presented earlier finally made this possible, allowing quantitative comparisons to be made between small design alterations.

In this section the effect of normalisation is illustrated, demonstrating its ability to control for slight variations in master mix preparations as well as remove minor signal artefacts. Normalised data are then compared between the T6T0C probe, containing the CR0C26 'GtoC' operator, and the T6T1C probe, which reintroduces a single native base, to show how this point mutation can substantially improve CgArsR affinity without creating notable secondary structure. At this point a new Lprobe design is also presented and its benefits over the Tprobe for these later experiments are outlined.

5.4.1 Normalisation removes raw data complexity and provides reaction insights

The protocol improvements discussed earlier enabled the comparison of parallel master mixes, both within and across different experiments. The DSD5 protocol allowed 12 wells to be assayed per master mix (one row of a 96-well plate), comprising three replicates for each of four different displacing oligo formulations. Usually this would take the form of the two control oligos (pT-30 and X26) paired with either two concentrations or two different lengths of reversible DSD oligos, or even two replicate sets of a single X oligo formulation. Two or more master mixes could be compared in one assay, commonly taking the form of different CgArsR concentrations to understand how the repressor

affected the reaction equilibrium and to identify how much was required for good arsenite sensitivity. After 30 minutes, by which time equilibria should have been established, three arsenite concentrations (usually 0/10/100 µM) were added to each set of replicates, including the control reactions, resulting in separation between individual wells for the reversible reactions, observed for the following hour.

Figure 5.1 displays the raw data for such an assay carried out with the T-6T-R0C26-G6c probe (abbreviated to T6T0C) established at the end of the previous chapter, with the 6T spacer integrated into the TR0C26-G6c used in Figure 4.35. The colours represent the two master mixes containing either CgArsR or an equal volume of its protein storage buffer, while the three oligos are identified through their different shapes. Only one reversible oligo concentration (X19 8x) was used, displayed in the middle sandwiched between the polyT-30 and X26 control reactions demarcating the minimum and maximum RFUs for each master mix at each time point. For the initial 30 minutes there are therefore six replicates for each reversible reaction, later split into two replicates for each arsenite concentration per master mix.

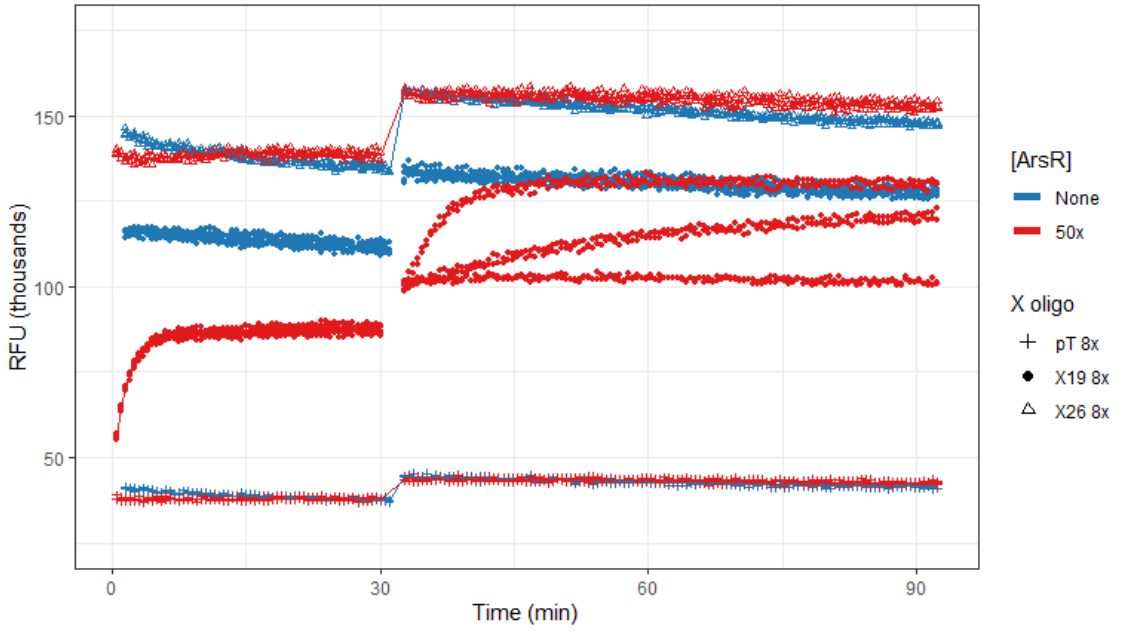


Figure 5.1: **T6T0C DSD reaction raw data.** Fluorescence time course displaying the RFU for T6T0C incubated with or without a 50x excess of CgArsR (over theoretical operator dsDNA), reacting to different X oligos, according to the DSD5 protocol with the DB6 buffer. Graph shows readings from individual wells as points and averages across control triplicates as lines. After 30 min, 50 µl 0/10/100 µM sodium arsenite was added to all triplicate sets. There were two sets of X19 reactions, both shown, giving two replicates for each arsenite concentration added to these test wells.

As it takes time to fill each row of the plate and switch between the master mixes the reads are staggered slightly, with the CgArsR-free reactions first read 90 seconds after initiation, rather than 30s for the wells containing repressor. Arsenite solutions were added in quick succession to all wells using a multichannel pipette, so for the second half of the assay all the samples are synchronised. This addition results in an RFU increase for all wells - a result of the increased volume bringing the solution surface closer to the fluorescence detector. Throughout the assay the samples display a gradual reduction in signal - this could be the result of photobleaching, micro-heating or evaporation, for example; oligo/fluorophore sequestration is unlikely due to the use of non-binding plates and the inclusion of a polyT-20 carrier oligo in all the wells.

Some rough comparisons can be made between the two master mixes using this raw data. Their pT-30 controls are very similar, while their X26 reactions show a slight disparity with the ArsR-free maximum control losing signal at a greater rate. This is likely due to a physical effect from the presence of CgArsR - their solutions may absorb/scatter/reflect/fluoresce light slightly differently, or perhaps heat up disparately. These trends are reflected in the behaviour of the reversible reactions, which also clearly show how CgArsR affects the equilibrium formed by the X19 oligos, shifting the reaction balance towards the quenched probe containing the (mutant) operator sequence, as expected. This enables these reactions to subsequently respond to arsenite addition, while neither the control wells nor the CgArsR-free reversible reactions are able to.

Figure 5.2 displays the same X19 data but normalised to either fluorophore concentration (a), or fluorophore content to give the percentage reaction balance (b). This procedure has the clear effect of removing the artefact of signal loss over time from all the samples - indicating this is happening similarly in all wells for each master mix, either to a shared background or proportional to the fluorophore activity. Both graphs show how the reversible reactions only take a few minutes to respond to the X oligos and reach stable equilibria, even when CgArsR is present. The repressor must be able to unbind and rebind rapidly for this to stabilise so quickly, but it still has a clear effect on the resulting reaction equilibrium. Figure 5.2 (a) shows how the concentration of fluorescent FGX product drops after arsenite addition, simply due to dilution. The CgArsR samples quickly respond to 100 µM arsenite, saturating within 30 minutes at the level of the ArsR-free wells - indicating the repressor effect has been completely abolished - while the response to 10 µM arsenite is slower.

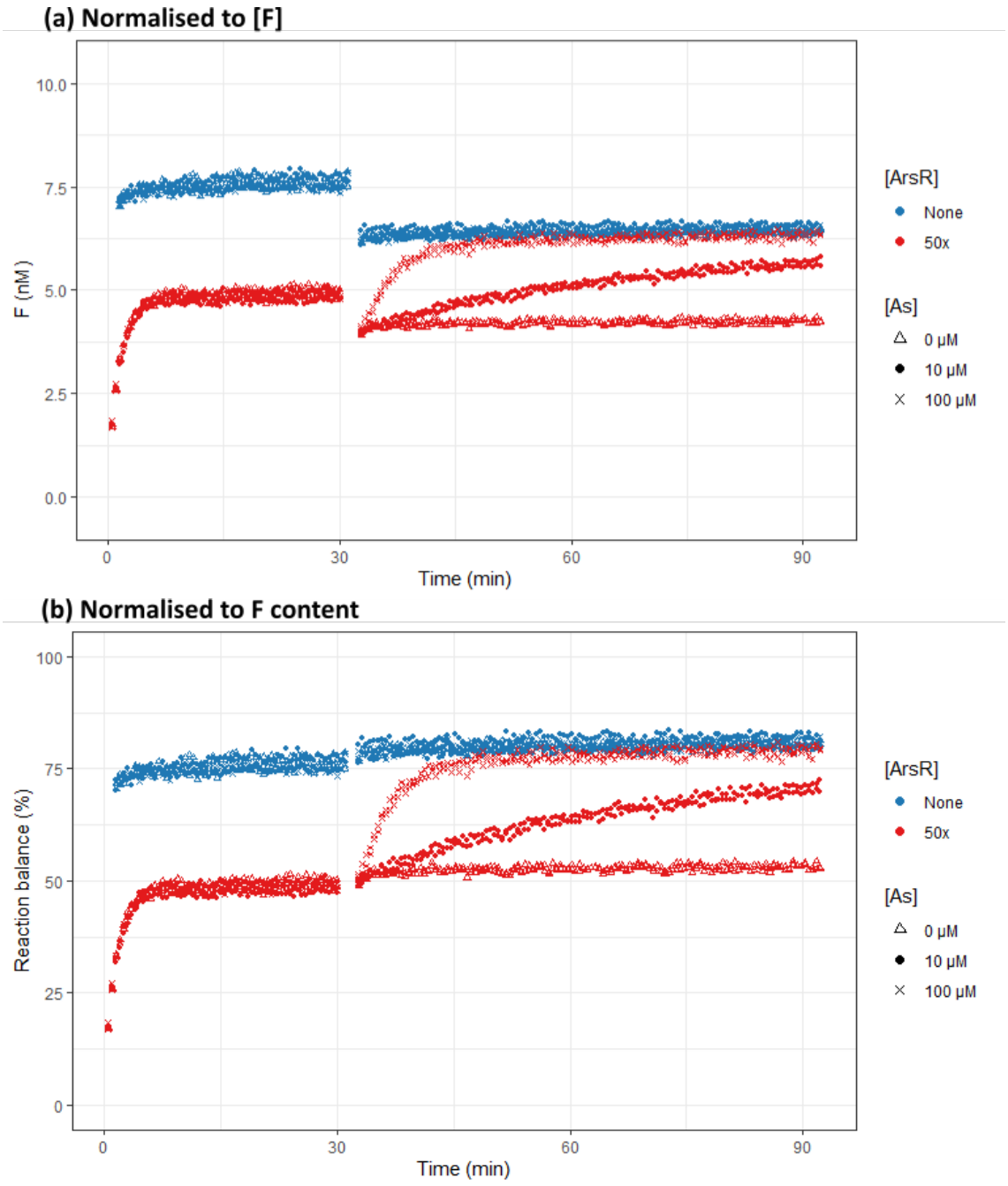


Figure 5.2: **T6T0C DSD reaction normalised data.** Fluorescence time course displaying the processed data for the assay displayed in Figure 5.1, with the reversible X19 reactions normalised to the pT-30 and X26 controls, as per the methodology presented in the previous section. (a) X19 8x data normalised to [F] according to Equation 5.4, with concentrations reduced following dilution by arsenite addition after 30 min. (b) X19 8x data normalised to F content giving the percentage reaction balance according to Equation 5.5.

Normalisation to fluorophore content rather than concentration, however, provides greater insight into the underlying dynamics by revealing the reaction balance between reactants and products irrespective of absolute concentrations.

Figure 5.2 (b) shows more clearly how the equilibria of each solution is subtly adjusted following dilution, before the slower response to arsenite kicks in for the CgArsR wells. This normalisation method, therefore, is the one primarily used to compare the differences in equilibria and arsenite response between the various master mixes and probe designs.

5.4.2 Point mutation to CR0C26 operator increases CgArsR affinity while maintaining displacement speed

The mutant CR0C26 operator within the T6T0C probe used above was designed primarily to remove secondary structure from the X displacing oligo. The EMSA in Figure 4.27, however, suggested this operator may have slightly reduced affinity for CgArsR than the native CR26 sequence, and the single mutant library screen pointed to bases C8 and C12 (in the non-reverse sequence) as being primarily responsible for this. To test if this was the case, the T6T1C probe was annealed containing a new CR1C26 operator, which included the C8G reversion to the native base at this position. Due to the subsequent operator reversal to make the displacing oligos C-rich, this meant the operator top sequence would now contain one cytosine base (hence 1C) instead of none.

While this reintroduction was expected to slightly increase CgArsR affinity, this would be at the cost of more structure. NUPACK predicts that the X6c-sCR1C26c oligo has an MFE of −0.80 kcal/mol, greater than the −0.18 kcal/mol for the X6c-sCR0C26c oligo (and 0.00 kcal/mol for shorter oligos), on account of formation of two sequential G-C bonds between the toehold and the operator. As a result, displacement speed may be reduced and dynamics could be more complicated, but it was hoped this structure would be sufficiently weak for effects to be negligible.

Figure 5.3 shows the normalised data for an assay comparing the behaviour of the two probes with and without CgArsR, and their resulting arsenite responses. The T6T0C data shown in Figure 5.2 was from this wider experiment. As can be seen, the ArsR-free T6T1C X19 reaction occurs just as rapidly as for T6T0C, reaching its equilibrium within 2 minutes, the time of the first read. This equilibrium reaction balance is almost identical across the two probes (75.7% versus 74.1%), to be expected with GC content unchanged and suggesting that minimal structure is introduced into the X oligos.

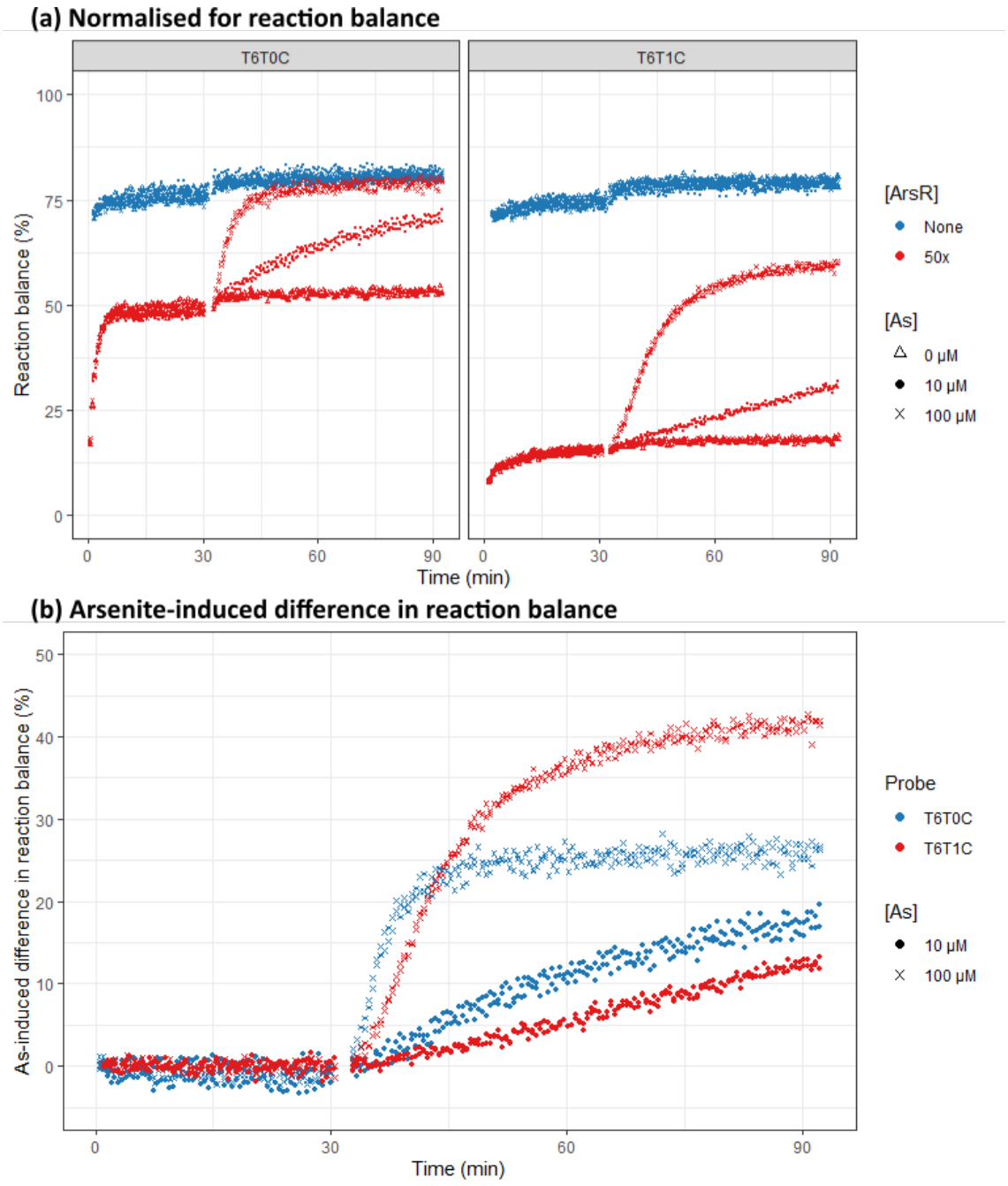


Figure 5.3: **T6T1C has a greater CgArsR affinity but may result in slower arsenite response than T6T0C.** Fluorescence time course displaying the processed data for an assay comparing T6T0C and T6T1C, according to the DSD5 protocol with the DB6 buffer, of which Figure 5.2 showed T6T0C data only. T6T1C wells with and without 50x CgArsR were first read 1 min and 2 min after addition to X19, respectively. (a) X19 8x data normalised for reaction balance. (b) Arsenite-induced difference in reaction balance, calculated by the subtraction of average normalised signal of the two arsenite-free wells per mix from each of the arsenite-added wells. Full raw data is shown in Figure A.5 (a).

More notably, the effect of the same amount of CgArsR is much greater on the new T6T1C probe. The pre-arsenite reaction balance equilibria for T6T0C with

50x CgArsR is 48.8% reacted, compared to only 15.8% displaced for T6T1C. This more than doubled the repressive effect of CgArsR from 26.9% to 58.3% of the DSD reaction, suggesting both an increase in the dynamic sensing range and greater resolution of arsenite concentrations.

Interestingly, despite this higher affinity, suggesting a greater proportion of CgArsR is actively binding T6T1C than T6T0C before arsenite addition, the response to the ion is slower. It was expected that if a greater DSD shift was achieved with the same amount of repressor then arsenic response would be faster, as more bound CgArsR would be released in any particular period. Figure 5.3 (b) compares the arsenite-induced difference in reaction balance for the CgArsR reactions for both probes. This clearly shows that while the eventual signal increase following arsenite addition is greater for T6T1C, this initial response is quicker with T6T0C for both concentrations. This suggests that CgArsR-operator affinity is tightly linked to arsenite-CgArsR affinity, such that more arsenite is required to dissociate a repressor more tightly bound to its operator. This may be a discrete change, with the stronger operator requiring two arsenite ions to bind CgArsR for dissociation rather than one.

This trade-off between dynamic range and detection speed can be altered not only by changing operator, but by manipulating CgArsR content. The weak CR0C26 operator in T6T0C required a large repressor excess for any DSD shift to be established at all, but much less CgArsR could be used with T6T1C. Lowering repressor content in a T6T1C reaction should reduce the dynamic range, but would be expected to boost detection speed as a result of increasing the ratio of arsenite to CgArsR. Sensitivity may also increase, but this would depend on the magnitude of shift change resulting from altering repressor content. This balance, while prohibitive to optimising all aspects of arsenite detection with a single formulation, does enable fine tuning to particular speeds or sensitivities.

5.4.3 Linear probe design removes Tprobe structure while allowing G strand flexibility

The Tprobe design was introduced in the previous chapter primarily to enable fast and cheap sequence alterations. It did this successfully - several operator sequences were tested in combination with different toeholds and polyT spacers.

It did have a couple of minor drawbacks, however. Using large strand excesses at each pairing away from the F strand does introduce some DSD inefficiencies, while the large quantity of quencher-modified Q strand required per assay is expensive in its own right. Also, it still seemed possible the physical structure of the Tprobe, with its universal arms forced outwards from the direction of the operator, may sterically inhibit CgArsR binding, even after the inclusion of the 6T spacer that should allow more flexibility.

An intermediate design was therefore proposed, shown in Figure 5.4. This used the same universal F strand and operator-specific G strand, including F adaptor, 6T-spacer and toehold domains, but replaced the QR duplex with a single quencher-modified operator oligo. Such a design should be more linear than the Tprobe, with some flexibility around the ssDNA polyT spacer, hence the 'Lprobe' name. DSD reactions would take place just like with the Tprobe, but releasing an ssDNA Q oligo instead of the QR complex as before. The clear drawback of this design is the use of a modified oligo containing one strand of the CgArsR operator - testing different operators would require a new modified strand to be designed and purchased each time. However, following the experiments in the previous chapter only a few different operator sequences were desired for the later quantitative comparisons here, so the cost of the handful of quencher oligos was deemed a worthwhile trade-off for the advantages presented above.

The use of a three-strand probe, rather than a conventional two-strand FQ reporter, still had a few benefits. First, this maintained the same DSD products compared at the end of the previous chapter that, with the inclusion of the 6T spacer, were shown to have similar inherent fluorescence levels due to their controlled fluorophore environments. Second, for a particular operator sequence the toehold domain could be altered, perhaps to test different reaction strengths, through the design of new unmodified G strands only - not requiring any new modified oligos. Finally, comparison between the different operator sequences required only new quencher-modified strands and not both quencher- and fluorophore-modified oligos, as would be the case for a two-stranded probe. The Lprobe, therefore, seemed a good compromise between cost, structure and functionality.

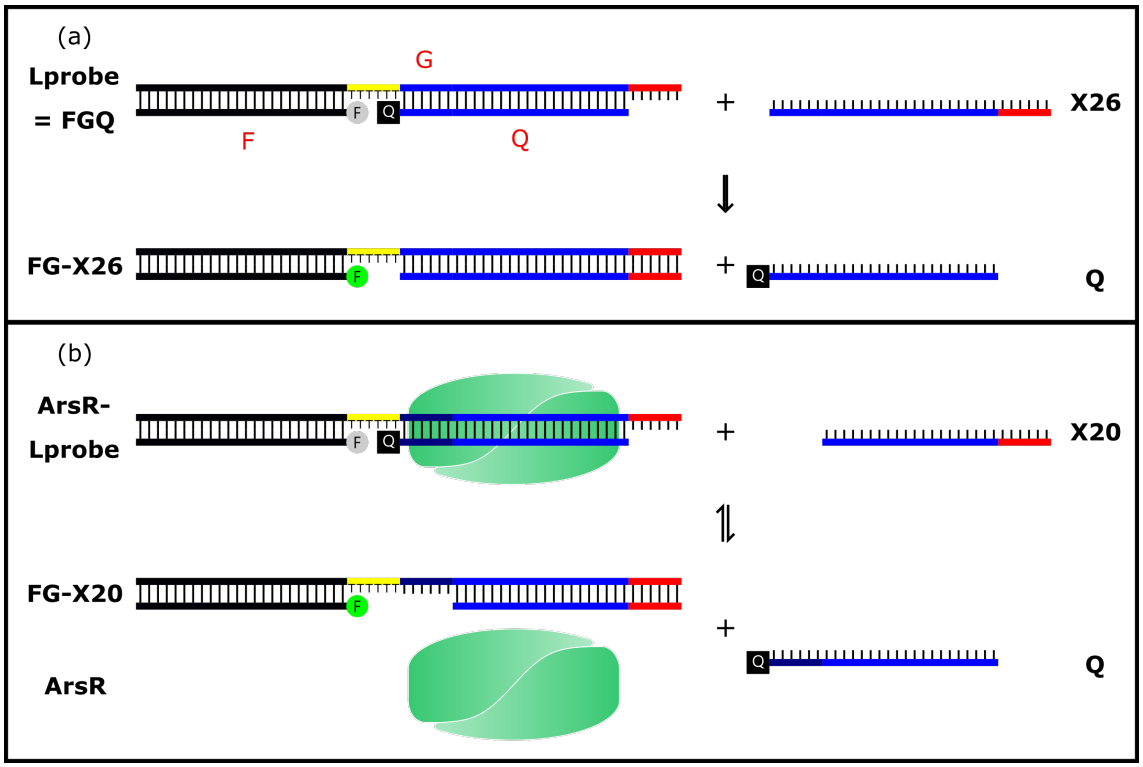


Figure 5.4: **Lprobe DSD reactions.** Schematic illustrating the structure of an Lprobe and how it undergoes DSD reactions. (a) The Lprobe is comprised of F, G and Q DNA strands (labelled with red letters), the first and last of which are modified by a fluorophore and quencher, respectively. When reacted with displacing oligo X (here, fully complementary X26), the products FGX and Q are formed. FGX is fluorescent as its fluorophore is no longer in the proximity of a quencher modification. (b) The Lprobe reacts similarly to the Tprobe in the context of a reversible, repressor-operator system. ArsR binds its operator sequence within the Lprobe (blue dsDNA), reducing the likelihood of DSD occurring with the shorter X20 oligo. As the ArsR will stochastically dissociate, some DSD will occur, forming the FG-X20 product without a full dsDNA operator. A dynamic equilibrium is therefore set up between the states, to be altered through addition of arsenite.

5.4.4 Lprobe further enables greater CgArsR affinity as well as faster arsenite responses

Figure 5.5 (a) displays the normalised data for an assay comparing T6T1C and L6T1C reactions, following incubation with or without 10x CgArsR. The five-fold reduction in repressor content compared to Figure 5.3 lowered the CgArsR-induced equilibrium shift to 21.9%, just over a third of the previous difference in reaction balance. The parallel reactions using the linear L6T1C resulted in a shift of 34.0%, more than 50% greater that the Tprobe. The ability of CgArsR to exert a greater effect on the Lprobe equilibrium despite the same operator sequence and displacing oligo suggests that the Tprobe structure did indeed hinder the

repressor slightly, and the change to the linear design enabled greater activity.

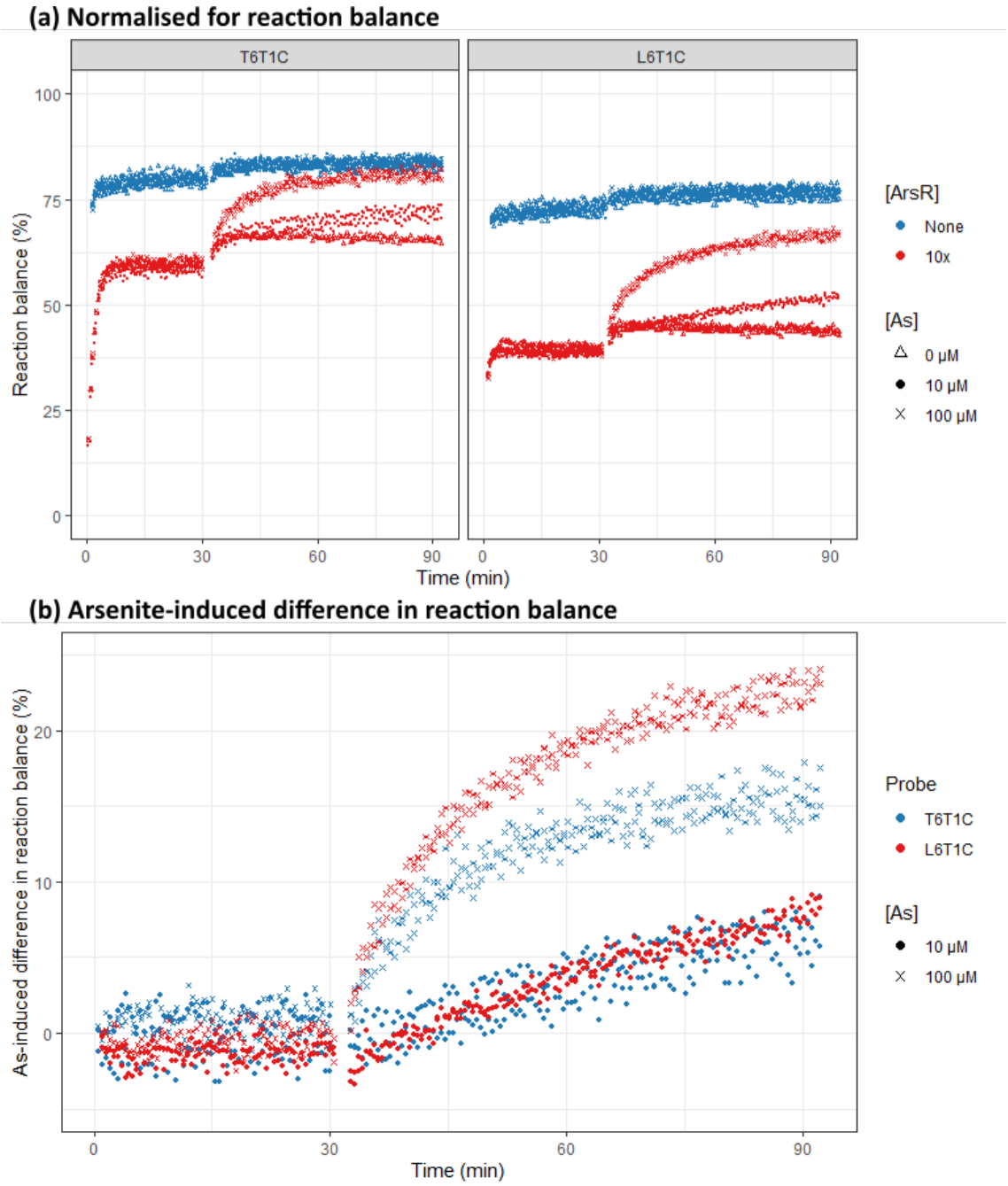


Figure 5.5: **L6T1C enables a greater CgArsR affinity and faster arsenite response than T6T1C.** Fluorescence time course displaying the processed data for an assay comparing T6T1C and L6T1C with and without 10x CgArsR, according to the DSD5 protocol with the DB6 buffer. Graph shows readings from individual wells as points and averages across control triplicates as lines. After 30 min, 50 µl 0/10/100 µM sodium arsenite was added to all triplicate sets. There were two sets of X19 reactions, both shown, giving two replicates for each arsenite concentration added to these test wells. (a) X19 8x data normalised for reaction balance. (b) Arsenite-induced difference in reaction balance, calculated by the subtraction of average normalised signal of the two arsenite-free wells per mix from each of the arsenite-added wells. Full raw data is shown in Figure A.5 (b).

In this case, however, the greater resulting affinity did not cause a loss of arsenite response speed. Figure 5.5 (b) displays the arsenite-induced difference in reaction balance for their CgArsR mixes, showing similar response to 10 µM arsenite but a faster increase for 100 µM using the linear probe. As the response to the lower arsenite concentration is so similar, the difference seen with the larger amount may not be due to a higher rate of CgArsR dissociation, but with faster basal DSD speed. The L6T1C-CgArsR X19 reaction reaches equilibrium within only a few minutes, compared to nearly 10 for the two Tprobes, due not to increased operator affinity but the DSD reaction itself. This difference is not clearly seen in the ArsR-free samples here due to their initial read delay, but can be seen later in Section 5.5, where the Lprobes consistently react in less than a minute. The Tprobe structure not only reduces CgArsR affinity, therefore, but affects DNA displacement.

Another subtle effect on displacement is the difference in CgArsR-free equilibrium. This is slightly reduced for L6T1C and this shift in favour of the reactants suggests the Lprobe is more stable and the Tprobe more prone to displacement, despite flexibility from 6T spacer. In addition, the raw data in Figure A.5 (b) showed a notable difference in the RFU of polyT-30 reactions with L6T1C compared to both Tprobes. After 30 minutes RFU was on average 28.6k, down from 40.7k across the Tprobes, a reduction of nearly 30%. As the maximum fluorescence of the L6T1C reactions was only marginally lower, this did not represent anomalously less fluorophore added to each mix, but a sharp drop in background signal. The most likely explanation seemed improved quenching efficiency from the Lprobe compared to the Tprobe structure, an unforeseen benefit of this change. This suggested that the orientation of the modifications in the Tprobe did not allow optimal quenching, but the simpler Lprobe structure improved their proximity.

Together these presented clear advantages to continued use of the linear probe design. Increased CgArsR affinity, faster DSD reactions and greater signal-to-noise were all beneficial properties, in addition to requiring a smaller quantity of quencher-modified oligo. As a result the remaining assays made use of the L6T1C probe.

5.5 Improving buffer composition

The normalised assays used in the previous section to compare probe designs could also be used to investigate the effects of altering buffer composition in a controlled manner. So far, most DSD reactions had been carried out in the same simple buffer that oligos were resuspended, annealed and stored in (buffer AB - 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$). The only deviations from this were the small volume of protein storage buffer added as a result of repressor addition (or a repressor-free control) and the recently inclusion of 1 µM polyT-20 as a carrier ssDNA oligo to reduce probe sequestration. Recent assays were therefore performed in solutions primarily composed of buffer DB1 - 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 1 µM polyT-20.

While this buffer clearly enabled DSD reactions to occur as well as allowed CgArsR to fold correctly, bind its operator, affect DSD equilibria and respond to arsenite, it seemed possible that increased CgArsR-operator affinity could be achieved through altering this buffer, enabling greater arsenite sensitivity or faster detection speeds. A balance would have to be struck, however, between aiding CgArsR activity and maintaining expected DSD performance.

5.5.1 Reducing NaCl concentration increases CgArsR affinity and arsenite sensitivity

Sodium chloride concentration was tested first, as it seemed the component most likely to influence CgArsR affinity. Some salt would be necessary to maintain protein structure but an excessively high concentration could inhibit the electrostatic interactions between repressor an operator - this does, however, leave a large concentration range within which to explore. Four NaCl concentrations were therefore tested: 50 mM, the standard 100 mM, 200 mM and 500 mM. To do this, L6T1C probe and CgArsR were incubated in buffer DB2, containing 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 20 mM $MgCl_2$, 2 µM polyT-20 - double every component except for NaCl.

Assays were then carried out with 3 µl pT-30/X19/X26 oligos loaded into wells followed by 100 µl of 50/150/350/950 mM NaCl. Reactions were then initiated by addition of 100 µl of the 2x concentrated probe mixture to each, so that the final NaCl concentrations of each mimicked those of 1x solutions containing the

desired concentrations. Each condition - in this case NaCl concentration - had two replicates of reactions with all three oligos, responding to both probe-repressor mixes and probe-only mixes. After 30 min, 50 µl of sodium arsenite or water were added to the replicate pairs for each of these, so that their responses to arsenic could be observed.

The inclusion of all these controls was very important. It enabled removal of the effect buffer changes may have had directly on fluorescence, while controlling the usual sources of variability in master mix preparation. Crucially, it also allowed separation of buffer effects on CgArsR behaviour from buffer effects on the underlying DSD reaction - the probe-repressor mix reactions would show the overall result of combining these effects.

The results of this assay is shown in Figure 5.6. In part (a), the normalised data for the repressor-free (left) and repressor-present (right) reactions are shown side by side. The left panel shows that altering NaCl concentration has a minimal effect on underlying DSD equilibrium. The 100 and 200 mM reactions are very similar, while the more extreme concentrations (especially 50 mM) may shift the equilibrium slightly towards the quenched probe. In the right panel, the green points show the data for the standard 100 mM buffer, with the difference between the pre-arsenite equilibria of the matched conditions between the two plots illustrating the effective dynamic range of each system.

Increasing NaCl concentration, therefore, seemed to reduce the dynamic range of the system presumably by reducing CgArsR-operator affinity. In these high salt conditions the repressor was unable to influence the DSD equilibrium much and so the following arsenite response is very limited. The lower 50 mM NaCl concentration, however, seemed to allow a much greater shift to be caused by CgArsR, presumably by increasing operator affinity.

In part (b) a further data processing step has been performed. This takes each pair of replicates (before they are differentially treated by water/arsenite addition) and subtracts the normalised reaction balance of the sample to which water is added from the sample to which arsenite is added. The result is, for each specific condition, the shift in reaction balance caused by addition of arsenite. This should be non-existent for the repressor-free samples, as arsenite does not affect the basal DNA, but for the reactions where CgArsR is present this shows the ultimate effect of buffer change on assay performance - whether speed or strength of response is altered.

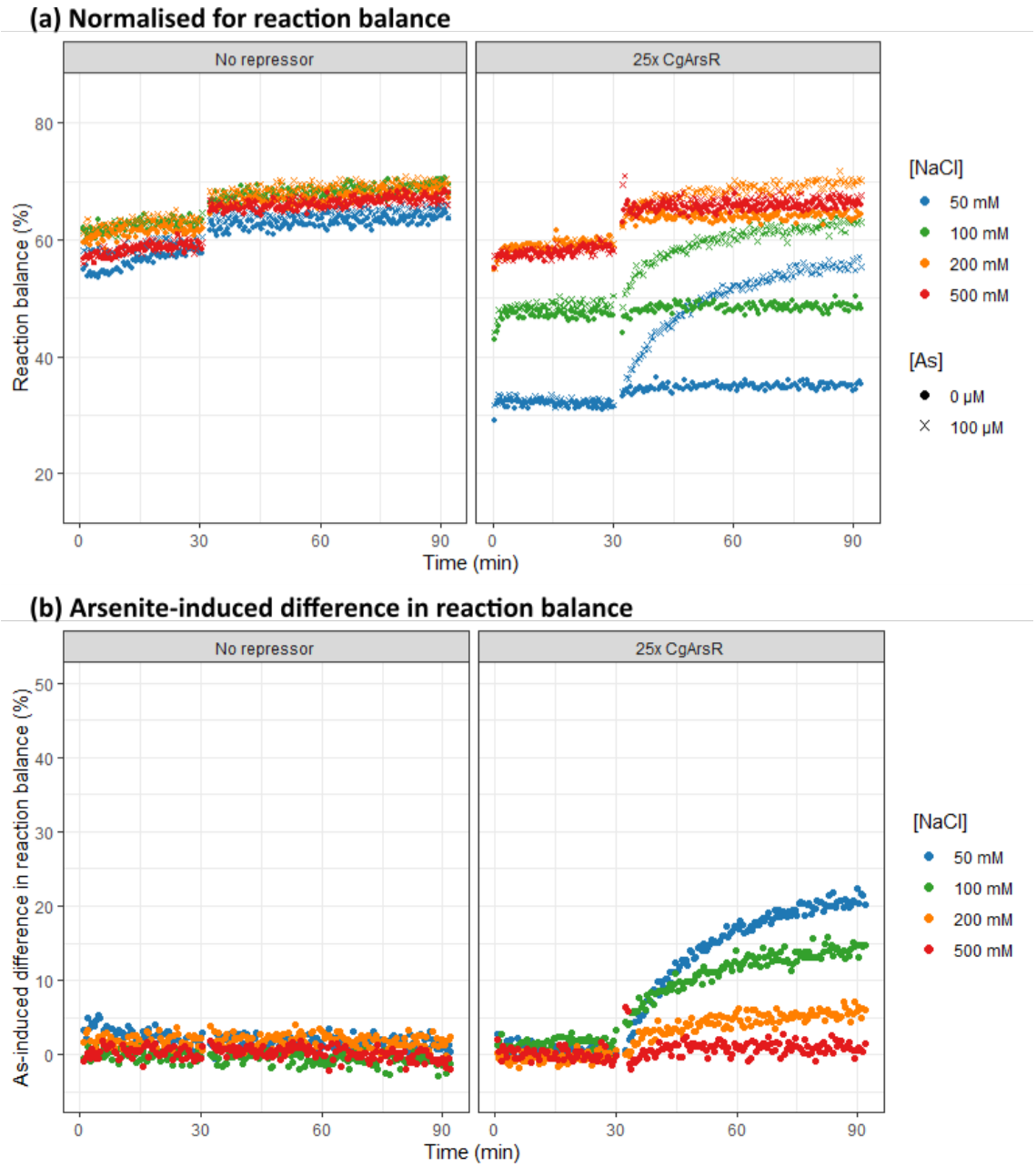


Figure 5.6: **Optimising NaCl concentration.** Fluorescence time course displaying the normalised data for L6T1C, incubated with and without CgArsR in buffer DB2, undergoing reversible DSD then responding to arsenite in solutions of different NaCl concentrations. Initial DSD was run for 30 min before arsenite was added and the assay continued for another hour, according to protocol DSD3. (a) X19 8x data normalised to F content giving the percentage reaction balance. (b) Arsenite-induced difference in reaction balance, obtained by calculating the difference between the paired arsenite-free and arsenite-added normalised data from (a), to compare the effect of arsenite addition for each particular solution.

For sodium chloride, this reveals a clear correlation between increasing concentration and reducing signal strength - thereby decreasing potential sensitivity. 200 mM NaCl enables roughly half of the signal increase as the

standard 100 mM, while response is effectively abolished in the 500 mM NaCl solution. In the 50 mM buffer, response speed is initially similar but by the end of the hour the arsenite-induced shift is roughly 20% compared to 15% previously, an impressive 33% increase.

Due to the multiple normalisation steps and controls required to produce this final data, many wells are needed for each eventual trace. As a result the data in Figure 5.6 (b) has no replicates and so caution must be used when comparing the precise numbers. However, the strength of the controls and the robust presentation of the correlation does suggest that reducing sodium chloride concentration below 100 mM would improve sensitivity. Later experiments, which accumulate particular changes, also serve as independent replicates for one condition per assay, suggesting the variation is at most 5% between assays and likely much less within each experiment.

5.5.2 Altering $MgCl_2$ concentration shifts the underlying DSD equilibrium

The same experimental setup was then used to test the effect of altering magnesium chloride concentration. Magnesium is important for stabilising the double helix structure of dsDNA, and so is a crucial component of DSD reaction buffers. Once again, however, there is a broad acceptable concentration range, across which it is possible an optimum could be found for CgArsR. For this assay a new DB3 buffer was used (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 2 µM polyT-20) so that when used in the probe master mixes and diluted 0.5x by 5/15/25/35 mM $MgCl_2$, the final concentrations would be 5/10/15/20 mM, in addition to the improved NaCl concentration of 50 mM.

The results are shown in Figure 5.7, where once again the standard concentration (10 mM) is shown in green. The right panel in part (a) reveals that, unlike with NaCl, reducing the standard concentration of $MgCl_2$ shifts the equilibrium with CgArsR up and increasing the concentration shifts the reaction balance down. In isolation this would suggest that more $MgCl_2$ is preferable and perhaps improves CgArsR-operator affinity, however the repressor-free control reactions enables the full picture to be seen.

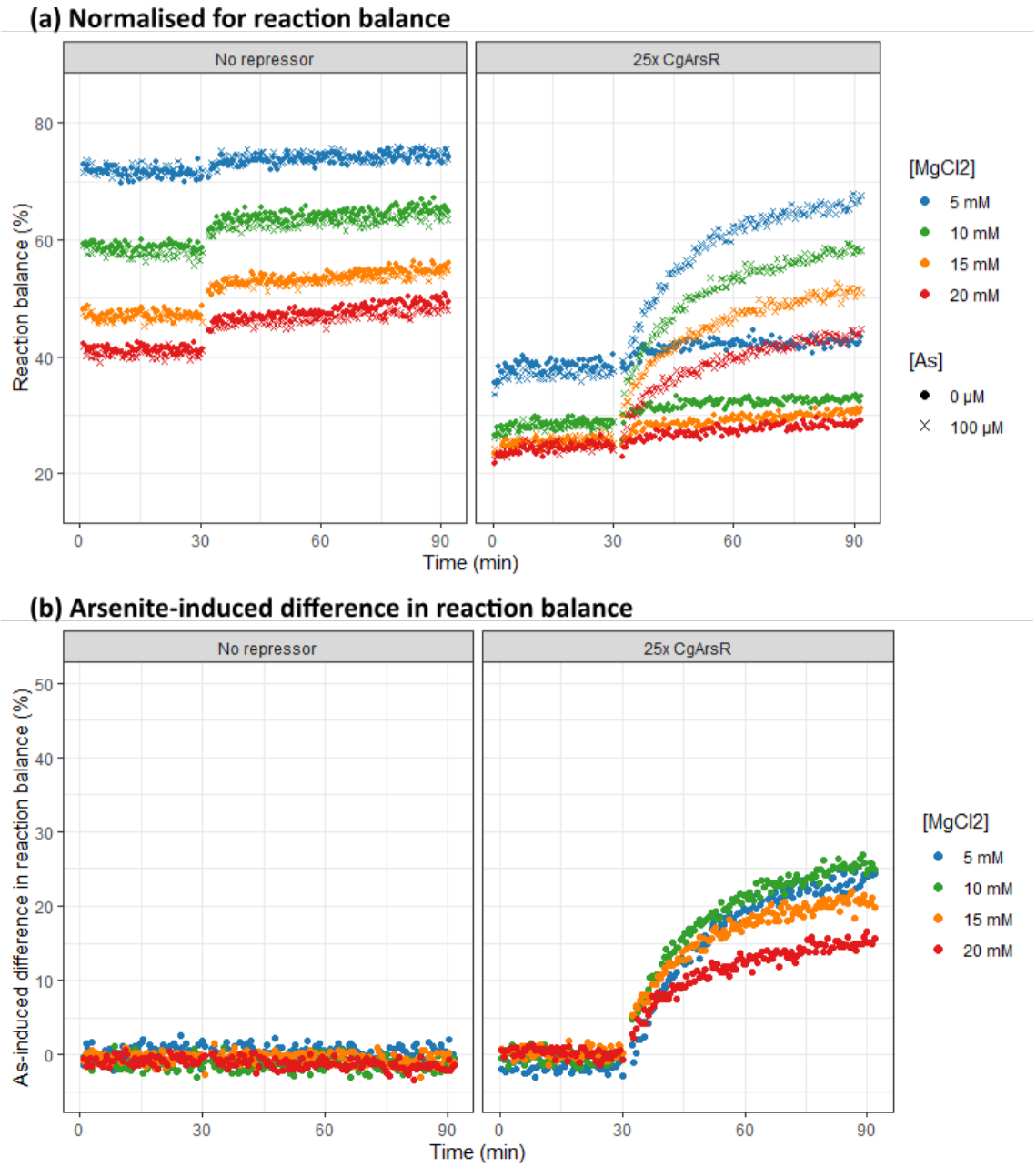


Figure 5.7: **Optimising $MgCl_2$ concentration.** Fluorescence time course displaying the normalised data for L6T1C, incubated with and without CgArsR in buffer DB3, undergoing reversible DSD then responding to arsenite in solutions of different $MgCl_2$ concentrations. Initial DSD was run for 30 min before arsenite was added and the assay continued for another hour, according to protocol DSD3. (a) X19 8x data normalised to F content giving the percentage reaction balance. (b) Arsenite-induced difference in reaction balance, obtained by calculating the difference between the paired arsenite-free and arsenite-added normalised data from (a), to compare the effect of arsenite addition for each particular solution.

The left panel shows how altering the $MgCl_2$ concentration has a profound effect on the underlying DSD equilibrium, before any CgArsR is present. Reducing magnesium content favours the products, while increasing magnesium favours

the reactants. Such a strong effect was very unexpected, with only a small difference in structure between the quenched probe and fluorescent product duplexes. With respect to base-pairing, a single reaction between a L6T1C probe and an X19 oligo would release a quencher oligo one nucleotide longer than X19 and result in the net loss of one base-pair in the FG duplex species. It would be remarkable if this minimal shift in the prevalence of dsDNA/ssDNA solely caused the effect on the equilibrium, but regardless of exact mechanism this effect was important to be aware of, and could be a useful tool to exploit.

Figure 5.7 (b) reveals that, despite the sizeable shifts in equilibria, changing magnesium content within this range had a relatively small effect on CgArsR functionality and arsenite sensitivity. The highest concentration of $MgCl_2$ tested (20 mM) displayed a reduced arsenite response, but this may be more due to its effect on the position of the original equilibrium (40% reaction), than a specific inhibition of CgArsR-operator affinity. The nature of mass action kinetics means that shifting an equilibrium 10%, from 55% to 45%, is much 'easier' than from 10% to 0%, to give an extreme example. It becomes progressively harder to force a dynamic equilibrium to either limit through the addition of a single reactant as the increasing concentration of the products will resist the change with greater strength, so comparisons between percentage shifts in equilibria, when the starting points are different, are complicated. Ideally the fundamental rate constants being altered by these interventions would be calculated and compared, but this would require an accurate model to be built.

In this scenario, however, the three lowest magnesium concentrations enabled similar arsenite responses, albeit the increase using the standard 10 mM was marginally greater than 5 mM. In spite of this, the lower concentration was carried forward. This was due to the overall equilibrium shift caused by the low magnesium buffer - it seemed likely that further CgArsR affinity 'gains' would be easier to exploit from a more moderate starting point (40% rather than 30% reaction balance with CgArsR), as highlighted above. More refined $MgCl_2$ adjustments would certainly be possible if manipulating the reaction equilibrium further was desired.

5.5.3 Altering pH affects both operator affinity and arsenite response speed

The final aspect of the original buffer to test was pH. Maintaining appropriate acidity/alkalinity is widely known to be important for protein functionality, however for *in vitro* assays focus tends to be on finding a pH acceptable for protein stability rather than optimising performance. Most of these systems have many proteins, so it can be difficult to find a pH optimal for all components, so solutions are kept roughly neutral, avoiding any isoelectric points that may cause a protein to precipitate.

DNA is often stored in pH8.0 buffers, and DSD reactions are commonly performed in these too. Purified CgArsR was stored at pH7.4, standard for many protein purification protocols, but while these enabled working assays it was possible that CgArsR functioned best in different conditions. A dummy solution was prepared containing all components but without any tris in the main buffer solution (a small quantity would be present from DNA and protein storage buffers) and pH was measured at 7.6, so it seemed feasible to prepare reactions master mixes like this and add them to other buffers of different pHs.

DB5 buffer was therefore used for this assay (100 mM NaCl, 10 mM $MgCl_2$, 2 µM polyT-20) so that when the probe master mixes were diluted 0.5x by 20 mM Tris-HCl pH 6.8/7.4/8.0/8.6, the final buffer concentration would be 10 mM, in addition to the halved concentrations of the DB5 constituents.

Figure 5.8 shows the result of this assay, this time with the standard pH8.0 condition in orange. The left panel in plot (a) shows that altering the pH has a minimal effect on the underlying DSD equilibrium. There appears to be a slight correlation with the more acidic solutions favouring the products, but if this effect is real its magnitude is very small. The right panel, however, shows a very interesting story. The samples at pH8.6 seem to reach the same equilibrium point as those for pH8.0, while the more acidic solutions settle at a shared, more quenched state. This suggests that CgArsR affinity for its operator is higher at these lower pHs, but this effect seems to be a discrete change rather than continuously varying over this range. It may be the case that, upon reducing the pH from 8.0 to 7.4, a single residue (or a small number of residues with similar pKa) becomes protonated and this change improves the repressor-operator affinity, while shifting the pH just outside this range does not.

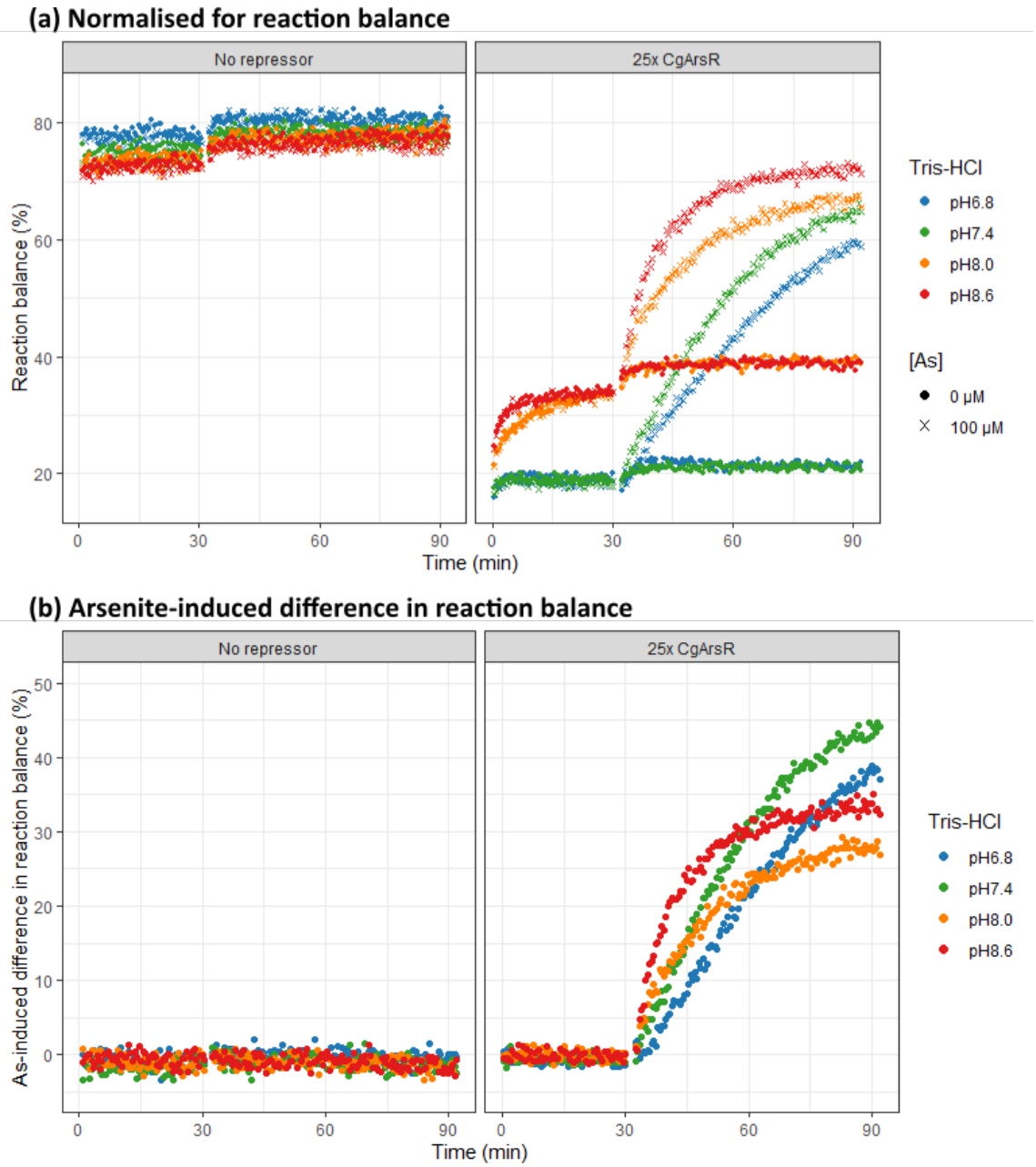


Figure 5.8: **Optimising pH.** Fluorescence time course displaying the normalised data for L6T1C, incubated with and without CgArsR in buffer DB5, undergoing reversible DSD then responding to arsenite in solutions of different pH, as buffered by 10 mM Tris-HCl. Initial DSD was run for 30 min before arsenite was added and the assay continued for another hour, according to protocol DSD3. (a) X19 8x data normalised to F content giving the percentage reaction balance. (b) Arsenite-induced difference in reaction balance, obtained by calculating the difference between the paired arsenite-free and arsenite-added normalised data from (a), to compare the effect of arsenite addition for each particular solution.

This result in isolation would suggest that use of the lower pHs was preferential - the larger dynamic range generated by the higher operator affinity should increase arsenite sensitivity. However, the signals following arsenite addition

complicated this picture. Within each pair of pHs settling at the distinct equilibria, the more alkaline of each responded to arsenite faster. Figure 5.8 (b) enables easier comparison between all four samples, and suggests the initial speed of signal increase correlates with higher pHs. This effect may be clearer if the samples were further normalised for the same dynamic range.

This second effect suggests that deprotonation is beneficial for speed of response to arsenite. This could be due either to deprotonation of the thiol groups within the arsenite-binding domains of CgArsR, or deprotonation of the arsenite ion itself - both could feasibly create more reactive species. Unfortunately, this creates a dilemma for pH optimisation as two system properties - speed and sensitivity - seem to be most improved by different conditions. Interestingly, Figure 5.8 (b) does suggest pH8.0 is the worst within this range, neither optimising speed nor sensitivity, so choosing either pH8.6 or pH7.4 would substantially improve the performance of the assay from the current setup through one metric.

The accumulation of improvements to NaCl, $MgCl_2$ and pH meant that the maximal arsenite-induced shift in reaction balance after one hour was tripled, from 15% to 45% (Figure 5.6 (b) vs Figure 5.8 (b)). Using pH8.6 instead, a greater than two-fold increase in signal would be maintained, alongside higher speed. Such results, and the inferences about effects on CgArsR behaviour, would have been much less clear without both the removal of secondary structure and the normalisation methodology presented earlier.

5.5.4 Molecular crowding agents have complex effects on CgArsR affinity and arsenite response speed

One last buffer component to test was the inclusion of crowding agents. Up to this point, assays have been performed in water-based solutions with only low solute concentrations to maintain system functionality. A cell, however, is an extremely busy environment, with hundreds or thousands of different proteins simultaneously present and functioning together to ensure viability, alongside numerous other macromolecules and smaller solutes [154]. While the *in vitro* system presented here is enriched for CgArsR and its operator above the natural concentrations expected in a *C. glutamicum* cell, the solutions are overall very sparse, containing much more water than normal for a living organism.

This high-water low-macromolecule environment is common for *in vitro* systems, and has been seen as a reason some underperform relative to genetically-engineered, living cells. Crowding agents - long, polar but unreactive macromolecules - have been suggested as solutions to this problem [155]. These highly soluble molecules effectively replace significant quantities of water and as most are linear or branched chains, they increase solution viscosity to more closely replicate cell cytoplasms. As a result, they are said to create solution pockets, increasing the local concentration of solutes within, enhancing the formation of complexes and improving reaction speeds [156, 157].

A number of different crowding agents exist and many were tested with this system, including BSA, dextran sulfate and Ficoll. These each displayed unique effects, but most experimentation was performed with PEG (polyethylene glycol), the crowding agent most commonly used in similar applications [158]. PEG forms linear chains of different lengths, several of which were compared but PEG8k (molecular mass of approximately 8000 g/mol) was primarily used.

Due to the complexity of the effects of macromolecular crowding, a slightly different assay format was used to those for the other buffer conditions. Here, each PEG8k concentration was tested on multiple CgArsR excesses as well as two arsenite concentrations (in addition to arsenite-free water). To allow for the number of wells on a plate these extra samples required, different PEG8k concentrations were investigated in separate experiments, although each were internally normalised as before to ensure controlled comparisons could be made.

Figure 5.9 shows the results of these experiments. In plot (a) the full normalised data are shown, coloured according to the CgArsR content in each sample. The blue data show that increasing the concentration of PEG8k seems to slightly favour the quenched probe in the CgArsR-free reactions, as the equilibrium reaction balance drops as PEG8k increases. As for the magnesium optimisation, this may be due to the single base-pair difference between the linear probe and the FG-X19 product.

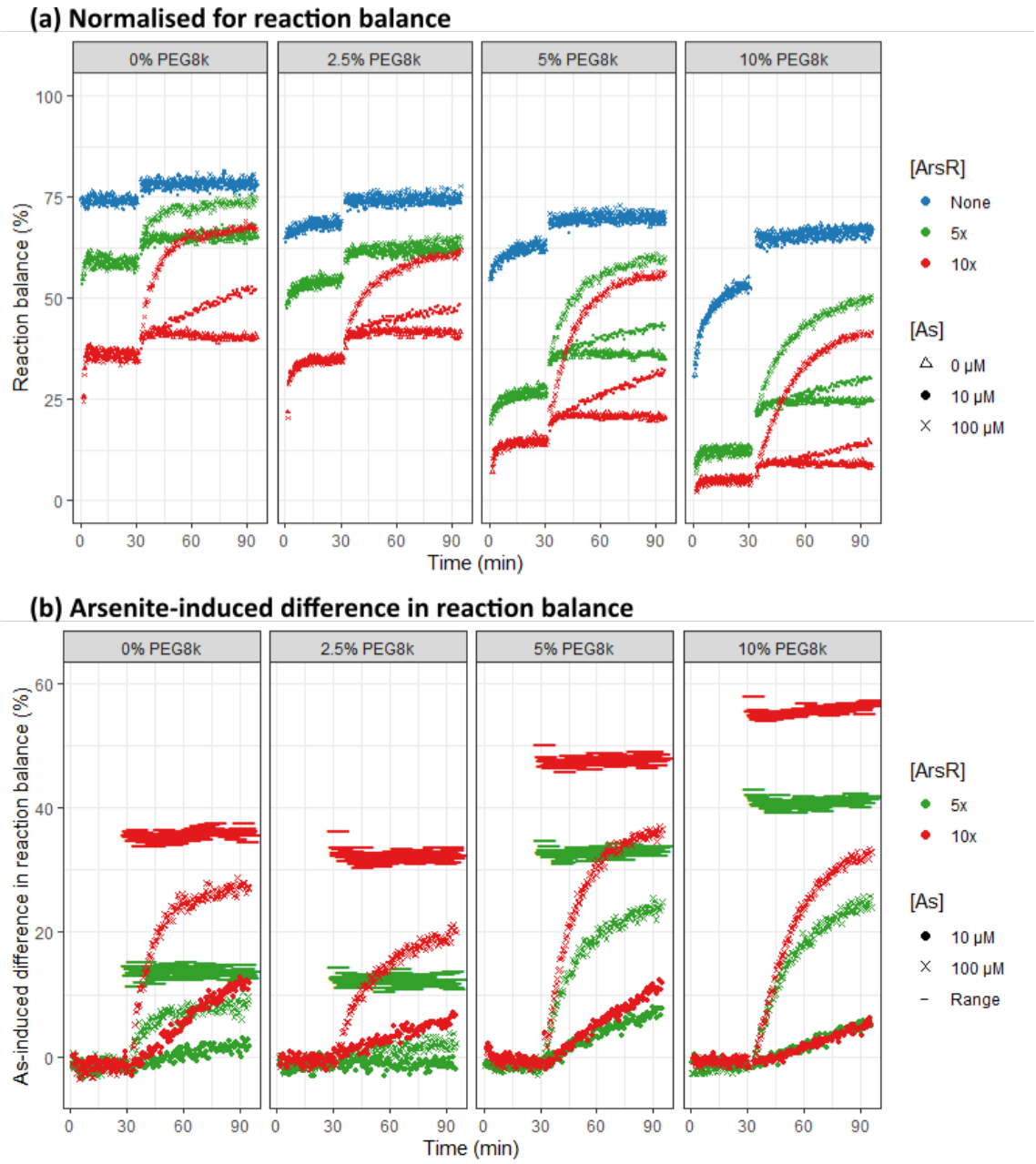


Figure 5.9: **Investigating effects of PEG8k addition.** Fluorescence time course displaying the normalised data for L6T1C, incubated with and without CgArsR in buffers DB6/7, undergoing reversible DSD then responding to arsenite in solutions containing various concentrations of the PEG8k crowding agent. Initial DSD was run for 30 min before arsenite was added and the assay continued for another hour, according to protocol DSD5. (a) X19 8x data normalised to F content giving the percentage reaction balance. (b) Arsenite-induced difference in reaction balance, obtained by calculating the difference between the paired arsenite-free and arsenite-added normalised data from (a), to compare the effect of arsenite addition for each particular solution. CgArsR-free data is not shown for clarity, but these all clustered around the 0% change level as expected. Horizontal lines indicate the theoretical maximum signal increase if the added arsenite induced dissociation of all the CgArsR in each sample, calculated through subtraction of the arsenite-free reaction balances in (a) from the average reaction balance of the ArsR-free samples.

Another trend here is slower initial reaction speeds with the higher concentration PEG - while the ArsR-free X19 DSD reactions equilibrated within a minute without PEG, the 10% PEG8k solution required the whole 30 minutes to stabilise and the intermediate concentrations displayed intermediate reaction speeds. In fact the 10% PEG8k solution slowed the X26 reaction so much that the controls for this experiment had to be triggered 30 minutes before the samples to ensure accurate normalisation (see subsection 5.3.4).

The data in red show the samples containing a 10x excess of CgArsR over the L6T1C probe. The pre-As equilibria follow the downwards trend of the ArsR-free samples, but this effect seems to be accentuated at the higher PEG concentrations, suggesting that the crowding agent is favouring the formation of the CgArsR-probe complex as hoped, effectively increasing the affinity of CgArsR for its operator. This increased range should translate into greater arsenite sensitivities.

The green data in Figure 5.9 represent samples containing only a 5x CgArsR excess. Without PEG these samples achieve less than half the reaction shift of the 10x CgArsR reaction, but as the PEG concentration increases the effect of the reduced repressor samples strengthens disproportionately. This reflects the fact that the combination of high PEG and 10x CgArsR is capable of almost fully quenching the X19 DSD reaction - shifting the reaction balance to almost 0%. CgArsR affinity has been artificially enhanced so much that most of this effect can be captured by half the repressor content. This is highly advantageous as for a particular arsenite concentration, roughly maintaining the DSD shift with half the repressor should close to double potential sensitivity.

Plot (b) displays the arsenite-induced reaction balance shift for all the samples with CgArsR, calculated through subtraction of the arsenite-added samples from their arsenite-free counterparts. The blue repressor-free samples are not displayed to improve clarity, but arsenite did not differentially affect these wells as expected (no separation in plot (a)) so data simply clustered around the 0% change level. Horizontal lines have been added to plot (b) - these indicate the theoretical maximum signal increase if the added arsenite induced dissociation of all the CgArsR in each sample (for each PEG and ArsR concentration), calculated through subtraction of the arsenite-free reaction balances in (a) from the average reaction balance of the ArsR-free samples for each respective PEG concentration.

This second analysis clarifies the actual signal increases caused by the two concentrations of arsenite, in the context of their maximum possible rise. For the same ArsR and arsenite concentrations, the same relative signal increase would be expected across the PEG concentrations if the crowding agent only affected repressor-operator affinity and nothing else. Comparing the maximum ranges, there is a trend towards boosting this potential increase with higher PEG concentrations, with the stark exception of the 2.5% PEG8k samples. This formulation seemed to hinder almost every assay performance metric compared to the PEG-free samples. The theoretical ranges were net slightly reduced and the actual arsenite-induced increases over one hour were much lower too. This was especially the case for the 5x CgArsR samples, which displayed almost no increase for either arsenite concentration - much worse than the PEG-free scenario. An independent repeat experiment was performed to check the validity of this result but almost identical data were obtained, suggesting this was not an anomalous effect but an accurate reflection of this particular formulation. This was at odds with expectations and at very least the lack of a consistent trend across the PEG concentrations suggested the crowding effects were complex and multi-faceted, with compensations between beneficial and disadvantageous consequences.

Excluding the 2.5% PEG8k experiment from further comparisons, the other PEG concentrations did seem to boost signal range relative to the PEG-free buffer, particularly for the lower 5x CgArsR excess. The actual arsenite response, however, did not consistently increase in proportion to these ranges. For the 10x CgArsR excesses, in the PEG-free solution the high arsenite sample caused a signal increase corresponding to 76% of the available range after one hour. This figure was the same for the 5% PEG buffer, while this dropped to 58% for 10% PEG. For the lower arsenite samples the increases were 33%, 24% and 9%, respectively, reflecting a similar real-term increase between no and 5% PEG, but a halving of actual signal increase in the 10% buffer. This suggests that while CgArsR-operator affinity may be improved by PEG addition, at high concentrations this may be outweighed by the effect of reducing reaction speeds (perhaps through lowering overall diffusion rates), slowing mixing or even diminished CgArsR-arsenite affinity.

The data using the 5x CgArsR content was perhaps even more surprising. Under no PEG/arsenite combination did the reduced repressor content outperform its 10x counterpart in actual signal increases over any time frame - at best the low

arsenite, 10% PEG condition resulted in matched signal increases across the two CgArsR concentrations. For the high arsenite samples the relative increases of 64%, 73% and 57% were very close to the percentages for the 10x CgArsR reactions. This perhaps suggests that arsenite-induced CgArsR dissociation was not the rate-limiting step in the signal increase, but maybe the subsequent DSD rebalancing. For the low arsenite samples the relative increases were 13%, 23% and 12% - an improvement only for the 10% PEG buffer. All these data suggested that improving repressor-operator affinity to reduce CgArsR content and thereby increase arsenite-repressor ratio, did not provide the expected performance improvements. Neither altering crowding agent concentration nor repressor excess seemed to follow straightforward patterns, so trends could not be extrapolated through untested conditions and optimisations could not be attempted in isolation of other variables.

5.5.5 Combining improvements generates more sensitive assay

While anticipating the effect of different PEG concentrations proved difficult, empirical testing could identify successful combinations of assay variables. It was noted above that the 10% PEG8k condition resulted in almost complete quenching of the 10x CgArsR X19 DSD reaction before arsenite addition. Reducing the repressor content was one way to shift the DSD equilibrium forwards; another is to increase the length or concentration of the reversible displacing oligo. High concentration PEG buffers were therefore tested with the X20 oligo, which had previously been too reactive for use in PEG-free solutions.

An example is shown below in Figure 5.10 for an X20 reaction in buffer DB7 (10 mM Tris-HCl pH 8.6, 50 mM NaCl, 5 mM $MgCl_2$, final 10% w/v PEG8k), containing many of the improved conditions explored earlier. In this version of the DSD5 protocol lower arsenite concentrations (0/1/10 µM) were added to the equilibrated system, with two replicates of each to gauge reliability. The arsenite-induced difference in reaction balance is presented here, with the raw and normalised data in Figure A.6.

The spread in data before arsenite addition is a result of the controls being triggered simultaneously to X20 wells, unlike in Figure 5.9. As mixing within the 10% PEG8k solutions is abnormally slow, the X26 controls are initially highly variable before settling at a consistent equilibrium after around 30

minutes (as shown in Figure A.6 (a)), after which time normalisation becomes more reliable. A strong signal increase is seen for the 10 µM arsenite samples, with the small variability between the replicates possibly deriving from slight variation in the volume of sodium arsenite solution added.

A consistent signal increase was also seen with the 1 µM arsenite samples. This sensitivity had not been reliably seen before, but these replicates seemed to be distinct from the water controls after one hour since addition and consistently rose for the remainder of the assay. The plate was sealed overnight and run further the next morning, with the right panel in Figure 5.10 demonstrating the stability of the system and the magnitude of the separation these low arsenite samples eventually achieved from the arsenite-free controls. An arsenite concentration of 1 µM is equivalent to roughly 75 ppb - above the WHO threshold but well within the range commonly found in Bangladesh groundwater samples, so robust sensitivity at this level is very useful.

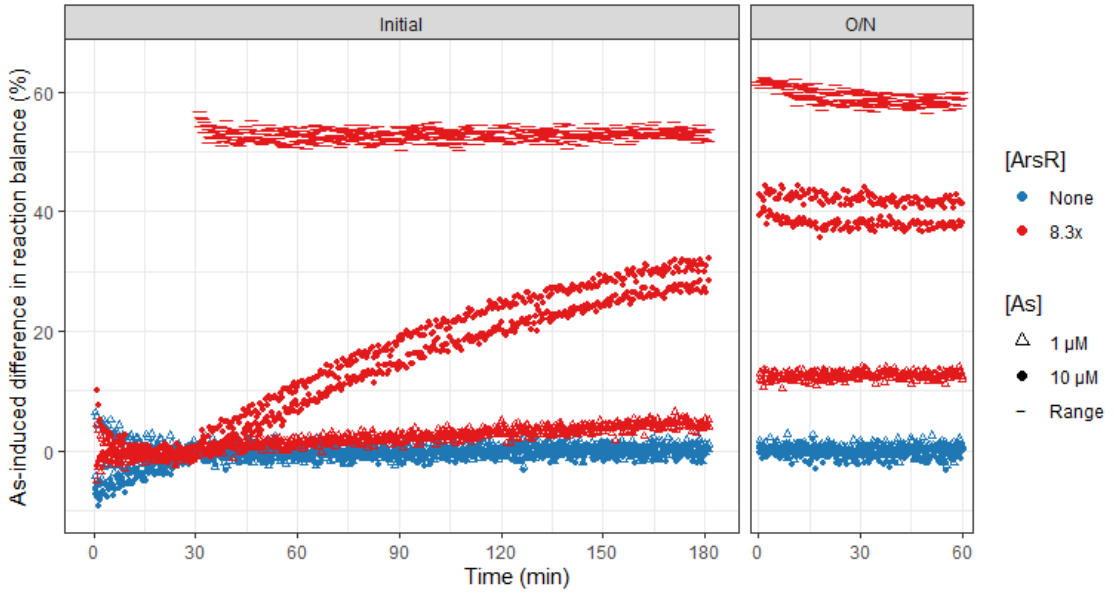


Figure 5.10: **Modified buffer and switch to X20 enables 1 µM arsenite sensitivity.** Fluorescence time course displaying data for L6T1C, incubated with and without CgArsR in the DB7 buffer, undergoing reversible X20 DSD then responding to 50 µl of 0/1/10 µM sodium arsenite, as in the DSD2 protocol. Initial DSD was run for 30 min before arsenite was added and the assay continued for another 2.5 hours. The plate was sealed then restarted the following morning for a further hour. Arsenite-induced difference in reaction balance is shown here, with the raw and normalised data shown in Figure A.6.

5.6 Visual DSD modelling

The potential for this repressor-operator biosensory system to be modelled has been suggested several times throughout this thesis. The minimal composition - only a single protein added to an otherwise exclusively DNA reaction network - made it feasible that creating such a model was realistic. Importantly, building a model would not be a superficial exercise - there were well-defined benefits for doing so, both for this immediate project and potential extensions.

First, the act of model creation often improves understanding of the underlying system. Do the components interact in the way we expect them to? What assumptions have we made in designing this system and are these valid? While basic reactions between structureless, 'non-biological' DNA sequences have been relatively well characterised, protein-DNA interactions, particularly in the context of DSD networks, are much less well understood. Can transcription factors simply block DSD reactions, or do they affect them in a more nuanced manner? Creating and evaluating such a model should allow these assumptions to be tested, either to confirm our suspicions or to force us to rethink our understanding.

One extension of this benefit is the ability of such a model to characterise repressor kinetics to a depth rarely achieved through other methods. Conventionally, affinity of a transcription factor to its operator and its ligand have been estimated through titrations on EMSA gels. These are notoriously imprecise, as well as only analysing single time points in an unnatural matrix environment. More modern methods of probing affinities in solutions do now exist, although even these tend to estimate association/dissociation constants rather than the more fundamental binding and unbinding rate constants. A kinetic DSD-repressor model has the potential to be a useful tool to infer such parameters.

Second, an accurate model could allow the optimisation of assay composition, substantially improving the performance of this particular biosensor. As has been outlined in recent sections, empirical experimentation has enabled notable advances in speed and sensitivity, however this is a slow iterative process and further gains are likely to be harder to find. There are simply too many variables in the design of this biosensor to ensure manual optimisation - the ability of an *in silico* model to rapidly simulate different formulations, evaluate their

performance and use a more holistic optimisation approach to identify ideal parameter combinations would significantly increase the chance of further assay improvements.

Finally, one major advantage to building an accurate kinetic model would be potential integration with existing DSD circuitry. Microsoft Research, alongside collaborating groups, developed the Visual DSD programming language as a tool to improve the design of DSD networks capable of computation. Since its inception, DSD reactions have been designed to perform simple mathematical operations, create logic gates and implement functions such as signal amplification or thresholding that could be useful to biosensor applications. As DSD technology progresses it may be possible to combine such motifs with the system presented here, simulate the effects of these expansions and design the additional DNA components to layer new functionality on top of the basal system, all *in silico*.

For example, an arsenic biosensor may be best constructed as a yes/no visual output on a test strip, like a pregnancy test. To transform a continuously variable output like that expected from the current assay design to a digital switch, a threshold gate or similar motif could be introduced into the assay. The ability to simulate such changes, optimise the design to cut-off at the 10ppb WHO threshold, then synthesise the new circuit for *in vitro* testing, would massively expand the functionality, range and development speed of new DSD devices, and bring design pipelines in line with non-biological technologies.

5.6.1 The basic M1 model

Models were constructed using Visual DSD to enable future integration as well as to make use of tools built into the software. An example script is shown in Figure 5.11. Visual DSD was developed in the functional F# language and requires the specification of program directives, parameters, domains, modules and concentrations. Importantly, reactions do not occur between arbitrary elements but between DNA strands and complexes with a defined domain structure and rate constants governing toehold interactions - the use of such a mechanistic model over a 'black box' approach helps to ensure components are always accounted for and side reactions are not missed, which becomes particularly useful as these networks scale.

```
directive simulator deterministic
directive deterministic {stiff = true}
directive simulation {
 initial = 0;
 final = 2700;
 plots = [FG_X()];
}

directive inference {burnin=5000; samples=5000; thin=1}
directive data [Data_191119_T6T0C_X19_normalised]
directive compilation default
directive rendering {classic = {mode = nucleotides}}
directive sweeps [X19_sweep=[x0=[2.0;2.0;2.0;2.0;4.0;4.0;4.0;4.0;8.0;8.0;8.0;8.0]]]

directive parameters [
k1 = 0.003, {interval=Real; distribution=Uniform(0.0001,0.01); variation=Random};
k2 = 0.003, {interval=Real; distribution=Uniform(0.0001,0.01); variation=Random};
u1 = 0.1126, {interval=Log; distribution=Uniform(0.001,10); variation=Random};
u2 = 0.1126, {interval=Log; distribution=Uniform(0.001,10); variation=Random};
x0 = 1.0;
d0 = 0.9852216749;
]

dom g6c = {bind = k1; unbind = u1; colour = "red"; seq = GGCGGG}
dom cg7r = {bind = k2; unbind = u2; colour = "blue"; seq = GGGTATA}
dom cg19 = {seq = TTGATAGGTGTGGATATTG}
dom t = {colour = "green"; seq = TTTTTT}
dom fl = {seq = T}

def Tprobe() = [fl]::<t^>[cg7r^ cg19]<g6c^>
def Gprobe() = <fl t^>[cg7r^ cg19]<g6c^>
def X() = <g6c^* cg19*>
def G_X() = <fl t^ cg7r^>[cg19 g6c^]
def FG_X() = [fl]::<t^ cg7r^>[cg19 g6c^]
def Quencher() = <cg19* cg7r^*>

(10*d0 Tprobe()
| 5*d0 Gprobe()
| 15*x0*d0 X() @ 0
| 0 G_X()
| 0 FG_X()
| 15*d0 Quencher()
| u1*<dummy>
| u2*<dummy>
\
```

Figure 5.11: **Visual DSD code for model M1.** All Visual DSD models were described using the structure and notation shown above. Experimental data was input using the 'directive data' command, with the 'directive sweeps' command specifying the X19 oligo concentration used to produce the normalised data in each column of the csv file. The *k* and *u* parameters were allowed to vary within the specified ranges, with the initial 'burnin' interactions picking values according to the specified distributions. The *d0* parameter adjusted the concentrations for the 3 µl volume increase from X oligo addition. Code was run through the interface at https://classicdsd.azurewebsites.net. Further notation explanation can be found in the user manual linked in the methods section.

DNA species are defined in the module section, using brackets and other punctuation characters to inform the single/double-stranded nature of their constituent domains and whether they are continuous or separated by nicks. Compilation into a chemical reaction network (CRN) initiates identification of unique ssDNA oligos from these structures, to be treated collectively.

Abstraction is a common feature of modelling, particularly of biological systems. Instead of simulating the movements and interactions of every molecule and atom, their behaviours are simplified and averaged across their population. The mechanics of toehold-mediated DNA strand displacement are non-trivial, and Visual DSD offers various levels of abstraction to tune model complexity. Instead of the simple swapping or displacement of ssDNA strands presented so far, in reality these species go through a three-step process of toehold binding, branch migration (the exchange of two identical strand domains) and unbinding of the displaced oligo. Each of these steps can proceed in a forwards or reverse direction, specified in Visual DSD's most complex 'detailed' model. The 'finite' abstraction merges unbinding and branch migration into a single parameter, while 'default' presumes migration to occur at an infinite rate. The most simple 'infinite' model assumes that all these steps can be combined into a single action modelled by one rate constant.

Although all these variations were tested, only simulations using the default compilation method are presented here as these seemed to provide the best balance of accuracy and model complexity. This meant that the reversible DSD reactions were modelled as a two-step process, displayed in Figure 5.12. This specifies that the X oligo binds its complementary toehold domain with rate constant *k1*, giving rise to a ProbeX complex, which unbinds with rate constant *u1*. As branch migration is assumed to occur at an infinite rate, this is equivalent to the species formed from the reverse reaction between fluorescent product and the quencher strands, governed by the *k2* and *u2* constants. All the rate constants used for the simulations in this section are listed in Table A.4 alongside the parameter ranges provided to model M1 in Table A.6.

While Visual DSD can be used to simulate reactions when components and parameters are all specified, its ability to perform inference is much more powerful. This is key when one of the aims is to estimate unknown parameters, such as rate constants, as in this work. The parameter space for each variable can be specified and the inference algorithm will score combinations based on

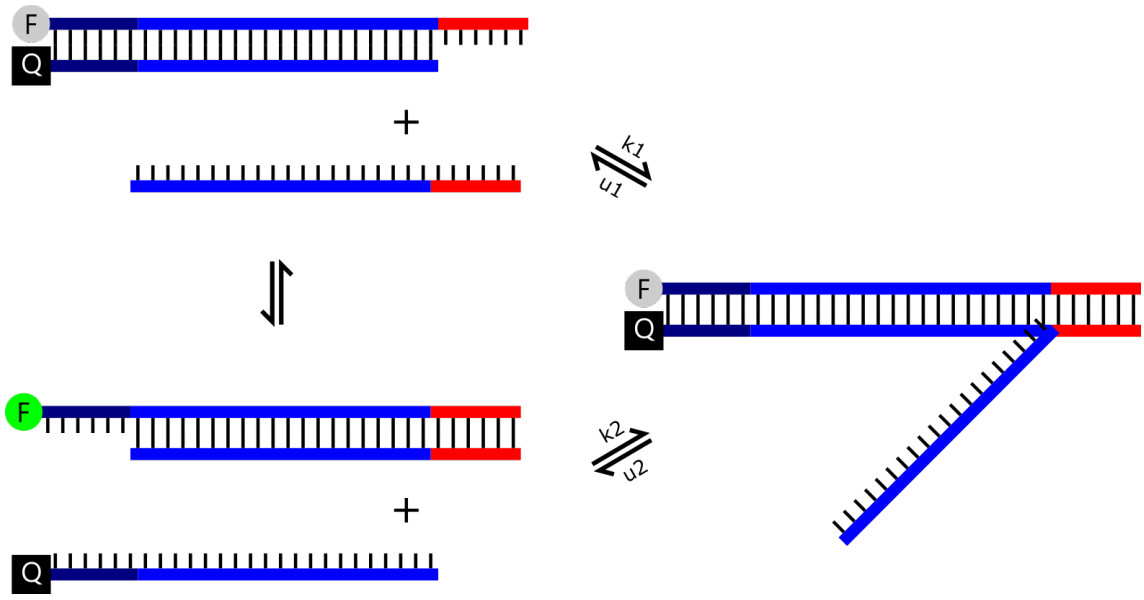


Figure 5.12: **Default Visual DSD compilation requires binding and unbinding rate constants.** The reversible reaction used for the assays are assumed to occur via a two-step mechanism, via a three-strand ProbeX intermediate. Binding events are governed by the bimolecular k constants; unbinding events by unimolecular u constants. The probe is shown as an FQ duplex for convenience - in reality this is either a Tprobe or Lprobe, but these are assumed to behave similarly.

how well the resulting simulations match experimental data; the model should iteratively converge on a set of parameters that best simulate the data provided. This is not the same as directly measuring rate constants, but allows many to be estimated in parallel in the context of actual reactions and the confidence in these to be quantified.

Under simple mass action kinetics, reaction rates are determined by the product of rate constants and the concentration of participating species. As a result, greater confidence in estimated rate constants is achieved when data is provided for multiple reactions over different concentrations of reagents. For DSD reactions, this usually means titrating the concentration of your input (displacing oligo) against a fixed concentration of your reporter (probe). Visual DSD, therefore, uses 'sweeps' to assess outputs deriving from different reagent concentrations.

5.6.2 Continuous normalisation enables more accurate inference

Modelling also requires the provision of meaningful data. As highlighted earlier, the RFU output of fluorescence readings does not directly relate to underlying chemistry - a normalisation method is required to infer the concentrations of different species from this raw data.

In Figure 5.13 an X19 titration of three different concentrations was carried out for the T6T0C probe, with the raw data shown in (a). Normalising to the initial values of the polyT and X26 controls (representing the fixed standards commonly used in other experiments) produces the data shown in the left panel of plot (b).

The lines overlaid are an attempt to infer the rate constants of the composite toeholds. As the signals reduce over time, due to some combination of photobleaching, micro-heating, sequestration or evaporation, the inference suggests rate constants with sizeable uncertainty (the noise parameter modelled by the grey shading), while the simulated traces clearly do not match the experimental data.

The normalisation protocol introduced at the start of this chapter, which continuously monitors minimum and maximum controls, instead produces the data in the right panel where the artefact of signal loss has been removed, which can be simulated with much greater confidence. This suggests that the normalisation methodology is accurately determining real concentrations from the raw data, and was crucial if any modelling attempt were to succeed.

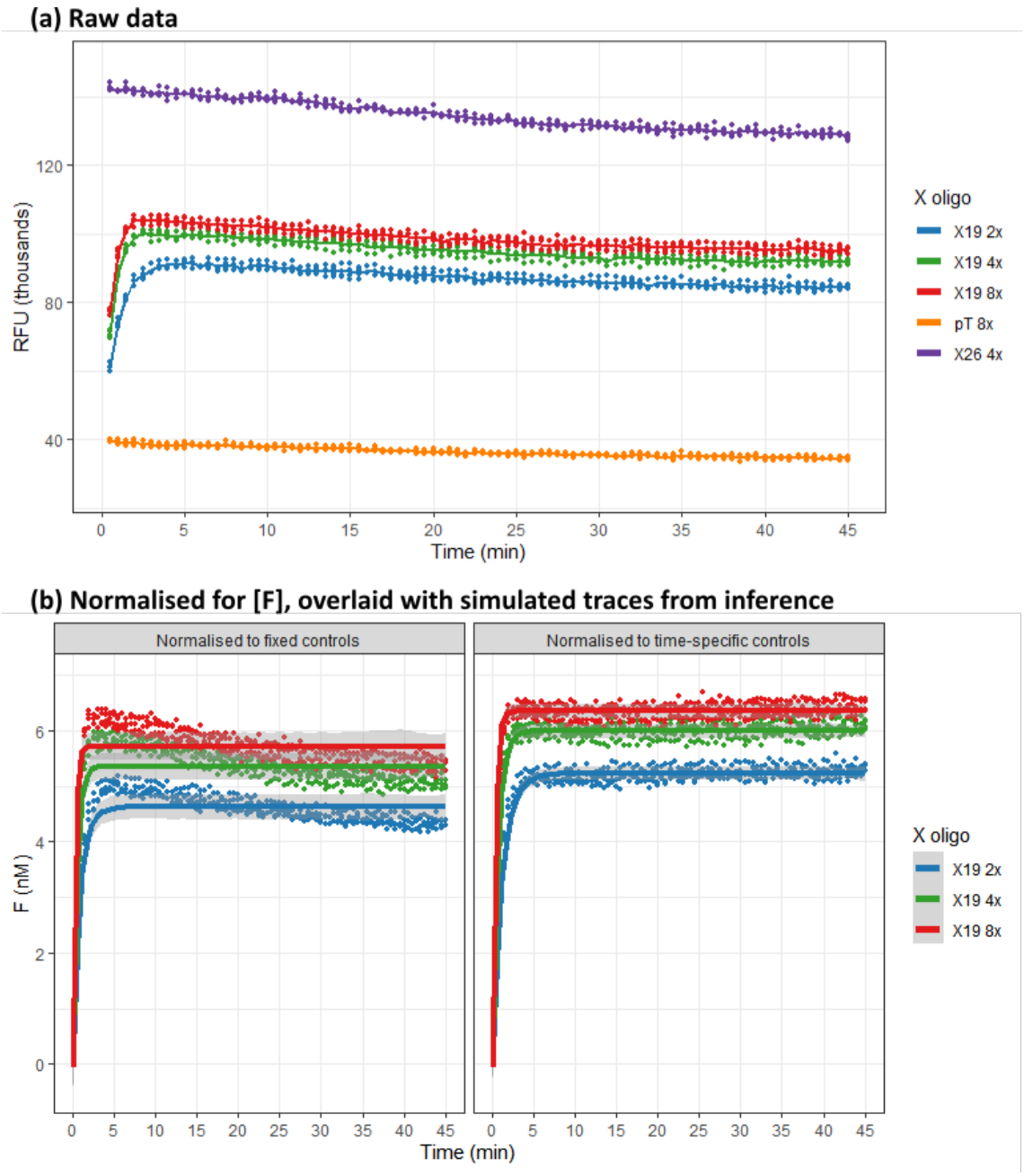


Figure 5.13: **Continuous normalisation enables more accurate inference.** Fluorescence time course displaying data for T6T0C reacting with corresponding X19 oligos under the DSD1 protocol in buffer DB1. (a) Raw data for X19 reactions as well as pT-30 and X26 controls. (b) X19 data normalised to the average initial RFUs of the controls (left) or time-specific control averages (right), as different methods to estimate the underlying concentration of unquenched fluorophore. These data were used to infer binding and unbinding rate constants for both the forward and reverse toeholds using Visual DSD model M1, and simulated reaction traces are overlaid for each including error margin (grey shading) as given by the 99% confidence interval of the simulation and parallel traces shifted either side by the estimated sigma noise parameter.

5.6.3 Structured probes are too complex for simple model

Normalisation methodology was not the only development required for sensible modelling. The advances in probe design discussed in the previous chapter - removal of secondary structure and FGX quenching - were also necessary. Figure 5.14 shows a titration carried out on the native operator TR26-G6a probe and its normalised data overlaid with a simulation using inferred rate constants. While the model is able to capture the general trajectories of the data, it clearly does not fully replicate the displayed dynamics.

On closer inspection there are further flaws in this inference. First, the assay controls are not robust enough for accurate normalisation. The X24 oligos react too slowly - they take several minutes to peak, meaning that the initial data for the X18 reactions were normalised to an artificially low maximum RFU, thereby inflating their initial reaction speeds. The polyT control also seems to creep up over time - an indication that perhaps these exhibited slight degradation or minor contamination. This would result in a small underestimate of reaction balance towards the end of the assay. Second, we now know that all the reaction products were slightly quenched using this probe, but the FG-X18 species probably slightly more so than FG-X24 due to the greater number of exposed residues adjacent to the fluorophore. This would have slightly underestimated the reaction balance throughout the whole assay.

The dominant factor in the dynamics of the native probe, however, was the substantial secondary structure present in the competing single-strands. The core 'GGTAT' inverted repeat in both the displacing oligos and the released QR complex meant that both the forward and reverse reactions were greatly slowed. This is reflected in the time taken to reach equilibrium for the TR26-G6a titration (hours versus minutes for T6T0C), but inevitably also in the rate constants inferred from Visual DSD. Without any mechanism for the effects of secondary structure in the simple model, the slow reaction speeds are reflected primarily in smaller binding rate constants.

The values suggested for the forward and reverse toehold binding constants of the native probe (2.70×10^{-5} and 8.70×10^{-3}) were orders of magnitude different to the T6T0C probe (4.29×10^{-4} and 3.14×10^{-4}). This clearly does not represent the toehold strengths alone - both the forward toeholds were fully GC (X6a: GCCCGC; X6c: CCCGCC), while the reverse toeholds had similar content with

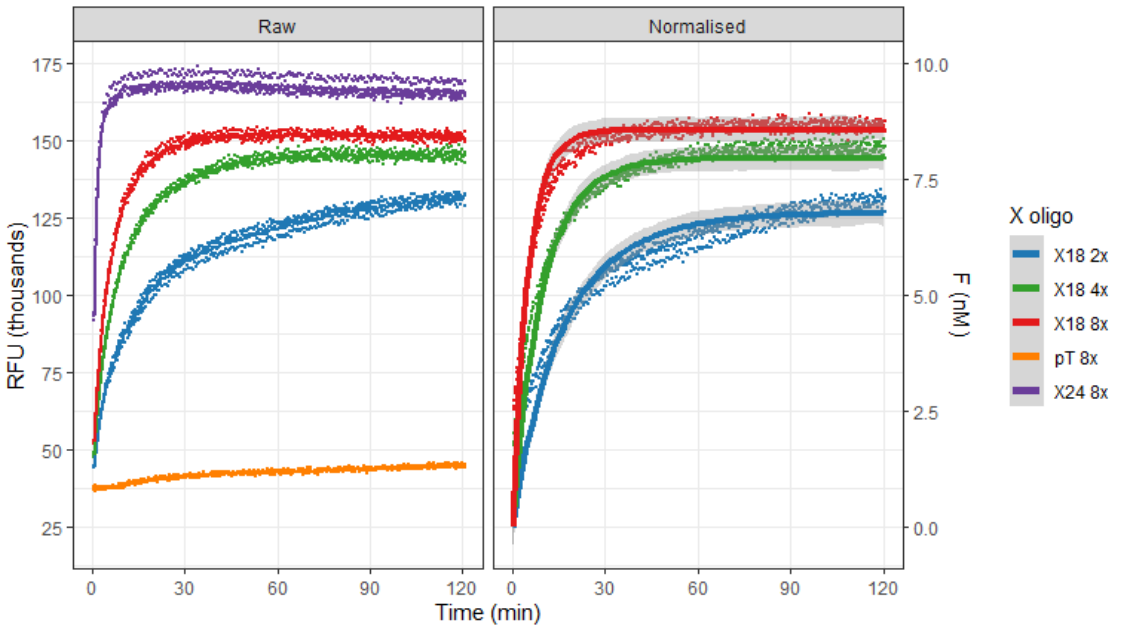


Figure 5.14: **Native probe is poorly simulated.** Fluorescence time course displaying data for TR26-G6a reacting with corresponding X18 oligos under the DSD1 protocol in buffer DB1. Raw data including polyT-30 and X24 controls are shown on the left, with normalised data on the right overlaid with lines from a simulation using parameters inferred by Visual DSD model M1.

only 1nt difference in length (native: GAATATCG; T6T0C: GGGTATA). This highlights the care needed when writing the models and examining the inference results - even if the simulation somewhat resembles the data (as for Figure 5.14) this does not mean the model is at all accurate. The parameters must be sensible and predictions must be outlined and tested.

The important question to ask is therefore, what are sensible values for these rate constants? This can be a difficult question to answer as reaction context is important - buffer, temperature and interrogation method could all influence measured constants for particular sequences. However, similar conditions were used in an analysis performed by Zhang *et al.* [57], who probed the relationship between toehold sequence and length on binding energies and rate constants. The X6c toehold was in fact based on their 'strong' toehold domain for this reason, known in their work as γs^6. They calculated the representative strong sequences of 6nt or greater to have rate constants of roughly $6 \times 10^{-3} \text{nM}^{-1}\text{s}^{-1}$. Average or weak sequences had binding rate constants of around $3 \times 10^{-3} \text{nM}^{-1}\text{s}^{-1}$ and $4 \times 10^{-4} \text{nM}^{-1}\text{s}^{-1}$, respectively.

This aligns with the inferred values for the native probe's reverse toehold and the forward toehold for T6T0C, but the others were further off, particularly for

the 6a toehold whose binding constant was more than 100-fold lower than expected. Of course, this was due to the structure of the native displacing oligos but, as shown in Figure 5.14, this abstraction (combining toehold strength and oligo structure) could not accurately capture reaction dynamics. This difficulty in modelling structural effects was a big reason for attempting to alter the native probe design.

Across different Tprobe simulations, binding constants were consistently estimated smaller than suggested by Zhang *et al.*. This was particularly the case for *k2*, which tended to be 5-10x lower than expected for a 7nt, medium strength toehold. One hypothesis was that the Tprobe structure inhibited the speed of reverse reaction, due to the steric hindrance as the QR complex accesses the reverse toehold of FGX. This was a likely reason the Lprobe reactions consistently reached equilibrium faster than the Tprobe reactions. Once again, the suggested rate constants could well be encompassing structural effects as well as toehold properties. For this reason, inference on Lprobe reactions was subsequently preferred.

5.6.4 Unbinding rate flexibility can enable overfitting

While the default compilation method offers a good balance between model complexity, risk of overfitting and accuracy, within this there are further levels of abstraction. Binding rates clearly vary between different toeholds, but the variation in unbinding rates is more subtle. Visual DSD allows the flexibility to assume all toeholds unbind at the same default rate ($0.1126\,s^{-1}$), they all unbind at the same, non-default rate (to be inferred), or they each unbind at their own rates. Once again, this most complex scenario is closest to the truth - unbinding rates are also dependant on sequence and length too and are therefore related to the toehold's corresponding binding rate.

Figure 5.15 shows simulations of an L6T1C X19 titration, using inferred data where the unbinding rates have been constrained to different extents. Plot (a) shows the result when both toeholds are forced to use the default unbinding rate, while (b) forces the two rates to be identical, but the value can vary over a couple orders of magnitude and (c) enables each rate to be independent. All three simulations fit the data well, although flexible unbinding increased likelihood scores slightly (see Table A.4).

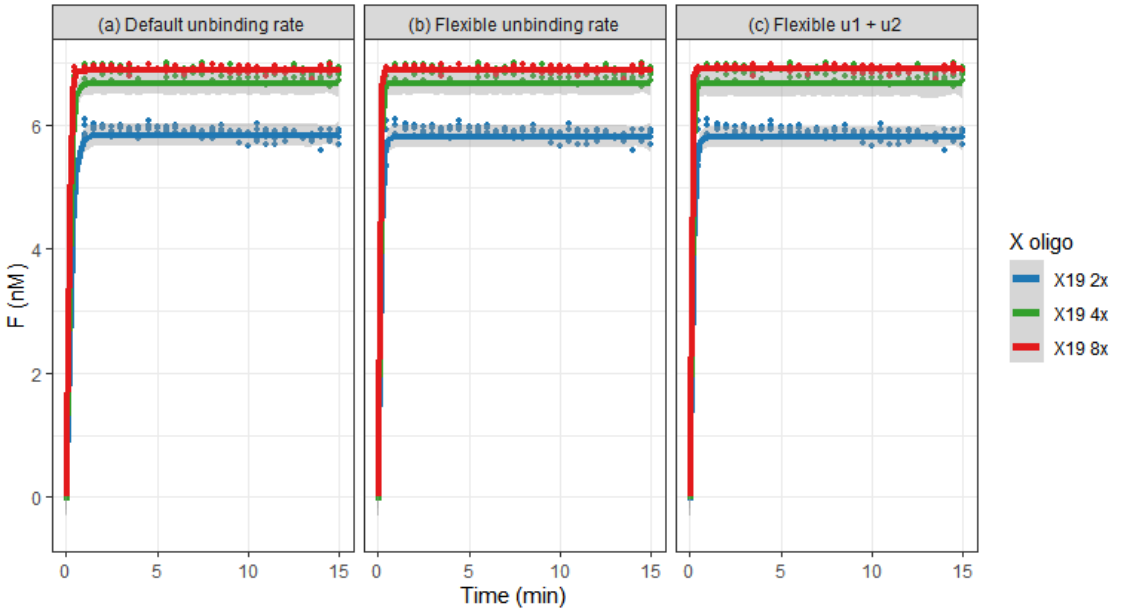


Figure 5.15: **Unbinding rate flexibility can enable overfitting.** Fluorescence time course displaying data for L6T1C reacting with corresponding X19 oligos under the DSD1 protocol in buffer DB1. Points are data from four replicates of each concentration continuously normalised to fluorophore concentration. Lines are from simulations using parameters inferred by Visual DSD model M1 run on the full 45 min assay, but only the first 15 min are shown for clarity. (a) Inference using the default toehold unbinding rate constant. (b) Inference allowing variation in a single unbinding constant for both toeholds. (c) Inference allowing variation in unbinding constant for each toehold.

Despite this similar fit, all three simulations had very different binding constant values ($k1 = 4.84/9.37/5.00 \times 10^{-3}$; $k2 = 1.44/2.93/4.21 \times 10^{-3}$), resulting from their flexibility in unbinding rate selection. This ability to fit the data almost perfectly with so many different parameter combinations suggests the model was insufficiently constrained, as demonstrated by the disparate distribution of posterior values in Figure A.7. To try to resolve this issue, further inferences were run specifying relationships between binding and unbinding constants.

One such relationship was used by Zhang *et al.*, relating the Gibbs free energy of sequence-specific toehold binding to the toehold's binding and unbinding rates, shown in Equation 5.6. Here, the unbinding rate constant (u) is dependent on the binding constant (k), the length of the recognition domain (b - m, in this case 19 nt), and the estimated free energy of toehold binding (ΔG). R and T are the gas constant and temperature in Kelvin, respectively:

$$u = k \times \frac{2}{b - m} \times e^{\Delta G/RT} \tag{5.6}$$

This was tested in Figure 5.16, using the estimated binding energies of −12.1 kcal/mol for the forward toehold and −9.2 kcal/mol for the reverse. Using these values in Equation 5.6 results in the following relationships: $u1 = 0.145 \times k1$ and $u2 = 19.3 \times k2$, tested in plot (a). This fixed relationship severely restricted ability to converge, with two orders of magnitude between these proportionality constants forcing either very different binding or unbinding constants, resulting in simulated reactions with a very poor fit. In particular both inferred binding constants were very close to the upper limit - too high to be reasonable - and giving high reverse and low forward unbinding constants.

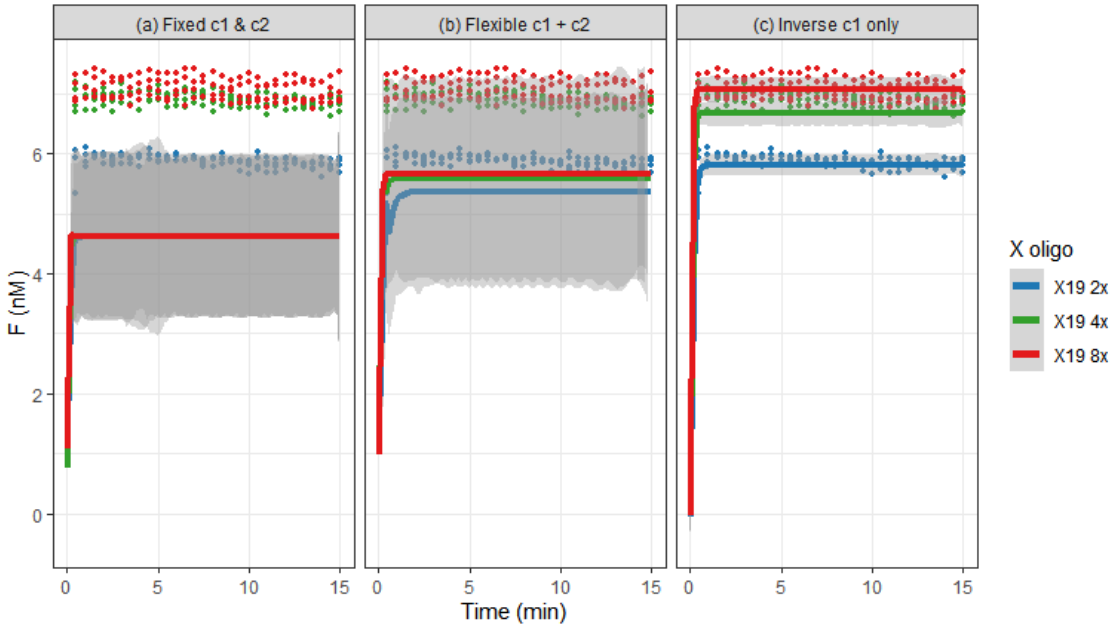


Figure 5.16: **Inverse *k*/*u* relationship better constrains parameters but enables same fit.** Fluorescence time course displaying data for L6T1C reacting with corresponding X19 oligos under the DSD1 protocol in buffer DB1. Points are data from four replicates of each concentration continuously normalised to fluorophore concentration. Lines are from simulations using parameters inferred by Visual DSD model M1 run on the full 45 min assay, but only the first 15 min are shown for clarity. (a) Inference using $u1 = 0.145 \times k1$ and $u2 = 19.3 \times k2$, as suggested by Zhang *et al.*. (b) Inference using the same relationship as in (a), but allowing variation in each constant relating the parameters, according to binding energies of toeholds +/-1nt. (c) Inference specifying an inverse relationship between k and u, related by a shared proportionality constant.

By allowing the *c1* and *c2* proportionality constants to vary between values for toeholds +/-1nt in length for their respective strengths, simulations were slightly improved but still very poor as seen in plot (b). The slightly different binding energies meant $0.00092 < c1 < 3.014$ and $0.203 < c2 < 87.9$, so only a minimal overlap existed between these constants. The inferred *c1* value was right against

its upper bound, suggesting that the low forward unbinding constant was the most limiting factor. The lack of any fit suggested the proposed relationship was invalid, either through poor data, incorrect binding energy estimates, or highly divergent experimental setups.

The general relationship proposed by Equation 5.6 was that the greater (more negative) the binding energy, the lower the unbinding constant relative to the binding constant. A simplified version of this was tested in plot (c), constraining k and u to be inversely proportional to each other, with a shared proportionality constant between the two toeholds ($c1$). This resulted in a fit just as good as for the fully flexible constants, suggesting the constants were not excessively constrained. Further to this, the posterior distribution was much smoother with a clear peak frequency, seen in Figure A.8, suggesting the model was not obviously overfitted, with the maximum likelihood parameter set identified with $c1 = 1.16 \times 10^{-3}$.

In addition, the inferred binding constants of $6.83{\times}10^{-3}$ and $3.81{\times}10^{-3}$ were very close to the values estimated by Zhang *et al.* for strong and moderate toeholds. While not conclusive it suggests the model may be accurately abstracting the fundamental dynamics of the system.

5.6.5 CgArsR kinetics may not be straightforward

As integration of transcription factors with DSD reactions had previously not been attempted, little is known about exactly how these components interact. Since the transition to the use of reversible reactions, and the appreciation that repressors will naturally unbind at an appreciable rate, the protein has been assumed to behave through the simplest possible mechanism - it can bind and unbind its operator sequence in the quenched probe and when bound, the repressor-probe complex cannot react with displacing oligos as the toehold domain is occluded. DSD reactions only occur on repressor-free probes, when they become available through natural or arsenite-induced dissociation.

This behaviour was added to the Visual DSD model by defining ROC formation as an arbitrary chemical reaction, controlled by the *k3* binding and *u3* unbinding constants, shown in Figure 5.17 (a), giving model M2.

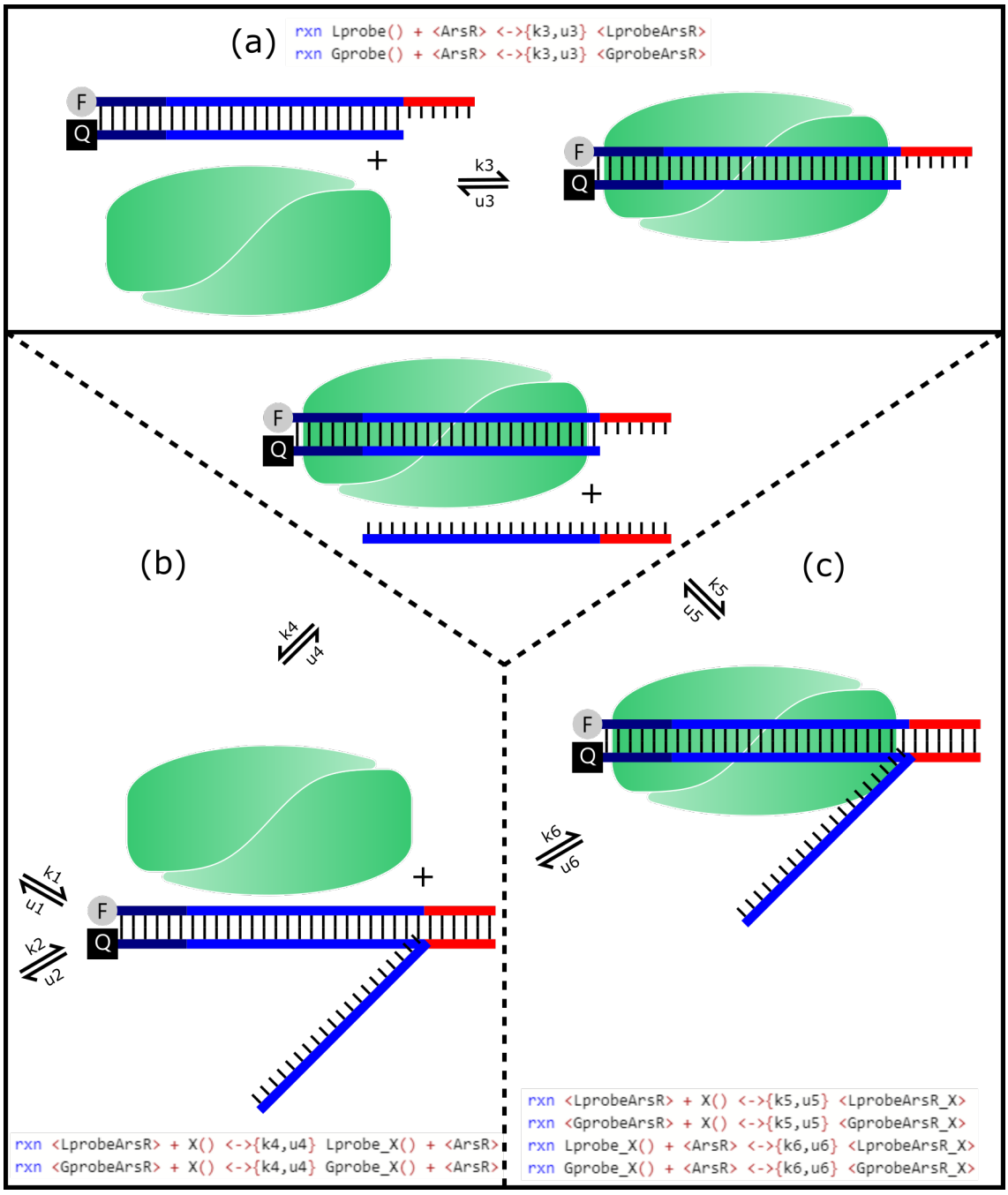


Figure 5.17: **ArsR reactions modelled in Visual DSD.** Schematic displaying the various ArsR reactions included in models M2-4. (a) The predominant expected ArsR behaviour - reversible binding of its operator sequence (in both L/Tprobes, as well as the Fstrand-free Gprobes for each), forming a species that does not undergo normal DSD reactions. (b) A supplementary theorised reaction included in model M3 the X oligo is able to react with the bound probe, but with rate constant $k4$, assumed to be low. The reverse bimolecular reaction is governed by $u4$. The ProbeX species can then react according to the $u1$ and $u2$ constants as well. (c) A more in-depth mechanism replacing the $k4/u4$ reaction, used in model M4. Here, a Probe-ArsR-X species can form, either from X binding the bound probe or from ArsR binding the ProbeX species. Two reactions ($k5/u5$ and $k6/u6$) replace $k4/u4$, but the unbinding reactions are unimolecular.

The difficulty is then how to accurately parameterise these constants. As seen earlier, excessively large ranges can enable model overfitting by giving implausible values, but setting limits that do not cover the real values is also clearly problematic. While many transcription factors have vague association/dissociation constants estimated for them, often based on unreliable data, rate constant evaluations are much harder to come by. A couple of examples were found however, suggesting binding and unbinding constants for the AP-1 transcription factor of roughly $1.4 \times 10^{-4}\text{nM}^{-1}\text{s}^{-1}$ and $7.5 \times 10^{-3}\text{s}^{-1}$ [159], and $2 \times 10^{-4}\text{nM}^{-1}\text{s}^{-1}$ and $5 \times 10^{-3}\text{s}^{-1}$ for the bacteriophage CI repressor [160]. These figures were surprisingly similar but due to the uncertainty in how representative these TFs are a 100-fold range either side was allowed.

Model M2 was used on data from an L6T1C X19 titration, similar to that seen in Figures 5.15 and 5.16, but where the probe had been pre-incubated with CgArsR for one hour before the reaction. This time gap was replicated in the model - setting the start time to '-3600' seconds and X oligos added at time 0 - to ensure that, like in the actual assay, an equilibrium had been reached between probe and repressor by the time DSD began. Values inferred in Figure 5.16 (c) for *k1*, *k2* and *c1* were then fixed, while *k3* and *u3* were allowed to vary. The data and the simulation using inferred *k3* and *u3* parameters of 3.53×10^{-3} and 2.01×10^{-2} are shown in Figure 5.18 (a).

This simulation roughly fits the data, but is not capable of capturing the different equilibria caused by the X concentrations. The initial speed of response to X addition promisingly showed the simulation could adapt quickly, but both the inferred binding and unbinding constants for ArsR were an order of magnitude higher that the values quoted earlier for other transcription factors. While it could be the case that CgArsR did indeed behave differently to these, the inference could have been inaccurate. Figure 5.18 (b) plots the same experimental data, but overlaid with a simulation for which both *k3* and *u3* have been manually reduced 10x from the values inferred for (a). As expected, this preservation of the ratio between the two constants maintains the placement of the equilibria, but the magnitude reduction slows the initial speed.

It could be the case, therefore, that the ArsR constants are of this slower magnitude but other factors cause the initial fast increase. Disturbance from the act of solution addition could be a possible explanation, while another could be alternative DSD mechanisms for the bound probe. Two such possibilities are

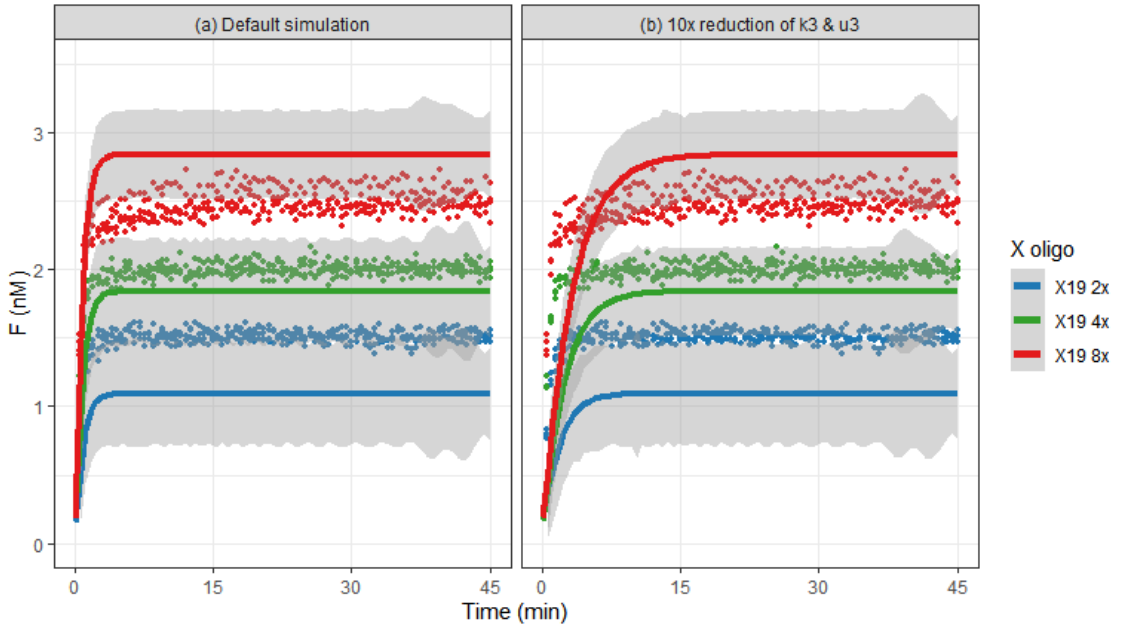


Figure 5.18: **M2 inference predicts higher than expected CgArsR rate constants.** Fluorescence time course displaying data for L6T1C, incubated with a 16.7x excess of CgArsR, reacting with corresponding X19 oligos under the DSD1 protocol in buffer DB1. Data normalised to fluorophore concentration are shown as points, with the simulated data from Visual DSD inference model M2 overlaid as lines in (a) and another simulation using 10x smaller k3 and u3 constants shown in (b).

presented in Figure 5.17. In (b), the X oligo is able to access the bound probe and force the repressor off. This would produce the same ProbeX intermediate as the usual DSD reactions, which could then unbind to reform the probe or form the fluorescent product. Such a mechanism would occur with rate constant *k4* - presumably much lower than *k1* due to the hindrance of the repressor, but perhaps large enough to be consequential, especially at the start of the reaction when probe concentration is at its highest. In this scenario, the X oligo and ArsR could not bind at the same time, but the reverse reaction could theoretically occur through another bimolecular reaction using *u4*.

A similar but more detailed mechanism is shown in (c). This allows an intermediate species to form, comprising the bound probe and X together and the *k4*/*u4* reaction is replaced by two more simple second order reactions. This may more accurately reflect this mechanism if it did indeed occur, but the extra parameters may enable overfitting. Models M3 and M4, respectively, contain these additional mechanisms and their results are shown in Figure 5.19. As can be seen, these produce almost identical simulations, both of which fit the data much better than the model M2 inference as the additional reaction mechanisms

provide alternative ways for the displacing oligo to be utilised.

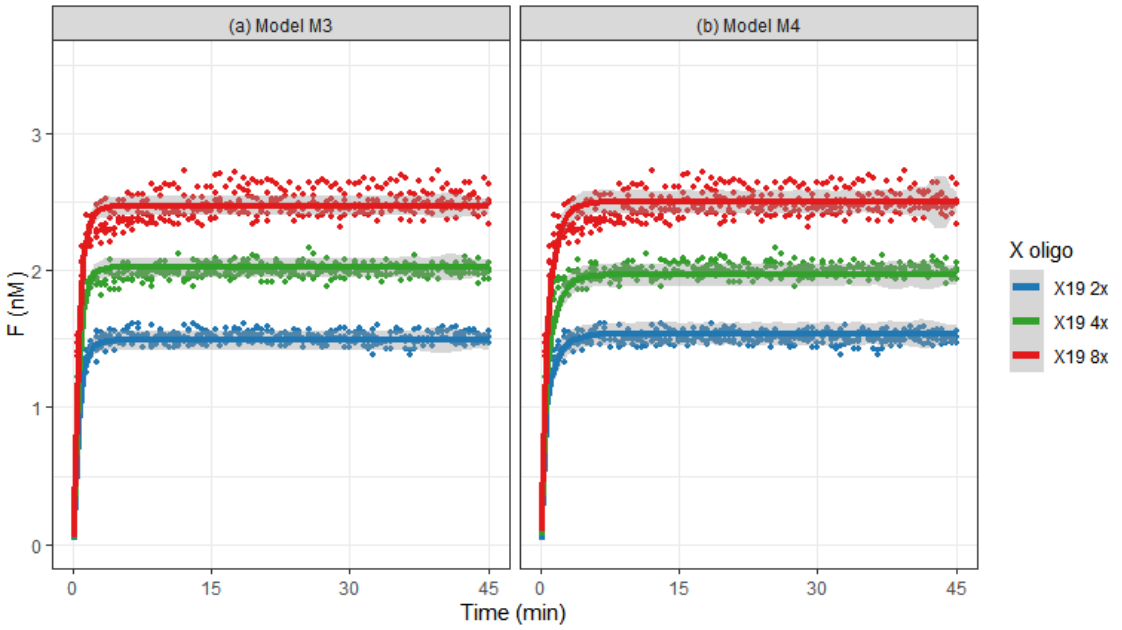


Figure 5.19: **Models M3 and M4 fit data much better, but could be overfitted.** Fluorescence time course displaying data for L6T1C, incubated with a 16.7x excess of CgArsR, reacting with corresponding X19 oligos under the DSD1 protocol in buffer DB1. Data normalised to fluorophore concentration are shown as points, with the simulated data from Visual DSD inference model M3 overlaid as lines in (a) and from model M4 overlaid as lines in (b).

Before taking the modelling further it is important to test the validity of these alternative mechanisms. To do so, both further modelling on different datasets and experimental validation of these possibilities is crucial. Inference on other assay formats have been attempted, with two X oligo and multiple ArsR concentrations within the same experiment, but the DNA-only simulations seem to suffer once fewer than three concentrations are provided, and the combinatorial increase of parameter space means the inferences take much longer and often do not obviously converge.

Future assay formats will attempt to combine more X oligo concentrations with different CgArsR master mixes, although careful planning will be needed to ensure neither reliability nor early time points are lost as a result. These experiments will hopefully produce more robust data on ArsR behaviour, so that modelling can progress towards inclusion of arsenite.

5.7 Conclusions

This chapter set out to enable system optimisation through improvements to protocol design, data processing methodology, buffer composition and modelling. These first three alterations were highly successful. Oligo sequestration and mixing were much better controlled through the introduction of new master mixes, polyT carriers and fine tuning of solution additions. Normalisation to continuously sampled, accurate controls enabled exposure of the underlying reaction balance, crucial for quantitative and kinetic analysis. Small changes to buffer constitution were able to improve both reaction sensitivity and speed, through manipulating conditions in favour of the repressor and arsenite itself.

The modelling did not yet progress to the original aims of enabling system level optimisation, or simulation of integration with other DSD motifs. Many of the other advances listed above (and the probe design improvements in the previous chapter) were necessary to establish collection of appropriate data so work with Visual DSD begun later in the project than anticipated. Despite this, headway has still been made towards making these aims eventually possible. The basic model has been built and parameter ranges and relationships that allow successful inference have been identified. Different compilations and levels of abstraction have been tested and the minimal level of detail required to accurately describe the base system has been identified.

Conclusions have also been made regarding the use of different probe designs. The structure within the Tprobes clearly seem to affect their kinetics, revealed in their inferred rate constants. While it is not known to what extent this behaviour can be accurately captured through altering these constants alone, the Lprobes seem to react in a more predictable manner, making it the more desirable construct to use for initial testing. This was able to make early attempts at modelling ArsR behaviour, and to test the assumptions and understanding over how the repressor may interact with DSD components. Other than testing for the presence of bound-probe DSD, other associations such as non-operator binding by ArsR may need to be investigated. Further work will surely shed light over the complexity of model required to capture its actions.

Some of the experiments showed that optimising repressor-operator affinity alone was not the best strategy. The goal of an arsenic biosensor is fast, sensitive

and reliable detection of arsenic - any optimisation must take place from the perspective of this species itself. Once ArsR is well described in the model then arsenic addition will be the last major hurdle. This is sure to present further challenges - in the current assay this will change solution composition, diluting all components including buffer salts. Rate constants would need to be inferred for this final state and not just the controlled conditions beforehand.

Although limited modelling progress has been made so far, the process of attempting it forced the creation of a more controlled system, leading to faster and more robust outputs. It has also challenged assumptions about how simple this system really is. With more time a complete model will be possible, but this may require a whole project of its own to tackle this problem now that a settled system has been established.

Chapter 6

Establishing a sensitive, selective, fast and robust final assay

6.1 Introduction

Work thus far has focussed on probe and assay developments enabling better understanding of the system at hand. In this chapter, the emphasis shifts towards more practical improvements useful for design of an actual arsenic biosensor. First, a new rate-of-change analysis combined with a more concentrated master mix is tested to improve speed and reliability of detection. Second, an arsenite-insensitive mutant repressor is expressed to allow better control of contaminants affecting the assay non-specifically. Third, assay lyophilisation is attempted in order to improve biosensor storage as well as to further increase sensitivity.

These advancements were highly successful and not only enhanced the assay's theoretical capabilities, but brought it much closer to real-world usage. Biosensors face many challenges in the field and identifying and overcoming these as early as possible in the developmental phase is hugely important for eventual success.

6.2 Implementing more concentrated assay with rate of change analysis

The protocols used in the previous chapter were designed for assessment of DSD kinetics and CgArsR behaviour. These insights proved very useful however, as discussed, it is response to arsenic that is the primary aim for this biosensor. With the change of emphasis from improving understanding of component interactions to testing the full capabilities of the current system, protocol setup was one of the clearest areas of interest. Adding 50 µl of arsenite solution to 200 µl DSD reactions enabled good resolution of initial DSD kinetics and equilibria, while allowing substantial changes to arsenic response to be noted, however using a more concentrated master mix and reversing these volumes should allow greater sensitivity.

This was partly enabled by the experiments in section 5.5 that revealed the repressor-probe system performed best in relatively low salt conditions, while being able to tolerate more concentrated solutions. Master mixes were therefore prepared more concentrated and already containing displacing X oligo, so that 50 µl solutions were four-fold concentrated over what used to be 5 µl X oligo + 195 µl MM reactions. 200 µl of arsenite solution could then be added to each well, bringing the total to 250 µl - the same volume and concentrations as before, save for four-fold more arsenite for each sample. As the composition should end up almost identical to the final stage of previous assays except for a larger ratio of arsenite to repressor, greater sensitivity would be expected.

Figure 6.1 shows the result of such an assay. Thirty-six wells were loaded with identical 50 µl master mix solutions and twelve different 200 µl arsenite samples were added in triplicate. Plot (a) shows the raw data coloured according to arsenite concentration, with the arsenite-induced signal increase for each below, calculated from subtraction of the average RFU of the arsenite-free solutions. This clearly shows that an equilibrium can be established through co-incubation of L6T1C probe, CgArsR repressor and a displacing oligo, in this case L6T1C's X19 oligo, to then respond to different arsenite solutions. The DSD reaction has to first adjust to the five-fold dilution, initially increasing in signal before slowly settling down to its baseline. Solutions containing arsenite are perturbed slightly through the gradual removal of ArsR, thereby increasing their signal relative to the arsenite-free samples.

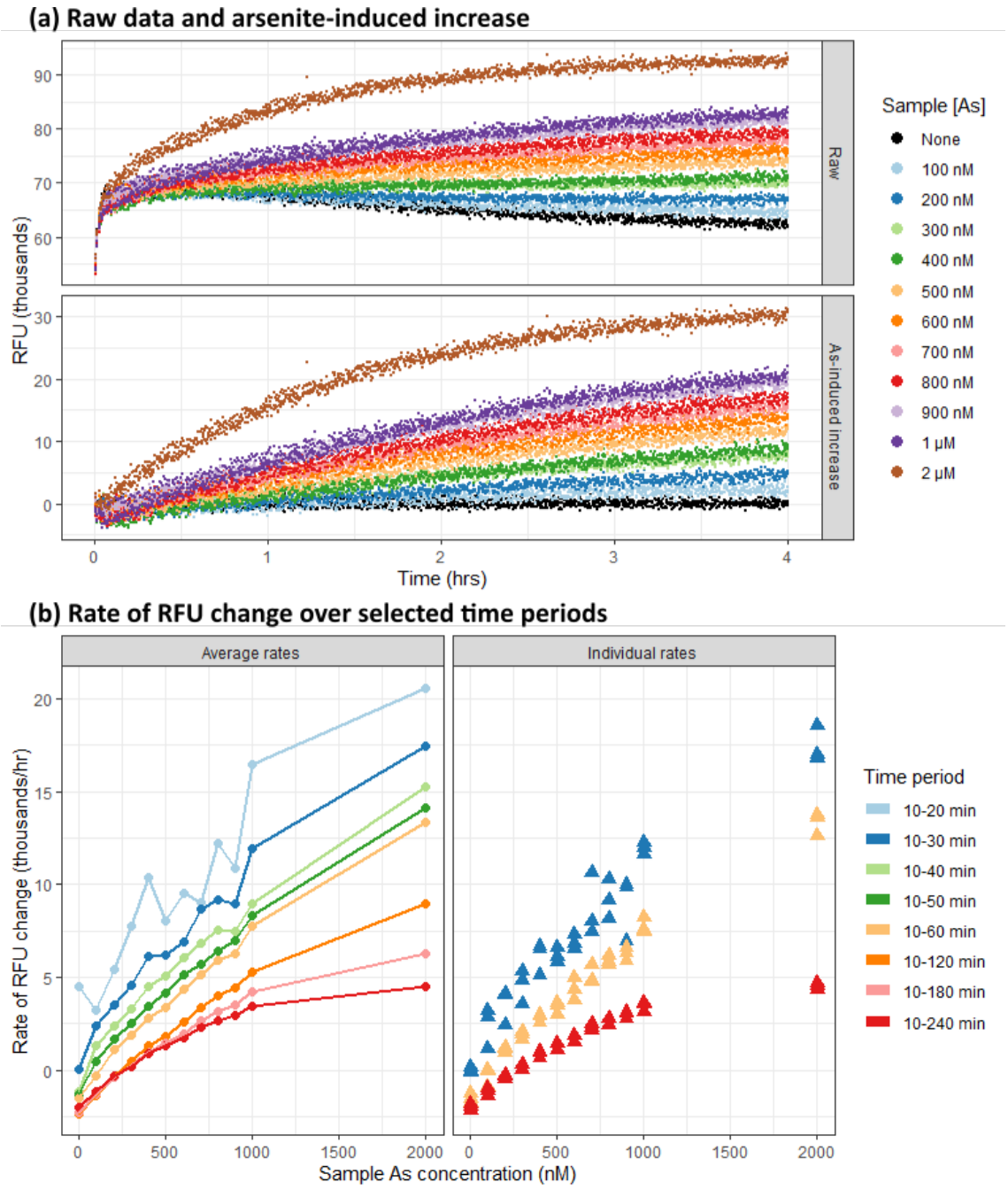


Figure 6.1: **Concentrated assay boosts sensitivity while rate analysis increases detection speed.** Fluorescence time course displaying data for an L6T1C-X19-CgArsR solution reacting with varying arsenite solutions under the DSD6 protocol in buffer DB8. (a) Raw data for these samples over 4 hours with points for individual wells in triplicate. Arsenite-induced signal increase is shown below, derived from subtraction of the average RFU of the arsenite-free solutions. (b) Average rate of RFU change for each sample calculated by creation of a linear model across different specified time periods, plotted by added arsenite concentration. The rates of individual wells for the 10-30, 10-60 and 10-240 min time periods are shown separately to showcase differences in variation.

As hoped, this effect was arsenite concentration dependent, with strong linearity over most the range tested. While the micromolar samples started to saturate over the second half of the experiment, the lower concentration samples displayed a consistent rate of increase above the baseline over the four hours.

The lowest arsenite concentration sampled, 100 nM, is roughly equivalent to 7.5ppb, under the 10ppb WHO threshold. The light-blue trace indicates that these samples can be distinguished from the arsenite-free solutions by the end of the assay, however the gradual separation makes it difficult to pin down after how long this result is significant. Using simple T-tests on the triplicates of each concentration (assuming one tail and equal variance), in the first hour 3/120 time points are significantly different ($p<0.05$), while in the second, third and fourth hours 13/120, 90/120 and 119/120 are significantly different, respectively. The small samples sizes and parametric assumptions make individual T-tests unreliable, however the significance of so many independent tests in the second half of the assay suggest that the signal of the 100 nM arsenite samples becomes significantly greater than the arsenite-free samples by roughly three hours. While a faster detection speed would clearly be desirable, achieving sensitivity below the WHO threshold is a major result.

Given the steady rate of RFU change for the low arsenite concentrations, an alternative method of judging difference between samples was proposed. Instead of comparing values at particular time points, the rate of change over a predetermined time period could assessed. While the data for the first 5-10 minutes of the assay primarily reflects the re-equilibration following well dilution, the subsequent divergence would seem to correlate with arsenite. Figure 6.1 (b) shows the rate of RFU change of sample averages over different time frames, plotted by their arsenite concentration. As expected, the earlier, shorter time periods are relatively noisy, however using this analysis a rough idea of arsenite concentration can be established within half an hour. Over longer time frames the correlation between arsenite concentration and rate of RFU change becomes more robust, enabling greater confidence in an inferred arsenic quantity. The rates of individual wells for a few time periods are shown alongside to display the underlying noise and how this reduces with longer windows. Within half an hour samples with 200 nM or more arsenite seem as though they could be distinguished from arsenite-free samples with great confidence, while the 100 nM solutions may be identified as containing low but non-zero quantities, with a longer wait enabling more accurate estimation.

Such an analysis could be used to threshold arsenic ranges with progressive confidence. For instance, after 20 minutes a rate of RFU change of +15 or more for the 10-20 min period could be used to indicate 'high' arsenic content in the tested sample, with a lower rate suggesting none-to-medium levels. By waiting longer these ranges could be refined and by an hour a rate of +7 or more for the 10-60 min period could be used to indicate 'high', with 1-7 'medium' and under 1 'low' or none. The gradual reduction of these rates over time reflect the imperfect linearity of the changing rates - perhaps a more complex function would give more consistent coefficients.

Another benefit of this analysis method, other than potentially increasing detection speed, is some level of contaminant screening. Real groundwater samples would not be perfect arsenite-only solutions, but messy suspensions that even when filtered would have many likely interfering solutes. In section 5.5, it was shown how both sodium and magnesium chloride concentrations can affect the equilibria of a DSD reaction as well as the affinity of the repressor for its operator within the probe. If the absolute magnitude of RFU or even normalised reaction balance was taken to infer a particular concentration, this metric would be easily fooled by varying levels of such salts. However, these ions seemed to rapidly affect equilibria rather than causing slow shifts, so looking at rate of changes following the initial 10 minutes instead of absolute values may be able to screen out these contaminants.

6.3 Integration of CgArsR-16S arsenite-insensitive mutant repressor

While this new analysis may be able to normalise the effects of particular solutes, or specific concentrations of solutes, it would likely be unable to guard against all, or all concentrations of solutes. Any other compound that affects the DSD reaction or repressor behaviour in a more progressive manner could manipulate the rates of RFU change away from accurately reflecting the underlying arsenite concentration. While the arsenite-binding domain of CgArsR is likely to be highly selective, integration with a DSD system means that anything affecting repressor-operator affinity or the DNA itself over a longer time period than the initial dilution adjustment would confound the assay.

Use of a second control repressor was identified at the start of the project as a potential method of addressing issues with assay robustness. Tetracycline repressors were suggested as good candidates due to their close relation to ArsRs but divergent operators and targets. However, since the adoption of parallel control DSD wells, which could be simultaneously filled in a microfluidic device, the needed for a control repressor with a distinct operator was removed, opening the door for a mutant ArsR instead.

In theory, if a mutant repressor could be created which maintained all the properties of wild-type CgArsR except for response to arsenite this would be a perfect control mechanism. Parallel systems using the two repressors could be compared and any deviations would be due to arsenite-specific activity, while signal increases seen in both would be put down to other non-specific solutes. However, designing such mutants and ensuring their lack of arsenite response and their equal affinity for the chosen operator are difficult tasks not certain to succeed. Any mutation would have structural consequences, potentially abolishing all DNA-binding capability let alone preserving a specific affinity.

One of the advantages of using the *Corynebacterium glutamicum* ArsR was the work with such mutants already carried out by Ordóñez *et al.* [130]. They assayed single, double and tripled cysteine-to-serine mutants for the residues within the arsenite-binding domain to establish their ligand sensitivity as well as any effects on operator affinity these mutations may have. Serine mutations seemed a good choice as the single atom change from sulfur to oxygen presented the minimal intervention hopefully capable of removing specific function while preserving residue size and polarity. Ordóñez *et al.* found that all these mutants resulted in loss of arsenite sensitivity, suggesting all three residues were required for ligand-induced dissociation, however they all suffered from reduced operator affinity. Estimates suggesting a dissociation constant increase of 3-10x depending on the mutation. The total loss of arsenite response was very promising, although the operator affinity reduction was unfortunate.

6.3.1 Characterisation of CgArsR-16S

Out of the three possible CgArsR cysteine-to-serine mutants, C16S was the one with the strongest evidence presented by Ordóñez *et al.* to be arsenite-insensitive. For this reason it was selected, alongside a preference to mutate one of the paired

C15/C16 residues over the lone C55 amino acid, to minimise potential folding mishaps. Primers were designed to carry out site-directed mutagenesis of the pET-28a-CgArsR expression plasmid and the new pET-28a-CgArsR-16S plasmid was sequence-verified before transformation into the *E. coli* expression strain, culture induction and protein purification.

The resulting CgArsR-16S samples were exchanged into the same storage buffer as CgArsR and matched to the wild-type A280 absorbance. A gel was run to confirm similarity in both concentration and purity, shown in Figure A.9. As can be seen, both glycerol stocks show a dominant band at the expected 15kDa molecular weight, alongside a weaker band at roughly double this mass. A couple of fainter bands appear in the WT lane and not the mutant, but for the most part these two preparations seemed very well matched. Following successful purification, the mutant repressor was compared to WT in a DSD assay - testing its ability to shift the reaction equilibrium, proxy for operator binding, and to respond to arsenite. The normalised data are shown in Figure 6.2.

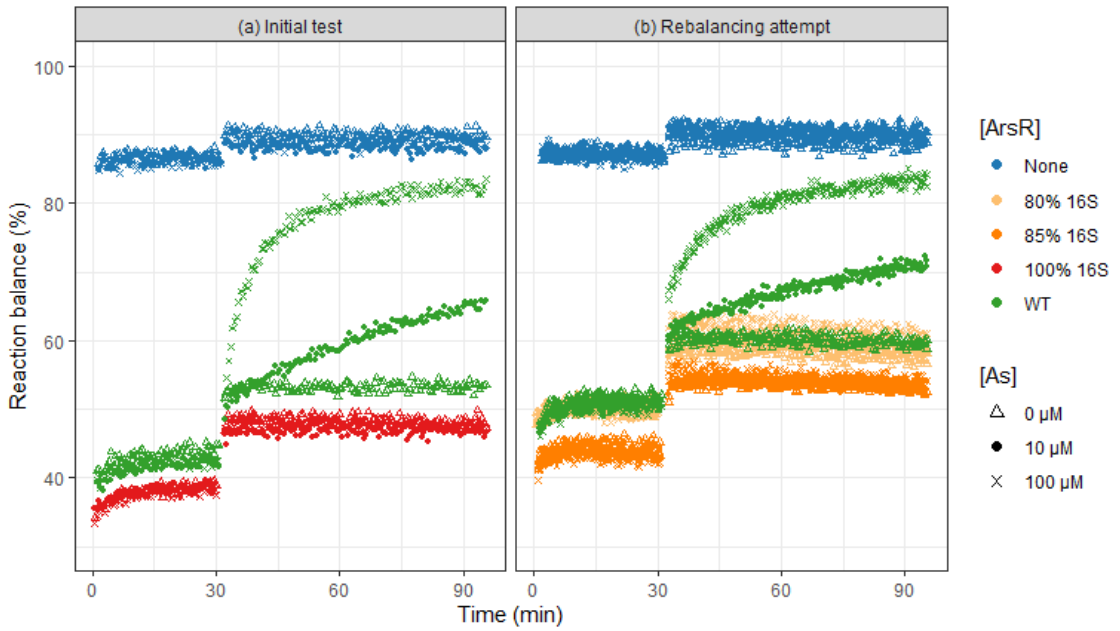


Figure 6.2: **CgArsR-16S is arsenite-insensitive but required dilution to balance operator affinity.** Fluorescence time course displaying data for different L6T1C preparations reacting with X20 under the DSD5 protocol in buffer DB6. (a) Normalised data for L6T1C reactions pre-incubated with equal volumes of wild-type CgArsR, mutant CgArsR-16S, or their common PSB. Different arsenite concentrations were added to each after 30 min. (b) Similar normalised data from a later experiment, although with a slightly different master mix formulation explaining the slight differences in equilibria. Here, smaller volumes of CgArsR-16S were added to the specified master mixes and made up to the consistent 'repressor volume' with PSB, thereby providing the stated percentages of CgArsR-16S relative to CgArsR. Raw data in Figure A.10.

Plot (a) shows the initial assay comparing samples containing equal volumes of the CgArsR and CgArsR-16S stocks, alongside a repressor-free control. The difference in reaction balance from the blue points down to the green in the first 30 minutes is indicative of the affinity CgArsR has for its operator in the L6T1C probe, while the separation of the triplicate wells following arsenite addition shows how the repressor dissociates at a rate depending on the concentration added, allowing the DSD reaction to shift closer to the repressor-free state. The red data are for the mutant repressor, and two features are of note. First, following arsenite addition the three wells do not separate at all, suggesting CgArsR-16S does not respond to arsenite as hoped. Second, the position of its equilibria below the green WT data, both before and after arsenite addition, suggests the mutant repressor is somehow able to shift the underlying DSD reaction more than CgArsR, contrary to expectations.

This scenario could be explained by a few hypotheses. On the surface, it suggests that the repressor-operator affinity is greater with CgArsR-16S than with CgArsR, however this would be contrary to the data presented by Ordóñez *et al.*. It is possible that instead, their affinities are almost identical and any difference seen is due to variance in repressor content added. Perhaps the seemingly slightly greater purity of the CgArsR-16S stock meant that for a fixed absorbance concentration more mutant repressor was added to its master mix than the parallel WT incubation. If the mutant repressor affinity was truly less than CgArsR an even larger disparity in amount added would be needed to cause this result.

An independent experiment confirmed the result in (a), so attempts were made to dilute the CgArsR-16S stock so that equal volumes of this and CgArsR could confer the same magnitude of equilibrium shift. A comparison of two such balance attempts is shown in plot (b), testing the addition of CgArsR-16S at 80% and 85% of the WT volumes. The master mix formulations were slightly different between the two assays to avoid problems encountered in the intervening period, but they were kept consistent within each experiment. Plot (b) shows how addition of 85% of the original volume still resulted in CgArsR-16S quenching the DSD reaction more than WT, but reducing this to 80% mimicked the base shift well, without responding to arsenite. The glycerol stock was therefore diluted in this ratio, so that future experiments could use equal volumes of each repressor.

6.3.2 CgArsR-16S normalisation exposes arsenite-induced signal increase for more consistent rate analysis

At the start of this section, the rate of RFU change was shown to correlate relatively well with arsenite concentration, enabling identification of sample content earlier than an endpoint analysis. However, for this method to work in the field, two features would be needed. First, accurate identification of the time required to adjust to solution dilution. The linear correlations were only possible if the first 10 minutes of readings were ignored, as this was the time taken for the water sample RFUs to peak, only identified after the assay. As unknown solutions cannot have a parallel arsenite-free control, it would not be possible to select times based on current data, so this time would have to be consistent for all tests and this may not be accurate for all solutions. Delaying the start of the analysed time period to be conservative would increase the total time required.

Second, it is unknown whether different solutions with the same arsenite content would display a consistent rate of RFU change. It would be hoped that non-specific contaminants like sodium and magnesium would cause rapid shifts in initial reaction balance as seen earlier, able to be screened out by delaying the start of the rate analysis, however it is very possible than smaller, more subtle readjustments to other solutes could take much longer than 10 minutes and thereby interfere with the rate measurements. Defining a relationship between arsenite concentration and rate of RFU change using only idealised solutions would be very unreliable.

The inclusion of the mutant repressor enables a proxy to be established for arsenite-free samples. This relies on the assumptions that all other solutes affect the WT and 16S CgArsR equally, and that unreacted arsenite does not then affect the 16S assay in any other way. Analysing the change in signal difference between these two parallel mixes could then highlight only the arsenite-induced effects, including normalisation of the initial dilution. With this subtraction a zero rate of change should be expected to meaningfully suggest presence of no arsenite. The relationship between actual arsenite concentration and rate of change of this difference would still need to be pre-calculated, but there should be greater confidence in its applicability across various solutions.

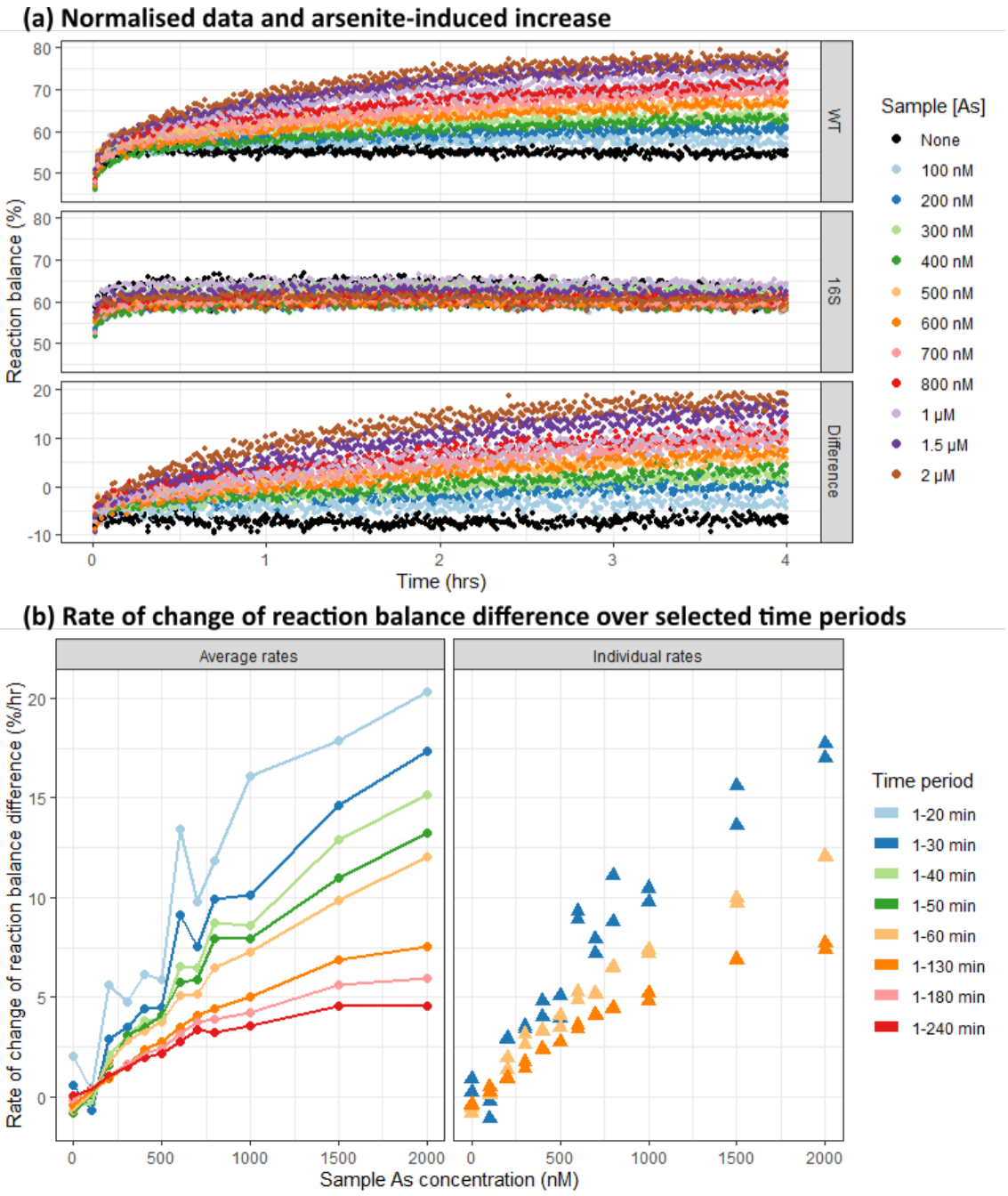


Figure 6.3: **CgArsR-16S acts as arsenite-free control and enables unbiased rate analysis.** Fluorescence time course displaying data for L6T1C-CgArsR DSD reactions responding to arsenite samples under the DSD7 protocol in buffer DB8. (a) Normalised data for the CgArsR solutions (WT, 16S and their difference) responding to these samples over 4 hours with points for individual wells and two replicates per arsenite concentration. The arsenite-induced signal increase was derived from subtraction of the average reaction balance of the CgArsR-16S solutions from each individual WT reaction balance, for each arsenite concentration. (b) Rate of change of the difference in reaction balance between the master mixes, calculated by creation of a linear model across different specified time periods, plotted by added arsenite concentration. The rates of individual wells for the 10-30, 10-60 and 10-130 min time periods are shown separately to showcase differences in variation. Raw data in Figure A.11.

An example of such an assay is shown in Figure 6.3. As explained in the methods, the DSD7 protocol made use of a full 96-well plate, with each column for a different arsenite solution added to two X20 'test' wells for both WT and 16S repressor, as well as polyT and X26 controls for each. This provided two replicates or normalised samples per repressor per arsenite sample. These are shown in plot (a) for the WT repressor and CgArsR-16S. Such a format allowed internal normalisation for each master mix/arsenite sample combination, as well as the paired repressor tests to be exposed to exactly the same original solution added using a multichannel pipette.

The important result of this bespoke assay format is the third plot in (a), displaying the difference in reaction balance between the paired samples. The non-zero starting point is the result of a slightly higher 16S equilibrium point than the WT repressor, but crucially this minor deviation does not affect the rate of change analysis, shown in (b). Unlike the assay in Figure 6.1 the first 10 minutes do not need to be ignored as the dilution effect is normalised out and the arsenite-free samples maintain a rate of change very close to 0 as desired. Being able to start the rate analysis from the first timepoint regardless of test sample removes bias and should enable much greater consistency between assays. As a result even without any arsenite-free control solutions the high arsenite concentrations should be rapidly discerned from others, while the correlation between rate and arsenite concentration will improve over time for more accurate quantification.

6.3.3 Arsenite signal correlation peaks after roughly two hours

Given this new method, one important parameter to establish was the overall best time period within which to analyse the data. Despite the 16S normalisation, is it still best to ignore the first few minutes? Is a longer assay always better? After how long can we start to classify unknown samples into qualitative arsenite concentration categories? To answer these questions the above dataset was analysed further, to calculate the rate of change of reaction balance difference for every sample, over every possible time period. For each window the correlation between these rates and the underlying arsenite concentration was assessed and displayed in Figure 6.4. As the higher concentrations clearly saturated, linear correlation was only investigated over the 0-800 nM range.

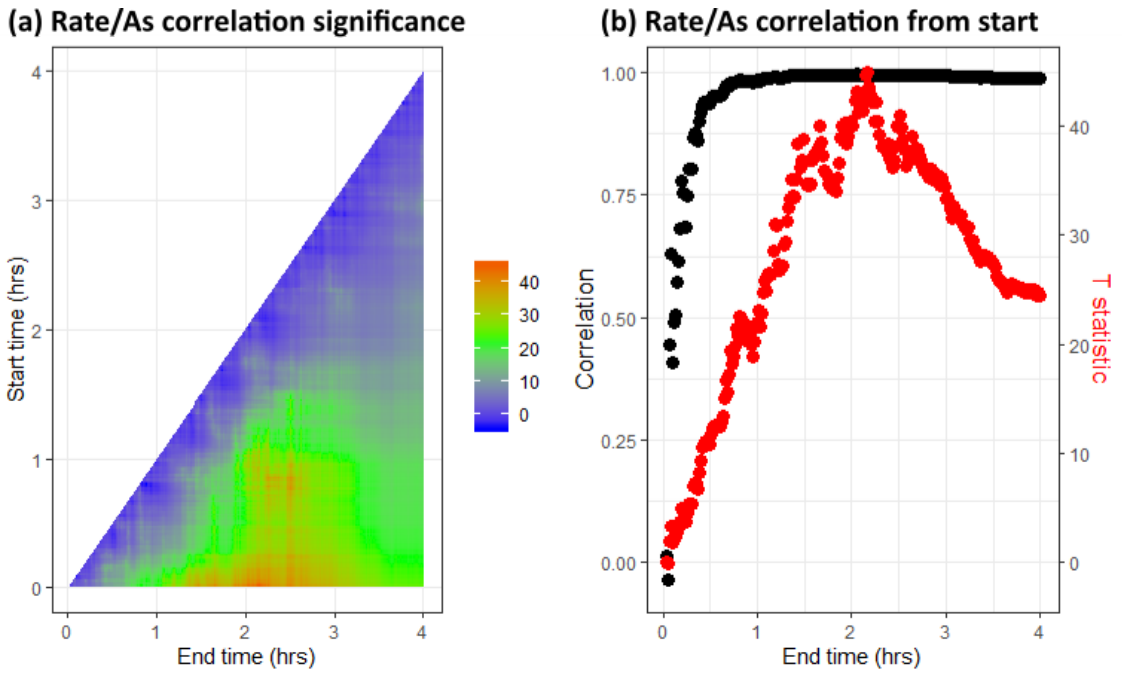


Figure 6.4: **Rate of difference correlates strongly with arsenite concentration.** (a) Heatmap displaying the significance of linear correlations made between arsenite concentration (for 0-800 nM samples) and rate of change of difference in reaction balance between WT and 16S samples, for every possible time period. (b) Comparison of all timeframe lengths starting from the first reading, showing both the calculated correlation coefficient and the significance of each correlation.

Plot (a) is a heatmap visualising the correlation significance between this rate and the arsenite concentration, over all combinations of one minute intervals. The very bottom line, therefore, represents windows beginning from the first read and ending at each later read time. Higher lines start later and so have fewer possible end times. Out of all these windows, the one with the greatest significance was the 1-130 minute period, with a Pearson correlation coefficient of 0.99605447 and a T statistic value of 44.8956549, giving a p-value of less than 2.2×10^{-16} - a highly significant linear correlation between the lower arsenite concentrations and their rates of signal change. The data for this window is shown in Figure 6.3 in orange instead of the default 1-120 minute period for this reason. The fact that inclusion of all the initial readings produced a more significant correlation than without them suggested that the 16S normalisation accurately removed the dilution artefact to expose only the underlying arsenite response.

Plot (b) shows the data for all windows starting with the first read in more detail. The estimated correlation coefficients are in black, with their significance

(the bottom line of (a)) overlaid in red. This reveals that while the correlation quickly climbs close to 1, correlation significance steadily increases until minute 130, after which it starts to decline. This suggests that the optimum assay time is around 2 hours, as beyond 130 minutes the correlation strength reduces so confidence in concentration estimation would lower.

Another important result from this analysis is identification of the times taken to pass various correlation thresholds. After 23 minutes the r coefficient first passes 0.90, after 32 minutes 0.95 and 70 minutes 0.99. This confirms that broad estimation of arsenite concentration is likely to be successful after only 20-30 minutes, while waiting an hour enables very precise evaluation.

6.4 Selectivity assessment

While the integration of the mutant repressor enabled a meaningful rate analysis, with good sensitivity and speed, the primary reason for its introduction was to improve assay selectivity. The biggest hurdle for *in vitro* assays tends to be dealing with other solutes - preventing them from either inducing a non-specific signal increase or from responding to their target as normal. Many are capable of detecting purified arsenite solutions but also respond to other ions or are sensitive to environmental variation that may confound quantification. By comparing the signal from the wild-type CgArsR reactions to those with the 16S mutant, it was hoped high selectivity could be achieved.

6.4.1 Non-arsenic compounds can produce false positive rates through 16S signal loss, but have distinct dynamics

In other systems arsenic specificity is typically demonstrated by assaying a selection of transition metal solutions at a chosen concentration and providing corresponding negative response data. While this is useful, it rarely tells the full picture. Arsenic, as a group 15 metalloid, is chemically most similar to phosphorus and antimony, preferentially forming negatively charged oxyanions in solution over positive cations. Phosphite and antimonite ions, therefore, are more likely to produce false positive readings but, depending on the biosensory transduction element, various metal cations can too through chemically

divergent mechanisms, and these are more common in groundwater.

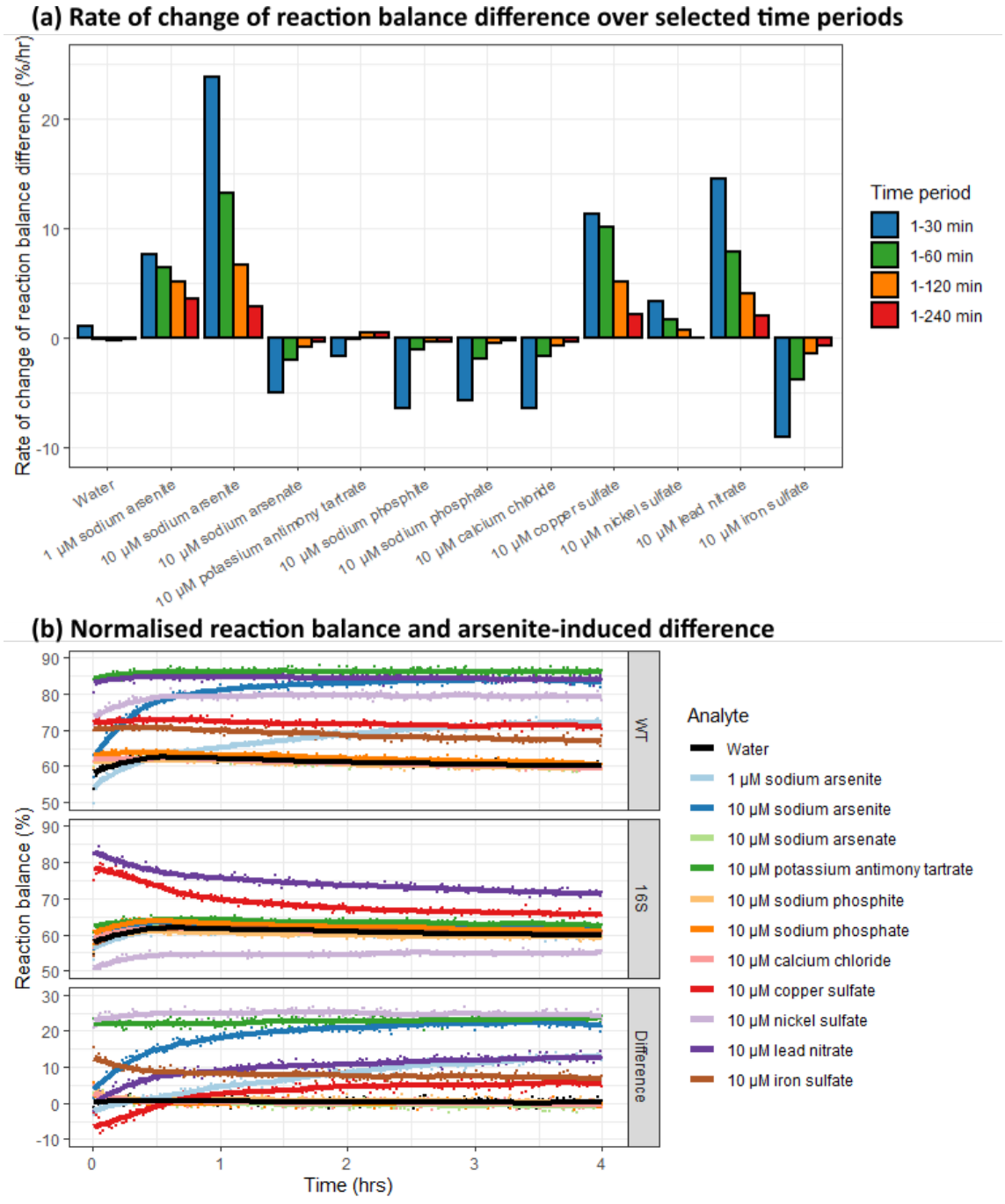


Figure 6.5: **Various compounds induce WT/16S differences but are distinguishable from arsenite's signature.** Fluorescence time course displaying data for L6T1C-X20-CgArsR master mixes reacting with different analyte samples under the DSD7 protocol in buffer DB8. (a) Rate of change of the difference in reaction balance between the master mixes, calculated by creation of a linear model across different specified time periods, plotted by analyte. (b) Normalised reaction balances for WT and 16S mixes and their difference, responding to the samples over 4 hours. Points are displayed for two replicates per analyte, overlaid by a smoothed average line. Raw data are shown in Figure A.12.

Figure 6.5 displays the results of an assay testing response to a selection of analytes at 10 µM. Plot (a) picks out the difference rate metric over four time periods for each, while plot (b) presents the full time course normalised data for both repressors and their difference. In (a) the arsenite response is clearly seen at both concentrations assayed, but false positive readings are seemingly given by copper sulfate, nickel sulfate and lead nitrate. On the other hand, sodium arsenite, potassium antimony tartrate (PAT - the most stable Sb(3+) compound), sodium phosphite, sodium phosphate, calcium chloride and iron sulfate seem to produce no response or even a slight negative signal.

These mixed results reflect an unexpectedly complex range of dynamics, seen in (b). The small positive reading from nickel sulfate derives from a surprisingly large difference between repressor equilibria, with the WT CgArsR shifted roughly 20% higher and the 16S repressor 5% lower. These initial alterations are captured by the early rate analysis, but then stay largely parallel so the difference rate substantially reduces with longer time periods. The copper and lead compounds, however, display much greater rates of change, but not from a steady increase in the WT signal as for arsenite, but from stably shifted WT signals against a steadily decreasing 16S response.

Non-arsenic compounds were expected to alter the reaction balances of the parallel repressors, either to similar or different magnitudes, but this dynamic 16S response was very surprising. This suggests that copper and lead non-specifically dissociated both repressors very quickly, but then the 16S repressor slowly rebound its operator while the WT CgArsR did not. This presents interesting hypotheses about the mechanism of these interactions and whether the mutant 16S arsenite-binding pocket inadvertently alters other affinities. These false positive readings, while able to confound a simple look at the reaction balance rate of change, could be easily removed with automated checks on the fuller dynamics. True positive signals should derive from sustained increase in the WT response until saturation, compared to a more stable 16S response, while an early drop in 16S could be flagged as false.

Calcium chloride and the sodium arsenate, phosphite and phosphate compounds minimally affected either repressor, in line with desired non-arsenite responses. All caused a small early drop in difference rate before settling in line with the water effect. Once again, such short-lived effects could be identified by scanning moving time periods - early changes that settle are more likely to be other ions

while arsenite responses tail off much more slowly unless the signal saturates.

The antimony response was also close to expectation. Ordóñez *et al.* suggested that PAT had a higher affinity for CgArsR than arsenite, perhaps as antimony is less easily taken up *in vivo*. 10 µM PAT had minimal effect on CgArsR-16S, but saturated WT signal almost immediately, whereas the same concentration of arsenite took roughly two hours. While this resulted in a negligible difference rate and therefore no false positive signal, it suggested that lower concentrations may be more problematic. This reinforces the need to test a spectrum of environmentally relevant concentrations for each analyte.

6.4.2 Antimony response too quick for false positives while EDTA boosts specificity

Lower antimony concentrations were tested in Figure 6.6. Surprisingly even these saturated the WT signal within the first minute, resulting in near-zero difference rates. The 100 nM concentration, with two antimony atoms per PAT complex, gave only a marginal final excess over CgArsR concentration, suggesting lower concentrations would likely increase signal as quickly but at lower equilibria. This also indicates an incredibly high antimony affinity by the WT repressor, completely abolished by the 16S mutation. While this removes the likelihood of false positive signals, one concern would be either completely lost or greatly limited sensitivity in mixed antimony/arsenic solutions. Natural antimonite prevalence in groundwater is low, however, at under 1ppb in most regions away from mines or industry, so perhaps this is not a crucial flaw but certainly warrants further investigation [161].

This effect was consistently replicated with nickel sulfate, albeit with both repressors at slightly lower equilibria. As the resulting maximum equilibrium difference was preserved this indicated both high specificity and affinity for the WT CgArsR as well as a more subtle effect on the underlying DSD reaction itself. Iron sulfate also caused a specific effect on the WT repressor, but to a much smaller degree over both concentrations assayed. Copper and lead previously generated false positive responses through 16S signal loss but at the lower concentration this was only replicated with copper, with 1 µM lead nitrate instead causing maximal dissociation with both repressors.

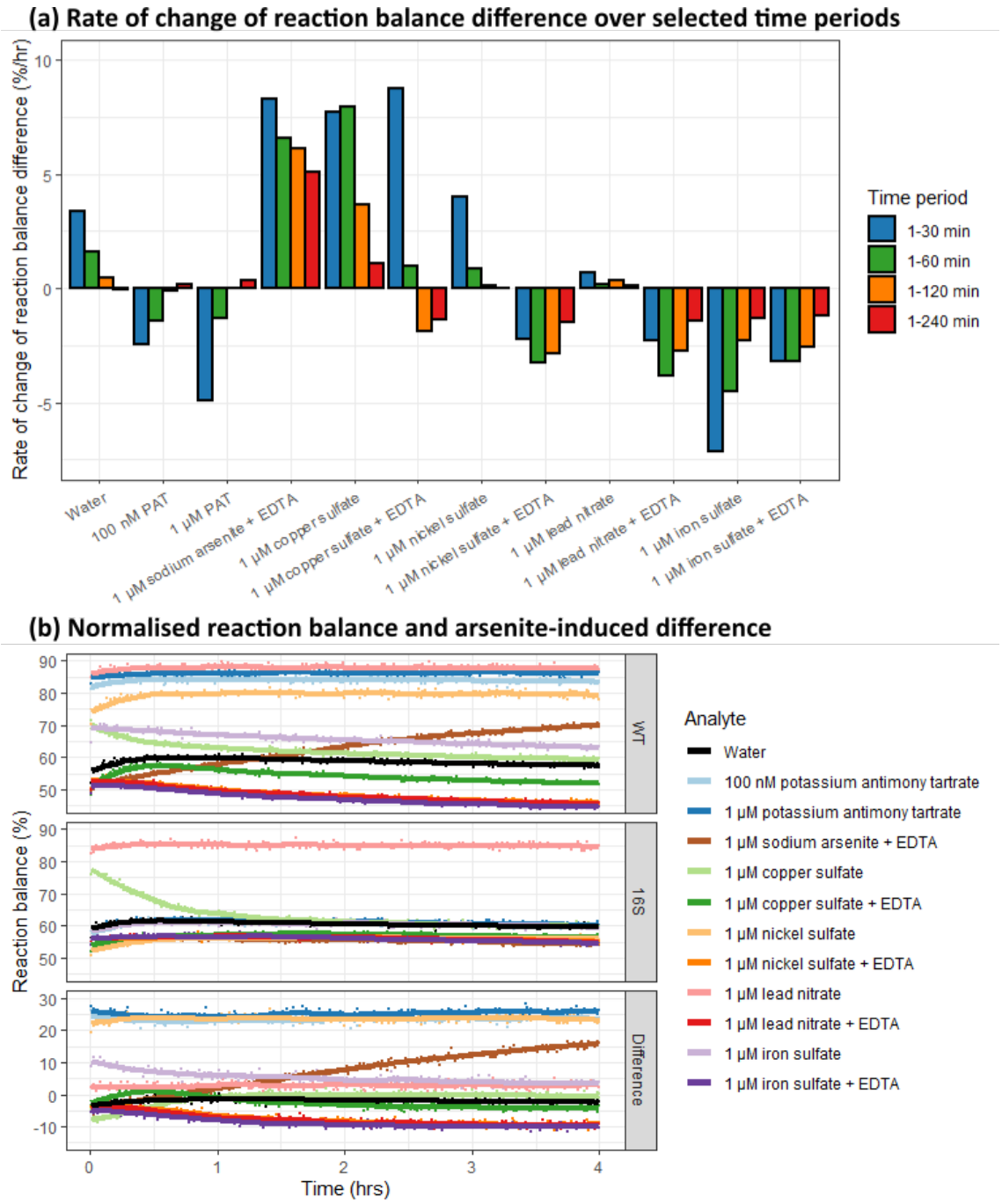


Figure 6.6: **Lower antimony concentrations distinct from arsenite while EDTA removes other non-specific effects.** Fluorescence time course displaying data for L6T1C-X20-CgArsR master mixes reacting with different analyte samples under the DSD7 protocol in buffer DB8. (a) Rate of change of the difference in reaction balance between the master mixes, calculated by creation of a linear model across different specified time periods, plotted by analyte. (b) Normalised reaction balances for WT and 16S mixes and their difference, responding to the samples over 4 hours. Points are displayed for two replicates per analyte, overlaid by a smoothed average line. Raw data are shown in Figure A.13.

Simultaneously tested was the effect of EDTA (ethylenediaminetetraacetic acid) addition. Chelating agents were deliberately avoided in all buffers until now, due to their ability to bind arsenite, a property that has been exploited in treatment of arsenic poisoning for many years [162]. EDTA itself, however, has a strong preference for multivalent cations and any potential benefit for arsenicosis patients is likely due to effects on calcium transport [163]. Some studies have suggested EDTA improves bioavailability of arsenic and Siegfried *et al.* claimed addition reduced the rate of false positives for their ARSOlux device.

Figure 6.6 shows pre-incubation with 100 µM EDTA had a dramatic effect on response to the metal compounds. Equilibria shifts caused by nickel, lead and iron were completely removed, replaced by a slow decrease below the water only level. Removal of magnesium would be expected to shift the DSD equilibria in the opposite direction (Figure 5.7), so EDTA may have affected the repressors instead. Almost the same effect was seen with copper but an initial WT rise was maintained before signal reduction hence its dramatic rate reversal.

Crucially, in addition to reducing cation effects, arsenite response was maintained with EDTA addition. The sizeable concentration of magnesium in the master mix may well prevent arsenite sequestration, while EDTA's higher affinity for the transition metals may cause their preferential chelation. For these experiments EDTA was pre-incubated with the analytes; further work will investigate the possibility of including EDTA in the master mix itself, which may remove the initial negative difference rate seen with these samples.

6.4.3 Final system is robust and has high selective potential

A summary of these selectivity tests is presented in Figure 6.7, which compares the difference rates over the first hour of each assay. It includes a further experiment investigating the effect of many of the same compounds at 1 mM concentration, to determine if the system could tolerate high levels of these ions. The raw and processed data for this are in Figures A.14 and A.15. Most compounds generated similar responses as before, but nickel, lead and iron all strongly affected the underlying DSD reaction at these concentrations.

The effect of 1 mM iron was most striking, causing total loss of fluorescence. As

this happened to the control as well as the test wells, normalisation produced meaningless noise data with extremely high variation, which could be easily identified and flagged as contamination. High nickel produced more subtle non-specific fluorescence loss, while lead initially caused strong quenching that progressively lessened over the course of the assay. Both normalised to relatively sensible, albeit low, reaction balance equilibria, whose effects could be removed by ignoring early time periods. For more sensitive contaminant detection, control RFU values outside expected ranges could be instantly flagged.

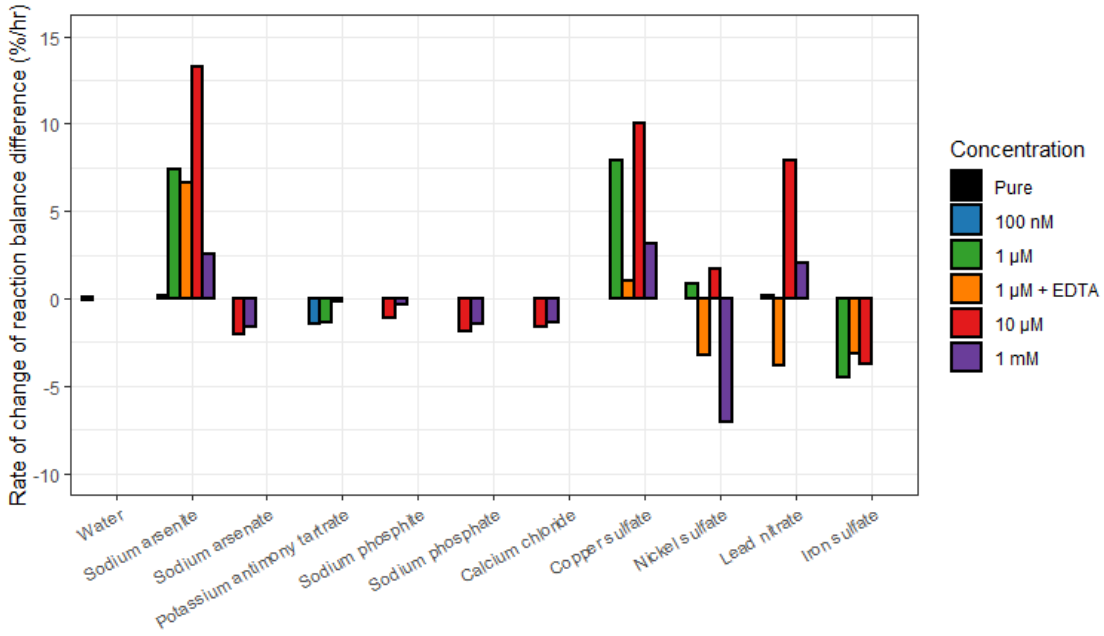


Figure 6.7: **Selectivity summary.** Compilation of the three selectivity assays from Figures 6.5, 6.6 and A.15, comparing rate of change of reaction balance difference over the 1-60 min time period for the different analytes and concentrations, including some with EDTA. The 100 nM sodium arsenite rate is from Figure 6.3, where the rate picked up over longer time periods. While this provides a general overview, it massively simplifies the spread of underlying dynamics and does not qualify the inferred rates with any confidence metric of the assumed linearity.

According to a BGS analysis of Bangladesh groundwater by Kinniburgh *et al.*, the only ion apart from sodium and magnesium likely to be found at millimolar concentrations was calcium, but across the different levels tested this had a negligible effect on the system. Likewise, out of the 3530 wells tested 95% contained 16 mM or less sodium. This concentration was assayed in Figure A.15 and resulted in the expected shift to both repressor equilibria and a minimal difference rate. These two ions were the most variable in the BGS analysis so these minimal effects suggested the system would be robust to their fluctuations. Effects and solutions to all these ions are summarised in Table 6.1.

Compound	Groundwater content		Assay response	Potential problems	Suggested resolutions
	50% level	95% level			
Sodium **arsenate**	<0.735 µM	<4.07 µM	No response from either reaction	Lack of detection	Arsenate reductase integration
Potassium **antimony** tartrate	<18.5 nM*	<47.8 nM*	Rapid WT repressor shift; 16S insensitive	Signal saturation preventing arsenite detection in mixed samples	Likely minimal bioavailability of Sb(3+). Other Sb species less likely to interfere
Sodium **phosphite**	<23.6 µM	<90.4 µM	Slight WT shift at high concentration	Slight loss in arsenite detection range	Likely minimal bioavailability of $[HPO_3]^{2-}$
Sodium **phosphate**	<23.6 µM	<90.4 µM	No response from either reaction		
Sodium chloride	3.87 mM	16.0 mM	Shift in both repressors at high concentration	Potential slight reduction in arsenite detection range	16S difference and rate analysis removes effect
Calcium chloride	1.29 mM	3.47 mM	No response from either reaction		
Nickel sulfate	0.137 µM*	0.705 µM*	Rapid WT shift; 16S insensitive. Small DSD down shift. Steady quenching at [high]	Limited arsenite detection in mixed samples	Low natural prevalence; EDTA inclusion in MM
Copper sulfate	0.012 µM	0.126 µM	Rapid shift up in WT+16S; subsequent 16S drop	False positive signal from 16S signal reduction	Low natural prevalence; EDTA inclusion in MM
Lead nitrate	4.83 nM*	19.8 nM*	Rapid shift up in WT+16S; subsequent 16S drop. Dynamic quenching at [high]	False positive signal from 16S signal reduction	Low natural prevalence; EDTA inclusion in MM
Iron sulfate	60.0 µM	245 µM	WT shift at [medium]; total non-specific signal loss at [high]	Reduced range at [medium]; no arsenite detection at [high]	Medium natural prevalence; EDTA inclusion in MM

Table 6.1: **Summary of selectivity effects and possible resolutions.** Comparison of the differing responses of the DSD7 assay to various compounds, problems their presence can cause and how these can be rectified, summarising Figures 6.5, 6.6 and A.15. The ion in each compound being primarily assessed is in bold, with estimated concentrations in Bangladesh groundwater taken from the 2001 BGS survey where available [67], or (*) from a separate analysis of Turkish groundwater [164]. Total arsenic, antimony and phosphorus content was quantified so concentrations represent upper limits for each oxyanion species.

While EDTA treatment looks highly promising for removing contaminant effects, an additional possibility would be to build in more detailed automated data analysis. Further to examining control RFUs, initial reaction balance changes could be excluded from the rate analysis by examining the linear model residuals across different time periods. Most contaminants seemed to present a rapid loss of linearity, indicative of almost oscillatory establishment of new DSD equilibria through non-specific interactions, while arsenite responses were consistently more stable. Detecting fast early changes could suggest ignoring early time periods and only starting attempts at arsenic quantification later once the baseline is established, leading to more robust results.

6.5 Lyophilisation

Freeze-drying is a widely-used technique to both improve longevity and reduce weight of goods. Most famous applications include astronaut meals and instant coffee, however pharmaceutical products are increasingly using this approach. Live cells and other organic material commonly require cold-chain transport and storage to preserve behaviour and limit degradation, but if successfully lyophilised can often be kept at room temperature for much longer, reducing expense and increasing the number of potential customers.

For these reasons, lyophilisation is highly desired for point-of-care diagnostics [165]. Many rural communities would struggle to store tests at low temperatures and the cost of regular transport would be prohibitive. If diagnostics can be provided in bulk and stored at ambient temperatures for months or years without loss of function then cost can be greatly reduced. Diagnostics in the form of test strips/sticks that only require addition of a solution, such as litmus or pregnancy tests, are also usually simple to handle which improves usability.

While this plate reader assay does not represent a finished diagnostic, demonstration of its ability to withstand lyophilisation is highly advantageous. Many cell-free or *in vitro* solutions struggle to lyophilise successfully, either losing performance or failing completely, as particular components denature in this process. Addition of cryopreservative agents can help, but not always, and they too can reduce assay functionality. Once again the minimal nature of this repressor-operator DSD system is beneficial as there are fewer species at risk of failing.

6.5.1 Lyophilisation increases assay variability, but functionality is preserved

Figure 6.8 presents normalised data from a lyophilised version of the standard DSD7 assay, responding to a spectrum of arsenite concentrations. The overall dynamics were very similar to those in the fresh setup, demonstrating the ability of the system to withstand this treatment and rapidly reactivate upon rehydration. Another benefit of lyophilisation is a theoretical improvement in sensitivity. Drying of an example plate reduced mass, from 64.767 g to 60.452 g, equivalent to 44.948 mg from each of the 96 wells. This enabled 245 µl arsenite solution to be added per well, rather than the usual 200 µl, to reach the same final volume and component concentrations, boosting potential sensitivity.

There were, however, a few minor differences. The raw data in Figure A.16 shows how the maximum fluorescence of the lyophilised system was about 130k, down from 150-160k previously. In addition the maximum control signal, which usually exhibits a steady decrease, was instead stable, suggesting lyophilisation either prevented or completed the primary effect responsible for this behaviour. All the samples also displayed greater than usual variation, seen most clearly in the maximum controls and the raw and normalised 16S data. As this noise dissipated over time this was not the result of permanent damage to components but temporarily altered mixing or behaviour. This increased initial variation reduced the reliability of early arsenite quantification, which may have been improved by starting once the noise decreased. Apart from anomalous behaviour of the 200 nM sample, however, the concentration-rate relationship remained very robust.

Successful freeze-drying was dependant on many factors. For this system a short, fast freeze at -80 °C was identified as important to ensure quick reactivation; slower cooling resulted in highly quenched initial states that were sluggish to respond. Also, the plates were sealed with a film during lyophilisation to limit contamination of open wells, with three small holes in each to enable drying. The size of these holes was unexpectedly important - too small and drying was incomplete even after an overnight incubation. Complete lyophilisation produced plates as in Figure A.17, with an even white sheet at the bottom of each well; other phenotypes indicated either uneven or incomplete drying, or insufficient pre-freezing. Further work will investigate how long these lyophilised plates can be kept before losing functionality, and whether

room temperature storage is possible. Cryopreservative agents, in addition to the master mix PEG, will also be tested to reduce initial variation.

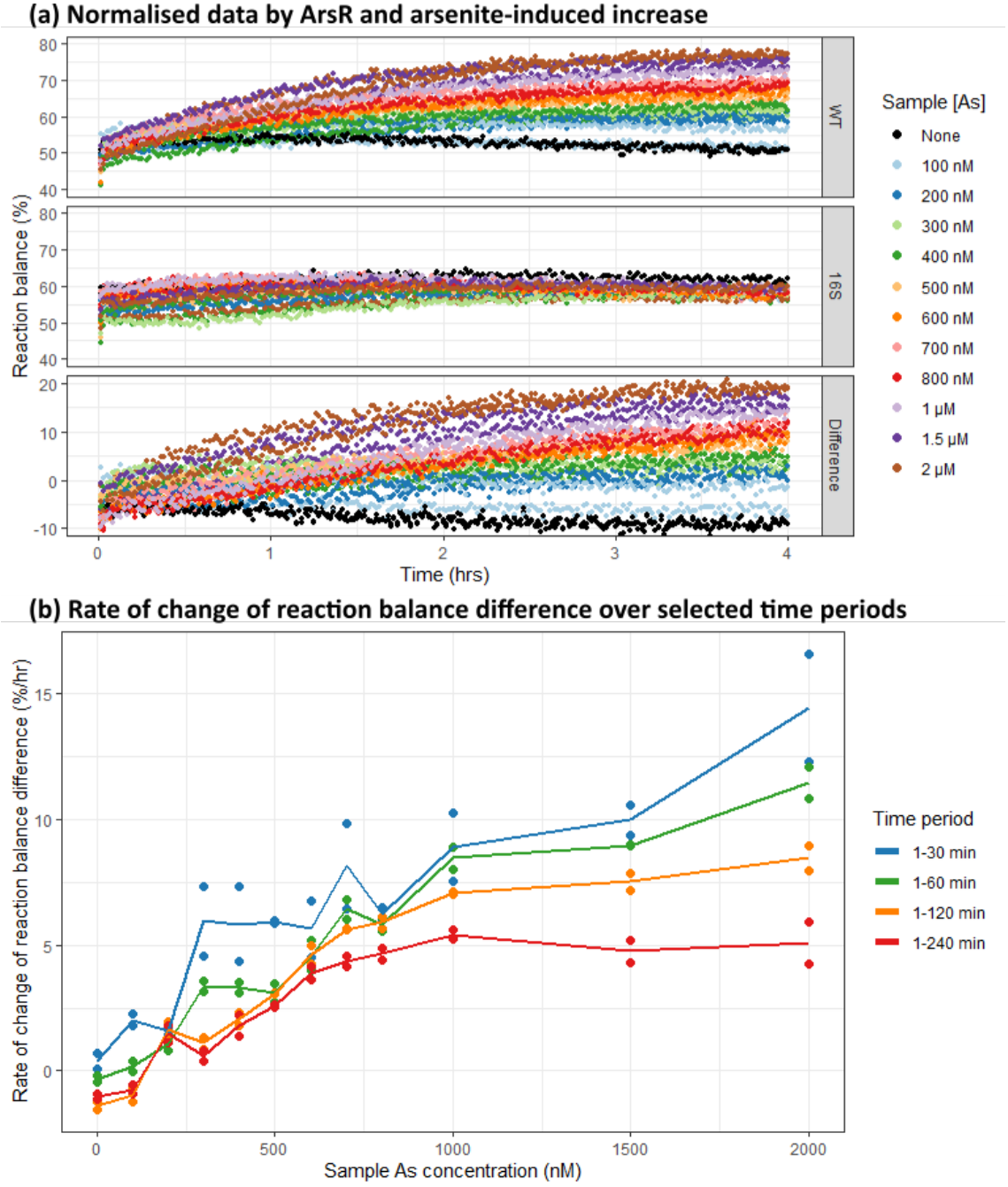


Figure 6.8: **Initial variability increase but arsenite detection correlation maintained.** Fluorescence time course displaying data for lyophilised L6T1C-CgArsR DSD reactions responding to arsenite samples under the DSD7 protocol in buffer DB8. (a) Normalised data for the CgArsR solutions (WT, 16S and their difference) responding to these samples over 4 hours with points for individual wells and two replicates per arsenite concentration. (b) Rate of change of the reaction balance difference, calculated by creation of a linear model across different specified time periods, plotted by added arsenite concentration. The rates of both individual wells are shown as points with their average as lines.

Chapter 7

Discussion

Substantial progress has been made towards the development of a functional arsenic biosensor during this project. While the repressor-DSD system presented here is not a finished device, the research conducted lays the foundation for this possibility. In its current format the assay surpassed targets for arsenite sensitivity and detection speed, identifying 100 nM/7.5ppb arsenite within an hour, and displayed a promising selectivity profile.

Integration of aTFs with DSD is a novel method of manipulating reaction dynamics and equilibria. Successful implementation enables applications in both biosensing as well as more general synthetic DNA computation, but required several notable developments. A mutant operator sequence was designed to provide a structureless, high-affinity DNA probe with an architecture enabling comparison of alternative sequences while maintaining a consistent fluorophore environment.

This was combined with a bespoke normalisation methodology for meaningful, reliable signal conversion to concentration or reaction balance metrics, allowing optimisation of both speed and sensitivity of arsenic quantification. Finally, integration of a parallel arsenite-insensitive mutant repressor and a differential rate analysis enabled high sensitivity and robust selectivity.

7.1 Arsenic biosensing potential

The number and efficacy of the different controls are perhaps this system's greatest strength, leading to reliable and reproducible arsenite quantification. Sensitivity and detection speed are also good, however there are areas in which this assay would need to improve to develop into a practical system.

An ideal arsenic biosensor would be able to detect all species capable of contributing to arsenic poisoning, not only arsenite. This is a clear limitation of depending on a single aTF with a particular analyte chemistry, but could still be achieved. Arsenic operons naturally include reductase enzymes to convert arsenate to arsenite, enabling detection of both, and such an enzyme or a chemical alternative could do the same for this *in vitro* system. The same ArsR seems to produce a specific response to alternative ions such as antimony and nickel, but the limited bioavailability of the former and chelation of the latter should enable robust performance. While quantification of arsenic in real groundwater samples was untested, the integrated controls and rate analysis instil confidence in this potential.

Practical considerations are crucial to bear in mind during research to give a system the best chance of success. The ASSURED (and subsequent REASSURED - real-time connectivity, ease of specimen collection, affordability, sensitivity, specificity, user friendliness, rapid and robust, equipment-free, deliverable to end users) criteria have informed guidelines for development of infectious disease diagnostics for many years, and the same ideas can be applied to POC devices for environmental sampling [166]. The only principles so far not directly addressed are those concerning cost and device usability. The minimal assay composition, not requiring any cell-free extracts or enzyme substrates, was a deliberate design feature to reduce expense as well as simplify dynamics. At the scales used for this work, the modified oligos cost roughly 3 pence per 8-well controlled reaction set required for analysis of a single sample, while unmodified oligos were a couple of orders of magnitude below this. Reagent and purified protein costs would likely take this up to 5-10 pence, but could well be noticeably reduced with scaled production.

By far the greatest expense for this system would be the physical device itself. Samples would need to be split into eight equal volume channels, likely using a microfluidic setup, before fluorescence excitation and emission detection. Signal

analysis plus reading display would require further electronics and hardware, especially with automated sample filtration, all more expensive than the assay itself. A reasonable format could use a permanent reader consolidating most of this cost into a one-off device, into which cheap disposable cartridges containing the lyophilised assay could be inserted.

While this results in great uncertainty over the cost and usability of an eventual device, these hurdles are similar to other contemporary systems. The assay itself is fundamentally inexpensive by comparison to many, and further design iterations could replace requirements for both fluorescence detection and fixed readers, the greatest sources of expense.

7.2 Comparison to other *in vitro* biosensors

To fully appreciate the strengths and weaknesses of the assay presented here, it is important to compare its properties to other published systems. The most advanced *in vitro* arsenic biosensor discussed in the introduction was the AquAffirm device based on work carried out by Santini *et al.*, replacing GOx in glucose sensors with an arsenite oxidase enzyme. This system claimed sufficient sensitivity, but its main strengths were a wide dynamic sensing range and an extremely fast detection speed, quantifying sample arsenite concentration in less than a minute. This latter property, in particular, is substantially better than the DSD assay presented here, but there are significant questions over AquAffirm's selectivity and tolerance of environmental variation.

Very limited data exist testing these properties, largely early work by Male *et al.* [101]. Good selectivity was claimed, but the assayed used bottled and river water samples containing only very low solute concentrations, far lower than expected in Bangladesh groundwater for most [67]. In particular, the large variation in sodium content could be expected to influence Aio enzymatic activity, thereby increasing the noise in an already noisy system. Male *et al.* tested only a maximum of 12ppb sodium, far less than the average 89ppb and 368ppb 95% levels, while greater cation concentrations could interfere with the complex electron transfer chain. The lack of methods to control effects of environmental variation on oxidase activity may explain why AquAffirm has still not reached the market, 8 years after patent approval. In addition, no antimony species were tested. As antimonite is so chemically similar to arsenite

there is a reasonable chance these ions could be similarly oxidised by Aio and produce a strong false positive signal. The repressor-DSD system, however, was shown to rapidly saturate in the presence of antimonite and while this would limit arsenite detection in mixed samples it would not produce false positives, an important advantage. AquAffirm is also inherently unable to detect arsenate alongside arsenite and would require a parallel Arr reaction to achieve detection of total inorganic arsenic content, while a reductase enzyme could be introduced to the repressor-DSD assay [103].

Since the inception of this project, a couple of other small molecule *in vitro* detection systems have been published and while they did not directly test for arsenic, their use of repressors as transduction element suggests they could be adapted for this analyte and serve as useful performance benchmarks. Crucially, neither embrace the reversible nature of aTF binding and so rely on suppressing background signal. The vulnerability of such systems to non-specific dissociation causing false positive signal accumulation stands out, and highlights the advantage the repressor-DSD assay presents.

ROSALIND is an IVT system, using aTFs to control transcription of the Broccoli RNA aptamer [167]. In the presence of the DFHBI-IT dye, the transcribed RNA forms a fluorescent complex that can be detected following ligand-induced TF dissociation. This system offers natural signal amplification and use of a transcription-based reporter greatly speeds response compared to a full TXTL method. Perhaps ROSALIND's most useful property is its modularity, with only a switch in dsDNA operator sequence and purified repressor required to sense new analytes. Identification of a high-affinity operator with structureless ssDNA would likely be a design bottleneck for this DSD system, and is not guaranteed to succeed, so ROSALIND's simplicity is a clear advantage.

While Jung *et al.* did not create an ArsR assay, they designed biosensors for 16 different organic molecules and metal ions. The cation sensors all had maximal sensitivities of over 1 µM after four hour incubations at 37 °C, both slower and less sensitive than this DSD system. Innovative feedback loop and logic gate circuits may improve sensitivity and selectivity, while altering aTF content could tune the response. The authors did suggest ROSALIND was best implemented as a thresholding device due to its limited dynamic range, however, and so would struggle to perform wide quantification. Jung *et al.* did demonstrate reasonable orthogonality between the 16 assays, but reliability did

look problematic. The ROSALIND copper sensor produced a greater signal for 139ppb in one river sample than 366ppb in another and some fluorescence inhibition was detected, suggesting the assay was not robust against environmental variation. Any solutes affecting aTF-operator affinity non-specifically could influence signal, while RNAP transcription rate could also be variable. This is a notable disadvantage of any such system relying on signal accumulation rather than signal shifts, as with this DSD assay. Integration of mutant repressors could reduce some of these problems, but would be unlikely to solve them.

Dependency on an RNA output and a 37 °C incubation is also problematic. The former makes the system vulnerable to contamination problems while the latter increases the expense and reduces the practicality of a device. The DSD system here performed assays at 25 °C instead and preliminary data suggested minor temperature fluctuations were well tolerated by the controls, so may not even require isothermal incubation. Signal normalisation to a fluorescein standard may also be unreliable for the ROSALIND system, as the intensity of this fluorophore is unlikely to be proportional to the Broccoli-dye complex in all conditions, highlighting another advantage of the DSD approach.

Another innovative avenue was taken by Hsing *et al.*, who developed a different biosensor based on competition between transcription factors and HgaI, a type IIS restriction enzyme [168]. The endonuclease was prevented from binding its adjacent recognition sequence with an aTF excess, but following ligand-induced dissociation the endonuclease could cut the aTF operator, releasing a DNA duplex with a 5nt ssDNA toehold. This separation enabled cyclical TMSD-dependant signal amplification, harnessing further HgaI activity to separate a fluorophore and quencher. This system claimed greater speed and sensitivity than ROSALIND's tetracycline biosensor, with the latter similar to the arsenite sensitivity shown here. Hsing *et al.* did minimal selectivity testing, however, and while their assays were orthogonal to other antibiotics, signal increase was notably reduced in environmental water samples, which would greatly limit their ability to quantify ligands reliably.

The HgaI assay represents an intermediate approach between ROSALIND and this project's system, making use of fluorophore-modified oligo reporting and DSD reactions, but still producing an irreversible, cumulative output. As a result its strengths (strong amplification, simple design modularity, good speed

and sensitivity) and weaknesses (lack of signal control, environmental variability, possible requirement to remove ssDNA operator structure) are also intermediate combinations. These first two shortcomings are shared by both alternate TF-dependant systems and are critically important to reliable water sampling, as well as hard problems to solve for these designs. For this reason the presented reversible DSD assay is advantageous and while its slower design cycle may limit development speed, it may have the potential for a higher quality product.

7.3 Future assay development

There are a number of areas in which the current composition could likely be improved. Several of these have already been outlined - integration of EDTA and cryopreservative agents into the reaction master mix, for example, as well as a reductase enzyme for possible arsenate detection. Use of purified probe complexes would also improve sensitivity - currently one third of the probe complexes with dsDNA ArsR operators did not contain a fluorophore, leading to many undetected displacement events. Omitting this step enabled much faster experimental turnover, but would certainly be included in a final device.

Composition optimisation through accurate mechanistic model simulation remains the technique likely to most improve assay performance. Finer tuning of repressor excess may be possible manually, but doing so in combination with DSD components would be much harder. Testing permutations *in silico* would be faster and more precise once an accurate model is achieved. Figure 5.3 also highlighted how single point mutations could substantially affect sensitivity, dynamic range and response speed. Further characterisation of different operator sequences could offer alternatives better suited for each, and combining manipulations in operator selection and repressor excess could tune the desired dynamics with high precision. A complete model could then be used to characterise the binding and unbinding rates of both aTF to its operator as well as for ligand to aTF. If done for a wide operator selection, the resulting parameter space could be explored *in silico* to optimise combinations for particular holistic features.

In the longer term, more ambitious changes could be tested. The initial choice of the *Corynebacterium glutamicum* ArsR was largely made through failure to

express other stable ArsRs in sufficient quantities. Early data suggested BsArsR may have been inherently more sensitive to arsenite and could be tested further if stable buffers were found. Many other ArsRs exist with uncharacterised properties, some of which are likely to outperform CgArsR in at least one of the metrics. An interesting possibility is the use of a thermophile ArsR - as they have evolved to function at high temperatures their operator affinities may be artificially high at room temperature [169]. Two of these were successfully expressed recently, but identification of their respective operator sequences would be necessary before full characterisation. Other ArsRs such as from *Chromobacterium violaceum* may also enable distinct dynamics if a thresholding response was desired [109].

Alternative reporting system may improve sensitivity as well as substantially reduce costs. Methylene blue-modified probes could enable an electrochemical instead of fluorescent output, and have been successfully integrated with DSD reactions for high sensitivity oligo detection [170, 171]. Using tethered oligos and even tethering the repressor itself could be used to increase local concentrations and therefore enable higher reaction speeds and sensitivities [172]. With an increased linear dynamic range and a cheap, easy-to-use physical device, this system could be developed into a legitimate arsenic biosensor.

7.4 Wider implications

Developing a viable arsenic biosensor was not an original goal for this project, but considering the requirements of such a detector provided a useful framework to explore the conceptualised system and guide progress. Functional implementation was therefore a minor result compared to the overall successful proof of principle demonstration that DSD reactions can be modulated by allosteric transcription factors. The ability to transform a DNA-only network into a system capable of quantifying presence of analytes other than nucleic acids with the addition of a single protein is novel and powerful.

As with the ROSALIND and HgaI systems, this suggests the reversible DSD assay could be adapted to other targets through the inclusion of alternative aTFs and their operators. It was incredibly important therefore, to develop techniques and identify design considerations that would be useful in a generic bioengineering pipeline for new analytes.

The adoption of polyT and irreversibly displacing oligos to act as continuous controls for signal normalisation was very significant in this project and would certainly be applicable to other systems. Standardisation to a 'reaction balance' metric also gave a more meaningful output than a concentration scale. Likewise, ensuring a consistent fluorophore environment would be important to other systems, which could use the same universal modified sequences and 6T spacer.

Removing operator secondary structure was another necessary step to minimise reaction times. Performing a single mutant EMSA screen following operator identification would be a recommended action for other aTFs to understand sequence dependency and manipulate binding affinities in combination with structure removal. Designing ligand-insensitive mutant TFs may be harder for repressors with less-studied allosteric domains, but if possible would likely also improve specificity, especially if the goal is to analyse samples with high variation in non-specific solutes. Examining rates of signal change instead of absolute values could also improve and quicken analyte quantification.

The depth of modelling achieved in this project was less than intended. Implementation of the normalisation methodology and the consistent fluorophore signal, which were achieved late on, delayed initial progress. Identifying reasonable, unbiased parameter distributions and relationships was also necessary, as well as understanding methods for assessing convergence. While the model was not able to be used for composition optimisation over the course of this work, it is poised to do so soon, and progress towards this aim did encourage the integration of the crucial controls. Models of repressor behaviour did suggest more complex displacement mechanisms may be present and, if confirmed, would be a highly useful insight to inform alternative designs and potential performance gains.

The original aim for the modelling was to aid development of biochemical signal processing methods to reduce unreliability of biosensors facing environmental variation. In the current biosensor design other features perform this role, but further modelling could certainly aid how well this is accomplished. Integration with existing DSD motifs was also considered, however the change to a reversible reaction limits the use of existing methods of signal amplification and thresholding due to their reliance on irreversible inputs. It is possible that identification of an artificially strong operator sequence could sufficiently reduce aTF unbinding rate that an irreversible

system could be trialled, in which case integration may be possible.

The devised experimental framework may still find utility beyond biosensing applications. Detailed kinetic characterisation of DNA-binding proteins and their affinities for both operators and ligands (for aTFs) could be interrogated through model inference, and would be much more informative than classical techniques such as EMSAs that only assess steady state ratios. This depth of characterisation, combined with operator sequence space evaluation and possible secondary structure removal, could improve the modularity and orthogonality of aTF-operator pairs, developing them into novel synthetic biology parts.

This could lead to use of aTFs not only in other diagnostics, but as stand-alone signal control elements in complex DSD computation. Layered riboswitch logic gates such as those by Green *et al.* are relatively slow, due to translational reporting, and could be replaced by repressors controlling release of different functional oligos. Reliance on dissociation and displacement processes alone could significantly speed up these systems, while integration with small molecule signal transduction would vastly expand the range of possible inputs. It is hoped, therefore, that this work will aid not only the development of future *in vitro* biosensors, but new methods of synthetic biological computation.

Appendix A

Appendix

A.1 Supplementary tables

List of oligonucleotides

Name	Sequence	Probe usage
sB22	CAAAATAAATTGATTTATTTGC	B22
sB22c	GCAAATAAATCAATTTATTTTG	B22
sB24	TCAAAATAAATTGATTTATTTGCT	B24
sB24c	AGCAAATAAATCAATTTATTTTGA	B24
sB26	ATCAAAATAAATTGATTTATTTGCTT	B26
sB26c	AAGCAAATAAATCAATTTATTTTGAT	B26
sB28	AATCAAAATAAATTGATTTATTTGCTTG	B28
sB28c	CAAGCAAATAAATCAATTTATTTTGATT	B28
sB30	TAATCAAAATAAATTGATTTATTTGCTTGC	B30, B30c-G6
sB30c	GCAAGCAAATAAATCAATTTATTTTGATTA	B30
sB36	AATTAATCAAAATAAATTGATTTATTTGCTTGCATT	B36
sB36c	AATGCAAGCAAATAAATCAATTTATTTTGATTAATT	B36
sB44	TTACAATTAATCAAAATAAATT GATTTATTTGCTTGCATTATTT	B44
sB44c	AAATAATGCAAGCAAATAAATC AATTTATTTTGATTAATTGTAA	B44
sB30c-G6	GCAAGCAAATAAATCAATTTATTTTGATTA-CACTGG	B30c-G6
X6-sB30	CCAGTG-TAATCAAAATAAATTGATTTATTTGCTTGC	B30c-G6
TYE-sB30c-G6	**/5TYE563/**GCAAGCAAATAAATCAATTTATTTTGATTA-CACTGG	TYE-B30-G6
sB30-IABkFQ	TAATCAAAATAAATTGATTTATTTGCTTGC**/3IABkFQ/**	TYE-B30-G6

Name	Sequence	Probe usage
sC70	TCCACTATATATTGACGAATGTCGATATTGAAAGT ATTTTGAATATCGACAGGTATCAATATACCGAAAG	C70
sC70c	CTTTCGGTATATTGATACCTGTCGATATTCAAAAT ACTTTCAATATCGACATTCGTCAATATATAGTGGA	C70
sCL30	TCCACTATATATTGACGAATGTCGATATTG	CL30
sCL30c	CAATATCGACATTCGTCAATATATAGTGGA	CL30
sCR30	GAATATCGACAGGTATCAATATACCGAAAG	CR30
sCR30c	CTTTCGGTATATTGATACCTGTCGATATTC	CR30, CR3-30, CR5-30, CR7-30, CR1-28, CR1-26, CR1-24, CR1-22
Fprobe	GCACGGCGTACGGACTCACGGTGT	All non-F Tprobes & Gprobes
Al488-Fprobe	**/Al488/**GCACGGCGTACGGACTCACGGTGT	All Tprobes & Gprobes
Fadapter (*Fa*)	ACACCGTGAGTCCGTACGCCGTGC	
Qprobe	TGGTCGGTGCTCGCAGGCTCGGCA	All non-Q Tprobes & Gprobes
Qprobe-IABkFQ	TGGTCGGTGCTCGCAGGCTCGGCA**/3IABkFQ/**	All Tprobes & Gprobes
Qadapter (*Qa*)	TGCCGAGCCTGCGAGCACCGACCA	
Fa-sCR30-G6a	*Fa*-GAATATCGACAGGTATCAATATACCGAAAG-GCGGGC	TR30-G6a
sCR30c-Qa	CTTTCGGTATATTGATACCTGTCGATATTC-*Qa*	TR30-G6a
X6a-sCR30c	GCCCGC-CTTTCGGTATATTGATACCTGTCGATATTC	TR30-G6a
Fa-sCL30c-G6a	*Fa*-CAATATCGACATTCGTCAATATATAGTGGA-GCGGGC	TL30c-G6a
sCL30-Qa	TCCACTATATATTGACGAATGTCGATATTG-*Qa*	TL30c-G6a
X6a-sCL30	GCCCGC-TCCACTATATATTGACGAATGTCGATATTG	TL30c-G6a
sCR3-30	ATATCGACAGGTATCAATATACCGAAAG	CR3-30
sCR5-30	ATCGACAGGTATCAATATACCGAAAG	CR5-30
sCR7-30	CGACAGGTATCAATATACCGAAAG	CR7-30
sCR1-28	GAATATCGACAGGTATCAATATACCGAA	CR1-28
sCR1-26	GAATATCGACAGGTATCAATATACCG	CR1-26
sCR1-24	GAATATCGACAGGTATCAATATAC	CR1-24
sCR1-22	ATATCGACAGGTATCAATATAC	CR1-22
Fa-sCR26-G6a	*Fa*-GAATATCGACAGGTATCAATATACCG-GCGGGC	TR26-G6a
sCR26c-Qa	CGGTATATTGATACCTGTCGATATTC-*Qa*	TR26-G6a, TR26-G6b, TR26-G4a

Name	Sequence	Probe usage
X6a-sCR26c	GCCCGC-CGGTATATTGATACCTGTCGATATTC	TR26-G6a
Fa-sCR24-G6a	*Fa*-GAATATCGACAGGTATCAATATAC-GCGGGC	TR24-G6a
sCR24c-Qa	GTATATTGATACCTGTCGATATTC-*Qa*	TR24-G6a, TR24-G6b, TR24-G4a
X6a-sCR24c	GCCCGC-GTATATTGATACCTGTCGATATTC	TR24-G6a
Fa-sCR54-G6a	*Fa*-TGTGTGGAGAGTGGTCAT-TATTTT-GAATATCGACAGGTATCAATATACCGAAAG-GCGGGC	TR54-G6a
sCR54c-Qa	CTTTCGGTATATTGATACCTGTCGATATTC-AAAATA-ATGACCACTCTCCACACA-*Qa*	TR54-G6a
X6a-sCR54c	GCCCGC-CTTTCGGTATATTGATACCTGTCGATATTC-AAAATA-ATGACCACTCTCCACACA	TR54-G6a
Fa-sCR26-G6b	*Fa*-GAATATCGACAGGTATCAATATACCG-ATAAAT	TR26-G6b
X6b-sCR26c	ATTTAT-CGGTATATTGATACCTGTCGATATTC	TR26-G6b
Fa-sCR26-G4a	*Fa*-GAATATCGACAGGTATCAATATACCG-GCGG	TR26-G4a
X4a-sCR26c	CCGC-CGGTATATTGATACCTGTCGATATTC	TR26-G4a
Fa-sCR24-G6b	*Fa*-GAATATCGACAGGTATCAATATAC-ATAAAT	TR24-G6b
X6b-sCR24c	ATTTAT-GTATATTGATACCTGTCGATATTC	TR24-G6b
Fa-sCR24-G4a	*Fa*-GAATATCGACAGGTATCAATATAC-GCGG	TR24-G4a
X4a-sCR24c	CCGC-GTATATTGATACCTGTCGATATTC	TR24-G4a
X6a-sCR17c	GCCCGC-CGGTATATTGATACCTG	TR26-G6a
X6a-sCR18c	GCCCGC-CGGTATATTGATACCTGT	TR26-G6a
X6a-sCR19c	GCCCGC-CGGTATATTGATACCTGTC	TR26-G6a
Fa-sCL26c-G6a	*Fa*-CAATATCGACATTCGTCAATATATAG-GCGGGC	TL26c-G6a
sCL26-Qa	CTATATATTGACGAATGTCGATATTG-*Qa*	TL26c-G6a
X6a-sCL16	GCCCGC-CTATATATTGACGAAT	TL26c-G6a
X6a-sCL17	GCCCGC-CTATATATTGACGAATG	TL26c-G6a
X6a-sCL18	GCCCGC-CTATATATTGACGAATGT	TL26c-G6a
X6a-sCL24	GCCCGC-CTATATATTGACGAATGTCGATAT	TL26c-G6a
pT-30	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	
sCR30-IR1	GAATATCGACAGGTATGTCGATATTGAAAG	CR30-IR1
sCR30-IR1c	CTTTCAATATCGACATACCTGTCGATATTC	CR30-IR1
sCR30-IR2	GGGTATATTGAGGTATCAATATACCGAAAG	CR30-IR2
sCR30-IR2c	GGGTATATTGAGGTATCAATATACCGAAAG	CR30-IR2
sCR30-6Space	GAATATCGACAGGTCTATCAATATACCGAA	CR30-6Space
sCR30-6Space.c	TTCGGTATATTGATAGACCTGTCGATATTC	CR30-6Space
sCR30-2Space	GAATATCGACAGATCAATATACCGAAAGTC	CR30-2Space
sCR30-2Space.c	GACTTTCGGTATATTGATCTGTCGATATTC	CR30-2Space
sCR30-SpaceACAT	GAATATCGACAACATTCAATATACCGAAAG	CR30-SpaceACAT

Name	Sequence	Probe usage
sCR30-SpaceACATc	CTTTCGGTATATTGAATGTTGTCGATATTC	CR30-SpaceACAT
sCR30-SpaceCAGC	GAATATCGACACAGCTCAATATACCGAAAG	CR30-SpaceCAGC
sCR30-SpaceCAGCc	CTTTCGGTATATTGAGCTGTGTCGATATTC	CR30-SpaceCAGC
sCR30-SpaceGTTG	GAATATCGACAGTTGTCAATATACCGAAAG	CR30-SpaceGTTG
sCR30-SpaceGTTGc	CTTTCGGTATATTGACAACTGTCGATATTC	CR30-SpaceGTTG
sCR30-SpaceTGCA	GAATATCGACATGCATCAATATACCGAAAG	CR30-SpaceTGCA
sCR30-SpaceTGCAc	CTTTCGGTATATTGATGCATGTCGATATTC	CR30-SpaceTGCA
sCR30-SpaceATTA	GAATATCGACAATTATCAATATACCGAAAG	CR30-SpaceATTA
sCR30-SpaceATTAc	CTTTCGGTATATTGATAATTGTCGATATTC	CR30-SpaceATTA
sCR30-SpaceGCCG	GAATATCGACAGCCGTCAATATACCGAAAG	CR30-SpaceGCCG
sCR30-SpaceGCCGc	CTTTCGGTATATTGACGGCTGTCGATATTC	CR30-SpaceGCCG
sCR30-T4C	GAACATCGACAGGTATCAATATACCGAAAG	CR30-T4C
sCR30-T4Cc	CTTTCGGTATATTGATACCTGTCGATGTTC	CR30-T4C
sCR30-A9G	GAATATCGGCAGGTATCAATATACCGAAAG	CR30-A9G
sCR30-A9Gc	CTTTCGGTATATTGATACCTGCCGATATTC	CR30-A9G
sCR30-C17T	GAATATCGACAGGTATTAATATACCGAAAG	CR30-C17T
sCR30-C17Tc	CTTTCGGTATATTAATACCTGTCGATATTC	CR30-C17T
sCR30-A21G	GAATATCGACAGGTATCAATGTACCGAAAG	CR30-A21G
sCR30-A21Gc	CTTTCGGTACATTGATACCTGTCGATATTC	CR30-A21G
sCR30-CtoG	GAATATGGAGAGGTATGAATATAGGGAAAG	CR30-CtoG
sCR30-CtoGc	CTTTCCCTATATTCATACCTCTCCATATTC	CR30-CtoG
sCR30-GtoC	CAATATCCACACCTATCAATATACCCAAAC	CR30-GtoC
sCR30-GtoCc	GTTTGGGTATATTGATAGGTGTGGATATTG	CR30-GtoC
sCR30-CtoG-A21G	GAATATGGAGAGGTATGAATGTAGGGAAAG	CR30-CtoG-A21G
sCR30-CtoG-A21Gc	CTTTCCCTACATTCATACCTCTCCATATTC	CR30-CtoG-A21G
sCR30-GtoC-T4C	CAACATCCACACCTATCAATATACCCAAAC	CR30-GtoC-T4C
sCR30-GtoC-T4Cc	GTTTGGGTATATTGATAGGTGTGGATGTTG	CR30-GtoC-T4C

Name	Sequence	Probe usage
Fa-CR(GtoC)26-G6d	*Fa*-CAATATCCACACCTATCAATATACCC-GCCCGC	TR(GtoC)26-G6d
sCR(GtoC)26c-Qa	GGGTATATTGATAGGTGTGGATATTG-*Qa*	TR(GtoC)26-G6d
X6d-sCR(GtoC)16c (or X6d-GtoC)	GCGGGC-GGGTATATTGATAGGT	TR(GtoC)26-G6d
X6d-sCR(GtoC)17c	GCGGGC-GGGTATATTGATAGGTG	TR(GtoC)26-G6d
X6d-sCR(GtoC)18c	GCGGGC-GGGTATATTGATAGGTGT	TR(GtoC)26-G6d
X6d-sCR(GtoC)24c	GCGGGC-GGGTATATTGATAGGTGTGGATAT	TR(GtoC)26-G6d
X6g-t16	GGGGGG-TTTTTTTTTTTTTTT	
X6g-g3t13	GGGGGG-GGGTTTTTTTTTTTT	
X6dt-GtoC	GTGGGT-GGGTATATTGATAGGT	
X6d-g3t13	GCGGGC-GGGTTTTTTTTTTTT	
X6d-t16	GCGGGC-TTTTTTTTTTTTTTT	
X6dt-g3t13	GTGGGT-GGGTTTTTTTTTTTT	
X6dt-t16	GTGGGT-TTTTTTTTTTTTTTT	
X6gc-GtoC	GCGCGC-GGGTATATTGATAGGT	
X6gc-g3t13	GCGCGC-GGGTTTTTTTTTTTT	
X6gc-t16	GCGCGC-TTTTTTTTTTTTTTT	
sCR0C26c-Qa	CAATATCCACACCTATCAATATACCC-*Qa*	TR0C26-G6m, TR0C26-G6a, T0T0C
Fa-sCR0C26-G6m	GGGTATATTGATAGGTGTGGATATTG-TGGAGA	TR0C26-G6m
X6m-sCR0C16c	TCTCCA-CAATATCCACACCTAT	TR0C26-G6m
X6m-sCR0C18c	TCTCCA-CAATATCCACACCTATCA	TR0C26-G6m
X6m-sCR0C24c	TCTCCA-CAATATCCACACCTATCAATATAC	TR0C26-G6m
Fa-sCR0C26-G6a	GGGTATATTGATAGGTGTGGATATTG-GCGGGC	TR0C26-G6a
X6a-sCR0C16c	GCCCGC-CAATATCCACACCTAT	TR0C26-G6a
X6a-sCR0C18c	GCCCGC-CAATATCCACACCTATCA	TR0C26-G6a
X6a-sCR0C24c	GCCCGC-CAATATCCACACCTATCAATATAC	TR0C26-G6a

Name	Sequence	Probe usage
Fa-sCR0C-G6c	GGGTATATTGATAGGTGTGGATATTG-GGCGGG	T0T0C
X6c-sCR0C16c	CCCGCC-CAATATCCACACCTAT	T0T0C
X6c-sCR0C18c	CCCGCC-CAATATCCACACCTATCA	T0T0C, T4T0C, T6T0C
X6c-sCR0C24c	CCCGCC-CAATATCCACACCTATCAATATAC	T0T0C, T4T0C, T6T0C
X6a-sCR20c	GCCCGC-CGGTATATTGATACCTGTCG	TR26-G6a
X6a-sCR25c	GCCCGC-CGGTATATTGATACCTGTCGATATT	TR26-G6a
Fa-4T-sCR0C26-G6c	*Fa*-TTTT-GGGTATATTGATAGGTGTGGATATTG-GGCGGG	T4T0C
Fa-6T-sCR0C26-G6c	*Fa*-TTTTTT-GGGTATATTGATAGGTGTGGATATTG-GGCGGG	T6T0C, L6T0C
X6c-sCR0C19c	CCCGCC-CAATATCCACACCTATCAA	T0T0C, T4T0C, T6T0C, L6T0C
X6c-sCR0C20c	CCCGCC-CAATATCCACACCTATCAAT	T0T0C, T4T0C, T6T0C, L6T0C
X6c-sCR0C21c	CCCGCC-CAATATCCACACCTATCAATA	T0T0C, T4T0C, T6T0C
X6c-sCR0C22c	CCCGCC-CAATATCCACACCTATCAATAT	T0T0C, T4T0C, T6T0C
X6c-sCR0C23c	CCCGCC-CAATATCCACACCTATCAATATA	T0T0C, T4T0C, T6T0C
X6c-sCR0C25c	CCCGCC-CAATATCCACACCTATCAATATACC	T0T0C, T4T0C, T6T0C
X6c-sCR0C26c	CCCGCC-CAATATCCACACCTATCAATATACCC	T0T0C, T4T0C, T6T0C, L6T0C
sCR1C26c-Qa	CAATATCGACACCTATCAATATACCC-*Qa*	T6T1C
Fa-6T-sCR1C26-G6c	*Fa*-TTTTTT-GGGTATATTGATAGGTGTCGATATTG-GGCGGG	T6T1C, L6T1C
X6c-sCR1C18c	CCCGCC-CAATATCGACACCTATCA	T6T1C, L6T1C

Name	Sequence	Probe usage
X6c-sCR1C19c	*CCCGCC*-CAATATCGACACCTATCAA	T6T1C, L6T1C
X6c-sCR1C20c	CCCGCC-CAATATCGACACCTATCAAT	T6T1C, L6T1C
X6c-sCR1C21c	CCCGCC-CAATATCGACACCTATCAATA	T6T1C, L6T1C
X6c-sCR1C22c	CCCGCC-CAATATCGACACCTATCAATAT	T6T1C, L6T1C
X6c-sCR1C23c	CCCGCC-CAATATCGACACCTATCAATATA	T6T1C, L6T1C
X6c-sCR1C24c	CCCGCC-CAATATCGACACCTATCAATATAC	T6T1C, L6T1C
X6c-sCR1C25c	CCCGCC-CAATATCGACACCTATCAATATACC	T6T1C, L6T1C
X6c-sCR1C26c	CCCGCC-CAATATCGACACCTATCAATATACCC	T6T1C, L6T1C
sCR0C26c-IABkFQ	CAATATCCACACCTATCAATATACCC**/3IABkFQ/**	L6T0C
sCR1C26c-IABkFQ	CAATATCGACACCTATCAATATACCC**/3IABkFQ/**	L6T1C

Table A.1: **List of oligonucleotides.** All the ssDNA oligonucleotides used and referenced in this thesis for *in vitro* assays. Sequences are given in 5' to 3' orientation, with distinct domains separated by hyphens. Probes listed are annealed species used for EMSAs or fluorescent DSD assays. Toehold sequences are underlined, functional modifications are in bold and abbreviations for universal sequences are italicised.

List of primers

Name	Sequence	Usage
T7F	TAATACGACTCACTATAGGG	pET-28a plasmid sequencing
T7R	GCTAGTTATTGCTCAGCGG	pET-28a plasmid sequencing
16SF	AACAGAATGCAGCACCTTGGC	CgArsR-16S mutagenesis PCR
16SR	GGGTTCGCCAACTGGATA	CgArsR-16S mutagenesis PCR

Table A.2: **List of primers.** All the ssDNA oligonucleotides used and referenced in this thesis for priming PCRs. Sequences are given in 5' to 3' orientation.

List of gBlocks

Name	Sequence
JC-AfArsR	ccagtttggtctca**cATG**GAGCCTTTACAAGACCCTGCTCAGATCGTTGCCCGCC TTGAGGCTTTGGCCTCCCCAGTGCGCTTAGAAATTTTCCGTTTGCTTGTG GAACAGGAACCTACCGGCCTGGTTAGCGGAGACATTGCAGAACACTTGG GACAGCCTCACAACGGCATCTCTTTTCATCTGAAGAACCTTCAACACGCG GGCTTGGTAACAGTACAACGTGAGGGTCGCTATCAGCGTTACCGCGCAGC AATGCCAGTTGTGCGTGCTTTGGTCGCGTACCTTACCGAGAACTGCTGTC ATGGTACTCGTGACTGTGCATTAAGCGGCGAGACTCGTTCACCCTCCGTT CAAGAGGGCAACCAA-*GAGAATCTGTATTTCCAAGGGGGTGGTGGTGGTT* *CGCACCATCACCACCATCATTAA***ggcc**tgagaccaaaggac
JC-BsArsR	ccagtttggtctca**cATG**GACGAGACAAAATCCGAATTGTTGCGTAAATACGAAC AGAAATTCAAAGCGTTGGCCGACCAAAAACGCTTGGAAATTATGTACGAG CTGTGCCAGCGCGGTAAGACGTGTGTATGCGATTTAACGGAGATTTTCGA GGTTACACAGAGTAAATTATCCTACCACTTAAAGATTCTGTTGGACGCGA ATTTGATTACTAAGGAGACGAAAGGTACCTGGTCGTATTATGATCTGAAT GATGAGGAAGTCAACGGGCTTCTGTCTGAGGAGCTTTGTTGTATCTTTCG CAAGAAAGGCGAGGGCGACTGCTGT-*GAAAATCTGTACTTTCAGGGTGG* *AGGTGGGGGAAGTCATCACCACCATCACCACTAA***ggcc**tgagaccaaaggac
JC-CgArsR	ccagtttggtctca**cATG**ACTACACTGCATACTATCCAGTTGGCGAACCCAACAGA ATGCTGCACCTTGGCAACTGGGCCATTAAGTTCCGACGAAAGCGAACACT ACGCGGATTTATTCAAGGTACTTGGTGACCCGGTGCGTCTTCGTATCTTG TCGCAGCTGGCAGCCGGTGGCTGTGGGCCGGTTTCGGTTAATGAACTGA CTGATCTTATGGGATTATCACAGCCGACGATCTCACACCACTTGAAGAAAA TGACCGAGGCGGGTTTCCTTGATCGTGTACCTGAAGGACGTGTGGTACTG CATCGTGTACGCCCGGAGCTTTTCGCCGAGTTGCGTACAGTTTTGCAAAT CGGTTCAATGGAGTTA-*GAGAACCTGTACTTTCAAGGGGGTGGAGGGGGA* *TCTCATCATCACCATCATCATTAA***ggcc**tgagaccaaaggac
JC-CmArsR	ccagtttggtctca**cATG**GAGACAGAAAACGCCCTGGAAGCACTGGCTGCCTTAG CGCACGGGATCCGTCTTGCAGTATTTCGCCTTTTGGTACAGGCCGGGCCC GAGGGCTTGCCCGCCGGTCGCATTGCTGAGCTGATGGAGATGCCCGCCT CATCACTTTCATTCCATTTGAAGGAGTTACACCGTGCTGGTCTTTTGGCAT CTCGCCAAGAGGGACGTTCTATCATCTATATGGCCCAATTCGAGACAATG AACGCGTTGTTAGGCTACTTGACTGAGAACTGCTGCGGTGGTGCTCCATG CTCCCCCGTTTCATCCTGTTCGGTTGCGACCGAGTCA-*GAAAACCTTTAC* *TTCCAGGGGGGCGGCGGAGGTAGCCATCACCACCATCACCATTAA***ggcc**tgag accaaaggac

Name	Sequence
JC-EcCArsR	ccagtttggtctca**cATG**AGCTTCCTTTTACCGATTCAGTTATTTAAGATTTTGGCT GATGAAACTCGCTTGGGCATTGTGCTTTTACTTTCCGAATTAGGAGAGCT GTGTGTCTGTGATCTTTGTACGGCTCTGGATCAAAGTCAGCCGAAAATCT CTCGCCATTTAGCATTATTGCGCGAATCAGGTCTTTTGCTTGACCGTAAA CAGGGGAAGTGGGTTCATTACCGCTTGTCACCCCATATCCCAGCGTGGGC GGCCAAAATTATTGACGAAGCGTGGCGCTGTGAACAGGAAAAAGTTCAA GCGATTGTGCGTAATTTAGCTCGCCAGAATTGTTCTGGCGACTCTAAAAA CATTTGCAGT-*GAAAATTTATACTTTCAAGGTGGGGGTGGCGGTAGCCATC* *ACCATCATCACCACTAA***ggcct**gagaccaaaggac
JC-EcRArsR	ccagtttggtctca**cATG**CTTCAGTTGACTCCTTTGCAGCTGTTTAAGAATTTAAG CGATGAGACGCGTCTGGGAATCGTCTTGTTACTTCGTGAGATGGGTGAGC TGTGCGTTTGCGACTTGTGTATGGCTCTTGACCAGTCGCAGCCGAAGATT TCCCGTCACTTGGCGATGCTGCGCGAGTCGGGGATCCTTTTGGACCGCAA ACAGGGCAAATGGGTCCACTATCGTCTTTCACCACACATCCCTTCGTGGG CAGCTCAAATCATTGAGCAGGCGTGGTTGTCACAACAAGATGATGTTCAG GTAATTGCTCGCAAATTAGCGAGCGTAAACTGTTCAGGGTCGAGCAAGGC CGTGTGTATT-*GAAAACTTATATTTTCAAGGGGGCGGGGGGGGATCGCATC* *ATCACCATCATCACTAA***ggcct**gagaccaaaggac
JC-PsArsR	ccagtttggtctca**cATG**TCATTTCTTCTTTTGCACGGAATCGTTCACCAATTTCTT CTGATGAAACCCGGTTGGGCCAGTTTCTACTGCTGGGCAAACTGGGAGT CTTATCCCAGCGCCATTAGTGCTCTTTTAGATCAGTCACAACCCAAAAAGT CGCGTCACTTGGCCTTGCTTCGTGAATCTGGGCTTTTACTTGATCGCAAA AAGCGTTACGTTGGAAGTCTTCCGCTGATCACAGCGTATTCTAGTATCTT CGGTGAGAATTATTGCTGTGGGTTAGCGATGTGTCATCGCAAACGCTTTC GCCGCTTATCGGCGCACCTGGCACGCCAGAATTGTTCAGGGGATTCGAAA AATATCTGTAGC-*GAAAACTTGTACTTCCAGGGGGGCGGCGGGGGTTCCC* *ACCATCATCATCACCACTAA***ggcct**gagaccaaaggac
JC-SxArsR	ccagtttggtctca**cATG**TCATACAAAGAATTGTCCACAATCTTAAAAGTACTGTC AGATCCTAGTCGCTTAGAGATTCTGGACTTACTGTCATGTGGCGAATTAT GTGCCTGCGACTTATTGGAACACTTTCAATTTTCGCAACCTACATTATCA CATCACATGAAATCATTAGTTGATAACGAGTTAGTGACTACACGTAAAAA CGGAAATAAACATATGTACCAGTTGAATCACGAGTTTTTAGATTATATCA ACCAAAACTTGGATATTATTAATACGTCCGATCAAGGATGTGCTTGCAAG AACATGAAAAGCGGCGAATGC-*GAGAATTTGTACTTCCAGGGCGGCGGGG* *GTGGATCGCATCACCATCATCACCATTAA***ggcct**gagaccaaaggac

Name	Sequence
JC-EcDTetR	ccagtttggtctca**cATG**GCGCGCCTGAATCGCGAATCAGTAATTGACGCAGCCC TGGAATTACTGAACGAAACGGGTATCGACGGATTAACGACTCGCAAGTTA GCGCAAAAACTTGGGATCGAGCAGCCCACGTTGTATTGGCACGTCAAGAA TAAGCGCGCCTTGCTGGATGCGCTTGCGGTTGAAATCCTGGCACGTCACC ATGACTATTCATTGCCTGCCGCAGGAGAAAGCTGGCAGTCTTTCCTGCGT AATAATGCGATGTCTTTCCGTCGCGCGTTGTTACGCTATCGTGACGGGGC CAAAGTACACCTGGGCACTCGTCCAGATGAAAAACAGTACGATACTGTTG AAACGCAATTGCGCTTCATGACGGAAAATGGATTCAGTCTGCGCGATGGG CTTTATGCAATTAGTGCAGTGTCGCACTTCACATTAGGGGCCGTTCTTGA GCAGCAGGAGCATACGGCAGCTTTAACTGATCGTCCCGCGGCTCCAGACG AGAATCTTCCCCCCTTGTTGCGTGAAGCCCTTCAGATCATGGATTCGGAT GACGGGGAGCAGGCATTTCTTCACGGCCTTGAATCGTTAATCCGTGGCTT TGAAGTCCAACTTACAGCTTTGTTACAAATCGTCGGAGGTGACAAACTTA TCATTCCCTTCTGT-*GAAAACCTTTATTTCCAAGGTGGTGGCGGAGGCTCG* *CATCATCATCACCATCACTAA***ggcc**tgagaccaaaggac

Table A.3: **List of gBlocks.** All the dsDNA gBlock gene fragments designed and used in this thesis for cloning repressor genes into the pET-28a expression plasmid. Sequences are given in 5' to 3' orientation with coding regions in upper case and non-coding bases in lower case. BsaI recognition sites are underlined with resulting overhangs in bold and C-terminal TEV protease recognition site, glycine-serine spacer and His-tag are italicised. gBlocks were synthesised by IDT at the 500 ng scale.

Model	Figure	Description	LL	sigma	k1 ($nM^{-1}s^{-1}$)	k2 ($nM^{-1}s^{-1}$)	u1 (s^{-1})	u2 (s^{-1})
M1v1	5.13 (b) left	T6T0C X19, fixed controls	-184	0.285	5.88×10^{-4}	3.43×10^{-3}	1.07×10^{-2}	1.30×10^{-1}
M1v2	5.13 (b) right	T6T0C X19, variable controls	495	0.155	4.29×10^{-4}	3.14×10^{-4}	8.09×10^{-2}	1.64×10^{-1}
M1v3	5.14 right	TR26-G6a X18	-562	0.314	2.70×10^{-5}	8.70×10^{-3}	2.95×10^{-3}	3.24×10^{-0}
M1v4	5.15 (a)	L6T1C X19, default u	309	0.181	4.84×10^{-3}	1.44×10^{-3}	1.13×10^{-1}	1.13×10^{-1}
M1v5	5.15 (b)	L6T1C X19, single variable u	364	0.173	9.37×10^{-3}	2.93×10^{-3}	2.35×10^{-1}	2.35×10^{-1}
M1v6	5.15 (c)	L6T1C X19, variable $u1$ + $u2$	365	0.172	5.00×10^{-3}	4.21×10^{-4}	1.25×10^{-1}	3.38×10^{-2}
M1v7	5.16 (a)	L6T1C X19, fixed $c1$ + $c2$	-2270	1.972	9.94×10^{-3}	8.95×10^{-3}	1.44×10^{-3}	1.73×10^{-1}
M1v8	5.16 (b)	L6T1C X19, flexible $c1$ + $c2$	-656	0.446	3.09×10^{-3}	9.71×10^{-3}	9.33×10^{-3}	5.12×10^{-1}
M1v9	5.16 (c)	L6T1C X19, $c1$ shared inverse constant	365	0.172	6.83×10^{-3}	3.81×10^{-3}	1.71×10^{-3}	3.05×10^{-3}

Table A.4: **Parameters inferred by Visual DSD model M1.** The simulations presented in section 5.6 used the above rate constants, as inferred by Visual DSD model M1 when given the described experimental data and parameter ranges shown in Table A.6. For models 7-9 unbinding constants were derived from inferred/specified k/c values. The simulations were given a LogLikelihood (LL) score dependant on how well they fit the experimental data, alongside a sigma noise parameter. All data are provided rounded to three significant figures.

Model	Figure	LL	sigma	k3 ($nM^{-1}s^{-1}$)	u3 (s^{-1})	k4 ($nM^{-1}s^{-1}$)	u4 ($nM^{-1}s^{-1}$)	k5 ($nM^{-1}s^{-1}$)	u5 (s^{-1})	k6 ($nM^{-1}s^{-1}$)	u6 (s^{-1})
M2v1	5.18 (a)	-348	0.334	3.53×10^{-3}	2.01×10^{-2}						
M2v2	5.18 (b)			3.53×10^{-4}	2.01×10^{-3}						
M3v1	5.19 (a)	1326	0.0719	5.17×10^{-4}	1.10×10^{-2}	3.19×10^{-6}	3.81×10^{-4}				
M4v1	5.19 (b)	1269	0.0743	6.85×10^{-5}	2.65×10^{-3}			8.71×10^{-5}	1.94×10^{-1}	3.59×10^{-4}	8.96×10^{-2}

Table A.5: **Parameters inferred by Visual DSD models M2-4.** The simulations presented in Figures 5.18 and 5.19 used the above rate constants, as inferred by Visual DSD models M2-4 when given the described experimental data, parameter ranges shown in Table A.7 and $k1/k2/c1$ values from M1v9. The simulations were given a LogLikelihood (LL) score dependant on how well they fit the experimental data, alongside a sigma noise parameter. All data are provided rounded to three significant figures. M2v2 was just a simulation with $0.1\times k3/u3$ values so model fit was not assessed.

Model	k1 range	k2 range	u1 range	u2 range	c1 range	c2 range
M1v1 + M1v2	R/U(0.0001,0.01)	R/U(0.0001,0.01)	L/U(0.001,10)	L/U(0.001,10)		
M1v3	L/U(0.00001,0.01)	L/U(0.00001,0.01)	L/U(0.001,10)	L/U(0.001,10)		
M1v4	R/U(0.0001,0.01)	R/U(0.0001,0.01)	0.1126	0.1126		
M1v5	R/U(0.0001,0.01)	R/U(0.0001,0.01)	L/U(0.001,10)	Same as $u1$		
M1v6	R/U(0.0001,0.01)	R/U(0.0001,0.01)	L/U(0.001,10)	L/U(0.001,10)		
M1v7	R/U(0.0001,0.01)	R/U(0.0001,0.01)	$u1 = c1 \times k1$	$u2 = c2 \times k2$	0.145	19.3
M1v8	R/U(0.0001,0.01)	R/U(0.0001,0.01)	$u1 = c1 \times k1$	$u2 = c2 \times k2$	L/U(0.000920,3.014)	L/U(0.203,87.9)
M1v9	R/U(0.0001,0.01)	R/U(0.0001,0.01)	$u1 = c1/k1$	$u2 = c1/k2$	L/U(0.00001,0.1)	

Table A.6: **Parameter ranges used by Visual DSD model M1.** The parameters listed in Table A.4 were inferred by providing model M1 with the appropriate experimental data and limiting each rate constant space to the above ranges. These were either fixed, able to vary logarithmically (L) or linearly (R) over a uniform distribution (L/R)/U(minimum, maximum).

Model	k3 range	u3 range	k4 range	u4 range	k5 + k6 ranges	u5 + u6 ranges
M2v1	L/U(0.000001,0.01)	L/U(0.0001,1)				
M3v1	L/U(0.000001,0.01)	L/U(0.0001,1)	L/U(0.00000001,0.001)	L/U(0.00000001,0.001)		
M4v1	L/U(0.000001,0.01)	L/U(0.0001,1)			L/U(0.00000001,0.001)	L/U(0.001,10)

Table A.7: **Parameter ranges used by Visual DSD models M2-4.** The parameters listed in Table A.5 were inferred by providing models M2-4 with the appropriate experimental data and limiting each rate constant space to the above ranges. These were able to vary logarithmically over a uniform distribution L/U(minimum, maximum) in addition to the fixed values of $k1/k2/c1$ inferred from M1v9.

A.2 Supplementary figures

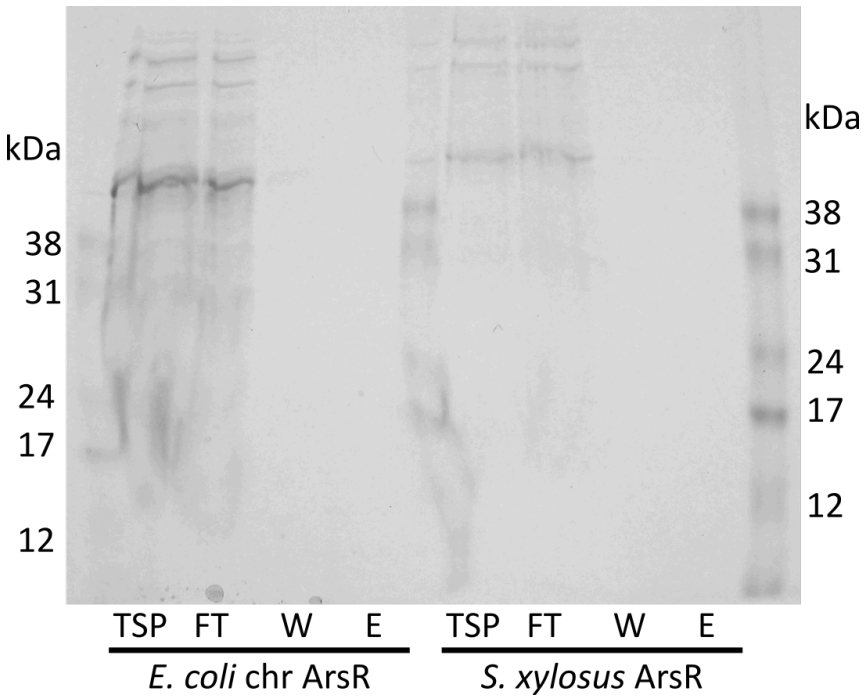


Figure A.1: **Expression attempt of EcCArsR and SxArsR.** Coomassie-stained SDS-PAGE gel of IPTG-induced 10 ml cultures, their total soluble protein (TSP) and the flow through (FT), wash (W) and eluate (E) stages of the attempted batch purification for each strain.

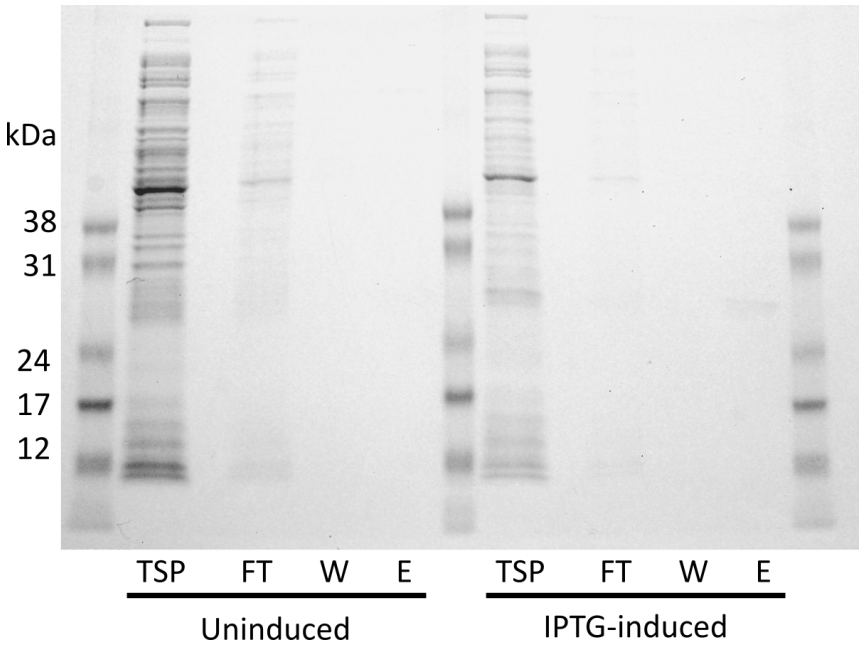


Figure A.2: **Large scale expression attempt of EcTetR.** Coomassie-stained SDS-PAGE gel of uninduced vs IPTG-induced 500 ml cultures for the EcTetR strain, with the total soluble protein (TSP), flow through (FT), wash (W) and eluate (E) stages of its attempted purification.

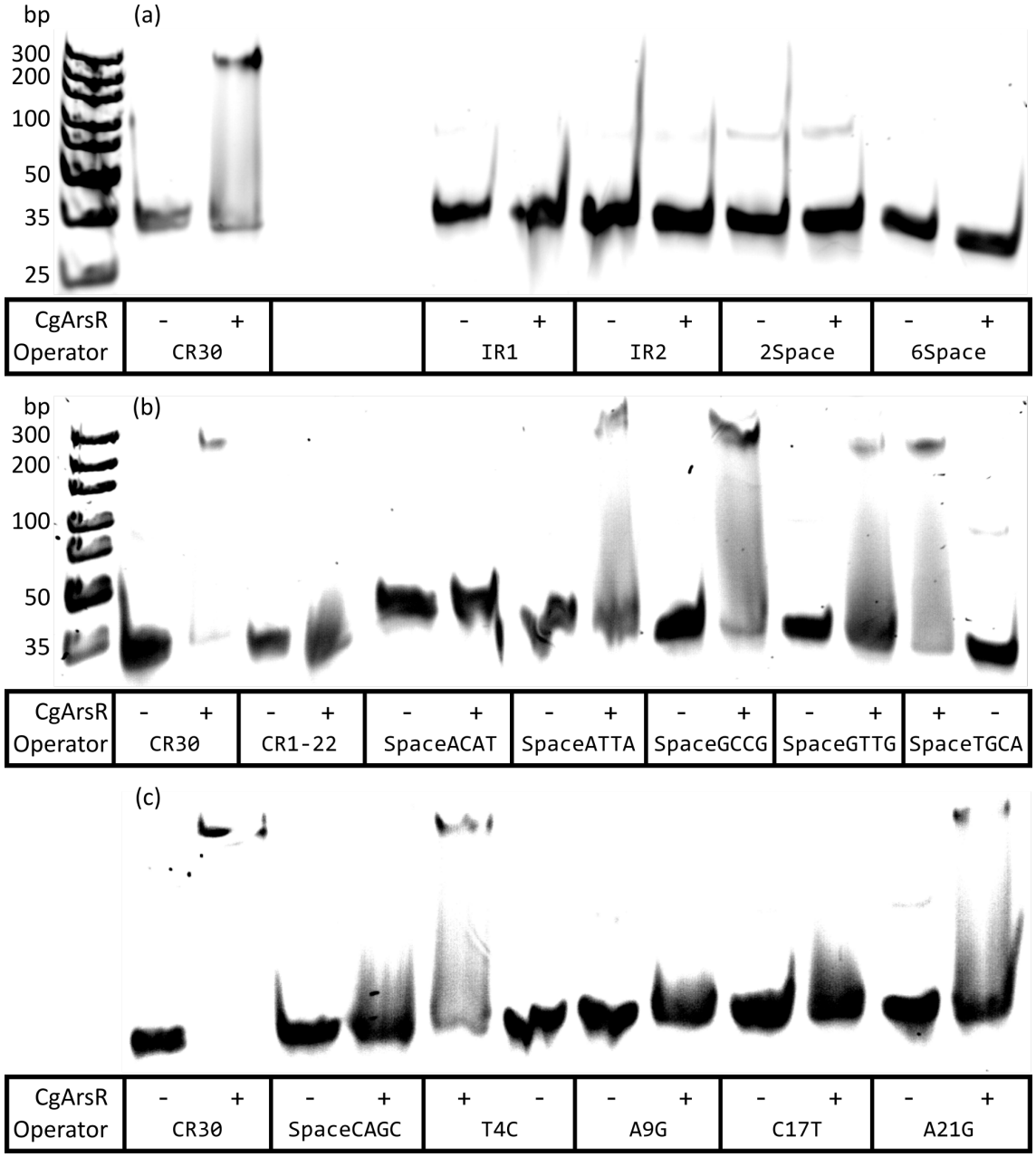


Figure A.3: **Testing CgArsR binding of mutant operators.** 20% PAGE EMSAs assaying the ability of CgArsR to bind various 30bp mutant operators. The sequence of each dsDNA construct is given in Figure 4.26. 15 pmol of each annealed duplex was incubated with 200 pmol CgArsR at room temperature for 30 min in a binding buffer containing 5 mM DTT before being run on the gel. All gels were run by B. Baker.

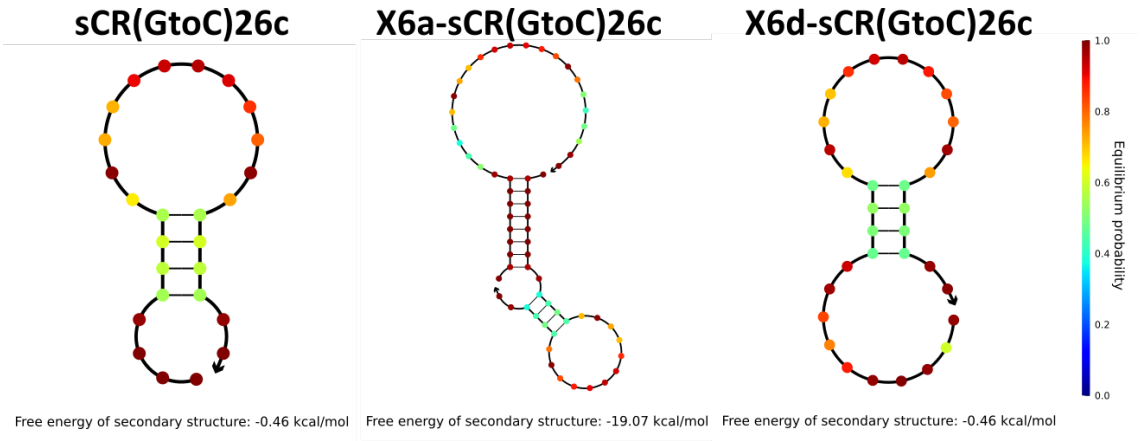


Figure A.4: **CR(GtoC)26 operator forms strong dimer with 6a toehold.** Minimum free energy secondary structures formed by single-stranded oligos comprised of sCR(GtoC)26c and displacing oligos with the X6a or X6d toeholds. The structures and associated minimum free energies were predicted using NUPACK at 1 µM and 25 °C, with individual bases coloured according to the probability of being in the suggested state using the scale on the right. The 3' end of each oligo is labelled with an outward arrow. 96% of X6a-sCR(GtoC)26c was predicted to form its dimer, with 4% forming a monomeric hairpin instead.

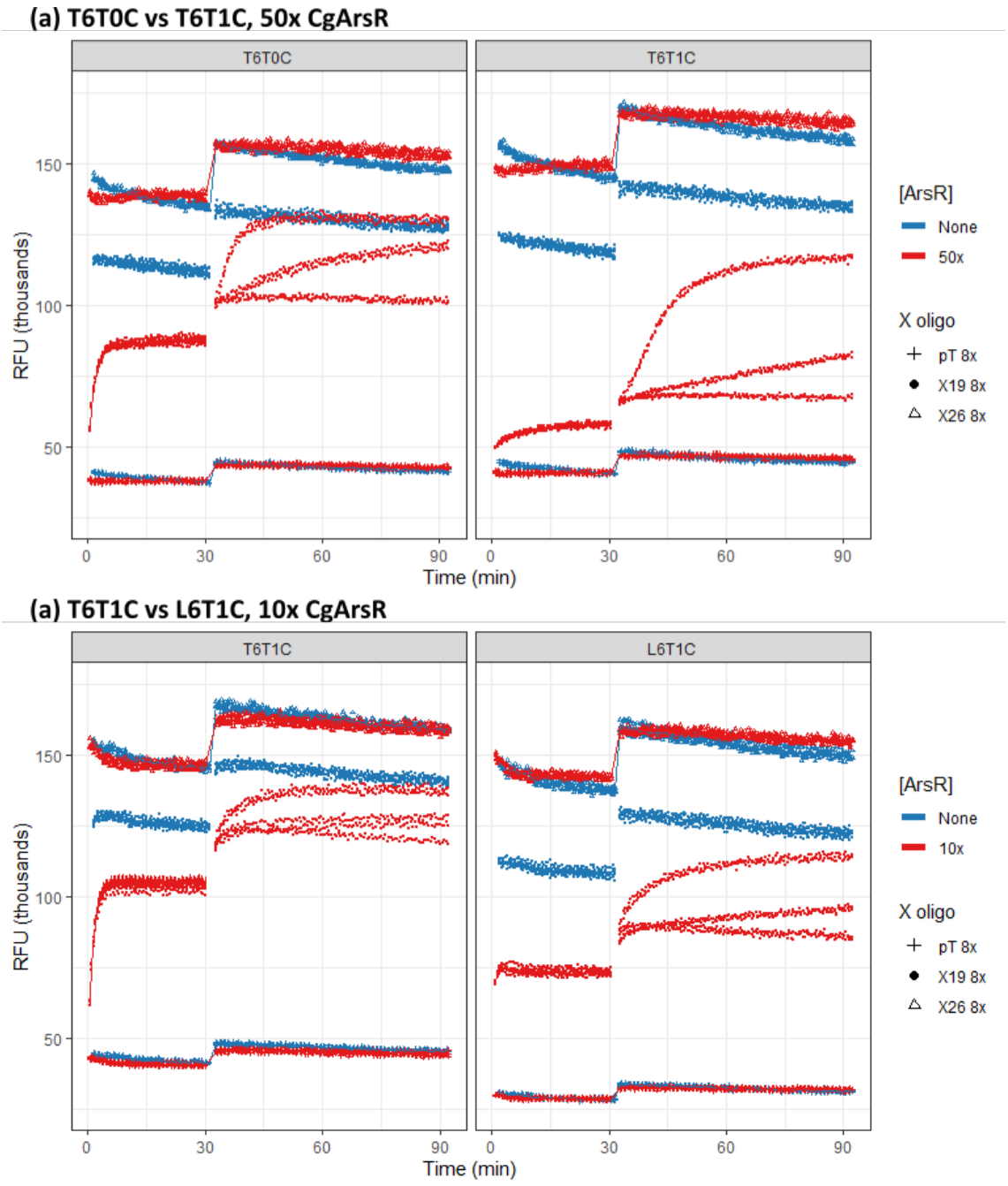


Figure A.5: **Probe comparison raw data.** Fluorescence time course displaying the raw data for the assays in Section 5.4. (a) Raw data for assay comparing T6T0C to T6T1C with or without 50x CgArsR, as shown in Figures 5.1-5.3. (b) Raw data for assay comparing T6T1C to L6T1C with or without 10x CgArsR, as shown in Figure 5.5.

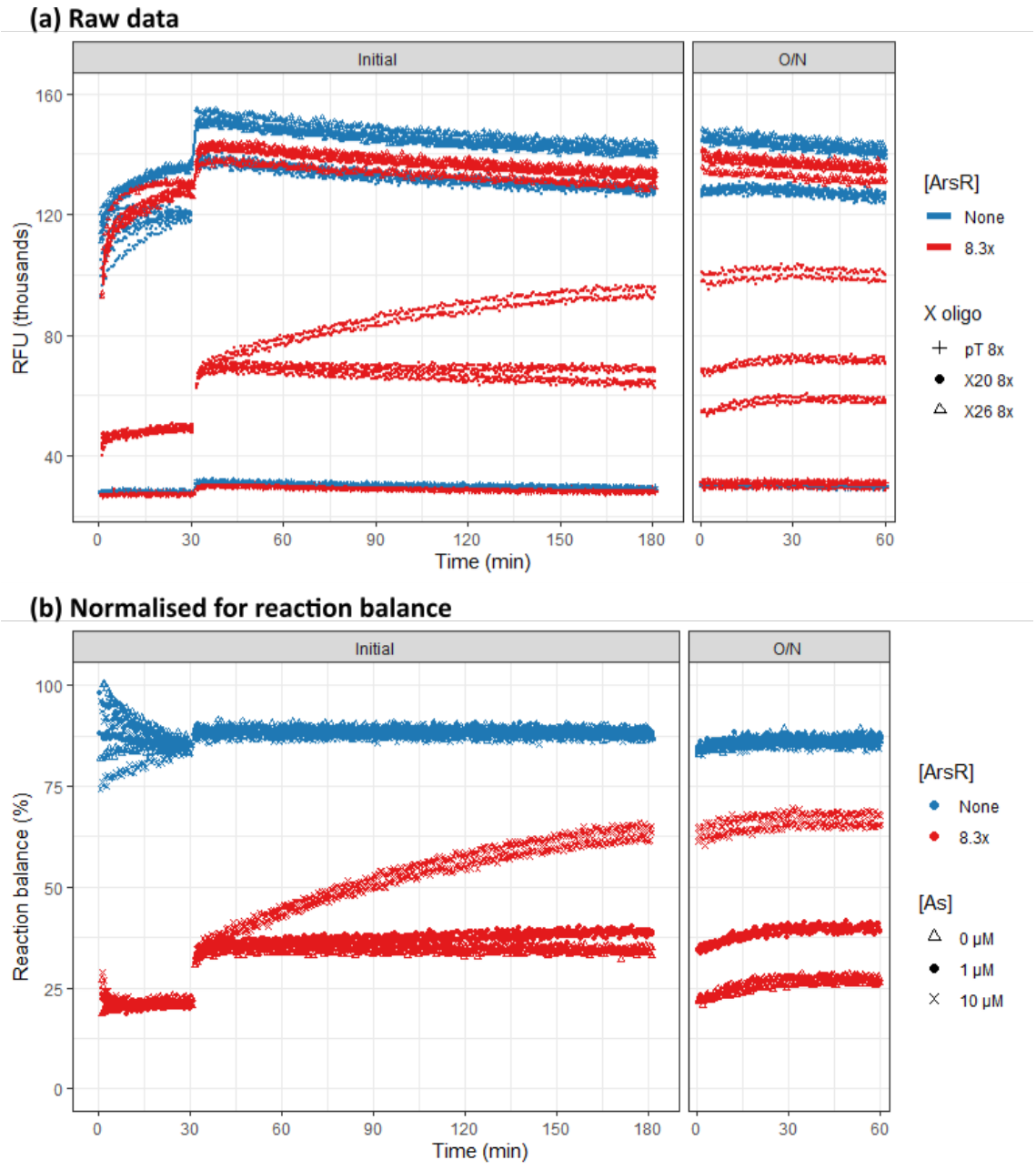


Figure A.6: **Modified buffer and switch to X20 enables 1 µM arsenite sensitivity.** Fluorescence time course displaying data for L6T1C, incubated with and without CgArsR in the DB6 buffer, undergoing reversible X20 DSD then responding to 50 µl of 0/1/10 µM sodium arsenite, under the DSD5 protocol in buffer DB6. Initial DSD was run for 30 min before arsenite was added and the assay continued for another 2.5 hours. The plate was sealed then restarted the following morning for a further hour. Raw (a) and normalised to reaction balance (b) data are shown here to complement the arsenite-induced difference in reaction balance data shown in Figure 5.10.

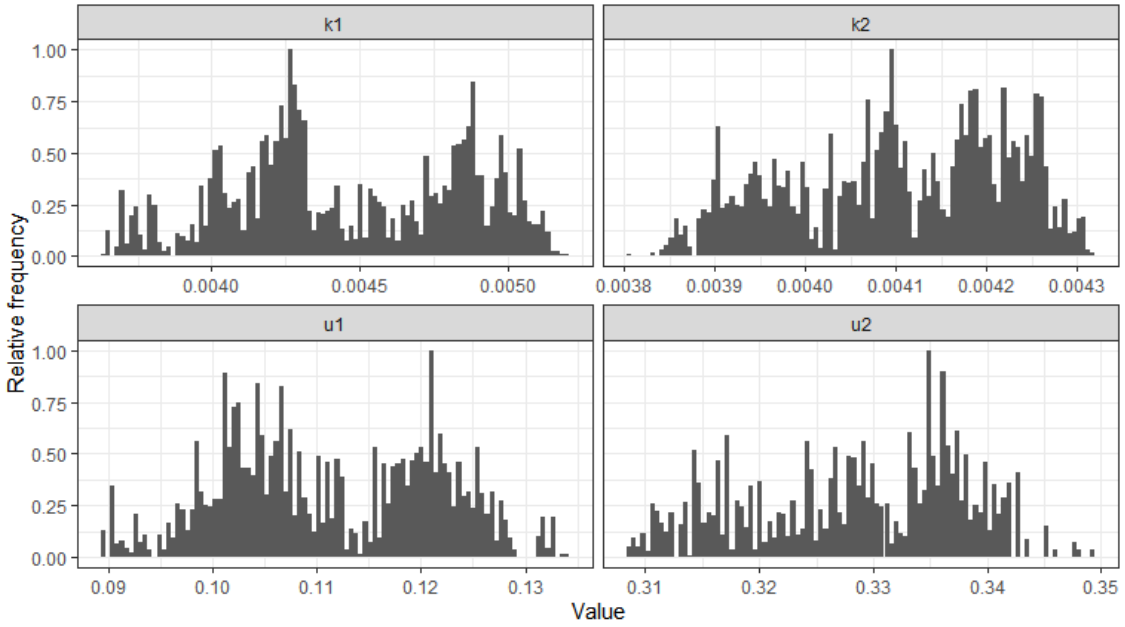


Figure A.7: **Posterior distribution of M1v6.** Histogram of binding constants inferred by Model M1v6, as seen in Figure 5.15 (c). 5,000 posterior iteration values were placed into 100 bins, scaled by relative frequency.

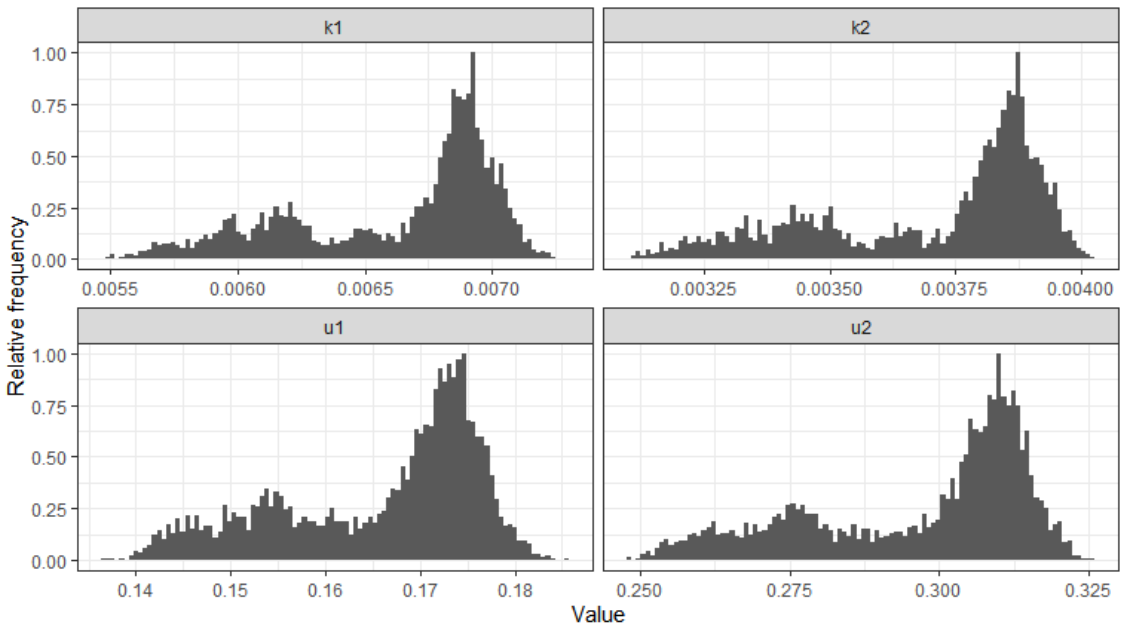


Figure A.8: **Posterior distribution of M1v9.** Histogram of binding constants inferred by Model M1v9, as seen in Figure 5.16 (c). Unbinding constants were not inferred directly, but calculated using the inferred $k1/k2/c1$ values. 5,000 posterior iteration values were placed into 100 bins, scaled by relative frequency.

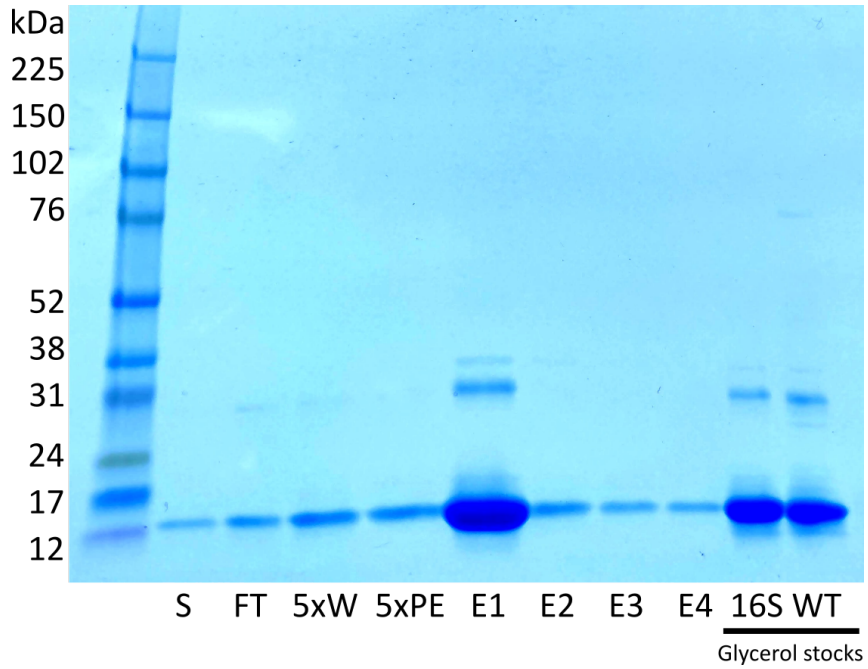


Figure A.9: **Purification of CgArsR-16S.** Coomassie-stained SDS-PAGE gel of a secondary purification of mutant CgArsR-16S. Much of the repressor eluted in the wash step of the initial purification, so samples were collated, dialysed to remove imidazole, and repurified with a lower imidazole wash step. Lanes corresponding to this dialysed sample (S), flow through (FT), wash (W), pre-eluate (PE) and eluate (E) stages are shown, with 5x corresponding to concentrated samples. The final glycerol stock, prepared by dilution of E1, is displayed alongside the wild-type CgArsR stock for comparison.

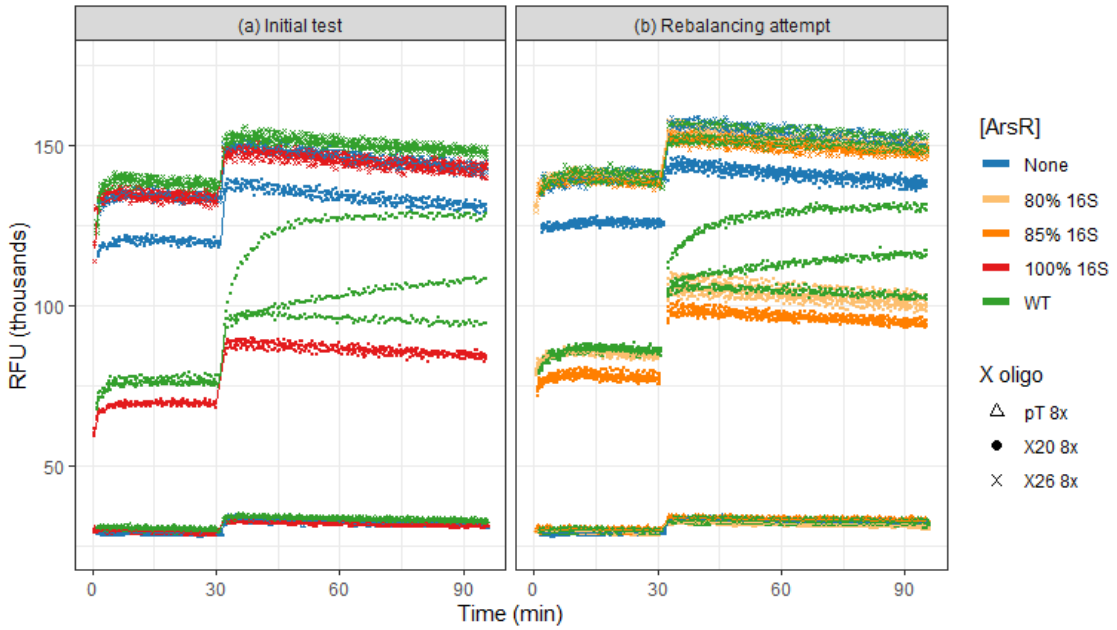


Figure A.10: **CgArsR-16S concentration balancing raw data.** Fluorescence time course displaying raw data for the assays shown in Figure 6.2.

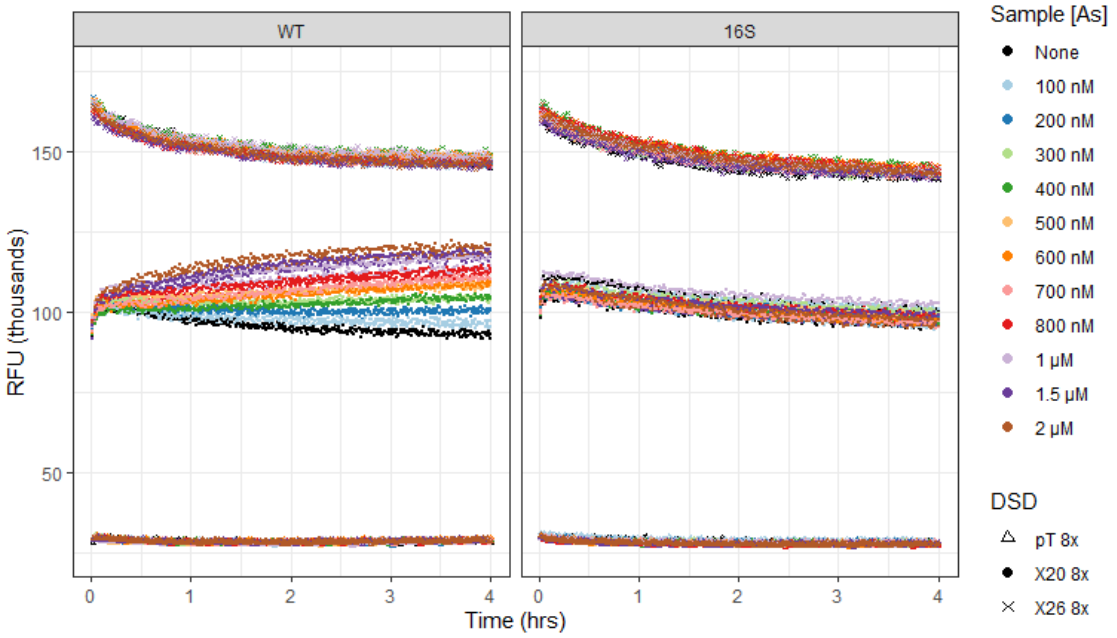


Figure A.11: **CgArsR-16S difference rate analysis raw data.** Fluorescence time course displaying raw data for the assay shown in Figure 6.3.

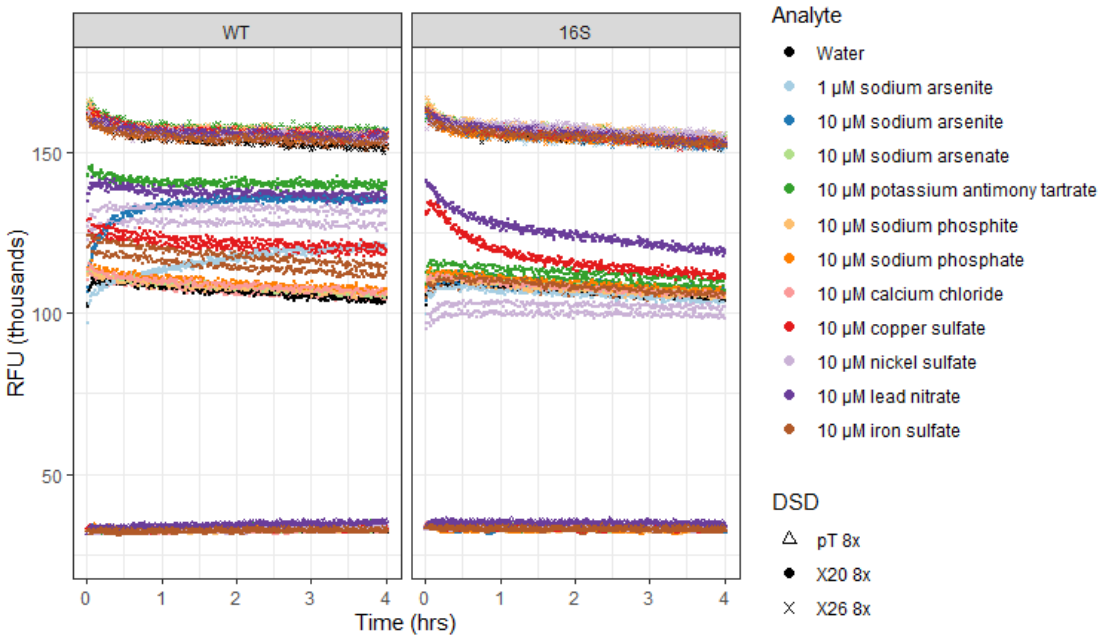


Figure A.12: **Medium concentration selectivity test raw data.** Fluorescence time course displaying raw data for the DSD7 assay in Figure 6.5, separated by ArsR.

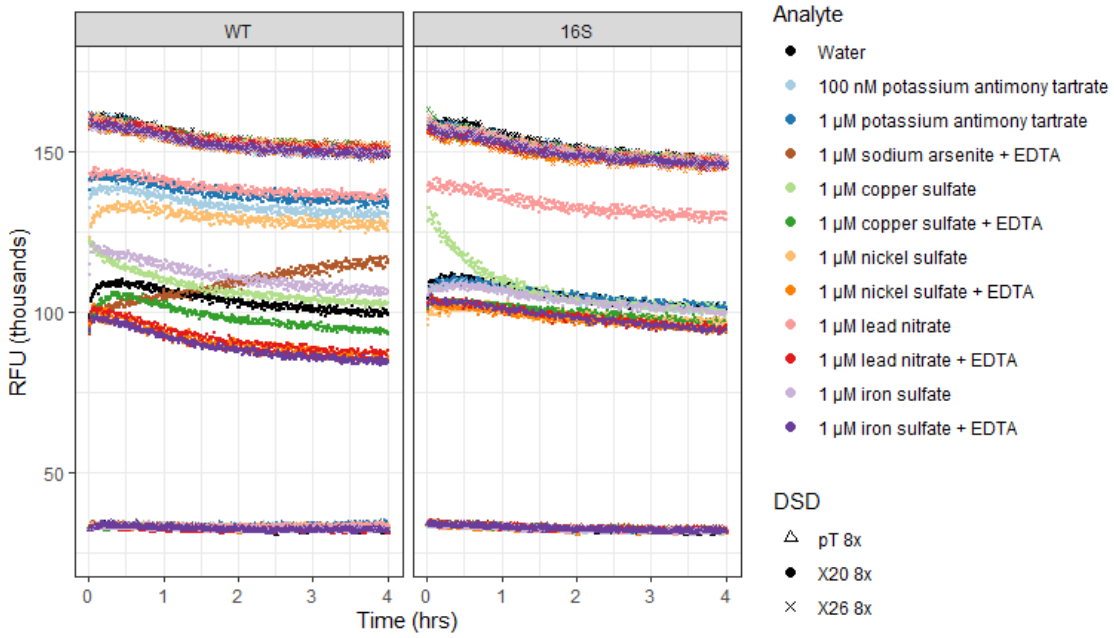


Figure A.13: **Low concentration selectivity test raw data.** Fluorescence time course displaying raw data for the DSD7 assay in Figure 6.6, separated by ArsR.

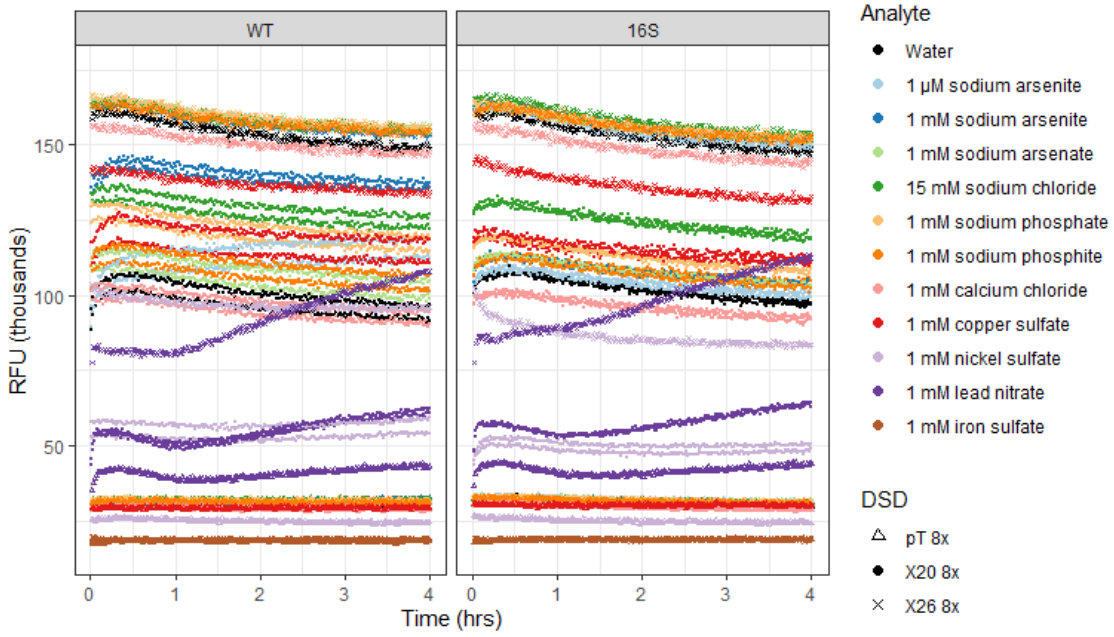


Figure A.14: **High concentration selectivity test raw data.** Fluorescence time course displaying raw data for the DSD7 assay in Figure A.15, separated by ArsR.

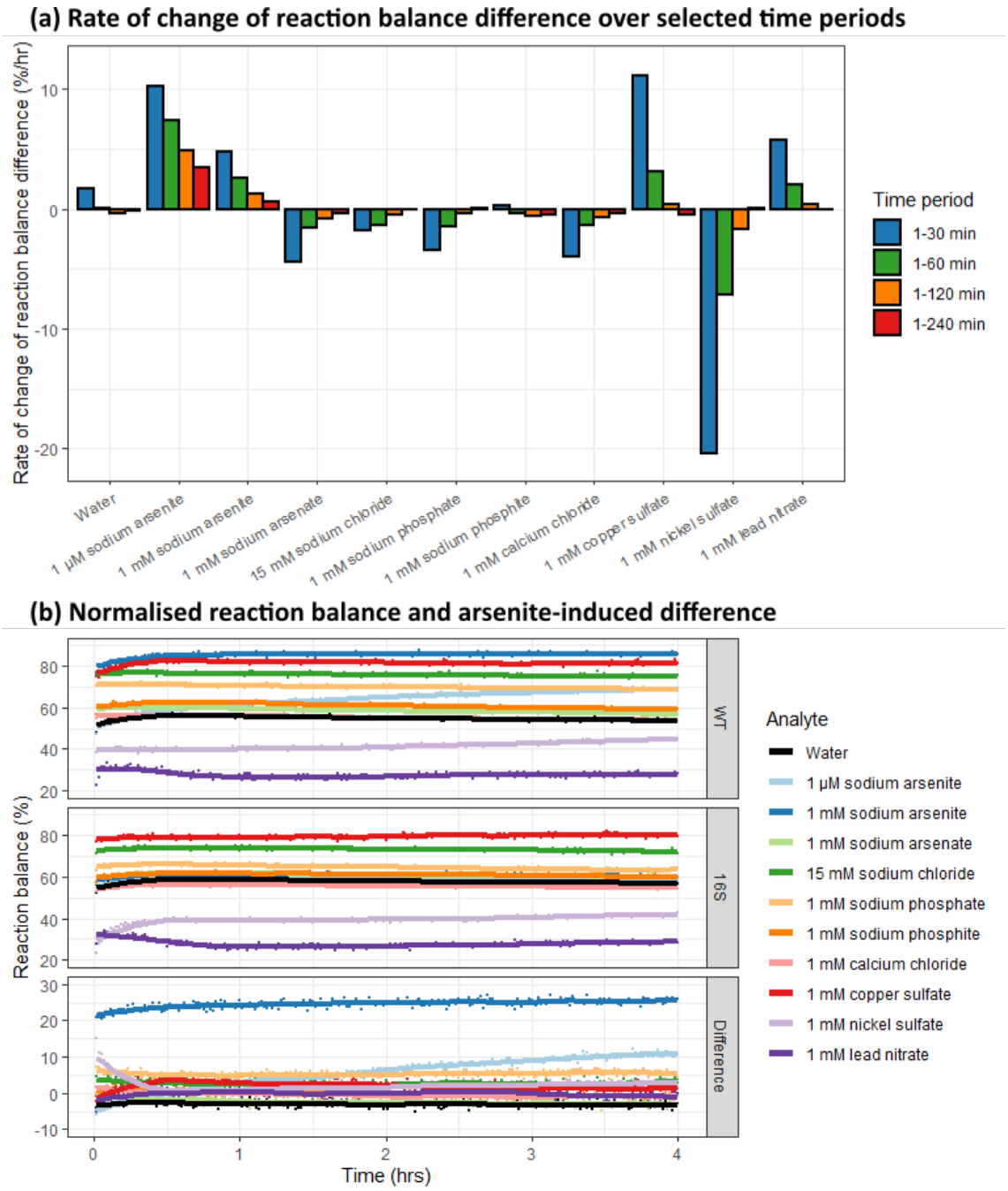


Figure A.15: **High concentration selectivity test normalised data.** Fluorescence time course displaying data for L6T1C-X20-CgArsR master mixes reacting with different analyte samples under the DSD7 protocol in buffer DB8. (a) Rate of change of the difference in reaction balance between the master mixes, calculated by creation of a linear model across different specified time periods, plotted by analyte. (b) Normalised reaction balances for WT and 16S mixes and their difference, responding to the samples over 4 hours. Points are displayed for two replicates per analyte, overlaid by a smoothed average line. Raw data are shown in Figure A.14.

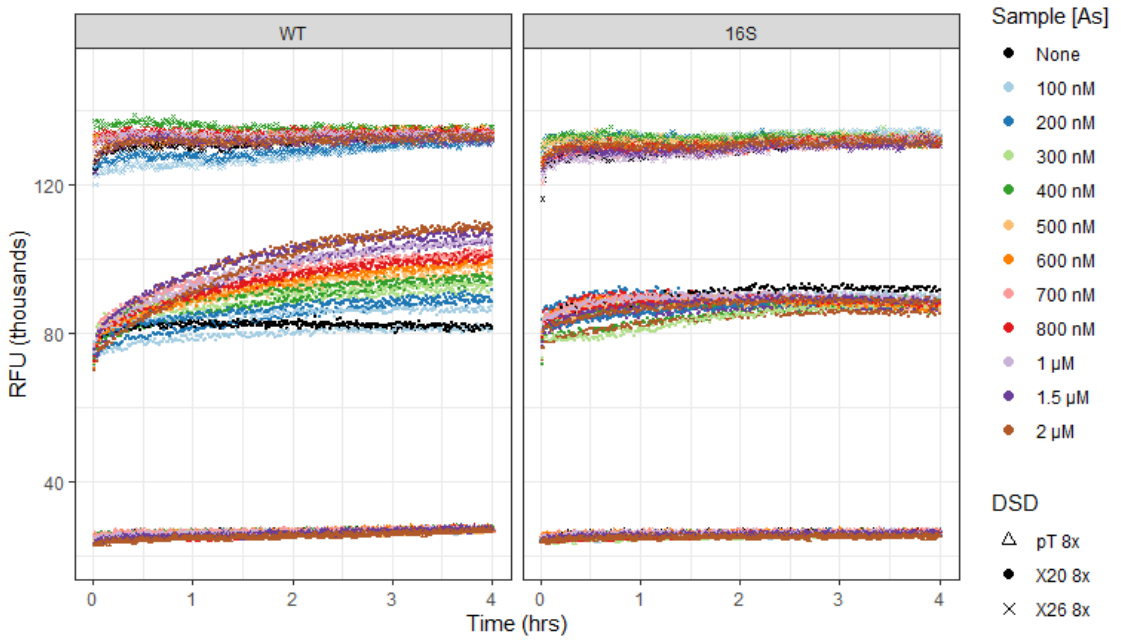


Figure A.16: **Lyophilised assay raw data.** Fluorescence time course displaying raw data for the lyophilised DSD7 assay in Figure 6.8, separated by ArsR.

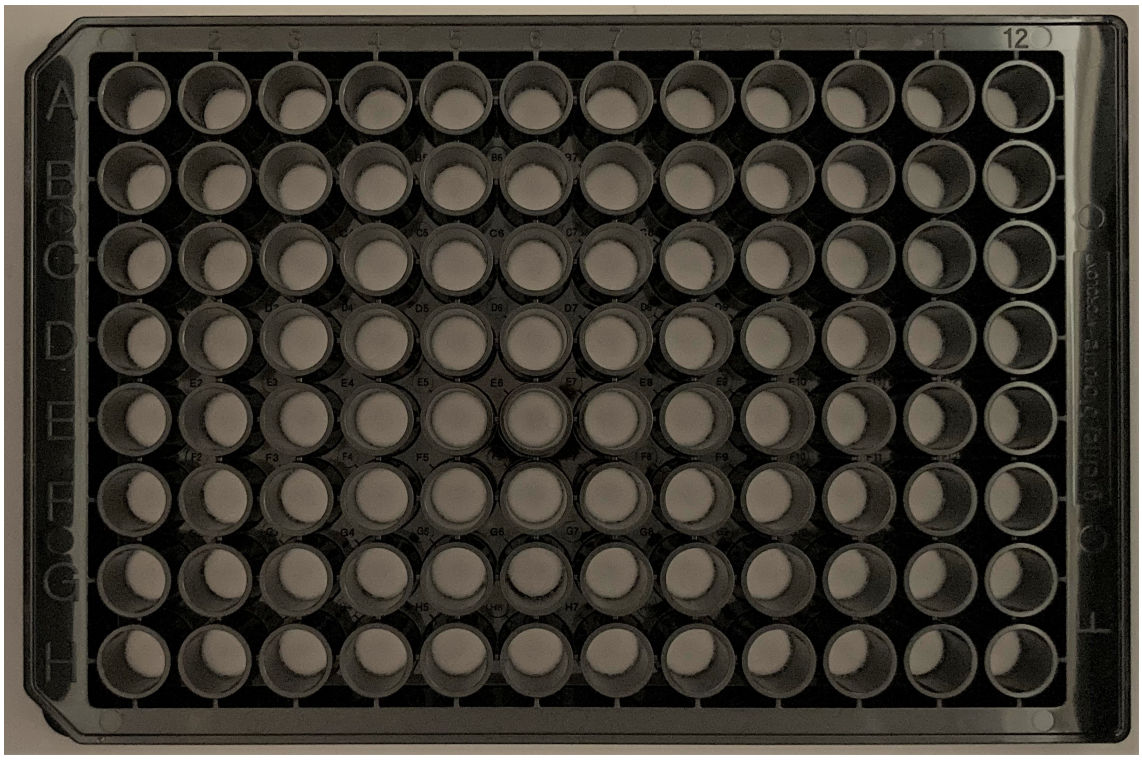


Figure A.17: **Assay lyophilised into stable white sheet.** Full 96-well plate of 50 µl L6T1C-CgArsR DSD reaction solutions were freeze-dried overnight, forming dry white powder sheets at the bottom of each well.

Bibliography

1. Kaur, H., Kumar, R., Babu, J. N. & Mittal, S. Advances in arsenic biosensor development - a comprehensive review. *Biosensors and Bioelectronics* **63,** 533–545 (2015).
2. Clarke, S. F. & Foster, J. R. A history of blood glucose meters and their role in self-monitoring of diabetes mellitus. *British Journal of Biomedical Science* **69,** 83–93 (2012).
3. Johnson, S. *100 Years of Human Chorionic Gonadotropin. Chapter 2.4: The home pregnancy test* (eds Cole, L. A. & Butler, S. A.) 107–121 (Elsevier, 2020).
4. Torjesen, I. Covid-19: How the UK is using lateral flow tests in the pandemic. *BMJ* **372** (2021).
5. Kitney, R. & Freemont, P. Synthetic biology-the state of play (2012).
6. Flores Bueso, Y. & Tangney, M. Synthetic Biology in the Driving Seat of the Bioeconomy. *Trends in Biotechnology* **35,** 373–378 (2017).
7. Khalil, A. S. & Collins, J. J. Synthetic biology: Applications come of age. *Nature Reviews Genetics* **11,** 367–379 (2010).
8. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
9. Martin, V. J., Piteral, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology* **21,** 796–802 (2003).
10. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2013).

11. Smanski, M. J., Zhou, H., Claesen, J., Shen, B., Fischbach, M. A., *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nature Reviews Microbiology* **14,** 135–149 (2016).

12. Tan, X., Letendre, J. H., Collins, J. J. & Wong, W. W. Synthetic biology in the clinic: engineering vaccines, diagnostics, and therapeutics. *Cell* **184,** 881–898 (2021).

13. Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., *et al.* Engineering Escherichia coli to see light. *Nature* **438,** 441–442 (2005).

14. Haellman, V. & Fussenegger, M. Synthetic Biology - Toward Therapeutic Solutions. *Journal of Molecular Biology* **428,** 945–962 (2016).

15. Bradley, R. W. & Wang, B. Designer cell signal processing circuits for biotechnology. *New Biotechnology* **32,** 635–643 (2015).

16. Johns, N. I., Blazejewski, T., Gomes, A. L. & Wang, H. H. Principles for designing synthetic microbial communities. *Current Opinion in Microbiology* **31,** 146–153 (2016).

17. Liu, C. C., Jewett, M. C., Chin, J. W. & Voigt, C. A. Toward an orthogonal central dogma. *Nature Chemical Biology* **14,** 103–106 (2018).

18. Taylor, A. I., Pinheiro, V. B., Smola, M. J., Morgunov, A. S., Peak-Chew, S., *et al.* Catalysts from synthetic genetic polymers. *Nature* **518,** 427–430 (2015).

19. Elowitz, M. & Lim, W. A. *Build life to understand it* 2010.

20. Kim, A. K., DeRose, R., Ueno, T., Lin, B., Komatsu, T., *et al.* Toward total synthesis of cell function: Reconstituting cell dynamics with synthetic biology. *Science Signaling* **9** (2016).

21. Bashor, C. J. & Collins, J. J. Understanding Biological Regulation Through Synthetic Biology. *Annual Review of Biophysics* **47,** 399–423 (2018).

22. Lee, J. W., Chan, C. T., Slomovic, S. & Collins, J. J. Next-generation biocontainment systems for engineered organisms. *Nature Chemical Biology* **14,** 530–537 (2018).

23. Lai, H. E., Canavan, C., Cameron, L., Moore, S., Danchenko, M., *et al.* Synthetic Biology and the United Nations. *Trends in Biotechnology* **37,** 1146–1151 (2019).

24. Khambhati, K., Bhattacharjee, G., Gohil, N., Braddick, D., Kulkarni, V., *et al.* Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Frontiers in Bioengineering and Biotechnology* **7,** 248 (2019).

25. Kitaoka, Y., Nishimura, N. & Niwano, M. Cooperativity of stabilized mRNA and enhanced translation activity in the cell-free system. *Journal of Biotechnology* **48,** 1–8 (1996).

26. Silverman, A. D., Kelley-Loughnane, N., Lucks, J. B. & Jewett, M. C. Deconstructing Cell-Free Extract Preparation for in Vitro Activation of Transcriptional Genetic Circuitry. *ACS Synthetic Biology* **8,** 403–414 (2019).

27. Kim, J., Campbell, A. S., de Ávila, B. E. F. & Wang, J. Wearable biosensors for healthcare monitoring. *Nature Biotechnology* **37,** 389–406 (2019).

28. Heineman, W. R. & Jensen, W. B. Leland C. Clark Jr. (1918–2005). *Biosensors and Bioelectronics* **21,** 1403–1404 (2006).

29. Su, L., Jia, W., Hou, C. & Lei, Y. Microbial biosensors: A review. *Biosensors and Bioelectronics* **26,** 1788–1799 (2011).

30. Virta, M., Lampinen, J. & Karp, M. A Luminescence-Based Mercury Biosensor. *Analytical Chemistry* **67,** 667–669 (1995).

31. Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., Daleykeyser, A., *et al.* Paper-based synthetic gene networks. *Cell* **159,** 940–954 (2014).

32. Lopatkin, A. J. & You, L. Synthetic biology looks good on paper. *Cell* **159,** 718–719 (2014).

33. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: De-novo-designed regulators of gene expression. *Cell* **159,** 925–939 (2014).

34. Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165,** 1255–1266 (2016).

35. Verosloff, M., Chappell, J., Perry, K. L., Thompson, J. R. & Lucks, J. B. PLANT-Dx: A Molecular Diagnostic for Point-of-Use Detection of Plant Pathogens. *ACS Synthetic Biology* **8,** 902–905 (2019).

36. Takahashi, M. K., Tan, X., Dy, A. J., Braff, D., Akana, R. T., *et al.* A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. *Nature Communications* **9** (2018).

37. Reboud, J., Xu, G., Garrett, A., Adriko, M., Yang, Z., *et al.* Paper-based microfluidics for DNA diagnostics of malaria in low resource underserved rural communities. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 4834–4842 (2019).

38. Hu, C., Kalsi, S., Zeimpekis, I., Sun, K., Ashburn, P., *et al.* Ultra-fast electronic detection of antimicrobial resistance genes using isothermal amplification and Thin Film Transistor sensors. *Biosensors and Bioelectronics* **96,** 281–287 (2017).

39. Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356,** 438–442 (2017).

40. Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Science* **360,** 439–444 (2018).

41. Lee, R. A., De Puig, H., Nguyen, P. Q., Angenent-Mari, N. M., Donghia, N. M., *et al.* Ultrasensitive CRISPR-based diagnostic for field-applicable detection of Plasmodium species in symptomatic and asymptomatic malaria. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 25722–25731 (2020).

42. Fozouni, P., Son, S., Díaz de León Derby, M., Knott, G. J., Gray, C. N., *et al.* Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell* **184,** 323–333.e9 (2021).

43. Özyurt, C., Üstükarcı, H., Evran, S. & Telefoncu, A. MerR-fluorescent protein chimera biosensor for fast and sensitive detection of Hg2+ in drinking water. *Biotechnology and Applied Biochemistry* **66,** 731–737 (2019).

44. Church, G. M., Gao, Y. & Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science* **337,** 1628 (2012).

45. Erlich, Y. & Zielinski, D. DNA Fountain enables a robust and efficient storage architecture. *Science* **355,** 950–954 (2017).

46. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467,** 167–173 (2010).

47. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491,** 249–253 (2012).

48. Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., *et al.* Genetic circuit design automation. *Science* **352** (2016).

49. Bradley, R. W., Buck, M. & Wang, B. Recognizing and engineering digital-like logic gates and switches in gene regulatory networks. *Current Opinion in Microbiology* **33,** 74–82 (2016).

50. Green, A. A., Kim, J., Ma, D., Silver, P. A., Collins, J. J., *et al.* Complex cellular logic computation using ribocomputing devices. *Nature* **548,** 117–121 (2017).

51. Kim, J., Zhou, Y., Carlson, P. D., Teichmann, M., Chaudhary, S., *et al.* De novo-designed translation-repressing riboregulators for multi-input cellular logic. *Nature Chemical Biology* **15,** 1173–1182 (2019).

52. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry* **3,** 103–113 (2011).

53. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314,** 1585–1588 (2006).

54. Zhang, D. Y., Turberfield, A. J., Yurke, B. & Winfree, E. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **318,** 1121–1125 (2007).

55. Soloveichik, D., Seelig, G. & Winfree, E. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 5393–5398 (2010).

56. Qian, L. & Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **332,** 1196–1201 (2011).

57. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society* **131,** 17303–17314 (2009).

58. Phillips, A. & Cardelli, L. A programming language for composable DNA circuits. *Journal of The Royal Society Interface* **6,** S419–S436 (2009).

59. Lakin, M. R., Youssef, S., Cardelli, L. & Phillips, A. Abstractions for DNA circuit design. *Journal of the Royal Society Interface* **9,** 470–486 (2012).

60. Lakin, M. R., Youssef, S., Polo, F., Emmott, S. & Phillips, A. Visual DSD: A design and analysis tool for DNA strand displacement systems. *Bioinformatics* **27,** 3211–3213 (2011).

61. Chen, Y.-J., Dalchau, N., Srinivas, N., Phillips, A., Cardelli, L., *et al.* Programmable chemical controllers made from DNA. *Nature Nanotechnology* **8,** 755–762 (2013).

62. Yordanov, B., Kim, J., Petersen, R. L., Shudy, A., Kulkarni, V. V., *et al.* Computational design of nucleic acid feedback control circuits. *ACS Synthetic Biology* **3,** 600–616 (2014).

63. Bi, S., Yue, S. & Zhang, S. Hybridization chain reaction: A versatile molecular tool for biosensing, bioimaging, and biomedicine. *Chemical Society Reviews* **46,** 4281–4298 (2017).

64. De Puig, H., Bosch, I., Collins, J. J. & Gehrke, L. Point-of-Care Devices to Detect Zika and Other Emerging Viruses. *Annual Review of Biomedical Engineering* **22,** 371–386 (2020).

65. Wang, J. S. & Zhang, D. Y. Simulation-guided DNA probe design for consistently ultraspecific hybridization. *Nature Chemistry* **7,** 545–553 (2015).

66. Li, B., Ellington, A. D. & Chen, X. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic Acids Research* **39** (2011).

67. Kinniburgh, D. G. & Smedley, P. L. *Arsenic contamination of groundwater in Bangladesh* tech. rep. (2001).

68. Chowdhury, A. M. R. Arsenic crisis in Bangladesh. *Scientific American* **291,** 86–91 (2004).

69. Oremland, R. S. & Stolz, J. F. The Ecology of Arsenic. *Science* **300,** 939–44 (2003).

70. Oremland, R. S. & Stolz, J. F. Arsenic, microbes and contaminated aquifers. *Trends in Microbiology* **13,** 45–49 (2005).

71. World Health Organisation. *Preventing Disease Through Healthy Environments* tech. rep. (2011), 1–6.

72. Barkat, A. & Hussam, A. *Provisioning of Arsenic-free Water in Bangladesh: A Human Rights Challenge* tech. rep. (2008), 1–12.

73. Smith, A. H., Lingas, E. O. & Rahman, M. Contamination of drinking-water by arsenic in Bangladesh: A public health emergency. *Bulletin of the World Health Organization* **78,** 1093–1103 (2000).

74. Flanagan, S. V., Johnston, R. B. & Zheng, Y. Arsenic in tube well water in Bangladesh: health and economic impacts and implications for arsenic mitigation. *Bulletin of the World Health Organization* **90,** 839–846 (2012).

75. World Health Organisation. *Arsenic in Drinking-water* tech. rep. (2011), 1–24.

76. Ahmed, F. *Arsenic Mitigation Technologies in South and East Asia* tech. rep. (2005), 44.

77. Sanger, C. R. & Black, O. F. The Quantitative Determination of Arsenic by the Gutzeit Method. *Proceedings of the American Academy of Arts and Sciences* **43,** 297 (1907).

78. Siegfried, K., Endes, C., Bhuiyan, A. F. M. K., Kuppardt, A., Mattusch, J., *et al.* Field testing of arsenic in groundwater samples of Bangladesh using a test kit based on lyophilized bioreporter bacteria. *Environmental Science and Technology* **46,** 3281–3287 (2012).

79. Merulla, D., Buffi, N., Beggah, S., Truffer, F., Geiser, M., *et al. Bioreporters and biosensors for arsenic detection. Biotechnological solutions for a world-wide pollution problem* 2013.

80. Maghsoudi, A. S., Hassani, S., Mirnia, K. & Abdollahi, M. Recent advances in nanotechnology-based biosensors development for detection of arsenic, lead, mercury, and cadmium. *International Journal of Nanomedicine* **16,** 803–832 (2021).

81. Novick, R. P. & Roth, C. Plasmid-linked resistance to inorganic salts in Staphylococcus aureus. *Journal of bacteriology* **95,** 1335–1342 (1968).

82. Hedges, R. W. & Baumberg, S. Resistance to arsenic compounds conferred by a plasmid transmissible between strains of Escherichia coli. *Journal of Bacteriology* **115,** 459–460 (1973).

83. Messens, J., Martins, J. C., Van Belle, K., Brosens, E., Desmyter, A., *et al.* All intermediates of the arsenate reductase mechanism, including an intramolecular dynamic disulfide cascade. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 8506–8511 (2002).

84. Kalia, K. & Joshi, D. N. in *Handbook of Toxicology of Chemical Warfare Agents* 1083–1100 (2009).

85. Busenlehner, L. S., Pennella, M. A. & Giedroc, D. P. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiology Reviews* **27,** 131–143 (2003).

86. Jia, X., Bu, R., Zhao, T. & Wu, K. Sensitive and specific whole-cell biosensor for arsenic detection. *Applied and Environmental Microbiology* **85** (2019).

87. Siegfried, K., Hahn-Tomer, S., Koelsch, A., Osterwalder, E., Mattusch, J., *et al.* Introducing simple detection of bioavailable arsenic at Rafaela (Santa Fe Province, Argentina) using the ARSOlux biosensor. *International Journal of Environmental Research and Public Health* **12,** 5465–5482 (2015).

88. Wang, B., Barahona, M. & Buck, M. A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosensors and Bioelectronics* **40,** 368–376 (2013).

89. Wang, B. & Buck, M. Rapid engineering of versatile molecular logic gates using heterologous genetic transcriptional modules. *Chemical Communications* **50,** 11642–11644 (2014).

90. Wang, B., Barahona, M. & Buck, M. Amplification of small molecule-inducible gene expression via tuning of intracellular receptor densities. *Nucleic Acids Research* **43,** 1955–1964 (2015).

91. Wan, X., Volpetti, F., Petrova, E., French, C., Maerkl, S. J., *et al.* Cascaded amplifying circuits enable ultrasensitive cellular sensors for toxic metals. *Nature Chemical Biology* **15,** 540–548 (2019).

92. Roggo, C. & van der Meer, J. R. *Miniaturized and integrated whole cell living bacterial sensors in field applicable autonomous devices* 2017.

93. Wu, Y., Zhan, S., Xing, H., He, L., Xu, L., *et al.* Nanoparticles assembled by aptamers and crystal violet for arsenic(iii) detection in aqueous solution based on a resonance Rayleigh scattering spectral assay. *Nanoscale* **4,** 6841 (2012).

94. Wu, Y., Zhan, S., Wang, F., He, L., Zhi, W., *et al.* Cationic polymers and aptamers mediated aggregation of gold nanoparticles for the colorimetric detection of arsenic(iii) in aqueous solution. *Chemical Communications* **48,** 4459 (2012).

95. Kim, M., Um, H. J., Bang, S., Lee, S. H., Oh, S. J., *et al.* Arsenic removal from Vietnamese groundwater using the arsenic-binding DNA aptamer. *Environmental Science and Technology* **43,** 9335–9340 (2009).

96. Zong, C. & Liu, J. The arsenic-binding aptamer cannot bind arsenic: critical evaluation of aptamer selection and binding. *Analytical Chemistry* **91,** 10887–10893 (2019).

97. Santini, J. M., Sly, L. I., Schnagl, R. D. & Macy, J. M. A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: Phylogenetic, physiological, and preliminary biochemical studies. *Applied and Environmental Microbiology* **66,** 92–97 (2000).

98. Corsini, P. M., Walker, K. T. & Santini, J. M. Expression of the arsenite oxidation regulatory operon in Rhizobium sp. str. NT-26 is under the control of two promoters that respond to different environmental cues. *MicrobiologyOpen*, e00567 (2017).

99. Badilla, C., Osborne, T. H., Cole, A., Watson, C., Djordjevic, S., *et al.* A new family of periplasmic-binding proteins that sense arsenic oxyanions. *Scientific Reports* **8,** 6282 (2018).

100. Cass, A. E., Santini, J. & Johnson, C. J. *Modified arsenite oxidase and a biosensor for detecting arsenite* 2013.

101. Male, K. B., Hrapovic, S., Santini, J. M. & Luong, J. H. Biosensor for arsenite using arsenite oxidase and multiwalled carbon nanotube modified electrodes. *Analytical Chemistry* **79,** 7831–7837 (2007).

102. Cass, A. E. *Dual mediator biosensor* 2017.

103. Poddar, N., Badilla, C., Maghool, S., Osborne, T. H., Santini, J. M., *et al.* Structural and Functional Investigation of the Periplasmic Arsenate-Binding Protein ArrX from Chrysiogenes arsenatis. *Biochemistry* **60,** 465–476 (2021).

104. Kawakami, Y., Siddiki, M. S. R., Inoue, K., Otabayashi, H., Yoshida, K., *et al.* Application of fluorescent protein-tagged trans factors and immobilized cis elements to monitoring of toxic metals based on in vitro protein-DNA interactions. *Biosensors and Bioelectronics* **26,** 1466–1473 (2010).

105. Siddiki, M. S. R., Kawakami, Y., Ueda, S. & Maeda, I. Solid phase biosensors for arsenic or cadmium composed of a trans factor and cis element complex. *Sensors* **11,** 10063–10073 (2011).

106. Siddiki, M. S., Ueda, S. & Maeda, I. Fluorescent bioassays for toxic metals in milk and yoghurt. *BMC Biotechnology* **12,** 76 (2012).

107. Siddiki, M. S. R., Shimoaoki, S., Ueda, S. & Maeda, I. Thermoresponsive magnetic nano-biosensors for rapid measurements of inorganic arsenic and cadmium. *Sensors (Switzerland)* **12,** 14041–14052 (2012).

108. Taniguchi, M., Siddiki, M. S. R., Ueda, S. & Maeda, I. Mercury (II) sensor based on monitoring dissociation rate of the trans-acting factor MerR from cis-element by surface plasmon resonance. *Biosensors and Bioelectronics* **67,** 309–314 (2015).

109. Arruda, L. M., Monteiro, L. M. O. & Silva-Rocha, R. The Chromobacterium violaceum ArsR arsenite repressor exerts tighter control on its cognate promoter than the Escherichia coli system. *Frontiers in Microbiology* **7,** 1851 (2016).

110. Qin, J., Fu, H. L., Ye, J., Bencze, K. Z., Stemmler, T. L., *et al.* Convergent evolution of a new arsenic binding site in the ArsR/SmtB family of metalloregulators. *Journal of Biological Chemistry* **282,** 34346–34355 (2007).

111. Ye, J., Kandegedara, A., Martin, P. & Rosen, B. P. Crystal structure of the Staphylococcus aureus pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. *Journal of Bacteriology* **187,** 4214–4221 (2005).

112. Valle, S., Lale, R. & Walker, J. M. *DNA Cloning and Assembly Methods* (2014).

113. Carlin, A., Shi, W., Dey, S. & Rosen, B. P. The ars operon of Escherichia coli confers arsenical and antimonial resistance. *Journal of Bacteriology* **177,** 981–986 (1995).

114. Diorio, C., Cai, J., Marmor, J., Shinder, R. & DuBow, M. S. An Escherichia coli chromosomal ars operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. *Journal of Bacteriology* **177,** 2050–2056 (1995).

115. Cai, J., Salmon, K. & DuBow, M. S. A Chromosomal Ars Operon Homolog of Pseudomonas-Aeruginosa Confers Increased Resistance to Arsenic and Antimony in Escherichia-Coli. *Microbiology* **Vol 144, I,** 2705 (1998).

116. Xu, C., Shi, W. & Rosen, B. P. The chromosomal arsR gene of Escherichia coli encodes a trans-acting metalloregulatory protein. *Journal of Biological Chemistry* **271,** 2427–2432 (1996).

117. Xu, C. & Rosen, B. P. Dimerization is essential for DNA binding and repression by the ArsR metalloregulatory protein of Escherichia coli. *Journal of Biological Chemistry* **272,** 15734–15738 (1997).

118. Christen, E. H., Karlsson, M., Kämpf, M. M., Weber, C. C., Fussenegger, M., *et al.* A general strategy for the production of difficult-to-express inducer-dependent bacterial repressor proteins in Escherichia coli. *Protein Expression and Purification* **66,** 158–164 (2009).

119. San Francisco, M. J. D., Hope, C. L., Owolabi, J. B., Tisa, L. S. & Rosen, B. P. Identification of the metalloregulatory element of the plasmid-encoded arsenical resistance operon. *Nucleic Acids Research* **18,** 619–624 (1990).

120. Wu, J. & Rosen, B. P. The ArsR protein is a trans-acting regulatory protein. *Molecular Microbiology* **5,** 1331–1336 (1991).

121. Wu, J. & Rosen, B. P. Metalloregulated expression of the ars operon. *Journal of Biological Chemistry* **268,** 52–58 (1993).

122. Shi, W., Wu, J. & Rosen, B. P. Identification of a putative metal binding site in a new family of metalloregulatory proteins. *Journal of Biological Chemistry* **269,** 19826–19829 (1994).

123. Shi, W., Dong, J., Scott, R. A., Ksenzenko, M. Y. & Rosen, B. P. The role of arsenic-thiol interactions in metalloregulation of the ars operon. *Journal of Biological Chemistry* **271,** 9291–9297 (1996).

124. Sato, T. & Kobayashi, Y. The ars operon in the skin element of Bacillus subtilis confers resistance to arsenate and arsenite. *Journal of Bacteriology* **180,** 1655–1661 (1998).

125. Rosenstein, R., Peschel, A., Wieland, B. & Gotz, F. Expression and Regulation of the Antimonite, Arsenite, and Arsenate Resistance Operon of Staphylococcus xylosus Plasmid pSX267. *Journal of Bacteriology* **174,** 3676–3683 (1992).

126. Rosenstein, R., Nikoleit, K. & Götz, F. Binding of ArsR, the repressor of the Staphylococcus xylosus (pSX267) arsenic resistance operon to a sequence with dyad symmetry within the ars promoter. *Molecular & general genetics* **242,** 566–72 (1994).

127. Patel, P. C., Goulhen, F., Boothman, C., Gault, A. G., Charnock, J. M., *et al.* Arsenate detoxification in a Pseudomonad hypertolerant to arsenic. *Archives of Microbiology* **187,** 171–183 (2007).

128. Ordóñez, E., Letek, M., Valbuena, N., Gil, J. A. & Mateos, L. M. Analysis of genes involved in arsenic resistance in Corynebacterium glutamicum ATCC 13032. *Applied and Environmental Microbiology* **71,** 6206–6215 (2005).

129. Letek, M., Valbuena, N., Ramos, A., Ordóñez, E., Gil, J. A., *et al.* Characterization and use of catabolite-repressed promoters from gluconate genes in Corynebacterium glutamicum. *Journal of bacteriology* **188,** 409–23 (2006).

130. Ordóñez, E., Thiyagarajan, S., Cook, J. D., Stemmler, T. L., Gil, J. A., *et al.* Evolution of metal(loid) binding sites in transcriptional regulators. *Journal of Biological Chemistry* **283,** 25706–25714 (2008).

131. Zhang, Y. B., Monchy, S., Greenberg, B., Mergeay, M., Gang, O., *et al.* ArsR arsenic-resistance regulatory protein from Cupriavidus metallidurans CH34. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **96,** 161–170 (2009).

132. Butcher, B. G., Deane, S. M. & Rawlings, D. E. The chromosomal arsenic resistance genes of Thiobacillus ferrooxidans have an unusual arrangement and confer increased arsenic and antimony resistance to Escherichia coli. *Applied and Environmental Microbiology* **66,** 1826–1833 (2000).

133. Butcher, B. G. & Rawlings, D. E. The divergent chromosomal ars operon of Acidithiobacillus ferrooxidans is regulated by an atypical ArsR protein. *Microbiology* **148,** 3983–3992 (2002).

134. Chen, J., Nadar, V. S. & Rosen, B. P. A novel MAs(III)-selective ArsR transcriptional repressor. *Molecular Microbiology* **106,** 469–478 (2017).

135. Suzuki, K., Wakao, N., Kimura, T., Sakka, K. & Ohmiya, K. Metalloregulatory properties of the ArsR and ArsD repressors of Acidiphilium multivorum AIU 301. *Journal of Fermentation and Bioengineering* **85,** 623–626 (1998).

136. Suzuki, K., Wakao, N., Kimura, T., Sakka, K. & Ohmiya, K. Expression and regulation of the arsenic resistance operon of Acidiphilium multivorum AIU 301 plasmid pKW301 in Escherichia coli. *Applied and Environmental Microbiology* **64,** 411–418 (1998).

137. Unger, B. & Hillen, W. Nucleotide sequence of the repressor gene of the RAl tetracycline resistance determinant: structural and functional comparison with three related Tet repressor genes. *Nucleic acids research* **12,** 7693–7703 (1984).

138. Hinrichs, W., Kisker, C., Düvel, M., Müller, A., Tovar, K., *et al.* Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science (New York, N.Y.)* **264,** 418–420 (1994).

139. Ettner, N., Müller, G., Berens, C., Backes, H., Schnappinger, D., *et al.* Fast large-scale purification of tetracycline repressor variants from overproducing Escherichia coli strains. *Journal of Chromatography A* **742,** 95–105 (1996).

140. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature structural biology* **7,** 215–9 (2000).

141. Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., *et al.* The TetR Family of Transcriptional Repressors. *Microbiology and Molecular Biology Reviews* **69,** 326–356 (2005).

142. Volkers, G., Petruschka, L. & Hinrichs, W. Recognition of drug degradation products by target proteins: Isotetracycline binding to tet repressor. *Journal of Medicinal Chemistry* **54,** 5108–5115 (2011).

143. Werten, S., Dalm, D., Palm, G. J., Grimm, C. C. & Hinrichs, W. Tetracycline repressor allostery does not depend on divalent metal recognition. *Biochemistry* **53,** 7990–7998 (2014).

144. Rosinski, J. A. & Atchley, W. R. Molecular evolution of helix-turn-helix proteins. *Journal of Molecular Evolution* **49,** 301–309 (1999).

145. Nelson, M., Hillen, W. & Greenwald, R. A. *Tetracyclines in Biology, Chemistry and Medicine* 336 (Birkhauser Verlag, 2001).

146. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols* **2,** 1849–1861 (2007).

147. Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. & Selvin, P. R. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Analytical Biochemistry* **273,** 73–80 (1999).

148. Han, J. C. & Han, G. Y. A procedure for quantitative determination of tris(2- carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Analytical Biochemistry* **220,** 5–10 (1994).

149. Zhang, J. X., Fang, J. Z., Duan, W., Wu, L. R., Zhang, A. W., *et al.* Predicting DNA hybridization kinetics from sequence. *Nature Chemistry* **10,** 91–98 (2018).

150. Stellwagen, E., Abdulla, A., Dong, Q. & Stellwagen, N. C. Electrophoretic mobility is a reporter of hairpin structure in single-stranded DNA oligomers. *Biochemistry* **46,** 10931–10941 (2007).

151. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Research* **34,** 5402–5415 (2006).

152. Zhang, X.-B., Kong, R.-M. & Lu, Y. Metal Ion Sensors Based on DNAzymes and Related DNA Molecules. *Annual Review of Analytical Chemistry* **4,** 105–128 (2011).

153. Noble, J. E., Wang, L., Cole, K. D. & Gaigalas, A. K. The effect of overhanging nucleotides on fluorescence properties of hybridising oligonucleotides labelled with Alexa-488 and FAM fluorophores. *Biophysical Chemistry* **113,** 255–263 (2005).

154. Dauty, E. & Verkman, A. S. Molecular crowding reduces to a similar extent the diffusion of small solutes and macromolecules: Measurement by fluorescence correlation spectroscopy. *Journal of Molecular Recognition* **17,** 441–447 (2004).

155. Ge, X., Luo, D. & Xu, J. Cell-free protein expression under macromolecular crowding conditions. *PLoS ONE* **6,** e28707 (2011).

156. Tokuriki, N., Kinjo, M., Negi, S., Hoshino, M., Goto, Y., *et al.* Protein folding by the effects of macromolecular crowding. *Protein Science* **13,** 125–133 (2004).

157. Zhou, H. X. Rate theories for biologists. *Quarterly Reviews of Biophysics* **43,** 219–293 (2010).

158. Ganji, M., Docter, M., Le Grice, S. F. & Abbondanzieri, E. A. DNA binding proteins explore multiple local configurations during docking via rapid rebinding. *Nucleic Acids Research* **44,** 8376–8384 (2016).

159. Kwon, H., Park, S., Lee, S., Lee, D. K. & Yang, C. H. Determination of binding constant of transcription factor AP-1 and DNA: Application of inhibitors. *European Journal of Biochemistry* **268,** 565–572 (2001).

160. Wang, Y., Guo, L., Golding, I., Cox, E. C. & Ong, N. P. Quantitative transcription factor binding kinetics at the single-molecule level. *Biophysical Journal* **96,** 609–620 (2009).

161. Long, X., Wang, X., Guo, X. & He, M. A review of removal technology for antimony in aqueous solution (2019).

162. Flora, S. J. S. & Pachauri, V. Chelation in Metal Intoxication. *Int. J. Environ. Res. Public Health* **7,** 7 (2010).

163. Bjørklund, G., Oliinyk, P., Lysiuk, R., Rahaman, M. S., Antonyak, H., *et al.* Arsenic intoxication: general aspects and chelating agents. *Archives of Toxicology* **94,** 1879–1897 (2020).

164. Aksoy, N., Şimşek, C. & Gunduz, O. Groundwater contamination mechanism in a geothermal field: A case study of Balcova, Turkey. *Journal of Contaminant Hydrology* **103,** 13–28 (2009).

165. Pardee, K. Perspective: Solidifying the impact of cell-free synthetic biology through lyophilization. *Biochemical Engineering Journal* **138,** 91–97 (2018).

166. Land, K. J., Boeras, D. I., Chen, X. S., Ramsay, A. R. & Peeling, R. W. REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. *Nature Microbiology* **4,** 46–54 (2019).

167. Jung, J. K., Alam, K. K., Verosloff, M. S., Capdevila, D. A., Desmau, M., *et al.* Cell-free biosensors for rapid detection of water contaminants. *Nature biotechnology*, accepted (2020).

168. Hsing, I. M. & Rodríguez-Serrano, A. F. Allosteric regulation of DNA circuits enables minimal and rapid biosensors of small molecules. *ACS Synthetic Biology* **10,** 371–378 (2021).

169. Gihring, T. M., Druschel, G. K., Mccleskey, R. B., Hamers, R. J. & Banfield, J. F. Rapid arsenite oxidation by Thermus aquaticus and Thermus thermophilus: Field and laboratory investigations. *Environmental Science and Technology* **35,** 3857–3862 (2001).

170. Wang, T., Zhou, L., Bai, S., Zhang, Z., Li, J., *et al.* Ultraspecific electrochemical DNA biosensor by coupling spontaneous cascade DNA branch migration and dual-signaling sensing strategy. *Biosensors and Bioelectronics* **78,** 464–470 (2016).

171. Huang, X., Li, J., Zhang, Q., Chen, S., Xu, W., *et al.* A protease-free and signal-on electrochemical biosensor for ultrasensitive detection of lead ion based on GR-5 DNAzyme and catalytic hairpin assembly. *Journal of Electroanalytical Chemistry* **816,** 75–82 (2018).

172. Chatterjee, G., Dalchau, N., Muscat, R. A., Phillips, A. & Seelig, G. A spatially localized architecture for fast and modular DNA computing. *Nature Nanotechnology* **12,** 920–927 (2017).