Repurposing the mammalian RNA-binding protein Musashi-1 as an allosteric translation repressor in bacteria

Roswitha Dolcemascolo[1,2,#], María Heras-Hernández[1,#], Lucas Goiriz[1,3,#], Roser Montagud-Martínez[1,2], Alejandro Requena-Menéndez[1], Rosa Márquez-Costa[1,2], Raúl Ruiz[1], Anna Pérez-Ràfols[4,5], R. Anahí Higuera-Rodríguez[6,7], Guillermo Pérez-Ropero[8,9], Wim F. Vranken[10,11], Tommaso Martelli[4], Wolfgang Kaiser[6], Jos Buijs[8,12], and Guillermo Rodrigo[1,*]

[1]Institute for Integrative Systems Biology (I2SysBio), CSIC – University of Valencia, 46980 Paterna, Spain

[2]Department of Biotechnology, Polytechnic University of Valencia, 46022 Valencia, Spain

[3]Department of Applied Mathematics, Polytechnic University of Valencia, 46022 Valencia, Spain

[4]Giotto Biotech SRL, 50019 Sesto Fiorentino, Italy

[5]Magnetic Resonance Center (CERM), Department of Chemistry Ugo Schiff, Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine (CIRMMP), University of Florence, 50019 Sesto Fiorentino, Italy

[6]Dynamic Biosensors GmbH, 82152 Planegg, Germany

[7]Department of Physics, Technical University of Munich, 85748 Garching, Germany.

[8]Ridgeview Instruments AB, 75237 Uppsala, Sweden

[9]Department of Chemistry – BMC, Uppsala University, 75123 Uppsala, Sweden.

[10]Structural Biology Brussels, Vrije Universiteit Brussel, 1050 Brussels, Belgium

[11]Interuniversity Institute of Bioinformatics in Brussels, Université Libre de Bruxelles – Vrije Universiteit Brussel, 1050 Brussels, Belgium

[12]Department of Immunology, Genetics, and Pathology, Uppsala University, 75185 Uppsala, Sweden

[#]Equal contribution to this work

[*]Correspondence: guillermo.rodrigo@csic.es

Keywords: Binding kinetics; Dynamic systems; Genetic circuits; Mathematical modeling; Post-transcriptional regulation; RNA recognition motif; Synthetic biology.

ABSTRACT

The RNA recognition motif (RRM) is the most common RNA-binding protein domain identified in nature. However, RRM-containing proteins are only prevalent in eukaryotic phyla, in which they play central regulatory roles. Here, we engineered an orthogonal post-transcriptional control system of gene expression in the bacterium *Escherichia coli* with the mammalian RNA-binding protein Musashi-1, which is a stem cell marker with neurodevelopmental role that contains two canonical RRMs. In the circuit, Musashi-1 is regulated transcriptionally and works as an allosteric translation repressor thanks to a specific interaction with the leader coding region of a messenger RNA and its structural plasticity to respond to fatty acids. We fully characterized the genetic system at the population and single-cell levels showing a significant fold change in reporter expression, and the underlying molecular mechanism by assessing the *in vitro* binding kinetics and *in vivo* functionality of a series of RNA mutants. Moreover, the dynamic response of the system was well recapitulated by a bottom-up mathematical model. This work illustrates how RRM-based regulation can be adapted to simple organisms, thereby paving the way for engineering more complex circuits in prokaryotes by combining transcription and translation control with proteins.

INTRODUCTION

Gene regulation at the post-transcriptional level is pervasive in living organisms of ranging complexity [1-4]. Indeed, the ability to regulate the genetic information flow at different points appears instrumental to maximize the integration of intrinsic and extrinsic signals, which enables an efficient information processing by the organisms. However, the solutions implemented in prokaryotes and eukaryotes greatly differ. In prokaryotes, small RNAs (sRNAs) regulate messenger RNA (mRNA) stability and translation initiation [1], supported by a series of RNA-binding proteins that act globally [2]. Regulatory proteins of specific scope in these simple organisms mainly operate in the transcriptional layer [5], what is aligned with the models presented in the early times of molecular biology [6]. By contrast, eukaryotes deploy a sizeable number of RNA-binding proteins with a variety of functions [4] that participate in the regulation of mRNA turnover, transport, splicing, and translation in a gene-specific manner and also at a global scale. In animals, in particular, most RNA-binding proteins contain RNA recognition motifs (RRMs) [7]. RRMs are small globular domains of about 90 amino acids that fold into four antiparallel β-strands and two α-helices, which can bind to single-strand RNAs with sufficient affinity and specificity to control biological processes [8]. Yet, while important to attain functional diversity in the post-transcriptional layer in animals, RRMs are not prevalent in all organisms. In fact, the scarcity of RRM-containing proteins in prokaryotes and the often-unknown functional role of those identified by bioinformatic methods [9] question if RRMs can readily work in organisms with much simpler gene expression machinery and intracellular organization. If so, this would raise the potential to use RRM-RNA interactions as an orthogonal layer to engineer gene regulation in prokaryotes.

To address these intriguing questions, we adopted a synthetic biology approach where a specific RRM-containing protein was incorporated in a bacterium in order to engineer a post-transcriptional control module. Synthetic biology has highlighted how living

77 cells can be (re)programmed through the assembly of independent genetic elements into
78 functional networks for a variety of applications in biotechnology and biomedicine [10]. Yet,
79 synthetic biology can also be used to disentangle natural systems and to probe
80 hypotheses about biological function [11]. In previous work, some proteins with the ability
81 to recognize RNA have been exploited as translation factors in bacteria for a gene-specific
82 regulation [12-14]. The first instance was the tetracycline repressor protein (TetR), which
83 naturally functions as a transcription factor, by means of the selection of synthetic RNA
84 aptamers [12]. The bacteriophage MS2 coat protein (MS2CP) [13] and eukaryotic Pumilio
85 homology domains [14] were also used in synthetic circuits. Alternatively, a wide palette of
86 post-transcriptional control systems based on sRNAs have been developed in recent years
87 to program gene expression in bacteria [15]. Of note, these systems are amenable to be
88 combined with regulatory proteins to attain complex dynamic behaviors [16]. A
89 heterologous RRM-containing protein with definite regulatory activity, in addition to provide
90 empirical evidence on the adaptability of such RNA-binding domains to different genetic
91 backgrounds, would enlarge the synthetic biology toolkit [17], boosting applications in
92 which high orthogonality, expression fine-tuning, and signal integrability are required
93 features. In addition, RRMs can themselves be allosterically regulated, opening up new
94 avenues for post-transcriptional regulation by small molecules.

95 In this work, the mammalian RNA-binding protein Musashi-1 (MSI-1) [18] was used
96 as a translation repressor in the bacterium *Escherichia coli* (**Fig. 1a**). MSI-1 belongs to an
97 evolutionarily conserved family of RRM-containing proteins, of which a member was first
98 identified in *Drosophila melanogaster* [19]. MSI-1 contains two RRMs in the N-terminal
99 region (RRM1 and RRM2) and recognizes the RNA consensus sequence RU_nAGU on the
100 nanomolar affinity scale [20]. Importantly, MSI-1 can be allosterically inhibited by fatty
101 acids (in particular, 18-22-carbon ω-9 monounsaturated fatty acids) [21]. In mammals,
102 MSI-1 is mainly expressed in stem cells of neural and epithelial lineage and plays crucial

103 roles in differentiation, tumorigenesis, and cell cycle regulation [18]. Notably, MSI-1
104 regulates Notch signaling by repressing the translation of a key protein in the pathway [20].
105 Hence, rather than moving genetic elements from simple to complex organisms, as it is
106 normally done (*e.g.*, the TetR-aptamer module was implemented in simple eukaryotes
107 [22]), we reversed the path by moving an important mammalian gene (from *Mus musculus*)
108 to *E. coli*. Some eukaryotic factors have already been implemented in bacteria to regulate
109 gene expression at different levels [14,23], but the case of RRM-containing proteins has
110 remained elusive. In the following, we present quantitative experimental and theoretical
111 results on the response dynamics of a synthetic gene circuit in which MSI-1 works as an
112 allosteric translation repressor. There, MSI-1 is transcriptionally controlled by the lactose
113 repressor protein (LacI), and translation regulation by MSI-1 is accomplished by means of
114 a specific interaction with an mRNA (encoding a reporter protein) that harbors a suitable
115 binding motif in its leader coding region.
116

RESULTS

A Musashi protein can down-regulate translation in bacteria. From the amino acid sequence of *M. musculus* MSI-1, we generated a nucleotide sequence with codons optimized for *E. coli* expression. Knowing that the C-terminus of MSI-1 is of low structural complexity [24], we cloned a truncated version of the gene encompassing the first 192 amino acids, which include the two RRMs, to implement our synthetic circuit (**Fig. 1b**). The resulting protein (termed MSI-1*) was expressed from a synthetic PL-based promoter repressed by LacI (termed PLlac) [25] lying in a high copy number plasmid. This allowed controlling the expression of the heterologous RNA-binding protein at the transcriptional level with lactose or isopropyl β-D-1-thiogalactopyranoside (IPTG) in a genetic background over-expressing LacI. As a regulated element, we used the superfolder green fluorescent protein (sfGFP) [26], which was expressed from a constitutive promoter (J23119) lying in a low copy number plasmid (**Suppl. Fig. 1**). An RNA motif obtained by affinity elution-based RNA selection (SELEX) containing two copies of the consensus recognition sequence (*viz*., GUUAGU and AUUUAGU) [20] was placed in frame after the start codon of *sfGFP*. This motif folds into a stem-loop structure that allows stabilizing the exposure of the recognition sequence to the solvent. In this way, MSI-1* can block the progression of the ribosome on the regulated gene in the initial phase. This mode of action differs from the natural one in mammals, in which MSI-1 binds to the 3' untranslated region of its target mRNA (*Numb*) to repress translation by disrupting the activation function of the poly(A)-binding protein [27]. Here, considering lactose (or IPTG) and oleic acid as the two inputs and sfGFP as the output, MSI-1* being an internal allosteric regulator operating at the post-transcriptional level, an IMPLY gate would model the logic behavior of the resulting circuit (*i.e*., sfGFP would only turn off with lactose and without oleic acid in the medium).

We first characterized by bulk fluorometry the dose-response curve of the system using a lactose concentration gradient up to 1 mM. Our data show that MSI-1* down-

143 regulated sfGFP expression by 2.5-fold (**Fig. 1c**). Fitting a Hill equation, we obtained a
144 regulatory coefficient of 99 μM (lactose concentration at which the repression is half of the
145 maximal) and a Hill coefficient of 1.7 (**Suppl. Note 1**). We also observed that IPTG (a
146 synthetic compound) triggered a very similar response. To further inspect the activity of the
147 RNA-binding protein, we filtered out the transcriptional regulatory effect. For that, we
148 expressed the enhanced blue fluorescent protein 2 (eBFP2) [28] from the PLlac promoter
149 to obtain the corresponding dose-response curve with lactose. In this way, eBFP2
150 expression was a proxy of MSI-1* expression, which allowed representing the transfer
151 function of the engineered regulation (**Fig. 1d**). A Hill equation with no cooperative binding
152 (*i.e.*, Hill coefficient of 1) explained the data with sufficient agreement, suggesting that only
153 one protein interacted with a given mRNA (*i.e.*, each RRM of MSI-1* binds to a consensus
154 sequence repeat, in agreement with a previous structural model [24]). We also measured
155 the cell growth rate for all induction conditions, finding that the values were almost
156 constant. This indicates that the expression of the mammalian protein did not produce a
157 significant burden to the bacterial cell.

158 In simple terms, protein expression comes from the product of the transcription and
159 translation rates of the gene. Hence, we examined such a decomposition in the case of
160 sfGFP expression regulated by MSI-1*. Of note, the low copy number plasmid harbors an
161 additional transcriptional unit to express the monomeric red fluorescent protein mScarlet
162 [29] from a constitutive promoter (J23119). We then monitored its expression profile with
163 lactose. Assuming that *sfGFP* and *mScarlet* were equally transcribed, as they were
164 expressed from the same promoter, and that the translation rate of *mScarlet* was constant,
165 the product of mScarlet expression and cell growth rate was considered a proxy of the
166 transcription rate of *sfGFP*. Moreover, the ratio of sfGFP and mScarlet expressions was a
167 proxy of the translation rate of *sfGFP* [30]. This served us to represent the dynamics of the
168 system in a plane defined as translation rate *vs.* transcription rate (termed Crick space

169 [31]), highlighting that the change in sfGFP expression with lactose comes indeed from
170 translation regulation (**Fig. 1e**). Finally, to evaluate the heterogeneity of the response
171 within a bacterial population, we performed single-cell measurements of sfGFP expression
172 by flow cytometry. Unimodal distributions able to shift in response to lactose were
173 observed (**Fig. 1f**). Setting a threshold to categorize expression, we found that the
174 percentage of cells in the ON state dropped from 87% to 15% upon addition of 1 mM
175 lactose. In sum, our results show that MSI-1* can regulate translation in a specific manner
176 in *E. coli*, and hence that eukaryotic regulators can be borrowed to be functional elements
177 in prokaryotes.
178
179 **Mechanistic insight into the engineered regulation based on a protein-RNA**
180 **interaction.** We then introduced a series of point mutations into the SELEX RNA motif to
181 assess their effect over the regulatory activity of the RRM-containing protein (**Fig. 2a**).
182 These mutations change the consensus recognition sequence of at least one repeat. A
183 characterization of all systems revealed that the mutations affected both the maximal level
184 and fold change of sfGFP expression (**Fig. 2b**). Of note, a single point mutation in one
185 repeat leading to RU_nCGU (mutant 1) was quite detrimental for the MSI-1*-based
186 regulation (only 1.4-fold reduction in sfGFP expression). This agrees with the prior
187 observation that, within the consensus sequence, UAG is a minimal core that determines
188 the specific recognition by MSI-1 [32]. A double point mutation changing the minimal cores
189 of the two repeats (UAC rather than UAG; mutant 5) also resulted in a detrimental action,
190 but not to a greater extent. We also engineered a new reporter system with a minimal RNA
191 motif consisting of a single copy of the shortest possible consensus sequence (AUAGU),
192 but its characterization showed no apparent regulation by MSI-1* (**Suppl. Fig. 2**). Taken
193 together, two copies of the consensus sequence seem necessary for a successful
194 regulation of protein expression.

195 To relate the cellular effects with protein-RNA interactions, we obtained a purified
196 MSI-1* preparation in order to perform *in vitro* binding kinetics assays (**Suppl. Fig. 3**). For
197 that, a gene coding for a truncated version of the human MSI-1 was expressed from a T7
198 polymerase promoter in *E. coli*. With respect to the *M. musculus* version, this protein only
199 differs in one residue of RRM2, which is the subsidiary domain for RNA recognition (note
200 also that the human and mouse proteins recognize the same consensus sequence [32]).
201 To avoid the necessity of labelling the molecules of interest and allow working with very
202 low amounts of protein and RNA, we used the switchSENSE technology, which allows
203 measuring molecular dynamics on a chip (**Fig. 2c**) [33]. **Fig. 2d** summarizes the resulting
204 protein-RNA association and dissociation rates (k_{ON} and k_{OFF}, respectively; see also
205 **Suppl. Fig. 4**). In the case of the original RNA motif, we found an association rate of 1.1
206 $nM^{-1}min^{-1}$, which means that a single regulator molecule would take 1-3 min to find its
207 target in the cell, and a residence time of the protein on the RNA of 1.5 min (given by
208 $1/k_{OFF}$). Of note, the reported value of k_{ON} is relatively close to the upper limit imposed by
209 the diffusion rate (~1 $nM^{-1}s^{-1}$). This fast rate suggests that MSI-1* is able to find its target
210 mRNA in *E. coli*, competing with ribosomes and ribonucleases, and then achieve
211 translation regulation. We also found that a single mutation in one of the two UAG minimal
212 cores (mutants 1 and 2) led to similar association but faster dissociation (almost 4 times
213 faster dissociation), whereas a double mutation affecting the two cores (mutant 5)
214 disturbed both phases (almost 15 times slower association and 10 times faster
215 dissociation). The dissociation constant ($K_D = k_{OFF}/k_{ON}$) was 0.62 nM for the original
216 system, while 87 nM for mutant 5. The switchSENSE technology allowed revealing that
217 affinity on the subnanomolar scale, refining a previous estimate of 4 nM obtained by gel
218 shift assays [20]. To contextualize these values, we compared to the binding kinetics of
219 MS2CP, a phage RNA-binding protein that has evolved in a prokaryotic context and that
220 we recently exploited to study how expression noise emerges and propagates through

221 translation regulation [34]. Previous work disclosed an association rate to the cognate
222 RNA motif of 0.032 $nM^{-1}min^{-1}$ and a residence time of 12 min, leading to a dissociation
223 constant of 2.6 nM [35]. Thus, MSI-1* would target RNA faster than MS2CP, but once this
224 happened the phage protein would remain bound longer.

225 Next, we tried to predict the impact of the mutations on sfGFP expression. On the
226 one hand, we used an empirical free-energy model (RBS calculator) to obtain an estimate
227 of the mRNA translation rate from the sequence [36]. However, only a poor correlation (R^2
228 = 0.16) with the maximal expression level was observed (**Fig. 2e**), suggesting that
229 additional variables should be considered. For example, it was surprising the higher
230 expression level in the case of mutant 4, despite a minimal change in the structure of the
231 RNA motif (**Suppl. Fig. 4a**). On the other hand, when the fold change was correlated with
232 the inverse of the dissociation constant ($1/K_D$, *i.e.*, the equilibrium constant) better results
233 were obtained (R^2 = 0.75; **Fig. 2f**). Mutant 1 is illustrative in this case because, even
234 though a fast association rate was preserved (1.6 $nM^{-1}min^{-1}$), it displayed a marginal
235 regulatory activity as a result of a shorter residence time (0.41 min). This indicates that the
236 underlying protein-RNA interaction in the bacterial circuit was close to thermodynamic
237 equilibrium.

238

239 **A mathematical model captured the dynamic response of the system.** Translation
240 regulation is more challenging than transcription regulation because mRNA is unstable
241 compared to DNA, especially in bacteria. In *E. coli*, in particular, the average mRNA half-
242 life is about 5 min [37]. However, it is possible to derive a common mathematical
243 framework from which to analyze the dynamics of both regulatory modes (**Fig. 3a**). The
244 fold change in protein expression is a suitable mesoscopic parameter that is directly
245 related to the kinetic parameters that characterize the interaction in the cell [38]. Using
246 mass action kinetics, we obtained a general mathematical description of the fold change

247 as a function of the regulator concentration (R), the association and dissociation rates, the
248 leakage fraction of RNA/peptide-chain elongation, and the nucleic acid degradation rate
249 (**Suppl. Note 2**). To visualize the impact of the different parameters, we represented the
250 fold change equation as a heatmap. When there is no nucleic acid degradation (DNA), a
251 linear dependence between the first-order association rate ($k_{ON}R$) and k_{OFF} is established
252 to maintain a given fold change value (**Fig. 3b**), which would correspond to the case of
253 transcription regulation. Accordingly, our model converges to the classical description of
254 fold = 1 + R/K_D. However, if the nucleic acid degrades quickly (mRNA), the dependence
255 between the first-order kinetic rates becomes nonlinear (**Fig. 3c**). Indeed, in the case of
256 translation regulation, it is important to note that when $k_{ON}R$ is lower than the mRNA
257 degradation rate (*i.e.*, the mRNA is degraded faster than the protein binds), the
258 functionality is greatly compromised. To overcome this barrier, the regulator needs to be
259 highly expressed, as MSI-1* is in our system (we estimate R > 1 μM with 1 mM lactose).
260 Furthermore, when the residence time is much longer than the mRNA half-life (*i.e.*, the
261 mRNA is degraded before the protein unbinds), K_D is not a suitable parameter to
262 characterize the regulation, which is solely association-dependent, resulting in non-
263 equilibrium thermodynamics [39]. According to the aforementioned kinetic rates, this would
264 be the case for MS2CP, but not for MSI-1* (*i.e.*, both k_{ON} and k_{OFF} are instrumental to
265 describe the regulation exerted by MSI-1*). Furthermore, given the 2.5-fold down-
266 regulation in our system, we estimated an elongation leakage fraction of 40% (using the
267 fold change equation in the limit $R \rightarrow \infty$). This leakage would come from the ability of
268 ribosomes to elongate even if MSI-1* is bound and their ability to bind sooner to the *sfGFP*
269 mRNA due to a conserved transcription-translation coupling mechanism [40].

270 In addition, we studied the transient response of the gene circuit with lactose, as
271 both MSI-1* and sfGFP expressions changed with time. For that, we quantified the total
272 red fluorescence of the cell population (**Fig. 3d**), which is an estimate of the total number

273 of cells, and the total green fluorescence (**Fig. 3e**), which comes from the composition of
274 population growth and gene regulation. We developed a bottom-up mathematical model
275 based on differential equations to predict sfGFP expression in the cell (**Suppl. Note 3**), as
276 well as a phenomenological model for the bacterial growth (**Suppl. Note 4**). The
277 parameter values were adjusted with the curves without and with 1 mM lactose. Then, we
278 used the mathematical model to predict the transient responses for different intermediate
279 lactose concentrations, finding excellent agreement with the experimental data (R^2 = 0.98;
280 **Fig. 3f**). We also characterized the time-course response of the circuit with IPTG,
281 encountering similar results (**Suppl. Fig. 5**). Moreover, to explore the maintenance of the
282 regulatory behavior when the cell physiology changes, we characterized cells growing in
283 solid medium with a repurposed LigandTracer technology, which initially was developed to
284 monitor molecular interactions in real time [41]. In this case, a significant difference in the
285 total red fluorescence was observed without and with 1 mM IPTG, suggesting that MSI-1*
286 expression was costly for the cell in these conditions. Besides, the total green
287 fluorescence of the growing population was recapitulated using the model with a 2.6-fold
288 down-regulation of cellular sfGFP expression, which is in tune with the results in liquid
289 medium (**Suppl. Fig. 6**). Subsequently, we analyzed the intracellular response. The time-
290 dependent ratio of total green and red fluorescence was used as a proxy of sfGFP
291 expression. A delay in the response is expected because MSI-1* needs to be produced
292 upon addition of lactose [42]. Nevertheless, our model predicted a faster response than
293 experimentally observed (**Fig. 3g**). Overall, this quantitative inspection of translation
294 regulation backs connections between molecular attributes and cellular behavior.
295
296 **Rational redesign of the targeted transcript to enhance the dynamic range of the**
297 **response.** The presence of stem-loop structures in the leader coding region contributes to
298 lower the expression level. The more stable and closer to the start codon, the greater the

299 impact on expression [43]. We hypothesized that, by destabilizing the RNA motif for MSI-1
300 binding, we would obtain an alternative regulatory system with higher expression levels.
301 Accordingly, a new reporter system was engineered removing three base pairs from the
302 stem, maintaining the two consensus recognition sequences. An experimental analysis
303 revealed a 4.9-fold increase of the maximal sfGFP expression level and a 2.0-fold down-
304 regulation with 1 mM lactose (**Fig. 4a**, redesign 1). We then investigated the possibility of
305 increasing the dynamic range of the response by placing three consecutive RNA motifs.
306 However, we did not observe a greater down-regulation with 1 mM lactose (**Fig. 4a**,
307 redesign 2), suggesting that the additional motifs far away from the start codon had no
308 effect; what was noticed is an effect on the maximal expression level. We also designed
309 fusion proteins, with eBFP2 or MS2CP connected with a glycine-serine linker in the C-
310 terminal region of MSI-1*, thereby envisioning a greater ability to interfere with the
311 ribosome due to a bigger size of the resulting protein. However, no suitable clones were
312 obtained despite multiple attempts, stressing the difficulty of recombinant protein
313 expression.

314 As a further strategy to enhance the dynamic range of the response, we redesigned
315 the 5' untranslated region (UTR) of *sfGFP* to accommodate two additional RU_nAGU
316 repeats (*viz*., GUUUAGU and AUUUAGU) flanking the ribosome binding site (RBS),
317 maintaining the original RNA motif after the start codon. Indeed, this is a widespread post-
318 transcriptional regulatory strategy in prokaryotes, as it happens *e.g.* with the MS2 phage
319 replicase [44]. We characterized by bulk fluorometry the dose-response curve of this new
320 system, revealing an 8.6-fold down-regulation of sfGFP expression by MSI-1* (**Fig. 4b**,
321 redesign 3). This was a substantial increase in performance with respect to the 2.5-fold
322 down-regulation of the system shown in **Fig. 1b**. Fitting a Hill equation, we obtained a
323 regulatory coefficient of 86 μM and a Hill coefficient of 4.5 (**Suppl. Note 1**). While the
324 regulatory coefficient was similar than in the original system (99 μM), the Hill coefficient

325 was significantly higher (compared to 1.7). Interestingly, an apparent cooperativity was
326 established between two MSI-1* proteins by binding to adjacent sites. The dynamics of the
327 system was also represented in the Crick space to highlight the change in translation rate.
328 At the single-cell level, we found a 91% of ON cells in the uninduced state that decreased
329 to 5.3% with 1 mM lactose (**Fig. 4c**). Taken together, our data present MSI-1* as a
330 powerful heterologous translation regulator in bacteria.
331
332 **The regulatory activity of a Musashi protein in bacteria can be externally controlled**
333 **by a fatty acid.** The ability of proteins to respond to small molecules is instrumental for
334 environmental and metabolic sensing. Previous work revealed that MSI-1 can be
335 allosterically inhibited by ω-9 monounsaturated fatty acids and, in particular, by oleic acid
336 [21], an 18-carbon fatty acid naturally found in various animal and plant oils (*e.g.*, olive oil).
337 Oleic acid binds to the RRM1 domain of MSI-1 and induces a conformational change that
338 prevents RNA recognition (**Fig. 5a**). To gain insight about the interactions between the
339 elements of our system, we performed gel electrophoretic assays using the purified MSI-1*
340 protein, the RNA motif as a label-free sRNA molecule, and oleic acid. The different mobility
341 of the nucleic acids upon binding to proteins and the coincident staining capacity of nucleic
342 and fatty acids were exploited. We confirmed the MSI-1*-RNA interaction using a protein
343 concentration gradient in this *in vitro* set up (**Suppl. Fig. 7a**), and we found that the
344 interaction was completely disrupted in presence of 1 mM oleic acid (**Fig. 5b**).
345 Furthermore, using an oleic acid concentration gradient, we obtained a half-maximal
346 effective inhibitory concentration of about 0.5 mM (**Suppl. Fig. 7b**).
347 Subsequently, we assessed the effect of oleic acid over the regulatory activity of
348 MSI-1* expressed in *E. coli*. This bacterium has evolved a machinery to uptake fatty acids
349 from the environment. FadL and FadD are two membrane proteins that act as
350 transporters, and FadE is the first enzyme that processes the fatty acid via the β-oxidation

351 cycle [45]. Because of the high turbidity of the cell culture observed in presence of oleic
352 acid, we characterized the system by single-cell measurements of sfGFP expression by
353 flow cytometry. The percentage of cells in the ON state increased from 10% (with 1 mM
354 lactose) to 49% upon addition of 20 mM oleic acid (**Suppl. Fig. 8**). However, the initial
355 93% of ON cells observed in absence of lactose was not recovered. Arguably, oleic acid
356 was partially degraded once it entered the cell. Moreover, a controlled heterogeneity of the
357 response with oleic acid was observed. In particular, we found a dispersion of the
358 unimodal distribution (quantified as the Fano factor, *i.e.*, the ratio between variance and
359 mean) 2.9-fold higher in the case of induction with 0.1 mM lactose than with 1 mM lactose
360 and 20 mM oleic acid, despite having similar mean expression levels. As the former
361 response sensitivity was dominated by transcription regulation and the latter by translation
362 regulation, that result agrees with previous work suggesting that molecular noise is
363 buffered when the control mechanism is post-transcriptional [34].

364 In addition, we investigated this allosteric regulation by imaging the fluorescence of
365 bacterial colonies grown in solid medium with different inducers. In stationary phase, FadE
366 and the rest of oxidative enzymes could be saturated with the fatty acids generated from
367 the membrane degradation [46], oleic acid then having more time to interact with MSI-1*.
368 Notably, we found a substantial inhibition of the repressive action of MSI-1* with 20 mM
369 oleic acid (**Fig. 5c,d**). The system implemented with the redesign-3 reporter displayed an
370 even better dynamic behavior in response to lactose and oleic acid (**Fig. 5e,f**).
371 Conclusively, these results illustrate how the plasticity of RRM-containing proteins (*e.g.*,
372 MSI-1) can be exploited to engineer, even in simple organisms, gene regulatory circuits
373 that operate in an integrated way at the transcriptional, translational, and post-translational
374 levels.

375

376 **DISCUSSION**

377 The successful incorporation of the mammalian MSI-1 protein as a translation factor in *E.*
378 *coli* highlights, in first place, the versatility of RRM-containing proteins to function as
379 specific post-transcriptional regulators in any living cell, from prokaryotes to eukaryotes.
380 Our data show that the protein-RNA association phase is very fast, which is suitable for
381 regulation even in cellular contexts in which RNA molecules are short-lived, such as in *E.*
382 *coli* [37]. Nonetheless, it is important to stress that the kinetic parameters *in vivo* might
383 differ from those measured *in vitro* due to off-target bindings and crowding effects [47].
384 Moreover, our data show that a down-regulation of translation rate up to 8.6-fold can be
385 achieved, with an appropriate design of the target mRNA leader region, and that the
386 engineered cell can sense oleic acid from the environment. Here, the C-terminal low-
387 complexity domain of the native MSI-1 was discarded to create MSI-1* [24], in order to
388 increase solubility, even though this domain might contribute to RNA binding [48].
389 Interestingly, proteins associated to clustered regularly interspaced short palindromic
390 repeats (CRISPR), which belong to the prokaryotic immune system, contain distorted RRM
391 versions [49]. Some CRISPR proteins might have evolved, for example, from an ancestral
392 RRM-based (palm) polymerase after duplications, fusions, and diversification. Noting that
393 the palm domain indeed presents an RRM-like fold [50], we hypothesize that a boost of
394 functionally diverse RRM-containing proteins took place once the polymerases were
395 confined into the nucleus, as the pressure for efficient replication was relieved in the
396 cytoplasm, which would provide a rationale on the unbalance noticed between eukaryotes
397 and prokaryotes [7,51].

398 In second place, our results pave the way for engineering more complex circuits in
399 bacteria with plastic and orthogonal RNA-binding proteins, such as MSI-1, capable of
400 signal multiplexing. Nature is a formidable reservoir of functional genetic material sculpted
401 by evolution that can be exploited to (re)program specific living cells [10]. However, to

402 overcome biological barriers, transgenes usually come from related organisms or cognate
403 parasites, at the cost of limiting the potential engineering. Therefore, efforts to borrow
404 functional elements from highly diverse organisms are suggestive (*e.g.*, regulatory proteins
405 from mammals to bacteria), with the ultimate goal of developing industrial or biomedical
406 applications.

407 Notably, advances in synthetic biology have pushed the bioproduction of a wide
408 variety of compounds in bacteria as a result of a better ability to fine tune enzyme
409 expression [52]. Translation regulation is instrumental to this end because in multiple
410 cases different enzymes are expressed from the same transcriptional unit (*i.e.*, operon).
411 Previous work exploited regulatory RNAs for such a tuning [53], but the use of RNA-
412 binding proteins as translation factors is also appealing. We envision the application of
413 MSI-1* as a genetic tool for metabolic engineering. Furthermore, MSI-1* is able to respond
414 to fatty acids, which are ideal precursors of potential biofuels due to their long hydrocarbon
415 chains. In particular, biofuel in the form of fatty acid ethyl ester, whose bioproduction in *E.*
416 *coli* can be optimized by reengineering the regulation of the β-oxidation cycle with the
417 allosteric transcription factor FadR [54]. Arguably, MSI-1* might be used in place of or in
418 combination with FadR for subsequent developments. However, engineering regulatory
419 circuits for efficient bioproduction is not evident in general as the enzymatic expression
420 levels may require fine tuning, so systems-level mathematical models need to be
421 considered for design along with a wide genetic toolkit for implementation [52]. We
422 anticipate that other animal RRM-containing proteins might be repurposed in *E. coli* as
423 translation factors. Moreover, protein design might be used to reengineer MSI-1* in order
424 to respond to new ligands, maintaining high specificity and affinity for a particular RNA
425 sequence, as previously done with the transcription factor LacI [55].

426 In addition, the Musashi protein family is of clinical importance, as in humans it is
427 involved in different neurodegenerative disorders (*e.g.*, Alzheimer's disease) and some

428 types of cancer [18,56,57]. Therefore, the development of simple genetic systems from
429 which to test protein mutants, potential target mRNAs, decoying RNA aptamers, and
430 inhibitory small molecules in a systematic manner is very relevant. Furthermore, isolating
431 human regulatory elements would help to filter out indirect effects that likely occur in the
432 natural context. This might lead to new therapeutic opportunities. Nevertheless, one
433 limitation of using *E. coli* as a chassis is that some post-translational modifications (PTMs)
434 may be lost, thereby compromising the functionality of the expressed proteins [58].
435 Fortunately, there are metabolic engineering efforts devoted to implement eukaryotic PTM
436 pathways in *E. coli*, such as the glycosylation pathway [59].
437 In conclusion, the functionalization of RRM-containing proteins in bacteria offers
438 exciting prospects, especially as more information becomes available on how individual
439 RRM domains bind to precise RNA sequences, interact with further protein domains, and
440 respond to small molecules through allosteric effects. This work illustrates how synthetic
441 biology, through the rational assembly of heterologous genes and designer *cis*-regulatory
442 elements into circuits, is useful to generate knowledge about the application range of a
443 fundamental type of proteins in nature.
444

445 **METHODS**

446 **Strains, plasmids, and reagents.** *E. coli* Dh5α was used for cloning purposes following
447 standard procedures. To express our genetic circuit for functional characterization, *E. coli*
448 MG1655-Z1 cells (*lacI*$^+$, *tetR*$^+$) were used. This strain was co-transformed with two
449 plasmids, called pRM1+ (KanR, pSC101-E93R ori; leading to ~230 copies/cell) [60] and
450 pREP6 (CamR, p15A ori; leading to ~15 copies/cell). On the one hand, pRM1+ was
451 obtained by cloning a truncated coding region of the *M. musculus* MSI-1 protein (the first
452 192 amino acids, containing the two RRMs; UniProt #Q61474; termed MSI-1*). This gene
453 was under the transcriptional control of the inducible promoter PLlac. On the other hand,
454 pREP6 was obtained by cloning the coding region of sfGFP with an RNA sequence motif
455 recognized by MSI-1. The coding region of mScarlet was also present in the plasmid.
456 These two genes were under the control of the constitutive promoter J23119 in two
457 different transcriptional units. In addition to the original RNA sequence motif, five point-
458 mutated sequences were designed and cloned in pREP6. Additional RNA sequence motifs
459 were cloned in front of *sfGFP* for control experiments (the resulting plasmids were named
460 pREP4, pREP4b, pREP4b3x, and pREP7). In particular, pREP4b3x incorporates three
461 RNA motifs in tandem after the start codon, and pREP7 has two RU_nAGU repeats flanking
462 the RBS and a full RNA motif after the start codon. Suitable genetic cassettes to obtain the
463 final constructions were synthesized by IDT. **Suppl. Table 1** lists all plasmids used in this
464 work. **Suppl. Table 2** presents the nucleotide sequences of the different genetic elements.

465 To perform the dynamic assays with LigandTracer (Ridgeview), *E. coli* BL21(DE3)
466 cells (*lacI*$^+$, *T7pol*$^+$) were used. This strain was also co-transformed with pRM1+ and
467 pREP6. To purify a recombinant Musashi protein, *E. coli* BL21-Gold(DE3) cells (*lacI*$^+$,
468 *T7pol*$^+$) were used. A truncated coding region of the human MSI-1 protein (the first 200
469 amino acids; UniProt #O43347; also termed MSI-1* abusing of notation) was cloned under
470 the control of a T7pol promoter into the plasmid pET29b (KanR, pUC ori).

471 Luria-Bertani (LB) medium was used for the overnight cultures and M9 minimal
472 medium (1X M9 minimal salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.05% thiamine, 0.05%
473 casamino acids, 1% glycerol or 0.4% glucose) for the characterization cultures. M9-
474 glucose medium was only used for real-time fluorescence quantification in liquid medium
475 with IPTG. LB-agar was used for real-time fluorescence quantification in solid medium.
476 Kanamycin and chloramphenicol were used at a concentration of 50 µg/mL and 34 µg/mL,
477 respectively. Lactose and IPTG were used as the inducers of the system (controlling the
478 expression of MSI-1* in *E. coli*) at a concentration of 5, 10, 20, 50, 100, 200, 500, or 1000
479 µM. Oleic acid was used as the allosteric inhibitor of MSI-1* at a concentration of 20 mM in
480 the *in vivo* assays (both in liquid and solid medium). In the *in vitro* assays, oleic acid was
481 used at a concentration of 0.01, 0.1, 0.2, 0.5, 0.7, 1, 1.5, or 2 mM. It was neutralized with
482 NaOH and used in a medium containing 0.5% tergitol NP-40. Compounds provided by
483 Merck.
484
485 **Bulk fluorometry.** Cultures (2 mL) inoculated from single colonies (in triplicate) were
486 grown overnight in LB medium with shaking (220 rpm) at 37 ºC. Cultures were then diluted
487 1:100 in fresh M9 medium (200 µL) with the appropriate inducer (lactose or IPTG). The
488 microplate (96 wells, black, clear bottom; Corning) was incubated with shaking (1300 rpm)
489 at 37 ºC up to 8-10 h (to reach an OD_{600} around 0.5-0.7). At different times, the microplate
490 was assayed in a Varioskan Lux fluorometer (Thermo) to measure absorbance (600 nm),
491 green fluorescence (excitation: 485 nm, emission: 535 nm), and red fluorescence
492 (excitation: 570 nm, emission: 610 nm). To characterize the time-course response of the
493 system, cultures were grown to exponential phase and then diluted before adding the
494 inducer (to minimize the response lag). Mean background values of absorbance and
495 fluorescence, corresponding to M9 medium, were subtracted to correct the signals.
496 Normalized fluorescence was calculated as the slope of the linear regression between

497 fluorescence and absorbance (assuming fluorophore maturation faster than cell doubling
498 time and no proteolytic degradation) [61]. The mean value of normalized fluorescence
499 corresponding to non-transformed cells was then subtracted to obtain a final estimate of
500 expression. In addition, cell growth rate was calculated as the slope of the linear
501 regression between the logarithm of background-subtracted absorbance and time in the
502 exponential phase.
503
504 **Real-time fluorescence quantification in solid medium.** Cultures (2 mL) inoculated
505 from single colonies (in triplicate) were grown overnight in LB medium with shaking (220
506 rpm) at 37 ºC. The overnight culture was plated (15 µL) in areas A and D of a MultiDish
507 2x2 plate (Ridgeview) coated with LB-agar. IPTG was added in areas A and B of the dish
508 at the final concentration of 1 mM. Area C was kept free of cells/inducers as a reference.
509 The dish was then placed in the rotating support of the LigandTracer instrument
510 (Ridgeview) and incubated at 37 ºC for 24 h. The fluorescence from sfGFP and mScarlet
511 was quantified with time in the seeded areas of the dish using the BlueGreen (excitation:
512 488 nm, emission: 535 nm) and OrangeRed (excitation: 568 nm, emission: 620 nm)
513 detectors. The readouts of the opposite parts of the dish were subtracted to correct the
514 signals.
515
516 **Flow cytometry.** Cultures (2 mL) inoculated from single colonies (three replicates) were
517 grown overnight in LB medium with shaking (220 rpm) at 37 ºC. Cultures were then diluted
518 1:100 in fresh LB medium (200 µL) to load a microplate (96 wells, black, clear bottom;
519 Corning) with the appropriate concentrations of lactose (0, 100, 1000 µM) and oleic acid
520 (0, 20 mM). The microplate was then incubated with shaking (1300 rpm) at 37 ºC until
521 cultures reached a sufficient OD_{600}. Cultures (6 µL) were then diluted in PBS (1 mL).
522 Fluorescence was measured in an LSRFortessa flow cytometer (BD) using a 488 nm laser

523 and a 530 nm filter for green fluorescence. Events were gated by using the forward and
524 side scatter signals and compensated (~10^4 events after this process). The mean value of
525 the autofluorescence of the cells was subtracted to obtain a final estimate of expression. Data
526 analysis performed with MATLAB (MathWorks).
527
528 **Purification of a Musashi protein.** Cells were grown in LB medium with shaking at 37 ºC
529 until OD_{600} reached 0.6-0.8. Subsequently, the expression of MSI-1* was induced with 0.5
530 mM IPTG. Cells were incubated at 37 ºC for 4 h and harvested by centrifugation at 7500
531 rpm for 15 min at 4 ºC. The cell pellet was resuspended in a lysis buffer (50 mM Tris-HCl,
532 pH 8.0, 500 mM NaCl, 10% glycerol, with protease inhibitor cocktail), ruptured by
533 sonication, and separated by centrifugation at 30,000 rpm for 35 min at 4 ºC. The soluble
534 fraction was collected and treated with a 5% polyethylenimine solution in order to remove
535 DNA/RNA attached to the protein. Resuspension of the protein was done in 20 mM Tris-
536 HCl, pH 9.0, with protease inhibitor cocktail. Soluble protein was filtered with a 0.22 µm
537 membrane and purified by ion exchange chromatography using an Anion exchange Q FF
538 16/10 column previously equilibrated in alkaline buffer. The protein was collected on the
539 flow-through. The protein was filtered and further purified to homogeneity by size exclusion
540 chromatography using a Hi load 26/60 Superdex 75 pg column previously equilibrated in
541 alkaline buffer with NaCl. The purified fractions were collected and buffer exchange
542 chromatography was performed using a HiPrep 26/10 Desalting column previously
543 equilibrated with the final buffer (20 mM MES, pH 6.0, 100 mM NaCl, 0.5 mM EDTA, with
544 protease inhibitor cocktail). Purification performed at Giotto.
545
546 **Binding kinetics assays of protein-RNA interactions.** Binding experiments of a purified
547 MSI-1* protein against different RNA ligands were performed using the switchSENSE
548 proximity sensing technology [33,62] and a suitable adapter chip on the heliX biosensor

549 platform (Dynamic Biosensors). The adapter chip consists of a microfluidic channel with
550 two gold electrodes functionalized with fluorophore-decorated DNA nanolevers that serve
551 as linkers between the gold surface and the ligand of interest. A constant negative voltage
552 is applied to the electrodes to keep the DNA nanolevers in an upright position. Binding
553 between the injected analyte (MSI-1*) and the ligand attached to the sensor surface (RNA)
554 leads to the alteration of the chemical surrounding of the dye, which results in a
555 fluorescence change. Fluorescence change of the dye in real time describes the binding
556 kinetics of the molecule of interest. Kinetic experiments consisted of a protein association
557 phase (5 min) and a dissociation phase (15 min) in which the chip was rinsed with a buffer
558 (50 mM Tris-HCl, 0.5 mM EDTA, 140 mM NaCl, 0.05% Tween 20, 1 mM TCEP, pH 7.2). A
559 flow rate of 100 µL/min was applied and a sampling rate of 1 Hz was used.

560 Six different RNA ligands (original and 5 mutants) were attached to the 5' end of a
561 generic 48 nt DNA ligand strand, which is part of the DNA linker system on the heliX
562 adapter chip surface. All oligonucleotides were synthesized by Ella Biotech. The ligand
563 strand was hybridized with an adapter strand carrying the fluorophore. Different
564 fluorophores were tested towards their sensitivity for protein-RNA interactions. The green
565 fluorophore Gb showed the most significant signal change. The other half of the adapter
566 strand is complementary to a DNA anchor strand, which is pre-attached to the chip
567 surface. The immobilization of the RNA used a standard functionalization procedure on the
568 heliX device. Kinetic rate constants and affinities were obtained by fitting the experimental
569 data with theoretical binding models implemented in the heliOS software (Dynamic
570 Biosensors). Exponential decay models were used. As a negative control to check for
571 unspecific protein-RNA binding, the single-strand RNA sequence CGGCGCCGC was
572 used (without any binding motif). All data were referenced with a blank run and with the
573 negative control.
574

575 **Gel electrophoresis.** Mobility shift assays with a purified MSI-1* protein and its cognate
576 RNA motif were performed. The RNA motif was generated by *in vitro* transcription with the
577 TranscriptAid T7 high yield transcription kit (Thermo) from a DNA template. It was then
578 purified using the RNA clean and concentrator column (Zymo) and quantified in a
579 NanoDrop. Bovine serum albumin (BSA) was used as a control protein (at 30 μM).
580 Reactions with different combinations of elements were prepared (MSI-1* at 45 μM, RNA
581 at 11 μM, and oleic acid at 1 mM). Reactions with concentration gradients of MSI-1* (from
582 0 to 45 μM) and oleic acid (from 0 to 2 mM) were also performed. Reactions were
583 incubated for 30 min at 37 ºC. Reaction volumes were then loaded in 3% agarose gels
584 prepared with 0.5X TBE and stained using RealSafe (Durviz). Gels ran for 45 min at room
585 temperature applying 110 V. The GeneRuler ultra-low range DNA ladder (10-300 bp,
586 Thermo) was used. This staining served to reveal the RNA and oleic acid (free or in
587 complex with the MSI-1* protein) [63,64]. In addition, gels were soaked for 10 min in the
588 Coomassie blue stain (Fisher) at room temperature with shaking to reveal the proteins.
589 Gels were then soaked in a destaining solution overnight to remove the excess of blue
590 stain. Pictures were taken with the Imager2 gel documentation system (VWR).
591
592 **Microscopy.** LB-agar plates seeded with *E. coli* MG1655-Z1 cells co-transformed with
593 pRM1+ and pREP6 or pREP7 were grown overnight at 37 ºC. Lactose (1 mM) and oleic
594 acid (20 mM) were used as supplements. The plates were irradiated with blue light and
595 images were acquired with a 2.8 Mpixel camera with a filter for green fluorescence in a
596 light microscope (Leica MSV269). The commercial software provided by Leica was used to
597 adjust the visualization of the differential fluorescence among plates. The fluorescence
598 intensity of the colonies was quantified with Fiji [65].
599
600 **Mathematical modeling.** On the one hand, Hill equations were used to empirically model

601 sfGFP expression with lactose/IPTG, eBFP2 expression with lactose, and sfGFP
602 expression with eBFP2 expression (see **Suppl. Note 1** for details). On the other hand, a
603 system of ordinary differential equations was developed to model the dynamic response of
604 the synthetic gene circuit from a bottom-up approach. The system accounted for the
605 intracellular mRNA and protein concentrations, considering a scenario of equilibrium to
606 model both LacI-DNA and MSI-1*-RNA binding (see **Suppl. Note 3** for details). Parameter
607 values were obtained by nonlinear fitting against our experimental data.
608
609 **Molecular visualization *in silico*.** The RMM1 of MSI-1 protein structure determined by
610 nuclear magnetic resonance was downloaded from the UniProt database
611 (www.uniprot.org) [66]. A 3D structure of the RNA motif subsequence involving the two
612 RU_nAGU repeats was predicted with the RNAComposer software [67]. The oleic acid
613 molecule was downloaded from the ChemSpider database (www.chemspider.com). All the
614 molecules were loaded, visualized, colored, trimmed (where necessary), and manually
615 docked using the open source PyMol software (Schrödinger; pymol.org).
616
617 **Resources availability.** The sequences of all genetic elements used in this work are
618 presented in the Supplementary Material. Plasmids available upon request to the
619 corresponding author.

ACKNOWLEDGEMENTS

We thank M. Sattler (TUM) for useful discussions. This work was supported by the grants H2020-MSCA-ITN-2018 #813239 (RNAct) from the European Commission and PGC2018-101410-B-I00 (SYSY-RNA) from the Spanish Ministry of Science and Innovation (co-financed by the European Regional Development Fund). RD, APR, RAHR, and GPR acknowledge each a Marie Curie fellowship. LG was supported by a predoctoral fellowship from the Valencia Regional Government (ACIF/2021/183). RMC was supported by a predoctoral fellowship from the Spanish Ministry of Science and Innovation (PRE2019-088531).

AUTHORS' CONTRIBUTION

GR designed the research. RD, MHH, RMM, ARM, RMC, and RR made the genetic constructions and performed the functional characterization experiments under the supervision of GR. APR purified a Musashi-1 protein under the supervision of TM. RAHR performed the binding kinetics assays under the supervision of WK. RD and GPR performed the dynamic analysis under the supervision of JB and GR. LG developed the mathematical models under the supervision of GR. RD, MHH, LG, and GR analyzed the data and prepared the figures with input of RAHR, GPR, WFV, WK, and JB. GR wrote the manuscript with input of RD, MHH, LG, and WFV. All authors revised the manuscript.

640 **REFERENCES**

641 1. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136: 615-628.

642 2. Holmqvist E, Vogel J (2018) RNA-binding proteins in bacteria. *Nat Rev Microbiol* 16:
643 601-615.

644 3. Jonas S, Izaurralde E (2015) Towards a molecular understanding of microRNA-
645 mediated gene silencing. *Nat Rev Genet* 16: 421-433.

646 4. Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-
647 transcriptional gene regulation. *FEBS Lett* 582: 1977-1986.

648 5. Babu MM, Teichmann SA, Aravind L (2006) Evolutionary dynamics of prokaryotic
649 transcriptional regulatory networks. *J Mol Biol* 358: 614-633.

650 6. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins.
651 *J Mol Biol* 3: 318-356.

652 7. Maris C, Dominguez C, Allain FHT (2005) The RNA recognition motif, a plastic RNA-
653 binding platform to regulate post-transcriptional gene expression. *FEBS J* 272: 2118-
654 2131.

655 8. Messias AC, Sattler M (2004) Structural basis of single-stranded RNA recognition. *Acc*
656 *Chem Res* 37: 279-287.

657 9. Maruyama K, Sato N, Ohta N (1999) Conservation of structure and cold-regulation of
658 RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic
659 glycine-rich RNA-binding proteins. *Nucleic Acids Res* 27: 2029-2036.

660 10. Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. *Nat Rev*
661 *Genet* 11: 367-379.

662 11. Bashor CJ, Collins JJ (2018) Understanding biological regulation through synthetic
663 biology. *Annu Rev Biophys* 47: 399-423.

664 12. Belmont BJ, Niles JC (2010) Engineering a direct and inducible protein-RNA
665 interaction to regulate RNA biology. *ACS Chem Biol* 5: 851-861.

666 13. Katz N, Cohen R, Solomon O, Kaufmann B, Atar O, Yakhini Z, Goldberg S, Amit R
667 (2019) Synthetic 5' UTRs can either up- or downregulate expression upon RNA-
668 binding protein binding. *Cell Syst* 9: 93-106.
669 14. Cao J, Arha M, Sudrik C, Mukherjee A, Wu X, Kane RS (2015) A universal strategy for
670 regulating mRNA translation in prokaryotic and eukaryotic cells. *Nucleic Acids Res* 43:
671 4353-4362.
672 15. Qi LS, Arkin AP (2014) A versatile framework for microbial engineering using synthetic
673 non-coding RNAs. *Nat Rev Microbiol* 12: 341-354.
674 16. Rosado A, Cordero T, Rodrigo G (2018) Binary addition in a living cell based on
675 riboregulation. *PLoS Genet* 14: e1007548.
676 17. Shotwell CR, Cleary JD, Berglund JA (2020) The potential of engineered eukaryotic
677 RNA-binding proteins as molecular tools and therapeutics. *Wiley Interdiscip Rev RNA*
678 11: e1573.
679 18. Fox RG, Park FD, Koechlein CS, Kritzik M, Reya T (2015) Musashi signaling in stem
680 cells and cancer. *Annu Rev Cell Dev Biol* 31: 249-267.
681 19. Nakamura M, Okano H, Blendy JA, Montell C (1994) Musashi, a neural RNA-binding
682 protein required for Drosophila adult external sensory organ development. *Neuron* 13:
683 67-81.
684 20. Imai T, Tokunaga A, Yoshida T, Hashimoto M, Mikoshiba K, Weinmaster G, Nakafuku
685 M, Okano H (2001) The neural RNA-binding protein Musashi1 translationally regulates
686 mammalian numb gene expression by interacting with its mRNA. *Mol Cell Biol* 21:
687 3888-3900.
688 21. Clingman CC, Deveau LM, Hay SA, Genga RM, Shandilya SMD, Massi F, Ryder SP
689 (2014) Allosteric inhibition of a stem cell RNA-binding protein by an intermediary
690 metabolite. *eLife* 3: e02848.
691 22. Ganesan SM, Falla A, Goldfless SJ, Nasamu AS, Niles JC (2016) Synthetic RNA-

692 protein modules integrated with native translation mechanisms to control gene
693 expression in malaria parasites. *Nat Commun* 7: 10727.
694 23. MacDonald CI, Seamons TR, Emmons JC, Javdan SB, Deans TL (2021) Enhanced
695 regulation of prokaryotic gene expression by a eukaryotic transcriptional activator. *Nat*
696 *Commun* 12: 4109.
697 24. Iwaoka R, Nagata T, Tsuda K, Imai T, Okano H, Kobayashi N, Katahira M (2017)
698 Structural insight into the recognition of r(UAG) by Musashi-1 RBD2, and construction
699 of a model of Musashi-1 RBD1-2 bound to the minimum target RNA. *Molecules* 22:
700 1207.
701 25. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in
702 Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.
703 *Nucleic Acids Res* 25: 1203-1210.
704 26. Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and
705 characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24: 79-88.
706 27. Kawahara H, Imai T, Imataka H, Tsujimoto M, Matsumoto K, Okano H (2008) Neural
707 RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G
708 for PABP. *J Cell Biol* 181: 639-653.
709 28. Ai HW, Shaner NC, Cheng Z, Tsien RY, Campbell RE (2007) Exploration of new
710 chromophore structures leads to the identification of improved blue fluorescent
711 proteins. *Biochemistry* 46: 5904-5910.
712 29. Bindels DS, Haarbosch L, Van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier
713 S, Gotthard G, Royant A, Hink MA, Gadella TWJ Jr (2017) mScarlet: a bright
714 monomeric red fluorescent protein for cellular imaging. *Nat Methods* 14: 53-56.
715 30. Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene
716 expression in bacteria. *Cell* 139: 1366-1375.
717 31. Hausser J, Mayo A, Keren L, Alon U (2019) Central dogma rates and the trade-off

718 between precision and economy in gene expression. *Nat Commun* 10: 68.
719 32. Zearfoss NR, Deveau LM, Clingman CC, Schmidt E, Johnson ES, Massi F, Ryder SP
720 (2014) A conserved three-nucleotide core motif defines Musashi RNA binding
721 specificity. *J Biol Chem* 289: 35530-35541.
722 33. Cléry A, Sohier TJ, Welte T, Langer A, Allain FH (2017) switchSENSE: A new
723 technology to study protein-RNA interactions. *Methods* 118: 137-145.
724 34. Dolcemascolo R, Goiriz L, Montagud-Martinez R, Rodrigo G (2022) Gene regulation
725 by a protein translation factor at the single-cell level. *PLoS Comput Biol* 18: e1010087.
726 35. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf
727 WJ (2014) Quantitative analysis of RNA-protein interactions on a massively parallel
728 array reveals biophysical and evolutionary landscapes. *Nat Biotechnol* 32: 562-568.
729 36. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding
730 sites to control protein expression. *Nat Biotechnol* 27: 946-50.
731 37. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN (2002) Global analysis of
732 mRNA decay and abundance in Escherichia coli at single-gene resolution using two-
733 color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* 99: 9697-9702.
734 38. Garcia HG, Phillips R (2011) Quantitative dissection of the simple repression input-
735 output function. *Proc Natl Acad Sci USA* 108: 12173-12178.
736 39. Goiriz L, Rodrigo G (2021) Nonequilibrium thermodynamics of the RNA-RNA
737 interaction underlying a genetic transposition program. *Phys Rev E* 103: 042410.
738 40. Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P (2017) Architecture of a
739 transcribing-translating expressome. *Science* 356: 194-197.
740 41. Björke H, Andersson K (2006) Measuring the affinity of a radioligand with its receptor
741 using a rotating cell dish with in situ reference area. *Appl Radiat Isot* 64: 32-37.
742 42. Rosenfeld N, Alon U (2003) Response delays and the structure of transcription
743 networks. *J Mol Biol* 329: 645-654.

744 43. Paulus M, Haslbeck M, Watzele M (2004) RNA stem-loop enhanced expression of
745 previously non-expressible genes. *Nucleic Acids Res* 32: e78.
746 44. Babitzke P, Baker CS, Romeo T (2009) Regulation of translation initiation by RNA
747 binding proteins. *Annu Rev Microbiol* 63: 27-44.
748 45. Fujita Y, Matsuoka H, Hirooka K (2007) Regulation of fatty acid metabolism in
749 bacteria. *Mol Microbiol* 66: 829-839.
750 46. Navarro-Llorens JM, Tormo A, Martínez-García E (2010) Stationary phase in gram-
751 negative bacteria. *FEMS Microbiol Rev* 34: 476-495.
752 47. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J (2012) The lac
753 repressor displays facilitated diffusion in living cells. *Science* 336: 1595-1598.
754 48. Järvelin AI, Noerenberg M, Davis I, Castello A (2016) The new (dis)order in RNA
755 regulation. *Cell Commun Signal* 14: 9.
756 49. Koonin EV, Makarova KS (2013) CRISPR-Cas: evolution of an RNA-based adaptive
757 immunity system in prokaryotes. *RNA Biol* 10: 679-686.
758 50. Anantharaman V, Iyer LM, Aravind L (2010) Presence of a classical RRM-fold palm
759 domain in Thg1-type 3'-5'nucleic acid polymerases and the origin of the GGDEF and
760 CRISPR polymerase domains. *Biol Direct* 5: 43.
761 51. Koonin EV, Krupovic M, Ishino S, Ishino Y (2020) The replication machinery of LUCA:
762 common origin of DNA replication and transcription. *BMC Biol* 18: 61.
763 52. Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY (2019) Systems metabolic
764 engineering strategies: integrating systems and synthetic biology with metabolic
765 engineering. *Trends Biotechnol* 37: 817-837.
766 53. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY (2013) Metabolic engineering of
767 Escherichia coli using synthetic small regulatory RNAs. *Nat Biotechnol* 31: 170-174.
768 54. Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator
769 system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol*

770 30: 354-359.
771 55. Taylor ND, Garruss AS, Moretti R, Chan S, Arbing MA, Cascio D, Rogers JK, Isaacs
772 FJ, Kosuri S, Baker D, Fields S, Church GM, Raman S (2016) Engineering an
773 allosteric transcription factor to respond to new ligands. *Nat Methods* 13: 177-183.
774 56. Montalbano M, McAllen S, Puangmalai N, Sengupta U, Bhatt N, Johnson OD, Kharas
775 MG, Kayed R (2020) RNA-binding proteins Musashi and tau soluble aggregates
776 initiate nuclear dysfunction. *Nat Commun* 11: 4305.
777 57. Kang MH, Jeong KJ, Kim WY, Lee HJ, Gong G, Suh N, Gyorffy B, Kim S, Jeong SY,
778 Mills GB, Park YY (2017) Musashi RNA-binding protein 2 regulates estrogen receptor
779 1 function in breast cancer. *Oncogene* 36: 1745-1752.
780 58. Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins
781 through bacterial expression systems: a review of the existing biotechnology
782 strategies. *Mol Cell Biochem* 307: 249-264.
783 59. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, Heiss
784 C, Azadi P, Aebi M, DeLisa MP (2012) An engineered eukaryotic protein glycosylation
785 pathway in Escherichia coli. *Nat Chem Biol* 8: 434-436.
786 60. Peterson J, Philips GJ (2008) New pSC101-derivate cloning vectors with elevated
787 copy numbers. *Plasmid* 59: 193-201.
788 61. Leveau JH, Lindow SE (2001) Predictive and interpretive simulation of green
789 fluorescent protein expression in reporter bacteria. *J Bacteriol* 183: 6752-6762.
790 62. Langer A, Hampel P, Kaiser W, Knezevic J, Welte T, Villa V, Maruyama M, Svejda M,
791 Jahner S, Fischer F, Strasser R, Rant U (2013) Protein analysis by time-resolved
792 measurements with an electro-switchable DNA chip. *Nat Commun* 4: 2099.
793 63. Perea W, Greenbaum NL (2020) Label-free horizontal EMSA for analysis of protein-
794 RNA interactions. *Anal Biochem* 599: 113736.
795 64. Fessenden-Raden JM (1972) Effect of fatty acids on the movement and staining of

796 membrane proteins in polyacrylamide gel electrophoresis. *Biochem Biophys Res*
797 *Commun* 46: 1347-1353.
798 65. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch
799 S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K,
800 Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image
801 analysis. *Nat Methods* 9: 676-682.
802 66. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E,
803 Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N,
804 Yeh LS (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res* 33 D154-
805 D159.
806 67. Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz
807 J, Adamiak RW (2012) Automated 3D structure composition for large RNAs. *Nucleic*
808 *Acids Res* 40: e112.
809

FIGURE LEGENDS

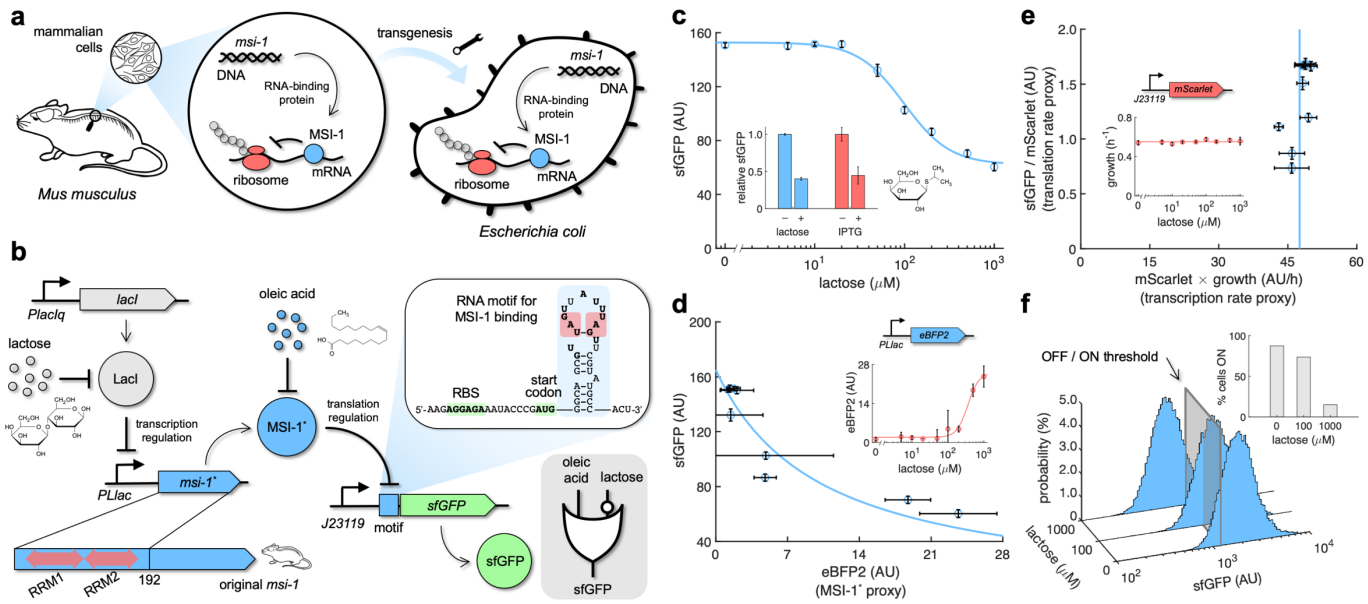

Figure 1: Musashi-1 can down-regulate translation in bacteria.

a) Overview of the biotechnological development. In mammals, MSI-1 binds to the 3' UTR of its target mRNA to repress translation. Here, the *M. musculus* gene coding for MSI-1 was moved to *E. coli* (transgenesis) to implement a synthetic regulation system at the level of translation. **b)** Schematic of the synthetic gene circuit engineered in *E. coli*. A truncated version of MSI-1 (termed MSI-1*) was expressed from the PLlac promoter to be induced with lactose (or IPTG) in a genetic background over-expressing LacI. sfGFP was used as a reporter expressed from a constitutive promoter (J23119) and under the control of a suitable RNA motif recognized by MSI-1* in the leader coding region of the transcript (*viz.*, located after the start codon). The activity of MSI-1* could in turn be allosterically inhibited by oleic acid. In electronic terms, this circuit implements an IMPLY logic gate. The inset shows the predicted secondary structure of the leader region of the reporter mRNA. Within the motif (blue shaded), the consensus recognition sequences (RU_nAGU) are bolded and the minimal cores (UAG) are marked in red. System implemented with pRM1+ and

828 pREP6. **c)** Dose-response curve of the system using lactose as inducer (up to 1 mM).
829 MSI-1* down-regulated sfGFP expression by 2.5-fold. The inset shows the dynamic range
830 of the response using lactose or IPTG (1 mM). **d)** Transfer function of the system (between
831 sfGFP and MSI-1*). The inset shows the dose-response curve of eBFP2 expressed from
832 the PLlac promoter (proxy of MSI-1* expression) with lactose. **e)** Scatter plot of the
833 dynamic response of the system in the Crick space (translation rate *vs*. transcription rate).
834 The dose-response curve of mScarlet expressed from the J23119 promoter with lactose
835 was used to perform the decomposition (vertical line fitted to 48 AU/h). The inset shows
836 the growth rate of the cells for each induction condition (horizontal line fitted to 0.55 h^{-1}). In
837 all cases, points correspond to experimental data, while solid lines come from adjusted
838 mathematical models. Error bars correspond to standard deviations (n = 3). **f)** Probability-
839 based histograms of sfGFP expression from single-cell data for different lactose
840 concentrations. The inset shows the percentage of cells in the ON state (sfGFP
841 expressed), according to a specified threshold, for each lactose concentration. AU,
842 arbitrary units.
843

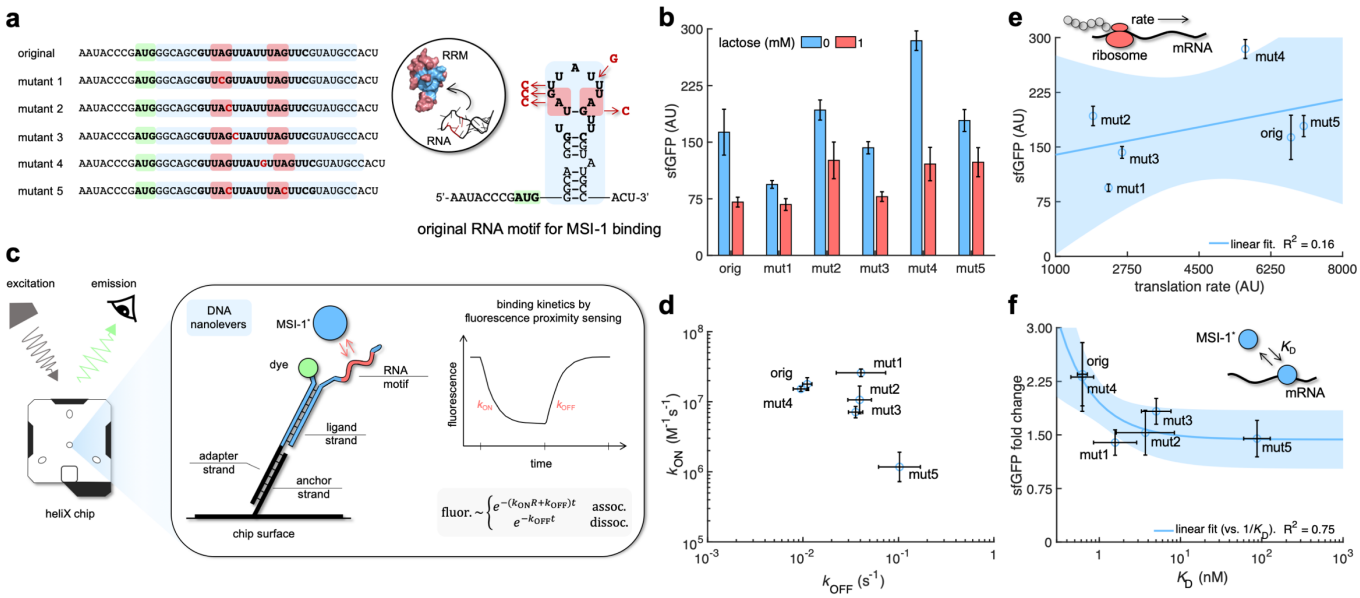

Figure 2: Mechanistic characterization of the Musashi-1-mRNA interaction. **a)** Sequences and predicted secondary structures of the different RNA motif variants for MSI-1 binding analyzed in this work. Point-mutations indicated in red. Three-dimensional representations of the RRM1 and RNA motif are also shown. Within the RRM1, the region that recognizes the RNA is shown in blue. **b)** Dynamic range of the response of the different genetic systems using lactose (1 mM). **c)** Schematics of the heliX biosensor platform. A double-strand DNA nanolever was immobilized on a gold electrode of the chip. The nanolever carried a fluorophore in one end and the RNA motif for MSI-1 binding in the other. Binding between MSI-1* (injected analyte) and RNA led to a fluorescence change, whose monitoring in real time served to extract the kinetic constants that characterize the interaction. **d)** Scatter plot of the experimentally-determined kinetic constants of association and dissociation between the protein and the RNA for all systems (original and 5 mutants). Means and deviations calculated in log scale (geometric). **e)** Correlation between the maximal sfGFP expression level (in absence of lactose) and the translation rate predicted with RBS calculator. Linear regression performed. **f)** Correlation between the fold change in sfGFP expression and the dissociation constant (K_D). Deviations

862 calculated by propagation. Linear regression performed (*vs*. $1/K_D$). Blue shaded areas
863 indicate 95% confidence intervals. In all cases, error bars correspond to standard
864 deviations (n = 3). AU, arbitrary units.
865

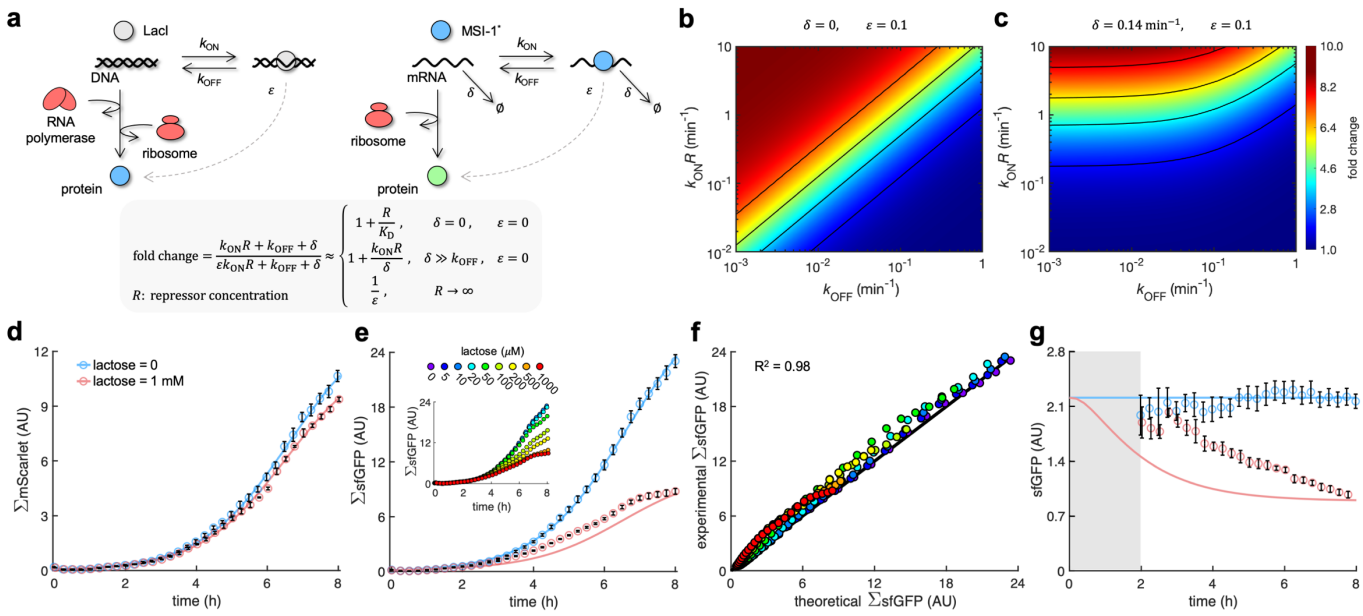

Figure 3: A mathematical model captures the dynamic response of the system. **a)** Schematics of gene regulation at different levels with proteins that bind to nucleic acids (DNA or RNA). On the left, schematic of transcription regulation (*e.g.*, LacI regulating MSI-1* expression). On the right, schematic of translation regulation (*e.g.*, MSI-1* regulating sfGFP expression). A general mathematical expression (grey shaded) was derived to calculate the fold change in protein expression as a function of the regulator concentration (R), the association and dissociation rates (k_{ON} and k_{OFF}), the elongation leakage fraction (ε), and the nucleic acid degradation rate (δ). **b)** Heatmap of the fold change as a function of $k_{ON}R$ and k_{OFF} (*i.e.*, the first-order kinetic rates that characterize the protein-DNA/RNA interaction) when $\delta = 0$ and $\varepsilon = 0.1$. This would correspond to transcription regulation. **c)** Heatmap of the fold change when $\delta = 0.14$ min^{-1} and $\varepsilon = 0.1$. This would correspond to translation regulation. **d)** Total red fluorescence of the cell population (ΣmScarlet) over time without and with 1 mM lactose. In this case, the cell growth rate was fitted to 0.80 h^{-1}. **e)** Total green fluorescence of the cell population (ΣsfGFP) over time without and with 1 mM lactose. The inset shows the dynamic response for different lactose concentrations. **f)**

883 Correlation between the experimental values of ΣsfGFP at different times and for different
884 lactose concentrations and the predicted values from a mathematical model that accounts
885 for population growth and gene regulation. Data for $t > 2$ h. Linear regression performed.
886 **g)** Ratio of total green and red fluorescence as a proxy of cellular sfGFP expression over
887 time. Ratio not represented at early times due to the high error obtained given the low
888 number of cells present in the culture (grey shaded area). Deviations calculated by
889 propagation. In all cases, points correspond to experimental data, while solid lines come
890 from an adjusted mathematical model. Error bars correspond to standard deviations (n =
891 3). AU, arbitrary units.
892

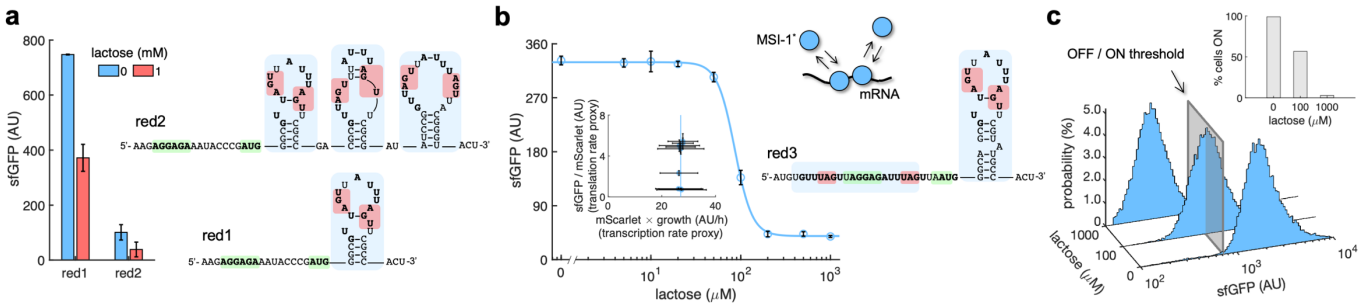

Figure 4: mRNA redesign to enhance the down-regulation by Musashi-1.

a) Dynamic range of the response of two redesigned genetic systems using lactose (1 mM). The predicted secondary structures of the leader regions of the reporter mRNAs are shown on the right. Redesign 1 (red1) was implemented with pREP4b and redesign 2 (red2) with pREP4b3x, which contains three MSI-1 binding sites. These stem-loop structures are less stable than the original one. **b)** Dose-response curve of another redesigned genetic system (redesign 3, red3) using lactose as inducer (up to 1 mM). MSI-1* down-regulated sfGFP expression by 8.6-fold. The inset shows the scatter plot of the dynamic response in the Crick space (translation rate *vs*. transcription rate; vertical line fitted to 27 AU/h). The predicted secondary structure of the leader region of the reporter mRNA containing two MSI-1 binding sites (blue shaded) is shown on the right. In the 5' UTR, the binding site is formed by two RU_nAGU repeats that flank the RBS. In the leader coding region, the binding site is the original one. The minimal cores (UAG) are marked in red. Redesign 3 was implemented with pREP7. Points correspond to experimental data, while the solid line comes from an adjusted mathematical model. In all cases, error bars correspond to standard deviations (n = 3). **c)** Probability-based histograms of sfGFP expression from single-cell data for different lactose concentrations (redesign 3). The inset shows the percentage of cells in the ON state (sfGFP expressed), according to a specified threshold, for each lactose concentration. AU, arbitrary units.

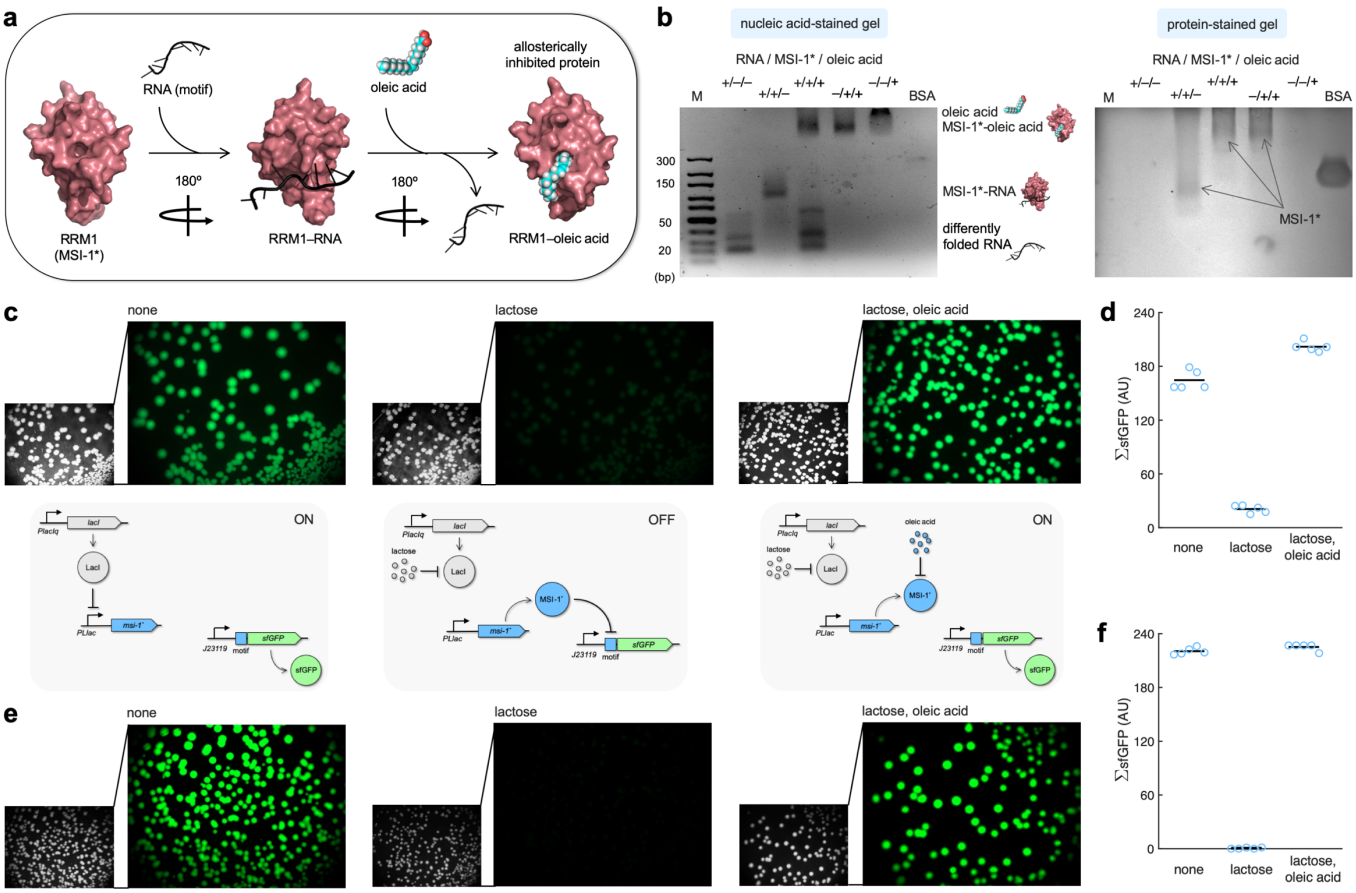

Figure 5: Oleic acid inhibits the regulatory activity of Musashi-1 in bacteria.

a) Three-dimensional structural schematic of the allosteric regulation. RRM1 of MSI-1 is shown alone, in complex with the RNA motif, and in complex with oleic acid. **b)** Gel electrophoretic assay to test the allosteric inhibition of MSI-1* with oleic acid. A purified MSI-1* protein (45 μM), the RNA motif as a label-free sRNA molecule (11 μM), and oleic acid (1 mM) were mixed in a combinatorial way *in vitro*. On the left, nucleic acid-stained gel. On the right, protein-stained gel (Coomassie). The different formed species are indicated. M denotes molecular marker (GeneRuler ultra-low range DNA ladder, 10-300 bp, Thermo). BSA was used as a control. **c)** On the top, images of *E. coli* colonies harboring pRM1+ and pREP6. Bacteria were seeded in LB-agar plates with suitable inducers (1 mM lactose or 1 mM lactose + 20 mM oleic acid). Fluorescence and bright field images are shown. On the bottom, schematics of the working mode of the synthetic gene

929 circuit according to the different induction conditions. **d)** Quantification of the green
930 fluorescence of the colonies from panel c (denoted by ΣsfGFP as it is from populations; *n*
931 = 5). **e)** Images of *E. coli* colonies harboring pRM1+ and pREP7. **f)** Quantification of the
932 green fluorescence of the colonies from panel e. AU, arbitrary units.
933