Supplementary Information

Analog nitrogen sensing in *Escherichia coli* enables high fidelity information processing via an efficient post-translational modification system

Michał Komorowski[1,*,#], Jörg Schumacher[2,*,#], Volker Behrends[3,*], Tomasz Jetka[1], Mark H. Bennett[2] Angelique Ale[2,4], Sarah Filippi[2,4], John Pinney[2,4], Jake Bundy[3,4], Martin Buck[2,4,#] and Michael P.H. Stumpf[2,4,5,#]

1. Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland.
2. Department of Lifesciences, Faculty of Natural Sciences, Imperial College London, London, UK.
3. Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK.
4. Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London, UK.
5. Institute of Chemical Biology, Imperial College London, London, UK.

* - these authors contributed equally to this work
# - correspondence: m.stumpf@imperial.ac.uk, m.buck@imperial.ac.uk, m.komorowski@ippt.gov.pl, j.schumacher@imperial.ac.uk

This is the supplementary information for the paper *Analog nitrogen sensing in Escherichia coli enables high fidelity information processing via an efficient post-translational modification system* which is henceforth referred to as the main paper (MP). We provide here both mathematical and experimental details behind the reported results. First, our general optimality criterium is derived. It is followed by formulae to calculate information capacity of signal transduction networks. Thereafter, the deterministic and stochastic models of nitrogen sensing mechanism in *Escherichia coli* are presented. Subsequently we present experimental procedures to quantify dynamic response of the nitrogen sensing mechanism to changes in nitrogen availability. The handling of experimental data, interpretation and model calibration are explained thereafter. Information-theoretic analysis of the nitrogen sensing mechanism concludes the supplement.

1 Derivation of the optimality criterium and information capacity formulae

We consider a general biochemical reaction network represented as the conditional probability distribution of the output, Y, given the input, X,

$$P_\Theta(Y|X), \tag{1}$$

where Θ is a vector of chemical reaction rate parameters. The information capacity, C_Θ, of the network $P_\Theta(Y|X)$ is defined as the maximal mutual information, $I(X,Y)$, between input, X, and output, Y, over a set of continuous and strictly positive probability distributions. Precisely,

$$C_\Theta = \max_{P_\Theta(X)} I(X,Y) = \max_{P_\Theta(X)} \int P_\Theta(X) \int P_\Theta(Y|X) \log \frac{P_\Theta(Y|X)}{P_\Theta(Y)} dY dX. \tag{2}$$

We are interested in calculating C_Θ and finding the probability distribution $P^*_\Theta(X)$, for which the maximal mutual information is achieved.
General approaches to solve this problem are based on the discretization of signal and state spaces.

The maximisation algorithm proposed by Blahut and Arimoto [1, 2] is then used to find maximum of the expression (2). More general optimisation schemes like steepest decent method and Kuhn-Tucker conditions [3, 4] can also be used. These approaches are present in many applications across different fields [5–7]. In the context of biochemical networks, under the assumption of small noise, the problem can be solved using variational methods involving integral expansion, which leads to analytical solutions [3].
The existing methods, however, either lack generality or are computationally infeasible. Specifically, we currently lack examples where the information capacity is calculated for stochastic systems with multiple inputs and multiple outputs.
Here we take a different approach compared to conventional methods as we do not need to employ optimisation techniques explicitly. Instead we use available results from the statistical theory of reference priors [8–12] that provide a general characterisation of optimal input distributions. As we demonstrate this provides a general optimality criterium described in the MP. Moreover, the characterisation leads to an efficient computation algorithm to determine C_Θ and $P^*_\theta(X)$ for biochemical systems.

1.1 Optimal distribution $P^*(X)$

In the statistical literature the problem of finding the distribution $P^*(X)$ arises in the context of non-informative priors. For explanation consider a general probability distribution $P(X|Y)$. Assume X is a parameter vector to be inferred from data Y. In Bayesian statistics the prior distribution $P(X)$ of parameters X is translated into a posterior distribution, $P(X|Y)$, that represents the knowledge about parameters, X, after the data, Y, have been collected. If no prior knowledge about parameters, X, is available it is desirable to use a non-informative prior distribution, i.e. a prior that impacts the results of estimation as little as possible. Non-informativeness of a prior distribution can be expressed in terms of the mutual information. The mutual information, $I(X,Y)$, can be written as [2]

$$I(X,Y) = H(X) - H(X|Y), \tag{3}$$

where $H(X)$ is the entropy of the input

$$H(X) = -\int P(X)\log(P(X))dX, \tag{4}$$

and $H(X|Y)$ is the average entropy of the input given the output

$$H(X|Y) = -\int P(Y)\left(\int P(X|Y)\log(P(X|Y))dX\right)dY. \tag{5}$$

The mutual information (3) can be interpreted as knowledge about X that can be gained from observing Y or, equivalently, the missing knowledge obtainable from Y. Therefore among all considered priors the one that maximises the missing knowledge can be seen as least informative. A prior distribution which maximises mutual information between X and Y is called the reference prior and this concept has been carefully characterised in the statistical literature [9]. Particularly, it has been shown [10–12] that under general regularity conditions (see Appendix 1 or the reference [11]) a prior that is asymptotically least informative is the Jeffreys prior [13]

$$P^*(X) \propto \sqrt{|FI(X)|}, \tag{6}$$

where $|FI(X)|$ is the determinant of the Fisher Information Matrix (FIM)

$$FI(X)_{i,j} = E\left(\frac{\partial \log(P(Y|X))}{\partial X_i}\frac{\log(P(Y|X))}{\partial X_j}\right), \tag{7}$$

and $E(\cdot)$ is the expected value with respect to Y at a fixed X. The characterisation of the optimal input distribution in terms of the Jeffreys prior holds under general conditions and does not require any approximations to the probabilistic input-output relation $P(Y|X)$. This is an asymptotic result and is derived under the assumption that output consist of a large number of copies of Y [10–12], which is, however, precisely the situation observed in cellular networks where both the signal and the output are temporally resolved variables and downstream effect results from temporal averaging [14].

1.2 Optimality criterium

Here we use the formula (6) to derive the optimality criterium described in the MP. Consider the Maximum Likelihood Estimator (MLE), $\hat{X}$, of the signal, X based on N independent measurements of Y from the distribution $P(Y|X)$. The asymptotic distribution of the multivariate MLE has the form of the Multivariate Normal distribution (MVN) [15]

$$\frac{1}{\sqrt{N}}(X - \hat{X}) \sim MVN(0, FI(X)^{-1}). \tag{8}$$

Denote $\sigma^2(X) = |FI(X)^{-1}|$. Then, the optimal input can be written as

$$P^*(X) \propto \frac{1}{\sigma(X)}. \tag{9}$$

The optimal input distribution is therefore described in terms of the asymptotic error of the estimator. We interpret the asymptotic error of the estimator as uncertainty about the network input, X, associated with an observed output, Y, and obtain the optimality criterium presented in the MP. The optimality criterium holds under general regularity conditions and does not involve any approximations. The calculation of the optimal input distribution is reduced here to the problem of calculating the FIM. This can be done for an arbitrary system via computationally expensive Monte Carlo Simulations. Efficient computational solution exists in the setting of the Linear Noise Approximation [16], precisely if $P_\Theta(Y|X)$ is approximated by the Multivariate Normal Distribution with mean $\mu(X,\Theta)$ and variance $\Sigma(X,\Theta)$

$$P_\Theta(Y|X) \approx MVN(\mu(X,\Theta), \Sigma(X,\Theta)). \tag{10}$$

Then the FIM is given by [17]

$$FI(X)_{ij} = \frac{\partial \mu(X,\Theta)}{\partial X_i}\Sigma(X,\Theta)^{-1}\frac{\mu(X,\Theta)}{\partial X_j} + \frac{1}{2}\mathrm{tr}\left(\Sigma(X,\Theta)^{-1}\frac{\partial \Sigma(X,\Theta)}{\partial X_i}\Sigma(X,\Theta)^{-1}\frac{\partial \Sigma(X,\Theta)}{\partial X_j}\right), \tag{11}$$

where i, j index elements of the input vector, X. In this setting the optimal input distribution can be found by differentiating the mean and the variance of the output Y with respect to X.
The formula (9) looks mathematically similar to the equations (6) and (7) in [18]. The two results, however, are different. We characterise the optimal distribution in terms of an inference error given by the Fisher information. The result of [18] characterises the optimal distribution in terms of noise in the network and formulae (6) and (7) in [18] are the relevant approximations to our expression (6).

1.3 Information capacity

Now we show how the characterisation of the optimal input (6) provides novel approaches to calculate the information capacity of biochemical networks. We derive two formulae. The first calculates the exact mutual information but approximates $P_\Theta(Y|X)$ via $MVN(\mu(X,\Theta), \Sigma(X,\Theta))$. The second, approximates mutual information equation but does not require any approximations of $P_\Theta(Y|X)$. We call these expressions exact and approximate, respectively.

1.3.1 Exact formula

In this derivation we use the representation of the mutual information as

$$I(Y,X) = H(Y) - H(Y|X). \tag{12}$$

Suppose that $P_\Theta(Y|X) = MVN(\mu(X,\Theta), \Sigma(X,\Theta))$, as it is in the LNA, and the optimal distribution formula has the form (2). The definition (12) of mutual information yields

$$C_\Theta = -\int P_\Theta(Y)\log P_\Theta(Y)dx - \int \frac{1}{V}\sqrt{|FI(X)|}\frac{1}{2}\log\left((2\pi e)^k|\Sigma(X,\Theta)|\right)dX, \tag{13}$$

where

$$P_\Theta(Y) = \int P_\Theta(Y|X)P^*_\Theta(X)dX, \tag{14}$$

and V is the normalising constant of (6). The equation (13) allows for a direct numerical computation of information capacity given that $\mu(X,\Theta)$, $\Sigma(X,\Theta)$ and $FI(X)$ can be evaluated.

1.3.2 Approximate solution

In this derivation we extend the results presented in [19], making use of the explicit characterisation of the optimal input (6). The concept behind this derivation is to replace the mutual information $I(X,Y)$ between the signal, X, and output, Y, with the mutual information $I(X,\hat{X})$ between the signal, X, and the maximum likelihood estimator of X denoted here by $\hat{X}$. In general $I(X,Y) \geq I(X,\hat{X})$: an estimator cannot contain more information about the signal than the output. Nevertheless in the asymptotic limit (i.e. $\hat{X}$ is calculated based on sufficiently many copies of Y) both quantities are equal [20]. Therefore if Y contains sufficient information about X, $I(X,\hat{X})$ should serve as a good proxy for $I(X,Y)$.
Therefore we calculate

$$I(X,\hat{X}) = H(\hat{X}) - H(\hat{X}|X).$$

The asymptotic theory of the MLEs states that for $\hat{X}$ obtained from N independent copies Y has the MVN distribution

$$\hat{X} \sim MVN(X, \frac{1}{N}FI(X)^{-1}).$$

Assuming $N = 1$ and using the formula for the entropy, $H(Z) = \frac{1}{2}\log((2\pi e)^k|\Sigma|)$ of a k-dimensional random variable $Z \sim MVN(\mu,\Sigma)$, we have

$$I(X,\hat{X}) \approx H(\hat{X}) - \int P(X)\frac{1}{2}\log\frac{(2\pi e)^k}{|FI(X)|}dX.$$

If the variability of the signal, X, is large compared to uncertainty in the estimator, which is expected to be for most of the signal transduction systems, then $H(X) \approx H(\hat{X})$ and we can write

$$I(X,\hat{X}) \approx H(X) - \int P(X)\frac{1}{2}\log\frac{(2\pi e)^k}{|FI(X)|}dX.$$

Mutual information is maximised for $P^*_\Theta(X) \propto \sqrt{|FI(X)|}$ therefore substituting

$$P^*_\Theta(X) = \frac{1}{V}\sqrt{|FI(X)|},$$

where V is the normalising constant, yields

$$\begin{aligned} C_\Theta &= -\int \frac{1}{V}|FI(X)|^{\frac{1}{2}} \log(\frac{1}{V}|FI(X)|^{\frac{1}{2}})dX - \int \frac{1}{V}|FI(X)|^{\frac{1}{2}} \log \frac{(2\pi e)^{\frac{k}{2}}}{|FI(X)|^{\frac{1}{2}}}dX \\ &= -\int \frac{1}{V}|FI(X)|^{\frac{1}{2}} \log \frac{1}{V}dX - \int \frac{1}{V}|FI(X)|^{\frac{1}{2}} \log |FI(X)|^{\frac{1}{2}}dX \\ &\quad -\int \frac{1}{V}|FI(X)|^{\frac{1}{2}} \log((2\pi e)^{\frac{k}{2}})dX + \int \frac{1}{V}|FI(X)|^{\frac{1}{2}} \log |FI(X)|^{\frac{1}{2}}dX \\ &= \log V - \log(2\pi e)^{\frac{k}{2}}. \end{aligned}$$

Finally we obtain

$$C_\Theta = \log\left(\frac{1}{(2\pi e)^{\frac{k}{2}}}\int |FI(X)|^{\frac{1}{2}}dX\right). \tag{15}$$

Interpretation of the formula (15)
The above provides an intuitive interpretation of the information capacity. The uncertainty about the signal is associated here with the estimator's distribution $\hat{X} \sim MVN(X, FI(X)^{-1})$. In the one dimensional scenario the integration $\int \frac{1}{\sigma(X)}dX = \int |FI(X)|^{\frac{1}{2}}dX$ measures how many standard deviations of the estimator can be fitted into the volume of possible signals.

1.3.3 Comparison of both solutions

In section 5 we calculate information capacities for the *E. Coli* nitrogen sensing mechanism using both approaches: formula (13) and (15). The two results are in excellent agreement (see Table 3).

1.4 Novelty of the approach

The novelty of our approach results from linking the statistical theory of reference priors [10–13, 19] with information-theoretic analysis of biochemical reaction networks. Together with our previous framework to calculate the Fisher information [16, 21] this constitutes an efficient computational approach to analyse information processing in biochemical networks. The approach is efficient enough to deal with complex multiple input – multiple output systems, which have constituted a substantial obstacle to previous approaches [2, 22]. Moreover, establishing a link between statistical inference and biochemical signalling has the potential to provide further insight into the function and evolutionary drivers of biochemical reaction networks.

2 Model of the nitrogen sensing mechanism in *Escherichia coli*

The nitrogen status is predominantly sensed by the abundance of assimilated nitrogen in the form of glutamine (GLN), while relative nitrogen shortage is signaled by Krebs cycle metabolite α-ketoglutarate (αKG). A bi-cyclic protein modification system integrates these signals to control the activity of the nitrogen assimilation protein glutamine synthase (GS). GLN binds to the bifunctional uridylyl transferase/removase (UT/UR), stimulating the deuridylylation of the trimeric PII. PII acts on the bifunctional adelylyl-transferase/removase (AT/AR) of GS, where PII stimulates the transfer and PII-UMP the removase activity of AT/AR.
αKG binds to PII, decreasing its regulatory capacity [23, 24]. The mathematical model that we adopt here was proposed in ref. [25] based on the authors' experimental work and other results including these described in [23, 24, 26–28].

In the considered model the state of the sensing mechanism can be therefore mathematically defined as the vector

$$Y = (Y_1, Y_1^*, Y_2, Y_2^*) = (PII, PII - UMP, GS - AMP, GS)$$

containing the number of molecules involved. The state of the system is enzymatically controlled via four reactions

$$\begin{array}{rccc} R_1: & PII & \xrightarrow{v_{UT}} & PII - UMP \\ R_2: & PII - UMP & \xrightarrow{v_{UR}} & PII \\ R_3: & GS & \xrightarrow{v_{AT}} & GS - AMP \\ R_4: & GS - AMP & \xrightarrow{v_{AR}} & GS \end{array}$$

with the following stoichiometry

$$\begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

The reaction rates

$$F(Y, X, \Theta) = (v_{UT}, v_{UR}, v_{AT}, v_{AR})$$

can be written as Michaelis-Menten terms

$$\begin{aligned} v_{UT} &= V_{max_{UT}} \left(\frac{[PII]}{K_{PII_{UT}} + [PII]} \right) \left(\frac{K_{GLN_{UT}}}{K_{GLN_{UT}} + [GLN]} \right), \qquad (16) \\ v_{UR} &= V_{max_{UR}} \left(\frac{[PII]_{\text{Tot}} - [PII]}{K_{PII_{UR}} + [PII]_{\text{Tot}} - [PII]} \right) \left(\frac{[GLN]}{K_{GLN_{UR}} + [GLN]} \right), \\ v_{AT} &= V_{max_{AT}} \left(\frac{[GS]}{K_{GS_{AT}} + [GS]} \right) \left(\frac{[PII]}{K_{PII_{AT}} \left(1 + \frac{[KG]}{K_{KG_{AT}}} \right) + [PII]} \right) \left(\frac{[GLN]}{K_{GLN_{AT}} + [GLN]} \right), \\ v_{AR} &= V_{max_{AR}} \left(\frac{[GS]_{\text{Tot}} - [GS]}{K_{GS_{AR}} + [GS]_{\text{Tot}} - [GS]} \right) \left(\frac{[PII]_{\text{Tot}} - [PII]}{K_{PII_{AR}} + [PII]_{\text{Tot}} - [PII]} \right) \left(\frac{K_{GLN_{AR}}}{K_{GLN_{AR}} + [GLN]} \right), \end{aligned}$$

involving the parameters

$$\begin{aligned} \Theta = (&V_{max_{UT}}, K_{GLN_{UT}}, K_{PII_{UT}}, V_{max_{UR}}, K_{GLN_{UR}}, K_{PII_{UR}}, V_{max_{AT}}, \qquad (17) \\ &K_{GLN_{AT}}, K_{PII_{AT}}, K_{GS_{AT}}, K_{KG_{AT}}, V_{max_{AR}}, K_{GLN_{AR}}, K_{PII_{AR}}, K_{GS_{AR}}) \end{aligned}$$

and constraints

$$\begin{aligned} [PII]_{\text{Tot}} &= [PII] + [PII - UMP] \\ [GS]_{\text{Tot}} &= [GS] + [GS - AMP] \end{aligned}$$

2.1 Deterministic model

The reaction rates and stoichiometry matrix define the deterministic description of the system [29]

$$\frac{dY}{dt} = f(Y, X, \Theta), \qquad (18)$$

where $f(Y, X, \Theta) = S \cdot F(Y, X, \Theta)$ and $\cdot$ denotes the dot product.

2.2 Stochastic model

Model stoichiometry and rates define a model description via the Chemical Master Equation (CME) that can be approximated via the Linear Noise Approximation [29–31]. In simple terms in the LNA the state Y is divided into a deterministic part ϕ and a stochastic part ξ so that

$$Y = \phi + \xi. \tag{19}$$

The deterministic state is described by an equation equivalent to (18)

$$\frac{d\phi}{dt} = S\ F(\phi(s), X, \Theta) \tag{20}$$

The stochastic part is given by the linear stochastic differential equation

$$d\xi = A(\phi, X, \Theta)\xi(t)dt + \sum_{j=1}^{4} S{\cdot}j\sqrt{F_j(\phi, X, \Theta)}dW_j, \tag{21}$$

where

$$\{A(\phi, X, \Theta)\}_{ik} = \sum_{j=1}^{4} s_{ij}\frac{\partial f_j(\phi, X, \Theta)}{\partial \phi_k}, \tag{22}$$

s_{ij} is the ij element of the matrix S, $S_{\cdot j}$ is its jth column, and dW_j are increments of the independent Wiener processes. Equations (20)-(21) imply that the distribution of Y at all times is multivariate normal

$$Y \sim MVN(\phi(t), \Sigma(t)) \tag{23}$$

with the mean $\phi(t)$ given by a solution of (20), and variance $\Sigma(t)$ described as

$$\frac{d\Sigma}{dt} = A(\phi, X, \Theta)\Sigma + \Sigma A(\phi, X, \Theta)^T + D(t), \tag{24}$$

where $D(t) = S \cdot E \cdot S^T$, and E is a diagonal matrix with F on the diagonal i.e. $E = \mathrm{diag}(F(\phi, X, \Theta))$.

2.2.1 Calculation of the optimal input distribution and information capacity for the stochastic model

Having a model of the nitrogen sensing mechanism written in the LNA we can calculate the FIM using the formula (11) as it requires only differentiation of mean ϕ and variance Σ with respect to the input signal X. This is implemented in the StochSens matlab package [21]. Calculation of the information capacity via formulae (13) and (15) is then performed by numerical integration.

3 Experimental setup and data acquisition

In order to examine experimentally the information processing capacity of the nitrogen sensing mechanism in the context of the developed methodology we designed an experimental setting in which *E. coli* are exposed to a broad range of dynamically varying nitrogen availability conditions. The collected data allowed to calibrate the model (18), confirm its validity and demonstrate the analogous character of information processing in the system. The details of the experimental methodology are described below.

3.1 Bacterial strains

The main experiment were performed using wild type *E. coli* K12 (NCM3722) [32]. Isogenic strains within frame, single gene knock out deletions of glnK and glnD were used to confirm that within our experimental conditions the PII homologous GlnK does not influence the adenylylation state of GS; and that PII uridylylation depends exclusivley on UT/UR in response to glutamine levels. NCM3722 glnK and glnD deletions were obtained through P1 phage transduction [33], using BW25113 glnk knockout (JW0440) and BW25113 glnD knockout (JW0162) as donor strains, respectively, provided by the Keio collection [33]. Following phage transduction and selection for kanamycin resistance, deletions were verified by locus and kanamycin specific PCR and genomic sequencing.

3.2 Nitrogen starvation experiment: growth conditions and sampling

Time course experiments were designed to capture two transitions, from nitrogen-rich to nitrogen-poor and vice versa. *E. coli* NCM3722 WT cells were grown in Gutnick minimal media, supplemented with 10 mM NH4Cl, 0.4% (w/v) glucose and Ho-Le trace elements. Main cultures were inoculated from this media into nitrogen-limiting media (3 mM NH4Cl), so that nitrogen starvation and growth arrest occurred during otherwise mid-exponential growth at an OD600 of 0.6. The point of growth arrest (tOD) defined subsequent sampling time points. After approximately one generation time as recorded for Gutnick media [34], fresh NH4Cl was added to again obtain a NH4Cl concentration of 3 mM (Figure 1). Samples for proteomic and metabolic analysis were taken during initial exponential growth at OD600 of 0.15 (t_0, proteomics only), 0.3 (t_1), 0.5 (t_2, proteomics only) and after growth arrest, at an approximate OD600 of 0.6 (t_{OD}). To investigate changes throughout starvation, samples were taken at 20 minutes (t_3, proteomics only) and 40 minutes (t_4) initial growth arrest. After fresh NH4Cl was added, samples were taken at 30 seconds (t_5), 2 min (t_6), 5 min (t_7), 10 min (t_8) and 30 (t_9) minutes. Sampling time points are shown relative to t_5 i.e $t_5 = 0$. (see Figure1A in the MP).

3.3 Quantification of GS, PII, GS-AMP and PII-UMP via MRM-MS

Multiple reaction monitoring mass spectrometry (MRM-MS) in conjunction with protein standard purification (PSAQ) is a robust method to accurately determine the absolute protein concentration of targeted proteins from complex biological samples [34, 35]. MRM-MS has been previously successfully employed to quantify post-translational phosphorylation states [36]. We describe a MRM-MS with PSAQ approach to directly quantify post-translational protein uridylylation and adenylylation. We measured uridylylated and adenylylated peptides of PII and GS. Their relationships to other signature tryptic peptides derived from PII and GS allowed determining the absolute concentrations of post translationally modified and unmodified species. Preparation of protein standards: GS and PII were doubly labelled with L-Arginine (13C6, 15N4) and L-Lysine (13C6, 15N4) in vivo, over-expressed, purified and protein concentrations determined as described in [34]. Purity of GS (93%) and PII (91%) was assessed by quantification of fluorescence following SDS-PAGE and staining with SYPRO Ruby Protein Gel Stain (Invitrogen) (Figure 1). Isotopic labelling efficiency of purified GS (97%) and PII (100%) protein standard peptides were determined by MRM-MS.

Total GS and PII concentrations were determined using isotopically labelled GS and PII as internal standards (IS) as described in [34]. Absolute protein concentrations for GS and PII (the analytes, A) were derived as follows:

$$[A] = [IS]\frac{[A] - IS_{\text{unlabeled}}}{IS_{\text{labeled}}} \tag{25}$$

Were [A] is the analyte concentration, [IS] the concentration of the labelled standard, A, $IS_{\text{unlabelled}}$ and IS_{labelled} are the MS measured signals of the analyte peptide, the unlabelled IS and the labelled IS, respectively (see Figure 2). Quantity $A - IS_{\text{unlabeled}}$ simply corrects for the small amount of unlabelled

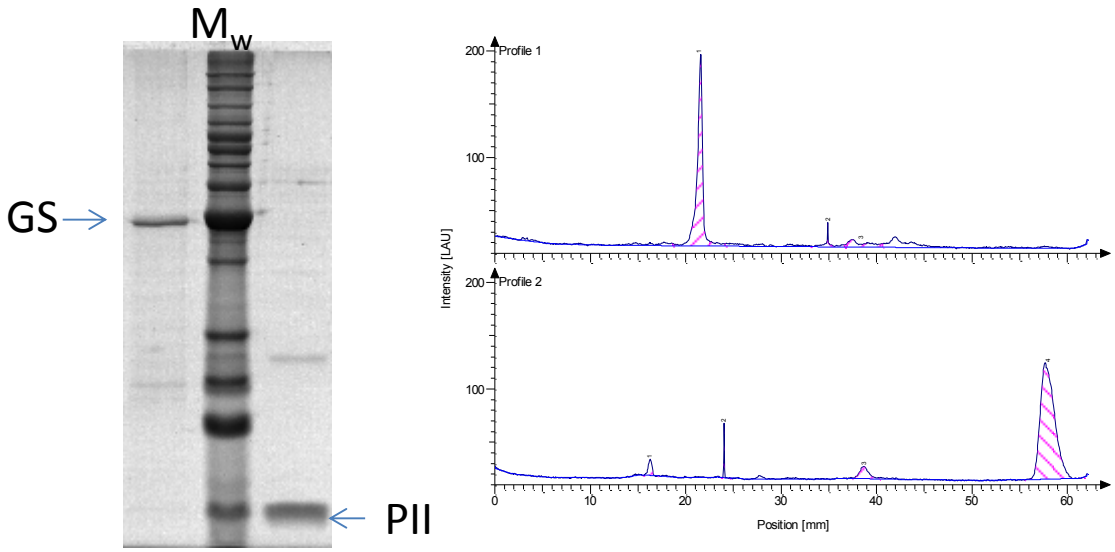

Figure 1: Purity of labelled GS and PII standards. A) SYPRO Ruby stained SDS-PAGE gel fluorescence images. Purified histidine tagged GS and PII migrated at their expected molecular masses of 53469 Da and 13991 Da, respectively, as judged from a molecular weight ladder (Mw, Invitrogen, Novex Sharp). B) Lane band intensity profiles for 1D evaluation for GS (Profile 1) and PII (profile 2) using the Aida software (Fuji) with standard settings. Protein purities were calculated from integrated peak areas.

IS in the standard mix (3% in the case of GS and 0% for PII). Cellular protein copy numbers were calculated from the protein concentrations in the MS samples and using the literature values for *E. coli* cell volume (1 fl) and cell density of 9.1x109 cells/OD600 [34].

3.3.1 Quantification of GS-AMP and PII-UMP

Although the total concentration of GS and PII in the internal standard [IS] is known, the post-translational state is not. However, in each sample the total protein concentration should be equal to the sum of the unmodified and post-translationally modified protein concentrations ([peptide tot]= [peptide-ptm] + [peptide +ptm]). We used this relationship to determine the standard concentrations c_{ptm} for converting the MS signals $[S]_{\text{ptm}}$ measured for GS-AMP and PII-UMP to absolute concentrations, by solving $[A] = c_{\text{ptm}}[S]_{\text{ptm}}$ for a collection of n samples using matrix inversion, with $[A]$ a $1 \times n$ vector of the total concentrations of PII or GS for a collection of n samples, $[S]_{\text{ptm}}$ is a $2 \times n$ matrix, consisting of two vectors of the corresponding MS signals of [PII-UMP] and unmodified [PII], respectively [GS-AMP] and unmodified [GS], in the same collection of samples, and c_{ptm} is a 1×2 vector of the standard concentrations. We obtained the following standard concentrations in the MS samples: [GS] 12.55 nM, [GS-AMP] 19.86 nM, [PII] 0.43 nM, [PII-UMP] 0.78 nM. These concentrations were directly used to determine the respective peptide concentrations. Copy numbers per cell were determined in the same way as for total concentrations of GS and PII.

3.4 Quantification of GLN and αKG

Glutamine and αKG were quantified as previously described in [34].

Peptide-protein-transition-(internal standard, is)	retention time (min)	Q1 m/z	Q2 m/z	Q3 collision energy
IPVVSSPK-GS-1a	21.4	413.8	517.3	20
IPVVSSPK-GS-1b	21.4	413.8	616.4	20
IPVVSSPK-GS-1c	21.4	413.8	713.4	20
IPVVSSPK-GS-1a-is	21.4	417.8	525.3	20
IPVVSSPK-GS-1b-is	21.4	417.8	624.4	20
IPVVSSPK-GS-1c-is	21.4	417.8	721.4	20
NLYDLPPEEAK-GS-1a	28.7	644.8	670.3	25
NLYDLPPEEAK-GS-1b	28.7	644.8	1061.3	25
NLYDLPPEEAK-GS-1c	28.7	644.8	228.1	28
NLYDLPPEEAK-GS-1a-is	28.7	648.8	678.3	25
NLYDLPPEEAK-GS-1b-is	28.7	648.8	1069.3	25
NLYDLPPEEAK-GS-1c-is	28.7	648.8	228.1	28
NLY(a)DLPPEEAK-GS-1a	26.3	809.8	670.3	25
NLY(a)DLPPEEAK-GS-1b	26.3	809.8	136.1	30
NLY(a)DLPPEEAK-GS-1a-is	26.3	813.8	678.3	25
NLY(a)DLPPEEAK-GS-1b-is	26.3	813.8	136.1	30
IFVFDVAR-PII-2a	38	483.8	706.4	25
IFVFDVAR-PII-2b	38	483.8	607.3	25
IFVFDVAR-P-11-2c	38	483.8	853.5	25
IFVFDVAR-PII-2a-is	38	488.8	716.4	25
IFVFDVAR-PII-2b-is	38	488.8	617.3	25
IFVFDVAR-PII-2c-is	38	488.8	863.5	25
GAEYMVDFLPK-PII-1a	40.6	635.3	718.4	30
GAEYMVDFLPK-PII-1b	40.6	635.3	849.5	30
GAEYMVDFLPK-PII-1c	40.6	635.3	244.2	30
GAEYMVDFLPK-PII-1a-is	40.6	639.3	726.4	30
GAEYMVDFLPK-PII-1b-is	40.6	639.3	857.5	30
GAEYMVDFLPK-PII-1c-is	40.6	639.3	252.2	30
GAEY(U)MVDFLPK-PII-1a	39	788.4	244.2	30
GAEY(U)MVDFLPK-PII-1b	39	788.4	718.4	30
GAEY(U)MVDFLPK-PII-1a-is	39	792.4	252.2	30
GAEY(U)MVDFLPK-PII-1b-is	39	792.4	726.4	30
GAEYM(O)VDFLPK-PII-1b	38	643.3	244.2	30
GAEYM(O)VDFLPK-PII-1b-is	38	647.3	252.2	30
GAEY(U)M(O)VDFLPK-PII-1a	37	796.4	244.2	30
GAEY(U)M(O)VDFLPK-PII-1a-is	37	800.4	252.2	30

Table 1: Signature peptides and transitions of GS and PII used for MRM-MS with PSAQ. Measurements were carried out on a Applied Biosystems QTrap in triple quadrupole mode (Q1,Q2,Q3) as described in [34]. The first column shows the peptide amino acid sequence, the parent protein, the detected transition identifier and indicating peptides from internal standards (-is). Chromatographic retention times and other MS characteristics are as indicated. We verified that the methionine comprising uridylylatable PII peptide was not oxidised.

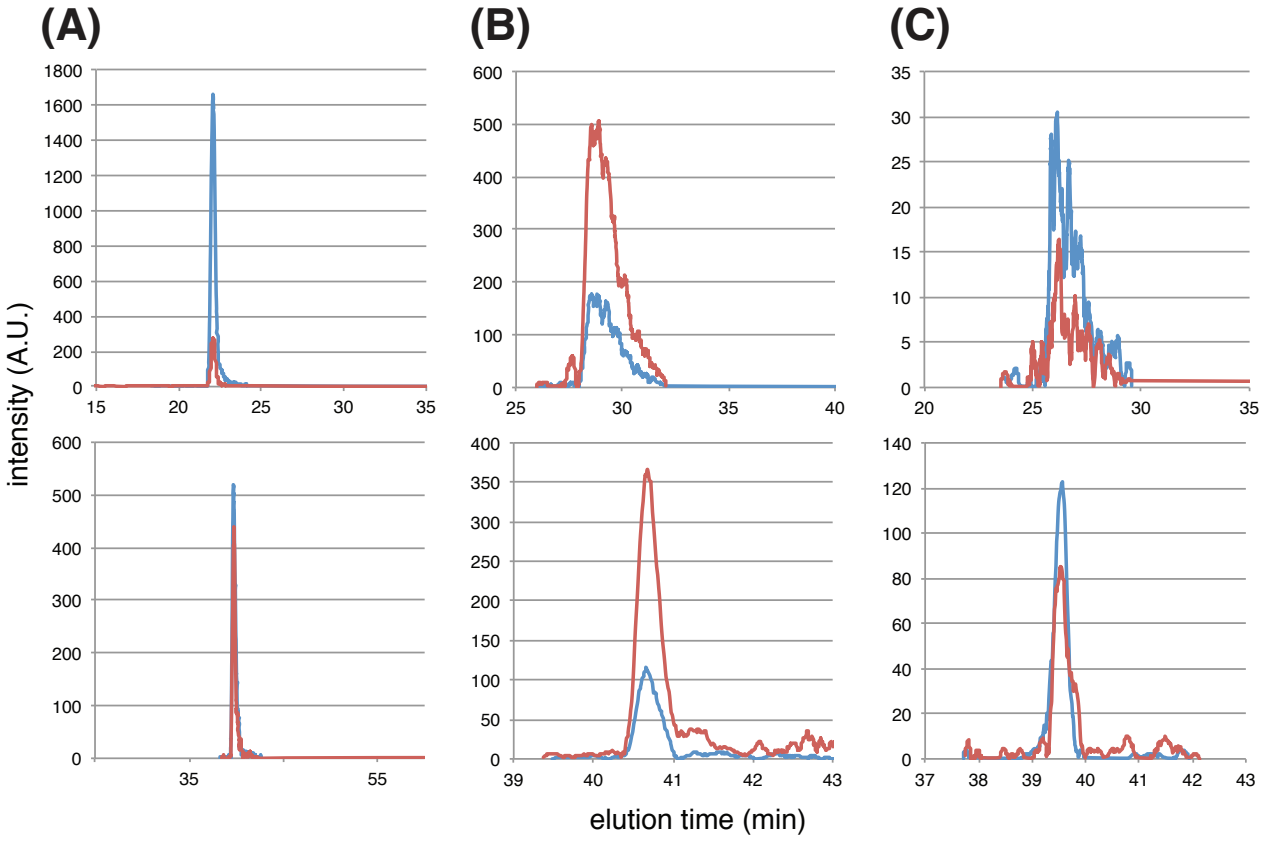

Figure 2: Extracted ion chromatograms of representative peptides of GS (upper panel) and PII (lower panel) derived from isotopic labelled internal standards (blue) or sample proteins (red). A) GS non-adenylylated peptide IPVVSSPK-1a, B) GS non-adenylylated peptide NLYDLPPEEAK-1a, C) GS adenylylated peptide NLY(AMP)DLPPEEAK-1a, D) PII non uridylylated peptide uridylylated peptide IFVFDVAR-2a, E) PII non uridylylated peptide GAEYMVDFLPK-1a, F) PII uridylylated peptide GAEY(UMP)MVDFLPK-1a.

3.5 Uridylylation and adenylylation in glnK and glnD knockout

E. coli encodes two PII proteins, PII and its homologue GlnK. Both proteins were proposed to have redundant functionalities and both are specifically uridylilated and de-uridylylated by the bifunctional uridylyl transferase/removase, encoded by glnD. In order to validate our model, which does not include a signalling contribution by GlnK, we measured PII, PII-UMP, GS and GS-AMP in NCM3722 glnK knockout in the same way as we did for NCM3722 (Figures 3A-C). Under our conditions, we found no significant differences between NCM3722 glnK knockout and NCM3722 with regards to the total copy numbers of PII-UMP and GS-AMP. These results indicate that GlnK has no discernable effect on PII uridylylation and GS adenylylation.
In order to confirm that PII uridylylation depends exclusively on UT/UR in response to glutamine levels we measured PII, PII-UMP in NCM3722 glnD knockout, again in the same way as we did for NCM3722. In the knockout strain PII remains in de-uridylylated over the course of the experiment (Figure 3D), which confirmed that the action of UT/UR controls PII uridylylation in response to changes in glutamine levels.

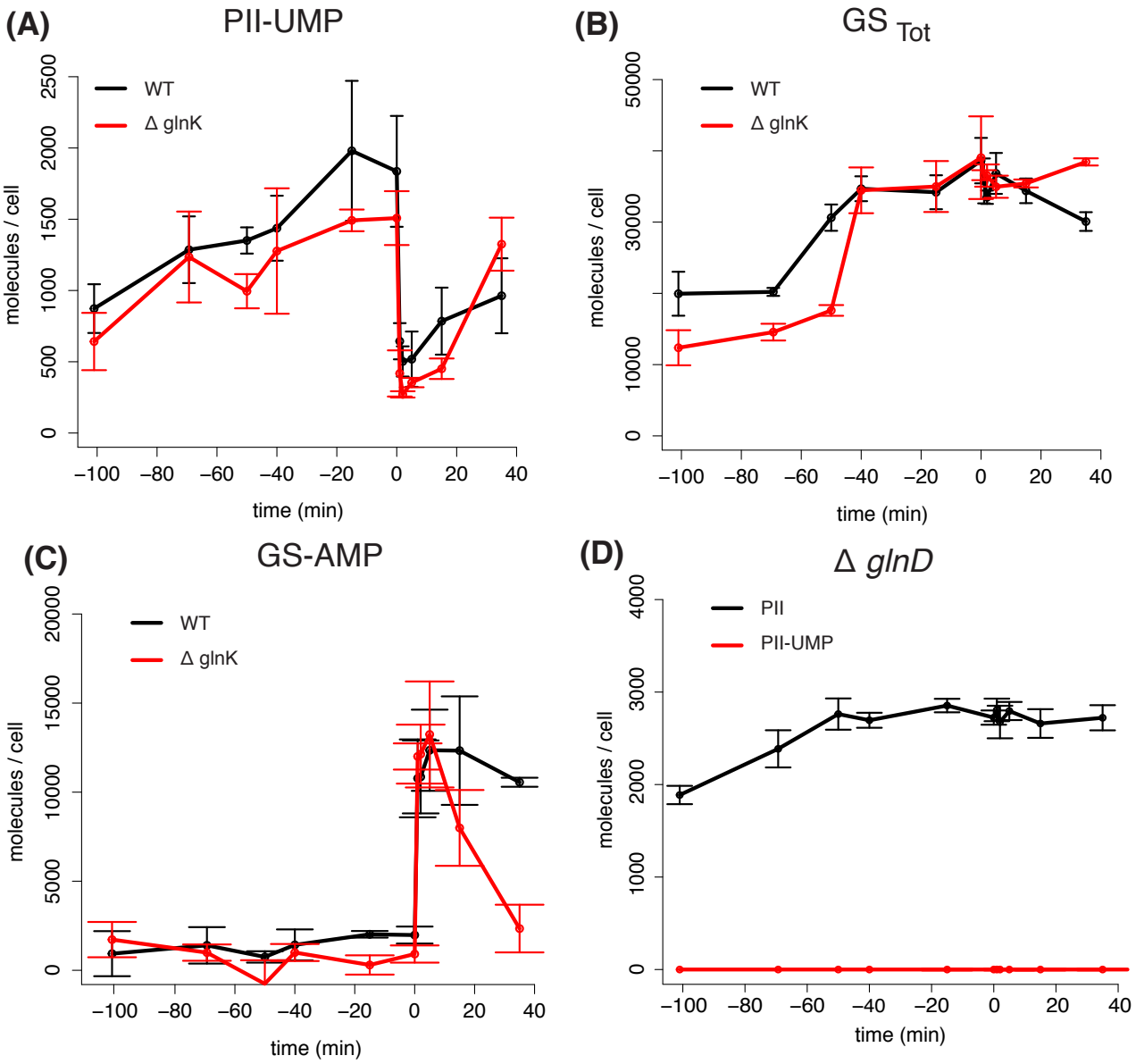

Figure 3: (A) PII is equally uridylylated in wild type (WT) NCM3722 and in glnK knockout (Δ glnK). (B,C) Total levels of GS and GS-AMP exhibit similar dynamics in WT and in (Δ glnK). (D) PII and is not uridylated in UT/UR (glnD) knockout (Δ glnD). (A,B,C) demonstrate that the PII homologous GlnK does not influence the adenylylation state of GS whereas (D) shows that PII uridylylation depends exclusively on UT/UR. Error bars were obtained using three replicates.

4 Parameter inference

The information capacity C_Θ and the optimal input distribution $P_\Theta(X)$ depend on model parameters Θ. Therefore we employed a Bayesian inference implemented via Approximate Bayesian Computation [37] in order to determine the posterior distribution of model parameters $P(\Theta|\text{Data})$ based on the data presented in the Figure 2B in the MP and on the deterministic model (18). More precisely we used the software ABC-SysBio [38, 39]. The initial conditions $Y_1^*(0)$ and $Y_2^*(0)$ were inferred simultaneously to the parameters whereas the initial conditions of $Y_1(0)$ and $Y_2(0)$ were set to be equal to the total level of PII and GS concentrations minus $Y_1^*(0)$ and $Y_2^*(0)$, respectively. The input signals $X(t) = (X_1(t), X_2(t))$ were spline interpolated based on αKG and GLN experimental measurements (see Figure 2B in the MP). Similarly the total amount of GS and PII molecules, $[PII]_{\text{Tot}}$, $[GS]_{\text{Tot}}$ needed to evaluate reaction rates were also spline interpolated (see Figure 4). We used uniform prior distributions for each of the model parameters; the lower and upper bounds of theses distributions are described in Table 2. The obtained marginal posterior distributions are plotted in the Figure 5 and the parameters providing the best fit to the data are presented in the Table 2. The Figure 2B in the MP shows the fit to the data for parameters sampled from the posterior distribution.

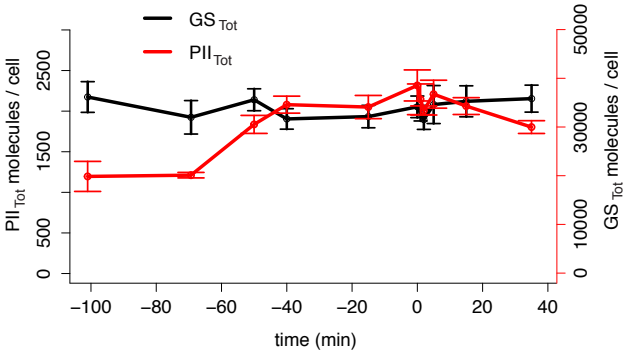

Figure 4: Experimentally measured total levels of PII and GS proteins. Total levels were spline interpolated to calculate reaction rates (16) required to simulate the model (18). Error bars were obtained using three replicates.

5 Information-theoretic properties of the nitrogen sensing mechanism

The analysis of the information-theoretic properties of the nitrogen sensing mechanism involved calculation of the optimal input distribution $P^*_\Theta(X)$ and information capacity C_Θ for parameter values given by the posterior distribution. Precisely, we calculated $P^*_\Theta(X)$ for maximum posterior estimates given in the Table 2 and posterior distribution of capacity $P(C|\text{Data})$ using $P(\Theta|\text{Data})$ and functional dependence of capacity, C_Θ, on Θ. We have assumed that output Y is defined as a sample from the stationary distribution of the system for a given level of stimulation X.

We first considered the case of GLN sensed through PII (see Figure 1 in the MP). Thereafter, the same analysis was performed considering both of the two inputs $(\alpha KG, GLN)$ and two outputs (see Figure 1 in the MP). For comparison we also analysed αKG sensing via GS at low, medium and high glutamine levels. Inferred capacities are presented in the Figure 3 in the MP, Figure 6 and Table 3. Calculations of the information capacity using both, the exact (13) and approximate (15), formulae give virtually identical results.

Parameter	Prior (l.b.)	Prior (u.b.)	Best Fit	Posterior (0.1-q.)	Posterior (0.9-q.)	Units
UT						
$V_{max_{UT}}$	1×10^2	1×10^8	7.7×10^7	3.9×10^7	9.4×10^7	molecules / min
$K_{GLN_{UT}}$	1	1×10^8	1.6×10^6	1.7×10^6	7.1×10^7	molecules
$K_{PII_{UT}}$	1×10^2	1×10^9	6.3×10^7	1.9×10^7	1.3×10^8	molecules
UR						
$V_{max_{UR}}$	1×10^2	1×10^8	8×10^7	3.5×10^7	9.3×10^7	molecules / min
$K_{GLN_{UR}}$	1	1×10^8	1.6×10^6	1.7×10^6	7.1×10^7	molecules
$K_{PII_{UR}}$	1×10^2	1×10^8	2.3×10^7	4.0×10^5	1.6×10^7	molecules
AT						
$V_{max_{AT}}$	1×10^4	1×10^6	3.1×10^5	1.8×10^5	6.7×10^5	molecules/min
$K_{GLN_{AT}}$	1×10^5	1×10^7	9.8×10^6	4.5×10^6	9.8×10^6	molecules
$K_{PII_{AT}}$	1×10^{-1}	1×10^2	8×10^{-1}	8.3×10^{-1}	2.4×10^1	molecules
$K_{GS_{AT}}$	1×10^{-1}	1×10^2	9.9×10^1	1.3×10^1	9.8×10^2	molecules
$K_{KG_{AT}}$	1×10^2	1×10^4	4.3×10^3	3.6×10^3	9.6×10^3	molecules
AR						
$V_{max_{AR}}$	1×10^5	1×10^7	7.1×10^6	3.9×10^6	9.4×10^6	molecules/min
$K_{GLN_{AR}}$	1×10^5	1×10^7	9.8×10^6	4.5×10^6	9.8×10^6	molecules
$K_{PII_{AR}}$	1×10^{-1}	1×10^3	5.6×10^2	2.6×10^1	5.9×10^2	molecules
$K_{GS_{AR}}$	1×10^5	1×10^7	1.3×10^6	7.5×10^5	2.3×10^6	molecules

Table 2: Summary of the inference results. We obtained posterior distribution of the model parameters from experimental data presented in the Figure 2B in the MP. The lower bound (l.b.) and the upper-bound (u.b.) for the uniform prior for all the parameters; the 0.1- and 0.9-quantiles of the posterior distribution as well as the parameters providing the best fit to the data are presented.

GLN-PII		(GLN,αKG)-(PII,GS)		αKG-GS					
				low GLN		medium GLN		high GLN	
Exact	Approx.	Exact	Approx.	Exact	Approx.	Exact	Approx.	Exact	Approx.
4.14 (0.04)	4.11 (0.02)	8.29 (0.57)	8.15 (0.63)	2.99 (0.23)	2.9 (0.25)	1.37 (0.14)	1.07 (0.18)	1.27 (0.14)	0.94 (0.18)

Table 3: Capacities for sensing of 1) GLN via PII; 2) (αKG, GLN) via (PII,GS); and 3) αKG via GS at low GLN ($1 \cdot 10^5$), medium GLN ($2.5 \cdot 10^5$) and high GLN ($6 \cdot 10^5$). Numbers are posterior means (standard deviations in brackets) expressed in bits.

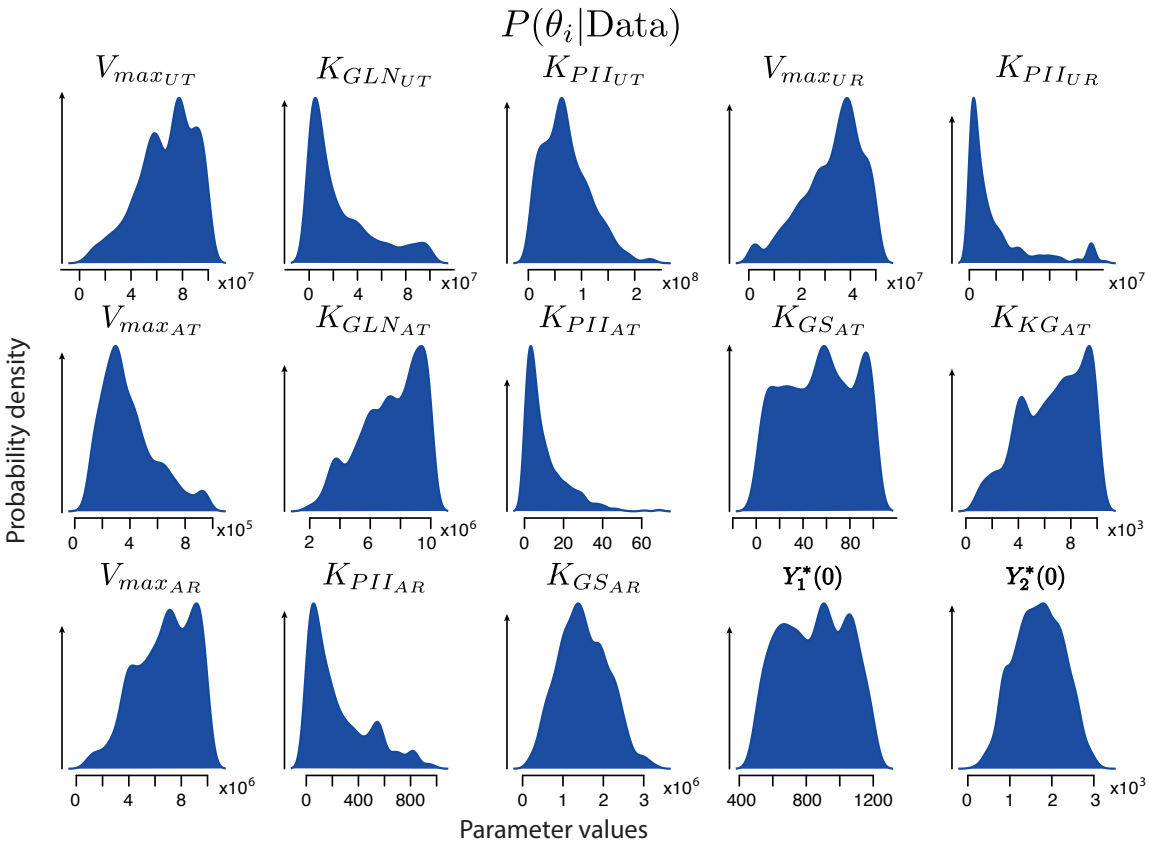

Figure 5: Inferred marginal posterior distribution for all the model parameters. The parameters $K_{gln_{UT}}$ and $K_{gln_{UR}}$ as well as $K_{gln_{AT}}$ and $K_{gln_{AR}}$ are equal, therefore only the distribution over $K_{gln_{UT}}$ and $K_{gln_{AT}}$ is shown here. In addition, the posterior distribution over the initial conditions are represented.

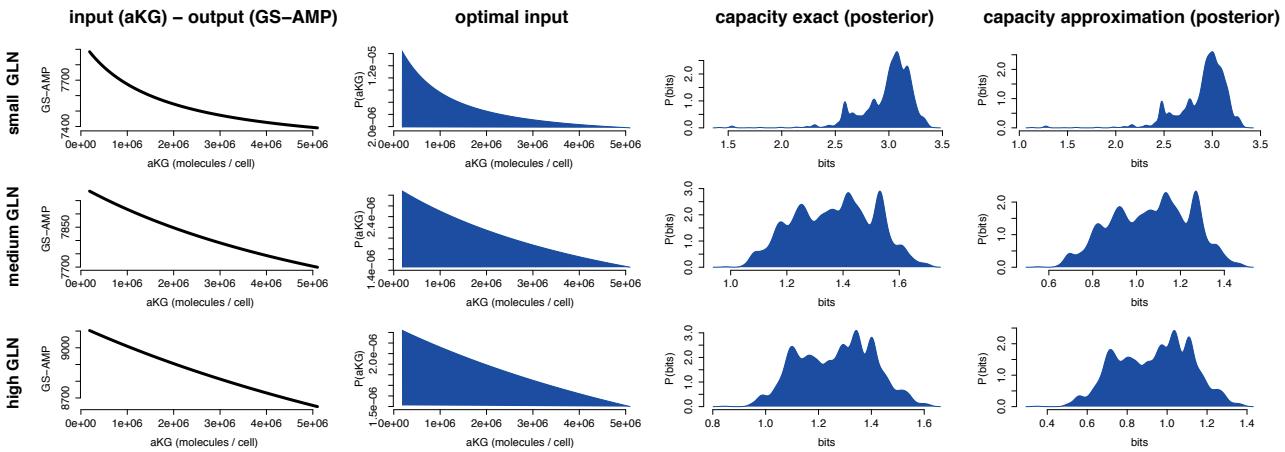

Figure 6: αKG sensing via GS: input-output relation, optimal input distribution, posterior of capacity using the exact (13) and approximate (3) formulae at low GLN ($1 \cdot 10^5$), medium GLN ($2.5 \cdot 10^5$) and high GLN ($6 \cdot 10^5$).

Appendix 1

The regularity conditions required for the validity of the formula (6)

Condition 1 The density $P(Y|X)$ is twice continuously differentiable in X for almost every Y and there exists a $\delta = \delta(X)$ such that $\forall_{j,k}$:

$$\mathbb{E} \sup_{X':|X'-X|<\delta(X)} \left| \frac{\partial^2}{\partial X'_j X'_k} \log P(Y|X') \right|^2$$

is finite and continuous as a function of X and, additionally, $\forall_j \forall_{\epsilon>0}$:

$$\mathbb{E} \left| \frac{\partial}{\partial X_j} \log P(Y|X) \right|^{2+\epsilon}$$

is finite and continuous as a function of X.
Condition 2 $FI(X)$ is positive definite for all $X \in \Xi$.
Condition 3 The parametrization of the family $\{P(Y|X) : X \in \Xi\}$ is one-to-one.
Condition 4 Ξ is a compact set.

The above general conditions, that ensure optimality of the distribution (6), are expected to be satisfied for all commonly used probability distributions used to model biochemical reaction networks.

References

[1] Blahut, R. E., 1972. Computation of channel capacity and rate-distortion functions. *Information Theory, IEEE Transactions on* 18:460–473.

[2] Cover, T. M., and J. A. Thomas, 2012. Elements of information theory. John Wiley & Sons.

[3] Tkačik, G., C. G. Callan Jr, and W. Bialek, 2008. Information capacity of genetic regulatory elements. *Physical Review E* 78:011910.

[4] Cheong, R., A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, 2011. Information transduction capacity of noisy biochemical signaling networks. *science* 334:354–358.

[5] Schreiber, S., C. K. Machens, A. V. Herz, and S. B. Laughlin, 2002. Energy-efficient coding with discrete stochastic events. *Neural Computation* 14:1323–1346.

[6] Balasubramanian, V., D. Kimber, and M. J. Berry II, 2001. Metabolically efficient information processing. *Neural Computation* 13:799–815.

[7] Dupuis, F., W. Yu, and F. Willems, 2004. Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information. *In* IEEE International Symposium on Information Theory. 179–179.

[8] Berger, J. O., and J. M. Bernardo, 1992. On the development of reference priors. *Bayesian statistics* 4:35–60.

[9] Bernardo, J., and A. Smith, 2001. Bayesian theory. *Measurement Science and Technology* 12:221–222.

[10] Bernardo, J. M., 1979. Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)* 113–147.

[11] Clarke, B., and A. Barron, 1994. Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference* .

[12] Berger, J. O., J. M. Bernardo, and D. Sun, 2009. The formal definition of reference priors. *The Annals of Statistics* 905–938.

[13] Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 186:453–461.

[14] Berg, H. C., and E. M. Purcell, 1977. Physics of chemoreception. *Biophysical journal* 20:193–219.

[15] Silvey, S., 1975. Statistical inference. Chapman & Hall.

[16] Komorowski, M., M. J. Costa, D. A. Rand, and M. P. H. Stumpf, 2011. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proc.Natl.Acad.Sci. USA* 108:8645–8650.

[17] Scharf, L. L., 1991. Statistical signal processing, volume 98. Addison-Wesley Reading.

[18] Tkačik, G., C. Callan, and W. Bialek, 2008. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences* 105:12265.

[19] Brunel, N., and J.-P. Nadal, 1998. Mutual information, Fisher information, and population coding. *Neural Computation* 10:1731–1757.

[20] Le Cam, L., 1986. Asymptotic methods in statistical decision theory. Springer, springer series in statistics edition.

[21] Komorowski, M., J. Žurauskienė, and M. P. Stumpf, 2012. StochSens–matlab package for sensitivity analysis of stochastic chemical systems. *Bioinformatics* 28:731–733.

[22] Ash, R. B., 1965. Information Theory. Dover Publications, Inc., New York.

[23] Jiang, P., J. A. Peliska, and A. J. Ninfa, 1998. Enzymological characterization of the signal-transducing uridylyltransferase/uridylyl-removing enzyme (EC 2.7. 7.59) of Escherichia coli and its interaction with the PII protein. *Biochemistry* 37:12782–12794.

[24] Jiang, P., A. E. Mayo, and A. J. Ninfa, 2007. Escherichia coli glutamine synthetase adenylyltransferase (ATase, EC 2.7. 7.49): Kinetic characterization of regulation by PII, PII-UMP, glutamine, and α-ketoglutarate. *Biochemistry* 46:4133–4146.

[25] Yuan, J., C. D. Doucette, W. U. Fowler, X.-J. Feng, M. Piazza, H. A. Rabitz, N. S. Wingreen, and J. D. Rabinowitz, 2009. Metabolomics-driven quantitative analysis of ammonia assimilation in E. coli. *Molecular systems biology* 5.

[26] Atkinson, M. R., T. A. Blauwkamp, V. Bondarenko, V. Studitsky, and A. J. Ninfa, 2002. Activation of the glnA, glnK, and nac promoters as Escherichia coli undergoes the transition from nitrogen excess growth to nitrogen starvation. *Journal of bacteriology* 184:5358–5363.

[27] Ventura, A. C., P. Jiang, L. Van Wassenhove, D. Del Vecchio, S. D. Merajver, and A. J. Ninfa, 2010. Signaling properties of a covalent modification cycle are altered by a downstream target. *Proceedings of the National Academy of Sciences* 107:10032–10037.

[28] Bruggeman, F. J., F. C. Boogerd, and H. V. Westerhoff, 2005. The multifarious short-term regulation of ammonium assimilation of Escherichia coli: dissection using an in silico replica. *FEBS Journal* 272:1965–1985.

[29] Van Kampen, N., 2006. Stochastic Processes in Physics and Chemistry. North Holland.

[30] Komorowski, M., B. Finkenstadt, C. Harper, and D. Rand, 2009. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics* 10:343.

[31] Thomas, P., H. Matuschek, and R. Grima, 2013. How reliable is the linear noise approximation of gene regulatory networks? *BMC Genomics* 14:S5.

[32] Vo, J., W. Inwood, J. M. Hayes, and S. Kustu, 2013. Mechanism for nitrogen isotope fractionation during ammonium assimilation by Escherichia coli K12. *Proceedings of the National Academy of Sciences* 110:8696–8701.

[33] Miller, J. H., et al., 1972. Experiments in molecular genetics, volume 60. Cold Spring Harbor Laboratory Cold Spring Harbor, New York.

[34] Schumacher, J., V. Behrends, Z. Pan, D. R. Brown, F. Heydenreich, M. R. Lewis, M. H. Bennett, B. Razzaghi, M. Komorowski, M. Barahona, et al., 2013. Nitrogen and Carbon Status Are Integrated at the Transcriptional Level by the Nitrogen Regulator NtrC In Vivo. *mBio* 4:e00881–13.

[35] Picotti, P., and R. Aebersold, 2012. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods* 9:555–566.

[36] Unwin, R. D., J. R. Griffiths, and A. D. Whetton, 2009. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). *Nature protocols* 4:870–877.

[37] Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf, 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* 6:187–202.

[38] Liepe, J., C. Barnes, E. Cule, K. Erguler, P. Kirk, T. Toni, and M. Stumpf, 2010. ABC-SysBio—approximate Bayesian computation in Python with GPU support. *Bioinformatics* 26:1797–1799.

[39] Liepe, J., P. Kirk, S. Filippi, T. Toni, C. P. Barnes, and M. P. H. Stumpf, 2014. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature protocols* 9:439–456.