Engineering of Second-Generation Acoustic Reporter Genes

Thesis by
Robert Cooper Hurt

In Partial Fulfillment of the Requirements for the Degree
of
Doctor of Philosophy in Neurobiology

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
(Defended March 23, 2023)

ACKNOWLEDGEMENTS

It’s hard to imagine adequately thanking everyone who has contributed to my personal and professional success up to this point. Here is my attempt to acknowledge at least those whose influence has substantially shaped my path thus far.

First, I must extensively thank my PhD advisor, Mikhail Shapiro. Mikhail has been an unwavering source of scientific and professional support, and an inexhaustible source of knowledge and creativity. Mikhail is one of the most driven and passionate people I’ve met, and his optimism and confidence inspire the same in others. It would likely take several PhDs for me to learn everything that I could from him, but I am comforted to know that he will remain a source of guidance and support for the rest of my career.

I also owe a great debt of thanks to my committee members: Henry Lester, Doug Rees, and Richard Murray. When I selected my committee members, I looked for professors with relevant and extensive expertise, but equally importantly, I looked for ones who were admirable people independent of their professional accomplishments. They of course contributed in important ways to the scientific content of this work, but that isn’t the primary thing I’ll take from my interactions with them; instead, it’s their ability to achieve great things while treating others with great compassion, understanding, patience, and respect that has impacted me the most. I’ve always left interactions with them feeling lifted up, and they will forever be sterling reminders of how I seek to treat others in my work and life.

The co-first authors on my publications—Zhiyang Jin, Marjorie Buss, and Tom Duan—have been critical to bringing those projects to fruition, and it’s been a pleasure to work closely and learn with them.

At Caltech, I’ve felt tremendously supported at almost every step of my journey. Caltech is an extremely special place, as I learned quickly from Henry Lester, its greatest fanatic. Its tight-knit, collaborative, vibrant atmosphere has never left me wanting for friendship, intellectual stimulation, and almost any personal or scientific resource I could need. At Caltech, I’ve benefitted greatly from various types of institutional support, including the Biotechnology Leadership Training Program directed by Frances Arnold and run by Kim Mayer, Julie Kelly, and Ngozi Eze; the Center for Environmental-Microbial Interactions led by Victoria Orphan; and the Neurobiology option, led by Markus Meister, Viviana Gradinaru, Henry Lester, and Marianne Bronner.

I would also like to thank my rotation supervisors, David Prober and Viviana Gradinaru, as well as my rotation mentors, Grigorios Oikonomou, Nick Flytzanis, and Di Wu. They not only taught valuable technical skills, but also supported me as I tried to find my path in the beginning of my Ph.D.

A great number of friends have supported me on this journey as well. First, I want to thank my "second family" in Pasadena: Joe, Neena, Pee Wee, and Penny. In addition to being my closest friends, they have been tremendous sources of love and support in all the aspects of my life, and have celebrated my successes and comforted me in my disappointments. I also want to thank the other members of my Neurobiology cohort—Yu-Li, Michael, and Ke—for making the difficult first year of our PhDs as enjoyable an experience as it could have been.

I'm also indebted to Mohamed Abedi, Di Wu, Dan Piraner, Bill Ling, and Hunter Davis for setting the culture of the Shapiro Lab in its earliest days as one of support, inclusion, and rigorous scientific discussion, but also as one of fun, mischief, and levity. While the latter elements have faded as most of you have left and the rest of us have grown older, you can be proud that the Shapiro Lab whose culture you helped define is still attracting great new people.

The fun, rotating cast of characters in my office, including Zhiyang, Claire, Anu, Hunter, Danny, Yuxing, and Arash, as well as a few pets and plants that are no longer with us, have made for a fun, supportive, and intellectually stimulating atmosphere where anyone would be lucky to work. Many other members of the Shapiro Lab have also contributed to making it a fun place to work and play, including Avinoam, The Davids, Ernesto, Ishaan, Abdullah, Erik, Sangjin, and Rohit.

I owe a great debt of gratitude to the amazing undergraduate and rotation students who have worked with me during my time at Caltech. Ramya Deshpande, Katie Wong, and Mohamed Soufi have been some of my closest collaborators, and in addition to producing excellent work for our papers, they have helped to remind me of the excitement I felt when starting out in science. They, along with Saren Daghlian, Christian Totoiu, Noah Epstein, and Hao Shen, have shown me how rewarding mentoring fantastic students can be.

It's hard to imagine how much extra work I would have had to do during my PhD if not for the Shapiro Lab's management team of Rosie, Margaret, Ruby, Gabrielle, Audrey, Pierina, and Dina. Thank you for keeping things running so smoothly, even in the face of supply chain issues and demanding personal obligations, and for being such team players and helping with all kinds of issues that you probably never expected to face.

Further, the Shapiro Lab members not mentioned above, including many with whom I only interacted in passing, have been excellent, both as scientists and as people. In the early days of the Lab, we spent a lot of time thinking about how we could attract the best people and train them to be careful and creative scientists, as well as generous and understanding coworkers. I feel that we've generally met those goals, and I'm very happy to see the new generation of Lab members exercising their prerogative to set the Lab's culture in ways that will serve them best.

Prior to Caltech, I had the good fortune of knowing and working with a number of brilliant and kind people at Emory University, both in my classes and in research labs. I want to thank Keith Easterling, Joe Manns, and Sam Sober both for being fantastic teachers, and for always being transparent about the realities of pursuing a career in research.

The single biggest influence on me during my time at Emory was Orion Keifer, the MD/PhD student I worked with in Kerry Ressler's lab. Orion taught me by example how to think like the consummate scientist, in ways that my classes had never forced me to do; while I have encountered and recognized these skills in others since then, I've never heard of anyone acquiring them except through close one-on-one training by a skilled and dedicated mentor, such as Orion. Orion's quick wit and great sense of humor made working with him fun, and his generosity with his time has been a trait that I have tried to incorporate into my own mentorship style.

Kerry Ressler himself was also a tremendous influence on me, both as a scientist and as a person. Kerry is one of the sharpest and kindest people I've met, and multiple people I encountered at Emory told me that they wanted to be like him when they grew up. I can't think of a better role model than him in every respect, and it was a great privilege to know and work with him. One of the most important development experiences for me in Kerry's lab was getting to work independently on CLARITY, a technique published less than a year earlier that no one in the lab had experience performing. Even though I'm sure he never imagined that I would get it to work, he supported me as I tinkered with it and gained valuable experience working independently on a project where I could rarely turn to others for answers. Through that project I got to work closely with Professor Tig Rainnie and with the talented microscopists Laura Fox-Goharioon and Neil Anthony, all of whom treated me as a peer despite my being a minimally-supervised undergrad, which helped to build my confidence as an independent researcher.

Finally, Hayk Harutyunyan was an excellent mentor and role model, and took a chance on hiring me to his fledgling nanophotonics lab considering my minimal physics background. I was very impressed by how well he made time for his lab members and took experimental failures in stride, despite being a brand new assistant professor.

Most importantly, I want to thank my family. I couldn't have asked for a more consistently loving and supportive group of people, who have taught me much more than I could ever learn in a PhD. Their excitement about my successes is a great source of pride, and their constant encouragement has helped me through the low points of this journey. I can't imagine going through this process without them.

Finally, there are many, many others in need of thanks, who helped me in ways large and small, whether they realized it or not.

ABSTRACT

A major outstanding challenge in the fields of biological research, synthetic biology, and cell-based medicine is visualizing the functions of natural and engineered cells noninvasively inside opaque organisms. Ultrasound imaging has the potential to address this challenge as a widely available technique with a tissue penetration of several centimeters and spatial resolution below 100 µm. Recently, the first genetically encoded acoustic reporters were developed based on bacterial gas vesicles (GVs) to link ultrasound signals to molecular and cellular function. However, the properties of these first-generation acoustic reporter genes (ARGs) resulted in limited sensitivity and specificity for imaging gene expression *in vivo*.

The goal of my thesis work has been to engineer second-generation ARGs with improved acoustic and expression phenotypes compared to the existing first-generation constructs. I took two complementary engineering approaches to developing these constructs: homolog screening and directed evolution, sometimes referred to as the “nature and nurture” of protein engineering. The resulting constructs offer major qualitative and quantitative improvements, including much stronger ultrasound contrast, the ability to produce nonlinear signals distinguishable from background tissue *in vivo*, stable long-term expression, and compatibility with *in vitro* multiplexed imaging. In collaboration with others in the lab, we demonstrate the capabilities of these next-generation ARGs by imaging *in situ* gene expression in mouse models of breast cancer and tumor-homing therapeutic bacteria, noninvasively revealing the unique spatial distributions of tumor growth and colonization by therapeutic cells in living subjects and providing real-time guidance for interventions such as needle biopsies.

This thesis is organized as follows: in the first two chapters, I introduce the key background needed to understand both the importance and properties of ARGS, and how they have been and could be engineered. In the next two chapters, I detail specific efforts to engineer them—one involving the construction of a high-throughput, semi-automated setup for acoustic phenotyping of cells and its application to ARG directed evolution, and another involving the screening of several GV cluster homologs to identify ones suitable for use as improved ARGs. Finally, I conclude with insights gleaned from these two ARG engineering projects and suggestions for future ones.

The approaches, results, and ideas presented in this thesis represent the current state-of-the-art in ARG engineering and application. While recent technology development in this field has unlocked exciting new use cases for ARGs in noninvasive biological imaging, most of their potential for basic science and disease diagnosis and treatment has yet to be realized.

PUBLISHED CONTENT AND CONTRIBUTIONS

Hurt R.C., Jin Z., Soufi M., Wong K., Sawyer D., Deshpande R., Shen H., Mittelstein D.R., Shapiro M.G. "Directed evolution of acoustic reporter genes using high-throughput acoustic screening." *In preparation*.

R.C.H. contributed to the study design, experiments, data analysis, and writing.

Hurt R.C.*, Buss M.T.*, Duan M.*, Wong K., You M.Y., Sawyer D.P., Swift M.B., Dutka P., Barturen-Larrea P., Mittelstein D.R., Jin Z., Abedi M.H., Farhadi A., Deshpande R., Shapiro M.G. "Genomically mined acoustic reporter genes for real-time *in vivo* monitoring of tumors and tumor-homing bacteria." *Nature Biotechnology*. 2023. DOI: 10.1038/s41587-022-01581-y

R.C.H. contributed to the study design, experiments, data analysis, and writing.

Dutka P., Metskas L.A., **Hurt R.C.**, Salahshoor H., Wang T.-Y., Malounda D., Lu G., Chou T.-F., Shapiro M.G., Jensen G.J.. "Structure of *Anabaena flos-aquae* gas vesicles revealed by cryo-ET." *bioRxiv*. 2022. DOI: 10.1101/2022.06.21.496981

R.C.H. contributed to the experiments performed in this study.

Yoo S., Mittelstein D.R., **Hurt R.C.**, Lacroix J., Shapiro M.G. "Focused ultrasound excites neurons via mechanosensitive calcium accumulation and ion channel amplification." *Nature Communications*. 2022. DOI: 10.1038/s41467-022-28040-1

R.C.H. contributed to the experiments performed in this study.

Bar-Zion A., Nourmahnad A., Mittelstein D.R., Shivaei S., Yoo S., Buss M.T., **Hurt R.C.**, Malounda D., Abedi M.H., Lee-Gosselin A., Swift M.B., Maresca D., Shapiro M.G. "Genetically encoded acoustic detonators for remote controlled mechanotherapy." *Nature Nanotechnology*. 2021. DOI: 10.1038/s41565-021-00971-8

R.C.H. contributed to the experiments performed in this study.

Dutka P., Malounda D., Metskas L.A., Chen S., **Hurt R.C.**, Lu G.J., Jensen G.J., Shapiro M.G. "Measuring gas vesicle dimensions by electron microscopy." *Protein Science*. 2021. DOI: 10.1002/pro.4056

Rabut C., Yoo S., **Hurt R.C.**, Jin Z., Li R., Guo H., Ling B.L., Shapiro MG. "Ultrasound Technologies for Imaging and Modulating Neural Activity." *Neuron*. 2020. DOI: 10.1016/j.neuron.2020.09.003

R.C.H. contributed to the writing of this review.

TABLE OF CONTENTS

Acknowledgements……………………………………………………………........................iii
Abstract …………………………………………………………………………..................... iv
Published Content and Contributions………………………………….............. vi
Table of Contents…………………………………………………….................…. vii
List of Illustrations and/or Tables……………………………………… .... …viii
Chapter I: Molecular Imaging of Acoustic Reporter Genes 1
- 1.1 The Importance of Noninvasive Molecular Imaging... 1
- 1.2 Existing Methods for Noninvasive Molecular Imaging 2
- 1.3 Gas Vesicles as Ultrasound Contrast Agents 5
- 1.4 Gas Vesicles as Acoustic Reporter Genes 9
- 1.5 References 11

Chapter II: Engineering Approaches for Acoustic Reporter Genes 15
- 2.1 Capabilities and Limitations of Existing Acoustic Reporter Genes 15
- 2.2 Previous Acoustic Reporter Gene Engineering Efforts 15
- 2.3 High-Throughput Approaches to Acoustic Reporter Gene Engineering......... 16
- 2.4 References 21

Chapter III: Directed Evolution of Acoustic Reporter Genes Using High-Throughput Acoustic Screening 23
- 3.1 Abstract.. 23
- 3.2 Introduction 24
- 3.3 Results 25
- 3.4 Discussion 33
- 3.5 Methods 34
- 3.6 Author Contributions 38
- 3.7 Acknowledgements 38
- 3.8 References 38
- 3.9 Supplementary Information 42

Chapter IV: Genomically Mined Acoustic Reporter Genes for Real-Time *In Vivo* Monitoring of Tumors and Tumor-Homing Bacteria...................................... 47
- 4.1 Abstract.. 47
- 4.2 Introduction 47
- 4.3 Results 48
- 4.4 Discussion 69
- 4.5 Methods 71
- 4.6 Data Availability 78
- 4.7 Code Availability 78
- 4.8 Acknowledgements 78
- 4.9 Author Contributions 79
- 4.10 Competing Interests 79
- 4.11 References 80
- 4.12 Index of Supplementary Information 86
- 4.13 Supplementary Information 86

Chapter V: Conclusions and Future Directions.. 106
- 5.1 Lessons learned from half a decade of ARG engineering 106
- 5.2 Future Directions for ARG Engineering... 107
- 5.3 References 111

LIST OF ILLUSTRATIONS AND/OR TABLES

Number *Page*

1. Figure 1-1: Properties and applications of ultrasound waves 4
2. Figure 1-2: Gas vesicle structural properties and *in vitro* GV US imaging....... 6
3. Figure 1-3: US interactions with GVs 8
4. Figure 1-4: ARG imaging *in vitro* and *in vivo* 9
5. Figure 3-1. High-throughput directed evolution workflow for ARGs............... 26
6. Figure 3-2. Optimization of GV expression from WT *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters 28
7. Figure 3-3. First round of directed evolution of *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters... 29
8. Figure 3-4. Second round of directed evolution of *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters 31
9. Figure 3-5. Non-destructive multiplexed imaging of two ARG types 32
10. Figure 3-S1. Detailed diagram of the acoustic plate reader workflow............ 42
11. Figure 3-S2. Details of *gvpA*/*gvpB* mutant library construction and screening43
12. Figure 3-S3. Characterization of the top mutants from Round 1 of evolution 44
13. Figure 3-S4. Characterization of the top mutants from Round 2 of evolution 45
14. Figure 3-S5. Additional characterization of the best mutants identified in this study 45
15. Figure 4-1: Genomic mining of gas vesicle gene clusters reveals homologs with nonlinear ultrasound contrast in *E. coli*... 49
16. Figure 4-ED1: Additional images and quantification of *E. coli* patches expressing select GV gene clusters... 50
17. Figure 4-ED2: Characterization of working GV clusters in BL21(DE3) *E. coli*52
18. Figure 4-2: Expression of bARG$_{Ser}$ in *E. coli* Nissle (EcN) and acoustic characterization *in vitro*.. 54
19. Figure 4-ED3: Testing bARG$_{Ser}$ expression in EcN with IPTG- and aTc-inducible gene circuits 55
20. Figure 4-ED4: Effect of induction on viability and OD$_{600}$ for bARG$_{Ser}$-expressing EcN in liquid culture.. 57
21. Figure 4-ED5: *In vitro* characterizations of bARG$_{Ser}$-expressing EcN with BURST ultrasound imaging.. 58
22. Figure 4-ED6: bARG$_{Ser}$ expression and acoustic characterization in *Salmonella enterica* serovar Typhimurium 59

23. Figure 4-3: *In situ* bARG_{Ser} expression enables ultrasound imaging of tumor colonization by EcN........61
24. Figure 4-ED7: xAM ultrasound imaging of mouse tumors colonized by EcN 62
25. Figure 4-4: Heterologous expression of the *Anabaena flos-aquae* GV gene cluster in mammalian cells........63
26. Figure 4-ED8: Additional data on heterologous expression of the *Anabaena flos-aquae* GV gene cluster in mammalian cells........65
27. Figure 4-5: *In situ* mARG_{Ana} expression enables nondestructive ultrasound imaging of orthotopic tumors........67
28. Figure 4-ED9: Tumor histology and imaging through thick tissue........68
29. Figure 4-6: xAM imaging of mARG_{Ana} enables ultrasound-guided genetically selective tumor biopsy........69
30. Figure 4-S1: 16S phylogenetic tree of all reported GV-producing organisms94
31. Figure 4-S2: Optimization of expression conditions for all tested clusters in BL21(DE3) *E. coli* 96
32. Figure 4-S3: Quantification of ultrasound signal for all samples shown in Fig. S2a-b 98
33. Figure 4-S4: Optical and xAM pre-minus-post-collapse difference images of all samples shown in Fig. S2........99
34. Figure 4-S5: Quantification of ultrasound signal for samples shown in Fig. S4c-d 100
35. Figure 4-S6: Effects of single-gene deletions on GV expression by the *Serratia* cluster 101
36. Figure 4-S7: Quantification and characterization of EcN mutants deficient in bARG_{Ser} expression isolated from daily subculturing *in vitro*........102
37. Figure 4-S8: xAM ultrasound signal versus time at varying acoustic pressures applied sequentially to the same sample versus separate samples........103
38. Figure 4-S9: Histology of MC26 tumor colonized with bARG_{Ser}-expressing EcN 104
39. Figure 4-S10: Screening for EcN mutants defective in bARG_{Ser} expression isolated from colonized tumors........105
40. Figure 4-S11: Flow cytometric gating strategy for chimeric tumor biopsy sample analyses 105

Chapter 1

MOLECULAR IMAGING OF ACOUSTIC REPORTER GENES

Sections of this chapter have been adapted from:
Rabut C, Yoo S, **Hurt RC**, Jin Z, Li R, Guo H, Ling BL, Shapiro MG. “Ultrasound Technologies for Imaging and Modulating Neural Activity.” *Neuron*. 2020. DOI: 10.1016/j.neuron.2020.09.003

1.1 The Importance of Noninvasive Molecular Imaging

Breakthroughs in biology can often be traced to new experimental methods. Diverse techniques ranging from electrophysiology and histology to optical imaging, magnetic resonance, optogenetics, and chemogenetics have provided new ways to study the structure and function of everything from neural circuits to tumor microenvironments to the gut microbiome. In particular, molecular imaging (MI)—defined as the use of imaging techniques to gain real-time information about molecular processes occurring in living specimens—has played a key role in dissecting the spatial and temporal contributions of a variety of biomolecules to various biological processes in health and disease. While a handful of label-free MI methods have been developed, most rely on imaging contrast agents: molecules that emit a signal detectable by a given imaging modality, and which may label a specific molecule of interest. The primary imaging modalities used for MI are positron emission tomography (PET), computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), bioluminescence imaging (BLI), 2-photon imaging, Raman imaging, and photoacoustics (PA). Of these, only PET and BLI strictly require contrast agents, though 2-photon imaging is almost always conducted with them. In the next section, I briefly discuss the strengths and weaknesses of the most popularly used of these methods, and of their associated contrast agents.

The ultimate goal of the MI field is to be able to noninvasively monitor any biochemical process occurring inside an organism in real time with high spatial, temporal, and molecular specificity. However, this is currently achievable for only a limited number of molecules/processes due to a lack of genetically encoded contrast agents, and many of the imaging modalities used for MI suffer from poor spatial or temporal resolution or penetration depth into tissue.

Currently, there is intense interest in expanding the number and types of molecules and molecular processes that can be imaged, as well as in improving various aspects of the imaging modalities themselves (primarily spatial and temporal resolution, sensitivity, and specificity). The former is being addressed by developing both functional contrast agents that can detect specific molecules of interest and contrast agent-free methods of detecting biomolecules, and by using genetically encoded contrast agents that can report gene expression in defined cell types. The latter is being addressed by hardware and software solutions specific to each imaging modality, but also by developing or using contrast agents that can extend the limits of these modalities (*e.g.*, sparsity-based super-resolution techniques) or that produce bright or bio-orthogonal signals for sensitive or specific detection.

In this chapter, I will discuss the current state-of-the-art MI methods that have been used to image gene expression and other cellular processes, with a particular focus on ultrasound and its contrast agents.

1.2 Existing Methods for Noninvasive Molecular Imaging

The history of MI has spanned many imaging modalities and molecular targets: neurotransmitters[1–3] and their receptors; glucose, kinases[4], amyloid beta, but also various tumor-related molecules. In most sub-fields of biology, PET and MRI contrast agents—once the state of the art—were largely supplanted by optical tools following the discovery and adaptation of GFP into a biotechnology tool. More recently, MRI contrast agents have made a resurgence due to the advent of genetically encoded probes, and other imaging modalities such as US and PA are gaining favor as their contrast agents are being developed.

The four most important considerations for the evaluation of a contrast agent's performance are: 1) its sensitivity (the ability to be detected in small quantities), 2) its specificity (the signal produced is well discriminated from other naturally- or artificially- occurring signals), 3) its stability (to not degrade spontaneously or to be actively degraded by either biological or imaging processes), and 4) its targetability (binding to the correct receptor, localization to the correct cell type or subcellular compartment, etc.). For example, microbubbles, the "gold standard" of US contrast agents, have been imaged with excellent sensitivity and specificity (in large part due to developments in US imaging software), but lack in targetability and stability.[5]

In the rest of this section, I discuss the four most commonly used MI modalities and their corresponding synthetic contrast agents.

1.2a Positron Emission Tomography and Its Synthetic Contrast Agents

Positron emission tomography images the gamma rays emitted when positrons (emitted from radiotracers via beta plus decay) annihilate with electrons present in tissue.[6] The contrast agents used in PET are typically a radioisotope conjugated to a drug or other macromolecule, making these probes highly targetable; further, because tissue does not emit gamma rays of its own, PET has the highest specificity of any clinically used imaging modality, and is considered “background-free.” PET is also considered the most sensitive MI modality, with its ability to detect analytes in nanomolar to picomolar concentrations.[7] However, PET contrast agents tend to be highly unstable, with half-lives on the order of minutes to hours,[6] and must be synthesized by cyclotrons that are often on-site at the imaging facility.[8] Further, PET (as well as CT) scans subject patients to ionizing radiation, which pose risks for patients who require many such scans.[9] Finally, the temporal resolution of PET is limited by the pharmacokinetics of these radioactive probes, and the spatial delocalization of their emissions limits its spatial resolution to ~1 mm.[10]

1.2b Magnetic Resonance Imaging and Its Synthetic Contrast Agents

Magnetic resonance imaging involves applying a strong magnetic field to align the spins of hydrogen protons in water molecules present in tissue and using pules of radiofrequency (RF) electromagnetic waves to excite them, then monitoring the current induced in an RF coil as these spins relax back to their ground state energy levels. Because it relies only on magnetic fields and RF waves, both of which have effectively unlimited penetration depth into tissue, MRI can produce whole-body images of large organisms noninvasively. Its moderate spatial and temporal resolutions (~100 μm and ~1 s, respectively) make it an excellent method for most types of clinical imaging, and its lack of ionizing radiation makes it extremely safe.[11] However, its high cost and low portability typically limit its use to hospital settings, and the long acquisition times it requires make it highly susceptible to motion artifacts.[10] A large number of contrast mechanisms and associated contrast agents are available for MRI,[11] including both synthetic and genetically encoded ones, and their development is the focus of intense research. While highly exciting in terms of the types of biological processes they can monitor noninvasively deep inside organisms, these contrast agents are fundamentally limited by the high cost, low portability, and unimpressive spatial and temporal resolutions of MRI.

1.2c Photoacoustic Imaging and Its Synthetic Contrast Agents

Photoacoustic imaging involves using optical excitation to create thermoelastic pressure waves that are then detected by ultrasound transducers.[12] The advantages of this approach are that: 1) it uses IR light and US, both of which penetrate deeply through tissue; 2) it can achieve high spatial and temporal resolutions;[13] 3) its contrast agents include existing molecular tools used in optical imaging, such as fluorescent proteins and light-absorbing pigments.[14,15] Its primary disadvantage is that despite the high penetration depth of US, the high absorbance of IR light by tissue means that it cannot image beyond a few cm of tissue without unacceptable phototoxicity. Similar to MRI, there is intense interest in developing functional contrast agents for PA, and these will surely represent very valuable additions to the MI toolbox.

1.2d Ultrasound Imaging and Its Synthetic Contrast Agents

Ultrasound imaging uses pulses of ultrasonic (>20 kHz) sound waves to generate images based on the backscattered signal intensity at either the transmitted US frequency or higher harmonics thereof. Because US easily penetrates soft tissues—the frequencies typically used in US imaging (3–25 MHz) correspond to wavelengths of 500–60 µm and penetration depths of 10–1 cm, respectively—US-based methods can generate images or deliver focused energy several cm into tissue with spatial precision corresponding to its wavelength.[16] Moreover, the speed of sound wave propagation—approximately 1.5 km/s in soft tissue—allows US to operate with temporal precision below 1 ms (Figure 1-1A). As US waves propagate through tissue, they reflect and scatter off interfaces between materials of differing densities and compressibilities (Figure 1-1B); light waves are similarly scattered at interfaces between materials with different refractive indices, but to a much greater extent that US waves (Figure 1-1C), with the result that light of most wavelengths penetrates only a few hundred microns into tissue. US imaging and optical imaging both suffer from the same fundamental tradeoff between spatial resolution and penetration depth (Figure 1-1D), due to the wavelength-dependent attenuation and diffraction of both types of waves, as enumerated in Abbe's diffraction limit. The fortuitous properties of US have made it one of the most widely used imaging technologies in clinical medicine, facilitated by its relatively low cost, high portability, and safety (Figure 1-1E). At the same time, as the only form of non-ionizing energy that can be focused in deep tissues, focused US (FUS) has become a rapidly growing modality for noninvasive therapy, used in the ablative treatment of cancer and neurological dysfunction, facilitated by advances in US hardware and image guidance.[17]

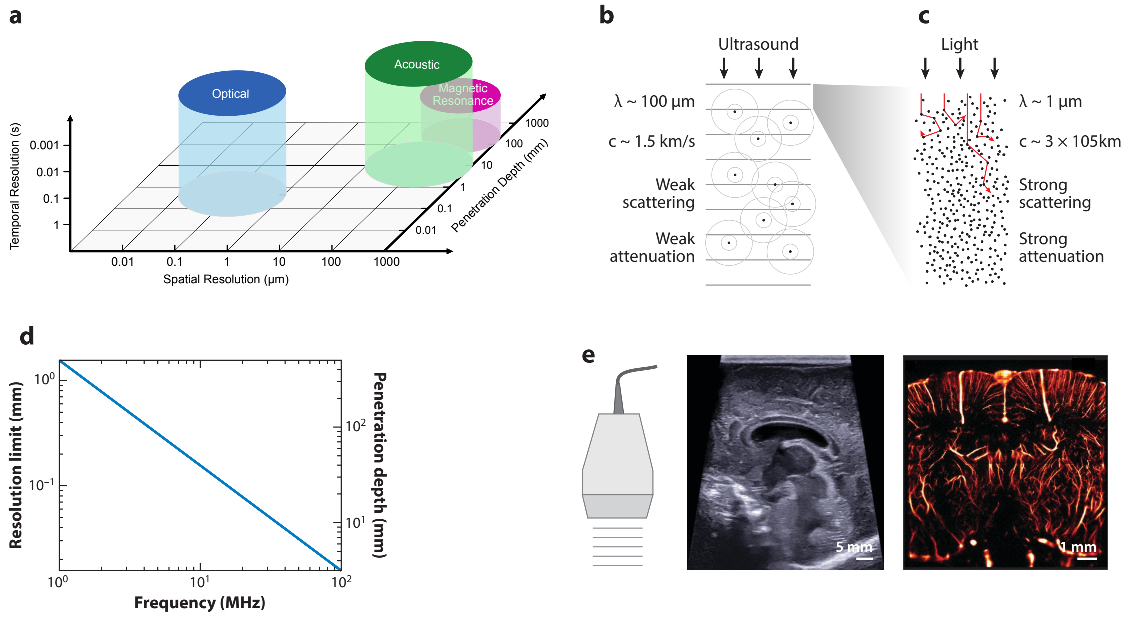


Figure 1-1: Properties and applications of ultrasound waves. (A) Approximate performance characteristics of common imaging modalities used for MI. (B) Physical properties of US in tissue. (C) Physical properties of light propagation through tissue. (D) Fundamental tradeoff between ultrasound resolution and penetration depth as a function of frequency in brain tissue (penetration depth was assessed based on a 60-decibel round-trip attenuation). (E) Demonstration of ultrasound imaging capabilities; conventional B-mode image of an infant human brain showing sub-millimeter resolution; 15 MHz super-resolution ultrasound image of rat brain vasculature with 8 μm resolution, breaking the classical tradeoff presented in (D). Panel A adapted from Rabut *et al.*[18] Panels B-E adapted from Maresca *et al.*[16]

A handful of synthetic contrast agents have been developed for US: primarily microbubbles and droplets/nanobubbles. These contrast agents are almost exclusively used for intravascular imaging, and their strengths and limitations have been reviewed.[19] Briefly, microbubbles are the most commonly used US contrast agents for clinical imaging, and can produce extremely strong and specific nonlinear signals when imaged at their resonant frequency. They are stable for a matter of minutes after injection into the blood with minimal toxicity, and they can be targeted to bind specific intravascular biomarkers.[20–24] Phase-shift droplets are less stable than microbubbles, and are more challenging to image.[19] They are not targetable, but are small enough to extravasate through leaky vasculature, leading to claims that they can label tumors specifically.[25,26] Nanobubbles, despite low per-particle scattering, can produce strong contrast because their small size allows them to reach high concentrations per unit volume.[19] They are also capable of extravasation into tumors.[27]

These synthetic contrast agents have been used for several exciting applications in intravascular imaging, namely tumor intravascular imaging, microbubble localization microscopy, molecular imaging of vascular markers of disease, microbubble-based therapy monitoring, and as enhancers of hemodynamic contrast in functional US imaging.[19] Despite these successful applications, the lack of genetic encodability of synthetic contrast agents prevents their use in monitoring some of the most important biological processes that scientists and physicians seek to study, such as gene expression and the sensing of intracellular signals such as Ca^{2+} waves and enzymatic activity.

1.3 Gas Vesicles as Ultrasound Contrast Agents

Gas vesicles (GVs) are genetically encoded gas-filled nanostructures (Fig. 1-2A) that evolved in aquatic photosynthetic microbes as a means to achieve buoyancy.[28,29] GV clusters are found in several species of photosynthetic bacteria and haloarchaea, and nonfunctional GV clusters are present in the genomes of many more. GVs comprise a 2-nm-thick protein shell, with a typical diameter of ~140 nm and length of 500 nm, enclosing a compartment of gas that is at equilibrium with the surrounding media (Fig. 1-2B).[16,28,29] The cigar-shaped protein shell that forms the core of the GV is composed primarily of monomers of a single protein—gas vesicle protein A (GvpA) (Fig. 1-2C)—with optional external scaffolding proteins (GvpC) that mechanically reinforce the GV shell (structure and properties reviewed in Maresca *et al.*[16]). They are encoded for by clusters of 8-20+

genes, only 1-2 of which (GvpA/B with the possible addition of GvpC) encode for major structural proteins, the others encoding for minor structural proteins, chaperones, or other essential “assembly factors” (Fig. 1-2D). GVs rely on a fascinating trick of chemistry to maintain their gas-filled interiors in aqueous environments—even though the pores between the monomers of the GV shell are large enough for water to diffuse through, the inner surface of the GV shell is so hydrophobic that any water vapor that diffuses in cannot nucleate into a liquid phase—therefore, GVs are filled with whatever gasses are present in their environment. Thus, GVs can be considered genetically encoded nanometric gas containers, and this property enables their use as US contrast agents because the gas phase inside them presents an interface off of which US waves propagating through the aqueous cellular environment can scatter. Critical for their use in MI, these gene clusters have been successfully transferred into commensal bacteria and mammalian cells, allowing these cells to be made visible under an acoustic beam, just as GFP once did for optical imaging.

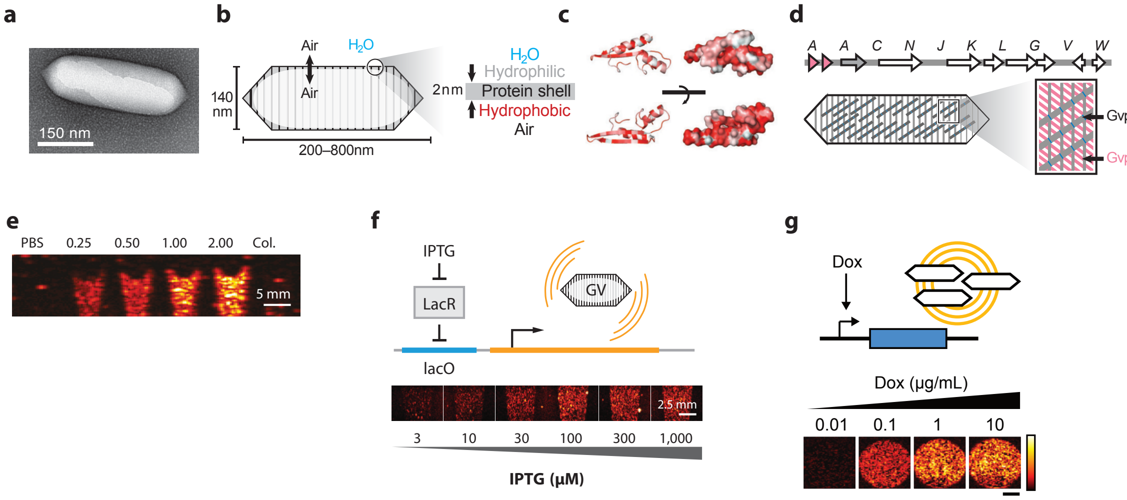


Figure 1-2: Gas vesicle structural properties and *in vitro* GV US imaging. (A) Transmission electron microscopy (TEM) image of a GV. (B) Diagram of a GV, showing the ability of gas to cross the shell, and the exclusion of water (left); physiochemical properties of the GV shell, namely the hydrophilic exterior that helps to solubilize the GV in the aqueous environment of the cell and the hydrophobic interior that prevents water from forming a liquid phase inside the GV (right). (C) Structural models of GvpA, the primary structural protein of the GV. Hydrophilic residues are colored white and hydrophobic residues red. The former can be seen to cluster on the convex side and the latter on the concave side. (D) (Top) Gene cluster from *Anabaena flos-aquae* encoding the formation of GVs, with each Gvp gene labeled. (Bottom) Diagram of a GV showing the relative contributions of the two main structural proteins to the structure of the GV. (E) US images of purified GVs in agarose gel at various optical densities and after hydrostatic collapse. (F) US images of bacterial cells in agarose gel expressing GVs under the control of an IPTG-inducible promoter. (G) US images of human HEK293T cells in agarose gel expressing GVs under the control of a doxycycline (Dox)-inducible promoter. Scale bar, 1 mm. All panels adapted from Maresca *et al.*[16]

When imaged in purified form *in vitro* (Fig. 1-2E),[30] or expressed in bacterial (Fig. 1-2F)[31] or mammalian cells (Fig. 1-2G),[32] GVs produce bright contrast from backscattered US. Specialized US imaging paradigms have been developed to detect GVs with increased sensitivity and specificity.[33–35] These methodologies can be destructive, in which case they are based on monitoring GV collapse under high US pressures;[30] conversely, they can leverage the reversible mechanical buckling of the GV shell which generates nonlinear responses to acoustic pressure,[36–38] enabling selective imaging of these particles using a tailored amplitude modulation strategy.[34,35] The unique possibility to tune the structural properties of GVs raise the opportunity to modulate their response to ultrasound and to maximally distinguish their signals from background. In addition to their use in imaging, GVs have been used to apply force to cells in the case of sonogenetics[39] and cell sorting,[40] and to tissues for blood-brain barrier disruption.[41]

Compared to the synthetic US contrast agents mentioned above, GVs have three primary advantages: 1) their genetic encodability, 2) their stability, 3) their small size and narrow size distribution. Like synthetic US contrast agents, GVs provide excellent sensitivity and specificity in US imaging, and can be targeted to specific receptors[37,42] and cell types. GVs' genetic encodability makes them fundamentally different from any other US contrast agent, and is critical for their use as Acoustic Reporter Genes (see below). Compared to microbubbles, which are stable for a matter of weeks, GVs can be stored for years with minimal degradation (from personal experience). As mentioned above, a major use of synthetic US contrast agents is as enhancers of hemodynamic contrast in functional US imaging; importantly for this use case, GVs' small size allows them to enter capillaries more effectively than microbubbles, and their narrow size distribution means that the hemodynamic enhancement they confer upon injection is smoother than that from microbubbles (despite being smaller in amplitude).[43]

1.3a US interactions with GVs

Our current model for the acoustic behavior of GVs includes three regimes (Fig. 1-3A).[34–37] At low pressures, GVs scatter "linearly," in the sense that the amount of signal detected is directly proportional to the input pressure. This signal is generated by the difference in acoustic impedance between the gas phase inside the GV and the liquid phase outside, as sound waves are known to scatter at interfaces between materials of differing acoustic impedances. At intermediate pressures, GVs scatter "nonlinearly," and as a result the signal detected becomes a higher percentage of the input pressure than what would be expected if the GVs were scattering linearly. This behavior is thought to result from reversible "buckling" of the GVs' protein shells, which makes them more efficient scatterers by changing the GV radius with each cycle of ultrasound.[34] Finally, at high pressures, the shell collapses and the gas inside escapes, which eliminates scattering.[37]

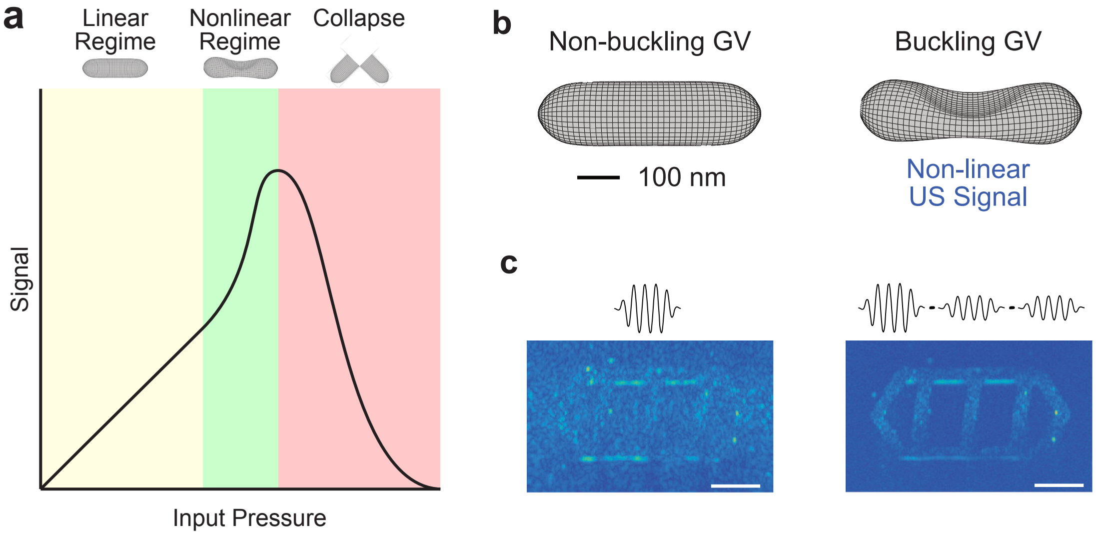


Figure 1-3: US interactions with GVs. (A) Three regimes of GV response to ultrasound. (B-C) Buckling and non-bucking GVs, and the corresponding linear and nonlinear US images acquired in each pressure regime. Panels B-C adapted from Rabut *et al.*[18]

In addition to the standard US imaging paradigms used for imaging tissue and synthetic contrast agents, two specialized paradigms have been developed that take advantage of GVs' nonlinear scattering and collapse regimes to detect them with high sensitivity and specificity: BURST imaging and xAM imaging.

Burst Ultrasound Reconstructed with Signal Templates (BURST) imaging leverages the unique temporal signal pattern produced by GVs as they collapse; by extracting the unique pattern of this signal from the total backscattered US signal, BURST increases the sensitivity GV imaging by more than 1,000-fold. In dilute cell suspensions, this improvement enabled the detection of ARG expression in individual bacterial and mammalian cells.[44] While extremely sensitive and highly specific, BUSRT imaging relies on GV collapse to generate images, making it undesirable for applications in which GVs are used to monitor cellular behavior over time.

Cross-Amplitude Modulation (xAM) imaging leverages GVs' tendency to scatter US more efficiently in the nonlinear regime due to the reversible buckling behavior of their shells (Fig. 1-3B): by subtracting an image collected in the linear scattering regime from one collected in the nonlinear regime, buckling GV signals can be specifically discriminated from tissue signals, which are primarily linear (Fig. 1-3C).[35] This high specificity makes xAM imaging the method of choice for GV detection, as it does not require GV collapse to generate images.

1.4 Gas Vesicles as Acoustic Reporter Genes

Reporter genes—genes coding for proteins with unique properties (typically enzymes or genetically encoded contrast agents) that make cells expressing them distinguishable from those that are not–have been used for decades to monitor the transcriptional activities of genes to which they are fused,[45] and have been developed for organisms in every kingdom of life.[46–51] In particular, there has been an explosion in the number of reporter genes available for bacterial and mammalian cells in the past few decades, and reporter genes have been developed for every commonly used imaging modality. Of particular interest is the use of GVs as reporter genes for US, which is possible only because of their genetic encodability.

Prior to the work described in this thesis, GVs had been successfully expressed heterologously in cultured cells from two bacterial species (*Escherichia coli* and *Salmonella typhimurium*) (Fig. 1-4A-B) as well as human cells, in which they were used to report transcription from exogenous promoters. When used as acoustic reporter genes (ARGs) *in vivo*, GVs benefit from the high penetration depth of US into tissue, as well as the high detection specificity and sensitivity afforded by the specialized US imaging paradigms developed for detecting GV-specific signals discussed above. Below, I will discuss the two previous studies in which GVs were used as ARGs, which directly enabled the work presented in this thesis.

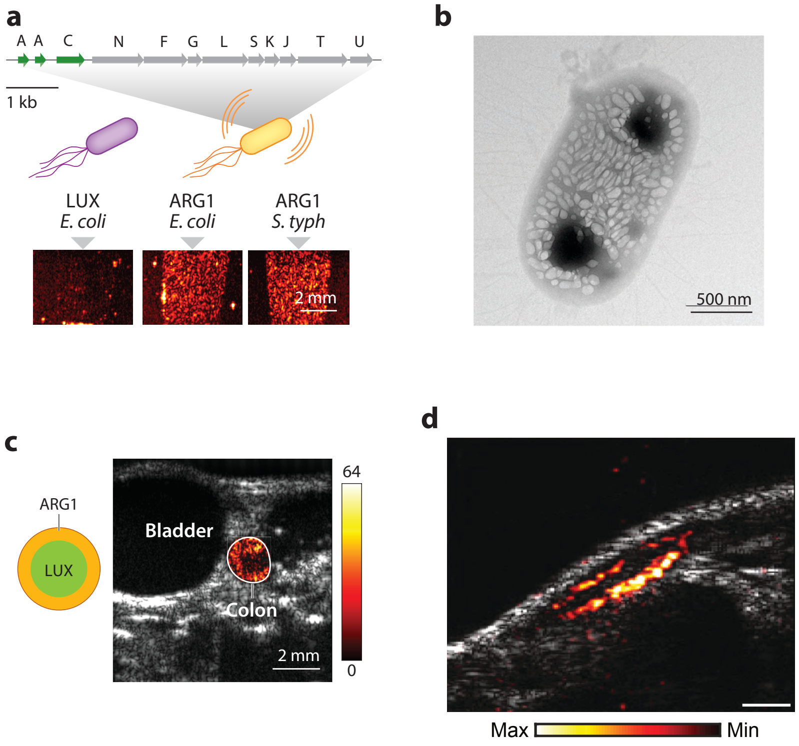


Figure 1-4: ARG imaging *in vitro* and *in vivo*. (A) US images of bacteria expressing either first-generation ARGs or a negative control LUX operon. (B) TEM image of an *E. coli* cell expressing ARGs. (C) Pre-/post-collapse difference image of ARG-expressing bacteria inside the colon of a mouse. (D) BURST US signal from HEK293T cells expressing ARGs under the skin of a mouse. All panels adapted from Maresca *et al.*[16]

The first demonstration of GVs as ARGs came in 2018, when Bourdeau and colleagues showed that GV-expressing bacteria could be imaged noninvasively inside the mouse colon using ultrasound (Fig. 1-4C).[31] Although previous work had established the ability to heterologously express GVs in bacteria,[52] this study expanded the number of GV types that could be expressed in *E. coli*, and demonstrated GV expression for the first time in *Salmonella typhimurium*. Further, it showed that specific GV detection was possible noninvasively in living mice, and that multiple GV types could be combined in multiplexed imaging, albeit in both cases using imaging methods that destroyed the GVs. While an extremely exciting proof of concept that GV-expressing bacteria could be successfully imaged *in vivo*, this study's most significant limitation was that all GV expression was performed *in vitro* under ideal culture conditions and not at physiological temperature, with GV-expressing bacteria being subsequently injected into mice prior to imaging. This study highlighted the need for ARGs that could be expressed *in situ* and imaged nondestructively, and which were compatible with nondestructive multiplexed imaging: goals which were achieved in the work described in this thesis.

Building on this work, Farhadi and colleagues showed in 2019 that GVs could be expressed by mammalian cells, and that GV-expressing mammalian cells could be imaged noninvasively inside mice using US (Fig. 1-4D).[32] In this case, GV expression was performed *in situ*, making it the first true use of GVs as *in vivo* ARGs. While very exciting, this study had two significant limitations: first, the GV type used was not compatible with the highly specific nonlinear GV imaging paradigm described above, which meant that it could only be detected using collapse-based imaging; second, cell-to-cell variability in its expression and burden meant that it could only be imaged robustly with US in clonally selected cell lines stimulated with potent epigenetic reagents. Thus, we sought a new GV type that produced considerable nonlinear signal when expressed in mammalian cells, and whose expression was robust enough that it could be detected in polyclonal cell lines prepared by sorting alone, without the need for screening of multiple monoclonal cell lines.

Based on the limitations of the ARGs presented in these two studies, we sought the following ARG types: 1) bacterial ARGs compatible with *in situ* expression and nondestructive imaging, 2) mammalian ARGs compatible with *in situ* expression and nondestructive imaging, and 3) bacterial ARGs compatible with nondestructive multiplexed imaging. In the following chapters I will detail our successful paths to engineering them.

1.5 References

1. Brustad, E. M. *et al.* Structure-Guided Directed Evolution of Highly Selective P450-Based Magnetic Resonance Imaging Sensors for Dopamine and Serotonin. *Journal of Molecular Biology* **422**, 245–262 (2012).
2. Lee, T., Cai, L. X., Lelyveld, V. S., Hai, A. & Jasanoff, A. Molecular-Level Functional Magnetic Resonance Imaging of Dopaminergic Signaling. *Science* **344**, 533–535 (2014).
3. Shapiro, M. G. *et al.* Directed evolution of a magnetic resonance imaging contrast agent for noninvasive imaging of dopamine. *Nat Biotechnol* **28**, 264–270 (2010).
4. Shapiro, M. G., Szablowski, J. O., Langer, R. & Jasanoff, A. Protein Nanoparticles Engineered to Sense Kinase Activity in MRI. *J. Am. Chem. Soc.* **131**, 2484–2486 (2009).
5. Paefgen, V., Doleschel, D. & Kiessling, F. Evolution of contrast agents for ultrasound imaging and ultrasound-mediated drug delivery. *Front. Pharmacol.* **6**, (2015).
6. Vaquero, J. J. & Kinahan, P. Positron Emission Tomography: Current Challenges and Opportunities for Technological Advances in Clinical and Preclinical Imaging Systems. *Annu. Rev. Biomed. Eng.* **17**, 385–414 (2015).
7. Vandenberghe, S., Moskal, P. & Karp, J. S. State of the art in total body PET. *EJNMMI Phys* **7**, 35 (2020).
8. Keppler, J. S. & Conti, P. S. A Cost Analysis of Positron Emission Tomography. *American Journal of Roentgenology* **177**, 31–40 (2001).
9. Wen, J. C., Sai, V., Straatsma, B. R. & McCannel, T. A. Radiation-Related Cancer Risk Associated With Surveillance Imaging for Metastasis From Choroidal Melanoma. *JAMA Ophthalmol* **131**, 56 (2013).
10. Pysz, M. A., Gambhir, S. S. & Willmann, J. K. Molecular imaging: current status and emerging strategies. *Clinical Radiology* **65**, 500–516 (2010).
11. Mukherjee, A., Davis, H. C., Ramesh, P., Lu, G. J. & Shapiro, M. G. Biomolecular MRI reporters: Evolution of new mechanisms. *Progress in Nuclear Magnetic Resonance Spectroscopy* **102–103**, 32–42 (2017).
12. Wang, L. V. & Yao, J. A practical guide to photoacoustic tomography in the life sciences. *Nat Methods* **13**, 627–638 (2016).
13. Xia, J., Yao, J. & Wang, L. H. V. PHOTOACOUSTIC TOMOGRAPHY: PRINCIPLES AND ADVANCES (Invited Review). *PIER* **147**, 1–22 (2014).
14. Jathoul, A. P. *et al.* Deep in vivo photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter. *Nature Photon* **9**, 239–246 (2015).
15. Yao, J. *et al.* Multiscale photoacoustic tomography using reversibly switchable bacterial phytochrome as a near-infrared photochromic probe. *Nat Methods* **13**, 67–73 (2016).

16. Maresca, D. *et al.* Biomolecular Ultrasound and Sonogenetics. *Annu. Rev. Chem. Biomol. Eng.* **9**, 229–252 (2018).
17. Escoffre, J.-M. & Bouakaz, A. *Therapeutic ultrasound*. (Springer, 2016).
18. Rabut, C. *et al.* Ultrasound Technologies for Imaging and Modulating Neural Activity. *Neuron* **108**, 93–110 (2020).
19. Yusefi, H. & Helfield, B. Ultrasound Contrast Imaging: Fundamentals and Emerging Technology. *Front. Phys.* **10**, 791145 (2022).
20. Kaufmann, B. A. *et al.* Molecular Imaging of Inflammation in Atherosclerosis With Targeted Ultrasound Detection of Vascular Cell Adhesion Molecule-1. *Circulation* **116**, 276–284 (2007).
21. Willmann, J. K. *et al.* Targeted Contrast-Enhanced Ultrasound Imaging of Tumor Angiogenesis with Contrast Microbubbles Conjugated to Integrin-Binding Knottin Peptides. *J Nucl Med* **51**, 433–440 (2010).
22. Hamilton, A. J. *et al.* Intravascular ultrasound molecular imaging of atheroma components in vivo. *Journal of the American College of Cardiology* **43**, 453–460 (2004).
23. Weller, G. E. R. *et al.* Ultrasound Imaging of Acute Cardiac Transplant Rejection With Microbubbles Targeted to Intercellular Adhesion Molecule-1. *Circulation* **108**, 218–224 (2003).
24. Ellegala, D. B. *et al.* Imaging Tumor Angiogenesis With Contrast Ultrasound and Microbubbles Targeted to $\alpha_v \beta_3$. *Circulation* **108**, 336–341 (2003).
25. Helfield, B. L. *et al.* Investigating the Accumulation of Submicron Phase-Change Droplets in Tumors. *Ultrasound in Medicine & Biology* **46**, 2861–2870 (2020).
26. Williams, R. *et al.* Characterization of Submicron Phase-change Perfluorocarbon Droplets for Extravascular Ultrasound Imaging of Cancer. *Ultrasound in Medicine & Biology* **39**, 475–489 (2013).
27. Pellow, C., Abenojar, E. C., Exner, A. A., Zheng, G. & Goertz, D. E. Concurrent visual and acoustic tracking of passive and active delivery of nanobubbles to tumors. *Theranostics* **10**, 11690–11706 (2020).
28. Pfeifer, F. Distribution, formation and regulation of gas vesicles. *Nature Reviews Microbiology* **10**, 705–715 (2012).
29. Walsby, A. E. Gas vesicles. *Microbiol Rev* **58**, 94–144 (1994).
30. Shapiro, M. G. *et al.* Biogenic gas nanostructures as ultrasonic molecular reporters. *Nature Nanotech* **9**, 311–316 (2014).
31. Bourdeau, R. W. *et al.* Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts. *Nature* **553**, 86–90 (2018).
32. Farhadi, A., Ho, G. H., Sawyer, D. P., Bourdeau, R. W. & Shapiro, M. G. Ultrasound imaging of gene expression in mammalian cells. *Science* **365**, 1469–1475 (2019).

33. Farhadi, A., Ho, G. H., Sawyer, D. P., Bourdeau, R. W. & Shapiro, M. G. Ultrasound imaging of gene expression in mammalian cells. *Science* **365**, 1469–1475 (2019).
34. Maresca, D. *et al.* Nonlinear ultrasound imaging of nanoscale acoustic biomolecules. *Appl. Phys. Lett.* **110**, 073704 (2017).
35. Maresca, D., Sawyer, D. P., Renaud, G., Lee-Gosselin, A. & Shapiro, M. G. Nonlinear X-Wave Ultrasound Imaging of Acoustic Biomolecules. *Phys. Rev. X* **8**, 041002 (2018).
36. Cherin, E. *et al.* Acoustic Behavior of Halobacterium salinarum Gas Vesicles in the High-Frequency Range: Experiments and Modeling. *Ultrasound in Medicine and Biology* **43**, 1016–1030 (2017).
37. Lakshmanan, A. *et al.* Molecular Engineering of Acoustic Protein Nanostructures. *ACS Nano* **10**, 7314–7322 (2016).
38. Zhang, S. *et al.* The Vibration Behavior of Sub-Micrometer Gas Vesicles in Response to Acoustic Excitation Determined via Laser Doppler Vibrometry. *Adv. Funct. Mater.* **30**, 2000239 (2020).
39. Hou, X. *et al. Ultrasound neuromodulation through nanobubble-actuated sonogenetics*. http://biorxiv.org/lookup/doi/10.1101/2020.10.21.348870 (2020) doi:10.1101/2020.10.21.348870.
40. Wu, D. *et al.* Biomolecular actuators for genetically selective acoustic manipulation of cells. *Sci. Adv.* **9**, eadd9186 (2023).
41. Zhang, J. *et al.* Biosynthetic Gas Vesicles Combined with Focused Ultrasound for Blood–Brain Barrier Opening. *IJN* **Volume 17**, 6759–6772 (2022).
42. Hao, Y., Li, Z., Luo, J., Li, L. & Yan, F. Ultrasound Molecular Imaging of Epithelial Mesenchymal Transition for Evaluating Tumor Metastatic Potential via Targeted Biosynthetic Gas Vesicles. *Small* 2207940 (2023) doi:10.1002/smll.202207940.
43. Maresca, D. *et al.* Acoustic biomolecules enhance hemodynamic functional ultrasound imaging of neural activity. *NeuroImage* **209**, 116467 (2020).
44. Sawyer, D. P. *et al.* Ultrasensitive ultrasound imaging of gene expression with signal unmixing. *Nat Methods* **18**, 945–952 (2021).
45. Serganova, I. & Blasberg, R. G. Molecular Imaging with Reporter Genes: Has Its Promise Been Delivered? *J Nucl Med* **60**, 1665–1681 (2019).
46. Herrera-Estrella, L., León, P., Olsson, O. & Teeri, H. T. Reporter genes for plants. in *Plant Molecular Biology Manual* (eds. Gelvin, S. B. & Schilperoort, R. A.) 139–170 (Springer Netherlands, 1994). doi:10.1007/978-94-011-0511-8_10.
47. Faktorová, D. *et al.* Genetic tool development in marine protists: emerging model organisms for experimental cell biology. *Nat Methods* **17**, 481–494 (2020).

48. Soll, D. R. & Srikantha, T. Reporters for the analysis of gene regulation in fungi pathogenic to man. *Current Opinion in Microbiology* **1**, 400–405 (1998).

49. Alam, J. & Cook, J. L. Reporter genes: Application to the study of mammalian gene transcription. *Analytical Biochemistry* **188**, 245–254 (1990).

50. Lindow, S. E. The use of reporter genes in the study of microbial ecology. *Mol Ecol* **4**, 555–566 (1995).

51. Atomi, H., Imanaka, T. & Fukui, T. Overview of the genetic tools in the Archaea. *Front. Microbio.* **3**, (2012).

52. Li, N. & Cannon, M. C. Gas vesicle genes identified in Bacillus megaterium and functional expression in Escherichia coli. *J. Bacteriol.* **180**, 2450–2458 (1998).

Chapter 2

ENGINEERING APPROACHES FOR ACOUSTIC REPORTER GENES

2.1 Capabilities and Limitations of Existing Acoustic Reporter Genes

While the first-generation bacterial and mammalian ARGs discussed in the previous chapter enabled exciting proof-of-concept imaging demonstrations, they had significant limitations preventing their use in most *in vivo* imaging contexts.

The most exciting capability of first-generation bacterial and mammalian ARGs was that cells expressing them could be detected noninvasively *in vivo* using US. Further, in the case of first-generation mammalian ARGs, their expression could be induced *in situ*, making them the first true acoustic reporters of *in vivo* gene expression. However, they had several significant drawbacks that limited their available applications. Both ARG types suffered from low overall signal (bacterial cells had to be concentration-enriched prior to injection for *in vivo* imaging) and no nonlinear signal, which meant that they could only be imaged *in vivo* using collapse-based imaging paradigms. Further, both showed high metabolic burden and low mutational stability, and first-generation bacterial ARGs had poor expression at 37C which prevented their *in situ* expression *in vivo*.

Based on these limitations, we sought to engineer the following four ARG phenotypes: 1) nonlinear scattering, 2) compatibility with *in situ* expression *in vivo*, 3) high mutational stability, and 4) compatibility with nondestructive multiplexed imaging. We sought ARGs with these properties that could be expressed in either bacterial or mammalian cells at physiological temperature.

In this chapter, I first describe the previous approaches used to engineer ARGs, including their advantages, limitations, and the results obtained from them. Then I review current approaches for high-throughput protein engineering, and discuss their applicability to ARG engineering.

2.2 Previous Acoustic Reporter Gene Engineering Efforts

Prior to the work described in this thesis, a small number of engineering efforts had attempted to improve certain acoustic and expression phenotypes of ARGs.

The first published GV engineering study not only showed that GVs could produce US contrast *in vitro*, but also that the level of this contrast could be increased by aggregating GVs together and decreased by lysing GV-expressing cells. Subsequent work showed that purified GVs could be engineered to collapse at varying US pressures and produce increased nonlinear US contrast,[1] and functionalized to target specific extracellular receptors in vitro[1] and in vivo.[2] However, as these studies used purified GVs rather than ones expressed in cells, they do not represent ARG engineering efforts.

There have been only three previous attempts to engineer ARGs: one that used cluster and expression engineering to alter ARGs' acoustic properties and enhance their expression levels in bacteria,[3] one that used only expression engineering to enable their expression in mammalian cells,[4] and one that used protein and expression engineering to convert ARGs into sensors of intracellular protease activity.[5] The first two of these efforts, while very impactful in their results, involved little to no protein engineering; the third involved substantial protein engineering, but only of the accessory structural protein GvpC.

There were three main limitations to the engineering efforts performed previously: 1) they were limited in their scope, using only a small number of GV gene cluster homologs, with most targeting GvpC and none targeting the primary GV structural protein GvpA; 2) they used only small, rationally-designed libraries; 3) they were conducted in low throughput, requiring manual loading and scanning of each sample that made these experiments highly time-consuming. We determined that before attempting to engineer the four ARG phenotypes listed above, we first needed to develop a high-throughput system for assaying the acoustic phenotypes of large libraries of ARG variants. After developing this system, we would use it to screen libraries of both GV gene cluster homologs and GV structural protein mutants in hopes of identifying improved ARGs with more desirable phenotypes for nondestructive *in vivo* imaging.

2.3 High-Throughput Approaches to Acoustic Reporter Gene Engineering

Here, we sought to move beyond the low-throughout, rational approaches to ARG engineering, towards a high-throughput engineering approach that could handle large variant libraries. In this section, I describe the library types and screening methods commonly used for protein engineering, and discuss the applicability of each to ARG engineering.

2.3a Screening vs. Selection

In a screen, phenotypic information is collected about each variant in a library; based on this information, most variants are discarded, and a small number of top performers are chosen for the next round of engineering. In a selection, all variants are exposed to a condition that will preserve or eliminate them based on an indicator of their "fitness," and only variants fit enough to pass selection are propagated to the next round. Many methods have been devised for screening or selecting mutants based on various phenotypes (see Table 2 of Packer and Liu 2015 for an extensive list),[6] each with different advantages and disadvantages; however, there are some general properties common to screening and selection strategies. Selections tend to have extremely high throughput, which is often limited only by the transformation/transfection efficiency of the library into the expression cells; however, selections generally provide a fairly coarse filter on a given phenotype, and there many phenotypes for which it is difficult or impossible to devise a suitable selection. Conversely, screens can be devised for any phenotype for which either qualitative or quantitative data can be collected, and some very creative phenotype proxy screens have been devised in cases where the phenotype of interest is difficult or slow to assay; however, the throughput of screens is limited by the rate at which this data can be collected for each member of a library, and depending on the screening strategy used, this can be very slow (see below).

While selections are almost always performed in pools in which all variants are mixed in the same sample, screening can be performed in either a pooled format or an arrayed format in which variants are separated into the wells of microplates.[7–9] Arrayed screening offers the significant advantages of easy genotype-phenotype correlation and compatibility with a large number of phenotypic assays, but is costly and time-consuming, and requires specialized equipment often including automation. Pooled screening is often much faster and simpler to implement experimentally, requiring less sample manipulation and no specialized equipment, but generates data that is more complex to process and is compatible with a much smaller number of phenotypic assays.[7]

2.3a(i) Screening and Selection Strategies for ARG Engineering

We considered the above options, and decided that there were two main strategies that could be used to engineer our ARG phenotypes of interest: a pooled selection strategy based on cellular buoyancy, and an arrayed screening strategy based on acoustic phenotypes.

In a pooled buoyancy selection, a library of ARG variants would be transformed into *E. coli*, and following expression, these cells would be centrifuged at a speed high enough to cause the buoyant cells to float to the top of the sample, but not so high that the ARGs collapse. Buoyant cells would be collected and sequenced, and unique mutants would be diversified in the next round of library generation. The advantages of this strategy are that it is extremely high-throughput and can be used to select from >10^8 cells per experiment, with multiple experiments doable in parallel, and that it can be tailored to select for mutants with high mutational stability. The main disadvantage is that it cannot select for all the phenotypes of interest to us: it only selects based on the gas volume fraction per cell, which is an indirect indicator of total nonlinear ARG contrast, and which has no bearing on the compatibility of a given ARG mutant with multiplexed imaging.

In an arrayed acoustic screen, libraries of ARG-expressing *E. coli* would be arrayed in 96-well format, and scanned using US to assay their acoustic phenotypes. These phenotypes would be quantified, and this information used to choose a subset of the top-performing variants to de diversified for the next round of screening. The advantage of this method is that it allows us to directly assay almost all the phenotypes of interest and to collect quantitative data on each. The primary disadvantages are that it is complex to implement, and its throughput is ~10^3 mutants per experiment due to the high screening time required per mutant. Thus, this method would be preferable to pooled buoyancy screening as long as the libraries we plan to screen have sizes on the order of 10^3 mutants.

2.3b Library Types for Protein Engineering

The two main types of variant libraries used for protein engineering experiments are mutant libraries and homolog libraries. In mutant libraries, mutants (either rationally designed or random) of one or more proteins are generated by any of several methods (see below), while in homolog libraries, existing wild type homologs of proteins are mined from metagenomic datasets and tested in place of the protein of interest. Both strategies lend themselves to iterative rounds of library generation and screening: with mutant libraries, the top variants from one round are mutagenized for screening or selection in the next; with homolog libraries, a diverse set of homologs can be sparsely sampled in one round, and homologs more closely related to those with favorable phenotypes can be densely sampled in the next. Both approaches have been used extensively and successfully to evolve a variety of proteins, and each has its strengths and limitations. Ultimately, the choice of library type should maximize the chance that variants with the desired phenotypes are present, while minimizing the number of undesirable mutants.

Below, I will describe the advantages and disadvantages of these library types as they relate to ARG engineering. Ultimately, we decided to construct both mutant and homolog libraries; our efforts to screen these libraries and characterize their best variants are the subjects of Chapters 3 and 4 of this thesis.

2.3b(i) Mutant Library Types and Considerations for ARG Engineering

There are several common methods for mutant library generation (see Table 1 of Packer and Liu 2015[6] for the most commonly used techniques), each with different advantages and disadvantages. Of these, the most ubiquitous ones currently are error-prone PCR, site-directed saturation mutagenesis, recombination, and rational or semi-rational design of computationally focused libraries.

The choice of which methods to use depends primarily on how much is known about the protein of interest. In particular, if there are certain residues that are known to interact with a molecule of interest (*e.g.*, a substrate or cofactor in the case of an enzyme), then a focused mutagenesis strategy (usually site saturation) is used. If a crystal structure is available, then either rational design or a combination of rational design and recombination is used to generate a focused library that is likely to exclude as many non-functional variants as possible. If the protein of interest has several orthologs with sufficiently different sequences (meaning that the sequence is not so highly conserved that recombining sequences from different orthologs would result in a library with very limited diversity) and different functions (which necessitates that the functions of these orthologs have actually been characterized and are interesting), then recombination is often used—even in the absence of a crystal structure—to generate libraries that cover portions of the sequence space that are especially likely to yield both functional and interesting variants. Further, if one or more rounds of library generation and screening have already been performed and interesting variants identified, then these mutants can be recombined to generate variants with pairs of interesting mutations, which are also more likely to retain functionality than a random set of variants. However, in the absence of either structural information, information on key residues known to affect the phenotype of interest, or orthologs with both sufficient sequence diversity and well-characterized interesting functions, a directed evolution experiment must start by using a mutagenesis method that generates a great deal of diversity (which is bound to result in many variants with decreased function but hopefully also a few with increased function), and to use the information gleaned from this library to generate more targeted ones in future rounds of engineering.

With respect to GVs, little useful prior information is available. There was no crystal structure available when these studies began, but computational models have been put forward[10,11] and a limited amount of targeted mutagenesis has been done to both validate these models and identify critical residues for GV formation.[12] There are also many known orthologs of GvpA/GvpB, but many of these have very high sequence similarity in the region predicted computationally to contain the most secondary structure, and little of the information necessary to us is available regarding the functionality of these other orthologs (namely, whether they can be expressed in *E. coli* and what their acoustic properties are). To our knowledge, the only such information available is that which has been collected by our lab, meaning that the only GV proteins that have been both successfully expressed in *E. coli* and characterized in terms of their ultrasound contrast are *Anabaena flos-aquae* GvpA/GvpC and *Megaterium* GvpB. Therefore, given the limited amount of useful information that could otherwise be used to generate a more targeted variant library, we needed to begin by using untargeted mutagenesis approaches and then using the data collected from screening these initial libraries to generate more targeted ones for subsequent rounds. This left us with two options for library generation: error-prone PCR and scanning site saturation.

Error-prone PCR (epPCR) is perhaps the most commonly used strategy for library generation. It relies on the use of a DNA polymerase with tunable fidelity that can be made to introduce mutations randomly into a DNA sequence during PCR. The significant advantages of this method are its ease of implementation and its ability to generate libraries that sample a wide range of sequence space, but its largest disadvantages are that it introduces bias toward certain codons (as not all singlenucleotide mutations are equally likely because of bias inherent to the polymerase) and cannot access all amino acids at each position (*i.e.*, assuming a maximum of one mutation per codon, it is not possible to make certain amino acid substitutions, for example F to G, V to C, M to N, and several others). Further, epPCR generates a range in the number of mutations per library member (*e.g.*, 0-3 mutations per gene), rather than a defined number; thus, in libraries where a small number of mutations per gene are desired, a significant percentage of the library will be the wild type sequence. Also, if the gene of interest is very sensitive to mutation, a large number of the mutant sequences may be nonfunctional. Considering these disadvantages, we chose to forego using epPCR in our experiments.

Scanning site saturation is a version of site-directed saturation mutagenesis in which multiple residues are saturated in the same library, leading to a final library containing a total of: [(# residues to be saturated) x (# amino acids to be substituted in each position)] variants. For example, for a protein of length 70 residues in which every residue is replaced with every amino acid (excluding the wild type amino acid at that position), the resulting library size will be 70 x 19 = 1,330. Because it is customary to substitute every possible amino acid in each desired position, these libraries typically only scale with the number of residues to be saturated, which may mean all the residues in the protein. Until recently, most scanning site saturation experiments have been limited to sub-regions of a protein of interest, as each site required a separate round of cloning. However, with recent advances in oligo synthesis, it is now possible to generate an entire scanning site saturation library as a pool of oligos, and to clone them into a target vector in a single step. This strategy has recently been used in combination with high-throughput screens to evolve a high error-rate DNA polymerase for continuous evolution[13] and to characterize RNA sequence elements that contribute to translation efficiency.[14] The largest advantages of a scanning site saturation library are that it provides a comprehensive view of the location of the starting point in sequence space (*i.e.*, it is equivalent to taking one “step” away from the parent in every direction in sequence space), which is very useful for characterizing a protein of unknown evolvability, and it helps to identify the residues that are most sensitive to mutation and most important for the phenotypes of interest. The main disadvantages of this approach are the technical difficulty in generating such libraries and the cost, but these were not sufficient to dissuade us from attempting to make such information-rich libraries.

2.3b(ii) Homolog Library Types and Considerations for ARG Engineering

Homolog libraries tend to be much smaller than mutant libraries, both because the number of homologous sequences available for a gene or gene cluster of interest in metagenomic databases is often relatively small (at least when compared to the diversity that can be generated for a single protein via mutagenesis), and because the cost per sequence tends to be much higher for homologs, which are obtained either through synthesis or through genomic PCR of an organism harboring them.

The homologous sequences that comprise these libraries are identified through a process of "genome mining,"[15,16] in which a target gene or gene cluster is searched against the sequence data present in databases such as the NCBI Reference Sequence database and the JGI genome database using BLAST. Identified homologs are then compared using ClustalW or antiSMASH and BiGSCAPE/CORASON, and the desired homologs for testing are chosen based on both this sequence comparison information and additional information about the species from which these sequences are derived, such as their growth conditions. These sequences are then obtained through synthesis or genomic PCR of organisms ordered from culture repositories, and cloned into expression vectors for testing. This approach has been used successfully to engineer a variety of genes and gene clusters.[17–22]

Homolog screening has never been applied to ARG engineering. However, because of its demonstrated success in other protein and cluster engineering efforts, as well as the fact that it generates libraries with sizes well within the range of our screening capacity, we saw fit to include an ARG cluster homolog library in our study.

2.4 References

1. Lakshmanan, A. *et al.* Molecular Engineering of Acoustic Protein Nanostructures. *ACS Nano* **10**, 7314–7322 (2016).
2. Hao, Y., Li, Z., Luo, J., Li, L. & Yan, F. Ultrasound Molecular Imaging of Epithelial Mesenchymal Transition for Evaluating Tumor Metastatic Potential via Targeted Biosynthetic Gas Vesicles. *Small* 2207940 (2023) doi:10.1002/smll.202207940.
3. Bourdeau, R. W. *et al.* Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts. *Nature* **553**, 86–90 (2018).
4. Farhadi, A., Ho, G. H., Sawyer, D. P., Bourdeau, R. W. & Shapiro, M. G. Ultrasound imaging of gene expression in mammalian cells. *Science* **365**, 1469–1475 (2019).
5. Lakshmanan, A. *et al.* Acoustic biosensors for ultrasound imaging of enzyme activity. *Nat. Chem. Biol.* **16**, 988–996 (2020).
6. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
7. Bock, C. *et al.* High-content CRISPR screening. *Nat. Rev. Methods Primer* **2**, 8 (2022).
8. Wilson, E. L. & Metzakopian, E. ER-mitochondria contact sites in neurodegeneration: genetic screening approaches to investigate novel disease mechanisms. *Cell Death Differ.* **28**, 1804–1821 (2021).
9. Agrotis, A. & Ketteler, R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Front. Genet.* **6**, (2015).

10. Bayro, M. J., Daviso, E., Belenky, M., Griffin, R. G. & Herzfeld, J. An Amyloid Organelle, Solid-state NMR Evidence for Cross-β Assembly of Gas Vesicles. *J. Biol. Chem.* **287**, 3479–3484 (2012).
11. Strunk, T. *et al.* Structural model of the gas vesicle protein GvpA and analysis of GvpA mutants in vivo: Structure of gas vesicle protein GvpA. *Mol. Microbiol.* **81**, 56–68 (2011).
12. Knitsch, R., Schneefeld, M., Weitzel, K. & Pfeifer, F. Mutations in the major gas vesicle protein GvpA and impacts on gas vesicle formation in *Haloferax volcanii*: Mutations in the major gas vesicle protein. *Mol. Microbiol.* **106**, 530–542 (2017).
13. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable continuous evolution of genes at mutation rates above genomic error thresholds. (2018) doi:10.1101/313338.
14. Cambray, G., Guimaraes, J. C. & Arkin, A. P. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4238.
15. Albarano, L., Esposito, R., Ruocco, N. & Costantini, M. Genome Mining as New Challenge in Natural Products Discovery. *Mar. Drugs* **18**, 199 (2020).
16. Scherlach, K. & Hertweck, C. Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* **12**, 3864 (2021).
17. Johns, N. I. *et al.* Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nat. Methods* **15**, 323–329 (2018).
18. Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10**, 99–105 (2014).
19. Vo, P. L. H. *et al.* CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* **39**, 480–489 (2021).
20. Williams, A. N., Sorout, N., Cameron, A. J. & Stavrinides, J. The Integration of Genome Mining, Comparative Genomics, and Functional Genetics for Biosynthetic Gene Cluster Identification. *Front. Genet.* **11**, 600116 (2020).
21. Mak, W. S. *et al.* Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat. Commun.* **6**, 10005 (2015).
22. Singh, G., Dal Grande, F. & Schmitt, I. Genome mining as a biotechnological tool for the discovery of novel biosynthetic genes in lichens. *Front. Fungal Biol.* **3**, 993171 (2022).

Chapter 3

DIRECTED EVOLUTION OF ACOUSTIC REPORTER GENES USING HIGH-THROUGHPUT ACOUSTIC SCREENING

Hurt R.C., Jin Z., Soufi M., Wong K., Sawyer D., Deshpande R., Shen H., Mittelstein D.R., Shapiro M.G. "Directed evolution of acoustic reporter genes using high-throughput acoustic screening." *In preparation*.

This chapter is a reformatted version of the above manuscript. My contributions to the work were designing and performing the study, analyzing the data, and writing the paper.

3.1 Abstract

A major challenge in the fields of biological imaging and synthetic biology is noninvasively visualizing the functions of natural and engineered cells inside opaque samples such as living animals. One promising technology that addresses this limitation is ultrasound, with its penetration depth of several cm and spatial resolution of tens of µm. Recently, the first reporter genes for ultrasound were developed to link ultrasound to cellular function via heterologous expression in commensal bacteria and mammalian cells. These acoustic reporter genes (ARGs) represent a novel class of genetically encoded ultrasound contrast agent, and are based on air-filled protein nanoparticles called gas vesicles (GVs). Just as the discovery of fluorescent proteins was followed by the improvement and diversification of their optical properties through directed evolution, here we describe the evolution of GVs as acoustic reporters. To accomplish this task, we introduce a strategy for high-throughput, semi-automated acoustic screening of ARGs in bacterial cultures and use it to screen mutant libraries to identify variants with increased nonlinear ultrasound scattering when expressed in *E. coli* at 37°C. Starting with scanning site saturation libraries for two homologs of the primary GV structural protein, GvpA, two rounds of evolution resulted in GV variants with 5 and 10-fold stronger acoustic signal than the two parent proteins, and identified variants with differential responses to acoustic pressure that enabled multiplexed imaging. To our knowledge, this is the first time that an acoustic biomolecule has been evolved in the laboratory, and ours is the first high-throughout, semi-automated setup to collect acoustic data on biomolecules expressed in cells. Finally, the ARG constructs evolved in this study enabled the first-ever demonstration of non-destructive multiplexed imaging of multiple ARG species in a single sample, adding multiple orthogonal "colors" of ARGs to the acoustic molecular imaging toolkit.

3.2 Introduction

Acoustic reporter genes (ARGs)—genetically encoded reporters that enable the imaging of gene expression using ultrasound—which were first introduced to bacteria in 2018[1] and subsequently to mammalian cells in 2019,[2] have been the subject of intense study[3–9], development,[10] and application[11–20] in recent years.[21–23] Based on genetically encoded, gas-filled protein nanoparticles called gas vesicles (GVs) that first evolved in buoyant microbes, ARGs have received considerable attention due to their ability to enable noninvasive, long-term, real-time imaging of gene expression in both bacterial and mammalian cells deep inside living organisms. In particular, ARGs have been used to image tumor growth[2,10] and colonization by therapeutic bacteria,[10] protease activity,[18] and phagolysosomal function,[16] and are anticipated to enable breakthroughs in microbiome imaging, cell tracking, and sonogenetics. However, despite several successful efforts to engineer the acoustic and expression properties of ARGs, further improvements to the performance of ARGs are needed to enable their most impactful applications, and the methods available for their acoustic characterization are low-throughput, complex to implement, and require a great deal of hands-on time per sample.

The exciting possible applications of improved ARGs necessitates a method for quickly and easily engineering their acoustic phenotypes in high throughput. Current methods for ultrasound-based characterization of ARGs are slow and labor-intensive, requiring manual loading and imaging of individual samples that limits their throughput to a handful of samples per day. In contrast, the state-of-the-art high-throughput methods used to engineer fluorescent proteins can process far larger libraries in shorter times, with less intervention from users: plate readers can assay thousands of samples per run, and flow cytometers have been used to screen libraries of 10^8 mutants in a single experiment[24]. In the past few decades, a growing suite of protein engineering techniques have been developed and applied with remarkable success to improving fluorescent proteins, opsins, Cas proteins, and other biotechnology tools, but these methods often require the screening of libraries containing thousands of members or more.[25] Thus, the low throughput of current acoustic screening methods prevents the effective use of most of the tools needed to unlock the full potential of ARGs.

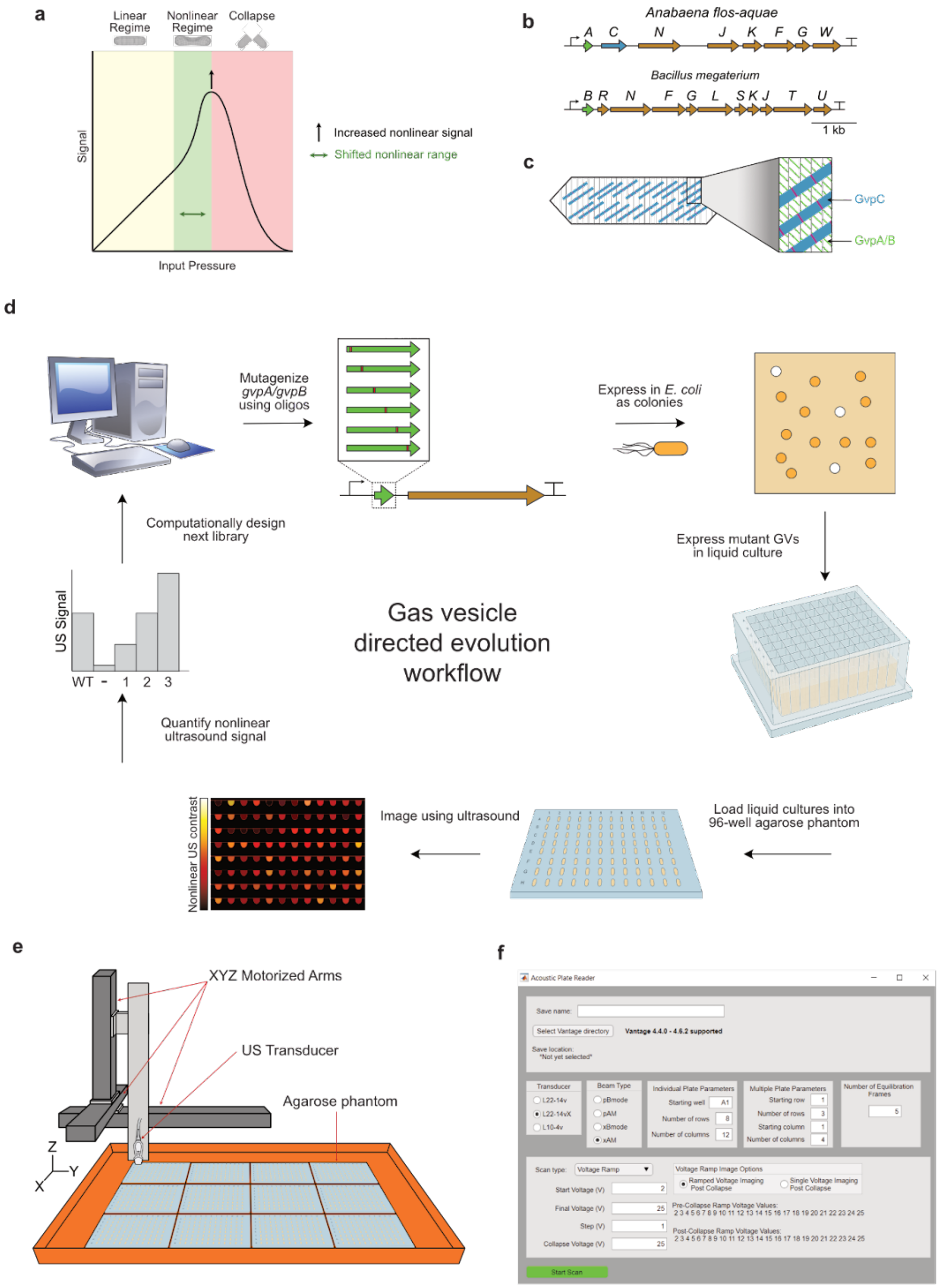
In this study, we developed a high-throughput, semi-automated pipeline for acoustic screening of ARGs, and used it to evolve two ARG clusters to improve their nonlinear acoustic signal and enable nondestructive multiplexing. Our system, which we call the Acoustic Plate Reader (APR), is capable of collecting acoustic data on up to 1152 ARG samples in a single automated scan, and includes graphical user interfaces (GUIs) for data collection and processing. This new workflow facilitates faster, more reliable, and more standardized acoustic screening of ARG samples, and requires significantly less hands-on time than current methods. Using this pipeline, we improved the nonlinear acoustic signal produced by two ARG clusters—the wild type *Anabaena flos-aquae* and *Bacillus megaterium* clusters—by 5- and 10-fold, respectively, when expressed at physiological temperature. We further demonstrated that these evolved ARGs can be multiplexed nondestructively, allowing us to separate distinct populations of cells in an imaging volume, each labeled with a different "color" of ARG.

3.3 Results

High-throughput directed evolution workflow for ARGs

GVs are known to respond to ultrasound in three regimes, depending on the input pressure applied: linear scattering, nonlinear scattering, and collapse[17] (Figure 3-1A). Of particular interest for *in vivo* imaging is the nonlinear scattering regime in which GVs produce significantly more contrast than tissue, putatively by "buckling" of their shells.[26] This effect has been exploited previously to non-destructively image GV-expressing bacterial cells *in vivo* with high specificity;[10] however, the bacterial ARG used has such high nonlinear scattering that it cannot be effectively used in combination with any other bacterial ARGs, which produce significantly less nonlinear signal. Therefore, we sought a pair of bacterial ARGs which: 1) produce considerable and comparable levels of nonlinear signal when expressed in *E. coli* at 37°C, 2) have distinct pressure-response properties that will make them separable during pressure-domain multiplexing, and 3) have compatible collapse pressures, such that imaging one species at its optimal imaging pressure will not substantially collapse the other.

As starting points for engineering the desired pair of bacterial ARGs, we chose two ARG gene clusters which: 1) have been successfully expressed in *E. coli*, but perform sub-optimally in terms of their US contrast and tolerability of expression; 2) have been successfully expressed in mammalian cells. These were the minimal versions of both the wild type *Bacillus megaterium* ATCC 19213 cluster[27] (lacking *gvpA*, *gvpP*, and *gvpQ*) and the wild type *Anabaena flos-aquae* cluster (with only one copy of *gvpA*, and lacking *gvpV*) (Figure 3-1B). The primary GV structural protein--GvpA or its homolog GvpB--creates the cigar-shaped body of the GV, and optionally GvpC may attach to the outside of this structure and reinforce it mechanically (Figure 3-1C). It has already been shown that engineering GvpC to reduce its binding to GvpA can result in GVs with increased nonlinear signal or decreased collapse pressure,[28] but GvpC serves as a limited target for engineering these phenotypes because: 1) not all GV types include GvpC, and 2) even the complete removal of GvpC only increases nonlinear signal by a factor of ~4; while it decreases the average collapse pressure by an impressive ~34%,[28] this is still ~6x higher than the theoretical minimum collapse pressure of 1 atm. Further, it would be desirable to engineer GVs with higher collapse pressures or ones whose collapse pressure is unchanged while having a significantly



Save name:

Vantage 4.4.0 - 4.6.2 supported

Save location:
"Not yet selected"

Transducer
○ L22-14v
● L22-14vX
○ L10-4v

Beam Type
○ pBmode
○ pAM
○ xBmode
● xAM

Individual Plate Parameters
Starting well: A1
Number of rows: 8
Number of columns: 12

Multiple Plate Parameters
Starting row: 1
Number of rows: 3
Starting column: 1
Number of columns: 4

Number of Equilibration Frames: 5

Scan type: Voltage Ramp

Start Voltage (V): 2
Final Voltage (V): 25
Step (V): 1
Collapse Voltage (V): 25

Voltage Ramp Image Options
● Ramped Voltage Imaging Post Collapse
○ Single Voltage Imaging Post Collapse

Pre-Collapse Ramp Voltage Values:
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Post-Collapse Ramp Voltage Values:
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 3-1. High-throughput directed evolution workflow for ARGs. (A) Three regimes of GV response to ultrasound. (B) Diagrams of the gene clusters used as starting points for evolution. (C) Roles of the structural proteins GvpA/B and GvpC in GV structure. (D) Schematic of directed evolution workflow for ARGs. The starting point GV structural protein is mutagenized, then expressed in *E. coli* as colonies on Petri dishes. Colonies that turn white are assumed to produce GVs, and are picked and expressed in liquid culture. Cultures of GV-expressing *E. coli* are then loaded into agarose phantoms and imaged using US. The resulting nonlinear US intensity data are used to rank the performance of mutants and select the most promising ones for further mutagenesis. (E) Schematic of the Acoustic Plate Reader (APR), which is used for automated US image collection of up to 1152 samples of GV-expressing *E. coli* arrayed in 96-well agarose phantoms. (F) Image of the graphical user interface for the APR.

lower buckling threshold. Therefore, we chose the GvpA/B protein as our target for engineering increased nonlinear scattering with or without altered collapse pressure.

To engineer the desired nonlinear signal and collapse pressure phenotypes, we developed a method for high-throughput, semi-automated characterization of US contrast and GV collapse pressure in *E. coli* (Figure 3-1D). After constructing scanning site saturation libraries of *gvpA* or *gvpB* in the starting point clusters (Figure 3-1B), we first performed a selection for high levels of GV expression by inducing transformants on Petri dishes and picking only colonies that appeared white (GV-expressing bacteria appear white because GVs scatter light, in addition to US). These mutants were then expressed in liquid cultures in 96-well format and loaded into agarose phantoms. These phantoms were imaged using an automated scanning setup in which a MATLAB-controlled 3D translating stage raster scans an US transducer above the submerged phantoms (Figure 3-1E), producing a set of US images in which samples with high GV expression appear bright. This pipeline allowed us to generate and acoustically screen several mutant libraries, from which we identified mutants with significantly enhanced acoustic phenotypes. We also created graphical user interfaces (GUIs) to simplify and standardize data acquisition (Figure 3-1F) and analysis. We termed this setup the “Acoustic Plate Reader,” the complete workflow for which is shown in Supplementary Figure 3-1.

Optimization of GV expression from WT* Anabaena flos-aquae *and* Bacillus megaterium *gene clusters

Before engineering the structural proteins in these ARG clusters, we optimized the expression of the WT *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters to increase nonlinear acoustic signal when expressed in liquid culture in *E. coli* at 37°C. For each cluster, we cloned three ORIs of different strengths (~40, ~20, and ~5 copies/cell) (Fig. 3-2A-B), and assessed their performance in liquid culture as a function of inducer concentration. For both clusters, the strongest ORI tested gave the highest nonlinear ultrasound signal (Fig. 3-2C-D), and was chosen for future experiments. With the optimal ORIs selected for expression (Fig. 3-2E-F), we then sought to optimize the autoinduction conditions to maximize nonlinear signal (in autoinduction media, increasing the concentration of glucose increases the OD at which induction occurs, while increasing the concentration of the inducing sugar increases the level to which the transcription unit is induced). We performed titrations of glucose and arabinose and assessed the resulting nonlinear signal from the expressed constructs (Fig. 3-2 G-H), and decided on concentrations of 0.25% glucose and 0.05% arabinose for induction of these constructs in future experiments. We observed that US signal from the *Anabaena flos-aquae* cluster peaked at a moderate arabinose concentration (Fig. 3-2C and G), while expression from the *Bacillus megaterium* cluster was highest at the maximum concentration (Fig. 3-2D and H). We suspect that the signal decline from the *Anabaena flos-aquae* cluster at high arabinose concentrations is due to the high metabolic burden associated with expressing so many non-native proteins in *E. coli*.

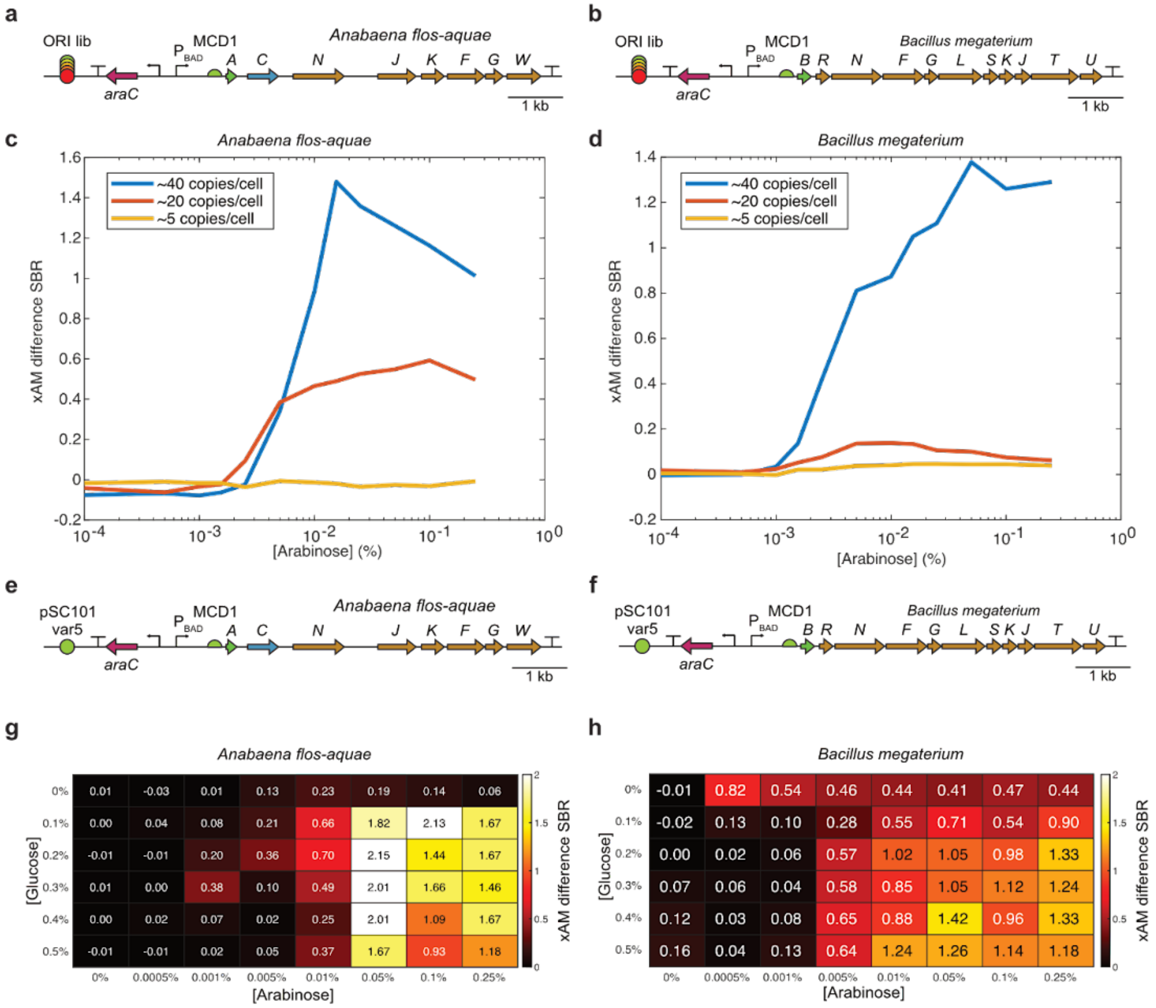


Figure 3-2. Optimization of GV expression from WT *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters. (A-B) Diagrams of the WT *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters with libraries of origins of replication (ORI) of different strengths. (C-D) Nonlinear US signal produced from expression of both clusters at three different copy numbers as a function of inducer concentration. (E-F) Diagrams of the optimized WT *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters used for directed evolution, both of which used the pSC101-var5 ORI (~40 copies/cell). (G-H) Nonlinear US signal quantifications for both WT clusters as a function of the concentrations of glucose and arabinose used for autoinduction. The concentrations selected for GV expression during library screening were 0.25% glucose and 0.05% arabinose.

Round 1 directed evolution of Anabaena flos-aquae GvpA and Bacillus megaterium GvpB

To further improve the nonlinear signal from the WT *Anabaena flos-aquae* and *B. megaterium* clusters in *E. coli*, we designed scanning site saturation libraries of the genes encoding the primary GV structural protein for each (*i.e.*, *gvpA* for *Anabaena flos-aquae*; *gvpB* for *B. megaterium*) (Figure 3-3A-B). This resulted in libraries containing 1400 and 1740 members for *gvpA* and *gvpB*, respectively (Table 1 Library Round 1). We constructed these libraries using a Golden Gate-based

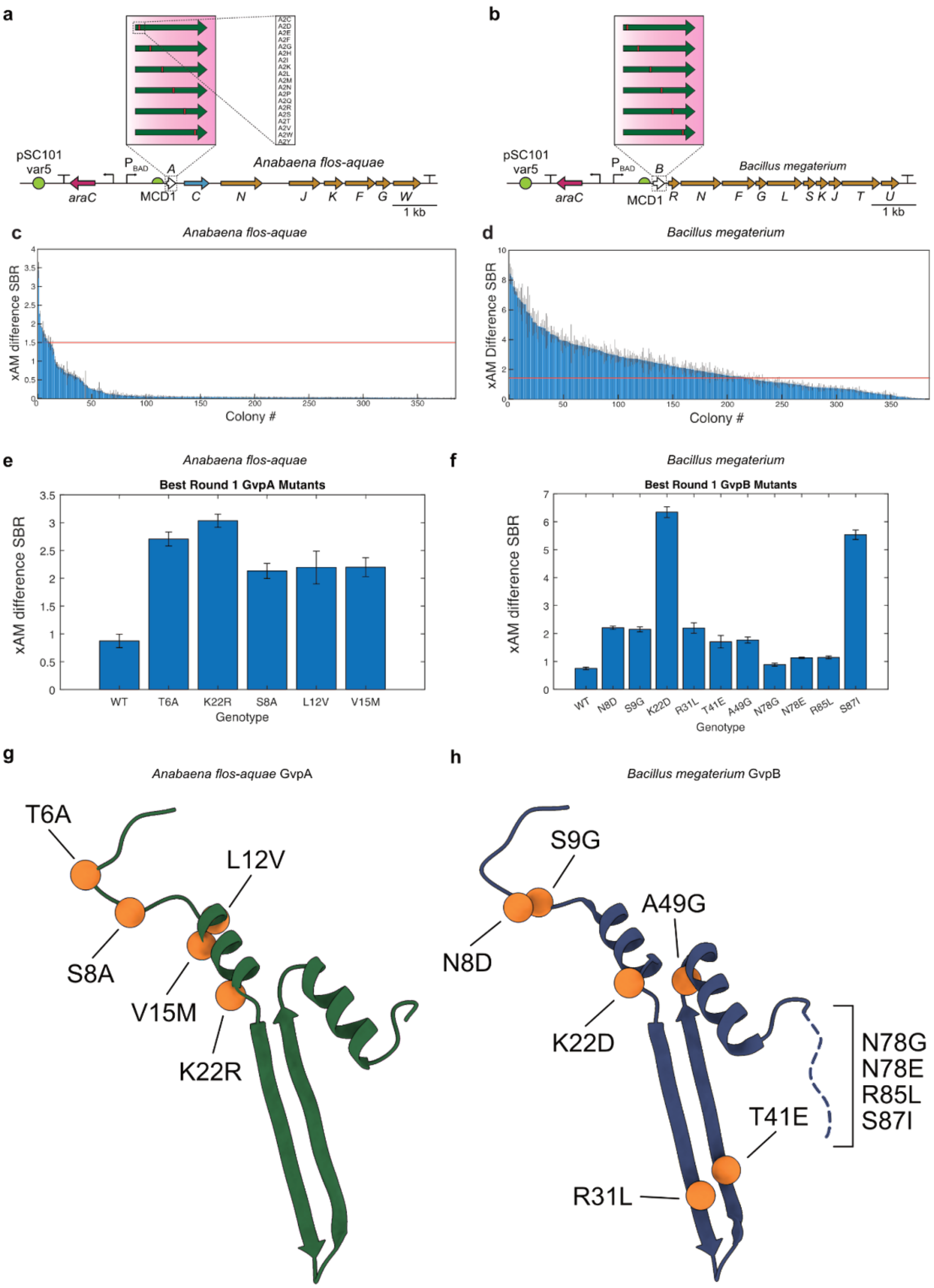


Figure 3-3. First round of directed evolution of *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters. (A-B) Diagrams of the mutagenized *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters,

depicting the scanning site saturation libraries screened in the first round of evolution. (C-D) Nonlinear US difference signal-to-background ratio (SBR) from all screened mutants of both clusters. Red lines indicate the difference SBR of the WT for that cluster. N=3 technical replicates of one biological sample. (E-F) Nonlinear US difference SBR for the WT and top mutants for each cluster. Error bars represent standard error. (G-H) Locations of top mutations from E-F in the GvpA/GvpB structure.

version of cassette mutagenesis, in which mutagenic oligos that tile the gene of interest are synthesized and cloned into an acceptor vector (Supplementary Figure 3-2A-B; see methods for details). We chose this approach over error-prone PCR because of its ability to generate defined libraries which have a controllable number of mutations per member and which lack unwanted mutants (*i.e.*, premature stop codons and multiple codons that code for the same mutant). When induced in solid culture, these libraries produce three distinct types of colonies: 1) blue colonies, in which the dropout chromoprotein was not excised during assembly, returning the original acceptor vector; 2) low-opacity colonies that lack GV expression or express small amounts, either because they contain a mutant that reduces GV expression or because the mutant gene did not insert correctly during assembly; 3) high-opacity colonies with high GV expression. We used this readout to select only the mutants with high GV expression for further study. We then expressed these mutants (384 from each of the two libraries) in 96-well liquid cultures, and imaged them in the Acoustic Plate Reader in 96-well agarose phantoms (Figure 3-1D and Supplementary Figure 3-1). We chose the 10 unique mutants with the highest US signal from each library and re-cloned them for validation and further characterization (see methods).

Round 2 directed evolution of Anabaena flos-aquae GvpA and Bacillus megaterium GvpB

We next performed a second round of directed evolution on these clusters by generating three distinct libraries: 1 and 2) scanning site saturation libraries of the top two mutants of *Anabaena flos-aquae gvpA*: T6A and K22R; 3) a paired recombination library of the top 10 unique mutants of *B. megaterium gvpB* (Figure 3-4A-B). We cloned and screened these libraries using the same methods described for the first first round of evolution (Figure 3-1D), and identified several mutants with greatly improved signal over their parents in both libraries (Figure 3-4C-D). We characterized the top 10 unique mutants from each library in terms of their nonlinear acoustic signal (Figure 3-4E-F) and OD600 (Figure 3-4G-H), and identified GvpA-T6A-L40A and GvpB-S9G-R31L-R85L as the top-performing variants.

Non-destructive multiplexed imaging of two ARG types

Finally, we sought to achieve nondestructive multiplexing of multiple GV types. For the two best-performing mutants from our screen, Ana-GvpA-T6A-L40A and Mega-GvpB-S9G-R31L-R85L (Figure 3-5A), we characterized their nonlinear acoustic signal as a function of the US pressure applied (Figure 3-5B), and observed that the pressure-response curves they produced were non-overlapping when normalized. This difference suggested that they were candidates for a pressure-

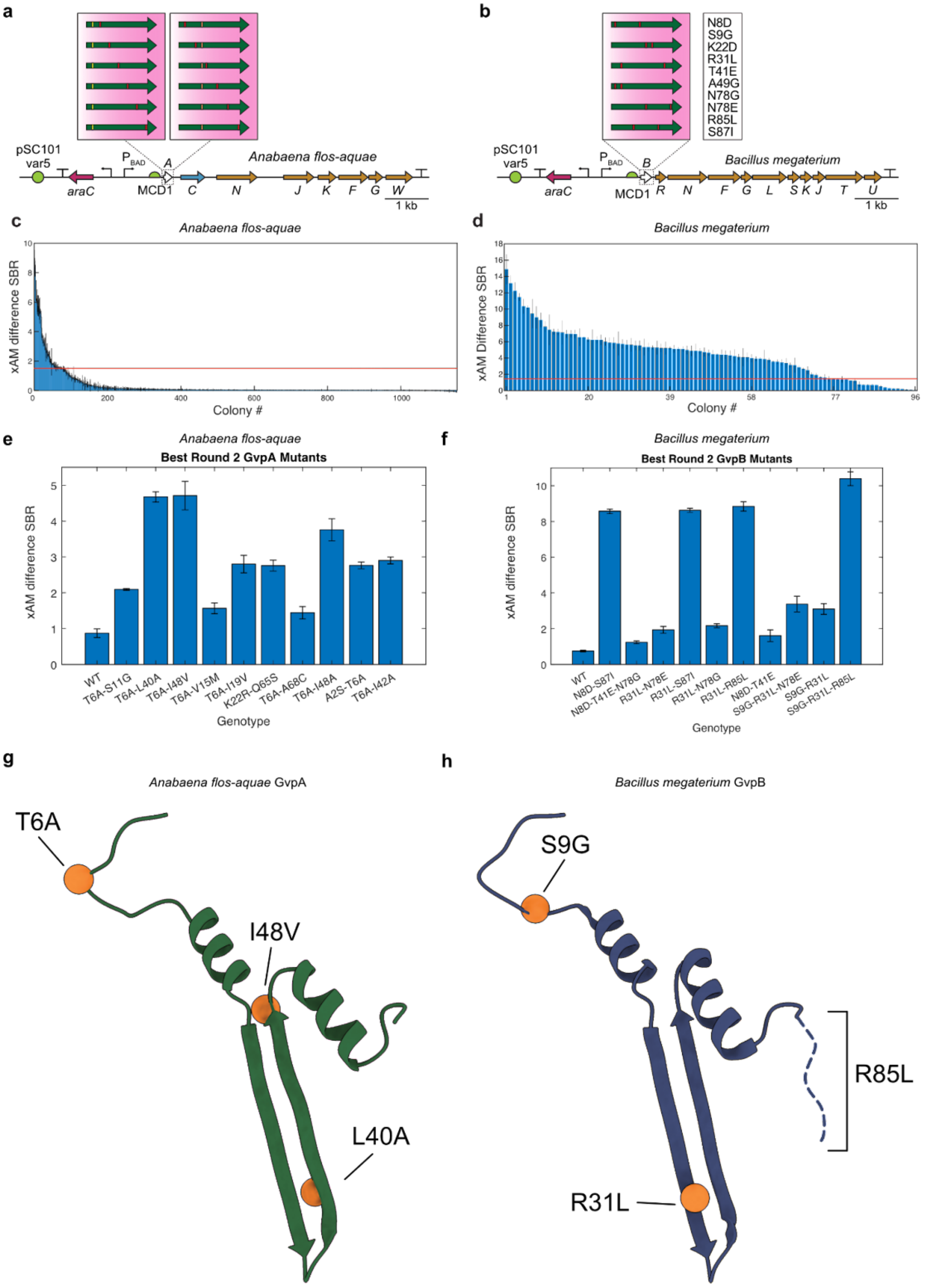


Figure 3-4. Second round of directed evolution of *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters. (A-B) Diagrams of the mutagenized *Anabaena flos-aquae* and *Bacillus megaterium* gene clusters

used in the second round of evolution. The best two mutants of *Anabaena flos-aquae* gvpA were used as parents for another scanning site saturation library, and the best ten mutants of *Bacillus megaterium* gvpB were used to create a paired recombination library. (C-D) Nonlinear US difference signal-to-background ratio (SBR) from all screened mutants of both clusters. Red lines indicate the difference SBR of the WT for that cluster. N=3 technical replicates of one biological sample. (E-F) Nonlinear US difference SBR for the WT and top ten mutants for each cluster. Error bars represent standard error. (G-H) Locations of mutations from the top mutants from E-F in the GvpA/GvpB structure.

domain multiplexing approach, in which a spectral unmixing technique is used to discriminate between distinct species within a mixed population based on their characteristic spectral "fingerprints."[29] We imaged samples of each ARG type both separately and mixed in a 1:1 ratio; to unmix their signals, we used a pressure spectral unmixing paradigm previously described,[28] but modified it to use a differential nonlinear turn-on matrix rather than a differential collapse matrix to discriminate between ARG types. Using this approach, we were able to separate the signals from these ARG types with moderate accuracy (Figure 3-5C) without the need for ARG collapse.

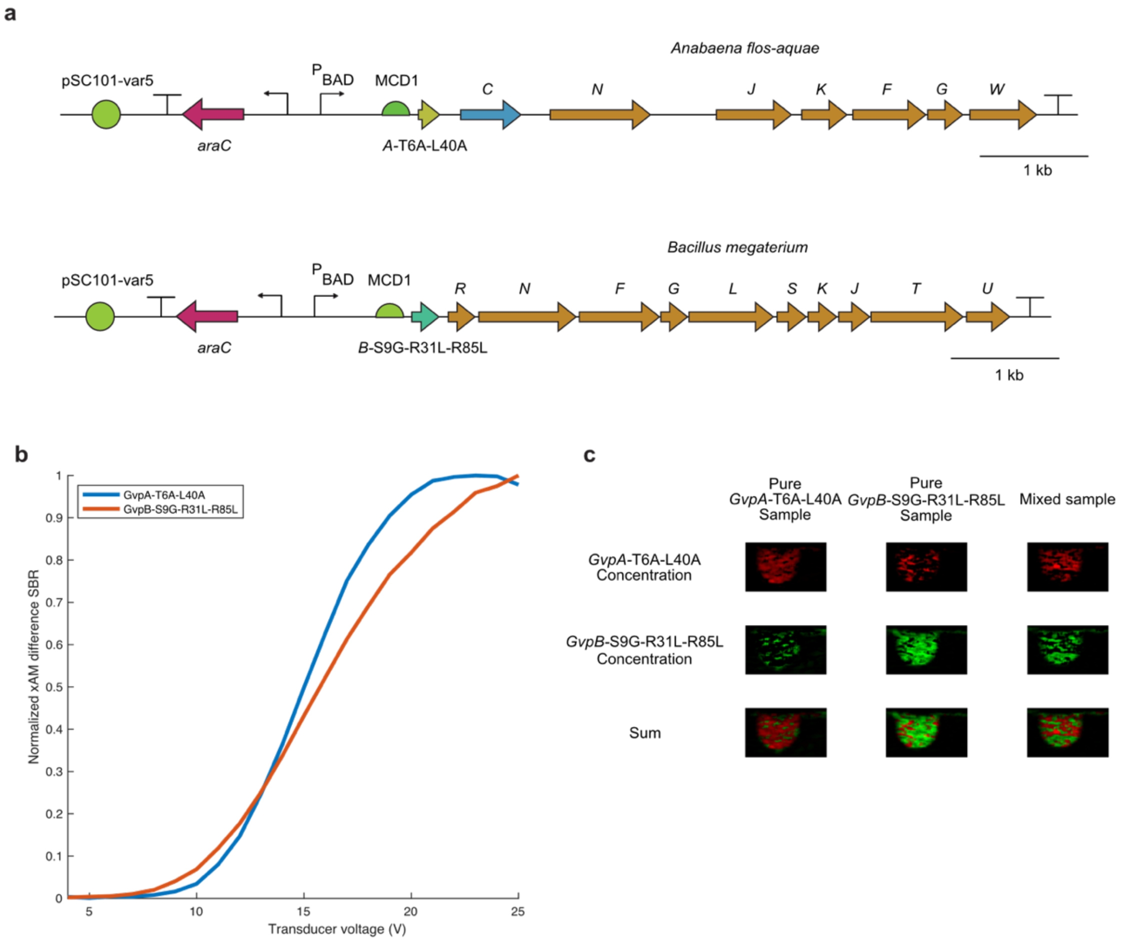


Figure 3-5. Non-destructive multiplexed imaging of two ARG types. (A) Diagrams of the evolved ARG clusters used for multiplexed imaging. (B) Normalized pressure response functions of both ARG types. (C) Non-destructive multiplexing of two ARG types. Row titles indicate the predicted species in the image, and column titles indicate the ground truth species.

3.4 Discussion

Our results establish the first ever method for high-throughput, semi-automated acoustic screening of biomolecules expressed in cells. When used to evolve two ARG clusters—those from *Anabaena flos-aquae* and *Bacillus megaterium*—this method yielded a pair of ARG constructs which not only show considerable improvements in their nonlinear acoustic scattering, but which also enabled—for the first time—the nondestructive multiplexed imaging of multiple ARG types in a single sample.

These results represent a major advance in the way that acoustic biomolecules can be engineered. In the same way that high-throughput screening tools like plate readers and flow cytometry revolutionized the engineering of fluorescent proteins and the many sensors derived from them (*e.g.*, those for ions, small molecules, and membrane voltage) by dramatically increasing the sizes of the libraries that can be screened in these experiments, so too will the increased throughput, reliability, and standardization introduced by the Acoustic Plate Reader enable the engineering of next-generation ARGs and their derivatives (*e.g.*, sensors of proteases, pH, and Ca^{2+}).

The constructs evolved in this study not only show considerably higher nonlinear scattering when expressed in *E. coli* than their parents, but they also have properties that make them compatible with non-destructive pressure-domain multiplexing: their levels of nonlinear scattering are similar to each other, and the optimal US pressure for imaging the Mega-GvpB-S9G-R31L-R85L does not substantially collapse Ana-GvpA-T6A-L40A. Previous approaches to pressure-domain multiplexing of ARGs were based on linear scattering and required ARG collapse,[1,28] making them incompatible with *in vivo* imaging or long-term monitoring of ARG signal; further, these demonstrations were performed on purified and concentration-enriched GVs, rather than on genetically encoded ARGs imaged *in situ* inside cells. This new nondestructive multiplexing capability, if extended to *in vivo* imaging contexts, will enable discrimination between bacterial and mammalian cells in close proximity, such as in a bacterially-colonized tumor, and between different strains of engineered bacteria in a consortium.[30–32] Finally, the high collapse pressure of Mega-GvpB-S9G-R31L-R85L ARGs will make them useful for acoustic cell sorting applications, in which high US pressures are used to separate ARG-expressing cells from non-expressing cells.[11]

While these evolved constructs represent substantial improvements over their parents, further improvements are required for some of their most exciting applications to be realized. First, both ARGs could benefit from further improvements in nonlinear contrast; this will likely be achieved through a combination of protein engineering (including not only the structural proteins engineered in this study, but also the assembly factors that assist in GV formation) and expression tuning (ORI, RBSs, and promoter) aimed at increasing both the amount of nonlinear contrast produced per GV and the number of GVs produced per cell. Second, the mutational stability of these constructs will likely need to be improved for them to be useful *in vivo* where selection antibiotics are usually not available at sufficient concentrations. The primary strategies used to increase the stability of bacterial expression constructs *in vivo* are chromosomal integration and inclusion of plasmid stability elements such as toxin/antitoxin or plasmid partitioning systems; while the latter can be easily added to plasmids bearing these ARGs, chromosomal integration will likely prove problematic because of the substantial decrease in copy number when moving from a medium-copy plasmid (~40 copies/cell in this case) to chromosomal integration (1-3 copies/cell). This change will likely necessitate further tuning of expression levels of all proteins in these clusters at both the transcriptional (via promoter tuning) and translational (via RBS tuning) levels; fortunately, high-throughput cloning methods such as MoClo and simplified systems for chromosomal integration enable the rapid generation of these constructs and cell lines, and when combined with the increased screening throughput offered by the Acoustic Plate Reader will greatly accelerate these and other engineering efforts. Third, the addition of a third ARG species multiplexable with the current two will enable labeling of an additional cell population; however, using the current pressure-domain multiplexing approach, care must be taken to avoid collapsing the species that buckle at low pressures while imaging those that buckle at higher pressures. To this end, the ideal ARG type for pressure-domain multiplexing is one that can buckle in a wide range of pressures without collapsing; fortunately, this phenotype can be easily screened for using the existing Acoustic Plate Reader imaging functions. Finally, it will be interesting to test these constructs *in vivo*: for example testing both mutants in tumor-homing bacteria (*e.g.*, *E. coli* Nissle 1917 and *Salmonella enterica* serovar Typhimurium) colonizing a tumor, or testing Ana-GvpA-T6A-L40A in mammalian cells to assess its acoustic and expression performance relative to the current mARG$_{\text{Ana}}$, which is based on the wild type *Anabaena flos-aquae* cluster.

This method will help high-throughput protein engineering to play as large a role in the engineering of acoustic biomolecules as it has in the development of their fluorescent counterparts.

3.5 Methods

Plasmid construction via MoClo

The EcoFlex MoClo system[33] was used to create all vectors cloned in this study, including existing parts (Addgene Kit # 1000000080) and custom-made parts (Table 2). Custom-made parts were introduced into the existing EcoFlex system as follows: 1) ORIs were selected from the pSC101-varX series[34]; promoters were selected from the Marionette series[35]; RBSs were selected from the MCDX series[36]; terminators were selected from the ECK and LXSXPX series[37]; 2) parts were either synthesized as fragments (Twist Bioscience) and subsequently PCRed using Q5 (NEB), or synthesized as duplex oligos (IDT); 3) parts were cloned into the corresponding part acceptor vector (Table 2) via Golden Gate to ensure that they received the appropriate assembly overhangs. EcoFlex assemblies were conducted as described in Supplementary Note 1 and electroporated into NEB Stable *E. coli* (except for the MetClo-based library acceptor vectors, which were transformed into DH10B-M.Osp807II[38]). Transformations were recovered for 2 hr in 1 mL of SOC at 37°C and 250 RPM, and plated on Petri dishes containing Lennox LB with 1% agar, 100 ug/mL kanamycin, and 1% glucose (for catabolite repression of the PBAD promoter). Colonies were picked into 1.5 mL liquid cultures of Lennox LB with 100 ug/mL kanamycin and 1% glucose in 96-well format and grown overnight to saturation. These cultures were then miniprepped using reagents and a protocol from Qiagen, a lysate clearing plate from Bio Basic (SD5006), and a DNA-binding plate from Epoch Life Sciences (2020-001). All constructs were verified by whole-plasmid sequencing (Primordium Labs).

Liquid culture GV expression in E. coli

GVs were expressed in *E. coli* liquid cultures in 96-well format according to the following general protocol, with modifications for specific experiments described below.

Miniprepped DNA was electroporated into NEB Stable *E. coli*, and transformations were recovered for 2 hr at 37°C in 1 mL of SOC. Transformations were then inoculated at a dilution of 1:100 into autoinduction Lennox LB containing 100 μg/mL kanamycin, 0.6% glycerol, and the appropriate concentrations of glucose and inducer for the experiment (see below). These expression cultures were set up in 500 uL volumes in deep-well 96-well plates (square wells used for maximum culture aeration; USA Scientific 1896-2800) sealed with porous tape (Qiagen 19571) and incubated at 37°C and 350 RPM for 20 hr. Cultures were stored at 4C until being loaded into phantoms for Acoustic Plate Reader scans. For the concentrations of glucose and arabinose described below, in experiments where titrations were used, 100X stocks of these sugars were prepared in 1X PBS and diluted 1:100 into the cultures when setting up the experiments.

The following concentrations were used for the ORI titration experiments shown in Fig. 3-2A-D: glucose: 0.25%; arabinose: 0, 0.0005, 0.001, 0.00155, 0.0025, 0.005, 0.01, 0.0155, 0.025, 0.05, 0.1, 0.25%.

The following concentrations were used for the parent expression optimization experiments shown in Fig. 3-2E-H: glucose: 0, 0.1, 0.2, 0.3, 0.4, 0.5%; arabinose: 0, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.25%.

The following modifications were made for the library screening experiments shown in Fig. 3-3A-C and 3-4A-C: 1) assembled libraries were transformed multiple times into NEB Stable *E. coli*, and it was ensured that the number of transformants produced was at least 100X the number of unique sequences expected in the library; 2) prior to expression in liquid culture, libraries were expressed in solid culture as colonies on Lennox LB with 100 ug/mL kanamycin, 0.6% glycerol, 0.25% glucose, and 0.05% arabinose at a density of ~100 colonies/dish. Colonies were grown for 48 hr at 37°C, and 380 opaque colonies were picked for each library, as well as 4 colonies for the library's parent, into the wells of 96-well PCR plates containing 100 uL of Lennox LB with 100 µg/mL kanamycin and 1% glucose, and grown to saturation overnight at 30°C. These saturated liquid cultures, rather than transformations, were used to set up expression cultures as described above; 3) 0.25% glucose, and 0.05% arabinose were used to induce expression in these experiments.

The following concentrations were used for the mutant expression experiments shown in Fig. 3-3E-H and 3-4E-H: glucose: 0.25%; arabinose: 0.05%.

The following concentrations were used for the multiplexing experiments shown in Fig. 3-5A-B: glucose: 0.25%; arabinose: 0.05%.

Scanning site saturation and recombination library generation

Scanning site saturation libraries were generated via a Golden Gate-based version of the cassette mutagenesis strategy previously described.[39] Briefly, the *Anabaena flos-aquae* GvpA and *Bacillus megaterium* GvpB coding sequences were divided into sections that tiled the gene, and oligos were designed to have a variable middle region with flanking constant regions against which PCR primers were designed (these primers also contain the evSeq[40] inner adapters for optional deep sequencing of the library). Depending on the library being created (*i.e.*, scanning site saturation or recombination), the variable region was designed to either sequentially saturate each residue or recombine pairs of the mutations listed in Fig. 3-4b (mutations identified during screening of the first round of scanning site saturation GvpB). The MATLAB scripts used to generate the oligo sequences for both the scanning site saturation and recombination libraries are available in Supplementary Code 1, and the oligo sequences themselves are listed in Table 1. Oligos were synthesized as a pool by Twist Biosciences or Integrated DNA Technologies, and were amplified by PCR (both to make them double-stranded and to generate enough DNA for Golden Gate assembly) using KAPA HiFi HotStart ReadyMix according to the manufacturer's instructions, but with 10 cycles, 100 ng of oligo pool template, and 1 uM of each primer. PCR products were run on a 2% agarose gel and purified using Qiagen reagents according to the manufacturer's instructions, but with a 5 uL final elution volume of water. Fragments were then assembled with the corresponding library acceptor vector (Table 2) in a Golden Gate reaction using reagents from New

England Biolabs according to the manufacturer's instructions. Assemblies were then expressed (first in solid culture and then in liquid culture) according to the protocol above.

Acoustic plate reader scans

The general protocol for preparing and scanning liquid cultures samples of GV-expressing *E. coli* in 96-well format is described in Fig. 3-S1 and the corresponding figure caption. Detailed instructions on how to build and use this system, as well as troubleshooting and bug-reporting information, are provided at https://github.com/shapiro-lab/acoustic-plate-reader.

The specific ultrasound pulse sequence parameters used for collecting the data shown in each figure are presented in Table 3.

Validation of best mutants

Selected mutants from each library were miniprepped and sequenced as described above. Unique mutants were then re-cloned using MoClo (see above) before undergoing validation testing to avoid the possibility that these plasmids accrued expression-reducing mutations during the GV expression steps performed during library screening. To prepare fragments for these MoClo assemblies, *gvpA*/*gvpB* mutant CDSs were PCRed using the primers described in Table 4 (which were selected based on the sequence of the mutant being amplified) and prepared for Golden Gate assembly as described above.

OD600 measurements

OD 600 culture measurements were performed on a Tecan Spark plate reader using the "Absorbance" protocol with the following settings: 600 nm measurement wavelength, 10 flashes, 50 ms settle time. Measurements were collected for 200 uL of culture and normalized to a 1 cm path length using the built-in "Pathlength Correction" feature.

Nondestructive multiplexed imaging of two ARG types

Ana-T6A-L40A and Mega-S9G-R31L-R85L ARGs were expressed in liquid culture as described above, with 0.25% glucose and 0.05% arabinose used to induce expression. Samples were loaded into phantoms and scanned in the Acoustic Plate Reader using the imaging parameters described in Table 3. See Supplementary Note 2 for the details of the spectral unmixing paradigm used to separate the two ARG types.

3.6 Author Contributions

R.C.H. and M.G.S. conceived and designed the study. R.C.H., Z.J., D.P.S., and D.R.M. designed and built the Acoustic Plate reader hardware. Z.J., D.P.S., and D.R.M. wrote the MATLAB scripts for data acquisition with the Acoustic Plate Reader. R.C.H., Z.J., and D.P.S. wrote the MATLAB scripts for data analysis from the Acoustic Plate Reader. M.S. designed the MATLAB graphical user interfaces for Acoustic Plate Reader data acquisition and analysis with direction from R.C.H. R.C.H., M.S., K.W., R.D., and H.S. performed directed evolution experiments. R.C.H. and M.S. performed the multiplexing experiments. R.C.H. analyzed all data. R.C.H. wrote the paper, with input from all authors. M.G.S. supervised the research.

3.7 Acknowledgements

The authors would like to thank Przemysław Dutka for creating the GvpA/GvpB structural diagrams in Fig. 3-3 and 3-4. This research was supported by the National Institutes of Health (R01-EB018975 to M.G.S.) and Pew Charitable Trust. R.C.H. was supported by the Caltech Center for Environmental Microbial Interactions. Related research in the Shapiro Laboratory is supported by the David and Lucille Packard Foundation, the Burroughs Wellcome Fund, the Heritage Medical Research Institute and the Chen Zuckerberg Initiative.

3.8 References

1. Bourdeau, R. W., Lee-Gosselin, A., Lakshmanan, A., Farhadi, A., Kumar, S. R., Nety, S. P., and Shapiro, M. G. (2018) Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts. *Nature 553*, 86–90.
2. Farhadi, A., Ho, G. H., Sawyer, D. P., Bourdeau, R. W., and Shapiro, M. G. (2019) Ultrasound imaging of gene expression in mammalian cells. *Science 365*, 1469–1475.
3. Dutka, P., Metskas, L. A., Hurt, R. C., Salahshoor, H., Wang, T.-Y., Malounda, D., Lu, G., Chou, T.-F., Shapiro, M. G., and Jensen, G. J. (2022) Structure of *Anabaena flos-aquae* gas vesicles revealed by cryo-ET. preprint, Molecular Biology.
4. Salahshoor, H., Yao, Y., Dutka, P., Nyström, N. N., Jin, Z., Min, E., Malounda, D., Jensen, G. J., Ortiz, M., and Shapiro, M. G. (2022) Geometric effects in gas vesicle buckling under ultrasound. *Biophys. J. 121*, 4221–4228.
5. Yao, Y., Jin, Z., Ling, B., Malounda, D., and Shapiro, M. G. (2021) Self-assembly of protein superstructures by physical interactions under cytoplasm-like conditions. *Biophys. J. 120*, 2701–2709.
6. Dutka, P., Malounda, D., Metskas, L. A., Chen, S., Hurt, R. C., Lu, G. J., Jensen, G. J., and Shapiro, M. G. (2021) Measuring gas vesicle dimensions by electron microscopy. *Protein Sci. 30*, 1081–1086.

7. Maley, A. M., Lu, G. J., Shapiro, M. G., and Corn, R. M. (2017) Characterizing Single Polymeric and Protein Nanoparticles with Surface Plasmon Resonance Imaging Measurements. *ACS Nano 11*, 7447–7456.
8. Cherin, E., Melis, J. M., Bourdeau, R. W., Yin, M., Kochmann, D. M., Foster, F. S., and Shapiro, M. G. (2017) Acoustic Behavior of Halobacterium salinarum Gas Vesicles in the High-Frequency Range: Experiments and Modeling. *Ultrasound Med. Biol. 43*, 1016–1030.
9. Zhang, S., Huang, A., Bar-Zion, A., Wang, J., Mena, O. V., Shapiro, M. G., and Friend, J. (2020) The Vibration Behavior of Sub-Micrometer Gas Vesicles in Response to Acoustic Excitation Determined via Laser Doppler Vibrometry. *Adv. Funct. Mater. 30*, 2000239.
10. Hurt, R. C., Buss, M. T., Duan, M., Wong, K., You, M. Y., Sawyer, D. P., Swift, M. B., Dutka, P., Barturen-Larrea, P., Mittelstein, D. R., Jin, Z., Abedi, M. H., Farhadi, A., Deshpande, R., and Shapiro, M. G. (2023) Genomically mined acoustic reporter genes for real-time in vivo monitoring of tumors and tumor-homing bacteria. *Nat. Biotechnol.*
11. Wu, D., Baresch, D., Cook, C., Ma, Z., Duan, M., Malounda, D., Maresca, D., Abundo, M. P., Lee, J., Shivaei, S., Mittelstein, D. R., Qiu, T., Fischer, P., and Shapiro, M. G. (2023) Biomolecular actuators for genetically selective acoustic manipulation of cells. *Sci. Adv. 9*, eadd9186.
12. Kim, W.-S., Min, S., Kim, S. K., Kang, S., Davis, H., Bar-Zion, A., Malounda, D., Kim, Y. H., An, S., Lee, J.-H., Bae, S. H., Lee, J. G., Kwak, M., Cho, S.-W., Shapiro, M. G., and Cheon, J. (2022) Magneto-acoustic protein nanostructures for non-invasive imaging of tissue mechanics *in vivo*. preprint, Biochemistry.
13. Bar-Zion, A., Nourmahnad, A., Mittelstein, D. R., Shivaei, S., Yoo, S., Buss, M. T., Hurt, R. C., Malounda, D., Abedi, M. H., Lee-Gosselin, A., Swift, M. B., Maresca, D., and Shapiro, M. G. (2021) Acoustically triggered mechanotherapy using genetically encoded gas vesicles. *Nat. Nanotechnol. 16*, 1403–1412.
14. Sawyer, D. P., Bar-Zion, A., Farhadi, A., Shivaei, S., Ling, B., Lee-Gosselin, A., and Shapiro, M. G. (2021) Ultrasensitive ultrasound imaging of gene expression with signal unmixing. *Nat. Methods 18*, 945–952.
15. Rabut, C., Wu, D., Ling, B., Jin, Z., Malounda, D., and Shapiro, M. G. (2021) Ultrafast amplitude modulation for molecular and hemodynamic ultrasound imaging. *Appl. Phys. Lett. 118*, 244102.
16. Ling, B., Lee, J., Maresca, D., Lee-Gosselin, A., Malounda, D., Swift, M. B., and Shapiro, M. G. (2020) Biomolecular Ultrasound Imaging of Phagolysosomal Function. *ACS Nano 14*, 12210–12221.
17. Maresca, D., Lakshmanan, A., Lee-Gosselin, A., Melis, J. M., Ni, Y.-L., Bourdeau, R. W., Kochmann, D. M., and Shapiro, M. G. (2017) Nonlinear ultrasound imaging of nanoscale acoustic biomolecules. *Appl. Phys. Lett. 110*, 073704.

18. Lakshmanan, A., Jin, Z., Nety, S. P., Sawyer, D. P., Lee-Gosselin, A., Malounda, D., Swift, M. B., Maresca, D., and Shapiro, M. G. (2020) Acoustic biosensors for ultrasound imaging of enzyme activity. *Nat. Chem. Biol. 16*, 988–996.
19. Farhadi, A., Bedrossian, M., Lee, J., Ho, G. H., Shapiro, M. G., and Nadeau, J. L. (2020) Genetically Encoded Phase Contrast Agents for Digital Holographic Microscopy. *Nano Lett. 20*, 8127–8134.
20. Lu, G. J., Farhadi, A., Szablowski, J. O., Lee-Gosselin, A., Barnes, S. R., Lakshmanan, A., Bourdeau, R. W., and Shapiro, M. G. (2018) Acoustically modulated magnetic resonance imaging of gas-filled protein nanostructures. *Nat. Mater. 17*, 456–463.
21. Rabut, C., Yoo, S., Hurt, R. C., Jin, Z., Li, H., Guo, H., Ling, B., and Shapiro, M. G. (2020) Ultrasound Technologies for Imaging and Modulating Neural Activity. *Neuron 108*, 93–110.
22. Lu, G. J., Farhadi, A., Mukherjee, A., and Shapiro, M. G. (2018) Proteins, air and water: reporter genes for ultrasound and magnetic resonance imaging. *Curr. Opin. Chem. Biol. 45*, 57–63.
23. Maresca, D., Lakshmanan, A., Abedi, M., Bar-Zion, A., Farhadi, A., Lu, G. J., Szablowski, J. O., Wu, D., Yoo, S., and Shapiro, M. G. (2018) Biomolecular Ultrasound and Sonogenetics. *Annu. Rev. Chem. Biomol. Eng. 9*, 229–252.
24. Yang, G., and Withers, S. G. (2009) Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem 10*, 2704–2715.
25. Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins. *Nat. Rev. Genet. 16*, 379–394.
26. Maresca, D., Sawyer, D. P., Renaud, G., Lee-Gosselin, A., and Shapiro, M. G. (2018) Nonlinear X-Wave Ultrasound Imaging of Acoustic Biomolecules. *Phys. Rev. X 8*, 041002.
27. Li, N., and Cannon, M. C. (1998) Gas vesicle genes identified in *Bacillus megaterium* and functional expression in Escherichia coli. *J. Bacteriol. 180*, 2450–2458.
28. Lakshmanan, A., Farhadi, A., Nety, S. P., Lee-Gosselin, A., Bourdeau, R. W., Maresca, D., and Shapiro, M. G. (2016) Molecular Engineering of Acoustic Protein Nanostructures. *ACS Nano 10*, 7314–7322.
29. Sadashivaiah, V., Tippani, M., Page, S. C., Kwon, S. H., Bach, S. V., Bharadwaj, R. A., Hyde, T. M., Kleinman, J. E., Jaffe, A. E., and Maynard, K. R. (2023) SUFI: an automated approach to spectral unmixing of fluorescent multiplex images captured in mouse and post-mortem human brain tissues. *BMC Neurosci. 24*, 6.
30. Alnahhas, R. N., Sadeghpour, M., Chen, Y., Frey, A. A., Ott, W., Josić, K., and Bennett, M. R. (2020) Majority sensing in synthetic microbial consortia. *Nat. Commun. 11*, 3659.
31. Hays, S. G., Patrick, W. G., Ziesack, M., Oxman, N., and Silver, P. A. (2015) Better together: engineering and application of microbial symbioses. *Curr. Opin. Biotechnol. 36*, 40–49.
32. Ziesack, M., Gibson, T., Oliver, J. K. W., Shumaker, A. M., Hsu, B. B., Riglar, D. T., Giessen, T. W., DiBenedetto, N. V., Bry, L., Way, J. C., Silver, P. A., and Gerber, G. K. (2019) Engineered

Interspecies Amino Acid Cross-Feeding Increases Population Evenness in a Synthetic Bacterial Consortium. *mSystems* (Collins, C. H., Ed.) *4*, e00352-19, /msystems/4/4/msys.00352-19.atom.

33. Moore, S. J., Lai, H.-E., Kelwick, R. J. R., Chee, S. M., Bell, D. J., Polizzi, K. M., and Freemont, P. S. (2016) EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology. *ACS Synth. Biol. 5*, 1059–1069.
34. Segall-Shapiro, T. H., Sontag, E. D., and Voigt, C. A. (2018) Engineered promoters enable constant gene expression at any copy number in bacteria. *Nat. Biotechnol. 36*, 352–358.
35. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., and Voigt, C. A. (2019) Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol. 15*, 196–204.
36. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods 10*, 354–360.
37. Chen, Y.-J., Liu, P., Nielsen, A. A. K., Brophy, J. A. N., Clancy, K., Peterson, T., and Voigt, C. A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods 10*, 659–664.
38. Lin, D., and O'Callaghan, C. A. (2018) MetClo: methylase-assisted hierarchical DNA assembly using a single type IIS restriction enzyme. *Nucleic Acids Res.*
39. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., and Liu, C. C. (2018) Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell 175*, 1946-1957.e13.
40. Wittmann, B. J., Johnston, K. E., Almhjell, P. J., and Arnold, F. H. (2022) evSeq: Cost-Effective Amplicon Sequencing of Every Variant in a Protein Library. *ACS Synth. Biol. 11*, 1313–1324.

3.9 Supporting Information

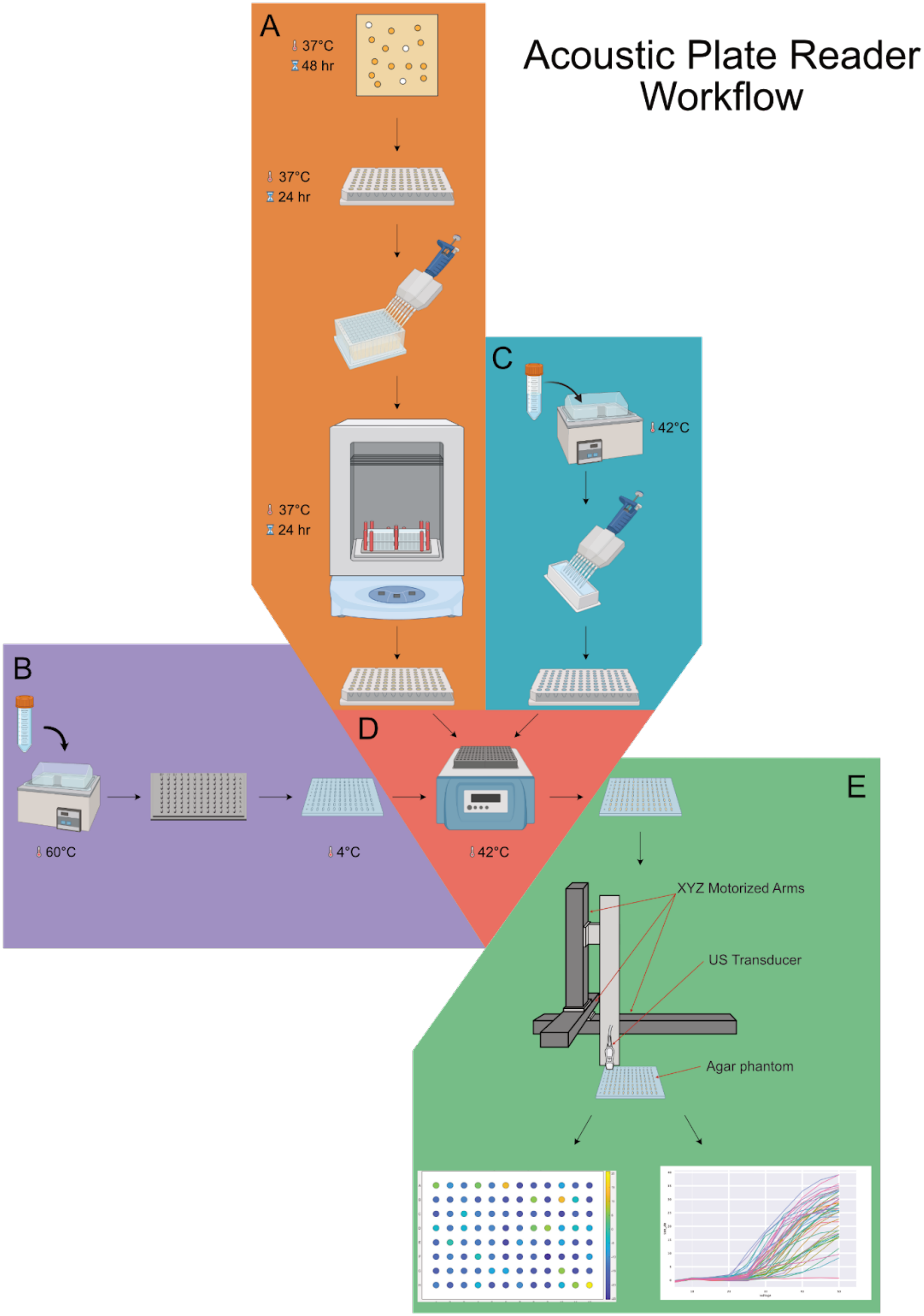


Figure 3-S1. Detailed diagram of the acoustic plate reader workflow. (A) GVs are expressed in *E. coli* as colonies on Petri dishes for 48 hr at 37°C, then colonies are picked into LB and grown to saturation in liquid culture for 24 hr at 37°C. These saturated liquid cultures are then diluted 1:100 into autoinduction LB and expressed for 24 hr at 37°C in 500 uL cultures in deep-well 96-well plates (square wells used for maximum culture aeration; USA Scientific 1896-2800). Aliquots of these cultures are aliquoted into an un-skirted 96-well PCR plate for subsequent loading into phantoms. (B) A solution of 2% Ultrapure Agarose (Invitrogen, 16500500) is prepared in 1X PBS and incubated at 60C for at least 12 hr to degas. Agarose phantoms are then made by pouring 75 mL of this solution into a 96-well phantom mold and incubating at 4C for 10 min. (C) A solution of 1% low-melting-temperature agarose (Goldbio, A-204-100) is prepared in 1X PBS and incubated at 60C for at least 12 hr to degas. This solution is then aliquoted into an un-skirted 96-well PCR plate to be used for phantom loading. (D) Phantoms from B are loaded by placing the 96-well PCR plates from A and C into 96-well heat block at 42C, and combining equal volumes of culture and agarose before pipetting into the empty phantom. (E) Phantoms from D are scanned using the acoustic plate reader, which generates US data for each sample and can image up to 12 96-well phantoms in a single scan.

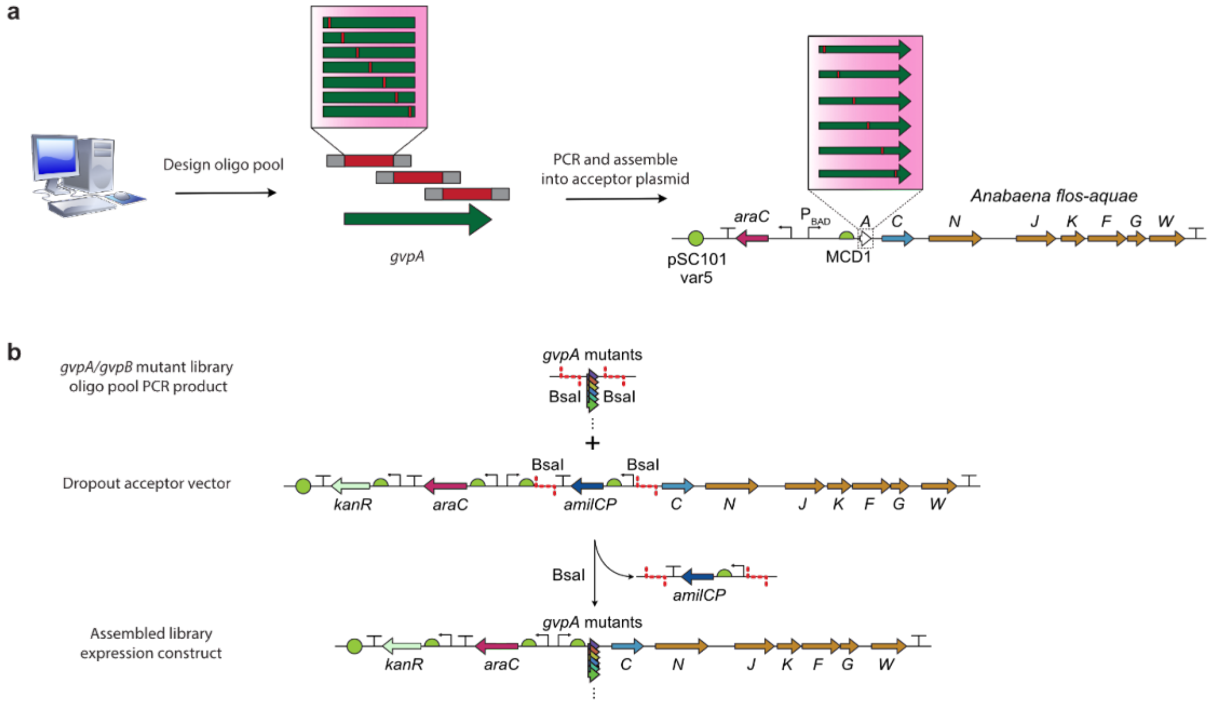


Figure 3-S2. Details of *gvpA*/*gvpB* mutant library construction and screening. (A) Overview of workflow for creating either scanning site saturation or recombination libraries. (B) Details of library assembly via a Golden Gate-based version of cassette mutagenesis (see Methods).

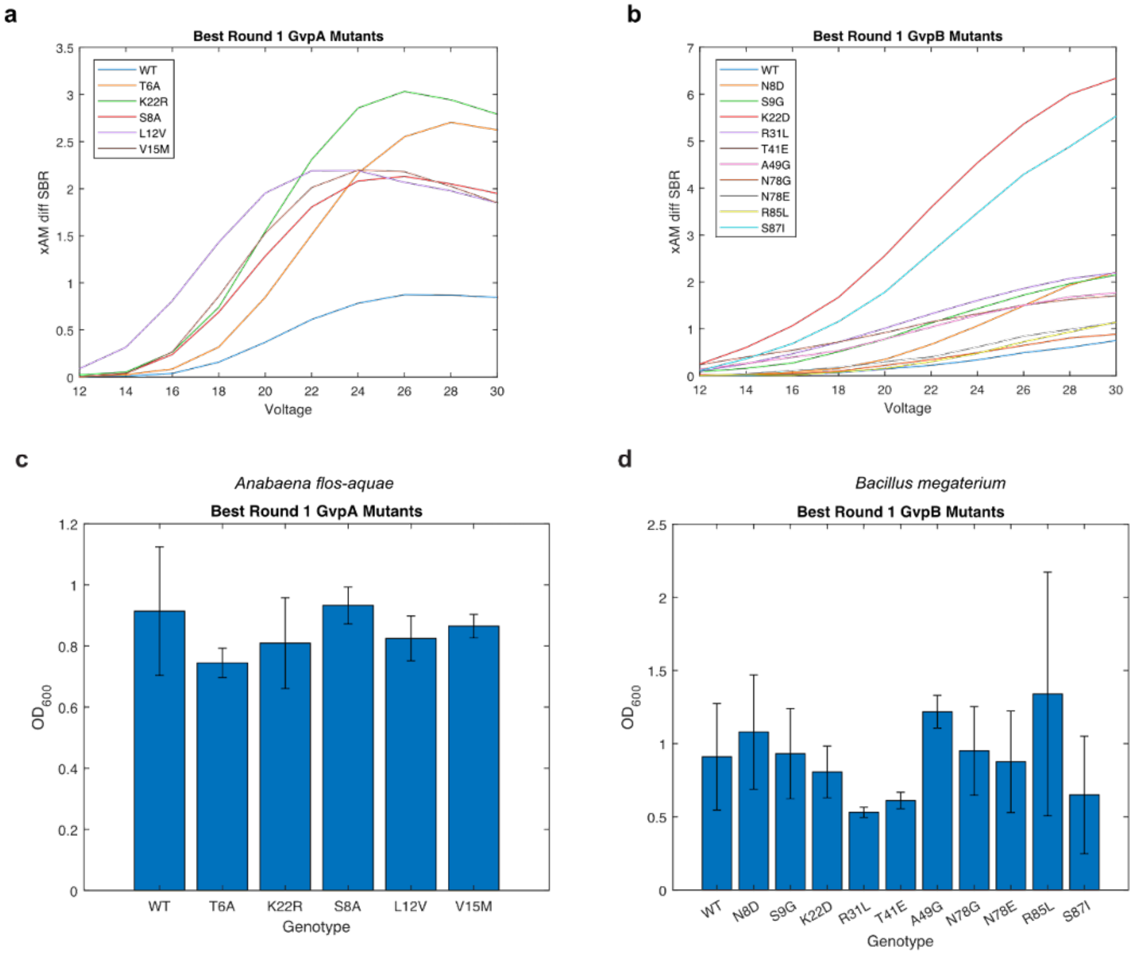


Figure 3-S3. Characterization of the top mutants from Round 1 of evolution. (A-B) xAM difference SBR as a function of pressure for each of the top mutants. (C-D) Collapse pressure curves for the mutants shown in A-B. (E-F) OD600 measurements for the mutants shown in A-B.

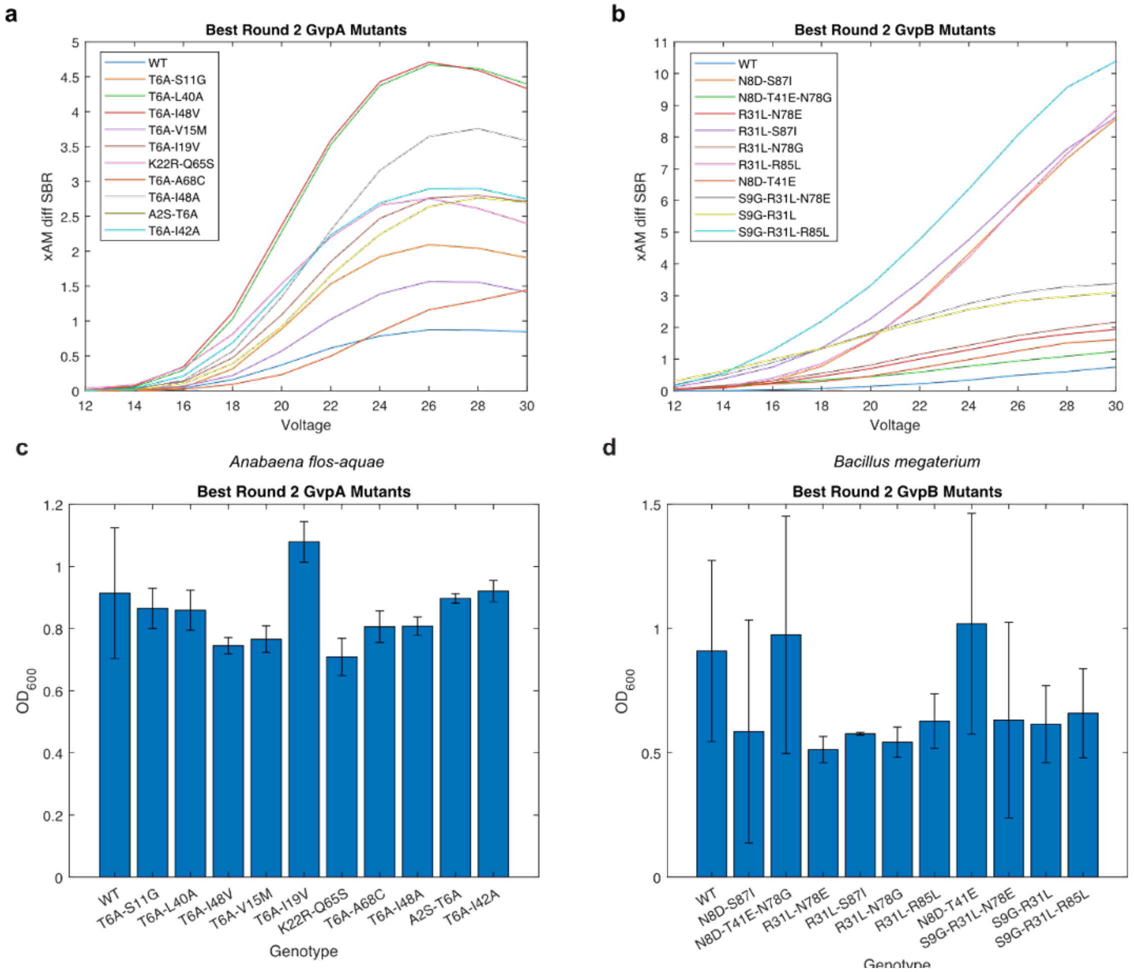


Figure 3-S4. Characterization of the top mutants from Round 2 of evolution. (A-B) xAM difference SBR as a function of pressure for each of the top mutants. (C-D) Collapse pressure curves for the mutants shown in A-B. (E-F) OD600 measurements for the mutants shown in A-B.

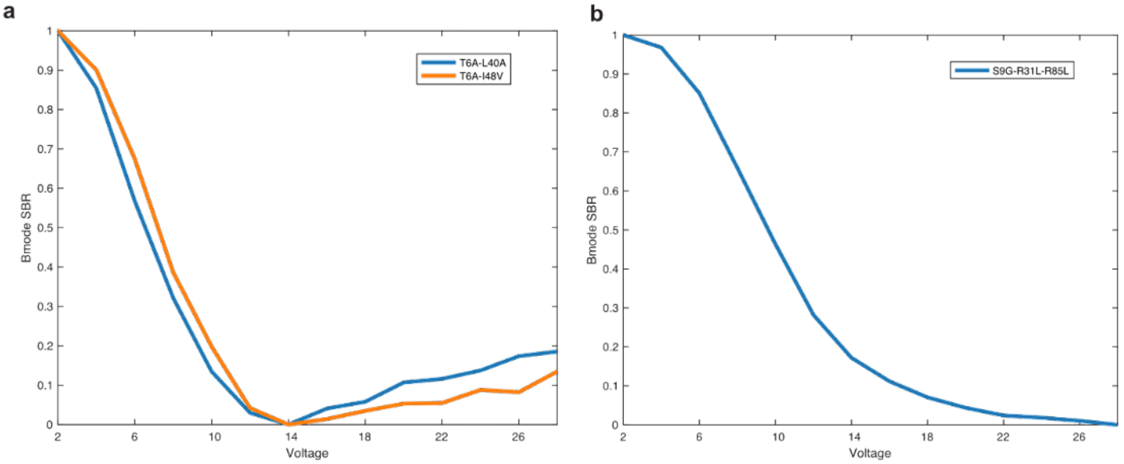


Figure 3-S5. Additional characterization of the best mutants identified in this study. (A-B) Bmode acoustic collapse pressure curves for the top-performing mutants identified in this study. We were not able to collect acoustic collapse pressure curves for the WT clusters because their Bmode signals were too low.

Supplementary Note 1: Golden Gate reactions
Master mix recipes

Component	Amount per reaction	Stock concentration (NEB)	For 66 reactions
T4 Ligase Buffer	1 uL	10X	66 uL
Hi-T4 DNA Ligase	500 U	400 U/uL	83 uL
BsmBI-v2	15 U	10 U/uL	100 uL
Water	to 5 uL		81 uL

Component	Amount per reaction	Stock concentration (NEB)	For 66 reactions
T4 Ligase Buffer	1 uL	10X	66 uL
Hi-T4 DNA Ligase	500 U	400 U/uL	83 uL
BsaI-HF-v2 or BbsI-HF	15 U	20 U/uL	50 uL
Water	to 5 uL		131 uL

To set up reactions, combine 75 ng of the backbone part with 150 ng of each insert part in a PCR tube with 5 uL of the appropriate master mix and fill to 10 uL with water. Miniprepped parts give higher assembly efficiencies than linear PCR products.

Golden Gate thermocycler protocol
20 min 37/42°C
3 minutes 37/42°C
4 minutes 16°C
Cycle 2-3 x30
10 minutes 50°C
10 minutes 80°C
Hold 4°C

C h a p t e r 4

GENOMICALLY MINED ACOUSTIC REPORTER GENES FOR REAL-TIME *IN VIVO* MONITORING OF TUMORS AND TUMOR-HOMING BACTERIA

Hurt R.C.*, Buss M.T.*, Duan M.*, Wong K., You M.Y., Sawyer D.P., Swift M.B., Dutka P., Barturen-Larrea P., Mittelstein D.R., Jin Z., Abedi M.H., Farhadi A., Deshpande R., Shapiro M.G. "Genomically mined acoustic reporter genes for real-time *in vivo* monitoring of tumors and tumor-homing bacteria." *Nature Biotechnology*. 2023. DOI: 10.1038/s41587-022-01581-y

This chapter is a reformatted version of the above publication. My contributions to the work were conceiving and performing the genomic mining portion of the study, analyzing data, and writing the paper together with my co-lead authors.

4.1 ABSTRACT

Ultrasound allows imaging at a much greater depth than optical methods, but existing genetically encoded acoustic reporters for *in vivo* cellular imaging have been limited by poor sensitivity, specificity, and *in vivo* expression. Here, we describe two acoustic reporter genes (ARGs)—one for use in bacteria and one for use in mammalian cells—identified through a phylogenetic screen of candidate gas vesicle gene clusters from diverse bacteria and archaea, that provide stronger ultrasound contrast, produce nonlinear signals distinguishable from background tissue, and have stable long-term expression. Compared to their first-generation counterparts, these second-generation bacterial and mammalian ARGs produce 83-fold and 38-fold stronger nonlinear contrast. Using the new ARGs, we noninvasively image *in situ* tumor colonization and gene expression in tumor-homing therapeutic bacteria, track the progression of tumor gene expression and growth in a mouse model of breast cancer, and perform gene expression-guided needle biopsies of a genetically mosaic tumor, demonstrating noninvasive access to dynamic biological processes at centimeter depth.

4.2 Introduction

Basic biological research, *in vivo* synthetic biology and the development of cell-based medicine require methods to visualize the functions of specific cells deep inside intact organisms. In this context, optical techniques based on fluorescent and luminescent proteins are limited by the scattering and absorption of light by tissue.[1] Ultrasound is a widely used technique for deep-tissue imaging, providing sub-100 μm spatial resolution and penetrating several cm into tissue.[2] Super-resolution methods[3,4] have pushed its spatial resolution below 10 μm. Recently, the first genetically encodable reporters for ultrasound[5–7] were introduced based on gas vesicles (GVs)—air-filled protein nanostructures encoded by clusters of 8-20+ genes, which evolved as flotation devices in diverse, mostly aquatic bacteria and archaea.[8,9] The low density and high compressibility of their air-filled interiors compared with surrounding tissues allow GVs to scatter sound waves and thereby produce ultrasound contrast when heterologously expressed as acoustic reporter genes (ARGs) in genetically engineered bacteria[6] or mammalian cells.[7]

Despite the promise of first-generation ARGs, their utility for monitoring bacterial or mammalian gene expression *in vivo* is limited. Bacterial ARGs[6] do not scatter ultrasound nonlinearly (making them difficult to distinguish from background tissues), express poorly at 37°C, and are too metabolically burdensome for *in situ* expression *in vivo*. Likewise, mammalian ARGs[7] produce only linear contrast, and cell-to-cell variability in their expression and burden has limited their use to clonally selected cell lines stimulated with potent epigenetic reagents. In both cases, the lack of nonlinear signal had to be circumvented by destructive ultrasound pulse sequences, which destroyed the GVs and limited dynamic imaging.[10]

We sought to make acoustic proteins widely useful in *in vivo* biological research and potential clinical applications by developing next-generation ARGs that, when expressed heterologously in either bacteria or mammalian cancer cell lines, could produce GVs with strong nonlinear ultrasound contrast and enable long-term expression under physiological conditions. We used a genomic mining approach—previously applied to improving fluorescent proteins[11–14], opsins[15–17] Cas proteins[18–22], and other biotechnology tools[23–28]—to identify ARGs with improved properties, which we further optimized through genetic engineering. By cloning and screening 15 distinct polycistronic operons chosen from a diverse set of 288 GV-expressing species representing a broad phylogeny, we identified two GV gene clusters – from *Serratia sp.* 39006 and *Anabaena flos-aquae*—that produce 83- or 38-fold more nonlinear acoustic contrast than previously tested clusters when expressed in several types of bacteria and mammalian cells, respectively. The bacterial ARG adapted from *Serratia* sp. 39006 (bARG_{Ser}), when expressed in the probiotic bacterium *E. coli* Nissle 1917 (EcN), enabled noninvasive ultrasound imaging of these bacteria colonizing tumors at depths greater than 1 cm, providing direct visualization of the microscale *in vivo* distribution of this potential anti-cancer therapy.[29–32] The mammalian ARG adapted from *Anabaena flos-aquae* (mARG_{Ana}), when expressed in human breast cancer cells, enabled both the noninvasive, *in situ* microscale imaging and long-term monitoring of heterologous gene expression in developing orthotopic tumors **at depths greater than 0.7 cm**, and the ultrasound-guided biopsy of a genetically defined subpopulation of these tumor cells. The properties and performance of these improved ARGs should facilitate a wide range of *in vivo* research.

4.3 RESULTS

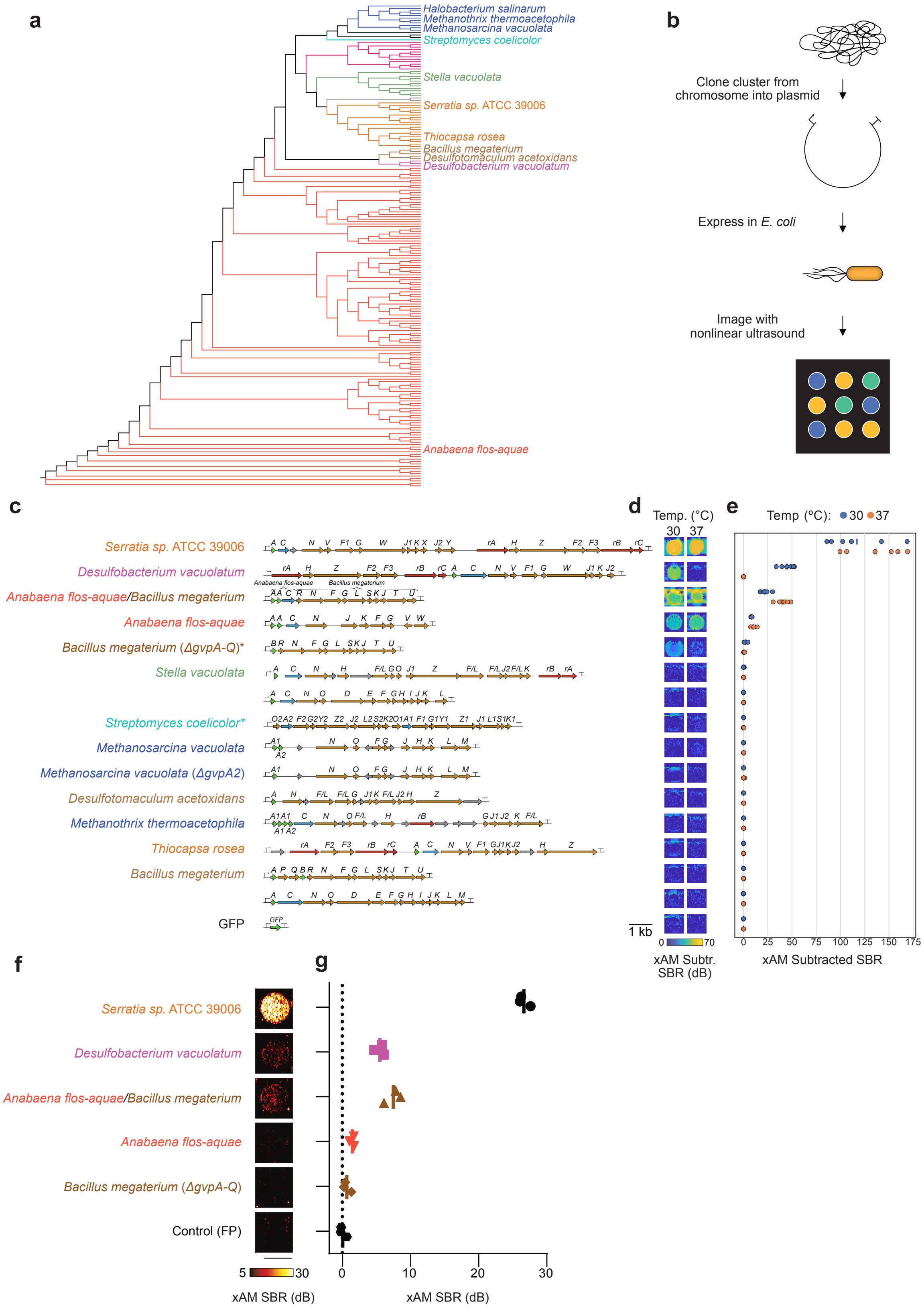

a
Halobacterium salinarum
Methanothrix thermoacetophila
Methanosarcina vacuolata
Streptomyces coelicolor
Stella vacuolata
Serratia sp. ATCC 39006
Thiocapsa rosea
Bacillus megaterium
Desulfotomaculum acetoxidans
Desulfobacterium vacuolatum
Anabaena flos-aquae
b
Clone cluster from chromosome into plasmid
Express in E. coli
Image with nonlinear ultrasound
c
Serratia sp. ATCC 39006
Desulfobacterium vacuolatum
Anabaena flos-aquae/Bacillus megaterium
Anabaena flos-aquae
Bacillus megaterium (ΔgvpA-Q)*
Stella vacuolata
Streptomyces coelicolor*
Methanosarcina vacuolata
Methanosarcina vacuolata (ΔgvpA2)
Desulfotomaculum acetoxidans
Methanothrix thermoacetophila
Thiocapsa rosea
Bacillus megaterium
GFP
Anabaena flos-aquae
Bacillus megaterium
1 kb
d
Temp. (°C)
30
37
0
70
xAM Subtr. SBR (dB)
e
Temp (°C): ● 30 ● 37
0
25
50
75
100
125
150
175
xAM Subtracted SBR
f
Serratia sp. ATCC 39006
Desulfobacterium vacuolatum
Anabaena flos-aquae/Bacillus megaterium
Anabaena flos-aquae
Bacillus megaterium (ΔgvpA-Q)
Control (FP)
5
30
xAM SBR (dB)
g
0
10
20
30
xAM SBR (dB)

Figure 4-1: Genomic mining of gas vesicle gene clusters reveals homologs with nonlinear ultrasound contrast in *E. coli*. (**a**) 16S phylogenic tree of known GV-producing organisms, with the species from which GV genes were cloned and tested in this study indicated by name. See **Fig. S1** for the fully annotated phylogenic tree. *B. megaterium* and *S. coelicolor* were not reported to produce GVs, but we tested their GV gene clusters here based on previous experiments in *E. coli*[3] and to broadly sample the phylogenetic space. (**b**) Workflow for testing GV clusters. Selected GV gene clusters were expressed in BL21(DE3) *E. coli* by growing patches of cells on plates containing the inducer IPTG, and the patches were then imaged with nonlinear ultrasound (xAM). (**c-e**) Diagrams of the GV gene clusters tested in *E. coli* (c), differential xAM images of representative patches (d), and quantification of the differential xAM signal-to-background ratio (SBR) of the patches (n=6 biological replicates) (e). (**f-g**) Representative xAM images (f) and quantification of the xAM SBR (n=3 biological replicates, each with 2 technical replicates; lines represent the mean) (g) for the top 5 GV-producing clusters expressed in *E. coli* at 30°C on solid media and normalized to 5 x 10^9 cells/mL in agarose phantoms, imaged at 1.74 MPa. See **Fig. 4-ED2a-b** for the ultrasound signal at varying acoustic pressures and **Fig. 4-ED2c** for the corresponding BURST data.

Genomically mined GV gene clusters with improved US contrast

GVs are encoded by polycistronic gene clusters comprising one or more copies of the primary structural gene *gvpA* and 7 to 20+ other genes encoding minor constituents, assembly factors or reinforcing proteins, which together help assemble the GVs' protein shells.[9] Starting with a list of organisms with confirmed GV production and sequenced operons (**Supplementary Table 1**), we cloned GV operons from 11 representative species providing a broad sampling of phylogenetic space, cluster architecture and organismal characteristics (*i.e.*, halophilic, thermophilic and mesophilic) (**Fig. 4-1a** and **Fig. 4-S1**).

We expressed each operon in confluent *E. coli* patches at several temperatures and inducer concentrations (**Fig. 4-1b**), comparing them to two bacterial ARG constructs previously shown to work in *E. coli*[6] – bARG1 (*Anabaena flos-aquae* (NCBI:txid315271)*/Bacillus megaterium* (NCBI:txid1404) hybrid) and *Bacillus megaterium ΔgvpA-Q*, as well as the full *Bacillus megaterium* gene cluster (**Fig. 4-1c-g**, **Fig. 4-ED1a-e**, **Fig. 4-S2a-c**, **Fig. 4-S3-5**). We scanned these patches using a home-built robotic ultrasound imaging apparatus, applying a cross-propagating amplitude modulation pulse sequence (xAM).[33] This pulse sequence enhances signals specific to nonlinear contrast agents such as GVs while cancelling linear background scattering. Unlike pulse sequences that rely on the irreversible collapse of GVs to obtain GV-specific contrast[6,7,10], xAM is nondestructive. In addition, we examined the optical opacity of the patches, which can be increased by sufficient levels of GV expression.

Of the 15 gene clusters tested, only 3 showed significant xAM signal when expressed at 37°C, and 5 showed significant xAM signal at 30°C (**Fig. 4-1, c-e**). Even though all operons tested are from organisms reported to produce GVs in their native hosts, only the *Anabaena flos-aquae*, *B. megaterium ΔgvpA-Q*, bARG1, *Desulfobacterium vacuolatum* (NCBI:txid1121400), and *Serratia sp.* 39006 (*Serratia*; NCBI:txid104623) clusters produced detectable GVs heterologously in *E. coli*. Several other operons produced a small amount of ultrasound contrast under certain conditions, which did not arise from GV expression but reflected a rough patch morphology likely due to cellular toxicity (**Fig. 4-S3d)**. The failure of most tested gene clusters to produce GVs in *E. coli* is not surprising given the complexity of polycistronic heterologous expression, which requires each component to fold and function properly in a new host with a potentially different cytoplasmic environment, growth temperature and turgor pressure.[34,35] In addition, it is possible that some genes included in the clusters act as cis-regulators,[9,34,36–38] limiting expression absent a specific trans input, or that some additional genes are required beyond the annotated operons.

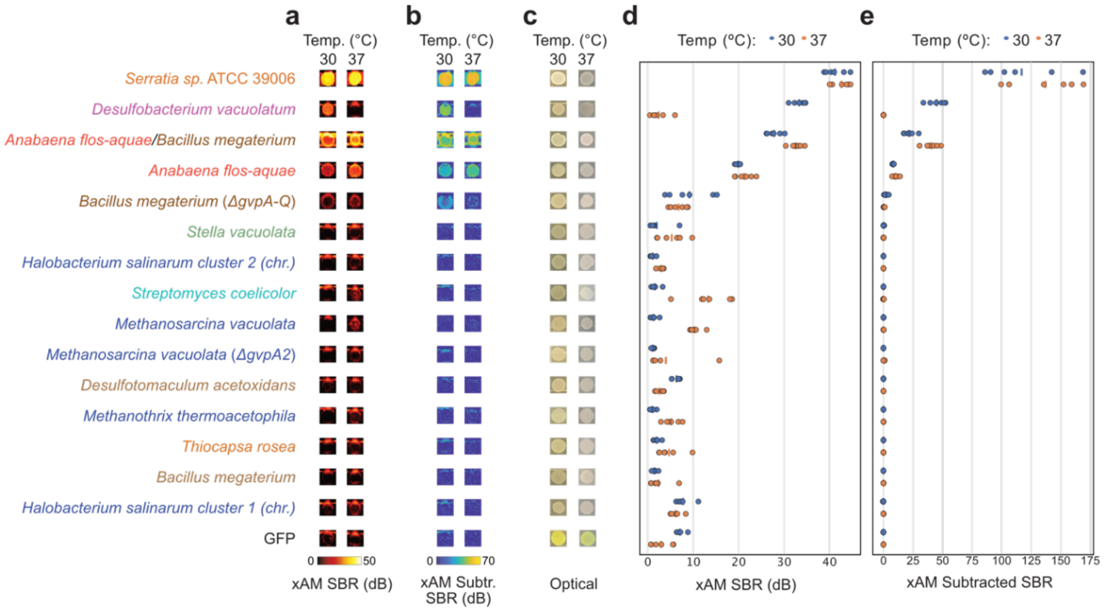


Figure 4-ED1: Additional images and quantification of *E. coli* patches expressing select GV gene clusters. (a-c) xAM images (a), pre-minus-post-collapse xAM images (b), and optical images showing opacity (c) of patches of *E. coli* expressing various GV gene clusters from the organisms listed on the left. (**d-e**) Quantification of the xAM (d) and pre-minus-post-collapse xAM (e) signals from images in (a-b) (n=6 biologically independent samples examined in one experiment). SBR, signal-to-background ratio.

In patch format, the strongest acoustic performance was observed with the genes from *Serratia*, bARG1, *Anabaena flos-aquae*, *B. megateriaum*, and *D. vacuolatum*. To compare these clusters while controlling for cell density, we resuspended cells expressing them in hydrogels at equal densities and imaged them using both xAM (**Fig. 4-1, f-g** and **Fig. 4-ED2, a-b**) and a more sensitive but destructive method called BURST[39] (**Fig. 4-ED2c**), and examined them optically with phase-contrast microscopy (PCM), which reveals the presence of GVs due to the refractive index difference between GVs and water[40,41] (**Fig. 4-ED2d**). Three of the clusters produced xAM signals, and all clusters produced BURST signals significantly stronger than the negative control. All clusters except *Anabaena flos-aquae* exhibited sufficient GV expression to be visible by PCM.

Cells expressing the *Serratia* cluster produced the strongest ultrasound signals, 19.2 dB above the next brightest cluster, bARG1, under xAM imaging at an applied acoustic pressure of 1.74 MPa: an 83-fold gain in signal intensity (**Fig. 4-1f**). Additionally, PCM and transmission electron microscopy

(TEM) images showed that cells expressing the *Serratia* cluster had the highest levels of GV expression (**Fig. 4-ED2, d-e**).

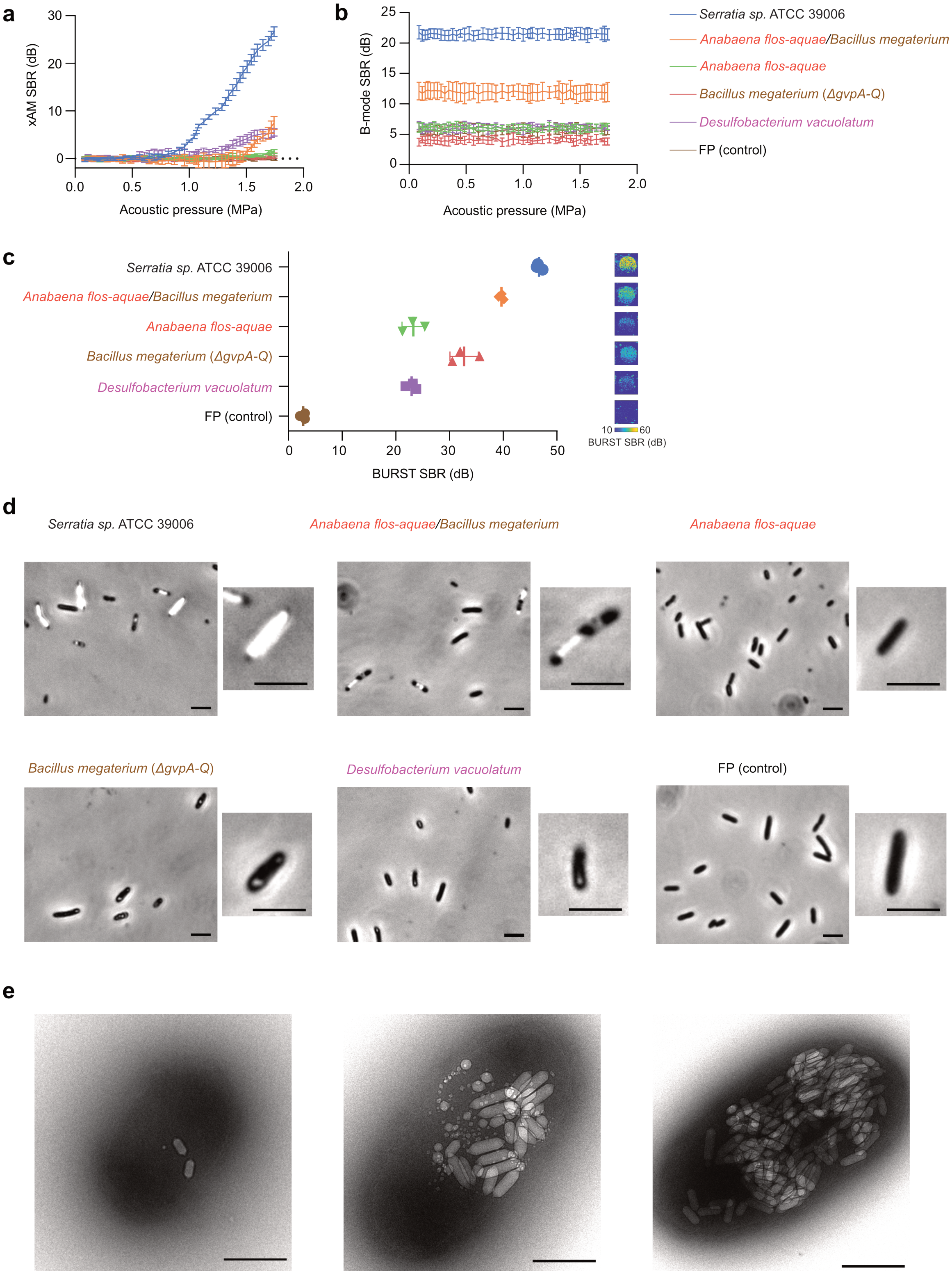


Figure 4-ED2: Characterization of working GV clusters in BL21(DE3) *E. coli*. (**a-d**) xAM signal-to-background ratio (SBR) as a function of acoustic pressure (a), B-mode SBR at a constant pressure of 0.15 MPa after each increase in acoustic pressure in a (b), BURST SBRs and corresponding representative images (c), and representative phase contrast microscopy (PCM) images (from at least n=3 biological replicates) (d) of the working GV clusters expressed in BL21(DE3) *E. coli* at 30ºC on solid media. For ultrasound imaging (a-c), samples were normalized to 5 x 10^9 cells/mL in agarose phantoms. Curves and error bars represent the mean (n=3 biological replicates each with 2 technical replicates) ± SD. (a-b) have the same legend. GV clusters in cells are visible by PCM for all clusters except for the cluster from *Anabaena flos-aquae* and the fluorescent protein (FP) control (d). (**e**) Representative TEM images of BL21(DE3) *E. coli* cells expressing Serratia GVs at varying levels of expression (from n=17 images from a single sample). Scale bars are 5 µm in (d) and 500 nm in (e).

Because overexpression of any protein imposes a metabolic demand on the host cell,[42–44] we reasoned that deletion of non-essential genes could improve GV expression from the *Serratia* cluster, and therefore the xAM signal. Previous work showed that deletions of *gvpC, gvpW, gvpX, gvpY, gvpH*, or *gvpZ* preserve GV formation in the native organism.[34] We tested these deletions, as well as the deletion of an unannotated hypothetical protein (Ser39006_001280) (**Fig. 4-S6a**). When expressed in *E. coli*, deletions of *gvpC, gvpH, gvpW, gvpY,* or *gvpZ* reduced or eliminated xAM signal (**Fig. 4-S6, b-c**) and patch opacity (**Fig. 4-S6d**). Deletion of *gvpX* increased xAM signal but decreased opacity, and deletion of Ser39006_001280 increased both xAM signal and opacity. Based on these results, we selected the *Serratia* ΔSer39006_001280 operon for subsequent *in vitro* and *in vivo* experiments. We call this genetic construct bARG_{Ser} – a bacterial acoustic reporter gene derived from *Serratia*.

bARG_{Ser} shows robust performance in E. coli Nissle

We transferred bARG_{Ser} into EcN, a strain of *E. coli* that is widely used in *in vivo* biotechnology applications due to its ability to persist in the gastrointestinal (GI) tract and colonize tumors[45–47] and deliver anti-tumor therapy.[29–32] We tested three different inducible promoter architectures and found that the L-arabinose-inducible pBAD promoter provided the most robust control over GV expression without obvious burden at 37ºC (**Fig. 4-2, a-b, Fig. 4-ED3**). To enable the the pBAD-bARG_{Ser} plasmid to be maintained without antibiotic selection, as required in certain *in vivo* applications, we added the toxin-antitoxin stability cassette Axe-Txe.[48] This allowed the plasmid to be maintained in EcN for up to 5 days of daily sub-culturing in liquid media without antibiotics, both with and without ARG induction (**Fig. 2c**).

The expression of most heterologous genes, including widely used fluorescent proteins (FPs), results in some degree of metabolic burden on engineered cells.[43,44,49] Consistent with this expectation, the induction of pBAD-bARG_{Ser} EcN resulted in reduced colony formation to an extent similar to the expression of a FP (**Fig. 4-ED4, a,c**), culture optical density remained relatively unchanged (**Fig. 4-ED4, b,d**). The growth curves of induced bARG_{Ser}-expressing and FP-expressing EcN were indistinguishable during the log phase (0 to 5 hours), indicating that the two strains have similar growth rates (**Fig. 4-ED4e**). Collectively, these results suggest that pBAD-driven expression of bARG_{Ser} in

EcN is not more burdensome than that of FPs, which are widely accepted as relatively non-perturbative indicators of cellular function.

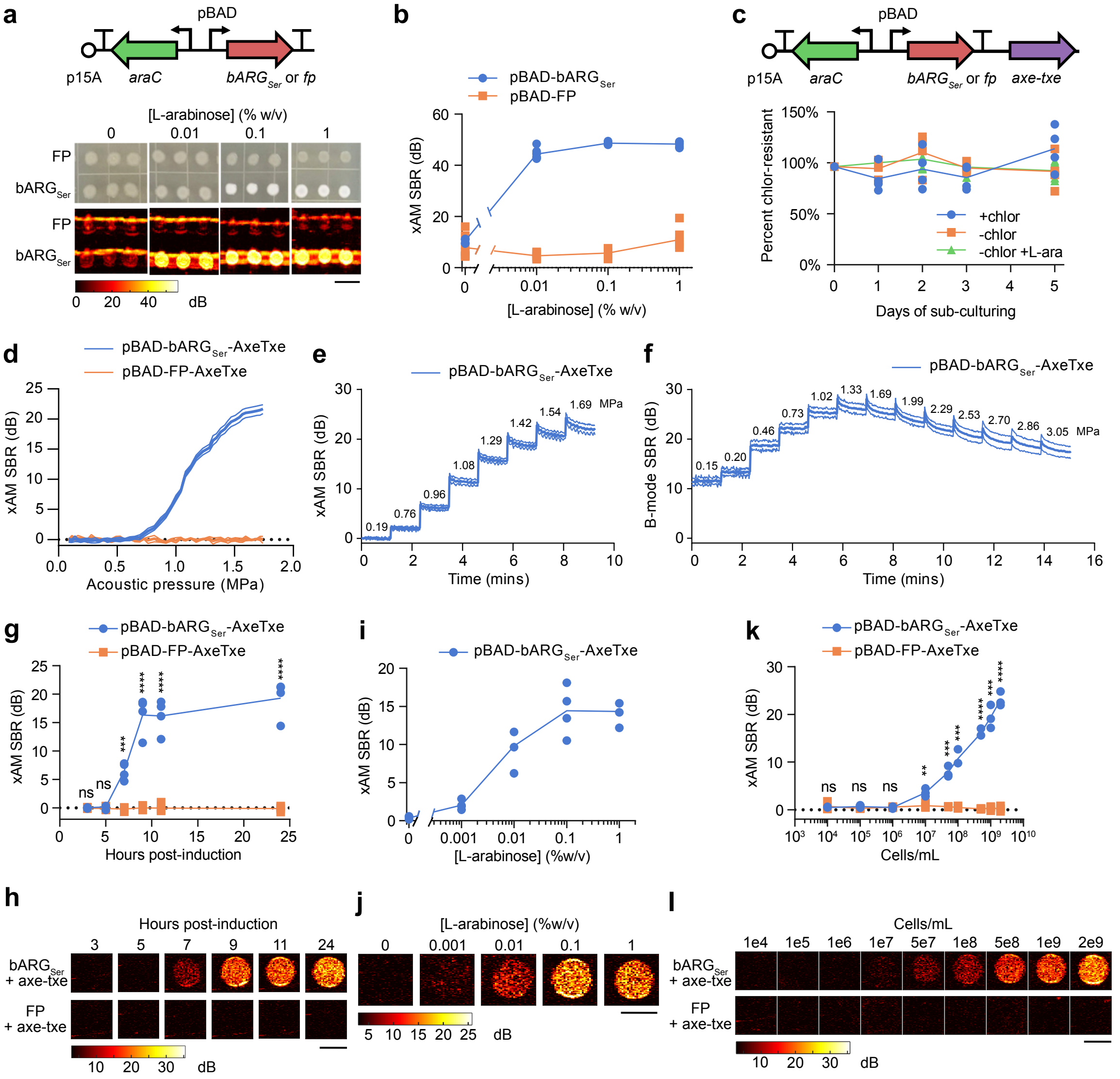


Figure 4-2: Expression of bARG$_{\text{Ser}}$ in *E. coli* Nissle (EcN) and acoustic characterization *in vitro*. (**a**) Diagram of the arabinose-inducible construct pBAD-bARG$_{\text{Ser}}$ used to express bARG$_{\text{Ser}}$ in EcN (top), and optical and xAM images of bARG$_{\text{Ser}}$-expressing or FP-expressing patches of EcN on solid media with varying L-arabinose concentrations (bottom). Scale bar is 1 cm. See **Fig. 4-ED3** for corresponding results with IPTG-inducible and aTc-inducible constructs. (**b**) Quantification of xAM SBR of patches from (a) versus the L-arabinose concentration (n=8). (**c**) Diagram of the construct from (a) with Axe-Txe[48] added, creating pBAD-bARG$_{\text{Ser}}$-AxeTxe, to enable plasmid maintenance in the absence of antibiotics (top), and verification of plasmid maintenance (bottom). Conditions were with chloramphenicol (+chlor), without chloramphenicol (-chlor), or without chloramphenicol and with 0.1% L-arabinose (-chlor +L-ara) using pBAD-bARG$_{\text{Ser}}$-AxeTxe EcN (n=4). (**d**) xAM SBR as a function of transmitted acoustic pressure. (**e-f**) xAM (e) and parabolic B-mode (f) SBRs measured over time when the transmitted acoustic pressure was increased every ~70 sec and the pulse repetition rate was 86.8 Hz. For (d-f), cells were induced with 0.1% L-arabinose for 24 hours. Bold lines represent the mean and thin lines represent ± standard deviation; n=3 biological replicates, each with 2 technical replicates. (**g-h**) xAM ultrasound SBR (g) and corresponding representative images (p values: 0.519494, 0.240386, 0.000120555, 6.818737e-5, 3.683585e-5, 2.325819e-5) (h) at several time points after inducing with 0.1% L-arabinose. (**i-j**) xAM SBR (i) and corresponding representative images (j) after inducing with varying L-arabinose concentrations

for 24 hours. (**k-l**) xAM SBR (p values: 0.699456, 0.0568424, 0.597418, 0.00906739, 0.00046697, 0.000456979, 4.937128e-6, 0.000183889, 1.708183e-5) (k) and corresponding representative images (l) of varying concentrations of cells induced for 24 hours with 0.1% L-arabinose in liquid culture. For (h, j, l) scale bars are 2 mm. For (d-j), cells were grown in liquid culture and normalized to 10^9 cells/mL for ultrasound imaging. For (g, i, k), each point is a biological replicate (n=4 for g and i; n=3 for k) that is the average of at least 2 technical replicates. Curves represent the mean for b-k. Asterisks represent statistical significance by two-tailed, unpaired Student's t-tests (**** = p<0.0001, *** = p<0.001, ** = p<0.01, ns = no significance).

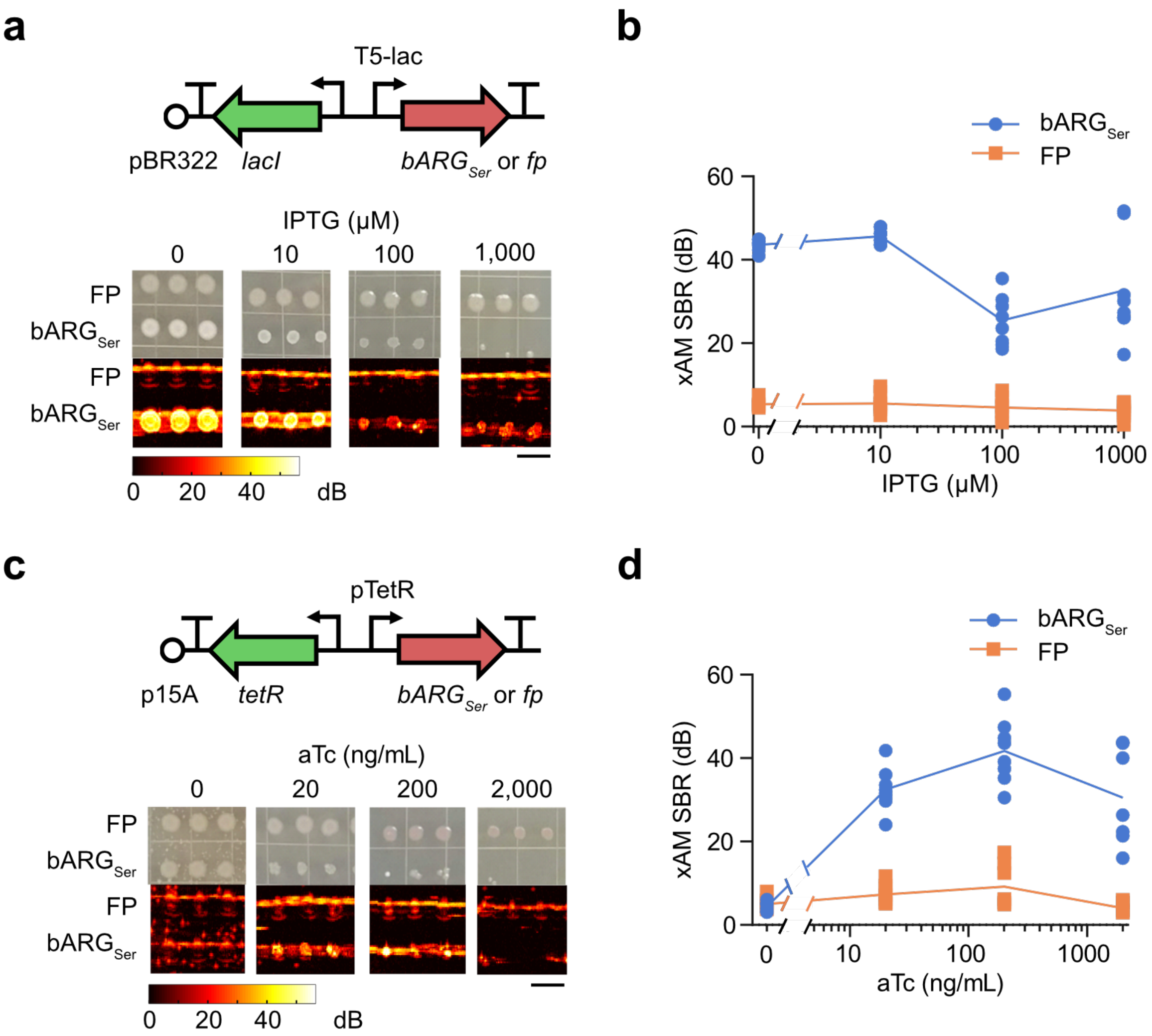


Figure 4-ED3: Testing bARG$_{\text{Ser}}$ expression in EcN with IPTG- and aTc-inducible gene circuits. (**a**) Diagram of the IPTG-inducible construct used to express bARG$_{\text{Ser}}$ in EcN (top), and representative optical and xAM images of bARG$_{\text{Ser}}$-expressing or fluorescent protein (FP)-expressing patches of EcN on solid media with varying IPTG concentrations at 37^{o}C (bottom). (**b**) Quantification of the xAM SBR of all patches from the experiment in (a) (n=8 biologically independent samples examined in one experiment; curves represent the mean). (**c-d**) Same as in (a-b) but for the aTc-inducible construct. The scale bars in (a,c) represent 1 cm.

To further examine the genetic stability of bARG$_{\text{Ser}}$ constructs, we plated cells from daily sub-cultures onto agar with 0.1% (w/v) L-arabinose and examined colony opacity (**Fig. 4-ED4f**) as a measure of retained GV expression. Of a total of 3824 colonies, nearly all were opaque (**Fig. 4-ED4f**), with GV expression confirmed by PCM and TEM (**Fig. 4-ED4, g-h**). Only 11 colonies (<0.3% after ~35 cell generations) exhibited a mutated phenotype with reduced opacity and GV production (**Fig. 4-S7, a-**

b). These results indicate that mutational inactivation of GV production is not a major issue for pBAD-bARG$_{Ser}$-AxeTxe EcN under typical conditions.

After establishing construct stability, we characterized the acoustic properties of bARG$_{Ser}$-expressing EcN. For cells induced in liquid culture for 24 hours and suspended at 10^9 cells/mL in agarose phantoms, an xAM signal was detected at acoustic pressures above 0.72 MPa, rising with increasing pressure up to the tested maximum of 1.74 MPa (**Fig. 4-2d**). The signal remained steady over time at pressures up to 0.96 MPa, above which we observed a slow decrease that followed an exponential decay (**Fig. 4-2e**, **Fig. 4-S8**, **Supplementary Note 2**), indicating that some of the GVs gradually collapsed despite sustained high xAM signals. We also imaged the cells with parabolic pulses, which can transmit higher pressures than xAM, and thus can be helpful *in vivo* to compensate for attenuation at tissue interfaces. When imaged with parabolic B-mode, the GVs started to collapse slowly at 1.02 MPa and more rapidly at 1.33 MPa and above (**Fig. 4-2f**). Based on these results, we chose an acoustic pressure of 1.29 MPa for xAM and 1.02 MPa for parabolic AM (pAM) in subsequent experiments to obtain the strongest signals while minimizing GV collapse.

Next, to characterize the sensitivity and dynamics of ultrasound contrast in bARG$_{Ser}$-expressing EcN, we measured xAM signal as a function of induction time, inducer concentration, and cell concentration. At a density of 10^9 cells/mL, xAM signal could first be observed 7 hours after induction with 0.1% L-arabinose and leveled off by 9 hours post-induction (**Fig. 4-2, g-h**). Keeping the induction time constant at 24 hours while varying the L-arabinose concentration, GV expression was detected with as little as 0.001% L-arabinose, and the highest ultrasound signal was observed for 0.1-1% L-arabinose (**Fig. 4-2, i-j**). When cells induced for 24 hours with 0.1% L-arabinose were diluted, they were detectable by ultrasound down to 10^7 cells/mL (**Fig. 4-2, k-l**). Critically, this detection was achieved non-destructively with nonlinear imaging, compared to previous bacterial ARGs, which required a destructive linear imaging approach.[6] The bARG$_{Ser}$ xAM signal was proportional to the cell concentration between 10^7 cells/mL and 2 x 10^9 cells/mL (**Fig. 4-2, k-l**). We also imaged the cells using BURST, which provides greater sensitivity at the cost of collapsing the GVs.[10] BURST imaging detected bARG$_{Ser}$ expression as early as 3 hours post-induction (**Fig. 4-ED5, a-b**), with as little as 0.001% L-arabinose (**Fig. 4-ED5, c-d**), and at a density as low as 10^5 cells/mL (**Fig. 4-ED5, e-f**).

Taken together, our *in vitro* experiments indicated that the reporter gene construct pBAD-bARG$_{Ser}$-AxeTxe is robust and stable in EcN and enables gene expression in these cells to be imaged with high contrast and sensitivity. Similar results were obtained in an attenuated strain of *Salmonella enterica* serovar Typhimurium (**Fig. 4-ED6**), another species used in bacterial anti-tumor therapies.[50,51]

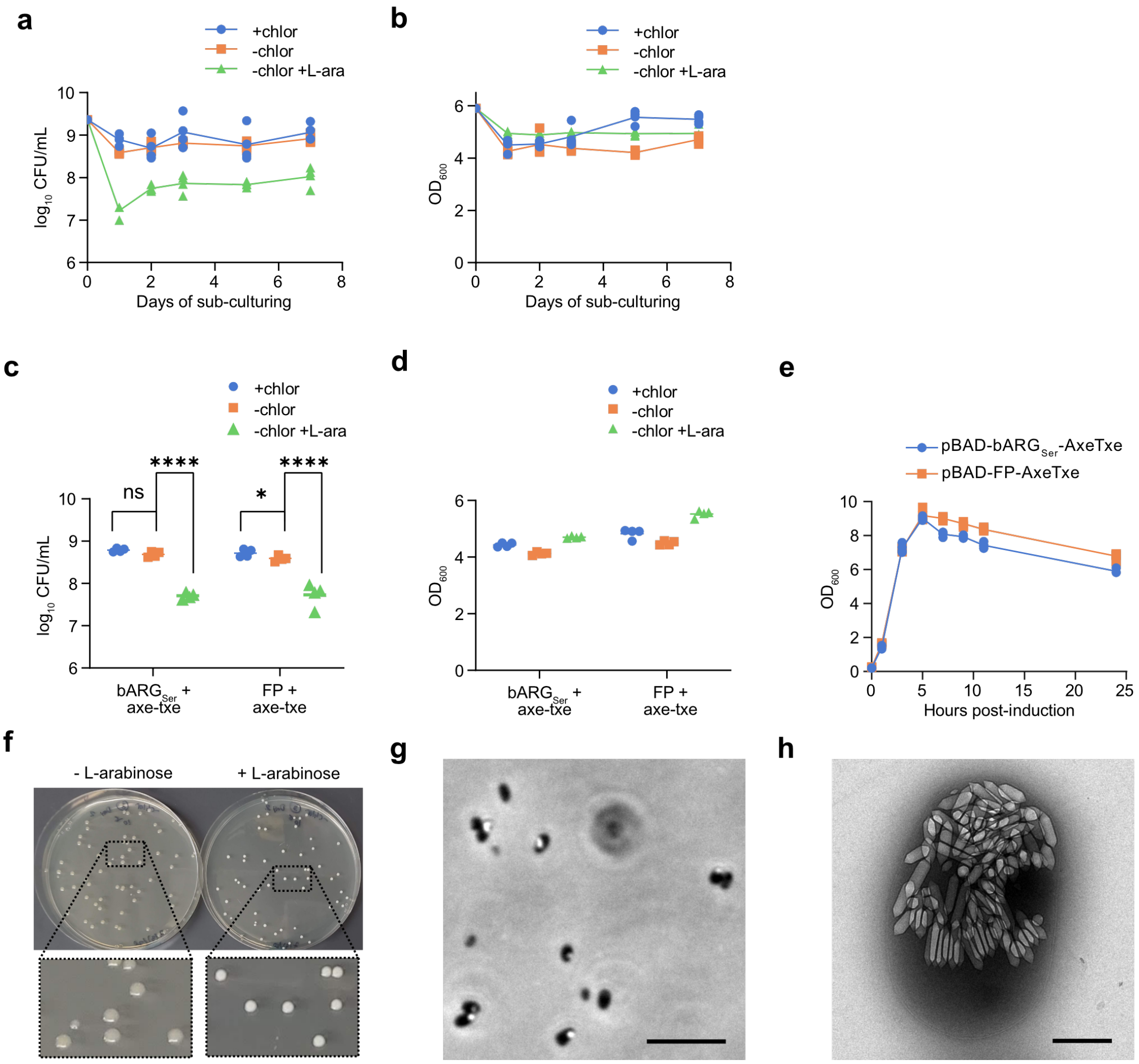


Figure 4-ED4: **Effect of induction on viability and OD_{600} for bARG$_{Ser}$-expressing EcN in liquid culture**. (**a-b**) Colony forming units (CFU) per mL of culture (a) and optical density at 600 nm (b) during daily sub-culturing into LB media with 25 µg/mL chloramphenicol (+chlor), without chloramphenicol (-chlor), or without chloramphenicol and with 0.1% (w/v) L-arabinose (-chlor +L-ara) using pBAD-bARG$_{Ser}$-AxeTxe EcN. Curves represent the mean of n=4 biological replicates. (**c-d**) Colony forming units (CFUs) per mL of culture on chloramphenicol plates (c) and optical density at 600 nm (p values from left to right: 0.0638556, 1.019913e-6, 0.0217493, 3.800448e-6) (d) of pBAD-bARG$_{Ser}$-AxeTxe EcN and pBAD-FP-AxeTxe EcN cultures 24 hours after sub-culturing into LB media with the same conditions as in (a-b) and in Fig. **4**-2c. Asterisks indicate statistical significance by two-tailed, unpaired Student's t-tests (**** = $p < 0.0001$, * = $p < 0.05$, ns = no significance); n=4 biological replicates. (**e**) OD_{600} versus time after inducing pBAD-bARG$_{Ser}$-AxeTxe and pBAD-FP-AxeTxe EcN strains with 0.1% (w/v) L-arabinose in liquid culture at 37°C (n=4 biological replicates). Between 5 and 24 hours post-induction, when the OD_{600} of all cultures decreased, the OD_{600} of FP-expressing cultures was slightly higher than that of the bARG$_{Ser}$-expressing cultures, likely due to expression of red fluorescent protein which is known to absorb light at 600 nm.[52] (**f**) Representative image of colonies from the experiment in **Fig. 4-2c** on chloramphenicol plates with (right) and without (left) 0.1% L-arabinose. The opacity of the colonies on plates with L-arabinose indicates bARG$_{Ser}$ expression and was used to screen for mutants deficient in bARG$_{Ser}$ expression (see **Fig. 4-S7**). (**g-h**) Representative phase contrast microscopy (from n=4 biological replicates) (g) and transmission electron microscopy (from n=23 images taken of a single sample) (h) images of pBAD-bARG$_{Ser}$-AxeTxe EcN cells grown on plates with 0.1% (w/v) L-arabinose at 37°C. Scale bars are 10 µm (g) and 500 nm (h). Curves and lines represent the mean for a-e.

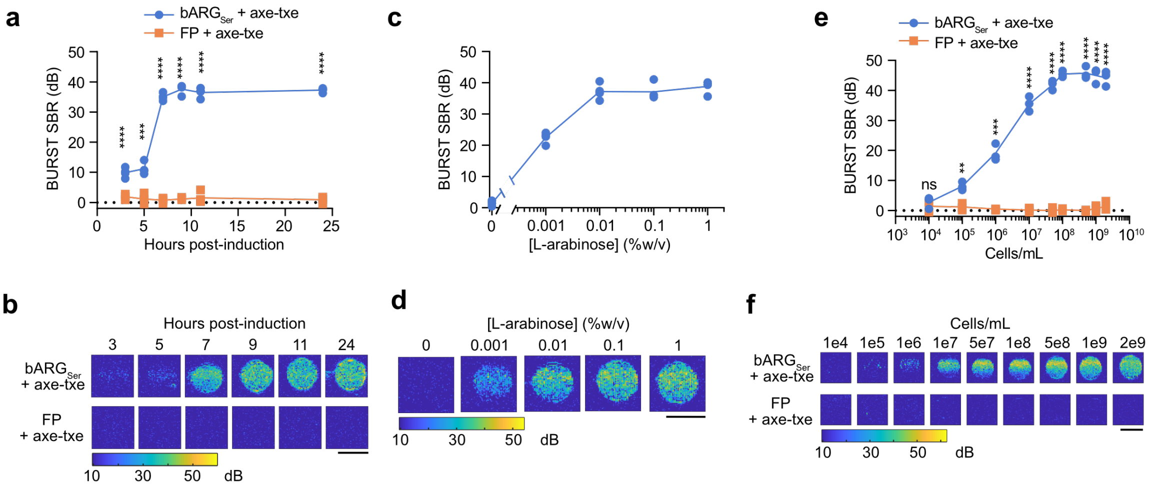


Figure 4-ED5: ***In vitro*** **characterizations of bARG$_{Ser}$-expressing EcN with BURST ultrasound imaging.** (**a-b**) BURST ultrasound signal-to-background ratio (SBR) versus time after inducing pBAD-bARG$_{Ser}$-AxeTxe and pBAD-FP-AxeTxe EcN strains with 0.1% L-arabinose in liquid culture at 37ºC (p values from left to right: 9.866231E-05, 0.00019939, 5.648535E-09, 9.858233E-09, 1.069657E-07, 1.161991E-09) (a) and the corresponding representative BURST images (b). (**c-d**) BURST ultrasound SBR versus L-arabinose concentration used to induce pBAD-bARG$_{Ser}$-AxeTxe EcN in liquid culture at 37ºC for 24 hours (c) and the corresponding representative BURST images. (**e-f**) BURST ultrasound SBR versus concentration of pBAD-bARG$_{Ser}$-AxeTxe or pBAD-FP-AxeTxe EcN cells induced for 24 hours at 37ºC with 0.1% L-arabinose in liquid culture (p values from left to right: 0.420535, 0.00261501, 0.000335619, 1.899074E-05, 2.348183E-06, 2.055919E-07, 2.793374E-06, 9.045661E-06, 1.487459E-05) (e) and the corresponding representative BURST images (f). Note that the BURST SBR saturated at 7 hours post-induction, 0.01% (w/v) L-arabinose, and 10^8 cells/mL. All scale bars are 2 mm. For (a-d), cells were normalized to 10^9 cells/mL in agarose phantoms for ultrasound imaging. For (a, c, e), each point is a biological replicate (n=4 for a and c; n=3 for e) that is the average of at least 2 technical replicates and curves indicate the mean. Asterisks represent statistical significance by two-tailed, unpaired Student's t-tests (**** = p<0.0001, *** = p<0.001, ** = p<0.01, ns = no significance).

bARG$_{Ser}$ enables in situ imaging of tumor-colonizing bacteria

To test the ability of bARG$_{Ser}$ to image the colonization and microscale distribution of bacteria inside tumors, we formed subcutaneous MC26 tumors in mice and intravenously injected EcN cells containing the pBAD-bARG$_{Ser}$-AxeTxe plasmid. Three days after injecting the bacteria, we induced GV expression and imaged the tumors with ultrasound, visualizing the tissue to a depth of 0.6 cm (**Fig. 4-3a**). In all tumors colonized by bARG$_{Ser}$-expressing EcN, we observed pAM, BURST, and xAM contrast one day after induction with L-arabinose (**Fig. 4-3b** and **Fig. 4-ED7a**). The signals were localized to the core of the tumor and concentrated at the interface between live and necrotic tissue, where the EcN primarily colonized, as confirmed with subsequent tissue histology (**Fig. 4-3f-l** and **Fig. 4-S9**). This biodistribution reflects the immune-privileged environment of the necrotic tumor core.[46,53]

After applying 3 MPa of acoustic pressure throughout the tumor to collapse the GVs, we re-injected the mice with L-arabinose, and found that similar ultrasound signals could be re-expressed in all tumors colonized by bARG$_{Ser}$-expressing EcN (**Fig. 4-3c** and **Fig. 4-ED7b**), showing that bARG$_{Ser}$ can be used to visualize dynamic gene expression at multiple timepoints. Absent L-arabinose induction, no xAM or pAM ultrasound signals were observed from bARG$_{Ser}$-containing EcN (**Fig. 4-3d** and **Fig. 4-ED7c**); likewise, no xAM or pAM ultrasound signals were seen in tumors

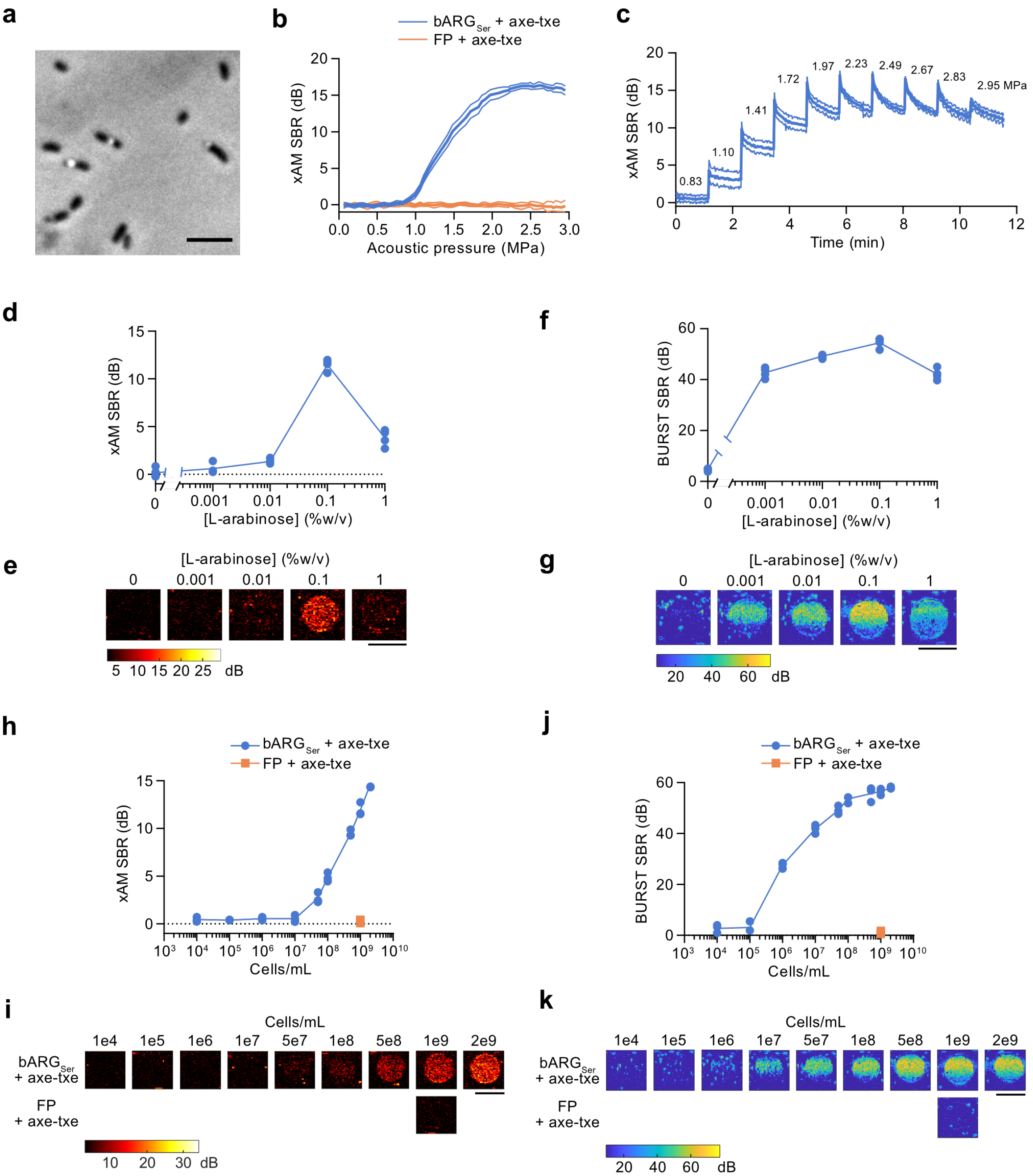


Figure 4-ED6: bARG$_{Ser}$ expression and acoustic characterization in *Salmonella enterica* serovar Typhimurium. (**a**) Representative phase contrast microscopy image of bARG$_{Ser}$-expressing *S. Typhimurium* cells (n=4 biological replicates). (**b**) xAM SBR as a function of transmitted acoustic pressure for bARG$_{Ser}$-expressing and FP-expressing *S. Typhimurium* cells. (**c**) xAM SBRs measured over time when the transmitted acoustic pressure was increased approximately every 70 sec as indicated by the numbers above the curve for bARG$_{Ser}$-expressing *S. Typhimurium*. Ultrasound was applied at a pulse repetition rate of 86.8 Hz. For (a-c), pBAD-bARG$_{Ser}$-AxeTxe *S. Typhimurium* cells were induced with 0.1% (w/v) L-arabinose for 24 hours at 37°C in liquid culture, and were then diluted to 10^9 cells/mL in agarose phantoms for ultrasound imaging. Bold lines represent the mean and thin lines represent ± standard deviation; n=3 biological replicates, each with 2 technical replicates. (**d-g**) xAM ultrasound SBR (d) and corresponding representative ultrasound images (e), and BURST

SBR (f) and corresponding representative images (g), using varying L-arabinose concentrations to induce pBAD-bARG$_{Ser}$-AxeTxe *S. Typhimurium* in liquid culture at 37°C for 24 hours. (**h-k**) xAM ultrasound SBR (h) and corresponding representative ultrasound images (i), and BURST SBR (j) and corresponding representative images (k), of varying concentrations of pBAD-bARG$_{Ser}$-AxeTxe or pBAD-FP-AxeTxe *S. Typhimurium* cells induced for 24 hours at 37°C with 0.1% (w/v) L-arabinose in liquid culture. Scale bars represent 5 µm in (a) and 2 mm in (e, g, i, k). For (d-g), cells were diluted to 10^9 cells/mL in agarose phantoms for ultrasound imaging. For (d, f, h, j), each point is a biological replicate (n=4 for d and f; n=3 for h and j) that is the average of at least 2 technical replicates, and curves represent the mean. All ultrasound imaging in this figure was performed with an L22-14vX transducer.

colonized by FP-expressing EcN (**Fig. 4-3e,j** and **Fig. 4-ED7, d-e**). Low levels of BURST signal were observed in uninduced animals (**Fig. 4-3k**), likely due to small amounts of L-arabinose present in the diet combined with BURST imaging's high sensitivity.

To quantify tumor colonization, on day 20 of the experiment we plated tumor homogenates on selective media. Tumors from all groups of mice contained more than 7 x 10^8 CFU/g tissue (**Fig. 4-3l**), indicating that the EcN can persist in tumors at high levels for at least 6 days after IV injection regardless of bARG$_{Ser}$ expression, collapse, and re-expression. The somewhat higher density of FP-expressing EcN suggests that maintenance of the smaller pBAD-FP-AxeTxe plasmid (7.2 kb versus 23.2 kb for pBAD-bARG$_{Ser}$-AxeTxe), which would impose less burden on the cell,[54–57] may be easier in this *in vivo* context where oxygen and other nutrients are limited. Negligible mutational silencing was observed in EcN plated from tumor samples (**Fig. 4-S10, a-c**).

Taken together, our *in vivo* experiments with EcN demonstrate that bARG$_{Ser}$ expression enables stable, non-destructive acoustic visualization of the microscale distribution of these probiotic agents in a therapeutically relevant context.

GV gene cluster with nonlinear contrast in mammalian cells

Next, we developed second-generation ARGs for mammalian cells. The first-generation mammalian ARGs were based on the GV gene cluster from *B. megaterium* (referred to here as mARG$_{Mega}$). mARG$_{Mega}$ expression could only be detected with destructive collapse-based imaging due to a low level of GV expression and the lack of nonlinear contrast.[7] Moreover, successful use of mARG$_{Mega}$ as a reporter gene required monoclonal selection of transduced cells and their treatment with a broadly-acting histone deacetylase inhibitor[7]. Seeking mARGs that are expressed more robustly and produce nonlinear contrast, we cloned mammalian versions of the genes contained in each of the three clusters that produced nonlinear signal in *E. coli* at 37°C: *Serratia*, *Anabaena flos-aquae*, and *Anabaena flos- aquae/B. megaterium* (**Fig. 4-1d-e**). Equimolar transient cotransfections of the monocistronic genes derived from each gene cluster into HEK293T cells yielded detectable BURST signal only for the *Anabaena flos-aquae* gene cluster and mARG$_{Mega}$[7] (**Fig. 4-4, a-b,** "1-fold excess"). Given the multiple *gvpA* copies contained in the native *Anabaena flos-aquae* GV operon,[8] we hypothesized that expressing GvpA at a higher stoichiometry relative to the other genes in this cluster

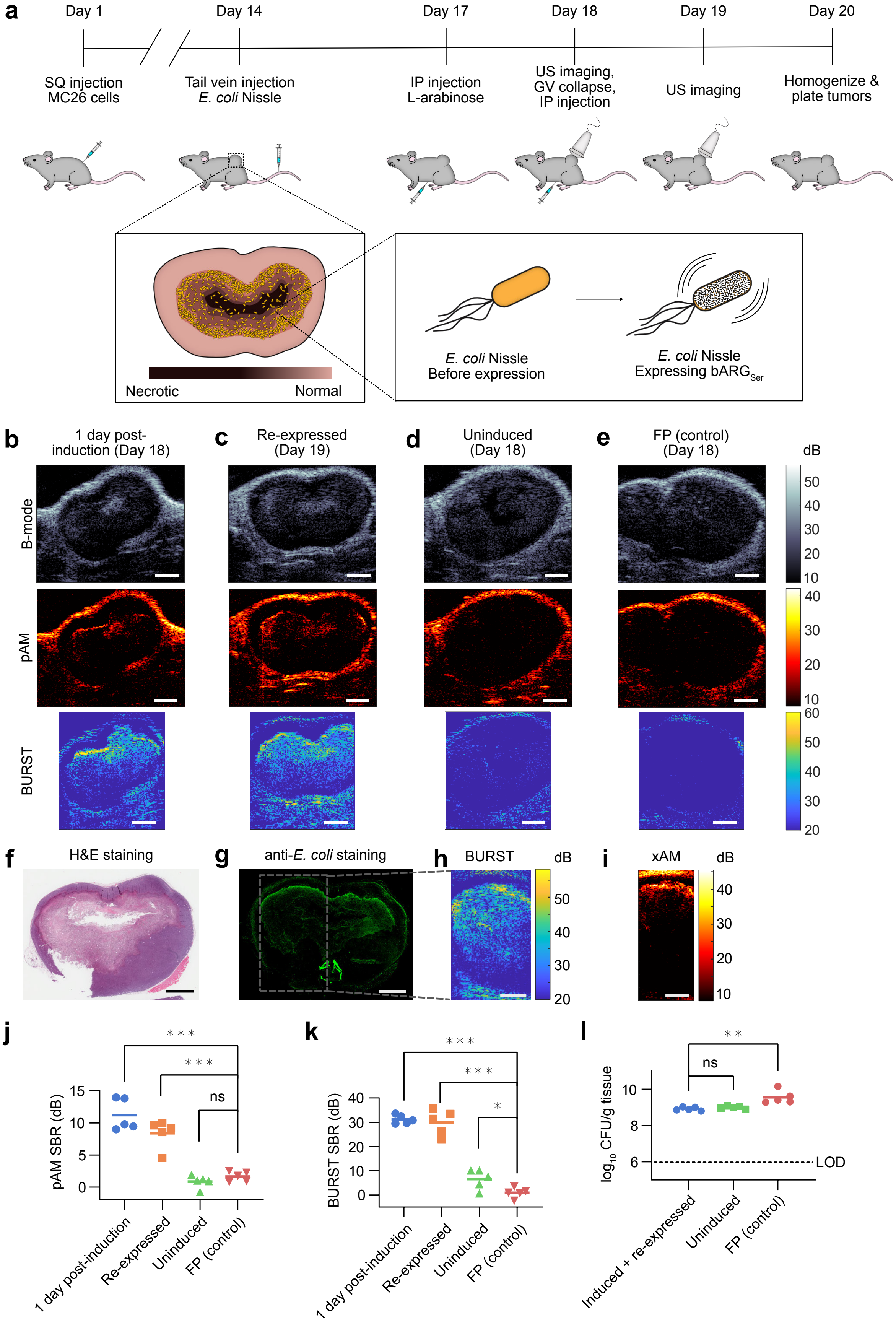

a
Day 1
Day 14
Day 17
Day 18
Day 19
Day 20
SQ injection
MC26 cells
Tail vein injection
E. coli Nissle
IP injection
L-arabinose
US imaging,
GV collapse,
IP injection
US imaging
Homogenize &
plate tumors
Necrotic
Normal
E. coli Nissle
Before expression
E. coli Nissle
Expressing bARG$_{Ser}$
b
1 day post-
induction (Day 18)
c
Re-expressed
(Day 19)
d
Uninduced
(Day 18)
e
FP (control)
(Day 18)
dB
B-mode
pAM
BURST
50
40
30
20
10
40
30
20
10
60
50
40
30
20
f
H&E staining
g
anti-*E. coli* staining
h
BURST
dB
50
40
30
20
i
xAM
dB
40
30
20
10
j
pAM SBR (dB)
15
10
5
0
***
***
ns
1 day post-induction
Re-expressed
Uninduced
FP (control)
k
BURST SBR (dB)
40
30
20
10
0
***
***
*
1 day post-induction
Re-expressed
Uninduced
FP (control)
l
$\log_{10}$ CFU/g tissue
10
8
6
**
ns
LOD
Induced + re-expressed
Uninduced
FP (control)

Figure 4-3: *In situ* bARG$_{Ser}$ expression enables ultrasound imaging of tumor colonization by EcN. (**a**) Diagram of the *in vivo* protocol for assessing *in situ* bARG$_{Ser}$ expression in tumors. Mice bearing subcutaneous (SQ) tumors were injected with EcN via the tail vein on day 14. bARG$_{Ser}$ or FP expression was then induced by injecting L-arabinose intraperitoneally (IP) on day 17; the next day, tumors were imaged with ultrasound. To check for re-expression, all the GVs in the tumors were collapsed, L-arabinose was re-injected, and tumors were imaged again the next day. On day 20, tumors were homogenized and plated to quantify the levels of EcN colonization. In separate experiments for histology, mice were sacrificed directly after imaging. (**b-d**) Representative B-mode, parabolic AM (pAM), and BURST ultrasound images of tumors colonized by pBAD-bARG$_{Ser}$-AxeTxe EcN 1 day after induction with L-arabinose on day 18 (b), at 1 day after collapse and re-induction (day 19) (c), or uninduced on day 18 (d). (**e**) Representative ultrasound images of tumors colonized by pBAD-FP-AxeTxe EcN 1 day after induction with L-arabinose on day 18. (**f-g**) Optical images of tissue sections stained with H&E (f) or anti-*E. coli* antibodies (g) from a tumor colonized by pBAD-bARG$_{Ser}$-AxeTxe EcN after imaging on day 19. (**h-i**) BURST (h) and xAM (i) images of the same tumor as in (f-g), with the boxed region showing the approximate BURST imaging region in the tissue section. Scale bars in (b-i) represent 2 mm. (**j-k**) Quantification of pAM (p values: 3.503E-05, 0.000209802, 0.19479281) (j) and BURST (p values: 8.19948E-09, 2.98285E-06, 0.023377359) (k) SBRs for the conditions in (b-e). (**l**) Colony forming units (CFUs) per gram of tumor tissue on day 20 (p values: 0.1911, 0.0055). For j-l, points represent each mouse (n=5), and lines represent the mean of each group. The dotted line indicates the limit of detection (LOD). Asterisks represent statistical significance by two-tailed, unpaired Student's t-tests (*** = p<0.001, ** = p<0.01, * = p<0.05, ns = no significance). See **Fig. 4-ED5** for representative xAM ultrasound images for the conditions in (b-d), and **Fig. 4-ED6** for more histological images of tissue sections.

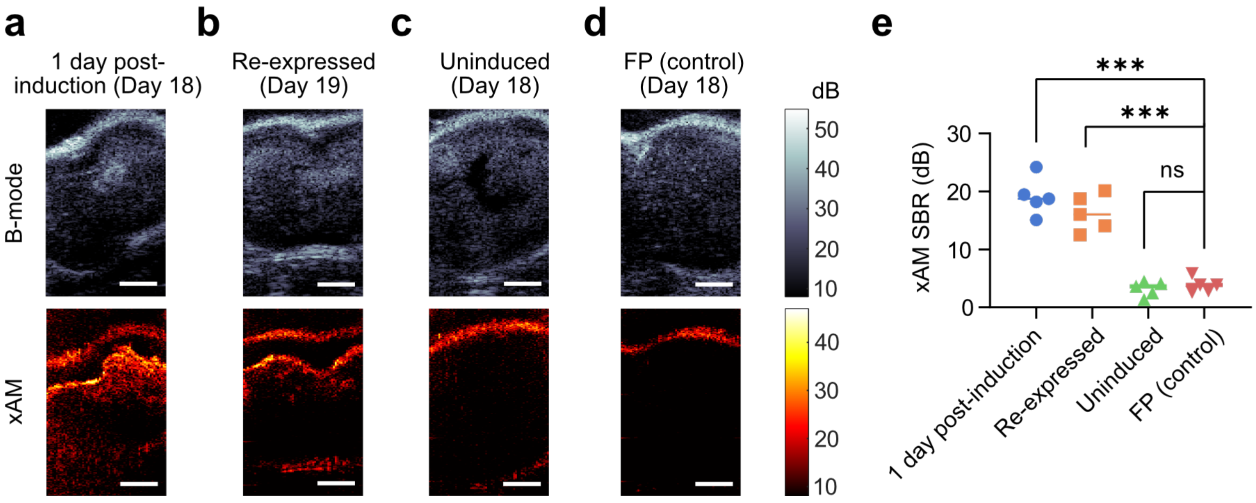


Figure 4-ED7: xAM ultrasound imaging of mouse tumors colonized by EcN. (**a-c**) Representative B-mode (top, grayscale) and xAM (bottom, hot-scale) ultrasound images of tumors colonized by pBAD-bARG$_{Ser}$-AxeTxe EcN at least 24 hours after induction with L-arabinose on day 18 (a), at least 24 hours after collapse and re-induction (day 19) (b), or uninduced on day 18 (c). (**d**) Representative ultrasound images of tumors colonized by pBAD-FP-AxeTxe EcN at least 24 hours after induction with L-arabinose on day 18. Scale bars in (a-d) represent 2 mm. (**e**) Quantification of the xAM SBR for the same conditions in (a-d) (p values from left to right: 1.0545E-05, 3.73902E-05, 0.487197948). Each group is n=5 mice and lines indicate the mean. Asterisks represent statistical significance by two-tailed, unpaired Student's t-tests (*** = p<0.001, ns = no significance). See **Fig. 4-3a** for the corresponding *in vivo* protocol.

could improve GV expression. To test this possibility, we titrated the amount of *gvpA* plasmid in the *Anabaena flos-aquae* plasmid set while keeping the DNA amount corresponding to other genes constant (the total DNA level was kept constant with a padding vector). We found that the BURST signal increased monotonically with increasing *gvpA* up to 8-fold *gvpA* excess (**Fig. 4-4b**). In contrast, the signal peaked at 2-fold excess of the homologous protein *gvpB* for the *B.megaterium* cluster. This suggests that the assembly factors in *the Anabaena flos-aquae* gene cluster may be more efficient

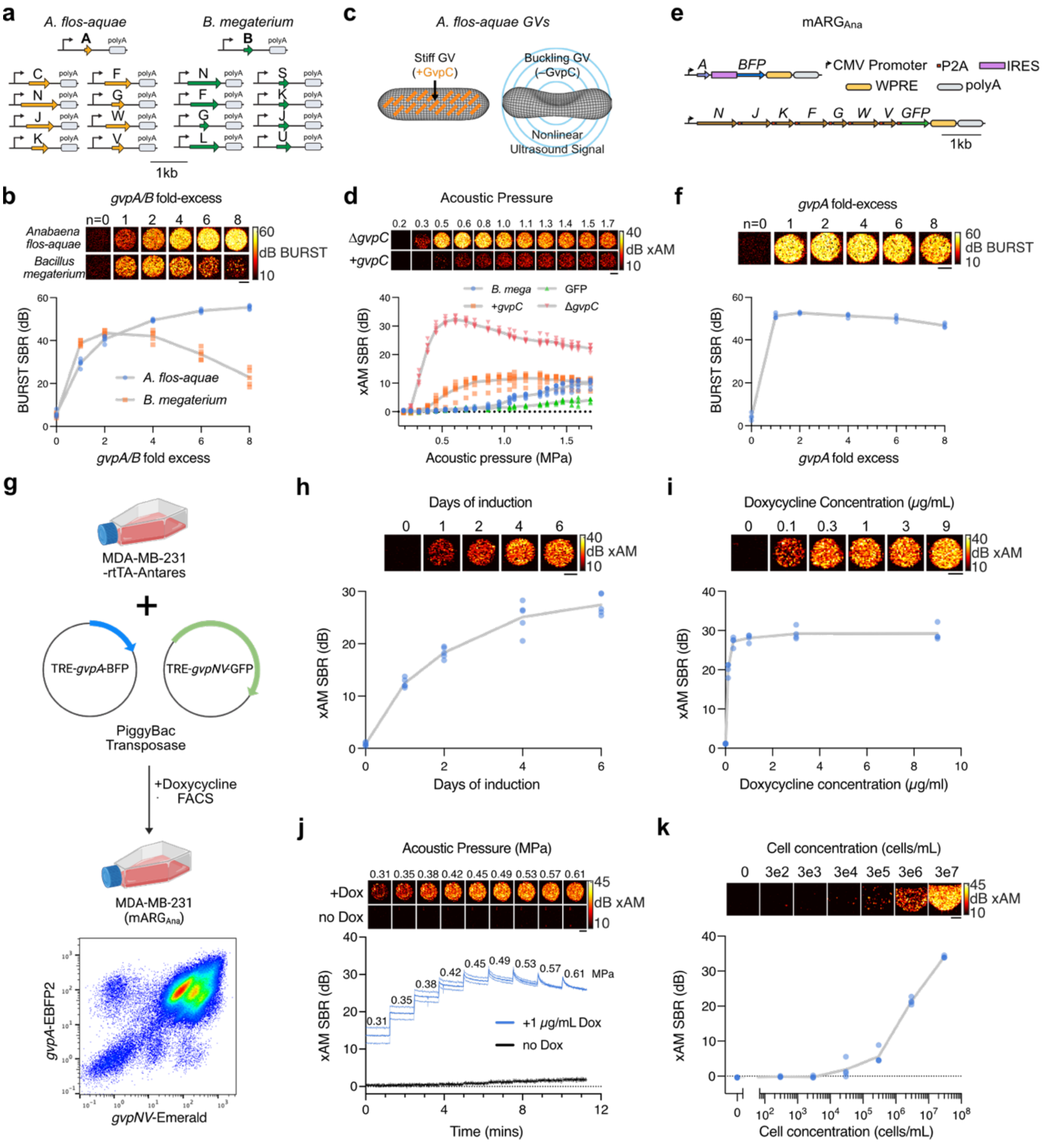


Figure 4-4: Heterologous expression of the *Anabaena flos-aquae* GV gene cluster in mammalian cells. (**a**) Schematic of the codon-optimized monocistronic plasmid sets used in this study. (**b**) Representative BURST images (top) and SBR quantification (n=5, bottom) of transient GV expression in HEK293T cells 3 days after co-transfection of mixtures with varying *gvpA/B* fold excess relative to their respective assembly factor plasmids. (**c**) Diagram of GV structure with GvpC highlighted in orange. (**d**) Representative xAM ultrasound images (top) and SBR quantification (n=6, bottom) of transient co-transfection experiments of *Anabaena flos-aquae* GV plasmids (4-fold *gvpA* excess) with and without *gvpC* at varying acoustic pressures. *B. megaterium* GV (at 2-fold *gvpB* excess) and GFP expression is included for quantitative comparison. (**e**) Schematic of the mARGAna polycistronic plasmids. (**f**) Representative BURST images (top) and SBR quantification (n=4, bottom) of transient GV expression in HEK293T cells 3 days after co-transfection of mARGAna mixtures with varying *gvpA* fold excess relative to the assembly factor plasmid. (**g**) Schematic of MDA-MB-231-mARGAna engineering (created with BioRender.com and FlowJo). The final population was ~95% double positive for *gvpA* and *gvpNJKFGWV* expression. (**h**) Representative xAM images (top) and SBR quantification (n=5, bottom) of MDA-MB-231-mARGAna cells at 0.54 MPa after 1, 2, 4 and 6 days of 1 μg/ml doxycycline induction. (**i**) Representative xAM images (top) and SBR quantification (n=4, bottom) of MDA-MB-231-mARGAna cells at

0.42 MPa as a function of doxycycline concentration after 4 days of expression. (**j**) Representative xAM images (top) and SBR quantification (n=4, bottom) of induced and uninduced MDA-MB-231-mARGAna cells as a function of time while imaged with sequentially increased acoustic pressures, starting at 0.31 MPa. (**k**) Representative xAM images (top) and SBR quantification (n=4, bottom) of induced MDA-MB-231-mARGAna cells at 0.61 MPa as a function of cell concentration. Limit of detection was 300k cells/mL (p=0.0181) by unpaired one-sided t-test. For j, thick lines represent the mean of 4 replicates and thin lines represent ± standard deviation. For b, d, f, h, i and k, gray lines connect the means of the replicates. All ultrasound image scalebars represent 1 mm.

at utilizing GvpA to form GVs or that excess GvpA may pose less of a burden to cells than excess GvpB. To further improve GvpA expression, we stabilized the *gvpA* transcript with WPRE-hGH poly(A) elements,[58] which resulted in peak signal at lower *gvpA*:chaperone ratios (**Fig. 4-ED8, a-b**).

We next looked for nonlinear ultrasound contrast. GvpC is a minor structural protein in *Anabaena flos-aquae* GVs that mechanically reinforces the GV shell (**Fig. 4-4c**).[8] Chemical removal of GvpC from purified GVs *in vitro* enhances nonlinear ultrasound scattering by allowing the GVs to deform more strongly in response to acoustic pressure.[33,59,60] When we omitted *gvpC* from our mammalian co-transfection mixture, we observed a dramatic enhancement of nonlinear signal in xAM, peaking at transmit pressure of around 0.6 MPa (**Fig. 4-4d**). By comparison, transfections including *gvpC* produced a much weaker xAM signal, while *B. megaterium* plasmids and GFP-expressing cells did not produce appreciable nonlinear contrast at any pressure. The omission of *gvpC* did not appreciably alter BURST contrast (**Fig. 4-ED8c**).

To create a convenient vector for mammalian expression of *Anabaena flos-aquae* GVs, we constructed a polycistronic plasmid linking the assembly factor genes *gvpNJKFGWV* through P2A co-translational cleavage elements. *gvpA* was supplied on a separate plasmid to enable stoichiometric tuning. The *gvpA* and *gvpNJKFGWV* plasmids were labeled with IRES-BFP and P2A-GFP, respectively, to allow for fluorescent analysis and sorting. Both transcripts were driven by CMV promoters. We termed this pair of plasmids mARG$_{\text{Ana}}$: mammalian ARGs adapted from *Anabaena flos-aquae* (**Fig. 4-4e**). mARG$_{\text{Ana}}$ produced nonlinear ultrasound contrast in HEK293T cells transiently co-transfected with a 1- to 6-fold molar excess of *gvpA* (**Fig. 4-4f**).

To produce a stable cancer cell line expressing mARG$_{\text{Ana}}$, we cloned our polycistronic constructs into PiggyBac integration plasmids under a doxycycline-inducible TRE promoter.[61,62] As a clinically relevant cancer model, we chose the human breast cancer cell line MDA-MB-231, which is widely used in tumor xenograft studies. We engineered these cells to constitutively co-express the rtTA transactivator and Antares optical reporter[63], transduced them with a mixture of mARG$_{\text{Ana}}$ and PiggyBac transposase plasmids at a 2:1 (*gvpA:gvpNJKFGWV*) molar ratio, and fluorescently sorted for co-expression of Antares, GFP and BFP (**Fig. 4-4g** and **Fig. 4-ED8d**).

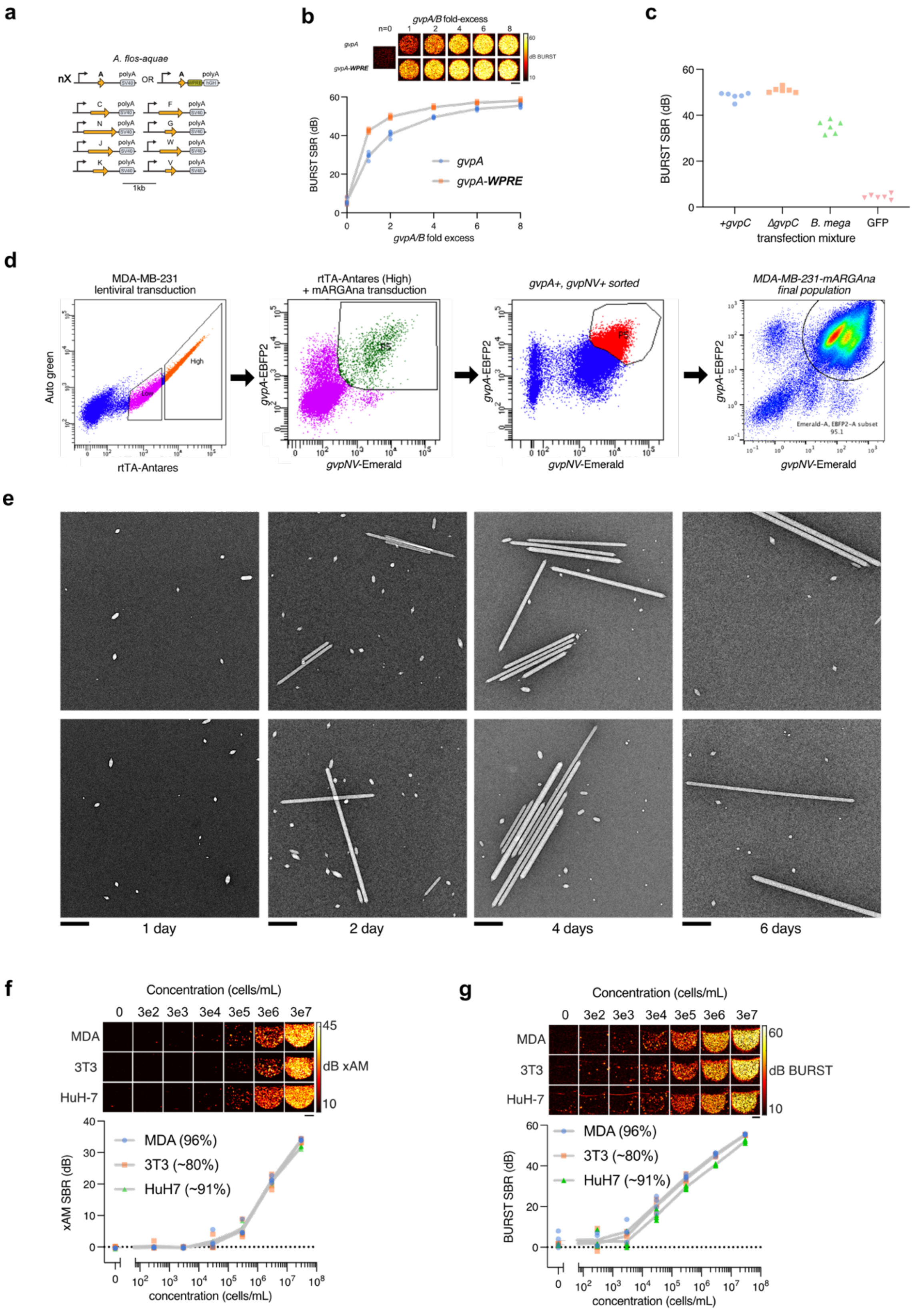


Figure 4-ED8: Additional data on heterologous expression of the *Anabaena flos-aquae* GV gene cluster in mammalian cells. (**a**) Schematic of the codon-optimized *Anabaena flos-aquae* monocistronic plasmid set

used in this study. (**b**) Representative BURST images (top) and SBR quantification (n=5, bottom) of transient GV expression in HEK293T cells 3 days after co-transfection of mixtures with varying *gvpA* fold excess relative to their respective assembly factor plasmids, with and without WPRE elements on the *gvpA* DNA. Gray lines connect the means of replicates. (**c**) BURST SBR quantification (n=6) of 293T cells expressing constructs tested in **Fig. 4-4d**. (**d**) FACS and flow cytometry data for production of the MDA-MB-231-mARGAna cell line. rtTA expressing "High" cells were collected for subsequent mARG/transposase transduction. Cells expressing *gvpA* and *gvpNV* were sorted twice to arrive at the final ~95% pure population. (**e**) TEM images of GVs purified from MDA-MB-231-mARGAna detergent lysates after 1, 2, 4 and 6 days of 1 µg/ml doxycycline induction (n=35 micrographs for day 1, 29 for day 2, 32 for day 4 and 7 for day 6). Scalebars represent 0.5 µm. (**f**) Representative xAM images (top) and SBR quantification (n=4 biologically independent samples examined in one experiment, bottom) of induced MDA-MB-231-mARGAna, 3T3-mARGAna and HuH7-mARGAna cells at 0.61 MPa as a function of cell concentration. Limit of detection was 300k cells/mL for MDA-MB-231 (p=0.0181) and 3T3 (p=0.0321), and 30k cells/mL for HuH7 (p=0.0158) by unpaired one-sided t-tests. (**g**) Representative BURST images (top) and SBR quantification (n=4, bottom) of induced MDA-MB-231-mARGAna, 3T3-mARGAna and HuH7-mARGAna cells as a function of cell concentration. Limit of detection was 30k cells/mL for MDA-MB-231 (p=0.011) and HuH7 (p=0.006), and 3k cells/mL for 3T3 (p=0.027) by unpaired one-sided t-tests. For (f) and (g), gray lines connect the means of the replicates and scalebars represent 1 mm. Percentages in parentheses represent mARGAna-positive cells determined by flow cytometry.

The resulting polyclonal MDA-MB-231-mARG$_{Ana}$ cells showed xAM contrast after a single day of doxycycline induction, which increased through day 6 (**Fig. 4-4h**). We confirmed the expression of GVs by electron microscopy (**Fig. 4-ED8e**). The ultrasound signal increased steeply with increasing doxycycline doses up to 1 µg/mL, then saturated (**Fig. 4-4i**). xAM signal was detected starting from an acoustic pressure of 0.31 MPa, while uninduced control cells did not produce signal at any pressure (**Fig. 4-4j**). At pressures above 0.42 kPa, the xAM signal gradually decreased over time, indicating the partial collapse of GVs. We chose 0.42 MPa as the xAM imaging pressure for subsequent experiments, providing the optimal balance of signal stability and strength.

We obtained similar results with mouse 3T3 fibroblasts and human HuH7 hepatocytes, with both cell lines able to be imaged with similar sensitivity to MDA-MB-231 cells when expressing mARG$_{Ana}$ (**Fig. 4-4k, Fig. 4-ED8f**). At 300k cells/mL for MDA-MB-231 and 3T3, and at 30k cells/mL for HuH7, the non-destructive xAM detection limits of these cells surpassed the destructive imaging sensitivity of first-generation mARG$_{Mega}$ by one to two orders of magnitude.[7] With BURST imaging, mARG$_{Ana}$ cells could be detected still more sensitively, at concentrations down to 3,000 cells/mL for 3T3 cells and 30,000 cells/mL for MDA-MB-231 and HuH7 **(Fig. 4-ED8g)**.

mARG$_{Ana}$-based imaging of* in vivo *gene expression patterns

We next tested the ability of mARG$_{Ana}$ to reveal the spatial distribution of gene expression in tumor xenografts. We formed orthotopic tumors by injecting MDA-MB-231-mARG$_{Ana}$ cells bilaterally in the fourth mammary fat pads of female immunocompromised mice. The mice were then split into doxycycline-induced and uninduced groups. We acquired ultrasound images of the tumors as they grew, with 3 imaging sessions distributed over 12 days (**Fig. 4-5a**). All induced tumors produced bright and specific xAM contrast starting from the first timepoint (day 4), whereas the uninduced tumors did not (**Fig. 4-5b**). The acquisition of adjacent planes allowed 3D visualization of expression patterns (**Supplementary Video 1**). The nonlinear xAM signal was highly specific to the viable tumor

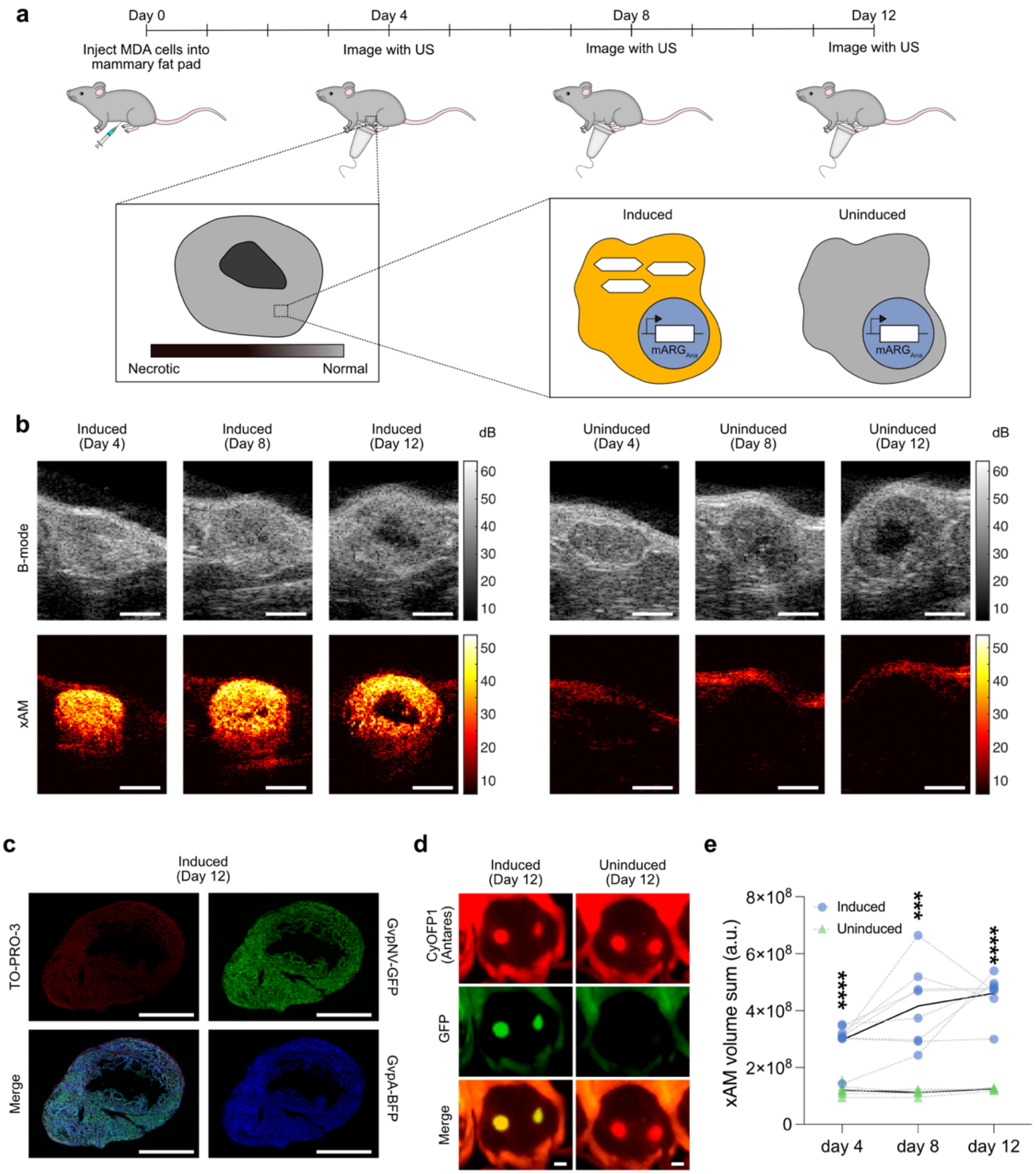


Figure 4-5: ***In situ*** **mARG$_{\text{Ana}}$ expression enables nondestructive ultrasound imaging of orthotopic tumors.** (**a**) Diagram of the *in vivo* protocol for assessing *in situ* mARG$_{\text{Ana}}$ expression in orthotopic tumors. Mice were injected bilaterally in 4th mammary fat pads with engineered MDA-MB-231-mARG$_{\text{Ana}}$ human breast adenocarcinoma cells on day 0. mARG$_{\text{Ana}}$ expression was induced by regular intraperitoneal (IP) doxycycline injections starting from the day of tumor injections. Tumors were imaged with ultrasound after 4, 8 and 12 days of expression. (**b**) Representative middle sections of B-mode and xAM ultrasound tomograms of MDA-MB-231 mARG$_{\text{Ana}}$ tumors induced with doxycycline (left; n=8 tumors from 4 mice) and uninduced control (right; n=7 tumors from 4 mice on day 4, n=5 tumors from 4 mice on days 8 and 12) imaged on day 4, 8 and 12. Scalebars represent 2 mm. See supplementary video for the full ultrasound tomogram of the induced tumor at day 12. (**c**) Fluorescence micrograph of a 100 nm thin tumor section (n=2 tumors). Green color shows GFP fluorescence, blue color shows BFP fluorescence, and red color shows TO-PRO-3 nuclear stain. See **Fig. 4-ED9** for the uninduced control (n=1 tumor). Scalebars represent 2 mm. (**d**) Whole-animal fluorescence imaging of induced

(left; n=8 tumors from 4 mice) and uninduced (right; n=5 tumors from 3 mice) mouse after 12 days of expression. All tumors are constitutively expressing CyOFP1 (Antares, red) whereas mARG$_{\text{Ana}}$ expression is linked to expression of GFP (green). The left (reader's right) tumors are shown in panel (b). Scalebars represent 5 mm. (**e**) Three-dimensional sum of xAM signal from ultrasound tomograms of induced (n=8) and uninduced (n=7 tumors from 4 mice on day 4, n=5 tumors from 4 mice on days 8 and 12) tumors from one experiment, plotted on a linear scale in arbitrary units. Black curves connect the means and grey curves connect points for each mouse. Asterisks represent statistical significance by two-tailed, unpaired Student's t-tests between induced and uninduced conditions (p values from left to right: 0.000622, 0.001554, 0.001554) (**** = $p<0.0001$, *** = $p<0.001$).

cells, being absent outside tumor boundaries and within the necrotic cores of the larger tumors. This observed spatial pattern was corroborated by fluorescence microscopy of fixed tumor sections obtained on day 12 (**Fig. 4-5c**, **Fig. 4-ED9a**), confirming the ability of mARG$_{\text{Ana}}$ to report microscale patterns of gene expression noninvasively in living animals at depths greater than 0.6 cm. In contrast, *in vivo* fluorescence images lacked information about the spatial distribution of gene expression within the tumor (**Fig. 4-5d**). The induced tumors had significantly higher total signal than the uninduced controls at all time points (**Fig. 4-5e**).

To demonstrate mARG$_{\text{Ana}}$ imaging at depths beyond what can be easily shown in mice, we successfully imaged MDA-MB-231-mARG$_{\text{Ana}}$ cells through a slab of beef liver thicker than 1 cm (**Fig. 4-ED9, b-c**).

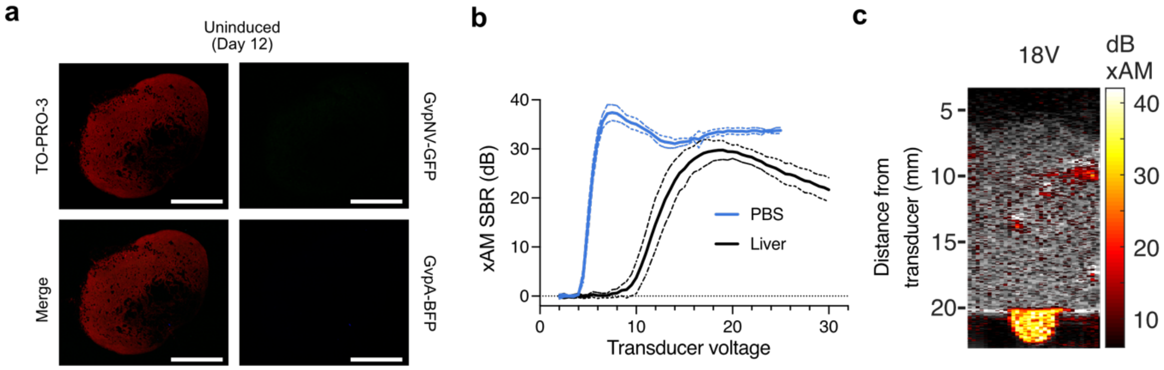


Figure 4-ED9: Tumor histology and imaging through thick tissue. (**a**) Fluorescence micrograph of a 100 nm-thick tissue section from an uninduced tumor. The red color shows TO-PRO-3 nuclear stain; GFP and BFP fluorescence were not observed for this control tumor. Scale bars represent 2 mm. (**b**) Quantification of xAM SBR of MDA-MB-231-mARG$_{\text{Ana}}$ cells imaged under >1 cm of beef liver tissue (n=7) or under PBS (n=4) as a function of transducer voltage. Thick lines represent the mean of replicates and thin lines represent ± standard deviation. (**c**) Representative xAM/B-mode overlay of MDA-MB-231-mARG$_{\text{Ana}}$ cells imaged under liver tissue using L10-5v transducer at 18V for xAM.

mARG$_{\text{Ana}}$ imaging enables real-time US-guided biopsy

One of the most common uses of ultrasound is to spatially guide procedures such as biopsies. ARGs provide the possibility for such procedures to be targeted based on gene expression. However, because real-time procedure guidance requires non-destructive imaging, such guidance was not possible with previous ARGs. To demonstrate an ultrasound-guided procedure with second-generation ARGs, we generated chimeric tumors in mice via adjacent subcutaneous injection of mARG_{Ana}-expressing and Antares-expressing MDA-MB-231 cells (**Fig. 4-6a**). After 3D tomograms confirmed chimeric tumor composition (**Supplementary Videos 2 and 3**) we targeted fine needle aspiration biopsies to either the xAM-positive or the xAM-negative regions of each tumor (**Fig. 4-6b, Supplementary Videos 4 and 5**). Flow cytometry showed a high success rate in obtaining cells with the expected mARG_{Ana}-positivity or negativity (**Fig. 4-6c, Fig. 4-S11)**.

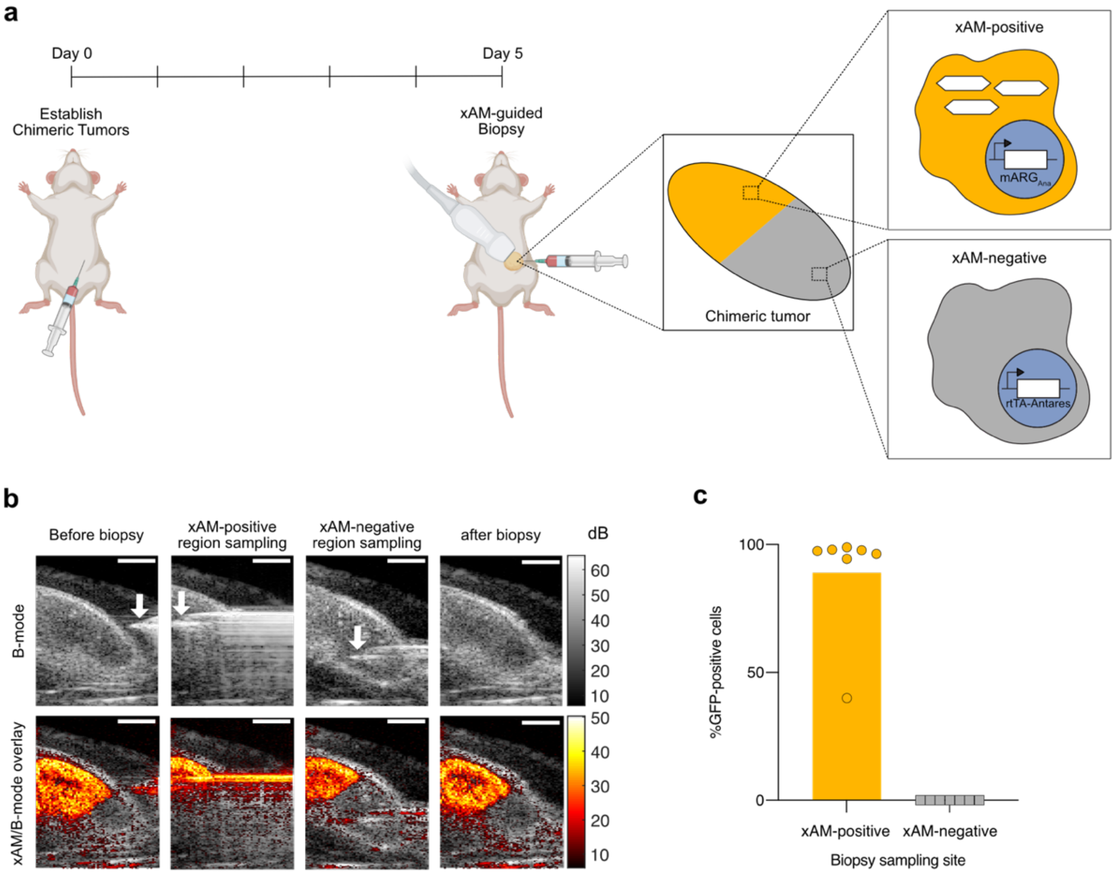


Figure 4-6: xAM imaging of mARG_{Ana} enables ultrasound-guided genetically selective tumor biopsy. (**a**) Diagram of the *in vivo* protocol for establishing chimeric tumors, *in situ* expression of GVs and tumor biopsy. Chimeric tumors were established on day 0. mARG_{Ana} expression was induced by regular intraperitoneal (IP) doxycycline injections starting from the day of tumor injections. Tumors were imaged with ultrasound and biopsied after 5 days of expression. (**b**) Representative B-mode and xAM an ultrasound-guided biopsy procedure. Scalebars represent 2 mm. See Supplementary Videos 2,3,4 and 5 for the full ultrasound tomogram of the induced chimeric tumor, 3D reconstruction of the chimeric tumor as well as the videos of the full procedure. (**c**) Results of flow cytometric analysis of biopsied samples from xAM-positive and xAM-negative regions of chimeric tumors (n=7 tumors from 4 mice). See **Fig. 4-S11** for flow cytometric gating strategy. Bar height represents the mean, circles represent individual data points.

4.4 DISCUSSION

Our results establish two second-generation ARG constructs—bARG_{Ser} and mARG_{Ana}—that provide unprecedented real-time detection sensitivity and specificity when expressed in bacteria and mammalian cells. These gene clusters produce bright nonlinear ultrasound contrast when expressed *in situ*, either by bacterial agents colonizing the necrotic core of a tumor, or by the tumor cells themselves. Their contrast enables the dynamic monitoring of the precise locations and transcriptional activities of these cells over multiple days. Furthermore, real-time nondestructive imaging of ARGs enables ultrasound-guided procedures such as biopsies targeted to genetically defined cells.

With the improvements described here, we anticipate that our ARGs will be useful for various applications that demand the noninvasive imaging of cells several mm or deeper inside the body. bARGs could be used to track therapeutic bacteria as they home to and proliferate in tumors or other target organs to assess whether the therapy is working as designed.[64–67] Furthermore, knowing their spatial distribution can help guide tumor-focused interventions such as focused ultrasound.[68–70] Other potential applications include microbiome research and tracking of probiotics designed to diagnose or treat GI conditions.[71,72]

Meanwhile, second-generation mARGs could be used to visualize the growth and viability of tissues such as tumors and precisely sample their contents under ultrasound guidance. Similar approaches could be applied to the immune system, brain and organismal development. In addition, mARG expression could enable the tracking and sampling of therapeutic cells, such as T-cells or stem cells. The fact that the mARGs are based on *Anabaena flos-aquae* GVs will allow these ARGs to benefit from extensive molecular engineering of these GVs [2,6,59,73] and development of acoustic biosensors of cellular signals such as enzyme activity.[73]

The envisioned applications of both bacterial and mammalian ARGs will benefit from the relative simplicity and low cost of ultrasound compared to other non-invasive techniques such as nuclear imaging and MRI, while providing *in vivo* resolution and potential for human translatability beyond what is currently possible with optical methods. In making cells detectable with ultrasound imaging, ARGs help pure ultrasound rival photoacoustic tomography, which has a larger number of contrast agents but is technically more complex to implement.[74]

Additional improvements to the technology can be envisaged. First, although expression kinetics on the order of one day, as observed with ARGs, are acceptable in numerous scenarios, such as tumor growth, mammalian development, and stem cell expansion and migration, they are slower than those of fluorescent proteins.[75,76] Faster expression could potentially be achieved by pre-expressing some of the genes in the ARG cluster (*e.g.*, assembly factors) and conditionally expressing the remainder (*e.g.*, structural proteins). Second, the ability to multiplex different “colors” of ARGs nondestructively would enable discrimination between different mammalian or bacterial cell types in a tissue. This could be accomplished by engineering versions of GvpC that respond differently to pressure[59]. Third, it would be helpful to reduce the size of the *Serratia* GV gene cluster to make it easier to clone and incorporate with other genetic elements. Similarly, consolidating the two-part $\mathrm{mARG_{Ana}}$ construct into a single plasmid would increase transfection efficiency, while shortening it could facilitate viral packaging for delivery to endogenous cells. Fourth, while the *in vivo* expression of both $\mathrm{bARG_{Ser}}$ and $\mathrm{mARG_{Ana}}$ yielded strong and specific ultrasound signals without obvious effects on cell migration and growth, strategies to reduce construct size and tightly regulate ARG expression could be beneficial to reducing burden under more stringent or resource-limited conditions. Finally, while mice are a key model in biomedical research and an established proving ground for reporter gene technologies[77–81], future work is needed to test ARGs in additional species. The ability of second-generation ARGs to function in cells from two different mammalian species and two different bacterial species is an encouraging sign for their broader utility.

Out of practical necessity, our phylogenetic screen subsampled the available genetic space, and testing additional GV-encoding gene clusters could reveal ARGs with new or further-improved properties. Improvements can also be made in the screening strategy by synthesizing multiple versions of each putative gene cluster to eliminate potential regulatory genes[9,34,36–38] and screening a larger number of clusters in higher throughput. It is likely that GV gene clusters from certain species will be more compatible with specific heterologous hosts due to similarity in growth temperature, turgor pressure and host factors, making it useful to perform genetic screens with new species of interest for imaging.

Just as improvements to and adaptations of fluorescent proteins enabled a wide range of microscopy applications that were mere speculations when GFP was first harnessed as a biotechnology, the systematic development of next-generation ARGs could help realize the promise of sensitive, high-resolution noninvasive imaging of cellular function inside intact mammals.

4.5 METHODS

Genomic mining of ARG clusters

A literature search was conducted to find papers reporting the production of gas vesicles in any species. Search terms included “gas vesicle,” “gas vacuole,” “aerotope,” and “aerotype.” All species found are listed in **Supplementary Table 1**. If the report did not include a strain name, then any available 16S rRNA gene sequence was used (as it was assumed that any other strain of the same species would fall in the same place on the phylogenetic tree), but no GV gene cluster sequence was used (even if it was available for one or more strains of that species) because it was found during our analysis that: 1) several reports describe species for which some strains produce GVs but others do not, and 2) comparison of GV gene cluster sequences of multiple strains of the same species almost always showed differences—often very significant ones. Further, even if a reference stating that a given organism produced GVs was not available, 16S rRNA gene sequences from all members of the following genera were included because GV production is a diacritical taxonomic feature for these genera: *Dolichospermum*,[82] *Limnoraphis*[83] and *Planktothrix*.[84]

GV clusters were identified in genomes through a combination of annotations and sequence similarity to known *gvp* genes. However, there were two challenges in identifying all *gvp*s in a given genome: 1) there is little to no annotation for many *gvp*s, and 2) GV gene clusters are not always contiguous in genomes, and *gvp*s can occasionally be found hundreds of kb away from the main cluster(s). We attempted to only select "well-behaved" GV clusters for testing (*i.e.*, ones in which all *gvp*s identified in that genome were present in clusters, and these clusters contained a minimum of non-*gvp* genes, which could increase the metabolic burden of cluster expression without increasing GV yield), but it is possible that even for these clusters, some *gvp*s were not cloned.

Of our list of 288 strains reported to form gas vesicles, 270 had 16S rRNA gene sequences available (**Supplementary Table 1**). These were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/sites/batchentrez) using a custom Python script, and a multiple sequence alignment was constructed using Clustal Omega.[85] This alignment was used to generate a phylogenetic tree file using ClustalW2,[86] which was rendered using EvolView.[87] Only unique species are displayed in the phylogenetic trees in **Fig. 4-1a** and **Fig. 4-S1**.

Bacterial plasmid construction and molecular biology

Organisms were obtained from commercial distributors as indicated in **Supplementary Table 2**. If an organism was shipped as a liquid culture, the culture was centrifuged and the pellet resuspended in ddH2O, as it was found that even trace amounts of certain culture media could completely inhibit PCR. Fragments were amplified by PCR using Q5 polymerase and assembled into a pET28a(+) vector (Novagen) via Gibson Assembly using reagents from New England Biolabs (NEB). Sub-cloning and other modifications to plasmids were also performed with Gibson Assembly using reagents from NEB. Assemblies were transformed into NEB Stable *E. coli*. All constructs were verified by Sanger sequencing.

Halobacterium salinarum has two chromosomal GV gene clusters (plus additional plasmid-borne ones), which were cloned and tested separately. *Methanosarcina vacuolata* has only one cluster, but while its genome sequence in the NCBI database has two copies of *GvpA1* and one copy of *GvpA2*, our genomic PCR yielded a product with only one copy of *GvpA1*. In a second cloning step, we added a copy of *GvpA2* to the cloned cluster. While we were able to PCR *GvpA2* from the genome, it was not contiguous with the rest of the cluster. Therefore, we speculate that either there was an error in the assembly of the genome sequence (likely caused by the high sequence similarity of the *GvpA* genes), or that the genotype of our strain differs slightly from that of the strain sequenced.

In vitro bacterial expression of ARGs

For initial testing, all constructs were expressed in BL21(DE3) *E. coli* (NEB). Fifty µL of electrocompetent *E. coli* were transformed with 1.5 µL of purified plasmid DNA (Econospin 96-well filter plate, Epoch Life Science), and 1 mL of SOC medium (NEB) was added immediately after electroporation. These cultures were incubated at 37°C for 2 hr, and 150 uL was inoculated into larger 1.5 mL LB cultures containing 100 ug/mL kanamycin and 1% (w/v) glucose (for catabolite repression[88] of the BL21(DE3) PlacUV5 promoter) in a deep-well 96-well plate and grown overnight in a shaking incubator at 30°C. Square dual-layer LB agar plates were prepared with varying concentrations of IPTG and 100 ug/mL kanamycin in the bottom layer, and 1% (w/v) glucose and 100 ug/mL kanamycin in the top layer. Each layer was 25 mL in volume, and the top layer was poured within 30 min of plating. LB agar was incubated at 60°C for 12-36 hr after dissolution to allow it to degas. After the agar solidified, plates were dried at 37°C to remove all condensation on the top layer that would cause the bacterial patches to run together. A multichannel pipette was used to thoroughly mix overnight cultures and drop 1 µL of each culture onto the surface of the dual-layer plates, with care taken to avoid puncturing the agar which results in artifacts during ultrasound scans. Low-retention pipette tips were used, as it was found that the small volumes of culture would wet the outsides of standard pipette tips, resulting in inaccurate volume dispensing. Patches were allowed to dry completely before plates were flipped and incubated at 37°C for 24 hr or 30°C for 48 hr.

For *in vitro* expression experiments in EcN, the appropriate plasmids were first transformed via electroporation and the outgrowth was plated on LB (Miller)-agar plates with the appropriate antibiotic (25 µg/mL chloramphenicol or 50 µg/mL kanamycin) and 1% (w/v) glucose. The resulting colonies were used to inoculate 2 mL LB (Miller) medium with the appropriate antibiotic and 1% (w/v) glucose, and these cultures were incubated at 250 rpm and 37°C overnight. Glycerol stocks were prepared by mixing the overnight cultures in a 1:1 volume ratio with 50% (v/v) glycerol and storing at -80°C. The night before expression experiments, glycerol stocks were used to inoculate overnight cultures (2 mL LB medium with the appropriate antibiotic and 1% (w/v) glucose) which were incubated at 37°C and shaken at 250 rpm. For expression on solid media, 1 µL of overnight culture was dropped onto square dual-layer LB agar plates with 2X the final inducer (IPTG, aTc, or L-arabinose) concentration in the bottom layer, 1% (w/v) glucose in the top layer, and the appropriate antibiotic in both layers (50 µg/mL chloramphenicol or 100 µg/mL kanamycin). Plates were allowed to dry, and then inverted and incubated at 37°C for 24 hours before imaging with ultrasound. For expression in liquid media, 500 µL of each overnight culture was used to inoculate 50 mL LB supplemented with 0.4% (w/v) glucose and 25 µg/mL chloramphenicol in 250 mL baffled flasks. Cultures were incubated at 37°C and 250 rpm until reaching at OD_{600} of 0.1 - 0.3. At this point, cultures were induced by addition of L-arabinose and placed back at 37°C and 250 rpm. For time titration experiments, 0.1% (w/v) L-arabinose was used for induction and 0.5 mL of each culture was removed at 0, 1, 3, 5, 7, 9, 11, and 24 hours post-induction for OD_{600} and ultrasound measurements. For L-arabinose titration experiments, L-arabinose concentrations ranging from 0 to 1% (w/v) were used for induction, and cultures were incubated for 24 hours at 37°C and 250 rpm after addition of L-arabinose before ultrasound imaging. For cell concentration titration experiments, cultures were incubated for 24 hours at 37°C and 250 rpm after addition of 0.1% (w/v) L-arabinose before ultrasound imaging. All cultures were stored at 4°C or on ice until casting in phantoms and imaging with ultrasound. In all liquid culture experiments, cultures were prescreened for the presence of GVs by phase contrast microscopy before being imaged with ultrasound.

In vitro expression experiments in *S. Typhimurium* were performed as described above for EcN, except plasmids were transformed into an attenuated version of *Salmonella enterica* serovar Typhimurium strain SL1344[50], 2xYT medium was used instead of LB medium, and induction in liquid culture was performed by adding L-arabinose to the medium at the time of inoculation at an OD_{600} of 0.05.

To assess plasmid stability of pBAD-$bARG_{Ser}$-AxeTxe in EcN, the glycerol stock of this strain was used to inoculate 2 mL LB (Miller) supplemented with 25 µg/mL chloramphenicol and 1% (w/v) glucose, and this culture was incubated at 37°C and 250 rpm overnight. Twenty µL of the overnight culture was subcultured into 2 mL LB with 25 µg/mL chloramphenicol, 2 mL LB without antibiotics, and 2 mL LB without antibiotics and with 0.1% (w/v) L-arabinose, each in quadruplicate. Every 24 hours, 20 µL of each culture was sub-cultured into fresh media of the same conditions. All cultures were incubated at 37°C and 250 rpm. On days 1-3, 5, and 7, serial dilutions of each culture were plated on LB-agar without antibiotics, LB-agar with 25 µg/mL chloramphenicol, and LB-agar with 25 µg/mL chloramphenicol + 0.1% (w/v) L-arabinose + 0.4% (w/v) glucose. Plates were incubated at 37°C for at least 16 hours and colonies were counted and screened manually. Plasmid retention was assessed by taking the ratio of CFUs on LB-agar plates with chloramphenicol to CFUs on LB-agar plates without antibiotics. The presence of mutations that disrupt the ability to express functional $bARG_{Ser}$ was assessed by a loss of colony opacity on LB-agar plates with 25 µg/mL chloramphenicol + 0.1% (w/v) L-arabinose + 0.4% (w/v) glucose.

Microscopy of bacteria

For TEM imaging, cells expressing GVs were diluted to OD_{600} ~1 in 10 mM HEPES (pH 7.5) or culture media. 3 µL of the sample was applied to a freshly glow-discharged (Pelco EasiGlow, 15 mA, 1 min) Formvar/carbon-coated, 200 mesh copper grid (Ted Pella) for 1 min before being reduced to a thin film by blotting. Grids with cells were washed three times in 10 mM HEPES (pH 7.5), blotted, air-dried, and imaged without the stain. Image acquisition was performed using a Tecnai T12 (FEI, now Thermo Fisher Scientific) electron microscope operated at 120 kV, equipped with a Gatan Ultrascan 2k X 2k CCD.

For phase contrast microcopy (PCM) imaging, cells expressing GVs were scraped off from plates and re-suspended in PBS at an OD_{600} of 1-2, or liquid cultures were used directly. Suspensions were transferred to glass slides and PCM images were acquired using a Zeiss Axiocam microscope with a 40X Ph2 objective.

In vivo bacterial ARG expression

All *in vivo* experiments were performed under protocol 1735 or 1761, approved by the Institutional Animal Care and Use of Committee (IACUC) of the California Institute of Technology. Animals were housed in a facility maintained at 71-75^oF and 30-70% humidity, with a lighting cycle of 13 hours on & 11 hours off (light cycle 6:00 AM - 7:00 PM). For experiments involving tumor colonization with EcN, MC26 cells (biohippo, 400156) were grown in DMEM media in T225 flasks. After trypsinization and resuspension in PBS + 0.1 mg/mL DNAseI, 5 x 10^6 MC26 cells were injected subcutaneously into the right flank of 6–8-week-old female Balb/cJ mice. Tumors were allowed to grow for 14 days (reaching sizes of 200-300 mm^3) before injecting 10^8 EcN cells suspended in PBS via the lateral tail vein. The day before injection of EcN, Ibuprofen was added to the drinking water at 0.2 mg/mL to ameliorate side effects of EcN injections. To prepare the EcN for injection, the appropriate glycerol stocks were used to inoculate 2 mL LB + 1% (w/v) glucose + 25 ug/mL chloramphenicol which was incubated at 37^oC and 250 rpm overnight. The overnight culture (500 µL) was used to inoculate 50 mL LB + 0.4% (w/v) glucose + 25 µg/mL chloramphenicol in 250 mL baffled flasks, which was grown at 37^oC and 250 rpm until reaching an OD_{600} of 0.3 - 0.6. This culture was pelleted, washed 4 times with PBS, resuspended in PBS at an OD_{600} of 0.625, and used for injection. Three days after injection of EcN, mice were injected intraperitoneally with 120 mg L-arabinose to induce the EcN. Starting 24 hours after induction, ultrasound images of tumors were acquired as described in **Supplementary Note 1**. After imaging, 3.0 MPa acoustic pressure was applied throughout the tumor to collapse GVs, and mice were injected again intraperitoneally with 120 mg L-arabinose. The next day, mice were imaged again with ultrasound for re-expression of GVs. The following day, all mice were euthanized and tumors were excised, homogenized, serially diluted, and plated on selective media (LB-agar + 25 µg/mL chloramphenicol) as well as on induction plates (LB-agar + 25 µg/mL chloramphenicol + 0.4% (w/v) glucose + 0.1% (w/v) L-arabinose). Colonies on plates with chloramphenicol were manually counted to quantify the levels of colonization, and colonies on induction plates were screened for a non-opaque mutant phenotype.

Histology of tumors colonized by bacteria

Tumors were colonized with pBAD-bARG$_{\mathrm{Ser}}$-AxeTxe EcN following the same protocol as described above. The day after inducing GV expression with IP injections of L-arabinose, BURST, xAM and B-mode images of tumors were acquired as described above. Shortly after imaging, mice were euthanized by sedation with isoflurane and cervical dislocation. Tumors were resected, placed in 10% buffered formalin for 48 hours, and then washed and stored in 70% ethanol. Tumors were then cut in half along the approximate plane of imaging, placed in tissue cassettes, and sent to the Translational Pathology Core Laboratory at UCLA, which embedded samples in paraffin and performed H&E staining, immunohistochemistry, and microscopy imaging. Immunohistochemistry was performed using Opal IHC kits (Akoya Biosciences) according to the manufacturer's instructions. Tissue sections were incubated with either polyclonal rabbit anti-*E. coli* antibody (Virostat; catalogue number 1001) or non-reactive rabbit IgG isotype control antibody as a negative control. All sections were then incubated with an Opal 520 polymer anti-rabbit HRP antibody (Akoya Biosciences) and counterstained with DAPI. Sections were imaged in the appropriate fluorescence or brightfield channels using a high throughput scanning system (Leica Aperio VERSA) with 40 µm resolution.

Mammalian plasmid construction

Monocistronic plasmids were constructed using standard cloning techniques including Gibson assembly and conventional restriction and ligation using primers listed in **Supplementary Table 3**. Coding sequences for the *Anabaena flos-aquae* GV genes were codon-optimized and synthesized by Integrated DNA Technologies, and subcloned into a pCMVSport backbone with a CMV promoter. *gvpA*-WPRE-hGH polyA was constructed by subcloning *gvpA* between PstI and MluI sites of pCMVSport vector with WPRE-hGH polyA.

Polycistronic mARG$_{\mathrm{Ana}}$ assembly factor genes *gvpNJKFGWV* were synthesized by Twist Bioscience in a pTwist-CMV vector. Emerald GFP was subcloned in-frame downstream of the *gvpNJKFGWV* ORF via a P2A linker and the entire *pNJKFGWV*-GFP ORF was subcloned into a pCMVSport backbone with WPRE-hGH-poly(A) elements using NEBuilder HiFi DNA Assembly (NEB). *gvpA*-IRES-EBFP2-WPRE-hGH polyA was constructed by Gibson assembly of a PCR-amplified IRES-EBFP2 fragment into the XbaI site of *gvpA*-WPRE-hGH polyA plasmid.

PiggyBac transposon plasmids were constructed by PCR amplifying the region between the start codon of *gvpNJKFGWV* or *gvpA* and the end of the hGH poly(A) from the pCMVSport plasmids. The amplified regions were Gibson-assembled into the PiggyBac transposon backbone (System Biosciences) with a TRE3G promoter (Takara Bio) for doxycycline-inducible expression.

The lentiviral transfer plasmid with constitutively expressed tetracycline transactivator (pEF1α-rtTA-Antares-WPRE) was constructed as follows: pNCS-Antares was obtained from Addgene (#74279) and P2A was added to the N-terminus of Antares with a primer overhang during PCR. This fragment was subcloned into the lentiviral transfer plasmid pEF1α-rtTA-WPRE between rtTA and WPRE in-frame with the rtTA ORF using NEBuilder HiFi DNA Assembly.

HEK293T cell culture and transient transfection

HEK293T cells (ATCC, CLR-2316) were cultured in 24-well plates at 37˚C, 5% CO2 in a humidified incubator in 0.5ml DMEM (Corning, 10-013-CV) with 10% FBS (Gibco) and 1x penicillin/streptomycin until about 80% confluency before transfection. Briefly, transient transfection mixtures were created by mixing of around 600 ng of plasmid mixture with polyethyleneimine (PEI-MAX; Polysciences Inc.) at 2.58 µg polyethyleneimine per µg of DNA. The mixture was incubated for 12 minutes at room temperature and added drop-wise to HEK293T cells. Media was changed after 12-16 hours and daily thereafter. For *gvpA* titration experiments, pUC19 plasmid DNA was used to keep the total amount of DNA constant.

After three days of expression, cells were dissociated using Trypsin/EDTA, counted using disposable hemocytometers (Bulldog), and centrifuged at 300g for 6 minutes at room temperature. Cells were resuspended with 1% low-melt agarose (GoldBio) in PBS at 40˚C at ~30 million cells/mL (**Fig. 4-4b**,**f**),~15 million cells/mL (**Fig. 4-4d**), or ~7.5 million cells/mL (**Fig 4-S18c**) before loading into wells of preformed phantoms consisting of 1% agarose (Bio-Rad) in PBS.

Genomic integration and FACS

MDA-MB-231 (ATCC, HTB26), 3T3 (ATCC, CRL-1658) and HuH7 (JCRB0403) cells were cultured in DMEM (Corning, 10-013-CV) supplemented with 10% FBS (Gibco) and 1x penicillin/streptomycin at 37˚C and 5% CO2 in a humidified incubator unless noted otherwise. Cells were lentivirally transduced with pEF1α-rtTA-Antares-WPRE and sorted based on strong Antares fluorescence (Ex: 488 nm, Em: 610/20BP + 595LP) using BD FACSAria II for MDA-MB-231 cells and MACSQuant Tyto (B2 channel) for 3T3 and HuH7 cells. MDA-MB-231-rtTA-Antares cells were then electroporated in 20ul format using 4D-Nucleofector using CH-125 protocol in SF buffer (Lonza) with 1 µg PiggyBac transposon:transposase plasmid mixture (2:1 PB-gvpA:PB-gvpNV transposons, 285ng PiggyBac transposase). 3T3-rtTA-Antares cells were transfected with the same PiggyBac plasmid mixture using PEI-MAX in a 12-well format and HuH7 cells were transfected using Lipofectamine 3000. Cells were expanded into surface-treated T75 flasks in TET-free media and were induced for 12 hours with 1 µg/mL doxycycline before sorting for triple positive cells (gated for Antares, then Emerald and EBFP2). The sorted cells were returned to DMEM with TET-free FBS (Takara). MDA-MB-231-mARG$_{\text{Ana}}$ was sorted twice. The first round of sorting was performed with permissive gates and the enriched population was ~50% double positive for Emerald and EBFP2 as analyzed with MACSQuant VYB (Miltenyi Biotec). This population was sorted again with stricter gates to ~95% purity. 3T3-mARG$_{Ana}$ cells were sorted for strong Emerald and EBFP2 fluorescence using MACSQuant Tyto only once, yielding a population that was ~80% double positive. HuH7 cells were sorted twice to ~91% purity using MACSQuant Tyto. Cells were expanded in TET-free media and frozen in Recovery Cell Culture Freezing Medium (Gibco) using Mr. Frosty cell freezing container (Nalgene) filled with isopropanol at -80˚C, and then stored in liquid nitrogen vapor phase until use.

TEM imaging of GVs expressed in mammalian cells

For TEM, cells were cultured in 6-well plates in 2 mL media. 1 µg/mL doxycycline was added to the wells at indicated times with daily media plus doxycycline changes thereafter until harvest. Cells were lysed by adding 400 µL of Solulyse-M (Genlantis) supplemented with 25 units/mL Benzonase Nuclease (Novagen) directly to the 6-well plates and incubating for 1 hour at 4˚C with agitation. The lysates were then transferred to 1.5 mL microcentrifuge tubes. Eight hundred µL of 10 mM HEPES pH 7.5 was added to each tube, and lysates were centrifuged overnight at 300g and 8˚C. Thirty µL of the supernatant was collected from the surface from the side of the tube facing the center of the centrifuge rotor and transferred to a new tube. Three µL of each sample was loaded onto freshly glow-discharged (Pelco EasiGlow, 15mA, 1 min) formvar/carbon 200 mesh grids (Ted Pella) and blotted after 1 minute then air-dried. The unstained grids were imaged on a FEI Tecnai T12 transmission electron microscope equipped with a Gatan Ultrascan CCD.

Whole-animal fluorescence imaging of mARGAna-expressing orthotopic tumors and tumor fluorescence microscopy

Tumor xenograft experiments were conducted in NSG mice aged 12-weeks and 6 days (The Jackson Laboratory). To implement an orthotopic model of breast cancer, all the mice were female. MDA-MB-231-mARG$_{Ana}$ cells were grown in T225 flasks in DMEM supplemented with 10% TET-free FBS and penicillin/streptomycin until confluency as described above. Cells were harvested by trypsinization with 6 mL Trypsin/EDTA for 6 minutes and quenched with fresh media. Cells were washed once in DMEM without antibiotics or FBS before pelleting by centrifugation at 300g. Cell pellets were resuspended in 1:1 mixture of ice-cold Matrigel (HC, GFR) (Corning 354263) and PBS (Ca^{2+}, Mg^{2+}-free) at 30 million cells/mL. Fifty µL Matrigel suspensions were injected bilaterally into the 4th mammary fat pads at 1.5 million cells per tumor via subcutaneous injection. 12 hours after tumor injection and every 12 hours thereafter (except the mornings of ultrasound imaging sessions) test mice were intraperitoneally injected with 150 µL of saline containing 150 µg of doxycycline for induction of GV expression. Control mice were not injected with doxycycline.

Mice were imaged using ultrasound as described in **Supplementary Note 1**. On the last day of ultrasound imaging, mice were anesthetized with 100 mg/kg ketamine, and 10 mg/kg xylazine and whole-body imaged in supine position using ChemiDoc MP imaging system with Image Lab software (BIO-RAD). Fluorescence channels were set as follows: blue epi illumination with 530/28 filter for Emerald/GFP and 605/50 filter for Antares/CyOFP1. Images were processed and merged using the FIJI package of ImageJ.

After whole-body fluorescence imaging, mice were euthanized and tumors were resected and placed in 10% formalin solution for 24 hours at 4˚C, after which they were transferred to PBS. Fixed tumors were embedded in 2% agarose in PBS and sectioned to 100 µm slices using a vibratome. Sections were stained with TO-PRO-3 nucleus stain, mounted using Prolong Glass (Invitrogen) and imaged using a Zeiss LSM 980 confocal microscope with ZEN Blue. Images were processed using the FIJI package of ImageJ. For micrographs of tumors from both induced and uninduced mice, the Emerald channel was capped between 0 and 25497, EBFP2 channel between 0 and 17233 and TO-PRO-3 channel between 5945 and 53136 for display.

4.6 Data availability

Plasmids will be made available through Addgene upon publication. All other materials and data are available from the corresponding author upon reasonable request. Genomic sequence information was downloaded from the NCBI sequence database via Batch Entrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez).

4.7 Code availability

Ultrasound data acquisition and analysis code will be made available on the Shapiro Lab GitHub at https://github.com/shapiro-lab upon publication.

4.8 ACKNOWLEDGEMENTS

The authors would like to thank Dianne Newman for a sample of *Streptomyces coelicolor* A3(2), and Yunfeng Li, Avinoam Bar-Zion and Hongyi (Richard) Li for help with tissue histology. Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. Mammalian cell sorting was performed at the Analytical Cytometry Core at City of Hope in Duarte, CA. Confocal microscopy was performed in the Beckman Institute Biological Imaging Center. This research was supported by the National Institutes of Health (R01-EB018975 to M.G.S.) and Pew Charitable Trust. R.C.H. was supported by the Caltech Center for Environmental Microbial Interactions. M.T.B. was supported by an NSF GRFP fellowship. Related research in the Shapiro Laboratory is supported by the David and Lucille Packard Foundation, the Burroughs

Wellcome Fund, the Heritage Medical Research Institute and the Chen Zuckerberg Initiative. M.G.S. is an investigator of the Howard Hughes Medical Institute.

4.9 AUTHOR CONTRIBUTIONS

R.C.H., M.T.B., M.D., K.W., M.B.S., P.D., Z.J., M.Y.Y., A.F. and R.D. planned and performed experiments. R.C.H. conceived and performed the phylogenetic screening experiments. M.T.B. and M.D. performed all *in vivo* experiments, with help from M.B.S and P.B.-L.. P.D. and M.D. performed TEM imaging. D.R.M. built the ultrasound plate-scanning setup, and D.R.M. and D.P.S. wrote the associated MATLAB scripts for controlling it. Z.J. and D.P.S. wrote the MATLAB scripts for ultrasound imaging of EcN *in vitro* and *in vivo*. Z.J. performed the calibration of the L22-14v transducer. M.H.A. provided the axe-txe stability cassette, and advised on tumor colonization experiments. R.C.H., M.T.B., M.D., D.P.S., P.D., and Z.J. analyzed data. R.C.H., M.T.B., M.D., and M.G.S. wrote the manuscript with input from all other authors. M.G.S. supervised the research.

4.10 COMPETING INTERESTS

R.C.H., M.T.B., M.D., K.W., A.F., M.Y.Y. and M.G.S. are co-inventors on two patent applications related to this work filed by and assigned to the California Institute of Technology.

4.11 REFERENCES

1. Piraner, D. I. *et al.* Going Deeper: Biomolecular Tools for Acoustic and Magnetic Imaging and Control of Cellular Function. *Biochemistry* **56**, 5202–5209 (2017).
2. Maresca, D. *et al.* Biomolecular Ultrasound and Sonogenetics. *Annu. Rev. Chem. Biomol. Eng.* **9**, 229–252 (2018).
3. Errico, C. *et al.* Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. *Nature* **527**, 499–502 (2015).
4. Christensen-Jeffries, K. *et al.* Super-resolution Ultrasound Imaging. *Ultrasound Med. Biol.* **46**, 865–891 (2020).
5. Shapiro, M. G. *et al.* Biogenic Gas Nanostructures as Ultrasonic Molecular Reporters. *Nat. Nanotechnol.* **9**, 311–316 (2014).
6. Bourdeau, R. W. *et al.* Acoustic reporter genes for noninvasive imaging of microbes in mammalian hosts. *Nature* **553**, 86–90 (2018).
7. Farhadi, A., Ho, G. H., Sawyer, D. P., Bourdeau, R. W. & Shapiro, M. G. Ultrasound imaging of gene expression in mammalian cells. *Science* **365**, 1469–1475 (2019).
8. Walsby, A. E. Gas vesicles. *Microbiol. Rev.* **58**, 94–144 (1994).
9. Pfeifer, F. Distribution, formation and regulation of gas vesicles. *Nat. Rev. Microbiol.* **10**, 705–715 (2012).
10. Sawyer, D. P. *et al.* Ultrasensitive ultrasound imaging of gene expression with signal unmixing. *Nat. Methods* **18**, 945–952 (2021).
11. Matz, M. V. *et al.* Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**, 969–973 (1999).
12. Lambert, G. G. *et al.* Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing. *PLOS Biol.* **18**, e3000936 (2020).
13. Rodriguez, E. A. *et al.* The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends Biochem. Sci.* **42**, 111–129 (2017).
14. Drepper, T. *et al.* Reporter proteins for in vivo fluorescence without oxygen. *Nat. Biotechnol.* **25**, 443–445 (2007).
15. Chow, B. Y. *et al.* High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98–102 (2010).
16. Klapoetke, N. C. *et al.* Independent optical excitation of distinct neural populations. *Nat. Methods* **11**, 338–346 (2014).
17. Govorunova, E. G. *et al.* RubyACRs, nonalgal anion channelrhodopsins with highly red-shifted absorption. *Proc. Natl. Acad. Sci.* **117**, 22833–22840 (2020).
18. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).

19. Harrington, L. B. *et al.* A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1424 (2017).
20. Burstein, D. *et al.* New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
21. Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 2577–2590 (2014).
22. Gasiunas, G. *et al.* A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nat. Commun.* **11**, 5512 (2020).
23. Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10**, 99–105 (2014).
24. Mak, W. S. *et al.* Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat. Commun.* **6**, 10005 (2015).
25. Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
26. Johns, N. I. *et al.* Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nat. Methods* **15**, 323–329 (2018).
27. Abel, A. J., Hilzinger, J. M., Arkin, A. P. & Clark, D. S. *Systems-informed genome mining for electroautotrophic microbial production*. http://biorxiv.org/lookup/doi/10.1101/2020.12.07.414987 (2020) doi:10.1101/2020.12.07.414987.
28. Durrant, M. G. *et al. Large-scale discovery of recombinases for integrating DNA into the human genome*. http://biorxiv.org/lookup/doi/10.1101/2021.11.05.467528 (2021) doi:10.1101/2021.11.05.467528.
29. Gurbatri, C. R. *et al.* Engineered Probiotics for Local Tumor Delivery of Checkpoint Blockade Nanobodies. *Sci. Transl. Med.* **12**, (2020).
30. Duong, M. T.-Q., Qin, Y., You, S.-H. & Min, J.-J. Bacteria-cancer interactions: bacteria-based cancer therapy. *Exp. Mol. Med.* **51**, 1–15 (2019).
31. Leventhal, D. S. *et al.* Immunotherapy with engineered bacteria by targeting the STING pathway for anti-tumor immunity. *Nat. Commun.* **11**, 2739 (2020).
32. Zhang, Y. *et al.* E. coli Nissle 1917-Derived Minicells for Targeted Delivery of Chemotherapeutic Drug to Hypoxic Regions for Cancer Therapy. *Theranostics* **8**, 1690–1705 (2018).
33. Maresca, D., Sawyer, D. P., Renaud, G., Lee-Gosselin, A. & Shapiro, M. G. Nonlinear X-Wave Ultrasound Imaging of Acoustic Biomolecules. *Phys. Rev. X* **8**, 041002 (2018).

34. Tashiro, Y., Monson, R. E., Ramsay, J. P. & Salmond, G. P. C. Molecular genetic and physical analysis of gas vesicles in buoyant enterobacteria. *Environ. Microbiol.* **18**, 1264–1276 (2016).
35. Oren, A., Pri-El, N., Shapiro, O. & Siboni, N. Gas vesicles isolated from Halobacterium cells by lysis in hypotonic solution are structurally weakened. *FEMS Microbiol. Lett.* **252**, 337–341 (2005).
36. Ramsay, J. P., Williamson, N. R., Spring, D. R. & Salmond, G. P. C. A quorum-sensing molecule acts as a morphogen controlling gas vesicle organelle biogenesis and adaptive flotation in an enterobacterium. *Proc. Natl. Acad. Sci.* **108**, 14932–14937 (2011).
37. Lee, C. M., Monson, R. E., Adams, R. M. & Salmond, G. P. C. The LacI–Family Transcription Factor, RbsR, Is a Pleiotropic Regulator of Motility, Virulence, Siderophore and Antibiotic Production, Gas Vesicle Morphogenesis and Flotation in Serratia. *Front. Microbiol.* **8**, (2017).
38. Li, N. & Cannon, M. C. Gas Vesicle Genes Identified in Bacillus megaterium and Functional Expression in Escherichia coli. *J. Bacteriol.* **180**, 2450–2458 (1998).
39. Sawyer, D. P. *et al.* Ultrasensitive ultrasound imaging of gene expression with signal unmixing. *Nat. Methods* **18**, 945–952 (2021).
40. Farhadi, A. *et al.* Genetically Encoded Phase Contrast Agents for Digital Holographic Microscopy. *Nano Lett.* **20**, 8127–8134 (2020).
41. Lu, G. J. *et al.* Genetically Encodable Contrast Agents for Optical Coherence Tomography. *ACS Nano* **14**, 7823–7831 (2020).
42. Li, Z. & Rinas, U. Recombinant protein production associated growth inhibition results mainly from transcription and not from translation. *Microb. Cell Factories* **19**, (2020).
43. Borkowski, O., Ceroni, F., Stan, G.-B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.* **33**, 123–130 (2016).
44. Sleight, S. C. & Sauro, H. M. Visualization of Evolutionary Stability Dynamics and Competitive Fitness of Escherichia coli Engineered with Randomized Multigene Circuits. *ACS Synth. Biol.* **2**, 519–528 (2013).
45. Schultz, M. Clinical use of E. coli Nissle 1917 in inflammatory bowel disease. *Inflamm. Bowel Dis.* **14**, 1012–1018 (2008).
46. Stritzker, J. *et al.* Tumor-specific colonization, tissue distribution, and gene induction by probiotic Escherichia coli Nissle 1917 in live mice. *Int. J. Med. Microbiol.* **297**, 151–162 (2007).
47. Sonnenborn, U. & Schulze, J. The non-pathogenic Escherichia coli strain Nissle 1917–features of a versatile probiotic. *Microb. Ecol. Health Dis.* **21**, 122–158 (2009).
48. Grady, R. & Hayes, F. Axe–Txe, a broad-spectrum proteic toxin–antitoxin system specified by a multidrug-resistant, clinical isolate of Enterococcus faecium. *Mol. Microbiol.* **47**, 1419–1432 (2003).

49. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat. Methods* **15**, 387 (2018).

50. Din, M. O. *et al.* Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536**, 81–85 (2016).

51. Broadway, K. M. & Scharf, B. E. Salmonella Typhimurium as an Anticancer Therapy: Recent Advances and Perspectives. *Curr. Clin. Microbiol. Rep.* **6**, 225–239 (2019).

52. Hecht, A., Endy, D., Salit, M. & Munson, M. S. When Wavelengths Collide: Bias in Cell Abundance Measurements Due to Expressed Fluorescent Proteins. *ACS Synth. Biol.* **5**, 1024–1027 (2016).

53. Chien, T., Doshi, A. & Danino, T. Advances in bacterial cancer therapies using synthetic biology. *Curr. Opin. Syst. Biol.* **5**, 1–8 (2017).

54. Weinberger, M. & Helmstetter, C. E. Chromosome replication and cell division in plasmid-containing Escherichia coli B/r. *J. Bacteriol.* **137**, 1151–1157 (1979).

55. Seo, J.-H. & Bailey, J. E. Effects of recombinant plasmid content on growth properties and cloned gene product formation in Escherichia coli. *Biotechnol. Bioeng.* **27**, 1668–1674 (1985).

56. Birnbaum, S. & Bailey, J. E. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant Escherichia coli. *Biotechnol. Bioeng.* **37**, 736–745 (1991).

57. Karim, A. S., Curran, K. A. & Alper, H. S. Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. *FEMS Yeast Res.* **13**, 10.1111/1567-1364.12016 (2013).

58. Donello, J. E., Loeb, J. E. & Hope, T. J. Woodchuck Hepatitis Virus Contains a Tripartite Posttranscriptional Regulatory Element. *J. Virol.* **72**, 5085–5092 (1998).

59. Lakshmanan, A. *et al.* Molecular Engineering of Acoustic Protein Nanostructures. *ACS Nano* **10**, 7314–7322 (2016).

60. Maresca, D. *et al.* Nonlinear ultrasound imaging of nanoscale acoustic biomolecules. *Appl. Phys. Lett.* **110**, (2017).

61. Zhou, X., Vink, M., Klaver, B., Berkhout, B. & Das, A. T. Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther.* **13**, 1382–1390 (2006).

62. Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.* **10**, 81 (2010).

63. Chu, J. *et al.* A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nat. Biotechnol.* **34**, 760–767 (2016).

64. Dang, L. H., Bettegowda, C., Huso, D. L., Kinzler, K. W. & Vogelstein, B. Combination bacteriolytic therapy for the treatment of experimental tumors. *Proc. Natl. Acad. Sci.* **98**, 15155–15160 (2001).
65. Westphal, K., Leschner, S., Jablonska, J., Loessner, H. & Weiss, S. Containment of Tumor-Colonizing Bacteria by Host Neutrophils. *Cancer Res.* **68**, 2952–2960 (2008).
66. Toley, B. J. & Forbes, N. S. Motility is Critical for Effective Distribution and Accumulation of Bacteria in Tumor Tissue. *Integr. Biol. Quant. Biosci. Nano Macro* **4**, 165–176 (2012).
67. Raman, V., Van Dessel, N., O'Connor, O. M. & Forbes, N. S. The motility regulator flhDC drives intracellular accumulation and tumor colonization of Salmonella. *J. Immunother. Cancer* **7**, 44 (2019).
68. Abedi, M. H. *et al.* Ultrasound-controllable engineered bacteria for cancer immunotherapy. *Nat. Commun.* **13**, 1585 (2022).
69. Bar-Zion, A. *et al.* Acoustically triggered mechanotherapy using genetically encoded gas vesicles. *Nat. Nanotechnol.* **16**, 1403–1412 (2021).
70. Piraner, D. I., Abedi, M. H., Moser, B. A., Lee-Gosselin, A. & Shapiro, M. G. Tunable thermal bioswitches for in vivo control of microbial therapeutics. *Nat. Chem. Biol.* **13**, 75–80 (2017).
71. Charbonneau, M. R., Isabella, V. M., Li, N. & Kurtz, C. B. Developing a new class of engineered live bacterial therapeutics to treat human diseases. *Nat. Commun.* **11**, 1738 (2020).
72. Jimenez, M., Langer, R. & Traverso, G. Microbial therapeutics: New opportunities for drug delivery. *J. Exp. Med.* **216**, 1005–1009 (2019).
73. Lakshmanan, A. *et al.* Acoustic Biosensors for Ultrasound Imaging of Enzyme Activity. *Nat. Chem. Biol.* **16**, 988–996 (2020).
74. Xia, J., Yao, J. & Wang, L. V. Photoacoustic tomography: principles and advances. *Electromagn. Waves Camb. Mass* **147**, 1–22 (2014).
75. Subramanian, S. & Srienc, F. Quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. *J. Biotechnol.* **49**, 137–151 (1996).
76. Hui, C.-Y., Guo, Y., Zhang, W. & Huang, X.-Q. Rapid monitoring of the target protein expression with a fluorescent signal based on a dicistronic construct in Escherichia coli. *AMB Express* **8**, 81 (2018).
77. Genove, G., DeMarco, U., Xu, H., Goins, W. F. & Ahrens, E. T. A new transgene reporter for in vivo magnetic resonance imaging. *Nat Med* **11**, 450–454 (2005).
78. Gilad, A. A. *et al.* Artificial reporter gene providing MRI contrast based on proton exchange. *Nat Biotechnol* **25**, 217–219 (2007).
79. Chu, J. *et al.* A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nat Biotech* **34**, 760–767 (2016).

80. Schilling, F. *et al.* MRI measurements of reporter-mediated increases in transmembrane water exchange enable detection of a gene reporter. *Nat Biotechnol* **35**, 75–80 (2017).
81. Allouche-Arnon, H. *et al.* Computationally designed dual-color MRI reporters for noninvasive imaging of transgene expression. *Nat. Biotechnol.* **40**, 1143–1149 (2022).
82. Wacklin, P., Hoffmann, L. & Komárek, J. Nomenclatural validation of the genetically revised cyanobacterial genus Dolichospermum (RALFS ex BORNET et FLAHAULT) comb. nova. *Fottea* **9**, 59–64 (2009).
83. Komárek, J. *et al.* Polyphasic evaluation of Limnoraphis robusta, a water-bloom forming cyanobacterium from Lake Atitlán, Guatemala, with a description of Limnoraphis gen. nov. *Fottea* **13**, 39–52 (2013).
84. Komárek, J. Planktic oscillatorialean cyanoprokaryotes (short review according to combined phenotype and molecular aspects). in *Phytoplankton and Equilibrium Concept: The Ecology of Steady-State Assemblages* (eds. Naselli-Flores, L., Padisák, J. & Dokulil, M. T.) 367–382 (Springer Netherlands, 2003). doi:10.1007/978-94-017-2666-5_30.
85. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
86. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
87. Subramanian, B., Gao, S., Lercher, M. J., Hu, S. & Chen, W.-H. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* **47**, W270–W275 (2019).
88. Ammar, E. M., Wang, X. & Rao, C. V. Regulation of metabolism in Escherichia coli during growth on mixtures of the non-glucose sugars: arabinose, lactose, and xylose. *Sci. Rep.* **8**, 609 (2018).

4.12 INDEX OF SUPPLEMENTARY INFORMATION

Supplementary Note 1: Ultrasound imaging procedures

Supplementary Note 2: Parameters for fitting xAM ultrasound signal versus time data to an exponential decay function

Supplementary Figures 1-11

Supplementary Tables 1-3

Supplementary Video 1: xAM/B-mode tomogram of induced orthotopic tumor on day 12

Supplementary Video 2: xAM/B-mode tomogram of induced chimeric tumor on day 5

Supplementary Video 3: xAM/B-mode 3D reconstruction of the tumor from Video 2

Supplementary Video 4: xAM/B-mode video of chimeric tumor biopsy procedure sampling of xAM-positive region

Supplementary Video 5: xAM/B-mode video of chimeric tumor biopsy procedure sampling of xAM-negative region

4.13 SUPPLEMENTARY INFORMATION

Supplementary Note 1: Ultrasound imaging procedures

In vitro ultrasound imaging of bacteria expressing ARGs on solid media

Ultrasound imaging of bacterial patches was performed using a Verasonics Vantage programmable ultrasound scanning system and an L10-4v 128-element linear array transducer (Verasonics) with a center frequency of 6 MHz and an element pitch of 300 µm. Image acquisition was performed using a custom imaging script with a 64-ray-lines protocol and a synthetic aperture of 65 elements. The transmit waveform was set to a voltage of 50 V and a frequency of 10 MHz, with 1 waveform cycle and 67% intra-pulse duty cycle. In xAM mode, a custom sequence detailed previously[33] was used with an angle of 19.5°. RF data from 4 repeated acquisitions was coherently averaged prior to beamforming for each image plane.

Agar plates containing bacterial patches were coated with a thin layer of LB agar and immersed in PBS to allow acoustic coupling to the L10-4v transducer. The transducer was connected to a BiSlide computer-controlled 3D translatable stage (Velmex) and positioned above the plane of the plate at an angle of 15° from the vertical (to minimize specular reflection from the plastic dishes and agar) and a distance of 20 mm from the bacterial patches. The imaging sequence was applied sequentially to acquire image planes covering the full area of all plates. A custom script was used to automate the scan by controlling the motor stage in tandem with the ultrasound system, translating 0.5 mm in the azimuthal direction between rows and 19.5 mm in the lateral direction between columns. In the case of differential imaging scans, the full scan sequence was repeated after returning the motor stage to its origin and adjusting the voltage of the transducer.

For image processing and analysis, custom beamforming scripts were applied on-line to reconstruct image planes from the acquired RF data at each location. The intensity data for each plane was saved for off-line processing. All image planes were concatenated to form a 3D volume with all plates and colonies. A 2D image of the colonies was extracted from the 3D volume by taking the maximum intensity over a manually-defined depth range for all voxel columns. 2D differential images were obtained by subtracting the post-collapse 2D image from the pre-collapse 2D image. Bacterial patch intensities were then quantified from these 2D images. Sample ROIs were drawn around the center of each patch to avoid artefacts from the edges, and background ROIs were drawn around representative regions without patches. The signal-to-background ratio (SBR) was calculated as the mean pixel intensity of the sample ROI divided by the mean pixel intensity of the background. Conversion to decibels (dB) was calculated as 20*log10(SBR). For display, images were normalized by dividing by the average background signal of all images being compared and setting the lower and upper limits of the colormaps to be the same, where the lower limit was equal to a constant A times the average background and the upper limit was equal to a constant B times the maximum pixel intensity out of all images being compared; images were then converted to dB. For xAM and differential xAM images of bacterial patches, A was set to 1 and B was set to 0.5.

In vitro ultrasound imaging of bacteria expressing ARGs suspended in agarose phantoms

To create phantoms for ultrasound imaging of bacteria from liquid cultures or suspended in PBS from patches on solid media, wells were cast with a custom 3D-printed mold using 1% (w/v) agarose in PBS, which was degassed by incubating at 65°C for at least 16 hours. Cultures or cell suspensions to be analyzed were diluted in ice-cold PBS to 2x the final desired cell concentration (calculated from the measured OD_{600}), incubated at 42°C for one minute, and mixed 1:1 with 1% (w/v) agarose in PBS at 42°C for a final concentration of 1x. This mixture was then loaded into the wells in duplicate and allowed to solidify. Care was taken not to introduce bubbles during this process. The phantoms were submerged in PBS, and ultrasound images were acquired using a Verasonics Vantage programmable ultrasound scanning system and an L22-14v 128-element linear array transducer with a center frequency of 18.5 MHz with 67%-6-dB bandwidth, an element pitch of 100 μm, an elevation focus of 8 mm, and an elevation aperture of 1.5 mm. The transducer was attached to a custom-made manual translation stage to move between samples. B-mode and xAM images were acquired using the same parameters as described previously:[73] the frequency and transmit focus were set to 15.625 MHz and 5 mm, respectively, and each image was an average of 50 accumulations. B-mode imaging was performed with a conventional 128-ray-lines protocol, where each ray line was a single pulse transmitted with an aperture of 40 elements. xAM imaging was performed using a custom sequence detailed previously[33] with an angle of 19.5° and an aperture of 65 elements. The transmitted pressure at the focus was calibrated using a Fibre-Optic Hydrophone (Precision Acoustics), and the peak positive pressure was used as the "acoustic pressure" in **Fig. 4-3**. BURST images were acquired as a series of pAM images as described previously,[7] except the focus was set to 6 mm, and the acoustic pressure was set to 0.15 MPa (1.6V) for the first 10 frames and 3.0 MPa (25V) for the last 46 frames.

To measure the xAM signal at varying acoustic pressures, an automated voltage ramp imaging script was used to acquire an xAM image at each voltage step (0.5 V increments from 2 to 25 V), immediately followed by a B-mode acquisition at a constant voltage of 1.6 V (0.15 MPa) before another xAM acquisition at the next voltage step; the voltage was held constant for 10 seconds at each step before the image was saved. To measure the xAM and B-mode signals over time at various acoustic pressures, another script was used to automatically save an xAM or B-mode image every second while the voltage was automatically increased by 2 V approximately every 70 seconds. Each frame consisted of 64 ray lines, which took 180 μs each to acquire, giving a pulse repetition rate of 86.8 Hz. Based on these results, all subsequent *in vitro* xAM images of bARG$_{Ser}$-expressing EcN were acquired at 18V (1.29 MPa).

For the experiments in **Fig. 4-S8** and **Fig. 4-ED6**, a different transducer, an L22-14vX transducer, was used which had a different pressure-to-voltage calibration. Consequently, for ultrasound imaging of *S. Typhimurium*, xAM imaging was performed at 1.72 MPa (14V), unless otherwise noted, and BURST was performed using 0.16 MPa (1.6V) for the first 10 frames and 3.7 MPa (25V) for the final 46 frames.

xAM and B-mode image processing and analysis were performed as described above, except that custom beamforming scripts were applied off-line to reconstruct images from the saved RF data for each sample, no 3D reconstruction was performed as images captured at single locations, circular ROIs were drawn around sample and background regions (taking care to avoid bubbles) to calculate SBRs, and values of A=1.4 and B=0.5 were used to normalize images for display. BURST images were reconstructed using the signal template unmixing algorithm as described previously[10]; as above, circular ROIs were then drawn around sample and background regions to calculate SBRs and values of A=3 and B=1 were used to normalize images for display. Data were plotted using GraphPad Prism (v9.4.1).

In vivo bacterial ARG ultrasound imaging

Mice were anesthetized with 2% isoflurane and maintained at 37°C using a heating pad. Images were acquired using the L22-14v transducer attached to a manual translation stage described above. Any hair on or around the tumors was removed with Nair, and Aquasonic 100 ultrasound transmission gel was used to couple the transducer to the skin. Parabolic B-mode and parabolic AM (pAM) images were first acquired using a custom 128 ray line script. Each image was formed from 96 focused beam ray lines, each with a 32-element aperture and 6 mm focus. The transmit waveform was set to a voltage of 1.6V in B-mode or 8V in pAM and a frequency of 15.625 MHz, with 1 waveform cycle and 67% intra-pulse duty. In B-mode, each ray line was a single transmit with all 32 elements, and in pAM each ray line consisted of one transmit with all 32 elements followed by 2 transmits in which first the odd and then the even-numbered elements are silenced.[60] Subsequently, xAM images, additional B-mode images, and finally BURST images were acquired at the same location without moving the transducer using the same parameters as described above for the *in vitro* experiments (*e.g.,* 18V for xAM, 1.6V for B-mode, and 1.6V to 25V for BURST). At least two separate locations spaced at least 2 mm apart in each tumor were imaged with B-mode, pAM, and xAM. Ultrasound images of tumors were quantified as described above where the sample ROIs were drawn around the necrotic cores in the tumors and the background ROIs were drawn around regions in the gel above the mouse. Images were normalized and plotted on a dB scale as described above except the scaling factors were A=2.5 and B=1 for xAM and pAM and the corresponding B-mode tumor images, and A=10 and B=0.5 for BURST images.

In vitro ultrasound imaging of transient expression of GVs in HEK293T cells suspended in agarose phantoms

Phantoms were imaged using L22-14v transducer (Verasonics) while submerged in PBS on top of an acoustic absorber pad. For BURST imaging, wells were centered around the 8 mm natural focus of the transducer and a BURST pulse sequence was applied in pAM acquisition mode as described above, except the focus was set to 8 mm, and the acoustic pressure was set to 0.26 MPa (1.6V) for the first 10 frames and 2.11 MPa (10V) for the remaining frames. The xAM voltage ramps and B-mode images were acquired concurrently using the same parameters as described above, except the transducer voltage was varied from 4 to 24V in steps of 0.5V for xAM, and 10 frames, each consisting of 15 accumulations, were acquired per voltage. The well depth and the B-mode transmit focus were set to 5 mm. All image quantification was performed as described above, where the sample ROIs were drawn inside the well and the background ROIs were drawn around an empty region in the agarose phantom for SBR calculation. All images were normalized and plotted on a dB scale as described above except the scaling factors were A=2 and B=0.5.

In vitro ultrasound imaging of MDA-MB-231 mARGAna cells suspended in agarose phantoms

For all *in vitro* experiments, MDA-MB-231-mARG$_{\text{Ana}}$ cells were cultured in DMEM supplemented with 10% TET-free FBS and penicillin/streptomycin. For xAM imaging of MDA-MB-231-mARG$_{\text{Ana}}$ cells suspended in agarose phantoms, cells were cultured in 24-well plates in 0.5 mL media. For **Fig. 4-5h**, cells were seeded at 7,500 cells per well and induced with 1 µg/mL doxycycline after an overnight incubation and at subsequent days as indicated (5 replicates per condition), except for the uninduced control which was grown in a 10 cm dish without doxycycline. Media was changed daily thereafter until cell harvest. Cells were trypsinized with 100 µL Trypsin/EDTA for 6 minutes at 37˚C, after which the trypsin was quenched by addition of 900 µL media. The cell number was equalized between different days of expression at 140,000 cells and pelleted at 300g for 6 minutes. Cells were then resuspended in 20 µl 1% low-melt agarose (GoldBio) in PBS at 40˚C and loaded into the wells of preformed 1% agarose (Bio-Rad) phantoms in PBS. Ultrasound images were acquired with L22-14v 128-element linear array transducer (Verasonics). xAM voltage ramp and B-mode images were acquired concurrently using the same parameters as described above (the transducer voltage was varied from 4 to 24V in steps of 0.5V for xAM and 10 frames, each consisting of 15 accumulations, were acquired per voltage. The B-mode transmit focus was set to 5 mm). Images taken at the voltage that produced peak xAM signal (9V, 0.54 MPa) were chosen for quantification. For **Fig. 4-4i,j**, cells were seeded at 66,666 cells per well and induced with the indicated doxycycline concentrations after an overnight incubation in TET-free media (4 replicates per doxycycline concentration). Cells were incubated for 4 days with daily media/doxycycline changes. Cells were harvested as above, and ~420,000 cells from each condition were loaded per agarose phantom well. xAM and B-mode images were acquired concurrently using the same parameters as described above except the transducer voltage was varied from 6V to 10V in steps of 0.5V for xAM and 120 frames, each consisting of 15 accumulations, were acquired per voltage (~75 seconds/voltage). The B-mode transmit focus was set to 6 mm. Images taken at 7.5V (0.42 MPa) were chosen for display and quantification in **Fig. 4-4i** (doxycycline response). For **Fig. 4-4k** and **Fig. 4-ED4-8f-g,** cells were seeded in 10-cm dishes and induced as above for 4 days. Cells were harvested as above and resuspended at 60,000,000 cells/mL. 10-fold serial dilutions were performed with each cell line. Each cell dilution was mixed 1:1 with 2% low-melt agarose before loading into agarose phantom wells. Cells were imaged with an L22-14vX transducer at 5.5V (0.61 MPa) for xAM: the highest pressure that produced stable signal over a 30-second exposure and using 2V to 15V pAM BURST.

For imaging of MDA-MB-231 cells under thick liver tissue, cells were induced with doxycycline in T225 flasks for 4 days. Cells were harvested as above and resuspended at 30,000,000 cells/mL in 1% low-melt agarose in PBS prior to loading into agarose phantom wells. >1 cm beef liver section (99 Ranch Market) was overlaid on top of the agarose phantom and secured with needles. The phantom and liver were submerged in a PBS bath and the transducer was positioned 20 mm away from the interface between the liver and the agarose phantom. Ultrasound imaging was performed using a L10-4v linear array transducer (Verasonics) using the same parameters as above, except the xAM voltage was varied between 2V (0.078 MPa) and 30V (2.51 MPa). B-mode was acquired at 1.6V (0.25 MPa). Each voltage was held for 5 frames, each consisting of 15 accumulations.

All image quantification was performed as described above where the sample ROIs were drawn inside the well and the background ROIs were drawn around an empty region in the agarose phantom for SBR calculation. All Images were normalized and plotted on a dB scale as described above except the scaling factors were A=2 and B=0.5. The xAM/B-mode overlay was made with the B-mode image as background. A binary alpha mask was applied to the xAM image, giving pixel values lower than 2x the average background a value of 0 and all values above this threshold a value of 1.

In vivo ultrasound imaging of mARGAna expressing orthotopic tumors

For ultrasound imaging, mice were depilated around the 4th mammary fat pads using Nair (Aloe Vera) for ultrasound coupling with Aquasonic 100 gel. Mice were anesthetized with 2.5% isoflurane and maintained at 37˚C in supine position on a heating pad. The first imaging session (day 4) consisted of 8 induced tumors from 4 mice and 7 uninduced tumors from 4 mice. One of the uninduced mice died during the first imaging session, which resulted in two fewer uninduced control tumors for the remaining imaging sessions.

Ultrasound images were acquired with an L22-14v 128-element linear array transducer. xAM and B-mode images were acquired concurrently using the same parameters as described in the *in vitro* section above except the transducer voltage was held at constant 7.5V (0.42 MPa) for xAM and 3 frames, each consisting of 15 accumulations, were acquired per section. A motor stage was programed to move 100 µm per section for a total of 150 sections per tumor. The B-mode transmit focus was set to 6 mm. Ultrasound images of tumors were quantified as described above where the sample ROIs were drawn around the tumors and the background ROIs were drawn around regions in the gel above the mouse. Images were normalized and plotted on a dB scale as described above except the scaling factors were A=2 and B=0.5 for both xAM and the corresponding B-mode tumor images. The xAM volume quantification was performed by summing all pixel values from all sections in each tomogram between 2 mm and 10 mm in depth.

In vivo ***ultrasound-guided biopsy of mARG$_{Ana}$-expressing chimeric tumors.***

Chimeric tumor biopsy experiments were conducted in female NCG mice aged 8-weeks (Charles River Laboratories). MDA-MB-231-mARG$_{Ana}$ and MDA-MB-231-rtTA-Antares cells were grown and harvested as above. Cell pellets were resuspended in a 1:1 mixture of ice-cold Matrigel (HC, GFR) (Corning 354263) and PBS (Ca^{2+}- and Mg^{2+}-free) at 30 million cells/mL. 100 µL Matrigel suspensions of MDA-MB-231-mARG$_{Ana}$ were injected bilaterally into the 4th mammary fat pads at 3 million cells per tumor lobe via subcutaneous injection. After 1 hr, additional 100 µL Matrigel suspensions of MDA-MB-231-rtTA-Antares were injected close to the edge of the blisters created by the first injections to create dual-lobed chimeric tumors with heterogeneous gene expression patterns. Mice were intraperitoneally injected with 150 µL of saline containing 150 µg of doxycycline for induction of GV expression starting 12 hours after tumor injection and then every 12 hours thereafter for 5 days. Mice were prepared for ultrasound imaging as above. Ultrasound images were acquired with an L22-14vX 128-element linear array transducer. xAM and B-mode imaging was performed as above except the transducer voltage was held at 5.5V (0.481 MPa) for xAM and 1.6V (0.161 MPa) for B-mode and the motor stage was programed to move either 200 µm per section for whole-tumor scans or was held stationary for biopsy video acquisition. Image normalization and scaling was performed as above. xAM/B-mode overlay was made as above.

To perform a fine-needle aspiration biopsy, a 23G needle was fitted to a 3 mL Luer-lock syringe prefilled with PBS. The syringe was mounted on a 3D-printed holder attached to a manual translation stage. Each biopsy attempt consisted of positioning the ultrasound probe over a tumor and moving the needle into the field of view. The needle was then inserted into either the xAM-positive or xAM-negative region of the tumor, guided by live xAM and B-mode imaging. The needle was wiggled back-and-forth a couple times before pulling the syringe plunger to aspirate cells. The tumor sample was then ejected into a tube with PBS. The biopsy was repeated for attempts that did not produce a visible cell pellet. Each sample was treated with Trypsin/EDTA for 6 minutes at 37˚C, then quenched with fresh media.

Flow cytometry was performed with MACSQuant 10 (Miltenyi Biotec). GFP was measured with the B1 channel and Antares using B2. All biopsy attempts for a given tumor/sampling condition were analyzed separately, but their resulting FCS data files were concatenated. Data analysis was performed in FlowJo. For quantification of biopsy samples, each population was first gated for Antares-positive cells to exclude endogenous mouse cells. Antares-positive cells were then gated based on FSC/SSC and single cells were gated using FSC-A vs FSC-W. The resulting populations contained on average 6947 cells with a SD of 6960 cells and range between 72 and 21958 cells. %GFP-positive (mARG$_{\text{Ana}}$-positive) was assessed based on these resulting populations.

Supplementary Note 2: Parameters for fitting xAM ultrasound signal versus time data to an exponential decay function

Data from Fig. S8b were first converted to linear scaling (not decibel) and then normalized so that at t = 0, the xAM signal y was equal to 1. The normalized data were then fit to the equation $y = A*\exp(-k*t) + B$, where the parameter k is the rate of exponential decay, B corresponds to the fraction of signal remaining at $t = \infty$, and A corresponds to the fraction of signal lost between $t = 0$ and $t = \infty$. As the applied acoustic pressure was increased, the fraction of signal lost A increased and the fraction of signal remaining B decreased, while the rate of decay remained relatively constant with a slight increase. See **Fig. 4-S8d** for a plot of the fitted curves overlayed with the data.

Pressure (MPa)	**A**		**k (min^{-1})**		**B**		**Goodness of fit (R^2)**
	Value	SEM	Value	SEM	Value	SEM	
1.10	0.113	0.009	5.085	0.873	0.857	0.004	0.327
1.41	0.191	0.008	5.777	0.483	0.764	0.003	0.658
1.72	0.259	0.012	5.703	0.527	0.692	0.004	0.613
1.97	0.307	0.011	5.499	0.374	0.620	0.004	0.748
2.23	0.346	0.009	6.622	0.323	0.574	0.003	0.843

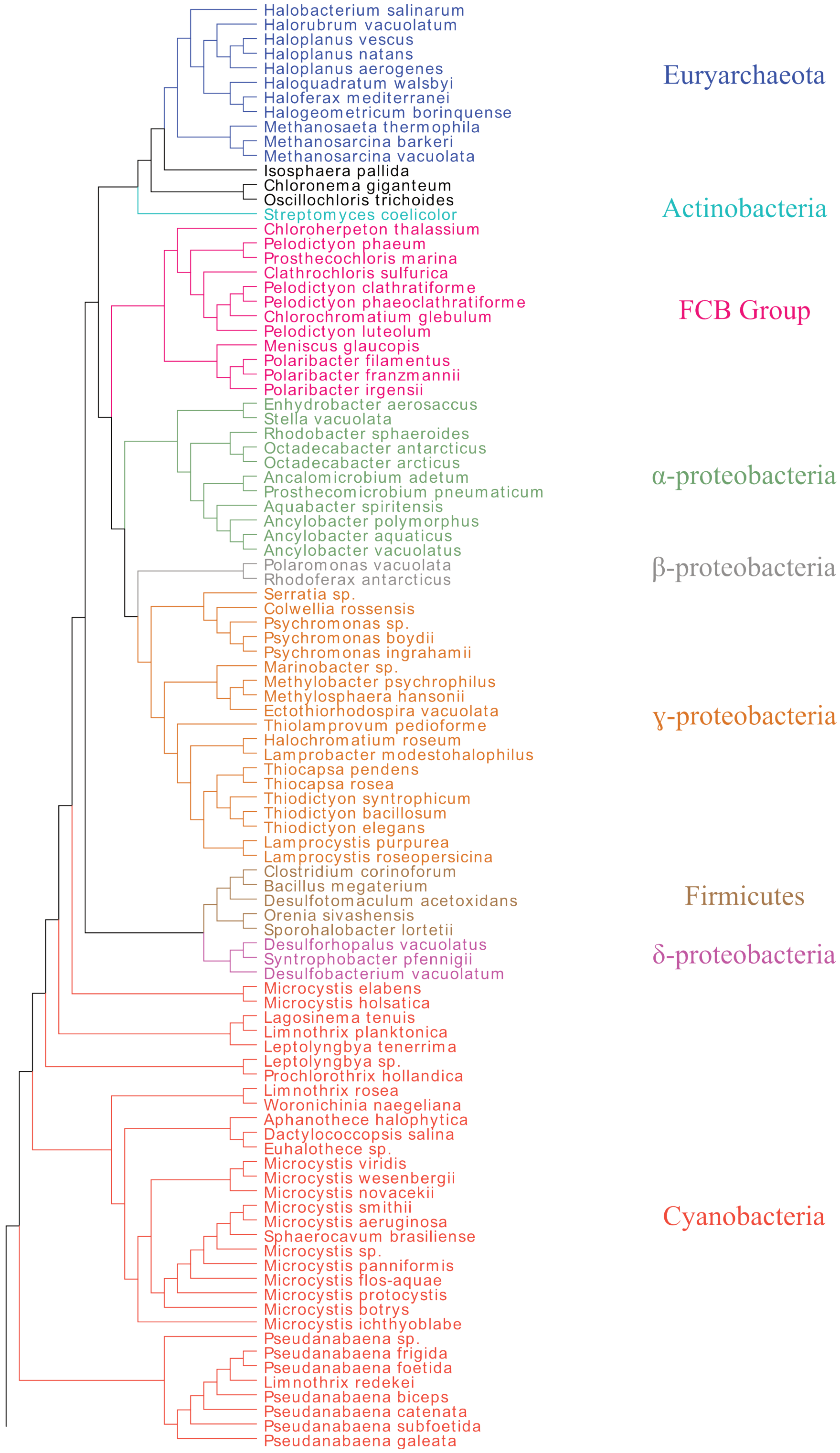

Halobacterium salinarum
Halorubrum vacuolatum
Haloplanus vescus
Haloplanus natans
Haloplanus aerogenes
Haloquadratum walsbyi
Haloferax mediterranei
Halogeometricum borinquense
Methanosaeta thermophila
Methanosarcina barkeri
Methanosarcina vacuolata
Isosphaera pallida
Chloronema giganteum
Oscillochloris trichoides
Streptomyces coelicolor
Chloroherpeton thalassium
Pelodictyon phaeum
Prosthecochloris marina
Clathrochloris sulfurica
Pelodictyon clathratiforme
Pelodictyon phaeoclathratiforme
Chlorochromatium glebulum
Pelodictyon luteolum
Meniscus glaucopis
Polaribacter filamentus
Polaribacter franzmannii
Polaribacter irgensii
Enhydrobacter aerosaccus
Stella vacuolata
Rhodobacter sphaeroides
Octadecabacter antarcticus
Octadecabacter arcticus
Ancalomicrobium adetum
Prosthecomicrobium pneumaticum
Aquabacter spiritensis
Ancylobacter polymorphus
Ancylobacter aquaticus
Ancylobacter vacuolatus
Polaromonas vacuolata
Rhodoferax antarcticus
Serratia sp.
Colwellia rossensis
Psychromonas sp.
Psychromonas boydii
Psychromonas ingrahamii
Marinobacter sp.
Methylobacter psychrophilus
Methylosphaera hansonii
Ectothiorhodospira vacuolata
Thiolamprovum pedioforme
Halochromatium roseum
Lamprobacter modestohalophilus
Thiocapsa pendens
Thiocapsa rosea
Thiodictyon syntrophicum
Thiodictyon bacillosum
Thiodictyon elegans
Lamprocystis purpurea
Lamprocystis roseopersicina
Clostridium corinoforum
Bacillus megaterium
Desulfotomaculum acetoxidans
Orenia sivashensis
Sporohalobacter lortetii
Desulforhopalus vacuolatus
Syntrophobacter pfennigii
Desulfobacterium vacuolatum
Microcystis elabens
Microcystis holsatica
Lagosinema tenuis
Limnothrix planktonica
Leptolyngbya tenerrima
Leptolyngbya sp.
Prochlorothrix hollandica
Limnothrix rosea
Woronichinia naegeliana
Aphanothece halophytica
Dactylococcopsis salina
Euhalothece sp.
Microcystis viridis
Microcystis wesenbergii
Microcystis novacekii
Microcystis smithii
Microcystis aeruginosa
Sphaerocavum brasiliense
Microcystis sp.
Microcystis panniformis
Microcystis flos-aquae
Microcystis protocystis
Microcystis botrys
Microcystis ichthyoblabe
Pseudanabaena sp.
Pseudanabaena frigida
Pseudanabaena foetida
Limnothrix redekei
Pseudanabaena biceps
Pseudanabaena catenata
Pseudanabaena subfoetida
Pseudanabaena galeata
Euryarchaeota
Actinobacteria
FCB Group
α-proteobacteria
β-proteobacteria
γ-proteobacteria
Firmicutes
δ-proteobacteria
Cyanobacteria

Figure 4-S1: 16S phylogenetic tree of all reported GV-producing organisms. Colors indicate groupings of phylogenetically similar organisms. Organisms from which GV genes were tested in *E. coli* are shown in **Fig. 4-1a**.

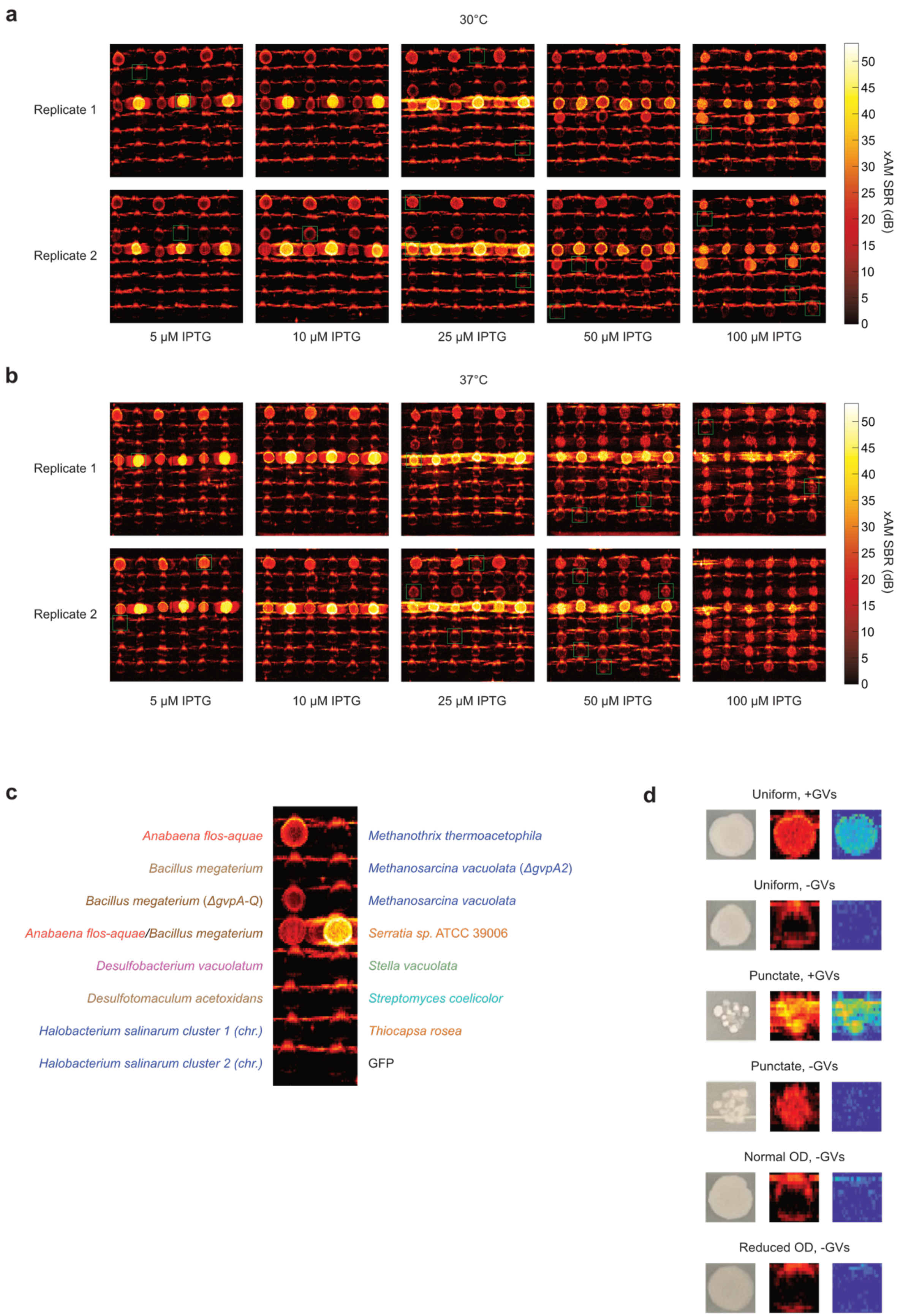

a
30°C
Replicate 1
Replicate 2
5 µM IPTG
10 µM IPTG
25 µM IPTG
50 µM IPTG
100 µM IPTG
50
45
40
35
30
25
20
15
10
5
0
xAM SBR (dB)
b
37°C
Replicate 1
Replicate 2
5 µM IPTG
10 µM IPTG
25 µM IPTG
50 µM IPTG
100 µM IPTG
50
45
40
35
30
25
20
15
10
5
0
xAM SBR (dB)
c
Anabaena flos-aquae
Bacillus megaterium
Bacillus megaterium (ΔgvpA-Q)
Anabaena flos-aquae/Bacillus megaterium
Desulfobacterium vacuolatum
Desulfotomaculum acetoxidans
Halobacterium salinarum cluster 1 (chr.)
Halobacterium salinarum cluster 2 (chr.)
Methanothrix thermoacetophila
Methanosarcina vacuolata (ΔgvpA2)
Methanosarcina vacuolata
Serratia sp. ATCC 39006
Stella vacuolata
Streptomyces coelicolor
Thiocapsa rosea
GFP
d
Uniform, +GVs
Uniform, -GVs
Punctate, +GVs
Punctate, -GVs
Normal OD, -GVs
Reduced OD, -GVs

Figure 4-S2: Optimization of expression conditions for all tested clusters in BL21(DE3) ***E. coli.*** (**a-b**) xAM images of bacterial patches expressing each GV cluster at varying inducer concentrations and temperatures. Green boxes indicate the patches shown in **Fig. 4-ED1a**. The IPTG concentration selected was the one that resulted in the highest xAM pre-minus-post-collapse difference signal (**Fig. 4-S5**) while not creating toxicity, as determined by whether the patch was uniform or punctate (**Fig. 4-S5a-b**). Some of the IPTG concentrations that led to toxicity also created significant xAM signal, but this signal did not originate from GVs, as indicated by the lack of xAM pre-minus-post-collapse signal difference (**Fig. 4-S4**). Further, there were some IPTG concentrations for certain genotypes that created significant xAM signal but no xAM pre-minus-post-collapse signal difference, and no visible toxicity (*e.g.*, *Streptomyces coelicolor*, *Thiocapsa rosea*, and GFP at 37˚C, 100 μM IPTG). This discrepancy was likely caused by subtle toxicity that is not apparent in optical images, but altered the texture of the patch enough to be detectable by US. (**c**) Key for genotypes tested in (a-b), with this pattern repeated in three pairs of columns replicated on each plate. (**d**) Examples of the effects of toxic genotypes on bacterial patches, and of artifacts that can appear in bacterial patch scan images. Bacteria themselves can produce significant xAM signal (especially when present in extremely high concentrations, as they are in the confluent patches imaged here), which can be seen in the forms of rings around all patches, regardless of GV expression status. Further, expression of toxic proteins (or of large amounts of otherwise non-toxic proteins, such as GFP) can interfere with bacterial growth; in extreme cases this results in significant cell death and a punctate appearance, and in less extreme cases it simply reduces the optical density of patches. GV expression can increase the optical density of patches, but only at high levels of GV expression. Punctate patches produce considerably more xAM signal than uniform ones, even in the absence of GV expression. The xAM pre-minus-post-collapse difference can be used to qualitatively determine if a patch produces GVs, but because collapse is incomplete in some cases, it is not an ideal method for quantitatively comparing genotypes.

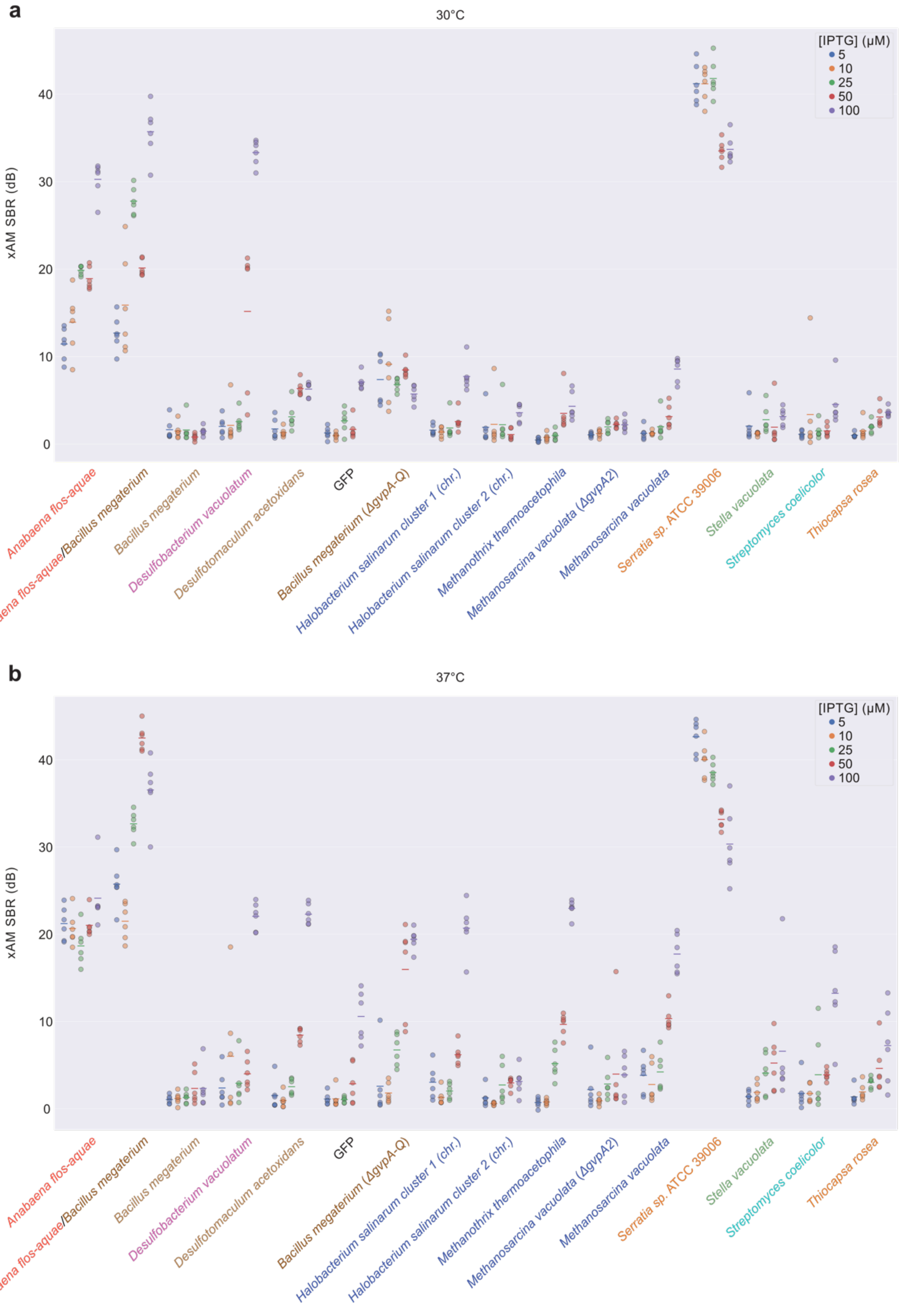


Figure 4-S3: Quantification of ultrasound signal for all samples shown in Fig. 4-S2a-b. (**a-b**) xAM SBR of the patches at 30°C (a) and 37°C (b) shown in **Fig. 4-S2a-b** (n=6; lines represent the mean).

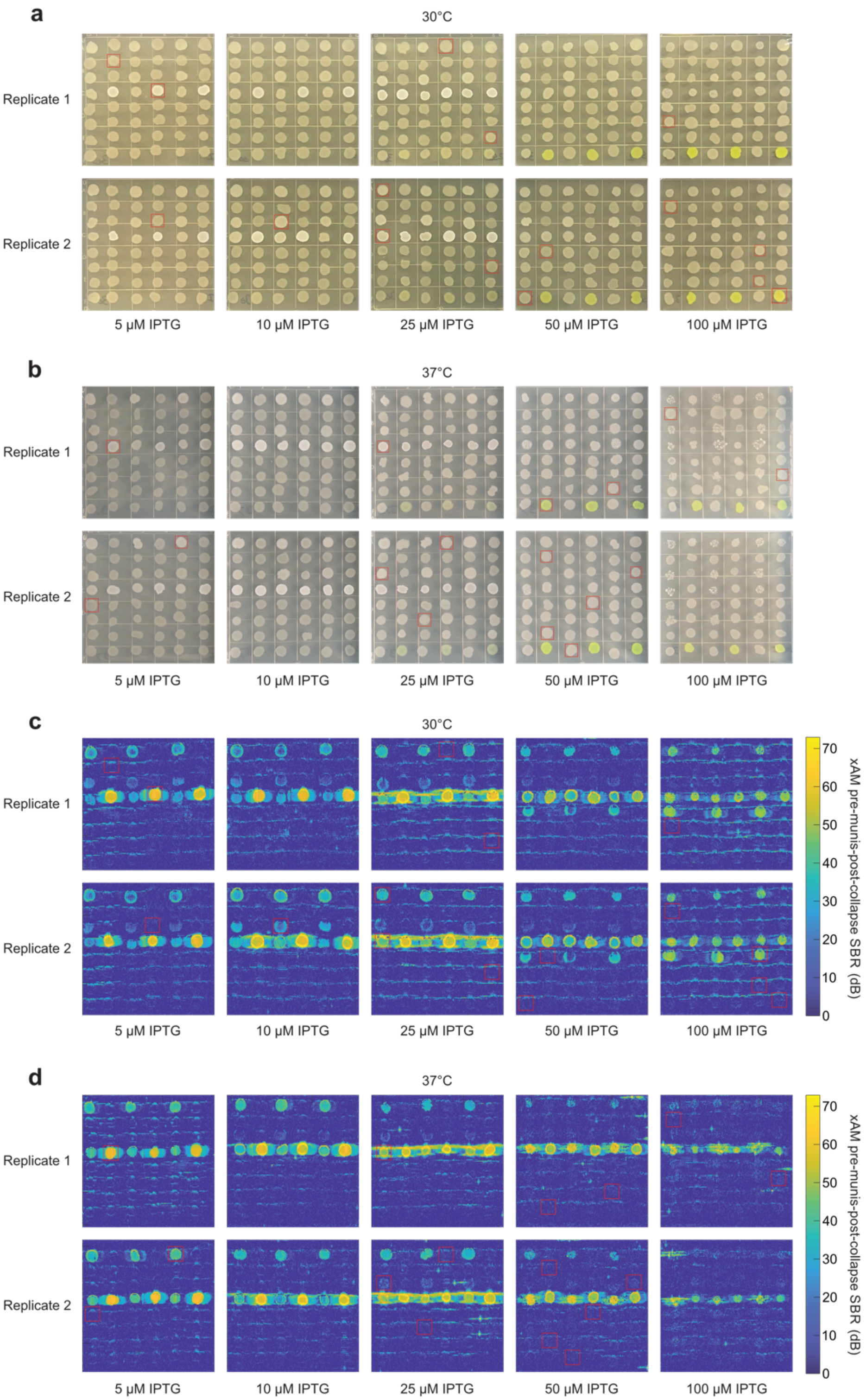


Figure 4-S4: Optical and xAM pre-minus-post-collapse difference images of all samples shown in Fig. 4-S2. (**a-b**) Optical images of patches at 30ºC (a) and 37ºC (b) shown in **Fig. 4-S2a-b**. (**c-d**) xAM pre-minus-post-collapse difference patches of samples at 30ºC (a) and 37ºC (b) shown in **Fig. 4-S2a-b**. Red boxes indicate the patches shown in **Fig. 4-ED1b-c**.

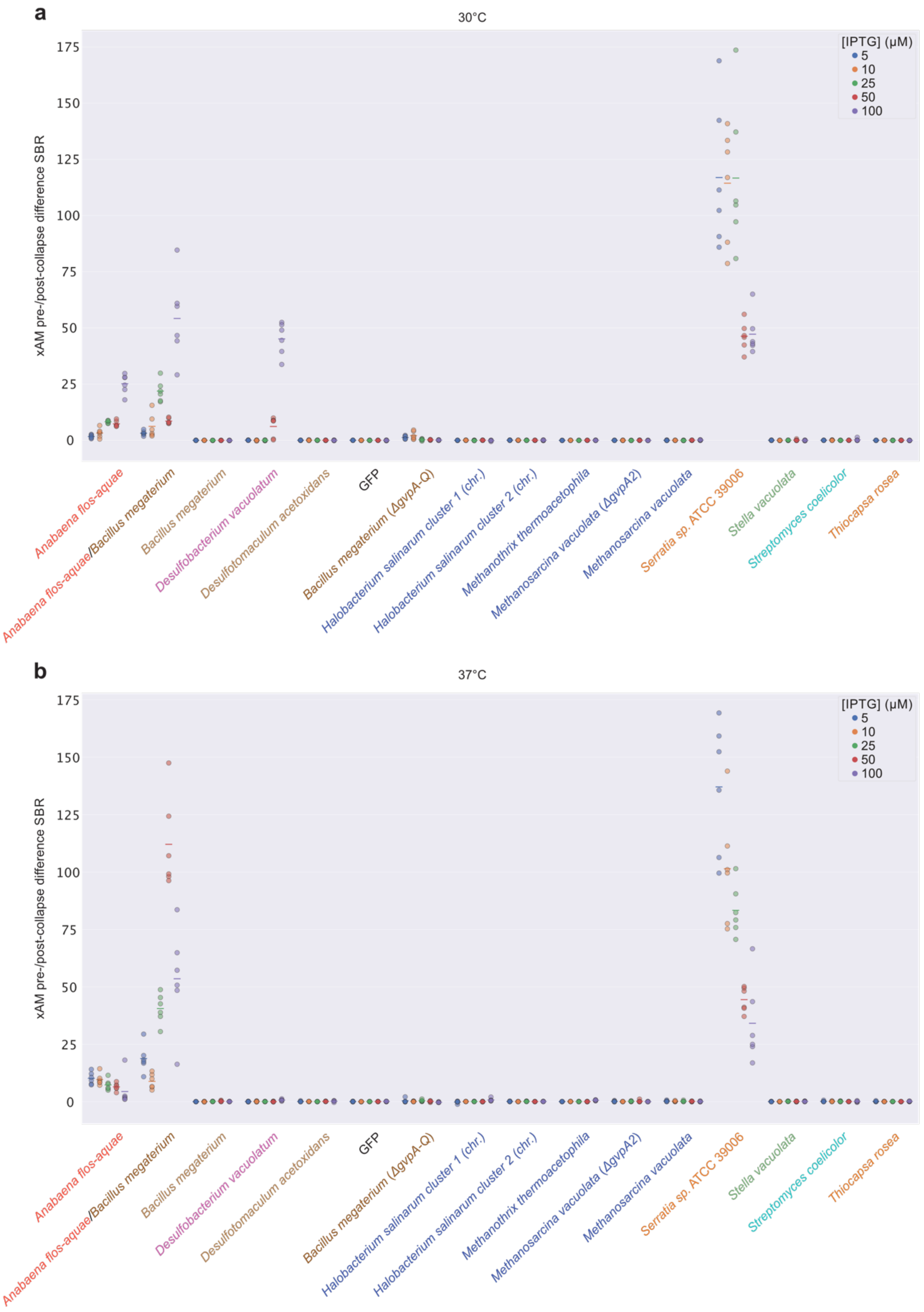


Figure 4-S5: Quantification of ultrasound signal for samples shown in Fig. 4-S4c-d. (**a-b**) xAM SBR for the patches at 30°C (a) and 37°C (b) shown in **Fig. 4-S4c-d** (n=6; lines represent the mean).

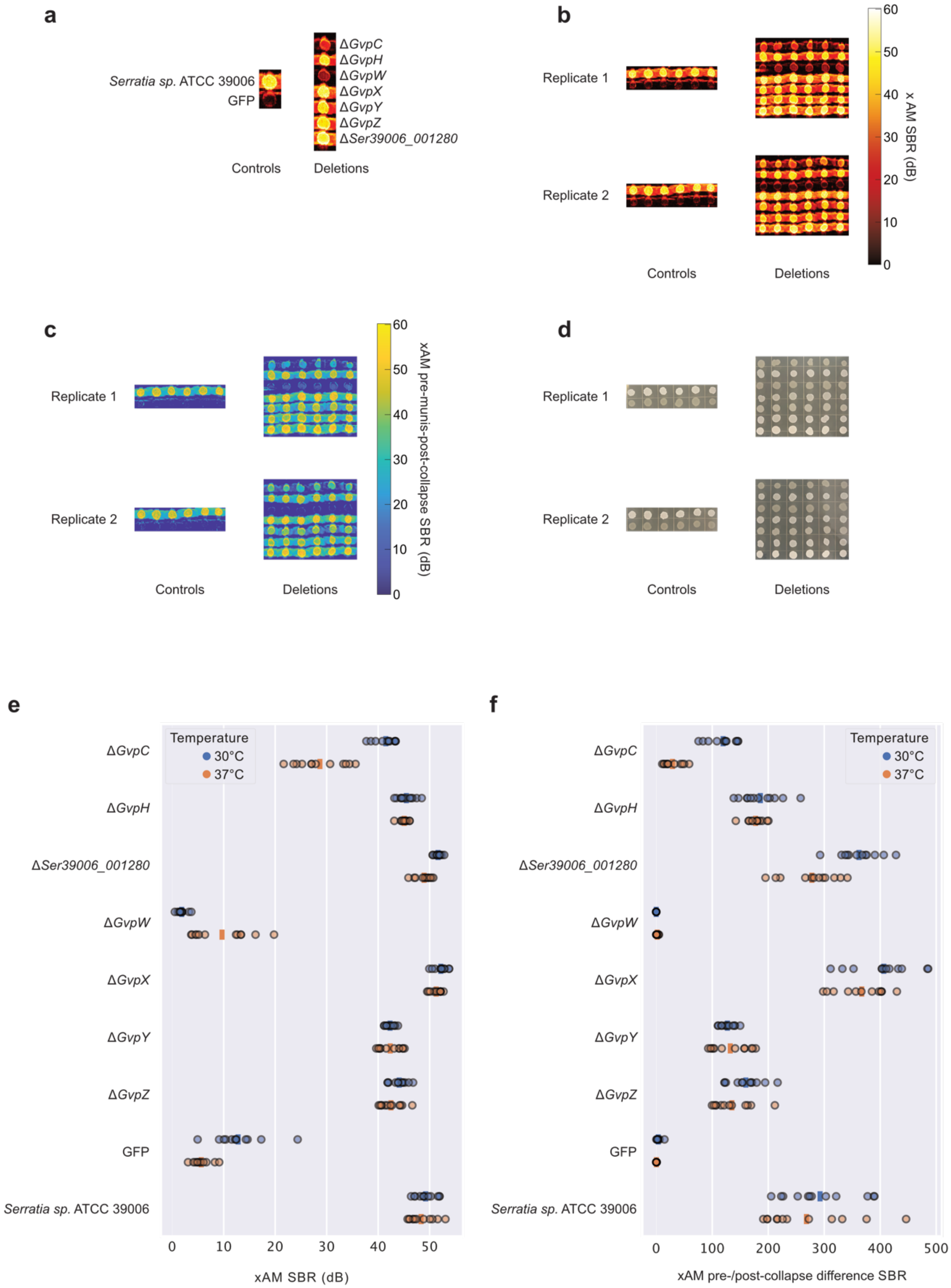


Figure 4-S6: Effects of single-gene deletions on GV expression by the *Serratia* cluster. (**a**) Key for genotypes tested, repeated in 6 replicate columns on each plate. (**b-d**) xAM images (b), pre-minus-post-collapse xAM images (c), and optical images (d) of bacterial patches expressing single-gene deletions of the Serratia cluster. (**e-f**) Quantification of the xAM images (e) and pre-minus-post-collapse xAM images (f) shown in (b-c) (n=12 biologically independent samples examined in one experiment).

a

			Non-white mutants / total colonies screened			
Condition	**Replicate**	**Starter**	**Day 1**	**Day 2**	**Day 3**	**Day 5**
+chlor	#1	0/226	0/68	0/69	0/60	0/40
+chlor	#2		0/55	0/47	0/355	0/54
+chlor	#3		0/48	0/39	0/98	0/42
+chlor	#4		0/56	0/21	0/49	0/233
-chlor	#1		0/39	0/54	0/70	0/64
-chlor	#2		0/28	0/46	0/65	0/46
-chlor	#3		0/47	0/57	0/75	0/52
-chlor	#4		0/36	0/57	0/57	0/38
-chlor +ara	#1		0/2	0/68	0/53	0/44
-chlor +ara	#2		0/2	0/62	2/86	0/68
-chlor +ara	#3		ND*	0/67	3/102	6/70
-chlor +ara	#4		0/1	0/76	0/34	0/56

*ND: no data because there were no colonies

b

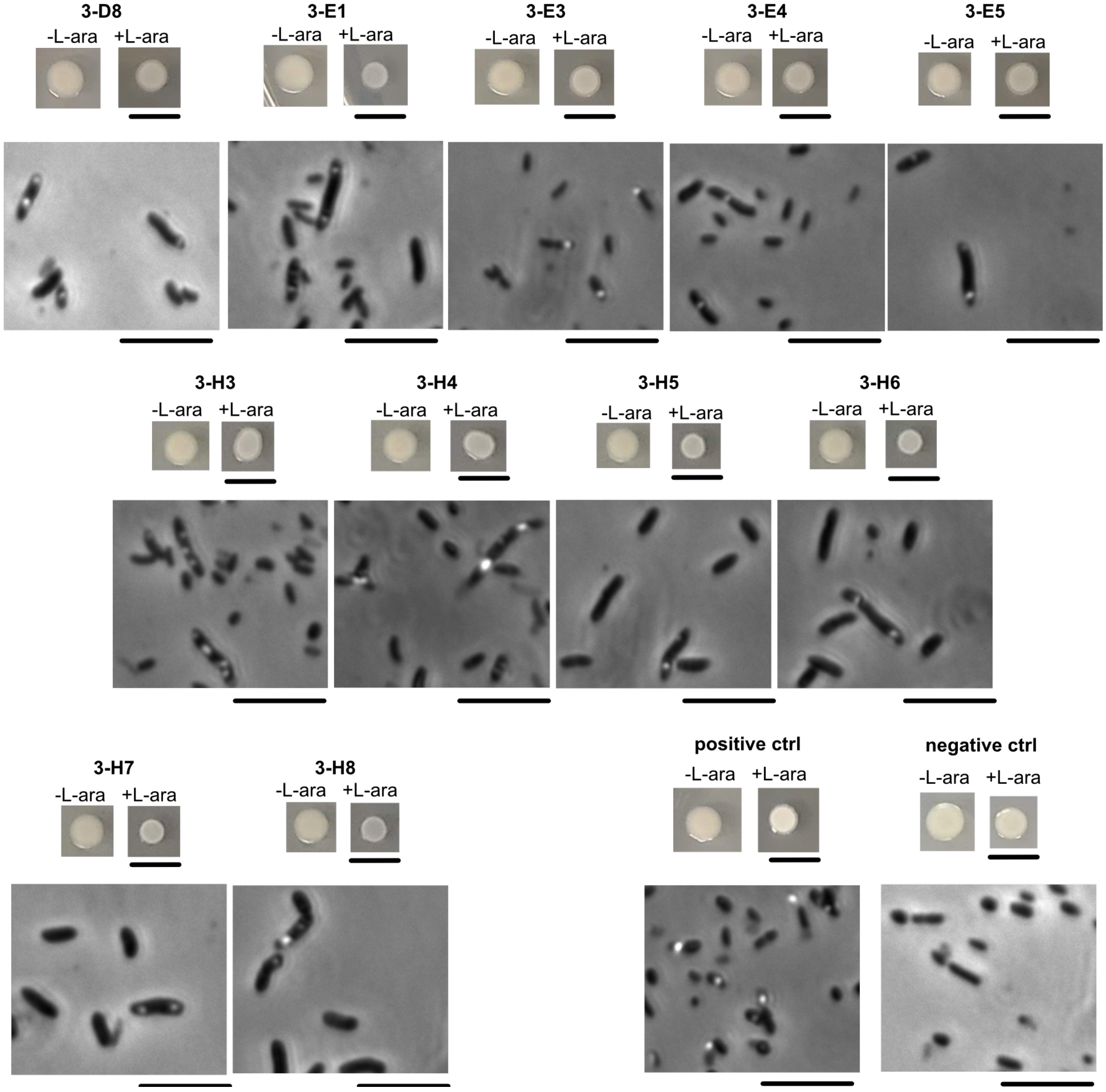


Figure 4-S7: Quantification and characterization of EcN mutants deficient in bARG$_{\text{Ser}}$ expression isolated from daily subculturing *in vitro*. (**a**) Numbers of non-white mutant colonies and total colonies screened on plates with 0.1% (w/v) L-arabinose from daily sub-culturing into LB media with 25 μg/mL chloramphenicol (+chlor), without chloramphenicol (-chlor), or without chloramphenicol and with 0.1% (w/v) L-arabinose (-chlor +L-ara) using pBAD-bARG$_{\text{Ser}}$-AxeTxe EcN. Cultures where mutants were found are indicated in red. (**b**) Optical images of patches (top rows) on fresh plates with 0.1% (w/v) L-arabinose (+L-ara) and without L-arabinose (-L-ara), and phase contrast microscopy images (bottom rows) from the 11 mutant colonies in (a). Mutants 3-D3 and 3-E1 were from the culture -chlor +ara, replicate #2, day 3; mutants 3-E3, 3-E4, and 3-E5 were from the culture -chlor +ara, replicate #3, day 3; and mutants 3-H3 through 3-H8 were from the culture -chlor +ara, replicate #3, day 5. The positive and negative controls were wild-type pBAD-bARG$_{\text{Ser}}$-AxeTxe EcN and pBAD-FP-AxeTxe EcN, respectively. Scale bars are 1 cm for images of patches and 10 μm for microscopy images.

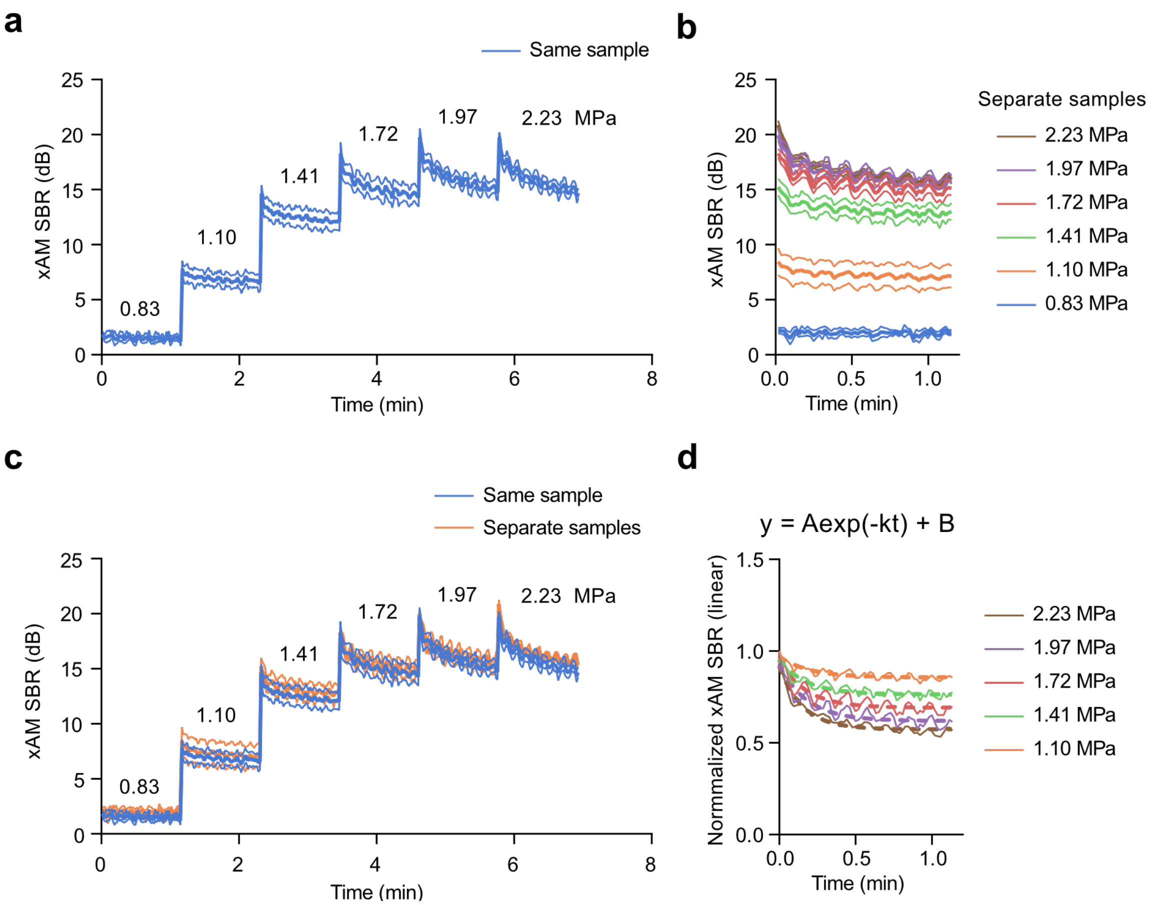


Figure 4-S8: xAM ultrasound signal versus time at varying acoustic pressures applied sequentially to the same sample versus separate samples. (**a-b**) xAM SBR of bARG$_{\text{Ser}}$-expressing EcN measured over time at various acoustic pressures. In (a), samples were subjected sequentially to 6 increasing acoustic pressures for approximately 70 sec each, whereas in (b) separate samples subjected to only one acoustic pressure for approximately 70 sec. (**c**) Overlay of xAM SBR curves for separate samples from (b) onto the curves for samples subjected to all pressures from (a). There is no difference between these curves, indicating that the xAM SBR measured at a certain pressure was not significantly affected by collapse at a previously applied pressure. For (a-c), pBAD-bARG$_{\text{Ser}}$-AxeTxe EcN were induced with 0.1% (w/v) L-arabinose for 24 hours at 37^{o}C in liquid culture, and were then normalized to 10^9 cells/mL in agarose phantoms for ultrasound imaging. Bold lines represent the mean and thin lines represent ± standard deviation; n=3 biological replicates, each with 2 technical replicates. Imaging was performed with an L22-v14X transducer, so the values for pressure and xAM SBR do not exactly match those in **Fig. 4-2d-f** where an L22-v14 transducer was used. (d) Normalized and linear-scaled data from (b) shown by thin lines which were fit to the exponential decay function y = Aexp(-kt) + B shown by thick dotted lines. See **Supplementary Note 2** for the fitted parameter values.

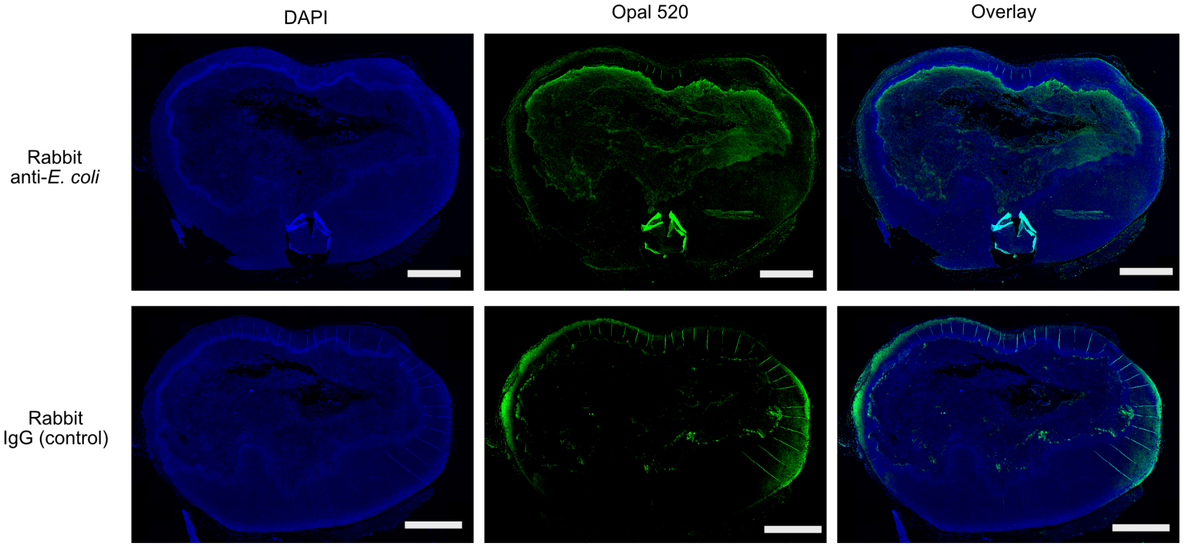
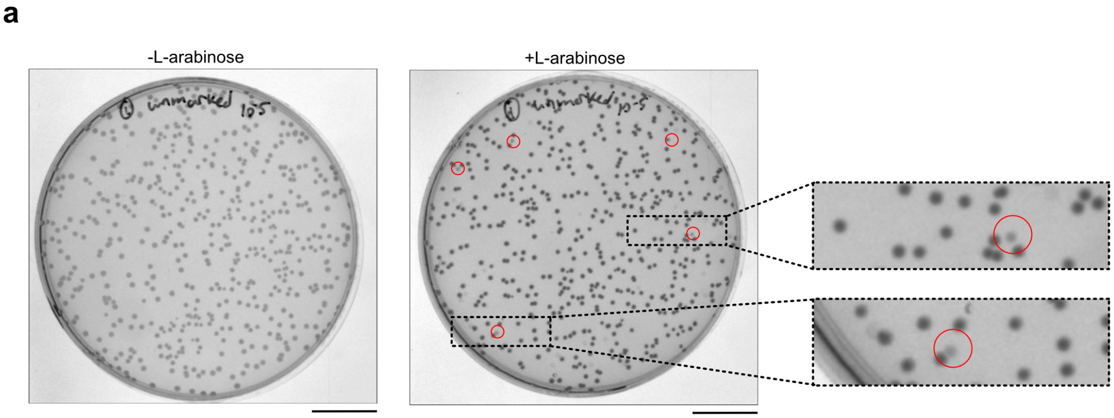


Figure 4-S9: Histology of MC26 tumor colonized with bARG$_{\text{Ser}}$-expressing EcN. Fluorescent images of tissue sections after ultrasound imaging on day 19 (see **Fig. 4-3a**). Sections were incubated with either polyclonal rabbit anti-*E. coli* antibodies (top row) or non-reactive rabbit IgG isotype control antibody (bottom row) as a negative control. All sections were then incubated with an Opal 520 polymer anti-rabbit HRP antibody (Akoya biosciences) and counterstained with DAPI. The EcN are visible in the necrotic core in the Opal 520 channel (top middle panel); the edges of the tissue exhibit a high degree of background staining (bottom middle panel). Histology was not repeated across multiple tumors because the other tumor tissues were instead homogenized and analyzed for colony forming units.



b

Mouse	Condition	Non-white mutants / total colonies screened
#1	Induced	0/687
#2	Induced	2/807
#3	Induced	0/341
#4	Induced	0/478
#5	Induced	5/585
#6	Uninduced	0/660
#7	Uninduced	0/567
#8	Uninduced	0/553
#9	Uninduced	0/556
#10	Uninduced	0/863
Total		7/6097 ~ 0.1%

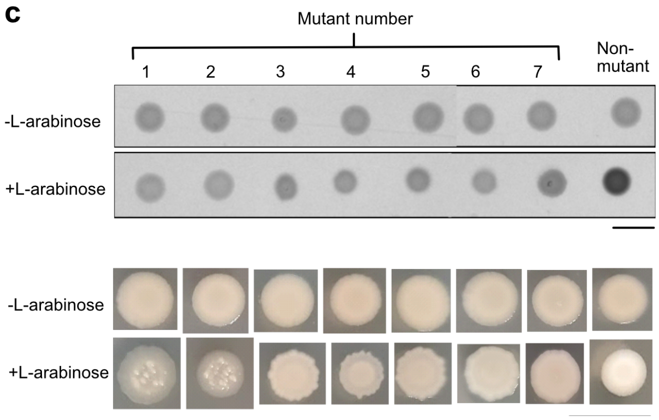


Figure 4-S10: Screening for EcN mutants defective in bARG$_{Ser}$ expression isolated from colonized tumors. (**a**) White light transmission images of plates with 0.1% (w/v) L-arabinose and without L-arabinose from plating a tumor (from mouse #5 in (b)) colonized by bARG$_{Ser}$-expressing EcN. Mutant colonies on the +L-arabinose plate appear lighter (more translucent) than wild-type opaque colonies and are indicated by red circles. (**b**) Numbers of non-white mutant colonies and total colonies screened on plates with 0.1% (w/v) L-arabinose for the ten mice injected with pBAD-bARG$_{Ser}$-AxeTxe EcN. **c**, White light transmission images (top) and photographs (bottom) of patches on fresh plates with 0.1% (w/v) L-arabinose and without L-arabinose made from the seven translucent mutant colonies in red in (b) and an opaque non-mutant colony as a control. Mutants 1-2 were from mouse #2 and mutants 3-7 were from mouse #5. Scale bars are 2 cm in (a) and 1 cm in (c).

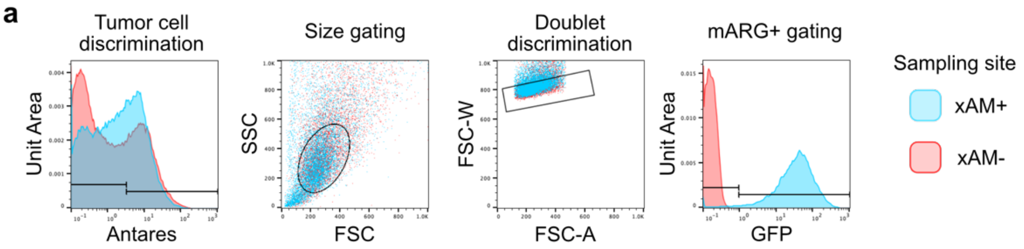


Figure 4-S11: Flow cytometric gating strategy for chimeric tumor biopsy sample analyses. Events were first gated based on Antares expression to exclude endogenous mouse cells. Antares-positive cells were then gated by size. Single cells were gated based on FSC-W vs FSC-A plot. GFP-positive (mARG$_{Ana}$-positive) cells were gated from the resulting histograms.

Chapter 5

CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, I report the development of a novel high-throughput pipeline for ARG engineering and describe its application to engineering a variety of ARG phenotypes. The resulting second-generation ARG clusters enabled the imaging of *in situ* gene expression in bacterial and mammalian cells in mice, and nondestructive multiplexed imaging of bacterial cells *in vitro*.

In this chapter, I compile a set of insights gleaned from these ARG engineering projects, and suggest some projects that could expand on their results.

5.1 Lessons Learned from Half a Decade of ARG Engineering

The main thing that I have learned from the past five years of ARG engineering projects is that although ARGs are considered the “GFP of US,” they are substantially harder to engineer than GFP. There are several reasons for this, but many stem from the fact that while GFP is a single highly soluble protein, GVs are assemblies of tens of thousands of copies of primarily a single protein and require a number of chaperones to coordinate their assembly, with incorrect assembly potentially resulting in cellular toxicity. Further, no reliable structural information was available for GvpA when these projects were initiated, and there was little infrastructure available for acoustic screening of GV phenotypes.

In the course of screening libraries of random GvpA mutants, I quickly realized that GvpA is a very challenging target for protein engineering, both because it is extremely sensitive to mutation (with <1% of random mutants producing significant amounts of GVs), and because engineering a single protein in a cluster of many whose functions depend on each other is likely to have unpredictable consequences. Further, there is no intermediate phenotype on the way to GV assembly that can be reliably assayed, and so any functional manipulations to the cluster that cause it to fall short of complete GV assembly do not yield actionable information to guide further engineering.

There are two limits on the amount of nonlinear US signal that GVs can be engineered to produce. The first, a hard limit, is based on the number of GVs that fit inside the cytoplasm of a cell, considering their relatively large size. The second, a soft limit, is based on the general trend observed in these and other GV engineering efforts that increasing nonlinear US scattering tends to be accompanied by decreasing collapse pressure. Thus, the environmental pressure GVs experience, set for example by the turgor pressure of the cell, limits the extent to which nonlinear scattering can be engineered while this relationship holds true. Decoupling this relationship would presumably require engineering the material properties of the GV shell to make it more flexible and prone to buckling in response to applied pressure without making it more brittle and prone to rupture. As this will likely be a rare phenotype that is difficult to engineer, machine learning methods and structure-guided mutagenesis may prove especially helpful for extracting relevant information from larger genotype-phenotype datasets and for designing focused libraries enriched in desirable mutants, respectively.

The methods developed and engineering approaches explored in this work set a path for future ARG engineering efforts. Constructing the Acoustic Plate Reader and using it to screen mutant libraries revealed that the throughput of 10^2-10^3 variants assayed per round of screening is sufficient to identify enhanced acoustic phenotypes, with the critical caveat that the mutants assayed are pre-screened for successful GV production through an opacity assessment in colony format. Screening homologs of GV clusters for successful expression in *E. coli* revealed the set of clusters that are useful as ARGs, establishing the most viable candidates for future engineering. Screening libraries of random mutants of GV structural proteins revealed that the acoustic and expression phenotypes of ARGs can be successfully engineered; with respect to the two limits mentioned above, while protein engineering alone is likely sufficient to maximize the amount of nonlinear scattering per GV, it is almost certainly not enough to maximize the number of GVs produced per cell, which will require tuning of promoters, RBSs, ORIs, and expression timing.

5.2 Future Directions for ARG Engineering

There are many exciting future directions for ARG engineering; in the next two sub-sections, I will detail a handful of them. I separate these potential projects based whether they can be attempted using the methods developed in this study (*i.e.*, they involve screening libraries of ~10^3 variants or fewer per experiment), or they will require the development of new methods (*i.e.*, higher-throughput screening or a novel selection strategy).

5.2A ARG Engineering Projects Attemptable Using Current Methods

Perhaps the most obvious follow-up to the work in this thesis is to screen homologs of GV gene clusters closely related to those that produced US signal when expressed in *E. coli* (Fig. 4-1d), and to test whether any of them have superior performance to the ones selected for use as second-generation ARGs; these clusters have a high likelihood of producing US contrast in *E. coli*, and a moderate likelihood of showing improved phenotypes. An extension of this approach is to swap in individual Gvp homologs from other clusters into the chosen second-generation ARG clusters to create “mosaic” clusters containing Gvps from multiple species; this approach was successful when combining structural proteins from *Anabaena flos-aquae* and assembly factors from *Bacillus megaterium*,[1] though further attempts to combine structural proteins from other species with *B. megaterium* assembly factors yielded only one combination with significant US contrast (data not shown).

Another important direction that will be fairly straightforward to implement using recently developed techniques is directed evolution of Gvps in mammalian cells. By combining novel methods for site-specific single-copy chromosomal integration of transgenes in mammalian cells[2] with monoclonal cell line generation achievable through FACS and with acoustic screening with the Acoustic Plate Reader, I estimate that a library of 1000 Gvp mutants could be screened in a few weeks. This approach would allow mutant ARGs to be screened directly in the cells in which they will be used, which is highly preferred to screening them in *E. coli* and testing only a small number of high-performing mutants in mammalian cells.

Relatedly, because most *in vivo* applications of bacterial therapeutics use chromosomal integration of transgenes to improve construct stability, it is sensible to attempt this for ARG clusters; fortunately, there is an increasing number of systems for the fast chromosomal integration of plasmid-borne cargo in bacteria. The largest challenge will be transitioning from a medium-copy expression plasmid (~40 copies/cell in the case of existing ARGs) to only 1-3 copies/cell for chromosomally integrated transgenes; this will certainly require promoter and RBS tuning, and may also require protein engineering.

Another straightforward project would be to improve ARG expression in bacteria from clusters lacking GvpC, as these ARGs are known to have more nonlinear signal than those with GvpC bound—this could be done either by using protein engineering techniques to modify GvpA, or by tuning the expression elements on the plasmid. The main challenge that will need to be overcome is that the turgor pressure of *E. coli* is higher than the collapse pressure of many GVs that lack GvpC, so these experiments may need to be performed in cells lacking turgor pressure such as cell wall-less *E. coli*[3] if the selected mutants are to be used in mammalian cells.

Our evolution pipeline could also be applied to the evolution of GvpC or functionalized derivatives of it, such as sensors of pH, enzymatic activity, or Ca^{2+}. In this case, it may be beneficial to use a GV type that cannot withstand the turgor pressure of the *E. coli* cytoplasm when GvpC is not bound to it, and to use a colony opacity pre-screen to eliminate nonfunctional GvpC mutants that do not bind strongly enough to the GV shell to increase its collapse pressure.

In the directed evolution campaign detailed in Chapter 3, we only obtained genotype information for a small number of the top-performing ARG mutants from each round of screening. However, new methods like evSeq[4] enable the fast and inexpensive genotyping of arrayed mutant libraries, and this information could be used in combination with novel methods for machine learning-guided directed evolution[5] to quickly evolve rare phenotypes—in particular, independently evolving nonlinear scattering and collapse pressure, which would be tremendously valuable when searching for variants that can be nondestructively multiplexed.

Finally, independent of ARG engineering, it will be important to conduct basic biochemical studies to better understand the mechanics of GV assembly and the roles of all Gvps in this process. The most challenging aspect of these studies is that GvpA cannot be purified—as one of the most hydrophobic known proteins, it is toxic to cells when expressed by itself or without enough of its chaperones, presumably because of aggregation. Therefore, developing the ability to express GVs in a cell-free transcription/translation system would facilitate studies in which the ratios of Gvps are varied and either US contrast or GV structure are assayed. In a similar vein, existing techniques like analytical ultracentrifugation[6] (available at Caltech) have yet to be applied to the study of Gvps, and can give important information about binding affinities and stoichiometries of protein-protein interactions.

5.2B ARG Engineering Projects Requiring Novel Methods

Just as the development of the Acoustic Plate Reader enabled the screening of libraries with thousands of members, which would have been nearly impossible using the existing low-throughput approach, further enhancements in ARG screening methods will facilitate even more extensive engineering efforts. In particular, ultra-high-throughput strategies such as selection by buoyancy separation or screening by cell sorting (either optical or acoustic) will enable experiments that are unfeasible using the Acoustic Plate Reader due to its throughput limit of ~10^3 variants per experiment. In this section, I describe engineering efforts that could be attempted using these approaches.

Beyond the homolog and scanning site saturation libraries constructed in this thesis, there are several other types of libraries that could be constructed for ARG clusters or individual Gvps if a sufficiently high-throughput screen or selection were available to test them.

First, error-prone PCR libraries, which tend to be large and contain a high percentage of undesirable variants, but also can contain tremendous diversity, could be generated for either GvpA or for an entire ARG cluster, and either screened for side-scattering via FACS or for US signal via acoustic sorting, or selected using buoyancy purification. In these cases, the phenotype being screened or selected for would be the total gas fraction of the cell (a combination of volume per GV and number of GVs per cell), and could give different information depending on the specifics of the experiment: for example, the library could be screened after multiple rounds of expression or with the cluster under the control of a leaky promoter to identify mutants with improved tolerability; it could be screened under high expression temperatures to identify mutants with improved thermostability or solubility; it could be screened as a neutral drift library to increase the evolvability of the starting point before subsequent evolution.[7] As an extension of this approach, continuous evolution methods such as EvolvR[8] could be used to generate diversity and these libraries could be selected using buoyancy separation before cells are passaged each time.

To take full advantage of this increased screening depth, methods such as deep mutational scanning[9] could be used to extract more information from these libraries than simply the genotypes of selected mutants: through a combination of mutant barcoding and deep sequencing, fitness landscapes can be constructed either densely for individual proteins or sparsely for multiple proteins in a cluster. For example, by separately barcoding buoyant and non-buoyant cell populations, mutants could be ranked based on their likelihood of conferring buoyancy to a cell expressing them, which would give information about the percentage of GV-expressing cells in the sub-population of cells that express that mutant (*i.e.*, whether all cells are expressing roughly the same amount of GVs, or if a small percentage of cells express a large number of GVs but most of the population expresses no GVs). Conversely, a sparse fitness landscape could be constructed for all the proteins in a cluster if an epPCR library were subjected to this protocol and the beneficial and deleterious mutations grouped to identify “hotspots” for targeted mutagenesis.

Beyond these large libraries of random mutants, combinatorial libraries can be created using MoClo, either of defined parts with characterized functions,[10,11] or of mutations in defined regions of one or more proteins:[12] the former could be used either to optimize expression of an ARG cluster by tuning the overall expression of the transcription unit (with a promoter library) or the ratios of individual Gvps in the cluster (with an RBS library for each gene); the latter could be used to generate either random or recombination libraries for one or more proteins in the cluster in a fast, combinatorial fashion. An ambitious extension of the former approach could involve placing each Gvp in a cluster under the control of a separate orthogonally inducible promoter and tuning either the ratios or relative timings of expression induction for all Gvps simply by varying the concentrations and delivery schedules of the inducers (see Fig. 4 of Meyer *et al.*[13] for an example of this approach). This approach would enable the fast and flexible tuning of the expression levels of all Gvps in a cluster, which will be useful when transferring a given cluster between two vastly different expression contexts, such as between two distantly related species of bacteria or from a plasmid to the chromosome.

Given the utility of GvpC as a functional handle for GVs, it would be useful to engineer orthogonal GvpA/GvpC pairs that bind specifically to each other with minimal cross-reactivity. This could be achieved either through screening homologs of both proteins against each other in purified form, or through mutagenesis of one or both binding partners, possibly using a surface display method for GvpC.[14] Finally, a more ambitious project could involve adding a completely new functionality to GvpA or GvpC: for example, engineering a "split-GvpA" or "split-GvpC" analogous to split-GFP[15] that only dimerizes and induces GV formation or a change nonlinear signal in response to a protein-protein binding event, or even in response to a small molecule.

5.3 References

1. Bourdeau, R. W. *et al.* Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts. *Nature* **553**, 86–90 (2018).
2. Xiong, K. *et al.* An optimized genome-wide, virus-free CRISPR screen for mammalian cells. *Cell Rep. Methods* **1**, 100062 (2021).
3. Errington, J. Cell wall-deficient, L-form bacteria in the 21st century: a personal perspective. *Biochem. Soc. Trans.* **45**, 287–295 (2017).
4. Wittmann, B. J., Johnston, K. E., Almhjell, P. J. & Arnold, F. H. evSeq: Cost-Effective Amplicon Sequencing of Every Variant in a Protein Library. *ACS Synth. Biol.* **11**, 1313–1324 (2022).
5. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
6. Edwards, G. B., Muthurajan, U. M., Bowerman, S. & Luger, K. Analytical Ultracentrifugation (AUC): An Overview of the Application of Fluorescence and Absorbance AUC to the Study of Biological Macromolecules. *Curr. Protoc. Mol. Biol.* **133**, (2020).
7. Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* **5**, 939–942 (2008).
8. Halperin, S. O. *et al.* CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252 (2018).
9. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
10. Moore, S. J. *et al.* EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology. *ACS Synth. Biol.* **5**, 1059–1069 (2016).

11. Iverson, S. V., Haddock, T. L., Beal, J. & Densmore, D. M. CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.* **5**, 99–103 (2016).
12. Püllmann, P. *et al.* Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep.* **9**, 10932 (2019).
13. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.* **15**, 196–204 (2019).
14. Naftaly, S. *et al.* Mapping protein selectivity landscapes using multi-target selective screening and next-generation sequencing of combinatorial libraries. *Nat. Commun.* **9**, 3935 (2018).
15. Romei, M. G. & Boxer, S. G. Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annu. Rev. Biophys.* **48**, 19–44 (2019).