A Nucleic Acid-based Bacterial Message Export
System for Cell-to-cell Communication

by

Lisa Nip

B.A. Biochemistry and Molecular Biology
M.A. Biotechnology
Boston University, 2012

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning
in Partial Fulfillment of the Requirements for the Degree

of

Master of Science

at the

Massachusetts Institute of Technology

February 2016

© 2016 Massachusetts Institute of Technology. All rights reserved.

Signature redacted

Signature of Author...

Signature redacted

Program in Media Arts and Sciences
MIT Media Lab
January 1, 2016

Signature redacted

Certified by.........

Signature redacted

Joseph M. Jacobson
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by........

Professor Pattie Maes
Academic Head
Program in Media Arts and Sciences
A Nucleic Acid-based Bacterial Message Export
System for Cell-to-cell Communication

by

Lisa Nip

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning
on January 1st, 2016 in Partial Fulfillment of the
Requirements for the Degree of Master of Science
at the Massachusetts Institute of Technology

ABSTRACT

Communication within natural systems of eukaryotes and prokaryotes typically entails
message transmission between and among cells via small-molecule messengers being
funneled from the sender to the receiver cell. Nucleic acids are rarely used as extracellular
messengers due to their labile nature and proclivity for enzymatic digestion. Eliminating
these obstacles will allow for a larger array of messages to be sent with minimal cellular
machinery. Exploiting the bacterial twin-arginine translocation (TAT) pathway and a
nucleic-acid binding protein sourced from bacteriophage MS2, we have engineered a
message-sending system in *Escherichia coli* capable of specifically exporting a "pre-written"
circularized RNA message to the extracellular environment. This RNA message maintains
its integrity over the course of at least four hours in extracellular growth medium, and this
system serves as the first demonstration of versatile, stable messaging with nucleic acids,
specifically with RNA, in the extracellular environment.

Thesis Supervisor: Joseph Jacobson
Title: Associate Professor of Media Arts and Sciences
A Nucleic Acid-based Bacterial Message Export
System for Cell-to-cell Communication

by

Lisa Nip

The following served as readers for this work

Signature redacted

Certified by

Signature redacted

Ed Boyden
Professor of Media Arts and Sciences
Thesis Supervisor

Shuguang Zhang
MIT Media Lab
Thesis Supervisor
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>6</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>14</td>
</tr>
<tr>
<td>Results</td>
<td>19</td>
</tr>
<tr>
<td>Discussion</td>
<td>23</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>27</td>
</tr>
<tr>
<td>Appendix/Supplement</td>
<td>28</td>
</tr>
<tr>
<td>References</td>
<td>36</td>
</tr>
</tbody>
</table>
INTRODUCTION
Extracellular communication
Existing extracellular communication between any types of cells (either eukaryotic or bacterial) is executed primarily via small molecules, or chemical signaling, such as in quorum sensing. In multicellular organisms, messengers include proteins, peptides, individual amino acids, nucleotides, and various lipids. The transmission of these signaling molecules (i.e., the “messages”) from their sender to the receiver requires protein exporters, carriers, and receptors that are highly specific to each of the chemical signaling molecules and cell system, and are therefore limited in the range of “messages” permitted for extracellular communication. Furthermore, the message must be able to sustain its integrity on its journey from the sender cell to the receiver, and as such, most messages are typically small molecules. The cells, even given these highly specific receptors, must also have the capacity to act upon the signal, should they be able to detect the messenger molecule among the sea of other molecules with which they are constantly bombarded. The reception of messages then results in a change in the behavior, or protein expression profile of a cell (i.e., they trigger genetic modifications). The extracellular communication scheme can be summed as the sending of a message that remains stable between the senders and receivers, and the matching of these small-molecule messages to target receptors to achieve a desired reaction.

Cellular communication networks typically consist of (A) a single molecule (i.e., the message) sent by a molecule-specific “transmitter” that has a molecule-specific “receiver,” (B) one message transmitter for sending one message to multiple receivers, or (C) multiple receivers each coupled to distinct messages that are sent by multiple transmitters. While such cellular communication arrangements were evolved to minimize crosstalk and extraneous noise, these analog methods require the building and maintenance of hundreds, if not thousands, of different sender and receiver proteins (Fig. 1 a-c).

![Figure 1. Analog cellular communication schemes require many types of messages, senders, and receivers.](image)
Cells may communicate using a single transmitter sending a single specific message to a specific receptor (a), or transmit a single message that can be interpreted by multiple receivers (b). Systems that attempt to minimize crosstalk will employ multiple transmitters coupled to multiple specific receivers for multiple specific messages (c). Recently, work has been done to decouple the messenger from the receiver, but requires significant protein encapsulation for protection (d). This figure is adapted from Ortiz et al. (2012). The potential of engineering cells to produce nucleic acid messages allows for novel ways of communicating with the body. For example, engineered bacteria could be ingested and send nucleic acid messages upon receipt of a signal in the gut, and such messages could be absorbed by the gut as recently described (e1). As the nucleic acid messages travel through the circulatory system, it is possible for detection by an implanted microfluidic chip (e2) containing short hairpins that fluoresce upon binding to the message (e3).

An attempt at improving these native systems has been accomplished by decoupling the receiver channel from any one message has been engineered; however this technique requires extra packaging with viral machinery from bacteriophage M13 in order to protect the message (Fig. 1d). Considering the importance of cell-cell communication, a system that relies on a minimal amount of cellular equipment or cell death could significantly benefit both in vivo and in vitro experimentation by minimizing the amount of energy allocated to producing such message-specific proteins. Furthermore, a cell-cell communication framework that allows for the free diffusion of messenger molecules from sender to receiver would greatly accelerate the speed of information exchange between and among cells. This work hopes to achieve a communication system as depicted in Fig. 1d without the extensive packaging.

**Nucleic acids as messengers**

Nucleic acids, though flexible in the messages carried (i.e., it is a single molecule that can have any sequence), are seldom used as an extracellular messenger for the reason that they are rapidly degraded both within and outside the cell. On the rare occasion that nucleic acids are used to transfer information “outside” the cell, pseudo-intracellular environments are created between sender and recipient. This type of exception occurs during times of high stress, such as when a cell is under viral attack or living in undesirable environmental conditions. In the former, viruses employ either DNA or RNA to hijack a recipient cell’s machinery via directly connecting to proteinaceous surface protrusions (e.g., F-pilus) on the recipient cell and misinforming the cell that it should divert its resources to generating viral products. In the latter, a stressor like heat shock would trigger an SOS response that involves a conjugation event where DNA is transported from a donor cell to a recipient cell via a temporary bridge made from the pilus of the donor cell, the overall mechanism of which yields horizontal gene transfer. In both cases, a physical, direct connection is required to transmit the message, which is akin to using a paper-cup walkie-talkie. Moreover, a viral-based message delivery system, while naturally efficient at commandeering host machinery, would be detrimental to the livelihood of the receiver cell.

The use of nucleic acids as the extracellular messenger have benefits that are two-fold: first, the message can be designed to have optimal expression in the host cell (i.e., pre-codon optimized) to minimize incompatibility during transcription or translation in the host, and second, a nucleic acid message can be highly programmable, assuming that the recipient cell has an innate process (e.g., an engineered or native protein receptor, engulfment) for receiving nucleic acid messages. These benefits lie in contrast to the aforementioned analog
chemical-messaging system, whose message transmission relies on the brute-force accumulation of a single molecule in the receptor cell. Still, nucleic acids as messengers have their own limitations to overcome both intra- and extracellularly. RNA turnover within the cell is evolutionarily made to be rapid for the reason that defective nucleic acid molecules could potentially introduce unwanted toxicity from mutations. Unmodified linear RNA in particular is known to be especially susceptible to degradation in the extracellular environment due to exogenous nuclease. Furthermore, evolution has brought forth enzymes that degrade extracellular nucleic acids as a defense mechanism against infection by foreign agents, and using unmodified nucleic acids as messengers requires that one find a way to protect the message.

Borrowing machinery
In the process of designing a more efficient cell-to-cell communication system, the goals are threefold: the first is to create “universal” sender machinery where the messages were the sole components required to be unique. Second, because nucleic acids are synthesized in the cytoplasm of prokaryotes, the sender machinery carrying the message must first be able to traverse the inner membrane, where it must endure the hazards of the periplasmic milieu, then navigate across the cell wall, and finally through the outer membrane. Finally, once these cellular barriers are crossed, the message must also be able to maintain its integrity and withstand harsh unknowns in the extracellular environment.

Engineering RNA message stability through autocatalytic RNA circularization.
Naked, linear RNA is rapidly degraded by ribonucleases (RNases) within the cytoplasm and in the periplasm. In the extracellular environment, hazards of enzymatic and chemical natures are present and ready to destroy RNA, and a typical _E. coli_ strain also releases many nuclease into the extracellular growth medium. The modification of RNA such that it can no longer be recognized as a substrate for most RNases became the first target. Efforts have previously been made to create more durable RNA that can endure the environment of human serum and cell extracts, and these studies have provided a method to make circular RNA molecules, which are found to be highly resistant to human RNases sans chemical modifications. Even to-be-excised sequences with higher order structures could be circularized both in vitro and in vivo, and simultaneously maintain enzymatic function as an RNA aptamer. Circular RNAs exist in nature, but primarily as discarded products after exon splicing; that is, until the splicing mechanism was found to be malleable by simply reordering the recognition sequences.

The _tetrahymena_ group I intron self-splicing RNA was the first ribozyme discovered and this group I intron sequence is shared among eukaryotes, prokaryotes, and some bacteriophages. The self-splicing mechanism is protein- or enzyme-independent and is executed chemically and driven by magnesium ions and guanosine nucleotides, usually resulting in circular introns. The making of circular exons was found to be possible by Umekage et al., who permutated the group I intron self-splicing sequence such that, instead of excising an intron, one can splice out an “exon” of one’s choosing in what they termed the “Permuted Intron-Exon Method” (herein referred to as PIE). Briefly, the self-splicing mechanism entails a two-step transesterification resulting in an arrangement into a higher-order structure that subsequently self-splices the exon. A guanosine nucleotide first makes
an attack on the downstream intron's 3' phosphate (Fig. 2, red line), after which the region between the exon (Fig. 2, green line) and downstream intron is cut, and the guanosine nucleotide is bonded to the 5' side of the downstream intron. The exon then circularizes when the hydroxyl group of the 3' end of the exon makes a nucleophilic attack on its 3' phosphorus (Fig. 2, bottom). The PIE sequence described by Umekage et al. was used in this study to create a stable RNA message.10

**Figure 2.** Graphical illustration of PIE sequence. By flanking an “exon,” or any message that requires circularization, with PIE sequences (black and red) and supplementing the reaction with GTP and magnesium ions, the sequence can be autocatalytically excised to yield two split introns and a circularized exon.
A carrier for the RNA message

Specific export of any desired RNA messages is necessary in designing a system in which the sender equipment is decoupled from any one message; that is, we need the system only to recognize that a message is a message. For this, we turned to the bacteriophage MS2 coat protein. The MS2 phage coat protein binds with high specificity and affinity to the RNA hairpin, with a $K_d$ of 0.4 nM and consists of the following minimal consensus sequence: 18

$$\text{NNNNANNAN}[\text{Py}][\text{Nc}][\text{Nc}][\text{Nc}][\text{Nc}][\text{Nc}][\text{Nc}]$$ (Fig. 3a).

This protein has been used as a method of imaging the distribution of intracellular RNA in living cells by fusing the MS2 coat protein to GFP. 18-20

![Graphical representation of the components of bacteriophage MS2 coat protein. An RNA hairpin is made from a sequence where N is any nucleotide, Py is any pyrimidine, and Nc is the nucleotide complement to N (a). RNA hairpin recognition triggers dimerization such that two coat protein monomers (blue and green) bind a single hairpin (orange) (b, image reprinted from Wikimedia Commons, Dr. Neil Ranson, University of Leeds, UK). Assembly of 90 coat protein dimers (red and gray) form a viral capsid (c, reprinted from Protein Data Bank, PDB, ID Code 1AQ3).](image)

When the coat protein monomer binds to its preferred RNA hairpin, a conformational change triggers dimerization (Fig. 3b) and capsid formation is stimulated such that a near-spherical shape is formed (Fig. 3c). 21 RNA molecules are subsequently encapsulated within the viral shell. A total of 180 coat protein monomers, or 90 coat protein dimers are present in this highly stable capsid state, whose formation is facilitated by the “FG loop” encoded by the amino acid sequence VATQTVGGVELPVAA present in each coat protein monomer. 22 This site has been mutated in previous studies to eliminate the capacity for capsid formation, but other studies have found that the addition of a peptide sequence as short as two amino acids to either the N or C terminus was sufficient to significantly reduce coat protein polymerization. 22,23 In order to minimize capsid formation and be able to select our desired RNA message from the cytoplasmic milieu, then, the MS2 coat protein was designed with a signal peptide fused to the N terminus. The signal peptide was selected for the protein transportation described below.
Overcoming intracellular barriers with the Tat system and kil protein

*E. coli*’s cell envelope is a tri-layer structure composed of the cytoplasmic (inner) membrane, the outer membrane, and the periplasmic space that lies between them (Fig. 4). The inner membrane is a lipid bilayer with many proteins for active solute transport. The outer membrane is a permeable barrier made of an asymmetric bilayer of lipopolysaccharides (LPS) and phospholipids that prevents periplasmic protein leakage, but some general diffusion pores allow for nutrient uptake and waste disposal. Covalently attached to the outer membrane is the periplasmic peptidoglycan layer that bestows the cell’s shape and rigidity.

Figure 4. Graphic depiction of the many layers of the *E. coli* cell envelope. The inner membrane of the envelope is made of a phospholipid bilayer in which various inner membrane proteins are present. The outer membrane is consists of lipopolysaccharides as well as a phospholipid monolayer, and like the inner membrane, contains membrane proteins. The inner and outer membranes are separated by a periplasmic space in which peptidoglycan (red) gives the cell its shape.

In prokaryotes, the twin-arginine translocation (Tat) system is dedicated to the specific transport of folded proteins across the cytoplasmic (inner) membrane and homologs are found across archaea, bacteria, and chloroplasts. As long as the folded protein contains a covalently attached N-terminal twin arginine signal peptide (SRRXFLK), even proteins containing co-factors are also transported across the inner bacterial lipid bilayer by means of the membrane-bound Tat transport apparatus. It remains unclear how the Tat machinery makes the distinction between folded and unfolded proteins, but its specific selectivity for folded proteins is an advantage in that it ensures that proteases in the periplasm are less likely to digest the Tat-transported protein cargo.

The Tat pathway consists of four genes, *tatA*, *tatB*, *tatC*, and *tatE*, that each encodes a critical inner membrane protein in *E. coli* where the TatABC complex acts as the twin-arginine signal peptide interaction site. While the composition of the Tat proteins differs across species, Tat proteins TatA, TatB, and TatC are minimally required for a successful Tat system in *E. coli*. 
TatC serves as the docking site for Tat signal peptides and is typically in complex with TatB. It has been hypothesized that the TatA protein of the TatABC complex is able to oligomerize and associate into structures of various sizes and shape before being “selected” by the TatBC complex to accommodate a substrate of a specific size, suggesting that substrates of different physical dimensions may be tolerated. A model of the Tat system posits that Tat proteins remain separate complexes during the resting state until a substrate binds, during which the proteins TatBC and TatA rearrange into a translocation pore (Fig. 5, steps 1 and 2). This multistep construction of the pore may be to keep unused space to a minimum and accommodate cargo of unpredictable shapes and sizes. The protein substrate is subsequently transported through the TatA-dense pore via a proton motive force (i.e., no ATP is directly expended), and the mature protein is released into the periplasm after the signal peptide is cleaved (Fig. 5, step 3). It is worth noting, however, that the calculated energy expenditure of Tat-dependent translocation averages to $7.9 \times 10^4$ protons, which is equivalent to about 10,000 ATP molecules for each translocated protein.

Figure 5. Schematic of Tat system mechanism of targeting and transport. At its inactive state, TatB and TatC are associated as a complex and TatA is spread out as protomers. (1) TatC of the TatBC complex first recognizes and binds to the twin-arginine (RR) motif in the signal peptide of the protein substrate. (2) TatA protomers are then recruited to the TatBC complex and polymerize to form a channel dependent on the proton motive force (PMF). (3) The substrate protein then traverses the membrane through the polymerized TatA complex and the signal peptide is cleaved.

As mentioned previously, the Tat system requires a signal peptide to initiate protein export. These signal peptides bear three domains: the n-region containing a positively charged N-terminal domain, the h-region containing the hydrophobic domain, and the c-region containing the C-terminal domain. Signal peptides carry the following consensus motif:

$$\text{Ser/Thr-Arg-Arg-X-Phe-Leu-Lys},$$

where X is any polar amino acid. Among the many prokaryotic twin-arginine signal peptides, the 39-amino-acid-long trimethylamine N-oxide reductase (TorA) from *E. coli* is the most heavily used for its high export selectivity, and the remainder of this study exclusively employs the TorA peptide as the signal peptide fused to the MS2 coat protein.

**Exploiting native *E. coli* colicin machinery to escape the periplasm and into the extracellular space**

Colicins are toxic proteins made by certain strains of *E. coli* as protection against related strains of *E. coli*, but self-immunity prevents self-destruction by those same toxic peptides. The genetic determinants of natural colicins were found on plasmids, which contain the colicin operon, and other genes for colicin activity. Plasmid colicin E1 encodes
an ionophoric, a pore-forming protein, and can inhibit macromolecular synthesis without hindering cellular respiration.\textsuperscript{33} Natively used as a method for transporting bacteriocins into the extracellular environment, bacteriocin release proteins (BRPs), such as the colicin lysis protein kil, range from 45 to 52 amino acid residues in length and contain a “lipobox” sequence (Leu-X-Y-Cys, where X and Y are small, neutral amino acid residues) necessary for lipid modification at the outer membrane.\textsuperscript{34}

Since the discovery of components of the colicin plasmid, numerous research groups have used the colicin lysis \textit{kil} (or \textit{celA}) gene of the plasmid for the export of various gene products produced by \textit{E. coli}. Among the first of such projects was the export of heterologous periplasmic protein by fusing the \textit{kil} gene with a growth-phase-dependent promoter. The result was a successful, near-triple increase in the secretion of a desired protein, beta-glucanase, by inducing a “quasi-lysis,” or decline in culture turbidity, with no lethality to the cells.\textsuperscript{35} Proteins with sizes ranging from 17 to 135 kilodaltons have been successfully secreted through the pores made by bacteriocin release proteins such as \textit{kil}.\textsuperscript{36} Weak and strong promoters for the \textit{kil} gene and the desired protein to be secreted can be used in different combinations to optimize the production and secretion of target proteins into the culture medium without cell lysis.\textsuperscript{37} Exploiting the capacity to tune the expression of the \textit{kil} and secretion protein, we fused the \textit{kil} gene to an IPTG-inducible promoter to allow for regulation of the periplasmic protein leakage from the cell.

This work describes the construction of a bacterial RNA message export system built thus: (1) any desired RNA message to be exported is (a) made with the hairpin sequence recognized by MS2 and (b) flanked with PIE sequences that allow for the message’s circularization; (2) the MS2 coat protein, which is fused to a TorA signal peptide for the Tat system, binds to the hairpin on the RNA message; (3) This TorA-MS2-RNA-message complex is then recognized by the TatABC protein situated on \textit{E. coli}’s inner membrane and is funneled out into the periplasm; and (4) the colicin kil protein is induced by IPTG to perforate the outer membrane and cell wall, allowing for the passive diffusion of the complex into the extracellular environment (Fig. 6).

\textbf{Figure 6.} Proposed model for the engineered message export system. Cells first produce the circularized RNA, whose hairpin is recognized by the MS2 coat protein monomer, which is tethered to a signal protein called TorA, and directed into the periplasm via the TatABC complex in the inner membrane (a). The Kil protein is then induced to perforate the \textit{E. coli} cell wall (b). Once the MS2TorA-message complex enters the periplasm, it escapes into the extracellular environment by passive diffusion (c). (TatABC is simplified as a channel protein in the image.)
MATERIALS AND METHODS

E. coli strains and plasmids

E. coli strains and plasmids used in these strains are described in Table 1 in the appendix.

Generation of an MS2-TorA-message and TorA-GFP fusion proteins

Construction of plasmid pMS2TorA-msg and pTorA-GFP. The MS2 coat protein sequence, TorA signal peptide sequence, and GFP sequence were purchased as E. coli codon-optimized gBlocks from IDT (Integrated DNA Technologies, Coralville, IA, USA) and introduced into a pUC19 backbone by Gibson Assembly and mobilized into E. coli MachI (C862003; ThermoScientific Fisher, Waltham, MA) via electroporation. The plasmids isolated from transformations growing on carbenicillin-containing media were sequence-verified by Sanger sequencing (Genewiz, Cambridge, MA, USA).

Two messages were designed to test the viability of the MS2 coat protein. The first nucleic acid message had a length of 124 nucleotides, and the second had a length of 150 nucleotides. Both were purchased as single-stranded DNA oligos or gBlocks, henceforth abbreviated as “message oligo 124,” and “message oligo 150,” respectively (IDT). The message oligos are designed to have a flanking 20-bp complementary overhang with the site of insertion in the pUC19-MS2TorA (abbreviated as pMS2-TorA) plasmid, and the message oligos were PCR-amplified with Q5 High-Fidelity Master Mix (New England BioLabs, NEB, Ipswich, MA, USA) using primers with sequences listed in Table 3 in the appendix. The message oligos were inserted into the pUC19-MS2TorA backbone by Gibson Assembly using the Gibson assembly master mix (E2611; NEB) according to the specification of the manufacturer. The plasmid containing message oligo 150 also contained an MS2 protein that had two amino acid residues between the TorA signal peptide and the MS2 protein itself as listed in Table 3 (MS2-bumper-for and MS2-bumper-rev). The resulting plasmids were electroporated into electrocompetent E. coli MachI and transformants were grown overnight on LB agar supplemented with carbenicillin (100 µg ml⁻¹) at 37°C. Constructs isolated from the transformants were confirmed by Sanger sequencing and subsequently mobilized into E. coli BL21(DE3) (C2527; NEB) by electroporation. Single transformants for pUC19-MS2TorA-msg-124 (abbreviated pMS2TorA-msg-124) and pUC19-MS2-TorA-msg-150 (abbreviated pMS2TorA-msg-150) were selected on carbenicillin-supplemented media and grown at 37°C.

Construction of plasmid to include pKil: the Kil gene was purchased pre-inserted into E. coli expression vector pD431-SR with Kanamycin resistance (DNA 2.0, Menlo Park, CA, USA). The Kil gene was inserted to the pUC19-MS2-TorA-msg-150 plasmid by the purchase of a gBlock (IDT) as listed in Table 3.

Double transformants were assayed for both the pKil, pMS2TorA-msg-124, pMS2TorA-msg-150, and pTorA-GFP plasmids by colony PCR with GoTaq G2 Green Master Mix (Promega, Madison, WI) and primers are listed in Table 3 in the appendix. Amplicons were analyzed by agarose gel electrophoresis: 2% TBE agarose gels were poured and run at 110 volts for 1 hour in 1X TBE (15581-044; Life Technologies, Grand Island, New York).
Media and growth conditions
For construct assembly, both strains of *E. coli* MachI and BL21(DE3) were grown at 37°C on LB agar (BD, Franklin Lakes, NJ, USA) supplemented with either carbencillin (100 μg ml⁻¹), kanamycin (30 μg ml⁻¹), or both. For double transformants and time course experiments using plasmid pMS2TorA-msg-124, where RNA absence is necessary, potentially contaminating RNA was removed thus: LB broth was treated with RNase A (12091-021; Life Technologies) at a concentration of 0.5 μg/ml for 1.5 hours at 37°C as described previously. The LB-RNase A solution was subsequently treated with DEPC to a final concentration of 1% and incubated overnight at room temperature remove any active RNases as suggested by Thermo Fisher Scientific. The DEPC-treated LB liquid medium was then autoclaved at 121°C, 15 PSI, for 15 minutes to inactivate the DEPC. This RNase-DEPC treated LB will be referred to as RNA-free LB, or rLB, in the remainder of this study.

Measuring GFP fusion protein export
Double transformants containing the pKil and pTorA-GFP plasmids, single transformants containing pTorA-GFP only or pKil only, and cells containing no plasmids were grow in rLB supplemented with antibiotics as described above and induced with IPTG (where pKil was present) when the cells reached an OD₆₀₀ of 0.6. A 1-ml aliquot of each of the cultures done in triplicate was spun down by centrifugation for 5 min at 9000 rpm and filtered with a 0.2-μm Nylon membrane. The supernatant (culture medium) was then assayed by absorbance readings on a spectrophotometer at 280 nm.

Message Circularization
The RNA message was circularized autocatalytically by means of the innate circularization sequences pre-designed to flank the message sequence (Fig. 7).

![Figure 7](image)

**Figure 7.** Schematic for flanking desired circularized sequence with intron sequences. Yellow is the RNA message of one’s choosing. As long as the blue sequences (upstream and downstream intron sequences) are flanking the 5’ and 3’ ends of the message, it will be circularized. The RNA message depicted in this figure contains an MS2 binding hairpin sequence near its 3’ end (Fig. 3a).

Isolation and purification of RNA from growth medium
Transformants containing both the pKil and pMS2TorA-msg-124, or pKil and pTorA-GFP
plasmids were grown overnight in rLB supplemented with carbenicillin (100 μg ml⁻¹) and kanamycin (30 μg ml⁻¹). Transformants containing pMS2-TorA-msg-150 only was grown in minimal media (no carbon), supplemented with carbenicillin (100 μg ml⁻¹). Fifty microliters of each of the pKil and pMS2TorA-msg-124, and pKil and pMS2TorA-GFP double-transformant overnight cultures were used to inoculate 50 ml of rLB supplemented with carbenicillin and kanamycin as described above and grown to an OD₆₀₀ of 0.600. Likewise, the single transformant pMS2TorA-msg-150, which contains the kil gene within the same plasmid, was grown instead with minimal media (no carbon) and carbenicillin as described above. OD measurements were recorded using the Ultrospec 2100 pro spectrophotometer (Amersham Biosciences). Each double-transformant cell culture’s kil gene was induced with IPTG at a final concentration of 1 mM and grown for four hours at 37°C after induction. Single transformants containing pMS2TorA-msg-150 were induced with rhamnose to a final concentration of 1%, and aliquots were taken hourly three hours after induction. Double transformants with no IPTG induction and single transformants with no rhamnose induction were also grown as controls under the same conditions as detailed in Table A. For the single transformant of pMS2TorA-msg-150, T₀ is the time course aliquot taken before addition of the rhamnose, and subsequently collections occurred hourly, three hours after addition of the rhamnose.

<table>
<thead>
<tr>
<th>Table A. Transformation and induction scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmids</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Msg w induction</td>
</tr>
<tr>
<td>Msg no induction</td>
</tr>
<tr>
<td>GFP w induction</td>
</tr>
<tr>
<td>GFP no induction</td>
</tr>
</tbody>
</table>

At every time point (i.e., every hour after induction, for four hours), 1 ml of cells was aliquoted from each culture, with each time point, message, and control done in triplicate. The cells were then pelleted with a microcentrifuge running for 5 minutes at 9000 rpm. The supernatant was filtered by pushing the liquid through a luer-lock syringe and a 0.2-μm Nylon membrane filter (4436; Pall, Port Washington, NY, USA). Pellets from pMS2TorA-msg-124 were set aside for TEM imaging described in the next section. Phenol:chloroform extraction was used to isolate the RNA from solution and the cell pellets. The following protocol was adapted from Cold Spring Harbor Laboratory Press. Briefly, a 1/9th volume of 10X SDS buffer (5% SDS, 10 mM EDTA, 200 mM Tris-HCl) was added to the supernatant and vortexed. Proteinase K (97026-670; VWR, Radnor, PA, USA) was then added to a concentration of 100 μg/ml and the solution was incubated for 20 min at 42°C. Subsequently, NaOAc (S7899; Sigma Aldrich, St. Louis, MO, USA) was incorporated to a final concentration of 0.3 M, pH 5.2. Two volumes of phenol:chloroform:isoamyl alcohol (25:24:1) (PCA) (77617; Sigma Aldrich) were added to the solution, which was then vortexed vigorously and centrifuged at 13,000 rpm (maximum speed of the centrifuge) for
2 min at room temperature. The upper aqueous phase was carefully transferred to a fresh tube, and two volumes of chloroform: isoamyl alcohol (C0549; Sigma Aldrich) were added to the aqueous phase and vortexed vigorously. The solution was centrifuged at the maximum speed of the centrifuge for another 2 minutes, and the upper aqueous layer transferred to a fresh tube.

The RNA in the aqueous layer was ethanol precipitated as previously described. Glycogen (97063-256; VWR) was added to a final concentration of 1 μg/μl, and the solution was once again adjusted with NaOAc to achieve a final concentration of 0.3 M. The RNA solution was then supplemented with 2.5 volumes of 100% ethanol and vortexed vigorously. This RNA-ethanol solution was frozen overnight at -80°C. The frozen solution was thawed quickly and pelleted by centrifugation at the maximum speed for 10 min. The ethanol was decanted and the nucleic acid pellet saved for further purification by washing twice with 500 μl pre-chilled 70% (v/v) ethanol and centrifuging for 5 min at 4°C. The washed pellet was air-dried, resuspended in RNase-free water, and residual DNase reaction components were removed and remaining RNAs purified via Qiagen RNeasy Mini kit (74104; Qiagen, Venlo, Netherlands) as described by the manufacturer. Residual contaminating DNAs of the kil-only and message-only controls were removed using the DNA-free™ DNA Removal Kit (AM1906; ThermoFisher) as described by the manufacturer, while pMS2TorA-msg-150 nucleic acid isolates were treated with RNase-free DNase I (89836; Life Technologies).

RNA Analysis
The presence of RNA was initially tested for by digestion with 1 μg/μl RNase A (12091-021; Life Technologies) at 37°C for 3 hours. In a separate reaction, the presence of lariat or circularized RNA in each sample was assayed by digestion with 1 unit RNase R for every 1 μg of purified RNA at 37°C for 3 hours (RNR07250; Epicentre Technologies, Madison, WI). Known linear DNA and known linear RNA were both used as controls to verify the specificity of RNase-free DNase I, RNase A, and RNase R.

The RNase-treated samples were analyzed via agarose gel electrophoresis: 4% agarose gels were made from NuSieve 3:1 (50091; Lonza, Basel, Switzerland) and formaldehyde as described by the manufacturer (Lonza). Briefly, 6 g NuSieve were dissolved in 150 ml Millipore-filtered water and cooled to 60°C in a hot water bath. Fifteen ml of prewarmed 10X MOPS Buffer (50876; Lonza), 8.35 ml 37% formaldehyde (252549; Sigma Aldrich), and 1X Sybr Green II (50523; Lonza) were added to the dissolved agarose, mixed, and cast in a fume hood. Samples were prepared by mixing a denaturing RNA loading dye to 1X concentration (B0363S; NEB) and heated at 65°C for 10 minutes. The agarose gels were then run at 100 volts for 1.5 hours in 1X MOPS.

Identification of the RNA message
Purified RNA from each time point was reverse-transcribed using SuperScript IV Reverse Transcriptase (18090010; Life Technologies) according to the specification of the manufacturer with primers listed in Table 3 of the appendix. The reverse transcribed products were then subjected to a qPCR cycle with Roche LightCycler 480 Instrument II (Roche, Indianapolis, IN, USA) with Power Sybr Green PCR Master Mix (4367659; Thermo Fisher Scientific).
Determination of RNA binding to MS2 capsid protein

Total-protein isolation
Single transformants containing pMS2-TorA-msg-124 were grown to an OD$_{600}$ of 0.6 in rFLB at 37°C. Cell were pelleted for 5 min at 9000 rpm, washed twice with 1X PBS, and resuspended in 100 µl of 25 mM Tris (pH 7.5), and 5 µl of PureLink DNase was added (100002884; Life Technologies). The samples were incubated without shaking for 15 min at room temperature, after which the cells were pelleted by centrifugation for 15 min at 4000 rpm. The supernatant was extracted and placed on ice; 90 µl of this was filtered using a 0.22 µm Nylon spin-filter. The pellet was resuspended in 25 mM Bug Buster Protein Extraction Reagent (70584; EMD Millipore, Billerica, MA, USA and 5 µl PureLink DNase, and incubated at room temperature for 15 min without shaking. The lysate was then treated with 475 µl of 25 mM Tris (pH 7.5) and the cell debris pelleted by centrifugation for 15 min at 4000 rpm. A 450- µl aliquot of the cleared lysate was filtered using a 0.22 µm Nylon spin-filter and placed on ice.

TEM
Precipitated MS2 protein was viewed by transmission electron microscope (TEM) analysis. A 5-µl aliquot of the sample was spotted onto formvar-carbon grids (Electron Microscopy Sciences), washed with Millipore water and stained with 1% uranyl formate before analysis on a JEOL 1200 TEM.
RESULTS
Viability of TorA-GFP fusion protein and kil double-transformants
First, the viability of the Tat system in E. coli was tested by fusing the TorA signal peptide to a GFP reporter protein. Measuring the absorbance at 280 nm of the pKil and TorA-GFP double-transformant’s growth medium over the course of four hours after induction with IPTG revealed an increase in the growth medium’s protein concentration (Fig. 8), though pKil alone also demonstrated a general increase in protein concentration that tapers out at the fourth and fifth hour after induction. It is worth noting that this protein increase happens at a “slower rate,” or there appears to be “less” protein leakage over time with pKil alone. Control cultures with neither plasmid and control cultures with pTorA-GFP alone have negligible changes in protein concentration of the growth medium over time.

![Figure 8. Absorbance readings of single and double transformants' growth media over time in arbitrary units (AU).](image)

**Figure 8.** Absorbance readings of single and double transformants' growth media over time in arbitrary units (AU). Transformants containing no kil gene (i.e., the pTorA-GFP and no-plasmid controls) demonstrated low absorbance readings throughout the four-hour time course. Transformants containing only the kil gene had low absorbance readings, and those containing both the kil gene and the TorA-GFP fusion had an apparently linear increase in absorbance over time.

Characterizing RNA from MS2-TorA-msg-124-and-kil double-transformants
Presuming that the Tat system was functional, constructs were made to bear the MS2 coat protein and an RNA message instead of a GFP. Instead of doing crude protein absorbance readings with a spectrophotometer, however, the nucleic acids from the medium were isolated for more detailed and specific characterization. The exported nucleic acids found in the growth media were found to consist of circularized RNA. Nucleic acids isolated from the growth medium and pretreated with RNase-free DNase I were enzymatically digested in separate reactions using RNase A and RNase R. The former resulted in total digestion of all nucleic acid content (Fig. 9, lanes 9-16). RNase R is an enzyme that digests all linear RNA and leaves lariat RNA and circularized RNA intact. Upon treatment with RNase R, the purified growth medium nucleic acid content reveals that there are RNA molecules that run at around 40 (Fig. 9, lanes 1-8). The known circular DNA and known linear RNA that were both used as controls to verify the specificity of RNase-free DNase I, RNase A, and RNase R,
resulted in the following:

(1) DNA known to be circular is digested with DNase I, but not with RNase A or RNase R;
(2) RNA known to be linear is cut with both RNase and RNase R, but not with DNase I (Fig. 9, Lanes 25-32).

**Figure 9.** Agarose gel electrophoresis analysis of RNA isolated from time course experiments. 4% NuSieve 3:1, 100 volts, 1.5 hours, 1X MOPS. Lanes 1-4 contain samples of induced double-transformants treated with RNase R for 3 hours. Lanes 5-8 contain samples of uninduced double transformants treated with RNase R for 3 hours. Lanes 9-12 contain RNase A-treated samples for induced and lanes 13-16 uninduced double-transformants. Lanes 17-24 contain single transformants of pMS2TorA only (two for each time point). Lane 25 contains untreated pKil circular DNA (i.e., plasmid), lane 26 is pKil after digestion with RNase-free DNase I, lane 27 contains pKil after treatment with RNase R, and lane 28 contains pKil after treatment with RNase A. Lane 29 contains untreated linear RNA of 40 nt, lane 30 is the linear RNA after treatment with RNase-free DNase I, lane 31 contains the linear RNA after treatment with RNase R, and lane 32 contains the linear RNA after treatment with RNase A.

To assay whether the RNA message was present in the pool of nucleic acids isolated from the time course, all samples were reverse transcribed and then PCR-amplified with divergent primers (msg-divergent-for and msg-divergent-rev) listed in Table 3 of the appendix. The PCR amplification produced amplicons of lengths that were multiples of approximately 125 bp, which is equivalent to the length of the RNA message. (Fig. 10, lanes 3-11). The no-template control containing only the PCR primers resulted in negligible amplification (Fig. 10, lane 1). Kil-only control also resulted in near-negligible amplification as in the no-template control (Fig. 10, lane 2). The single transformant containing only pMS2TorA-msg experienced some amplification (Fig. 10, lane 3).

**Figure 10.** Agarose gel electrophoresis analysis of reverse-transcribed and PCR-amplified nucleic acid products isolated from growth media. 2% agarose gel, 100 volts, 1.5 hours in 1X TBE. All samples were first
isolated from the growth medium, reverse transcribed with a specific primer, and then PCR amplified. The no template control contains a single band equivalent to length of small primer dimerization product (lane 1). A second control containing only pKil also only has what appears to be a primer dimerization product. The single transformant containing only the pMS2TorA-msg plasmid demonstrates some amplification products with a length of around 125 bp. Lanes 4-7 contain the PCR products of uninduced double transformants containing both pKil and pMS2TorA-msg, and lanes 8-11 contain the PCR products of the induced double transformants. While there are smears of DNA present, the primary bands are all approximately multiples of 125 bp: 125, 250, and 750.

**Determining the potential for in vivo MS2 capsid formation**

To understand what may be happening within the cells, cell contents were imaged with TEM. Viral capsids are 23-30 nm in diameter and icosohedral in appearance. Capsids encapsulating RNA have distinct characteristics when viewed by TEM: empty capsids have dark “ring-like” borders, whereas RNA-containing capsids appear uniformly granular. TEM analysis unveiled the presence of capsids present within cells (not growth medium) that feature the granular texture of RNA-containing capsids.

![Figure 11. TEM analysis of cellular contents. Scale bar is equivalent to 100 nm. Capsids found within pMS2TorA-msg-only transformants are spherical and granular in appearance.](image)

**Determining Controlled Export versus Cell Lysis**

In order to verify whether cells were truly exporting message, or simply lysing to expose cellular contents to the extracellular media, the plasmid pMS2TorA-msg-150 was constructed to contain a two-amino-acid bumper sequence between the TorA signal peptide and MS2 coat protein. RNA content was extracted from cell pellets and extracellular media, and reverse-transcribed with both a message-specific and 16S rRNA-specific primer. The cDNA generated from the reverse transcription was then analyzed using qPCR.

Based on the C\(\text{p}\) points, the cell interior of Kil-only expressing cells had much more 16S rRNA than did the message, though the “message” signal was not a true signal as deduced from the shape of the absorbance curve, which was flat (Table B, lines A and B, Supplementary figure 4). The growth media of the Kil-only expressing cells had
approximately 430 times (from $2^{(15.75-7)}$) less 16S rRNA than did the cell pellet (Table B, lines C and D). The cell interior of the message-only expressing cells had about 16 times (from $2^{(12-7.75)}$) more 16S rRNA than it did the message (Table B, line E). The growth media of the message-only expressing cells had approximately 100 times (from $2^{(14.55-7.75)}$) less 16S rRNA than did the cell interior (Table B, lines F and E), and the message was over 1600 times (from $2^{(22.7-12)}$) more prevalent in the extracellular media than within the cells (Table B, lines F and E).

For cells that contained the plasmid with a kil protein, a 150-nucleotide message, and two-amino-acid spacer between the TorA and MS2 protein (i.e., pMS2TorA-msg-150), the cell interior experiences a slight decrease of message (Table B, line G), and has twice as more 16S rRNA present over the course of three hours after induction (Table B, line G). The growth media of these cells had about the same amount of message and 16S rRNA over the course of three hours (Table B, line I), and the amount 16S rRNA remains fairly constant (Table B, line J).

Table B. qPCR Cp values of cell pellets and extracellular media from time course.

<table>
<thead>
<tr>
<th>Cell content</th>
<th>Fraction</th>
<th>Time elapsed</th>
<th>Amplification primers</th>
<th>Cp (avg)</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Kil</td>
<td>cell pellet</td>
<td>4 hours</td>
<td>universal primer</td>
<td>7</td>
<td>0.00</td>
</tr>
<tr>
<td>B Kil</td>
<td>cell pellet</td>
<td>4 hours</td>
<td>msg only primer</td>
<td>32.25</td>
<td>0.96</td>
</tr>
<tr>
<td>C Kil</td>
<td>media</td>
<td>4 hours</td>
<td>universal primer</td>
<td>15.75</td>
<td>1.71</td>
</tr>
<tr>
<td>D Kil</td>
<td>media</td>
<td>4 hours</td>
<td>msg only primer</td>
<td>33.75</td>
<td>4.03</td>
</tr>
<tr>
<td>E msg only</td>
<td>cell pellet</td>
<td>4 hours</td>
<td>universal primer</td>
<td>7.75</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>cell pellet</td>
<td>4 hours</td>
<td>msg only primer</td>
<td>12</td>
<td>0.82</td>
</tr>
<tr>
<td>F msg only</td>
<td>media</td>
<td>4 hours</td>
<td>universal primer</td>
<td>14.55</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>media</td>
<td>4 hours</td>
<td>msg only primer</td>
<td>22.7</td>
<td>2.12</td>
</tr>
<tr>
<td>G pMS2TorA-msg-150 (Msg+Kil)</td>
<td>cell pellet</td>
<td>2 hours</td>
<td>msg only primer</td>
<td>10.8125</td>
<td>3.49</td>
</tr>
<tr>
<td></td>
<td>pMS2TorA-msg-150</td>
<td>cell pellet</td>
<td>4 hours</td>
<td>msg only primer</td>
<td>14.75</td>
</tr>
<tr>
<td>H pMS2TorA-msg-150</td>
<td>cell pellet</td>
<td>2 hours</td>
<td>universal primer</td>
<td>18.5</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>pMS2TorA-msg-150</td>
<td>cell pellet</td>
<td>4 hours</td>
<td>universal primer</td>
<td>19.625</td>
</tr>
<tr>
<td>I pMS2TorA-msg-150</td>
<td>media</td>
<td>2 hours</td>
<td>msg only primer</td>
<td>16.3</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td>pMS2TorA-msg-150</td>
<td>media</td>
<td>4 hours</td>
<td>msg only primer</td>
<td>16.975</td>
</tr>
<tr>
<td>J pMS2TorA-msg-150</td>
<td>media</td>
<td>2 hours</td>
<td>universal primer</td>
<td>31.375</td>
<td>2.13</td>
</tr>
<tr>
<td></td>
<td>pMS2TorA-msg-150</td>
<td>media</td>
<td>4 hours</td>
<td>universal primer</td>
<td>31.3125</td>
</tr>
</tbody>
</table>
DISCUSSION

Our pTorA-GFP and pKil double transformants were our preliminary control meant to verify whether the native Tat system in our *E. coli* BL21(DE3) cells were functional in conjunction with pKil. Crude readings with the spectrophotometer were to inform whether there was protein leakage over time with (1) no plasmid, (2) pKil alone, (3) pTorA-GFP alone, and (4) pKil and pTor-GFP double transformants with induction. The absorbance readings indicate that pKil alone allows for a noticeable, but slow, leakage of proteins from the cells into the growth medium over the course of four hours (Fig. 8, red squares), but these absorbance readings were not as strong as those of the double transformants (Fig. 8, blue diamonds). This is likely because the TorA-GFP fusion proteins contribute to the population of proteins leaking out of the periplasm. Transformants that lacked the pKil plasmid had low protein absorbance readings (Fig. 8, green triangles, purple crosses), suggesting that the cell cultures dying at a certain rate may also be contributing to the leaked protein population. Because cells were not pulsed with IPTG (i.e., no cells were washed after pKil induction with IPTG), it is difficult to identify whether the leakage of proteins resulted from the continuous perforation of the membrane with the Kil protein in the presence of IPTG, or whether a constant number of cell wall holes simply allowed a steady stream of proteins to leak and accumulate in the growth medium. This experiment was an indicator, however, that the TorA and Kil combination were sufficient to allow for a detectable change in protein concentration in the growth medium.

In attempting to express an RNA message, we chose to start with a short sequence in consideration of the fact that nucleic acids can be large molecules (e.g., a single-stranded RNA of 124 nucleotides has a size of around 39 kDa, while the molecular weight of GFP is around 27 kDa, and the molecular weight of the MS2 coat protein monomer is 17.3 kDa). Because the sizes of cellular holes introduced by Kil are unknown, we opted to keep our message-protein (TorA-MS2) complex small (estimated to be 50-60 kDa), lest large RNA molecules lead to steric hindrance and impede protein exit from the cell. Nucleic acids extracted from the growth medium after purification indicates that we have “message delivery” to the extracellular environment to a degree; that is, the message is at least in RNA form, is intact and circular, and can be detected via PCR amplification after reverse transcription (Fig. 10, lanes 4-11).

The nucleic acids isolated from the growth medium were first treated with RNase A and RNase R in order to verify that our isolates were both RNA and circular. To ensure that our enzymes were functional and specific, plasmid DNA (that is, circular DNA) was used as a control and digested with RNase-free DNase I, RNase R, and RNase A (Fig. 9, lanes 18, 19, 20 respectively). Known linear RNA was then treated as well with RNase-free DNase I, RNase R, and RNase A (Fig. 9, lanes 30, 31, 32, respectively). As expected, DNase I only digested DNA, RNase A only digested RNA, and RNase R exclusively digests linear RNA. Treatment of both induced and uninduced double transformants with RNase R reveals little to no digestion of our isolates (Fig. 9, lanes 1-8), suggesting that our isolates were, indeed circular RNA. Treatment of the same samples with RNase A completely digests all the samples, further verifying that our isolates are RNA only (Fig. 9, lanes 9-16). It is noteworthy that single transformants containing only the message plasmid (i.e., pMS2TorA-msg) also produced circular RNA. In order to further understand why there was leakage of the message into the growth medium without Kil, TEM analysis was employed.

Since *in vitro* reverse transcription reactions only permit the use of a single primer, such a reaction with a circular template is anticipated to result in a type of “rolling-circle”
transcription, where there is not a termination point until the reverse transcriptase is heat-inactivated (Fig. 12a). Subsequent PCR amplification of such a reverse-transcribed product would be expected to yield amplicons with various lengths that are multiples of the length of a single copy (i.e., if a single PCR product were expected to give a 100-bp amplicon, a PCR amplification of a rolling-circle product with known primers would give products of lengths of 200 bp, 300 bp, 400 bp, etc.), and such is the result that we observe in our agarose gel electrophoresis analysis (Fig. 10, lanes 4-11). The absence of significant amplification products for the no-template and pKil-only controls confirm that the primers are at least specific to our message (Fig. 10, lanes 1-2). Because our single transformant (pMS2TorA-msg-only) demonstrates some amplification products, it is possible that low levels of leakage are occurring without induction. Indeed, there is very little difference between induced and uninduced samples of the transformants containing both the message and the Kil protein. A closer examination of our agarose gel also indicates that there are very small products that are about 50 bp in length. Investigation into relevant MS2 literature indicates that *E. coli* RNA is also sequestered to the coat protein with great affinity and suggests that the next iteration of message design ought to outcompete native *E. coli* RNA for MS2 coat protein binding sites. 47

**Figure 12.** Graphical depiction of reverse transcription of circular RNA. Because reverse transcription requires only one primer, a template that is circular RNA will result in a cDNA product of many lengths (a). PCR amplification with a single set of primers (any one red and any one black) will likewise produce amplicons of various lengths, but are likely to be multiple of the length of the circular template.

In spite of being unable to definitely quantify the amount of RNA message produced for export, that our message was able to find its way to the extracellular environment and remain intact is a critical first-step in the building of longer messages. To investigate why there was minor leakage of the message in single transformants without Kil and to better understand the delivery mechanism of our message-protein-fusion vehicle for the next experimental iteration using a longer message, single transformants only containing the message plasmid pMS2TorA-msg were grown to late log phase. The extracellular medium was washed away from the cells, and the cells chemically lysed. The contents of the cell lysis were then examined by TEM.

It was known to us that, when binding to an MS2 monomer, the MS2-specific RNA hairpin triggers dimerization and subsequent oligomerization of MS2 to form a capsid. 48 It is also known that fusing a peptide as short as two amino acids to either the N or C terminus of the
coat protein was sufficient to significantly hinder its oligomerization into a capsid. For these reasons, we designed a new MS2-TorA plasmid (pMS2TorA-msg-150) such that the TorA signal peptide was separated by two random amino acids.

An examination of our cell contents via TEM quickly revealed the presence of capsid clusters with encapsulated RNA in cells containing pMS2TorA-124, which did not have the two-amino acid spacer between the TorA peptide and MS2 protein (Fig. 11). Further investigation into our design indicated that the FG loop of the MS2 coat protein was present; we had used the native protein sequence instead of the sequence with a mutated FG loop presuming that the 30-aa signal peptide would have been sufficient. However, a more extensive review of Tat pathway literature indicated that the entire length of the TorA signal peptide is cleaved off as the protein is transported through the Tat machinery. This suggests that, while our MS2TorA-msg-124 fusion proteins may not be assembling in the bacterial cytoplasm, there is a high probability of their assembly within the periplasmic space after transport through the Tat protein complex because the absence of two-amino acid minimum is unable to stop oligomerization of the coat proteins. The excessive pressure of whole capsids in the periplasmic space might have been sufficient to lyse a small number of cells such that our reverse transcription reaction was able to detect the RNA message (Fig. 10, lane 3).

On the other hand, cells containing pMS2TorA-msg-150 and grown in minimal media had a less noticeable struggle in export of the message. By integrating the two amino acids between the TorA sequence and the MS2 coat protein, we mitigated the formation of entire capsids. Analysis of the cellular and extracellular contents over the course of three hours revealed that the ratios of 16S rRNA and message RNA varied somewhat predictably (Table B). The RT-qPCR read-out of kil-only expressing cells indicated an abundance of 16S rRNA within the cells (Table B, line A), and its noticeable, but less imposing presence in extracellular media (Table B, line C). Message-specific primers for kil-only cells did not make an appreciable signal in the RT-qPCR reaction (Table B, lines B and D). The message-only cells, on the other hand, had the same amount of 16S rRNA with the cells as the kil-only cells did (Table B, lines A and E).

Cells with pMS2TorA-msg-150, in which both the message and kil protein are expressed, were likewise analyzed by RT-qPCR using primers specific for the 16S rRNA and the message. The interior of these cells had a high expression of the RNA message and relatively less 16S RNA. Over the course of three hours, the 16S rRNA content with the cells remained high, though there is a slight decrease as time elapses. The amount of the message within the cells also slightly decreases over time, but does so more rapidly than the 16S RNA (i.e., the RNA message appears to be depleted from the cells more quickly). Even though there is an apparent “loss” of both 16S RNA and RNA message from the cell interior over time, the amount of both RNAs in the media remains nearly constant in the time course. This is probably due to RNA degradation occurring in the growth media from lysis debris, which may contain nucleases. The noticeably greater amount of RNA message in the growth media as compared to 16S rRNA strongly suggests that controlled export of RNA is happening more than cell lysis is occurring.

The controls in which we measured kil-only and message-only expressing cells seem to indicate that there is “more” RNA overall, relative to the experimental samples, with the exception of trying to amplify the kil-only material with message-specific primers for which there is poor amplification. Because the nucleic acid content of the kil-only and message-only controls were isolated at a later time and treated with a different DNase I, comparison of the
crossing point (Cp) values may only be objectively done by comparing the controls with each other, and the experimental data with each other, but not by comparing the controls directly with the experimental data. Because we were measuring the amount of RNA being exported to the extracellular growth media, and it is unknown whether *E. coli* has a standard RNA molecule that it exports regularly, a "housekeeping gene control" could not have been used as it normally is done with eukaryotic RNA extracts. Alternatively, the large variation between the controls (Table B, lines A-F) and the kil-msg-dual-transformants (Table B, lines G-J) may have been the result of more stable, or healthy, cells that were not required to export bulk RNA and RNA-carrying proteins. That the message was detectable in the extracellular environment suggests the possibility of sending other functional, short messages such as ribozymes. Further fine-tuning of each of the components (message affinity for MS2 coat protein, or other RNA-carrying proteins) may help make this system more efficient. This concludes the work done to demonstrate the use of circularized RNA as extracellular missives to be sent to the extracellular environment and its survival for up to three hours without significant cell lysis.

An attempt to further the communication system was executed by testing RNA uptake from growth media sans excessive treatment of cells. This method entailed the exposure of non-competent BL21(DE3) cells in late-log phase with high concentrations of short RNAs approximately 60 nucleotides in length. These short RNAs were designed as triggers to switch on the expression of GFP as developed previously. After incubation of these late-log cells that have had no chemical treatments for a period of 15 hours, a distinctive increase in GFP fluorescence is present (Supplementary figure 6, Table 4) after normalizing the cell populations across samples using measurements for optical density. Together, these data hold promise for a complete, engineered bacterial communication system that can both send and receive nucleic acid messages.
Acknowledgements

I would like to express my deep gratitude to Professor Joseph M. Jacobson, Dr. Shuguang Zhang, and Professor Ed Boyden, my research advisors, for their invaluable and constructive guidance in the planning and development of this research work. I thank Professor Neil Gershenfeld for providing the many resources in the laboratory. My special thanks are extended to Noah Jakimo, Peter Q. Nguyen, and Charles V. Fracchia, my collaborators and friends whose brilliance, sound advice, and countless hours of scientific discussion were indispensable to the initiation, execution, and completion of this work. I would also like to extend my thanks to Jeffrey Way, Thrasyvoulos Karydis, James Pelletier, Pranam Chatterjee, and Prashant Patil for their inspirational conversations and kind patience in protocol-debugging of the scientific and social variety. Lastly, I would like to send my sincere gratitude to Jamie Severson, Joe Murphy, Ryan Walsh, and Linda Peterson for their enthusiastic encouragement, administrative assistance, and behind-the-scenes work that has made my research possible.
### Table 1. *E. coli* cell strains used

<table>
<thead>
<tr>
<th>Strains</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td></td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>NEB</td>
</tr>
<tr>
<td>MachI</td>
<td>NEB</td>
</tr>
</tbody>
</table>

### Table 2. Plasmids and their antibiotic markers

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Details</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pKil</td>
<td>KanR</td>
<td>DNA 2.0</td>
</tr>
<tr>
<td>pMS2-TorA</td>
<td>puc19, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pTorA-GFP</td>
<td>puc19, AmpR</td>
<td>This study</td>
</tr>
<tr>
<td>pMS2-TorA-kil</td>
<td>puc19, AmpR</td>
<td>This study</td>
</tr>
</tbody>
</table>

### Table 3. Oligos used for the construction or sequencing of plasmids in this study

<table>
<thead>
<tr>
<th>Target</th>
<th>Oligo name</th>
<th>Oligo sequence (5'-3')</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Kil</em></td>
<td>kil-for</td>
<td>atgcgtaacagttttttgtgg</td>
<td>Segment amplification</td>
</tr>
<tr>
<td></td>
<td>kil-rev</td>
<td>tcttgaccgcgtacccc</td>
<td>Segment amplification</td>
</tr>
<tr>
<td><em>RNA message</em></td>
<td>msg-divergent-for</td>
<td>gatcaccatgtctggagag</td>
<td>Segment amplification of 124-nucleotide message</td>
</tr>
<tr>
<td></td>
<td>msg-divergent-rev</td>
<td>tctgcttggtcctactcc</td>
<td>Segment amplification of 124-nucleotide message</td>
</tr>
<tr>
<td></td>
<td>msg-150-for</td>
<td>ttgccttctaccctgc</td>
<td>Segment amplification of 150-nucleotide message</td>
</tr>
<tr>
<td></td>
<td>msg-150-rev</td>
<td>ttcacatttgtatcctttg</td>
<td>Segment amplification of 150-nucleotide message</td>
</tr>
<tr>
<td></td>
<td>Universal-primer-27F</td>
<td>agagtttgcatcggctcag</td>
<td>Amplification of 16S rRNA</td>
</tr>
<tr>
<td></td>
<td>Universal-primer-1492R</td>
<td>ggtacctgctttccttctt</td>
<td>Amplification of 16S rRNA, and for reverse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>transcription with SSIV</td>
</tr>
<tr>
<td>pMS2-TorA</td>
<td>ms2-for-seq-primer-1</td>
<td>taacgatcttttcagcatcag</td>
<td>Sequencing</td>
</tr>
<tr>
<td></td>
<td>ms2-for-seq-primer-2</td>
<td>gcagcgctgcagttcatcag</td>
<td>Sequencing</td>
</tr>
<tr>
<td></td>
<td>ms2-for-seq-primer-3</td>
<td>gcaccccgagcatcatcag</td>
<td>Sequencing</td>
</tr>
<tr>
<td></td>
<td>ms2-rev-seq-primer-1</td>
<td>cgaacaaggtcctcagtgtc</td>
<td>Sequencing</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
<td>Library Step</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>ms2-rev-seq-primer-2</td>
<td>ggtattgtctcatgagcggatac</td>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td>puc19-ms2-gibson-for</td>
<td>ttctaaatctcatgtgtaaccttggaagacaaaggcctegtgatacgcg</td>
<td>Backbone amplification for Gibson assembly</td>
<td></td>
</tr>
<tr>
<td>puc19-ms2-gibson-rev</td>
<td>cctagagctgagctagcgtcaagctgacc tgagggcctagcag</td>
<td>Backbone amplification for Gibson assembly</td>
<td></td>
</tr>
<tr>
<td><strong>GFP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>promoter-rbs-gblock</td>
<td>tgtgacgcctagctagctcttagtgtagctgtcgtagctctagagataagagaaatcctaatggcgactctatttc</td>
<td>Gibson promoter gBlock</td>
<td></td>
</tr>
<tr>
<td>gfp-insert-for</td>
<td>atgcgttaaggagaagaacttttc</td>
<td>Gibson assembly insert amplification</td>
<td></td>
</tr>
<tr>
<td>gfp-insert-rev</td>
<td>gactctagtagagacgttccagcgcg</td>
<td>Gibson assembly insert amplification</td>
<td></td>
</tr>
<tr>
<td>gfp-insert-terminator-</td>
<td>ggtggtcagctcttcttagtgtagctgtcgtagctctagagataagagaaatcctaatggcgactctatttc</td>
<td>Gibson assembly terminator insert</td>
<td></td>
</tr>
<tr>
<td>gblock</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>puc19-gfp-gibson-for</td>
<td>gacgaaagggcctctgtgatacgcc</td>
<td>Backbone amplification for Gibson assembly</td>
<td></td>
</tr>
<tr>
<td>puc19-gfp-gibson-rev</td>
<td>gcagccatgcaagcttgcc</td>
<td>Backbone amplification for Gibson assembly</td>
<td></td>
</tr>
<tr>
<td>gfp-for-seq</td>
<td>cacacaggaaacagctgtag</td>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td><strong>pMS2TorA-msg-150</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ms2-bumper-for</td>
<td>cggcgcatgtcgtactatgcgtcgaagacgtactt caccagcttcttcggt</td>
<td>For the insertion of 2 amino acids between TorA signal peptide and MS2 protein</td>
<td></td>
</tr>
<tr>
<td>ms2-bumper-rev</td>
<td>gtaggaatttagacgcccatagatgcagatgacgcgccgactt gcgcctttgcgcgcgcg</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**pKil sequence**

atgcgtaaacgtttttttgtgaggtgttgagttttggattgtaacgccatctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt...
Supplementary Figure 2. Plasmid map of pMS2TorA-msg-124.
Supplementary Figure 3. Plasmid map of pMS2TorA-msg-150.

pTorA-GFP sequence

gacgaaagggcctgctgatacgcctatttttataggttaatgtcatgatataatggtttcttagacgtcaggtggcacttttcggggaaatgtgccggaaccctattgtttttttcataacttcataatctgtgctgtaaatgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgcticaaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggcttacacttatgcttccggetcgtatgttgtggaattgtgagcggataaacaatttcacacaggaaacagctagtaccatgattacgccaagcttgcatgcctgcaggtcgacttgacggctagctcagtcctaggtacagtgctagctctagagaaagaggagaaatactagatgacaataacgatctctttcaggcatcacgtcggcgttttctggcacaactcggcggcttaaccgtcgccgggatgctggggccgtcattgttaacgccgcgacgtgcgactgcggcgcaagcggcgcatatgcgtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaatttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttattgcactactggaaaactacctgttccatggccaacactgtcactactttcggttatggtgttcaatgctttgcgagatacccagatcatatgaaacagcatgactttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatattttgtaacactcattagggggggggacgtgtcttttttcatctcactcttccagtggtagctctgcttcatcttctcccagaagtgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg
Supplementary Figure 4. Plasmid map of pTorAGFP.

Supplementary Figure 5. Example of qPCR amplification curves collected. Specifically, these are the curves resulting from the amplification of cDNA synthesized from cell pellets containing pMS2TorA-msg-150.
Supplementary Figure 6. Fluorescence analysis of cells whose GFP transcript were activated by extracellular RNA. PBS has little to no fluorescence (a), while unaltered cells (b) and control cells with no incubation using RNA (c) had dim autofluorescence. Cells incubated with extracellular RNA were appreciably brighter (d).

Table 4. Absorbance of cells containing a GFP switch, incubated with RNA triggers.

<table>
<thead>
<tr>
<th>Sample contents</th>
<th>Absorbance values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PBS</strong></td>
<td>6.719</td>
</tr>
<tr>
<td><strong>Control (no RNA incubation) for RNA trigger 1</strong></td>
<td>44.343</td>
</tr>
<tr>
<td><strong>Control (no RNA incubation) for RNA trigger 2</strong></td>
<td>44.583</td>
</tr>
<tr>
<td><strong>Control (no RNA incubation) for RNA trigger 3</strong></td>
<td>47.829</td>
</tr>
<tr>
<td><strong>Control (no RNA incubation) for RNA trigger 4</strong></td>
<td>44.788</td>
</tr>
<tr>
<td><strong>Control (no RNA incubation) for RNA trigger 5</strong></td>
<td>46.436</td>
</tr>
<tr>
<td><strong>Control (no RNA incubation) for RNA trigger 6</strong></td>
<td>42.593</td>
</tr>
<tr>
<td><strong>Control (no RNA incubation) for RNA trigger 7</strong></td>
<td>43.125</td>
</tr>
<tr>
<td><strong>Cells with RNA trigger 1</strong></td>
<td>54.428</td>
</tr>
<tr>
<td><strong>Cells with RNA trigger 2</strong></td>
<td>49.922</td>
</tr>
<tr>
<td><strong>Cells with RNA trigger 3</strong></td>
<td>49.473</td>
</tr>
<tr>
<td><strong>Cells with RNA trigger 4</strong></td>
<td>49.17</td>
</tr>
<tr>
<td><strong>Cells with RNA trigger 5</strong></td>
<td>50.557</td>
</tr>
<tr>
<td><strong>Cells with RNA trigger 6</strong></td>
<td>44.889</td>
</tr>
<tr>
<td><strong>Cells with RNA trigger 7</strong></td>
<td>44.993</td>
</tr>
</tbody>
</table>
REFERENCES

38. Pavankumar, A. R., Ayyappasamy, S. P. & Sankaran, K. Small RNA fragments in


