were reproduced with errors less than 0.01, while the forbidden transitions acquired transition probabilities of less than 0.012, consistent with our interpretation of the $1 \leftrightarrow 3$ interconversion in main text Fig. 4a as insignificantly different from zero.

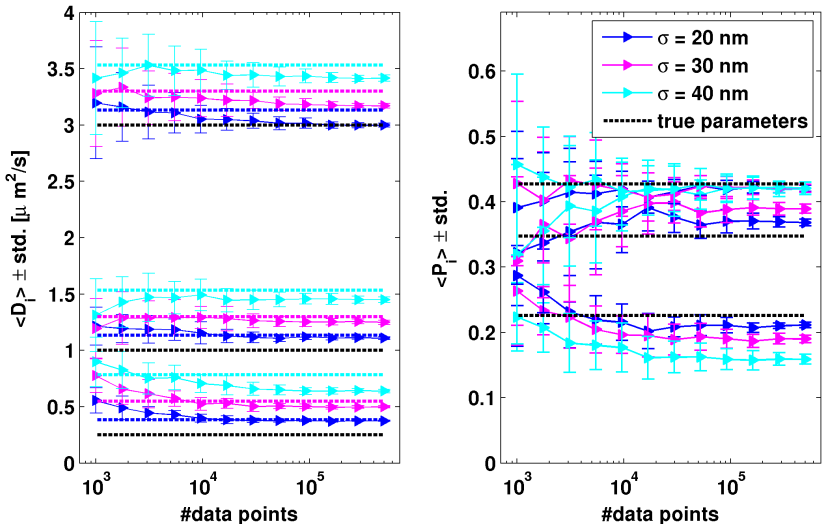

Figure 2: Convergence of diffusion constants and occupation probabilities with increasing amount of data for different localization errors, analyzed and presented as in main text Fig. 2. The dashed lines ($D_{\text{true}} + \sigma^2/\Delta\text{t}$) are colored in agreement with the corresponding diffusion constants for the different position errors.

Synthetic data with varying localization error

To investigate the importance of the localization error, we reran the analysis behind main text Fig. 2 on the same synthetic data, but with localization errors $\sigma = 30$ and 40 nm, in addition to the 20 nm analysis presented in the main text.

The results are shown in Figs. 2-3. An increased localization error increases the apparent diffusion constants by an amount in rough agreement with the simple correction, $D_{\text{true}} + \sigma^2/\Delta\text{t}$, although finite size effects due to the cell geometry, which tend to decrease the apparent diffusion constant, also increase with higher localization error. Low spatial resolution furthermore worsens the occupation probability estimates, and makes the lowest

Nature Methods: doi: 10.1038/nmeth.2367

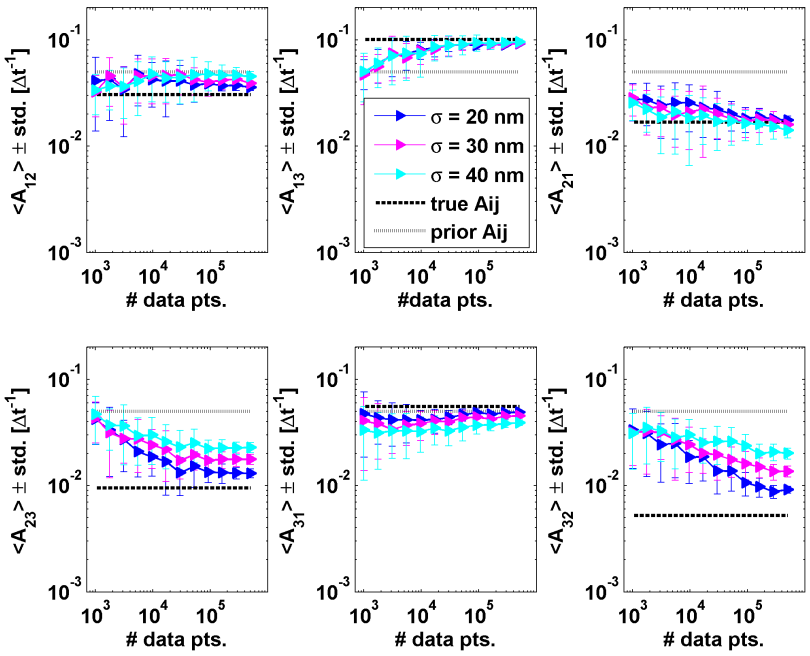

Figure 3: Convergence of transition probabilities with increasing amount of data for different localization errors. Parameters are computed and plotted exactly as in main text Fig. 2.

transition probabilities more biased by the prior probability. This can be expected since the information content of the data decreases.

Synthetic data with varying trajectory length

Another interesting question is the extent to which our analysis suffers from short trajectory lengths, or conversely, how urgent it is to work towards longer trajectories.

In Fig. 4, we extend the analysis of the system presented in main text Figs. 1 and 2, as well as the previous section, to tackle this question. Instead of varying the number of trajectories, we now vary the average trajectory length while keeping the number of diffusion steps constant.

To extract systematic effects rather than statistical fluctuations, we study the full synthetic data set used in Fig. 4, 50 single trajectories with 500 000 data points each. We then subdivide each such trajectory into shorter

Nature Methods: doi: 10.1038/nmeth.2367

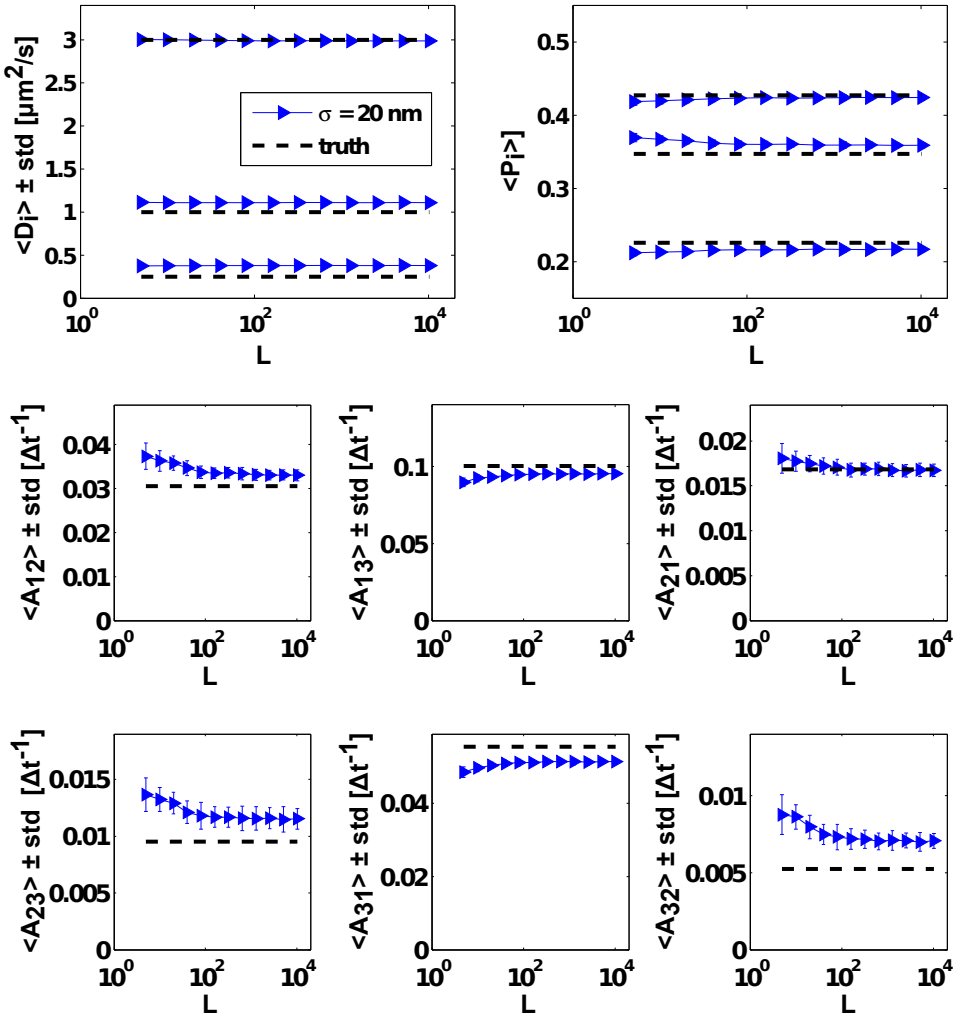

Figure 4: Inferred model parameters as a function of average trajectory length L (in number of data points), for a fixed amount of data corresponding to the largest data set in main text Fig. 2. Parameters are computed and plotted exactly as in main text Fig. 2.

Nature Methods: doi: 10.1038/nmeth.2367

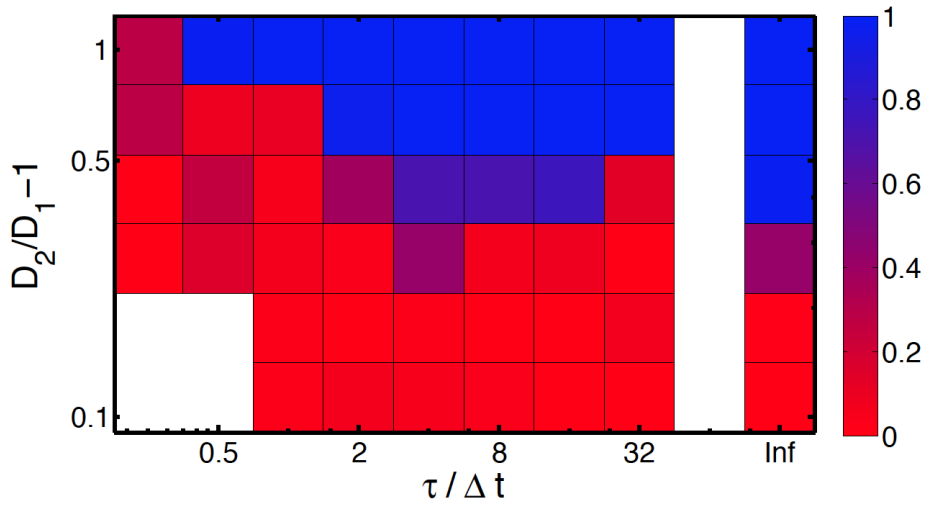

Figure 5: The probability, pBS, of correctly identifying a 2-state model and not a 1-state model as estimated from bootstrap analysis. The probability is estimated for different state lifetimes (x-axis) and different 'distances' between the diffusion constants D_1 and D_2 (y-axis).

pieces according to an exponential distribution with average length L, and add a 20 nm position error. Thus, each point in Fig. 4 is based on the same diffusion trajectories subdivided in different ways (the position errors differs however).

Diffusion constants, which converge fastest as a function of increasing amount of data, are insensitive to trajectory length in this test. The occupation probabilities similarly change only slightly. For the transition probabilities, significant improvements occur only in the rarest transitions, and saturate on trajectory lengths that are a few times the average state life times, indicating that rare events are more easily detected in longer trajectories.

Taken together, these tests indicate that while long trajectories are certainly favorable for inference, many of the difficulties generated by the short trajectories in our data can be compensated for by having a large enough data set. A notable exception is the detection of rare events, which is difficult in short trajectories.

Nature Methods: doi: 10.1038/nmeth.2367

Synthetic data with varying state separation

In Fig. 5 we present the probability of correctly classifying a 2-state system when we vary the 'distance' between the diffusion constants of the two states and the rate of transitions between the states. The plot has been generated with an experimental length distribution of 12130 trajectories and with spatial confinement by the cell geometry. For example, when the state lifetimes are on average ∼10 timesteps, the diffusion constants for the two states have to differ by approximately 50% in order for the two states to be identified. We note that for a 2-state model where the diffusion constants differ a factor of two or more the vbSPT analysis manages to identify two states even when the state lifetime is on average one timestep. For even faster transitions, the algorithm only identifies a single state with the diffusion constant corresponding to the average of the two underlying states.

Experimental data

Here we present tests on our experimental data in order to determine how the vbSPT analysis classify our biological system with a varying amount of experimental data as well as how the choice of prior distributions potentially can influence the result.

State splitting in experimental data

The RNA molecules that interact with Hfq display a broad length distribution and will have a very wide range of diffusion constants. For this reason we expect that the intermediate state observed in the analysis of the individual experiments (see main text Fig. 4) in reality corresponds to a range of diffusive states, grouped as one state by the algorithm due to limited evidence for additional states in the data. In order to test if this is the case we have pooled the data from Exp. 1 and 2, for the untreated cells, and analyzed all 23756 trajectories together (see Tab. 6). We have also added 18903 trajectories from an additional experiment (Exp. 3). In Fig. 6 we have plotted how the scores for 3-, 4- and 5-state models depend on the amount of data in this case. With more than ∼15000 trajectories the algorithm find sufficient evidence to score a 4-state model higher than the 3-state model. The number of trajectories in the individual experiments is near the point where the 4-state model gets the best score. This is also clear from Fig. 7 where we show the bootstrapping results from 100 bootstraps for individual and pooled experiments. In the inset (see Fig. 6) we have also plotted the best 3- and 4-state models in a diffusion - occupancy space. From this plot

Nature Methods: doi: 10.1038/nmeth.2367

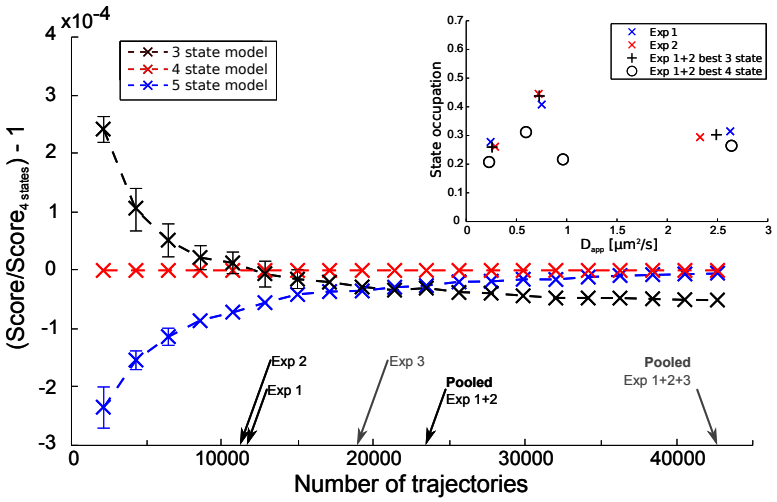

Figure 6: The influence of the amount of data on scores given by the analysis to a 3-, 4- and 5-state model. With sufficient data the analysis can identify an additional state. The data subsets were generated by randomly choosing trajectories from the whole data set for each subset. The errorbars correspond to ± one standard deviation from 20 trials. In the case where the error is within the marker, it has been omitted. For bootstrapping estimates on the probability of ending up in the different models see Fig. 7. **Inset:** The states found by the individual experiments as well as the best 3- and 4-state model for the pooled data (with the 4-state model having the highest score) plotted in an diffusion - occupancy space.

it is clear that it is the intermediate state that is split in two states given sufficient data. We do not observe state splitting when pooling the data from the rifampicin treated cells. This is most likely due to the diminished RNA associated pool of Hfq caused by the global RNA decay.

Choice of prior distributions

The Bayesian formalism offers a possibility to explicitly influence the outcome of the analysis by injecting knowledge about the process that comes from other sources than the data at hand through the specifications of prior distributions. In our case, we strive to minimize this effect, but some priors must still be chosen. Here, we explain and motivate the choices used in this work, and indicate how they can be transferred to other systems, for example ones using other units to measure length and time.

Nature Methods: doi: 10.1038/nmeth.2367

	Best 3 state model	Best 4 state model	Best 5 state model
Exp 1	 pBS = 0.60	 pBS = 0.40	pBS = 0
Exp 2	 pBS = 0.55	 pBS = 0.45	pBS = 0
Exp 3	 pBS = 0.01	 pBS = 0.96	pBS = 0.03
Exp 1+2	 pBS = 0.01	 pBS = 0.94	pBS = 0.05
Exp 1+2+3	 pBS = 0	 pBS = 0.64	pBS = 0.36

Figure 7: The probability of ending up in a 3-, 4- or 5-state model as determined from bootstrapping analysis from 100 bootstraps (pBS) of experimental data sets. For the visualized models the area of each state correspond to its determined occupancy and the dashed transitions represents transition probabilities lower than 0.01 per timestep.

Nature Methods: doi: 10.1038/nmeth.2367

- The **model size** prior $p(N)$ is constant, which is both uninformative and uncomplicated.

- The **diffusion constant** prior is given by an inverse gamma distribution,

$$p(D_j|N) = \frac{\beta_j^{\tilde{n}_j}}{\Gamma(\tilde{n}_j)} D_j^{-(\tilde{n}_j+1)} e^{-\beta_j/D_j}. \quad \beta_j = c_j/(4\Delta t), \tag{1}$$

with mean value and standard deviation

$$\langle D_j \rangle_{\text{prior}} = \frac{c_j}{4(\tilde{n}_j - 1)\Delta t}, \qquad \text{std}[D_j]_{\text{prior}} = \frac{\langle D_j \rangle}{\sqrt{\tilde{n}_j - 2}}. \tag{2}$$

In this case too, choosing an uninformative prior is uncomplicated. We need to choose $\tilde{n}_j, \tilde{c}_j$ such that we get broad distributions with a reasonable mean value. We use $\langle D_j \rangle_{prior} = 1\ \mu\text{m}^2\text{s}^{-1}$ and $\tilde{n}_j = 5$ throughout this work.

- The **initial state probability** vector has a Dirichlet distribution. This is is a multidimensional version of the beta distribution, with density

$$p(\vec{\pi}|N) = \text{Dir}(\vec{\pi}|\tilde{w}_j^{(\vec{\pi})}) = \frac{1}{B(\tilde{w}_j^{(\vec{\pi})})} \prod_j \pi_j^{(w_j^{(\vec{\pi})}-1)}, \quad B(\tilde{w}_j^{(\vec{\pi})}) = \prod_j \frac{\Gamma(w_j^{(\vec{\pi})})}{\Gamma(w_0^{(\vec{\pi})})}, \tag{3}$$

with the constraints $0 \leq \pi_j \leq 1$ and $\sum_j \pi_j = 1$, and $\tilde{w}_0^{(\vec{\pi})} = \sum_k \tilde{w}_k^{(\vec{\pi})}$ [3, 4]. The mean values of the individual components are $\langle \pi_j \rangle_{\text{Dir}} = \tilde{w}_j^{(\vec{\pi})}/\tilde{w}_0^{(\vec{\pi})}$, and we call $\tilde{w}_0^{(\vec{\pi})}$ the strength of the prior. Throughout this work, we use the symmetric choice $\tilde{w}_j^{(\vec{\pi})} = \tilde{w}_0^{(\vec{\pi})}/N$, with $\tilde{w}_0^{(\vec{\pi})} = 5$ for all N.

- The prior for the **transition probabilities** is a little more complicated. Partly because the transition probabilities show a clear bias from the prior values at small data sets in our convergence plots (see main text Fig. 2), and partly because it is not obvious what constitutes a weak prior, as explained below.

Each row of the transition matrix is Dirichlet distributed,

$$p(\mathbf{A}|N) = \prod_j \text{Dir}(A_{j,:}|\tilde{w}_{j,:}^{(\mathbf{A})}), \quad \langle A_{jk} \rangle_{\text{prior}} = \frac{\tilde{w}_{jk}^{(\mathbf{A})}}{\tilde{w}_{j0}^{(\mathbf{A})}}, \quad \tilde{w}_{j0}^{(\mathbf{A})} = \sum_k \tilde{w}_{jk}^{(\mathbf{A})}. \tag{4}$$

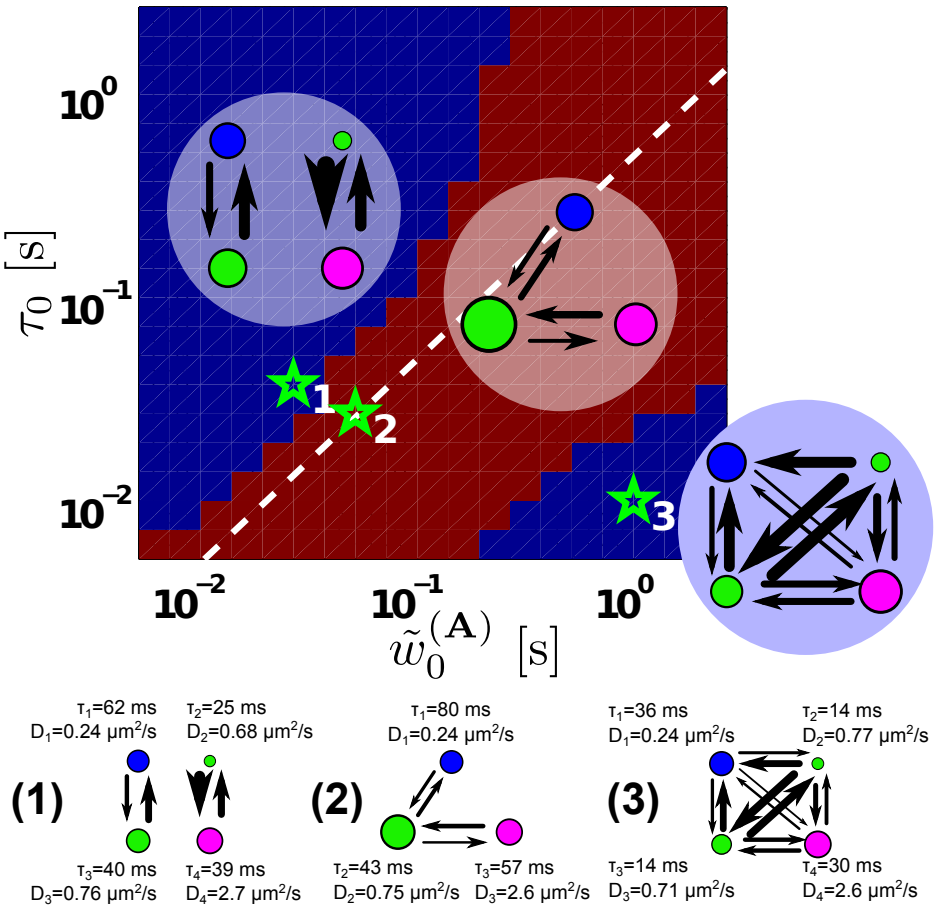

Figure 8: Optimal model for Hfq data without rif treatment (Exp. 1), for a broad range of transition probability prior distributions. The distributions are parameterized by their mean dwell time τ_0 and strength $\tilde{w}_0^{(\mathbf{A})}$. Note the log scales. We find that at extremely high or low strength, compared to the mean dwell time, the green state (with $D \approx 0.7$ μm^{2}s^{-1}) is split up into two states, as indicated by the insets. The white dashed line indicate our choice $\tilde{w}^{(\mathbf{A})} = 2\tau_0$.

Nature Methods: doi: 10.1038/nmeth.2367

We expect the lifetimes of the hidden states to be longer than one timestep, which means that the transition matrix is diagonally dominant. A symmetric choice that reflects this expectation is to choose the matrix $\tilde{w}^{(\mathbf{A})}$ to be the product of a transition matrix, where each state has a mean lifetime τ_0 and all interconversion probabilities are equal, and a row strength $\tilde{w}_{j0}^{(\mathbf{A})}$ (same for all j and N). With τ_0 in units of Δt, this leads to

$$\tilde{w}_{ij}^{(\mathbf{A})} = \begin{cases} \frac{\tilde{w}_0^{(\mathbf{A})}}{\tau_0}(\tau_0 - 1), & i = j, \\ \frac{\tilde{w}_0^{(\mathbf{A})}}{\tau_0(N-1)}, & i \neq j. \end{cases} , \tag{5}$$

with the requirement $\tau_0 > 1$. In this case, the prior choice can matter, and choosing an uninformative value of $\tilde{w}_0^{(\mathbf{A})}$ involves a trade-off. On the one hand, a high value of the strength means that the prior is not uninformative, so that the inferred transition probabilities are significantly biased towards the prior mean values. On the other hand, sample vectors from Dirichlet distribution with very low strengths tend to have most components near zero, which in this case means a bias towards models with few non-zero transitions out of each state.

To illustrate the limitations this places on the prior, we compared 3- and 4-state models from an experimental data set (that of main text Fig. 4a) for a wide range of $(\tau_0, \tilde{w}_0^{(\mathbf{A})})$-pairs shown in Fig. 8. Three different regions can be discerned. When the strength is $\sim 1 - 100$ times the prior mean lifetime τ_0, we get a sequential model (model 2 in Fig. 8), corresponding to the model in main text Fig. 4a. At even higher strengths, the intermediate state splits up into two quickly interconverting states with similar diffusion constants (model 3 in Fig. 8). When the strength is much lower than the mean lifetime, the middle state also splits, but now into two non-interconverting states. This is consistent with the expectation that low prior strength expresses a belief in sparse transition matrices.

Both of these extremes seem unlikely to us, and as an appropriate balance, we suggest $\tilde{w}_0^{(\mathbf{A})} = 2\tau_0$ (the white dashed line in Fig. 8), with $\tau_0 = 10\Delta t$, used throughout this work.

Nature Methods: doi: 10.1038/nmeth.2367

2 Supplementary Note 2. Strain Construction

The λ-Red mediated site-specific recombination [5] technique was used to generate the *hfq-dendra2* strain. The *dendra2* gene coupled to the chloramphenicol resistant cassette was PCR amplified from plasmid pDendra2-cm [6] using primers hfq-dendra fw (GCAGAATACTTCCGCGCAACAGGACAGCGAAGA AACCGAAggcatgagtgcgattaagccagac) and hfq-cm rev (ATCGCTGGCTCCCCGTGTAAAAAAACAGCCCGAAACCttatggtgtgtaggctggagct), containing flanking sites complementary to 3' end of *hfq*. In the resulting open reading frame, *hfq-glycine-dendra2*, the stop codon was replaced by the glycine codon GGC. The PCR product was recombined into strain BW25993 (*Δ(araD-araB)567, λ⁻, rph-1, Δ(rhaD-rhaB)568, hsdR514*). An *hfq* deletion strain was constructed by P1 transduction from MC4100 *hfq::tet* [9] into strain BW25993.

Nature Methods: doi: 10.1038/nmeth.2367

3 Supplementary Note 3. Growth Rate Measurements

Overnight cultures were grown in LB, diluted ∼1:100 (adjusted to the same OD_{600} for all strains included in an experiment) in fresh LB. From the diluted culture 100 µl-aliquots were transferred to 96 well-plates (Corning). Triplicate cultures were prepared for each strain and grown in an Infinite M-600 plater reader (Tecan) for 16 hours at 37°C with shaking (1 minute shaking every fifth minute). Measurements were taken every fifth minute (see Fig. 9).

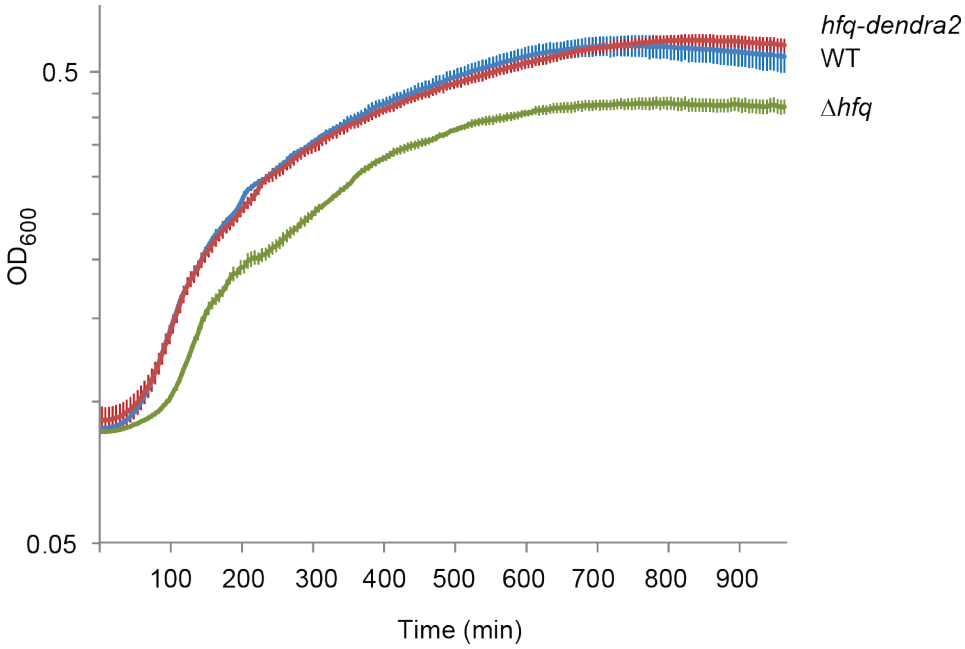

Figure 9: Diagram showing relative growth of three different isogenic strains; wild-type, *hfq*-deletion and *hfq-dendra2* strains. *hfq-dendra2* strain grows as wild-type strain and significantly different from the *hfq*-deletion strain (Δhfq). Cells were grown in LB at 37°C using a 96-well plate reader. An OD_{600} of 0.2 units in the plate reader correspond to approximately 1 OD_{600} standard unit. Error bars represent ± one standard deviation, estimated from triplicates.

Nature Methods: doi: 10.1038/nmeth.2367

4 Supplementary Note 4. Sample Preparation and Microscopy Experiments

Overnight cultures of strain BW25993 *hfq-dendra2* were diluted 1:1000 in fresh LB and grown aerobically at 37°C. At OD_{600}= 0.2, 1 µl of the culture was placed on an 2.5% agarose pad (50111 SeaPlaque GTG Agarose, Lonza) in M9 media supplemented with 0.4% glucose and RPMI amino acid solution (R7131, Sigma). The pad was mounted in a FCS2 growth chamber (Bioptech) as described in ref. [7]. To reduce the global RNA level, rifampicin (200 µg/ml, R3501, Sigma) was added for 15 minutes to the cultures at OD_{600} 0.2 before the cells were placed on rifampicin-containing agarose pads. Subsequently, the growth chamber was mounted on a Nikon Ti-E microscope and single Hfq-Dendra2 proteins was tracked using 2.4 $kWcm^{-2}$ excitation at 555 nm (Crystalaser, GCL-150-555), a high numerical aperture objective (Nikon 100VC PlanApo, 1.4NA) and a back-illuminated cooled EMCCD (Andor Ixon DV-897) at a frame rate of approximately 300 frames per second.

Nature Methods: doi: 10.1038/nmeth.2367

5 Supplementary Note 5. RNA Decay Measurements

Overnight cultures were diluted ∼1:100 in fresh LB and grown aerobically at 37°C. When OD_{600} reached 0.2 rifampicin was added (final concentration of 200 μg/ml). RNA samples were extracted prior to and after rifampicin treatment (1, 2, 4, 8, 16, 32 and 64 minutes after rifampicin addition). RNA was isolated using a hot phenol treatment [8] and real time qPCR was performed as described earlier [9], without normalization against tmRNA (see Fig. 10). The primers used are listed below.

Name	**Sequence (5'→3')**
RT tmRNA 1	GGCAAGCGAATGTAAAGACTGA
RT tmRNA 2	CCGCGTCCGAAATTCCTA
RT lrp 1	CAACGCCGTGCCTTGAG
RT lrp 2	AGCCCTGAATAAACCCTTGTCTT
RT livJ 1	AGCCCTTCCGCTGATTCG
RT livJ 2	CGGAAGGTGTGGCTAACGTT
RT ilvI 1	GGGCTGCCATCAGCAGTT
RT ilvI 2	CAAACAACGGGCAGATTCAA
RT ompA 1	GGGCTGGTCCCAGTACCAT
RT ompA 2	ATGGGTCGGGCCATTGT
RT ompC 1	CGTCTTGGCTTCAAAGGTGAA
RT ompC 2	ATTCCCACTGGCCGTAACC
RT gcvT 1	TGCAAGGGCCGAATGC
RT gcvT 2	CGCTGGGCGTCATTAAACA
RT gltB 1	CATGCGATTGCCAAAGATTATC
RT gltB 2	AAGCCTTTGTTGATGCCGTTA

Nature Methods: doi: 10.1038/nmeth.2367

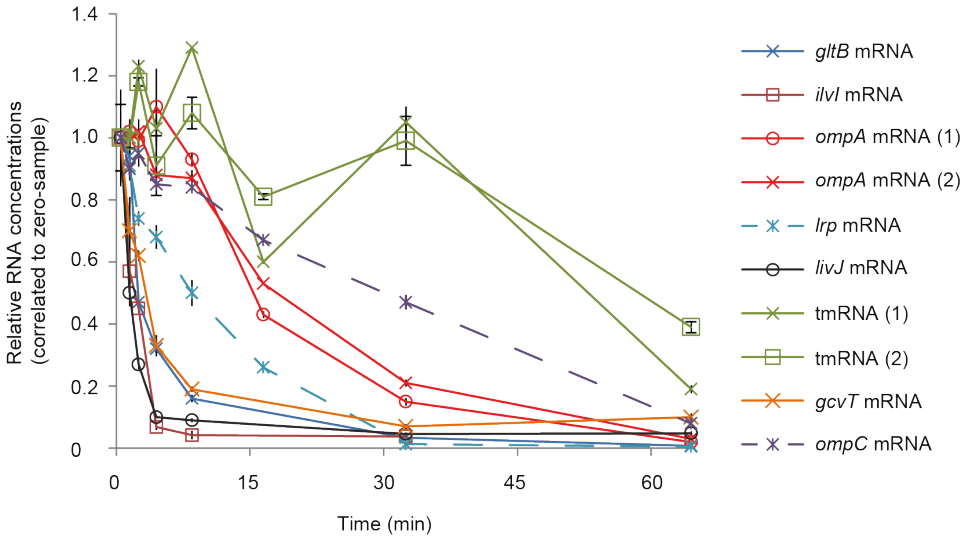

Figure 10: Diagram showing abundances of selected RNA species as a function of time after addition of rifampicin. RNA samples were extracted prior to and after rifampicin addition as specified in Section 5. Real time qPCR was used to measure the relative RNA abundances, error bars represent $\pm$ one standard deviation from qPCR triplicates. The levels for several RNA species decreased radically within the first five minutes, whereas other RNAs, such as *ompC* mRNA, *ompA* mRNA and especially tmRNA were more stable and could be detected a long time after rifampicin addition. *ompA* mRNA (1) and *ompA* mRNA (2) (and tmRNA (1) and (2)) represent two different real time qPCR runs.

Nature Methods: doi: 10.1038/nmeth.2367

References

[1] N G Kampen. *Stochastic processes in physics and chemistry*. Elsevier, Amsterdam Boston London, 3rd ed. edition, 2007.

[2] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[3] D. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, Cambridge UK, 2003.

[4] Matthew Beal. *Variational Algorithms for approximate Bayesian inference*. PhD thesis, University of Cambridge, UK, 2003.

[5] K. A. Datsenko and B. L. Wanner. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, 97(12):6640 – 6645, 2000.

[6] B. P. English, V. Hauryliuk, A. Sanamrad, S. Tankov, N. H. Dekker, and J. Elf. Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proc. Natl. Acad. Sci. U. S. A.*, 108(31):E365 – E373, 2011.

[7] J. Xiao, J Elf, G-W Li, J Yu, and S. Xie. *Imaging Gene Expression in Living Cells at the Single-Molecule Level*. Cold Spring Harbor Laboratory Press, 2008.

[8] P. Blomberg, E. G. Wagner, and K. Nordstrom. Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically in vivo and in vitro by RNase III. *EMBO J.*, 9(7):2331 – 2340, 1990.

[9] E. Holmqvist, J. Reimegard, M. Sterk, N. Grantcharova, U. Romling, and E. G. Wagner. Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J.*, 29(11):1840 – 1850, 2010.

Nature Methods: doi: 10.1038/nmeth.2367