SCIENTIFIC REPORTS
nature research

OPEN

Maximal information transmission is compatible with ultrasensitive biological pathways

Gabriele Micali[1,2,3,4] & Robert G. Endres[1,2*]

Cells are often considered input-output devices that maximize the transmission of information by converting extracellular stimuli (input) via signaling pathways (communication channel) to cell behavior (output). However, in biological systems outputs might feed back into inputs due to cell motility, and the biological channel can change by mutations during evolution. Here, we show that the conventional channel capacity obtained by optimizing the input distribution for a fixed channel may not reflect the global optimum. In a new approach we analytically identify both input distributions and input-output curves that optimally transmit information, given constraints from noise and the dynamic range of the channel. We find a universal optimal input distribution only depending on the input noise, and we generalize our formalism to multiple outputs (or inputs). Applying our formalism to *Escherichia coli* chemotaxis, we find that its pathway is compatible with optimal information transmission despite the ultrasensitive rotary motors.

Biological cells continuously process environmental cues, allowing them to make critical decisions quickly, e.g. whether to move or stay, whether to express a certain protein, or whether to divide[1]. These decisions are generally made based on the level of one or more key proteins, which are the internal representation of the extracellular stimulus. The higher the amount of environmental information encoded in the intracellular representation, the more reliable the response. In contrast, cell-external and internal noise may reduce the reliability. Hence, a biological system under evolutionary pressure is expected to evolve to optimally transmit information under biologically relevant constraints[2], at least when information is a limiting factor[3–5].

To formalize this optimization problem, cells can be considered input-output devices, where stimuli of extracellular concentrations are the input, receptors (or the entire pathway) are the communication channel, and the intracellular concentration of a key protein (or the final behavior of the cell) is the output. In contrast to engineered physical systems, the distinction between input, channel, and output is not always clear in biological systems with feedback. Take for instance *Escherichia coli* chemotaxis, a well-characterized pathway allowing bacteria to sense chemicals and to swim towards nutrients[6,7]. The intracellular level of the phosphorylated protein ($CheY_p$) represents the extracellular concentration of a chemical and regulates the motors (clockwise or counterclockwise rotation) and hence motility ('run' or 'tumble') (Fig. 1). The swimming behavior clearly affects the input as cells change their location, making information flow a circular problem. Hence, the question emerges how to tackle such problems.

Shannon's mutual information is generally used to quantify information transmission, capturing the statistical (linear and nonlinear) dependencies between inputs and outputs[8]. Specifically, the mutual information describes the ability on average to reconstruct the input distribution after repeatedly measuring the output[8,9]. Often maximal mutual information is assumed, either to reflect biological function or because the mutual information cannot be calculated otherwise[10,11]. How should mutual information be maximized? Maximizing with respect to the (generally unknown) distribution of inputs leads to the channel capacity. Such an approach was, e.g., used to study transcriptional regulation in the developing fruit-fly embryo[12,13]. While in this case, the mother organism may be able to tune its maternal factors to the optimal input distribution to match the 'expectation' of the embryo, the channel capacity may not generally be a valid approach. Alternatively, it is possible to assume a fixed input distribution and to maximize the mutual information with respect to the input-output curve[14,15]. The underlying

[1]Department of Life Sciences, Imperial College, London, UK. [2]Centre for Integrative Systems Biology and Bioinformatics, Imperial College, London, UK. [3]Department of Environmental Microbiology, Eawag, Dübendorf, Switzerland. [4]Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland. *email: r.endres@imperial.ac.uk

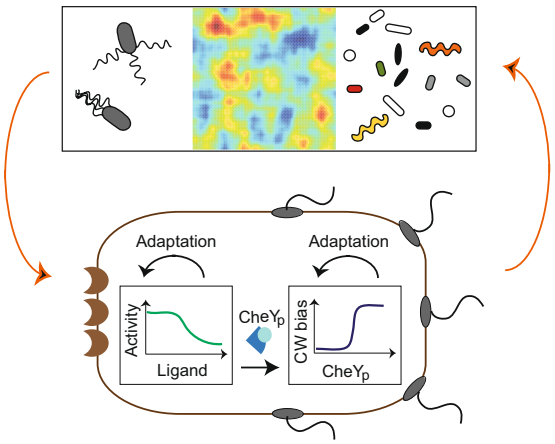

Figure 1. Connection between environmental cues and chemotactic response. Chemotactic bacteria live in complex microenvironments in which input distributions of chemical concentrations are shaped by the swimming behavior (top left), chemical sources and sinks (top middle), and competition with other bacteria (top right). Inputs are processed by the cell-internal chemotaxis pathway, which can be viewed as an input-output device (bottom). Specifically, input-output curves are measured in experiments by dose-response curves with noise. The resulting final behavior feeds back into the environment. Evolution is assumed to select the best input-output curves for maximizing fitness. The chemotactic pathway is a two-component system, and for modeling purposes, is divided into two information-transmission channels: receptors sense external concentration of stimuli and their activity regulates the protein $CheY_p$ (receptor channel, bottom left). $CheY_p$ is the internal representation of the external stimulus and regulates motor switching (clockwise or counterclockwise rotation) and thus bacterial motility (straight swimming via a 'run' or random reorientation via a 'tumble'; motility channel, bottom right). Note that there is additional adaptation both at the receptors[75] and the motors[52].

idea is that the input-output curve is adjusted by evolution to match the distribution of inputs. This approach might apply to organisms confined to certain environments, e.g. bacteria living in specific niches[15]. In addition to these two optimization procedures (and some attempts to combine them for specific types of input-output curves[16–18]), the Fisher information from estimation theory may also be used to predict the input distribution[19,20]. Hence, due to multiple available approaches there is considerable uncertainty to what optimization procedure to use.

Even the principle of maximal information transmission may be questioned. Recently, maximizing information transmission at the *E. coli* receptors led to maximal drift of cells swimming up a chemical gradient and hence optimal chemotactic behavior[15]. Although a reasonable result, there are two potential problems when considering the whole pathway: Firstly, the dose-response curves of the motors are measured to have Hill coefficients up to 20[21]. Hence, such switch-like motors may only transmit about one bit of information, distinguishing only low and high levels of $CheY_p$. Secondly, the $CheY_p$ level, which maximizes the drift is not in the sensitive region of the motor[22–24]. Both issues may lead one to suggest that high information transmission at the receptors is wasted downstream at the motors, and that an entirely different principle may guide cell behavior[25,26].

Here, we address two key questions: (1) How should the mutual information be calculated in a biological context, and (2) does the bacterial chemotaxis pathway maximize information transmission? Specifically, we reconcile the various optimization procedures, leading to a new way of maximizing the mutual information, particularly useful for biological systems. Assuming general external and internal noise, and a fixed range of sensitivity (as any biological or physical system is necessarily limited), we analytically derive both the optimal input distribution *and* input-output curve. Unlike previous approaches[16–18], this general solution does not assume specific input-output curves, such as Hill functions. Surprisingly, we find an universal optimal input distribution, only dependent on the input noise. Furthermore, numerically and with the help of simulations, we were able to extend our formalism to multiple outputs (or inputs), greatly extending the applicability of our formalism to biological systems. As an illustrative example, we focus on the *E. coli* chemotaxis pathway using previously estimated noise and measured dose-response curves at the receptors and the motors. By deriving analytical results for multiple output motors, we show that although the optimal response is not a Hill function, the measured Hill coefficients naturally emerge from our optimal prediction. Overall our results confirm the idea of maximal information transmission in *E. coli* chemotaxis, and prove that maximal information transmission at the receptors is critical for the whole pathway despite the ultrasteep motor dose-response curves.

Results

Maximizing mutual information: comparison of different approaches. Information transmission between an input, X, and an output, Y, is often quantified by the mutual information, which is a measure of statistical dependency and reflects the average ability for inferring the input after measurements of the output ([8–11,27–29] for extended reviews). For continuous random variables, X and Y, the mutual information is defined by

$$\mathcal{I}[X,\ Y] := \int \mathrm{d}y\ \mathrm{d}x\ p(y|x)p(x)\log_2\left[\frac{p(y|x)}{p(y)}\right], \tag{1}$$

where $p(y|x)$ is the conditional probability of observing $Y=y$ at given $X=x$, encoding the input-output curve and noise. Quantities $p(x)$ and $p(y)$ represent the input and output distributions, respectively, which are mathematically connected by the conservation of probability $p(x)\mathrm{d}x=p(y)\mathrm{d}y$, valid in the small-noise limit. Considering small Gaussian noise with conditional probability, $p(y|x) = \exp[-(y-\overline{y})^2/(2\sigma_\mathrm{T}^2)]/\sqrt{2\pi\sigma_\mathrm{T}^2}$, with mean $\overline{y}(x)$ and standard deviation $\sigma_\mathrm{T}(x)$, Eq. (1) becomes

$$\mathcal{I}[X,\ Y] = -\int_{x_\mathrm{on}}^{x_\mathrm{off}} \mathrm{d}x\ p(x)\ \log_2\left[\frac{\sqrt{2\pi e}\ \sigma_\mathrm{T}(x)}{\overline{y}'(x)}p(x)\right], \tag{2}$$

where x_on and x_off set the sensitive region, i.e. the dynamic range of inputs[14,15] (similar equations appear in[12,13,20]). In addition to $p(x)$, Eq. (2) depends on the gain $\overline{y}'(x)$, i.e. the first derivative of the input-output curve $\overline{y}(x)$, and total noise $\sigma_\mathrm{T}(x)$.

To understand if biological systems maximize information flow, we need to maximise the mutual information and derive general principles or compare with data. To maximise the mutual information, the channel capacity is often considered, i.e. the mutual information maximized with respect to the input distribution[12,13]. Alternatively, the mutual information can be maximized with respect to the input-output curve assuming a fixed input distribution[14,15]. The former method is an attempt to deal with the often unknown input distribution, while the latter is based on the idea that the biological channel can be modified by evolution. Additionally, the two approaches were combined for specific input-output Hill and Hill-like functions[16–18]. However, is there a general way to unify the different methods, without making assumptions about functional form of the input-output functions?

Formally, we maximize the mutual information, Eq. (2), with respect to $p(x)$ and $\overline{y}x)$ by writing

$$\max_{p(x)}\mathcal{I}[X,\ Y]\ \rightarrow\ \frac{\partial\mathcal{L}}{\partial p} = 0, \tag{3}$$

$$\max_{\overline{y}(x)}\mathcal{I}[X,\ Y]\ \rightarrow\ \frac{\partial\mathcal{L}}{\partial\overline{y}} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial\mathcal{L}}{\partial\overline{y}'} = 0, \tag{4}$$

where the right-hand side of Eqs. (3) and (4) are Euler-Lagrange equations from calculus of variations with Lagrangian $\mathcal{L}(x,\ p,\ \overline{y},\ \overline{y}') = p\ \log_2\left[\frac{\sqrt{2\pi e}\,\sigma_\mathrm{T}}{\overline{y}'}p\right]$, i.e. the integrand of Eq. (2). Equation (3) represents the channel capacity applied to a Gaussian channel (i.e. Gaussian conditional probability $p(y|x)$, see[12,13] for examples in gene regulation). In contrast, Eq. (4) is used to obtain the optimal input-output curve for a given input distribution (see[14,15] for examples in sensory systems). For completeness, we provide the solutions of the individual maximizations of Eqs. (3) and (4) in SI Text Sec. 1.1–1.2, with a discussion of the sensitive region in Sec. 1.8.

When the noise is uniform (σ_T constant) Eqs (3) and (4) coincide. Specifically, $\frac{\partial\mathcal{L}}{\partial\overline{y}} = 0$ in Eq. (4) so that $\overline{y}$ is a cyclic variable. As a result, $\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial p}{\partial\overline{y}'} = 0$ so that $p/\overline{y}'=$ const and hence is conserved (not in time but in input space), following the Emmy Noether theorem. In this case, maximizing the mutual information leads to a simple matching relationship ($\overline{y}'\propto p$), so that the input-output curve is the cumulative integral of the input distribution (see *S1 text*, Sec. 1.1 and Fig. S1)[30]. However, in general when both input and output noise matter the noise is a function of the input and input-output curve, given by $\sigma_\mathrm{T}=\sigma_\mathrm{T}(x,\ \overline{y},\ \overline{y}')$. Assuming independent cell-external and internal noise, we consider

$$\sigma_\mathrm{T}(x,\ \overline{y},\ \overline{y}') = \sqrt{\sigma_x^2\overline{y}'^2 + \sigma_y^2}, \tag{5}$$

which follows from error propagation. Specifically, σ_x is the input noise depending on x only, amplified by the gain $\overline{y}'(x)$, and σ_y is the output noise depending on $\overline{y}(x)$ only. In case of negligible input ($\sigma_\mathrm{x}\approx 0$) or output ($\sigma_\mathrm{y}\approx 0$) noise, Eq. (3) again converges to Eq. (4) and the system can be solved for any input-output curve $\overline{y}(x)$ (see *S1 text*, Sec. 1.2–1.3). As a result, the predicted input and output distributions from the two optimization approaches become identical (Fig. 2A). However, in general the two equations differ and the resulting optimal input and output distributions are very different in the two approaches (Fig. 2B). In particular, the output distributions can be uni- or bimodal with details described in *SI Text*, Sec. 1.3.

It is worth noting that, in Bayesian statistics the Fisher information is linked to the channel capacity[19,20]. A key problem in Bayesian statistics is choosing a prior distribution for a given stochastic process (i.e. $p(y|x)$)[31]. The idea of having a prior which does not affect the posterior distribution (i.e. $p(x|y)$) is linked to maximal mutual information, given by the average Kullback-Leibler divergence between prior and posterior distributions. This prior distribution is called the reference prior[19], given by $p(x) \propto \sqrt{\mathcal{F}(x)}$ with Fisher information $\mathcal{F}(x) = \int dy\ p(y|x)(\partial\log(p(y|x))/\partial x)^2$. As shown in ref.[20]. the Fisher information is the result of maximizing the equivalent of Eq. (2) for a general (not necessarily Gaussian) conditional probability distribution (see *S1 Text*, Sec. 2.). Hence, the channel capacity and the approach based on the Fisher information are equivalent.

Maximizing mutual information: a new approach. The difference between Eqs. (3) and (4) is that Eq. (3) assumes a fixed input-output curve and a variable input distribution, while Eq. (4) assumes a fixed input

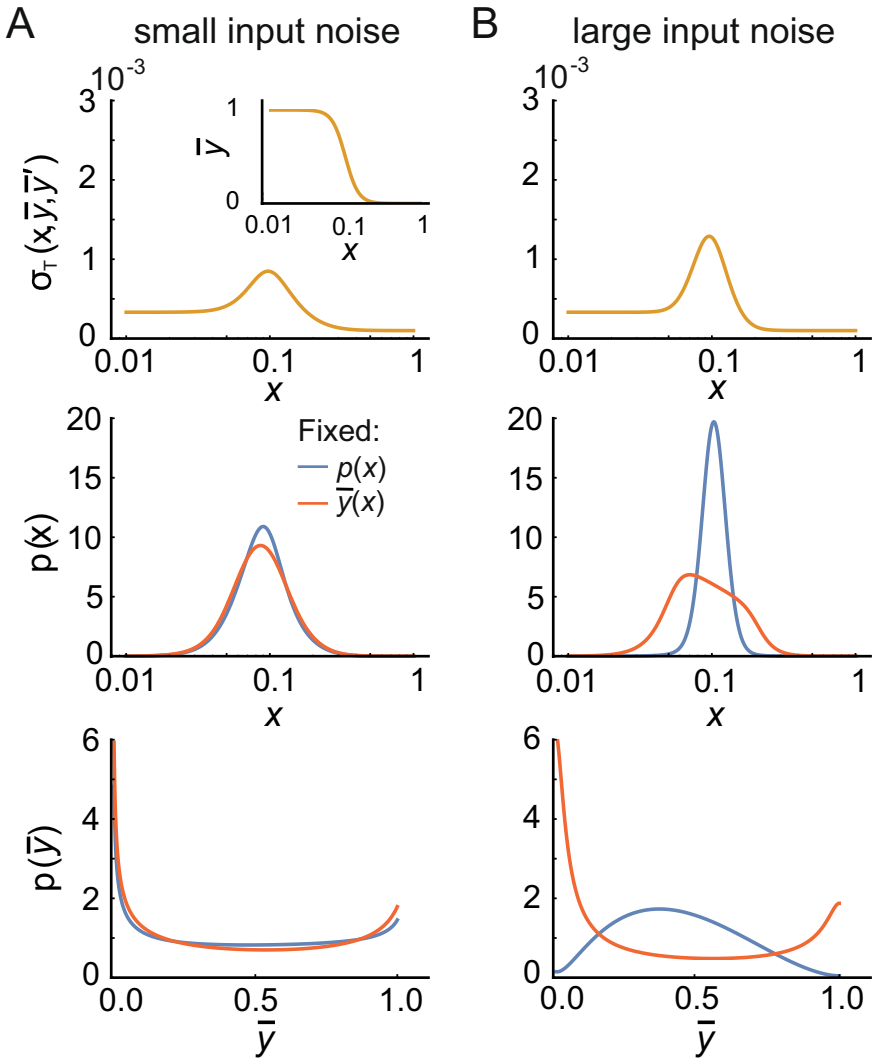

Figure 2. Conventional ways of maximizing mutual information. Comparison of the channel capacity, i.e. solving the maximization problem Eq. (3) for $p(x)$ using a fixed $\overline{y}(x)$ and noise (red), and maximization with respect to input-output curve, i.e. solving the maximization problem Eq. (4) for $\overline{y}(x)$ using a fixed $p(x)$ and noise (blue). (For specific examples of the solutions of Eqs (3) and (4) and a discussion of the bimodality of the output distribution, see *SI Text* Sec. 1). In both cases, the noise is provided as a function of the input distribution and input-output curve (top row), with the input-output curve $\overline{y}(x) = [1 + (x/k_d)^{n_H}]^{-1}$ assumed a Hill function for simplicity, with Hill coefficient n_H and threshold k_d (inset). (**A**) For small input noise, the two approaches converge, i.e. the red and the blue input (middle left) and output (bottom left) distributions match. (**B**) For large input noise, the two approaches predict different input (middle right) and output (bottom right) distributions. Using Eq. (5) for the noise, the input noise is $\sigma_x^2 = \alpha_1 x$, while the output noise $\sigma_y^2 = \alpha_2 \overline{y}(1-\overline{y}) + \alpha_3 \overline{y} + \alpha_4$ has three different contributions with $\overline{y}$ the input-output curve. The parameters are chosen to provide an overall similar level of noise, given by $\alpha_1 = 10^{-8}$, $\alpha_2 = 2 \cdot 10^{-6}$, $\alpha_3 = 10^{-7}$, $\alpha_4 = 10^{-8}$ (panel A) and $\alpha_1 = 10^{-7}$, $\alpha_2 = 10^{-8}$, $\alpha_3 = 10^{-7}$, $\alpha_4 = 10^{-8}$ (panel B). For the red model, $n_H = 5$ and $k_d = 0.1$. For the blue model, the input distribution is fixed by normalizing the derivative of a Hill function with $n_H = 5$ and $k_d = 0.1$, with the input-output curve free to change according to the maximization. The sensitive region is set by $x_{on} = 0$ and $x_{off} = +\infty$.

distribution and a variable input-output curve. There might be situations in which one approach is more appropriate than the other but in a general biological context the two are intrinsically connected (Fig. 1). From a mathematical point of view, Eqs (3) and (4) can be combined and solved together, i.e. $\mathcal{I}[x, y]$ maximized with respect to both $p(x)$ and $\overline{y}(x)$. Similar numerical double optimizations are common in rate distortion theory using, e.g., the Blahut algorithm[4,32].

In what follows, we provide the analytical solution for $p(x)$ and $\overline{y}(x)$ by solving Eqs. (3) and (4) together. We assume a fixed dynamical range of inputs set by x_{on} and x_{off} (given by the receptor sensitivity), leading in return to a fixed dynamical range of outputs from $y(x_{on}) = 1$ to $y(x_{off}) = 0$. We consider Eq. (2) with noise given by Eq. (5). After a first integration, we obtain the formal solution

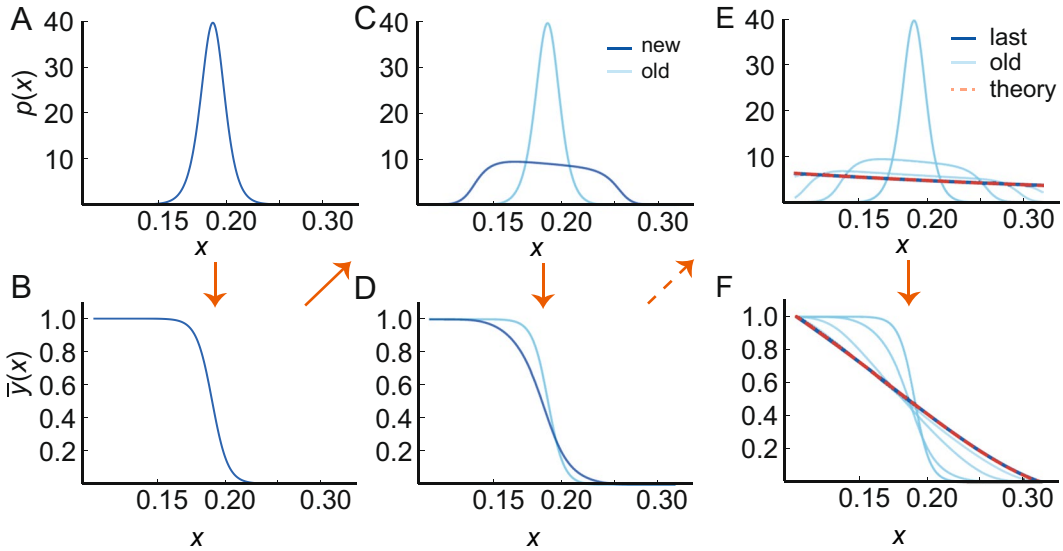

Figure 3. Adaptive evolutionary algorithm approaches analytical result. (**A**) Bacteria are assumed to live in a given environment and experience a given distribution of inputs. (**B**) Assuming that maximal information transmission enhances the chance to survive, evolution selects the mutations and hence the phenotype with optimal input-output curve. (**C**) At given input-output curve, the optimal input distribution generally (dark blue) differs from the initial input distribution (light blue) in (**A**). A change in behavior (and hence a change in the inputs) may provide an increase in information transmission (see Discussion section for more details). (**D**) The input-output curve changes again to maximise information transmission for the new set of stimuli. (**E**,**F**) This iterative cycle continues and eventually converges to the solution given by Eqs. (6a) and (6b) (dashed red line). Parameters: $\alpha_{1-3}=10^{-7}$, $\alpha_4=10^{-8}$, $x_{on}=0.115$, $x_{off}=0.323$.

$$p = \frac{\bar{y}'}{Z\sigma_T}, \tag{6a}$$

$$\sigma_x^2 \bar{y}'^2 = Q\sigma_y^2, \tag{6b}$$

where Z and Q are two constants set by normalization and boundary conditions, respectively (see *Materials and Methods* and *S1 Text*, Sec. 1.7).

Equation (6a) for the input distribution extends the matching relationship found in[30] to nonuniform noise. In the latter case the optimal input distribution weighs certain inputs more than uncertain inputs[13,20,33]. Equation (6b) determines the input-output curve which maximizes the mutual information given the noise. A solution of the system of equations exists if the transmitted input noise can be expressed in terms of the output noise or vice versa (see *S1 Text*, Sec. 1.7). While such a solution may seem very specific, it is certainly plausible, given enough time, that evolution eventually finds it.

How may evolution find the solution? To mimic evolution, we envision an adaptive algorithm, allowing the pathway to iteratively reach optimal information transmission. Given an environment and hence a distribution of inputs, $p_1(x)$ (Fig. 3A), evolution selects the optimal internal input-output curve, $y_1(x)$ (Fig. 3B). However, the distribution of inputs is susceptible to changes, which might be caused by a change of the organism's behavior, even in the same environment. The new input distribution, $p_2(x)$, may again lead to an increase in information transmission at fixed input-output curve, $y_1(x)$ (Fig. 3C). Subsequently, evolution will select a new input-output curve, $y_2(x)$, which enhances information transmission at a fixed input distribution, $p_2(x)$ (Fig. 3D). This cycle is repeated many times. If the optimal configuration is achievable and information transmission is a proxy for fitness, we expect that the solution of Eqs. (6a) and (6b) naturally emerges in the pathway. This is indeed the case for the examples studied here (see Fig. 3E,F).

Information transmission at *E. coli* chemoreceptors. To apply our new approach, we use the chemotaxis pathway of *E. coli* as an explicit example, since it is relatively simple and well characterized in its molecular components[6]. Briefly, chemoattractant (ligand) binding turns receptors off, inhibits the kinase CheA, and hence reduces the phospho-transfer from $CheA_p$ to CheY. This leads to 'runs' as only $CheY_p$ can bind the 6–8 motors to introduce 'tumbling'. There is also an adaptation mechanism, where addition of methyl groups to receptors compensates for increased attractant concentration by increasing the receptor activity and hence the $CheY_p$ level to induce cell tumbling. Removal of methyl group has the opposite effect[34,35]. In order to study signaling in fixed adaptational states, the adaptation enzymes can be removed from the chromosome and the receptor expressed with specific, genetically engineered, receptor modification levels to mimic receptor methylation (see *S1 Text*, Sec. 4.3)[34,35].

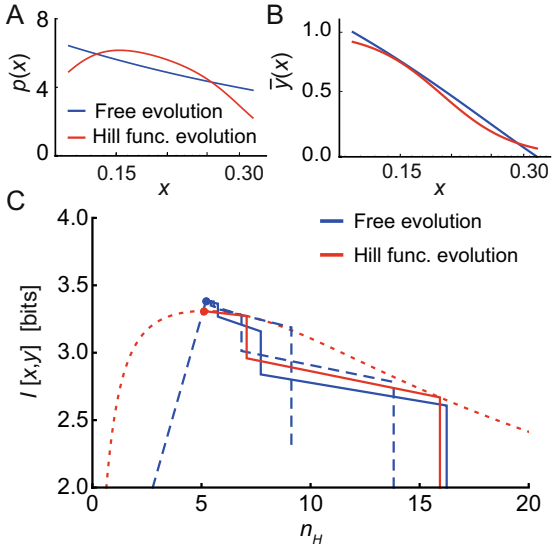

Figure 4. Robustness of the adaptive algorithm. Comparison between our analytical result (blue lines) and the corresponding result when the input-output curve is constrained to a Hill function with adjustable Hill coefficient n_H (red lines). (**A**,**B**) Input distributions (**A**) and the input-output curves (**B**). (**C**) Comparison between the adaptive algorithm for the unconstrained (blue lines) and Hill-function constrained (red solid line) optimizations. The adaptive algorithm moves the input-output curves to the optimal value $n_H \approx 5$, steadily increasing the mutual information. The convergence towards the analytical solution is robust to different initial Hill coefficients (dashed blue lines). The optimal Hill coefficient is compatible with the corresponding experimental curve from[35] (see also Fig. S9). When constraining the input-output curve to a Hill function, the highest mutual information at a given Hill coefficient is shown by the red dashed line. Noise parameters: $\alpha_{1-3} = 10^{-4}$, $\alpha_4 = 10^{-5}$, $x_{on} = 0.115$ and $x_{off} = 0.323$ (see *S1 Text*, Sec. 3 for details of the noise).

Specifically, we consider the instantaneous information transmission between the chemoattractant methylaspartate (MeAsp) as the input and the response regulator $CheY_p$ as the output. Hence, we consider the information transmitted by the initial (fast) response for a given adaptational state (which only changes slowly). (At a later time this response is removed by adaptation and hence is transient only). Note, unlike ref.[36] we do not assume small Gaussian inputs but natural stimuli drawn from broad, potentially asymmetric input distributions $p(x)$. When the input distribution of cells simulated in gradients of different strength match the optimal information-theoretical input distribution for the same receptor modification levels, the drift velocity up the gradient is maximized and, hence, this leads to optimal chemotaxis[15]. As this matching of input distributions occurs anywhere in the gradient, these initial responses describe chemotaxis in the whole gradient, and so implicitly include adaptation. Indeed, the predicted distributions of inputs are scale invariant (when normalized by the adapted concentration) and reproduce Weber's law and logarithmic sensing[15]. The latter can also be captured by the predictive mutual information[37].

The functional form of the noise is assumed to be known and derived from microscopic theory as in[38] (see *S1 Text*, Secs. 3 and 4.2 for noise estimation and sensitivity to noise parameters, respectively). In short, the input variance is considered to be proportional to the input strength, $\sigma_x^2 = \alpha_1 x$, with x in units of the ligand concentration the cell is adapted to. Furthermore, $\alpha_1 \propto (DN\tau)^{-1}$ is given by the Berg-and-Purcell limit[39], where D is the diffusion constant of the ligand molecules, N is the number of receptors acting cooperatively in a cluster, and τ is the averaging time, assuming a spherical cell[39]. The output noise has three contributions: signaling noise, switching noise due to *on*/*off* changes of the receptor state, and a constant background noise, leading to $\sigma_y^2 = \alpha_2 \bar{y}(1-\bar{y}) + \alpha_3 \bar{y} + \alpha_4$ with phosphorylated $\bar{y}$ in units of the total CheY level, Y_T, an intrinsic dependence on the (unknown) input-output curve, and α_{2-4} additional parameters defined in *S1 Text*, Sec. 3. Note these effective noise terms are time-averaged due to their dependence on chemical reactions based on finite rate constants[38]. Despite σ_y^2 being specific, this noise should apply to many receptor-signaling pathways, including other two-component pathways[40].

Using this noise, the explicit solution of Eq. (6b) is

$$\bar{y}(x) = \frac{\alpha_2 + \alpha_3 + \sqrt{(\alpha_2+\alpha_3)^2 + 4\alpha_2\alpha_4}\,\sin\left[\sqrt{\alpha_2}\left(C + 2\sqrt{\frac{Qx}{\alpha_1}}\right)\right]}{2\alpha_2}, \tag{7}$$

where C and Q are constants set by imposing fixed boundary conditions, $y(x_{off}) = 1$ and $y(x_{on}) = 0$ (Fig. 4A). Note that x appears with the prefactor Q/α_1. Thus, α_1 is set by the boundary conditions and it can be seen as the units of x. We consider two special cases: by simplifying the output noise for $\alpha_2 = 0$, we obtain

$$\overline{y}(x) = \frac{\alpha_3^2 C'^2 - 4\alpha_4}{4\alpha_3} + \alpha_3 C' \sqrt{\frac{Qx}{\alpha_1}} + \frac{\alpha_3 Qx}{\alpha_1}, \tag{8}$$

where the solution is again independent of α_1 after imposing the boundary conditions. For $\alpha_{2,3}=0$, we obtain

$$\begin{aligned} \overline{y}(x) &= 2Q\sqrt{\frac{\alpha_4}{\alpha_1}}\sqrt{x} + C'' \\ &= -\frac{\sqrt{x}}{\sqrt{x_{\text{on}}} - \sqrt{x_{\text{off}}}} + \frac{\sqrt{x_{\text{on}}}}{\sqrt{x_{\text{on}}} - \sqrt{x_{\text{off}}}}, \end{aligned} \tag{9}$$

which does not depend on both α_1 and α_4 once Q and C'' are set by the boundary conditions. The optimal input distribution obtained by inserting Eq. (6b) into Eq. (6a) is $p = (Z\sqrt{(1+Q)/Q}\,\sigma_x)^{-1}$. In particular, for our choice of external noise and fixed sensitivity range, the input distribution converges to $p(x) = [2(\sqrt{x_{\text{on}}} - \sqrt{x_{\text{off}}})\sqrt{x}]^{-1} \sim 1/\sqrt{x}$ independently of α_{1-4} (see *S1 Text*, Sec. 1.4.6 and Fig. S3). Importantly, this is a general result for the Berg-and-Purcell input noise, and hence should be valid for many signaling pathways. This result was previously found numerically[16], and such an input distribution of glutamine was suggested to optimize nitrogen sensing[33].

Extracting the sensitive regime of *E. coli* receptors for a fixed modification level (resembles receptor methylation level, see *S1 Text*, Sec. 4.3)[35,41], we test the convergence of the adaptive algorithm to the solution in Eq. (7). After a few iterative cycles the system indeed converges (Fig. 4A), increasing the mutual information at each step (Fig. 4C, blue line). This convergence to the analytical solution occurs when starting at different initial conditions, showing robustness of our algorithm. Note that in Fig. 4C the solution $\overline{y}(x)$ is fitted to Hill functions for convenience of presentation, allowing the mutual information to be plotted as a function of a single parameter (i.e. the Hill coefficient n). The optimal curve selects a Hill coefficient compatible to the experimental measurements from FRET data, at least for larger receptor modification levels (Fig. S9)[35]. Applying the adaptive algorithm instead to Hill-function constrained input-output curves produces the same optimal Hill coefficient n albeit with a smaller mutual information (Fig. 4C, red solid line, with Fig. 4B comparing the corresponding optimal input distributions and optimal input-output curves). Note that for a fixed Hill equation the optimal mutual information is calculated directly using the input distribution from Eq. (6a), resulting in

$$\mathcal{I}[x, y] = \log_2\left[\frac{Z}{\sqrt{2\pi \mathrm{e}}}\right] \quad \text{with} \quad Z = \int_{x_{\text{on}}}^{x_{\text{off}}} dx\, \frac{\overline{y}'(x)}{\sigma_{\text{T}}(x)}, \tag{10}$$

as shown in Fig. 4C (red dashed line).

However, unlike the *sine* function in Eq. (7), experimental dose-response curves of CheY_p are thought to be well approximated by Hill functions rather than Eq. (7)[34]. There are several possible reasons for this discrepancy. For instance, in our model receptors are either fully sensitive or fully insensitive, and the solution given by Eq. (7) is only valid in the sensitive region and constant otherwise. In reality, receptors have a smooth sensitivity curve spanning the ligand-dissociation constant of the *off* and *on* states (such as $\mathrm{d}F/\mathrm{dlog}(x)$ in[41], where F is the receptor free-energy difference between *on* and *off* states). This may lead to a smooth Hill-function-like response. One way of imposing smooth input-output curves is to introduce the additional constraint of zero first derivatives at the boundary. In this case, however, we only obtain sigmoidal input-output curves without internal switching noise ($\alpha_3=0$) (see *S1 Text*, Sec. 1.4.4 and Fig. S3E). Moreover, very asymmetric (or even bimodal) input distributions might be uncommon in natural environments[15], potentially favoring log-normal input distributions and hence Hill-function-like responses[42]. Finally, *E. coli* needs to account for many other constraints and the pathway performs other tasks at the same time, such as sensing temperature and pH[43–45]. Hence, the *E. coli* sensory system might be in a suboptimal configuration for transmitting information about chemicals in order to account for all the other tasks. In the *S1 Text*, Sec. 1.4.5 we also solve the inverse problem and derive the optimal noise, which leads to an exact Hill function (see Fig. S4). In this case, the predicted input and output noises are not independent anymore. In summary, the mismatch between the experimental Hill functions and the solution in Eq. (7) is not an artifact of our assumptions on the sources of noise; it emerges when considering independent input and output noise, and when maximizing the mutual information with respect to both the input distribution and the input-output curve.

Information transmission along the *E. coli* chemotaxis pathway. Now that we understand the optimization of the mutual information better, we can tackle the second problem: Does the chemotaxis pathway maximise information transmission? Previous work suggests that the higher the information transmission at the receptors, the higher the drift velocity in the direction of the gradient[15]. Is this finding compatible with the recent observation of the ultrasensitive response of the motor to changes in internal CheY_p[21], or does such a steep response prevent the cell from high information transmission? To answer this question we extend our analysis to the whole chemotaxis pathway.

We consider a minimal model of two channels: a receptor channel for sensing by the chemoreceptors and a motor channel for the flagellar motors. For the receptor channel, the external chemical concentration x is the input and the internal CheY_p concentration, y, is the output. For the motor channel, y is the input and the motor clockwise (CW) bias z (for tumble) is the output (Fig. 5A). To simplify the problem and to closely resemble the experimental dose-response curves, we now restrict the curves to Hill functions, with Hill coefficients n and m for receptors and motors, respectively. The noise expressions for the receptor and motor channels are given by

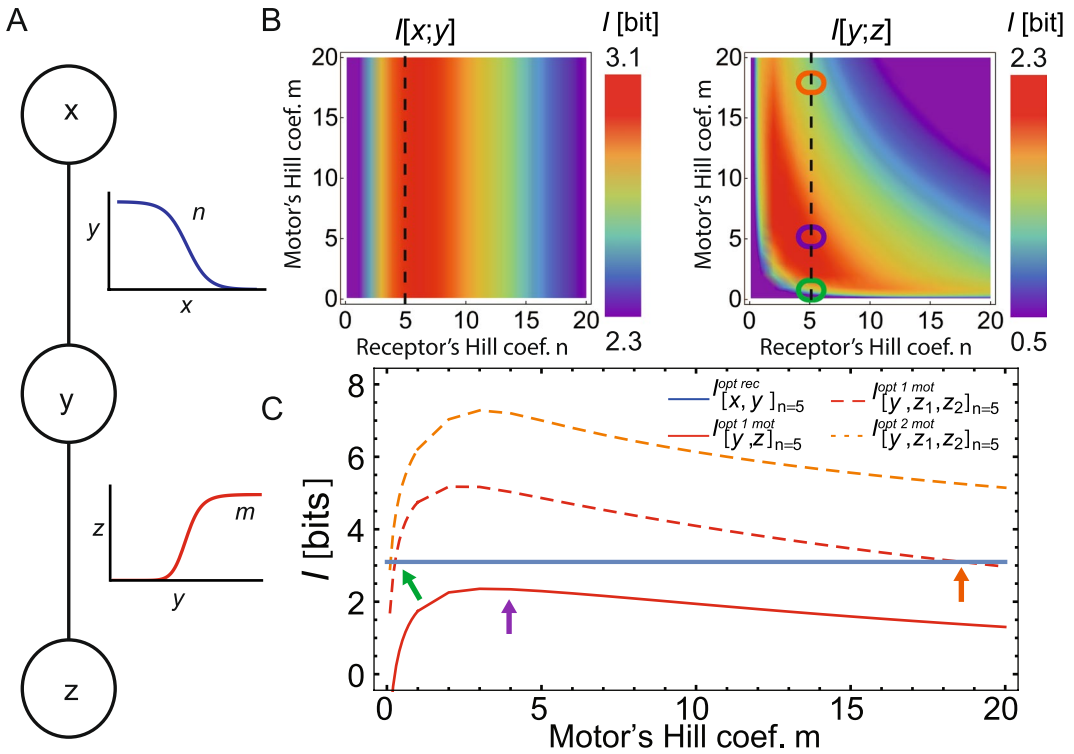

Figure 5. Optimal information transmission in the *E. coli* chemotaxis pathway. (**A**) Minimal model of the *E. coli* chemotaxis pathway, using two concatenated channels. The extracellular concentration of stimulus x is the input of the receptor channel and $CheY_p$, y, concentration is the output. For the motor channel, $CheY_p$ is the input and the tumble bias z the output. Hill functions with adjustable Hill coefficients n and m represent the input-output curves of the receptor and motor channels, respectively. The dissociation constants for the receptor and motor channels are $k_d^n = 0.189/(1\,\mu M)$ and $k_d^m = 0.39\,[Y_T]$, respectively. (**B**) Heat maps of the separately calculated mutual information of the receptor, $\mathcal{I}[x, y]$, (left) and of the motor, $I[y,z]$, (right) channels for a single motor as a function of Hill coefficients n and m. Vertical dashed lines correspond to maximal value of $\mathcal{I}[x, y]$ (circles corresponds to arrows in **C**). (**C**) The mutual information as function of Hill coefficient m at $n = 5$ (dashed black lines in panel B). While for a single motor the optimal mutual information of the receptor channel (*opt rec*, blue solid line) is higher than the optimal mutual information at the motor (*opt mot*, red solid line), increasing the number of motors (see legend) enhances the optimal mutual information of the motor channel (*opt 1 mot*, dashed red line, for two motors). Arrows point to the predicted m values. The purple arrow indicates the optimal m value for a single motor (corresponding to purple circle in **B**), the green and orange arrows point to the optimal m values for two motors (corresponding to green and orange circles in **B**). The mutual information is further increased when the optimal mutual information for two motors is calculated (*opt 2 mot*, orange dashed line, see *S1 Text*, Sec. 4). Parameters: $\alpha_{1-3} = 10^{-4}$, $\alpha_4 = 10^{-5}$, $\beta_2 = 7 \cdot 10^{-4}$, $\beta_3 = 7 \cdot 10^{-4}$, and $\beta_4 = 1 \cdot 10^{-4}$.

$\sigma_{yT} = \sqrt{\alpha_1 x G_y^2 + \alpha_2 \overline{y}(1-\overline{y}) + \alpha_3 \overline{y} + \alpha_4}$ and $\sigma_{zT} = \sqrt{\sigma_{yT}^2 G_z^2 + \beta_2 \overline{z}(1-\overline{z}) + \beta_3 \overline{z} + \beta_4}$, respectively, where G_y and G_z are the gains of the receptor and motor channels. Parameters β_{2-4} represent the noise of the motors and are kept generic due to lack of characterization, but may reflect analogous biological processes including adaptation of the motors[21,46,47] (see *S1 Text*, Sec. 3 for further details and robustness of results to changes in β values). Due to the immense gain at the motors[21], we generally have higher noise at the motor then at the receptor (see *S1 Text* Sec. 4.5 for the discussion of the two limits). Here only n and m are considered adjustable parameters in our model.

The data processing inequality, which characterizes the flow of information in a Markov chain, states that, at any additional processing step, information can only be lost, never gained[48]. For instantaneous information transmission this means that the mutual information between the external concentration and the motor bias cannot be higher than the minimum of the mutual informations of the receptor and the motor channels, i.e. $\mathcal{I}[x, z] \leq \min\{\mathcal{I}[x, y],\ \mathcal{I}[y, z]\}$. A strategy to possibly increase the mutual information is then to maximise the limiting mutual information.

We start by considering a single motor, represented by a single output z. We calculate the maximal mutual information at the receptor (Fig. 5B, left) and motor (Fig. 5B, right) channels, dealing with the optimization of the two channels separately. The mutual information at the motor, $\mathcal{I}[y, z]$, always limits the whole information transmission for any Hill coefficient n of the receptors and m of the motor. Hence, single-motor cells should optimize the motor rather than the receptor channel. This result is not unexpected for the chemotaxis pathways since the ultrasensitive motor enhances the downstream noise, which is generally larger than the upstream noise (here the

total $CheY_p$ noise is the input noise for the motor channel). The resulting optimal information transmission corresponds to relatively low n and m (≈6; red area in right panel of Fig. 5B). Experimentally, the Hill coefficient of the receptor channel agrees with our prediction, ranging from 6–12 in Tar-only cells[35]. In contrast, the ultrasteep motor response curve with $m \approx 20$ is in stark contradiction to our single-motor dose-response model. Hence, the single-motor model indicates higher information transmission at the receptors (see Figs. 5, S6, S7, and *S1 Text*, Secs. 3 and 4.2 for dependence on noise). However, *E. coli* has multiple motors, which might effect information transmission.

We now extend our model to multiple (K) motors, allowing the cell to make multiple measurements of the internal $CheY_p$ concentration. There are now a single input, y, and multiple outputs $z_1, ..., z_K$, of the CW biases. The chain rule for the mutual information allows us to calculate $\mathcal{I}[y; z_1, ..z_K]$. In particular, for the extreme case of fully coupled motors the mutual information does not increase with the growing motor number, $\mathcal{I}[y; z_1, ..z_K] = \mathcal{I}[y; z_1]$. In contrast, for completely uncoupled motors (i.e. for motors which are simultaneously independent and conditionally independent given y) the mutual information increases with the number of motors, $\mathcal{I}[y; z_1, ..z_K] = K\mathcal{I}[y; z_1]$ (see *S1 Text*, Sec. 4.5). Real motors show evidence of partial coupling[49], and thus we assume conditional independent motors, i.e. $p(z_1, ..z_K|y) = \Pi_i^K p(z_i|y)$. This means that all motors depend on the common y level but can independently select their CW bias. For two motors, the mutual information becomes $\mathcal{I}[y; z_1, z_2] = \mathcal{I}[z_1; y] + \mathcal{I}[z_2; y|z_1] = \mathcal{I}[z_1; y] - H(z_2|y) + H(z_2|z_1)$. For K motors this is generalized to

$$\mathcal{I}[y; z_1, \ldots, z_K] = \mathcal{I}[z_1; y] - KH(z_1|y) + \sum_{i=2}^{K} H(z_i|z_{i-1}, \ldots, z_1). \tag{11}$$

Using the small noise Gaussian approximation for $p(z_i|y)$, the conditional entropy is given by $H(z_1|y) = \int \mathrm{d}y \;\; p(y) \cdot \log(\sqrt{2\pi \mathrm{e}}\,\sigma_T(y))$, and $H(z_2|z_1) \approx 0$ (Fig. 5C, see also *S1 Text*, Sec. 4.5). Thus, Eq. (11) becomes

$$\mathcal{I}[y; z_1, \ldots, z_K] = \mathcal{I}[z_1; y] - K \int \mathrm{d}y \;\; p(y) \log(\sqrt{2\pi \mathrm{e}}\,\sigma_T(y)). \tag{12}$$

We numerically tested that the conditional independence of the motors holds for two motors, despite the fact that motors compete for the binding of internal $CheY_p$ molecules, which can introduce negative correlations (see *S1 Text*, Sec. 4.6 and Fig. S10). Therefore, $\mathcal{I}[y; z_1, z_2] > \mathcal{I}[y; z_1]$, and more generally $\mathcal{I}[y; z_1, ..z_K] > \mathcal{I}[y; z_1, ..z_{K-1}]$ for K conditionally independent motors[50].

Hence, for conditionally independent motors, the mutual information at the motors will eventually overtake the mutual information at the receptors when the number of motors increases. Consequently, the mutual information at the receptors becomes the limiting factor for information transmission (Fig. 5C, see[50] for the case of Gaussian input distributions). In other words, for a small number of motors the cell has high information transmission at the receptors, which will be wasted at the motors (cf. red and blue solid lines). In contrast, for a large number of motors the information transmission at the motors exceeds the information transmission at the receptors without overall improvement. However, in the intermediate case both receptors and motors equally limit the transmission of information (cf. red dashed and blue solid lines in Fig. 5C). *E. coli* chemotaxis seems to avoid bottlenecks and to optimally allocate resources the latter case is the most advantageous[51]. Hence, the ultrasteep Hill function of the motor ($m \approx 20$) can be explained by this matching of the information transmission at the receptors and motors (orange arrow in Fig. 5C). Note that in addition to the high Hill coefficient m of the motors there is also a corresponding low m solution (green arrow in Fig. 5C). However, the latter is not robust to changes in m, i.e. a small change in m can lead to a drastic reduction of information transmission, which can emerge from varying the number of FliM molecules of the motor[52]. In addition, note that the mutual information shown in Fig. 5C with a red dashed line is calculated assuming that the two motors are optimized separately. However, the mutual information is further increased by maximizing the two motors simultaneously (dashed orange line in Fig. 5C). Our overall result that a high mutual information can be achieved with a high Hill coefficient of the motors remains valid (see *S1 Text*, Sec. 4). In between $m \approx 1$ and ≈20, the information transmission of the motor is wasted as receptors are information-flow limiting. In conclusion, multiple ultrasensitive motors are only useful when motors are sufficiently independent. Any residual coupling among motors may be the result of close motor proximity or mechanical coupling of the flagella.

Discussion

This study presents a new approach to maximise the mutual information, particularly suitable for evolving biological systems subject to random mutations and selection. Previously, the channel capacity, i.e. the mutual information maximized with respect to the input distribution was widely used for electronic and biological communication channels[12,13,20,33]. However, this method fails to capture possible changes of the internal input-output curve (e.g. by mutations). Furthermore, the mutual information maximized with respect to the input-output curve neglects the biological relevant feedback of the output on the input[14,15]. Here, we reconciled these two approaches by maximizing the mutual information with respect to both the input distribution and the input-output curve for Gaussian channels with small noise. Only when the total noise is uniform, or when the input or the output noise is negligible, the two approaches are identical. Unlike previous joint optimizations[16–18], our input-output curves are not restricted to Hill or Hill-like functions. Our adaptive algorithm demonstrates how evolution might implement this iteratively.

Our analytical solution of the joint optimization provides a number of new insights into optimal information transmission. First, the optimal input distribution is universal, depending only on the input noise. For Berg-and-Purcell type input noise, we specifically obtain $p(x) \sim \sqrt{x}^{-1}$. Hence, organisms are optimized for environments in which low intensity stimuli occur with high frequency. This is sensible as their high frequency would

compensate for their large relative noise levels. Second, the optimal input-output curve is invariant to up or down scaling of the input noise (parameter α_1), which sets the units of the input. Hence, only the shape of the input noise (i.e. its functional dependence on input x) affects the input-output curve. Third, our optimal input-output curve is rather linear (Figs. 3D and 4A). While this does not match the sigmoidal Hill functions as suggested by models of *E. coli* chemotaxis[34,53], a near linear input-output curve makes best use of a given dynamic range. Furthermore, enforcing either zero slope of the input-output curve at the boundaries of the sensitive region or Hill functions as input-output curves leads to assumptions on the noise which are hard to justify biologically. Hence, Hill functions are incompatible with independent input and output noise (see Supplementary Information Sec. 1.7 for details).

How can cells actively influence and optimize their distribution of sensory input? Genetic changes in the downstream pathway and motor can clearly change chemotactic behavior and hence the experienced input stimuli. For instance, increases in the motor speed lead to larger changes in stimulus and hence broader distributions of inputs. Similarly, faster adaptation leads to narrower distributions. However, the notion that cells influence their microenvironments is most supported by the important role of *niches* in stem cell differentiation, cancer development, gut microbiota, and host-pathogen interactions[54–58]. Once inside the gut, *E. coli* related pathogen *C. rodentium* in mice (and similarly EPEC/EHEC in humans) injects effector proteins into the epithelial host cells. In response, these cells secrete increased levels of oxygen, allowing in return the pathogen to perform aerobic metabolism[59]. Hence, its aerotaxis ability, inherited from *E. coli* based on Aer and Tsr receptors, experiences an increased frequency of oxygen stimuli, which the pathogen actively stimulated. If we take the assumption of maximal information transmission seriously, then cells do not only actively influence but also optimize their environment.

To apply our information-theoretical approach, we analytically showed that the entire *E. coli* chemotaxis pathway can maximize the instantaneous mutual information between chemical concentration and motor bias despite the ultrasteep dose-response curve of the motors. Briefly, ultrasensitive motors do not restrict information transmission, since a collection of motors boosts information transmission, in addition to providing other chemotactic advantages in the soil or animal intestine[60]. In particular, our model identifies the number of motors and their conditional independence as key quantities to transmit large amounts of information in peritrichous bacteria.

What is the additional information at the motors used for if the ultimate behavioral output is just binary runs and tumbles? We speculate that the tumble angle, torque, and filament handedness could be regulated[61,62]. Indeed, real-time imaging of *E.coli* with fluorescent flagella showed that the tumble angles increased with the number of clockwise-turning motors, allowing for differential cell responses[61]. Having non-identical motors with different Hill coefficients and thresholds may further increase the information transmission (e.g. as produced by different number of FliM in the motor ring)[16] but this may not be feasible in the bacterial chemotaxis pathway, as the adapted activity set the operating point of the motors. For instance, different threshold values for the motor would lead to some motor always rotating clockwise and counter-clockwise. Our model also makes the prediction that chemotactic bacteria with a single motor should prefer a relatively low Hill coefficient at the motor or multiple response regulators feeding into a motor with a high Hill coefficient to highly transmit information. This prediction could be tested with the uni-flagellated bacterial species, such as *Rhodobacter sphaeroides*, *Pseudomonas aerugiuosa* or monotrichous marine bacteria[60,63]. In support of our theory, *R. sphaeroides* is known to have multiple CheY's[64,65].

While applicable to many biological systems, our model makes a number of simplifications (in addition to assumptions on noise and receptor sensitivity). Our results are based on the independent maximizations of the receptor and motor channels. However in *SI text*, Sec. 4.5, we discuss the general case, providing estimes of the mutual information $\mathcal{I}[x; z]$, between ligand input and final motor output. Our analysis suggests, once again, that high Hill coefficient for multiple motors can support high information transmission. In particular, we analytically identify two expected limits, (i) when the receptor noise is much smaller than the motor noise, we obtain $\mathcal{I}[x; z] \approx \mathcal{I}[y; z]$, and (ii) when the motor noise is smaller than the receptor noise, we have $\mathcal{I}[x; z] \approx \mathcal{I}[x; y]$. Our analysis over the chemotactic pathway primarily focuses on the Hill coefficient. However, the dissociation constant ${k_d}^m$ of the motor response is known to be larger than the adapted $CheY_p$ concentration. In *SI text*, Sec. 4.5.4, we explicitly study the role of the dissociation constant ${k_d}^m$, and found a relative weak dependence of the mutual information on it. We also found that after fixing the Hill coefficient m and using the optimized output distribution of the receptor channel, the ${k_d}^m$ that maximizes the mutual information at the motor matches the experimentally measured value (which is larger than the adapted $CheY_p$ level, see Fig. S8).

Another simplification is that our model deals with instantaneous information transmission, and hence does not explicitly include any history dependence[36,37,66,67]. Hence, our approach should be highly suitable for the slow genetic response in quorum sensing[68,69]. In this system, the input-output relation has been measured but input distributions were simply guessed, and not predicted. Another area of application is eukaryotic chemotaxis, where cells move slowly while actively shaping their chemical gradient by ligand secretion[70] and degradation[71]. In all these examples, the input distributions and cell behaviors need to match the input-output relations to allow for optimal information gathering. Nevertheless, our model is valid for information transmission by initial transient chemotactic responses, and as this applies anywhere in the gradient, our model describes chemotaxis even including adaptation[15]. We expect that our model even works in relatively steep gradients, where, in addition to adaptation, long-history effects are important, such as caused by receptor saturation and rotational diffusion[24]. The main assumption in[15] is that gradients can be linearized over the range of input distributions. However, we do not assume small Gaussian-distributed inputs. A drawback of our model is that we neglect any cell-to-cell variability, which can be substantial[72,73], so that in effect our theory focuses on a certain subpopulation of cells. This cell-to-cell variability may lead to advantages in terms of bet-hedging strategies not directly related to information processing[74].

In conclusion, we provided a biologically inspired adaptive algorithm with analytical solution for complex problems in information transmission in sensory systems. Future studies may need to account for time-dependencies explicitly, including adaptation both at the receptors and motors. This may be achieved by considering trajectories of molecular concentrations and/or cell behavior, which may also help establish a link between chemotactic performance (e.g. drift velocity), information transmission, and energetic cost of chemotaxis. This link may show interesting tradeoffs and new design principles[29]. A methodological contribution might be necessary as calculation of the mutual information based on trajectories are hampered by the high-dimensional phase space of all possible trajectories.

Materials and Methods

Maximization of mutual information with respect to input distribution and input-output curve. To maximize Eq. (2) assuming the noise in Eq. (5), we focus on the integrand $\mathcal{L}(x, p, \bar{y}, \bar{y}') = p \log_2\left[\frac{\sqrt{2\pi e}\,\sigma_T}{\bar{y}'}p\right]$ and use the well-known Lagrange formalism from calculus of variations, where x is the independent variable while $p = p(x)$, $\bar{y} = \bar{y}(x)$ and $\bar{y}' = \bar{y}'(x)$ are the dependent variables. Note that $p'(x)$ is not appearing in the Lagrangian $\mathcal{L}$. To find the maximum of I with respect to p and $\bar{y}$, we need to solve Eqs. (3) and (4) together.

Equation (6a) is the solution of Eq. (3), which can be rewritten through differentiation as

$$\frac{p'}{p} = \frac{\bar{y}''}{\bar{y}'} - \frac{\sigma'_T}{\sigma_T}. \tag{13}$$

To derive Eq. (6b), we evaluate the following derivatives

$$\frac{\partial \mathcal{I}}{\partial \bar{y}} = \frac{p}{\sigma_T}\frac{\partial \sigma_T}{\partial \bar{y}}, \tag{14}$$

$$\frac{\partial \mathcal{I}}{\partial \bar{y}'} = \frac{p}{\sigma_T}\frac{\partial \sigma_T}{\partial \bar{y}'} - \frac{p}{\bar{y}'}, \tag{15}$$

$$\frac{d}{dx}\frac{\partial \mathcal{I}}{\partial \bar{y}'} = -\frac{p'}{\bar{y}'} + \frac{p\bar{y}''}{\bar{y}'^2} - \frac{p'}{\sigma_T}\frac{\partial \sigma_T}{\partial \bar{y}'} + \frac{p\sigma'_T}{\sigma_T^2}\frac{\partial \sigma_T}{\partial \bar{y}} + \frac{p}{\sigma_T}\frac{d}{dx}\frac{\partial \sigma_T}{\partial \bar{y}'}. \tag{16}$$

Using Eqs (13)–(16), Eq. (4) becomes

$$2\frac{\bar{y}''}{\bar{y}'} - 2\frac{\sigma'_T}{\sigma_T} + \frac{\frac{d}{dx}\frac{\partial \sigma_T}{\partial \bar{y}'}}{\frac{\partial \sigma_T}{\partial \bar{y}'}} = -\frac{1}{\bar{y}'}\frac{\frac{\partial \sigma_T}{\partial x}}{\frac{\partial \sigma_T}{\partial \bar{y}'}}. \tag{17}$$

Now assuming independent cell-external and internal noises as in Eq. (5), Eq. (17) becomes Eq. (6b).

Received: 24 September 2018; Accepted: 29 October 2019;
Published online: 15 November 2019

References

1. Perkins, T. J. & Swain, P. S. Strategies for cellular decision-making. *Mol Syst Biol* **5**, 326–326 (2009).
2. Vergassola, M., Villermaux, E. & Shraiman, B. I. 'infotaxis' as a strategy for searching without gradients. *Nature* **445**, 406 (2007).
3. Lander, A. D. How cells know where they are. *Science* **339**, 923–927 (2013).
4. Taylor, S. F., Tishby, N. & Bialek, W. Information and fitness. *arXiv preprint* **0712.4382** (2007).
5. Rivoire, O. & Leibler, S. The value of information for populations in varying environments. *J Stat Phys* **142**, 1124–1166 (2011).
6. Berg, H. C. Motile behavior of bacteria. *Phys Today* **53**, 24–29 (2000).
7. Endres, R. *Physical Principles in Sensing and Signaling: With an Introduction to Modeling in Biology* (Oxford University Press, 2013).
8. Shannon, C. E. A mathematical theory of communication. *Bell Syst Tech J* **27**, 379–423 (1948).
9. Bowsher, C. G. & Swain, P. S. Environmental sensing, information transfer, and cellular decision-making. *Curr Opin Biotech* **28**, 149–155 (2014).
10. Tkačik, G. & Walczak, A. M. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter* **23**, 153102 (2011).
11. Tkačik, G. & Bialek, W. Information processing in living systems. *Annual Review of Condensed Matter Physics* **7**, 89–117 (2016).
12. Tkačik, G., Callan, C. G. Jr & Bialek, W. Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci USA* **105**, 12265–12270 (2008).
13. Tkačik, G., Callan, C. G. Jr & Bialek, W. Information capacity of genetic regulatory elements. *Phys Rev E* **78**, 011910 (2008).
14. Detwiler, P. B., Ramanathan, S., Sengupta, A. & Shraiman, B. I. Engineering aspects of enzymatic signal transduction: photoreceptors in the retina. *Biophys J* **79**, 2801–2817 (2000).
15. Clausznitzer, D., Micali, G., Neumann, S., Sourjik, V. & Endres, R. G. Predicting chemical environments of bacteria from receptor signaling. *PLoS Comput Biol* **10**, e1003870 (2014).
16. Tkačik, G., Walczak, A. M. & Bialek, W. Optimizing information flow in small genetic networks. *Physical Review E* **80**, 031920 (2009).
17. Walczak, A. M., Tkačik, G. & Bialek, W. Optimizing information flow in small genetic networks. ii. feed-forward interactions. *Physical Review E* **81**, 041905 (2010).
18. Tkačik, G., Walczak, A. M. & Bialek, W. Optimizing information flow in small genetic networks. iii. a self-interacting gene. *Physical Review E* **85**, 041903 (2012).
19. Bernardo, J. M. Reference posterior distributions for Bayesian inference. *J R Stat Soc Ser B Stat Methodol* **41**, 113–147 (1979).

20. Brunel, N. & Nadal, J. P. Mutual information, Fisher information, and population coding. *Neural Comput* **10**, 1731–1757 (1998).
21. Yuan, J. & Berg, H. C. Ultrasensitivity of an adaptive bacterial motor. *J Mol Biol* **425**, 1760–1764 (2013).
22. Dufour, Y. S., Fu, X., Hernandez-Nunez, L. & Emonet, T. Limits of feedback control in bacterial chemotaxis. *PLoS Comput Biol* **10**, e1003694 (2014).
23. Wong-Ng, J., Melbinger, A., Celani, A. & Vergassola, M. The role of adaptation in bacterial speed races. *PLoS Comput Biol* **12**, e1004974 (2016).
24. Micali, G., Colin, R., Sourjik, V. & Endres, R. G. Drift and behavior of e. coli cells. *Biophysical Journal* **113**, 2321–2325 (2017).
25. Skoge, M., Meir, Y. & Wingreen, N. S. Dynamics of cooperativity in chemical sensing among cell-surface receptors. *Phys Rev Lett* **107**, 178101–178101 (2011).
26. Thomas, P. J. & Eckford, A. W. Capacity of a simple intercellular signal transduction channel. *IEEE Transactions on Information Theory* **62**, 7358–7382 (2016).
27. Levchenko, A. & Nemenman, I. Cellular noise and information transmission. *Curr Opin Biotechnol* **28**, 156–164 (2014).
28. Mc Mahon, S. S. *et al*. Information theory and signal transduction systems: From molecular information processing to network inference. *Semin Cell Dev Biol* **35C**, 98–108 (2014).
29. Micali, G. & Endres, R. G. Bacterial chemotaxis: information processing, thermodynamics, and behavior. *Curr Opin Microbiol* **30**, 8–15 (2016).
30. Laughlin, S. A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch C* **36**, 910–912 (1981).
31. Jeffreys, H. An invariant form for the prior probability in estimation problems. *P Roy Soc Lond A Mat* **186**, 453–461 (1946).
32. Blahut, R. Computation of channel capacity and rate-distortion functions. *IEEE transactions on Information Theory* **18**, 460–473 (1972).
33. Komorowski, M. *et al*. Analog nitrogen sensing in escherichia coli enables high fidelity information processing. *bioRxiv* 015792 (2015).
34. Keymer, J. E., Endres, R. G., Skoge, M., Meir, Y. & Wingreen, N. S. Chemosensing in *escherichia coli*: two regimes of two-state receptors. *Proc Natl Acad Sci USA* **103**, 1786–1791 (2006).
35. Endres, R. G. *et al*. Variable sizes of *Escherichia coli* chemoreceptor signaling teams. *Mol Syst Biol* **4**, 211 (2008).
36. Tostevin, F. & ten Wolde, P. R. Mutual information between input and output trajectories of biochemical networks. *Phys Rev Lett* **102**, 218101–218101 (2009).
37. Becker, N. B., Mugler, A. & ten Wolde, P. R. Optimal prediction by cellular signaling networks. *Phys Rev Lett* **115**, 258103 (2015).
38. Clausznitzer, D. & Endres, R. G. Noise characteristics of the Escherichia coli rotary motor. *BMC Syst Biol* **5**, 151–151 (2011).
39. Berg, H. C. & Purcell, E. M. Physics of chemoreception. *Biophys J* **20**, 193–219 (1977).
40. Salazar, M. E. & Laub, M. T. Temporal and evolutionary dynamics of two-component signaling pathways. *Curr Opin Microbiol* **24**, 7–14 (2015).
41. Endres, R. G. & Wingreen, N. S. Precise adaptation in bacterial chemotaxis through "assistance neighborhoods". *Proc Natl Acad Sci USA* **103**, 13040–13044 (2006).
42. Frank, S. A. Input-output relations in biological systems: measurement, information and the Hill equation. *Biol Direct* **8**, 31–31 (2013).
43. Oleksiuk, O. *et al*. Thermal robustness of signaling in bacterial chemotaxis. *Cell* **145**, 312–321 (2011).
44. Yang, Y. & Sourjik, V. Opposite responses by different chemoreceptors set a tunable preference point in Escherichia coli pH taxis. *Mol Microbiol* **86**, 1482–1489 (2012).
45. Hu, B. & Tu, Y. Behaviors and strategies of bacterial navigation in chemical and nonchemical gradients. *PLoS Comput Biol* **10**, e1003672 (2014).
46. Tu, Y. & Berg, H. C. Tandem adaptation with a common design in escherichia coli chemotaxis. *Journal of molecular biology* **423**, 782–788 (2012).
47. Zhang, C., He, R., Zhang, R. & Yuan, J. Motor adaptive remodeling speeds up bacterial chemotactic adaptation. *Biophys J* **114**, 1225–1231 (2018).
48. Cover, T. M. & Thomas, J. A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, **99** edn (Wiley-Interscience, 1991).
49. Mears, P. J., Koirala, S., Rao, C. V., Golding, I. & Chemla, Y. R. *Escherichia coli* swimming is robust against variations in flagellar number. *eLife* **3**, e01916 (2014).
50. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358 (2011).
51. Govern, C. C. & ten Wolde, P. R. Optimal resource allocation in cellular sensing systems. *Proc Natl Acad Sci USA* **111**, 17486–17491 (2014).
52. Yuan, J., Branch, R. W., Hosu, B. G. & Berg, H. C. Adaptation at the output of the chemotaxis signalling pathway. *Nature* **484**, 233–236 (2012).
53. Sourjik, V. & Berg, H. C. Binding of the *Escherichia coli* response regulator CheY to its target measured *in vivo* by fluorescence resonance energy transfer. *Proc Natl Acad Sci USA* **99**, 12669–12674 (2002).
54. Hawkins, E. D. *et al*. T-cell acute leukaemia exhibits dynamic interactions with bone marrow microenvironments. *Nature* **538**, 518 (2016).
55. Tanaka, M. *et al*. Identification of anti-cancer chemical compounds using xenopus embryos. *Cancer science* **107**, 803–811 (2016).
56. Cao, Z. *et al*. Angiocrine factors deployed by tumor vascular niche induce b cell lymphoma invasiveness and chemoresistance. *Cancer cell* **25**, 350–365 (2014).
57. Plaks, V., Kong, N. & Werb, Z. The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? *Cell stem cell* **16**, 225–238 (2015).
58. Messer, J. S., Liechty, E. R., Vogel, O. A. & Chang, E. B. Evolutionary and ecological forces that shape the bacterial communities of the human gut. *Mucosal immunology* **10**, 567 (2017).
59. Berger, C. N. *et al*. Citrobacter rodentium subverts atp flux and cholesterol homeostasis in intestinal epithelial cells *in vivo*. *Cell metabolism* **26**, 738–752 (2017).
60. Stocker, R. Reverse and flick: Hybrid locomotion in bacteria. *Proc Natl Acad Sci USA* **108**, 2635–2636 (2011).
61. Turner, L., Ryu, W. S. & Berg, H. C. Real-time imaging of fluorescent flagellar filaments. *Journal of bacteriology* **182**, 2793–2801 (2000).
62. Darnton, N. C., Turner, L., Rojevsky, S. & Berg, H. C. On torque and tumbling in swimming escherichia coli. *J Bacteriol* **189**, 1756–1764 (2007).
63. Sourjik, V. & Armitage, J. P. Spatial organization in bacterial chemotaxis. *EMBO J* **29**, 2724–2733 (2010).
64. Porter, S. L., Wadhams, G. H. & Armitage, J. P. *Rhodobacter sphaeroides*: complexity in chemotactic signalling. *Trends in Microbiology* **16**, 251–260 (2008).
65. Porter, S. L., Wadhams, G. H. & Armitage, J. P. Signal processing in complex chemotaxis pathways. *Nature Reviews Microbiology* **9**, 153–165 (2011).
66. Tu, Y., Shimizu, T. S. & Berg, H. C. Modeling the chemotactic response of *Escherichia coli* to time-varying stimuli. *Proc Natl Acad Sci USA* **105**, 14855–14860 (2008).

67. Shimizu, T. S., Tu, Y. & Berg, H. C. A modular gradient-sensing network for chemotaxis in *Escherichia coli* revealed by responses to time-varying stimuli. *Mol Syst Biol* **6**, 382 (2010).
68. Mehta, P., Goyal, S., Long, T., Bassler, B. L. & Wingreen, N. S. Information processing and signal integration in bacterial quorum sensing. *Mol Syst Biol* **5**, 325 (2009).
69. Taillefumier, T. & Wingreen, N. S. Optimal census by quorum sensing. *PLoS computational biology* **11**, e1004238 (2015).
70. De Palo, G., Yi, D. & Endres, R. G. A critical-like collective state leads to long-range cell communication in Dictyostelium discoideum aggregation. *PLoS Biol* **15**, e1002602 (2017).
71. Tweedy, L., Knecht, D. A., Mackay, G. M. & Insall, R. H. Self-generated chemoattractant gradients: attractant depletion extends the range and robustness of chemotaxis. *PLoS Biol* **14**, e1002404 (2016).
72. Colin, R., Rosazza, C., Vaknin, A. & Sourjik, V. Multiple sources of slow activity fluctuations in a bacterial chemosensory network. *eLife* **6**, e26796 (2017).
73. Keegstra, J. M. *et al.* Phenotypic diversity and temporal variability in a bacterial signaling network revealed by single-cell fret. *eLife* **6**, e27455 (2017).
74. Frankel, N. W. *et al.* Adaptability of non-genetic diversity in bacterial chemotaxis. *eLife* **3**, e03526 (2014).
75. Sourjik, V. & Berg, H. C. Receptor sensitivity in bacterial chemotaxis. *Proc Natl Acad Sci USA* **99**, 123–127 (2002).

Acknowledgements

We thank Chiu Fan Lee, Thomas Ouldridge, Michael Stumpf and Peter Swain for valuable comments on the manuscript, Judith Armitage for valuable discussions about *Rhodobactersphaeroides*, and Victor Sourjik for collaborating with us on related projects. This work was supported by European Research Council Starting-Grant N. 280492-PPHPI. Additional support came from the Biotechnological and Biological Sciences Research Council grant BB/G000131/1 (RGE) and the Swiss National Science Foundation grant nr. 31003A_169978 to Martin Ackermann (GM).

Author contributions

Both authors (G.M. and R.G.E.) designed the study. G.M. performed the research. Both analyzed results and data, and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-53273-4.

Correspondence and requests for materials should be addressed to R.G.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.