



UvA-DARE (Digital Academic Repository)

Stochasticity in signal transduction pathways

Vidal Rodriguez, J.

Publication date
2009

[Link to publication](#)

Citation for published version (APA):

Vidal Rodriguez, J. (2009). *Stochasticity in signal transduction pathways*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 5

Processivity Effects of Gene Expression of Proteins in Bacterial Two-Component Signalling Systems

In the previous chapter we studied the structural factors that influence the response time in two-component signalling systems (TCS) typically found in bacteria. We found that the ratio between the number of histidine-kinases and response regulators (HK:RR) yields optimal response times when this ratio is lower than 1, and for the diffusion-limited kinetics of both HKs and RRs this optimal is 2:3. We also argued that in terms of cell's efficiency it is sensible that these systems work in proximity of that optimal ratio, as a way to obtain with just a few molecules short response times are obtained.

Besides the efficient use of the the molecules in TCS, we should expect that the mechanisms to regulate the synthesis of HKs and RRs are also optimised. Therefore, optimal functioning of the system should arise naturally from the system's architecture without requiring additional active regulatory mechanism. Prokaryotic organisms have much simpler mechanism for gene expression control than their complex eukaryotic counterparts. In prokaryotes many related genes are grouped into operons which form the expression unit (25% in *E. coli*). If the transcription and translation were faultless, the ratio of expressed proteins from the same operon would be constant at 1:1. Nevertheless, these processes are error prone and consequently due to the sequentiality of each process, the one that comes first has more chances to be completed than the subsequent. This phenomena was already observed in the *lac* operon and noticed that its

The contents of this chapter are based on the results published in

- Dobrzyński et al. Prokaryotic signaling has been optimized for quick but robust response. Submitted (2009)

three consecutive genes where expressed with decreasing quantity (Hartl and Jones, 2005). There are multiple reasons for which premature termination of both processes may occur, such as errors introduced during transcription or by some endogenous regulation mechanism.

As transcription proceeds from one end to the other, translation of polycistronic mRNA has a richer behaviour. In some cases, there may be a ribosome binding site before each cistron, so that *de novo* translation occurs in both cistrons in parallel. In other cases, there is only one ribosome binding site before the first cistron. Additionally, a ribosome translating an upstream cistron may re-initiate translation of the downstream cistron if the translational coupling is strong (Oppenheim and Yanofsky, 1980 ; Yoo and RajBhandary, 2008).

In this study we attempt to shed light in the effects that gene organisation have on the response time and efficiency of two-component signalling systems. We focus on systems with adjacent genes and a single ribosome binding site. The principal features to consider are on the premature stopping of gene expression which may occur during transcription and translation and resulting in a divergence in the final expression ratios HK:RR. Additionally, the implicit gene order in the operon, also influences the final expressed ratios of polypeptide chains which ultimately fold into active proteins.

In this chapter we first analyse the statistical properties of two-component signalling genes with regard to their order and length. We contrast the data from *E. coli* against that of over 600 microbial organisms. Thereafter we address the expression ratios in mRNA and polypeptide chains deriving a probabilistic model which accounts for gene ordering, processivity, gene length and lifetime of mRNA. Then we use the expression ratios to obtain the response time. In the last section we discuss our model and future extensions.

5.1 Statistical Properties of *Escherichia Coli* Gene Length

The information on gene organisation and gene length for two-component signal transduction systems (TCS) in *E. coli* was gathered from the The EcoCyc database (Keseler et al., 2005). This source lists a total of 21 signal transduction pathways. Table 5.1 shows a summary of the gene organisation, their length (in base pairs, bp) and histidine-kinase versus response regulator length ratio. Some of these systems, however, in spite of conforming to a typical two-component architecture — one HK and its cognate RR, as described in section 1.5.1 and chapter 4 —, share one of its components. For example, NarQ histidine-kinase pairs with two response regulators NarP and NarL. For the statistical analysis we consider that a TCS system is strictly composed of a single HK and a single RR. Therefore, some systems appear listed twice with one of the components being duplicated.

We distinguish three gene organisation classes. The first class contains genes found on the same operon and adjacent to each other. When adjacent we also take into account the order in which the operon is transcribed and, therefore,

distinguish two sub-classes: HK-RR and RR-HK, that distinguish the order in which an RNA polymerase transcribes the genes. (In The EcoCyc the order appears to be graphically depicted by a pointing box.) The third class includes those orphan genes, which, for the genes listed here, they all are found in different operons. Orphan genes are located in different operons and therefore the expression levels are regulated independently.

Another relevant feature is the length of each genes. The average length and standard deviation of each group of system is summarised in Table 5.2. The most salient feature is that histidine-kinases are between 0.6 and 5.8 times longer than response regulators (2.6 on average). Overall, the length ratios for adjacent genes is similar (1.96 and 2.0), whereas for the orphan class it differs significantly with an average of 3.1. This is true in general except for two cases: the nitrogen and RcsFB pathways. The former has an exceptionally long RR, while the latter has an exceptionally short HK. Histidine-kinases' longer genes may be attributed to a higher number of functional structures, which besides the sensing domain and output kinase domain, there are also a number, from 2 up to 20 transmembrane helices (Mascher et al., 2006). It is striking that the length of histidine-kinases with an adjacent response regulator is similar between the two orderings, whereas for orphan genes histidine-kinases are on average 400 bp longer. In the response regulator case, however, similar lengths occur between the RR-HK and the orphan groups. The ordering HK-RR has longer genes, although the length in the lower end range overlaps with the other classes.

5.1.1 Is *E. coli* Representative of Microbes?

One might wonder whether *E. coli* may be considered a representative case of microbes. We analysed the data obtained from the microbial MiST (Microbial Signal Transduction) database that contains 617 organisms, as of November 2008 (Ulrich and Zhulin, 2007). Nevertheless, we only took into consideration non-hybrid adjacent genes. In first instance, we find that the two-component signal transduction pathway systems retrieved from MiST represent a super set from that found in The EcoCyc, a specialised database in *E. coli*, and more concretely the strain K-12. Note in Table 5.1, that MiST lists 6 more pairs of two-component signalling systems than The EcoCyc. We shall mention that gene length data agrees in all but one case: CreB has a length of 1296 instead of 1425 as listed in EcoCyc. This difference, though, does not affect significantly the statistical analysis.

In MiST 59% of the systems belong the the RR-HK ordering. This suggest that this ordering might be favourable in the signalling system, as we shall show later. However, since the ordering is highly conserved among organisms, only four gene pairs out of 8348 were found to present different orders in different organisms, the result might be biased if one microbial family has evolved more into more organisms.

Second, we observe from the comparative statistics shown in Table 5.2, that the average gene length of *E. coli* follows the same pattern as that of all microbes in MiST: Histidine-kinases and response regulators are longer in the HK-RR

5. PROCESSIVITY EFFECTS IN GENE EXPRESSION

Table 5.1: List of the two-component signal transduction pathways found in *E. coli* databases *EcoCyc* and *MiST* (Refs. Keseler et al. (2005) ; Ulrich and Zhulin (2007)). The order of the gene is the transcription order, not their location on the genome. Some two-component signalling systems have dual input or output, but they are treated as independent pathways. Length is given in base pairs (bp).

order	source	System	HK	RR	length		ratio
					HK	RR	HK:RR
HK-RR	Ecocyc	Ato	AtoS	AtoC	1827	1386	1.32
		Bae	BaeS	BaeR	1404	723	1.94
		Kdp	KdpD	KdpE	2685	678	3.96
		NarX	NarX	NarL	1797	651	2.76
		Nitrogen ^a	NtrB	NtrC	1050	1410	0.74
	Zra	ZraS	ZraR	1398	1326	1.05	
	MiST		CitA	CitB	1659	681	2.44
			YojN	BcsB	2673	651	4.10
			YpdA	YpdB	1698	735	2.31
			YehU	YehT	1686	720	2.34
		DcuS	DcuR	1632	720	2.27	
RR-HK	Ecocyc	Bas	Bas-S	BarR	1092	669	1.63
		Cpx	CpxA	CpxR	1374	699	1.97
		Cre	CreC	CreB	1425	690	2.06
		Evg ^b	EvgS	EvgA	3594	615	5.84
		Osmoregularity	EnvZ	OmpR	1353	720	1.88
		PhoQ	PhoQ	PhoP	1461	672	2.17
		PhoR	PhoR	PhoB	1296	690	1.88
		Rst	RstB	RstA	1302	729	1.79
	Uhp	UhpB	UhpA	1503	591	2.54	
	MiST		YgiY	YgiX	1350	660	2.05
		CusS	CusR	1443	684	2.11	
		YedV	YedW	1359	672	2.02	
apart	EcoCyc	Arc	ArcB	ArcA	2337	717	3.26
		Bar	BarA	UvrY	2757	657	4.20
		Chemotaxis	CheA	CheY	1965	390	5.04
		Chemotaxis	CheA	CheB	1965	1050	1.87
		Cre	CreC	PhoB	1425	690	2.07
		NarQ	NarQ	NarP	1701	648	2.63
		NarQ	NarQ	NarL	1701	651	2.61
		NarX	NarX	NarP	1797	648	2.77
		Osmoregularity	BarA	OmpR	2757	720	3.83
		Rcs	RcsF	RcsB	405	651	0.62
		Rcs	RcsC	RcsB	2802	651	4.30
		Tor	TorS	TorR	2745	693	3.96

^a Not included in MiST

^b Hybrid histidine-kinase

5.1. STATISTICAL PROPERTIES OF *ESCHERICHIA COLI* GENE LENGTH

Table 5.2: Statistics for two-component signal transduction pathways in *E. coli* and a microbial database (MiST). Average length, in base pairs, and standard deviation of histidine-kinase sensor (HK), response regulator (RR) and the ratio (HK/RR).

	DB	gene	all	order		
				HK-RR	RR-HK	apart
<i>E. coli</i> K12	EcoCyc	HK	1811.78 ± 701	1693.5 ± 565	1600.0 ± 757	2029.75 ± 709
		RR	758.55 ± 243	1029.0 ± 380	675 ± 46	680.5 ± 144
		ratio	2.6	1.96	2.4 (2.0 ^a)	3.1
<i>E. coli</i> MiST	MiST	HK	1591.3 ± 404	1845.9 ± 461	1359.8 ± 110	
		RR	749.9 ± 204	827.1 ± 281	679.6 ± 36	
		ratio	2.2	2.4	2.0	
All	MiST	HK	1523.8 ± 415	1700.5 ± 530	1400 ± 243	
		RR	762.8 ± 238	821.5 ± 317	721.8 ± 148	
		ratio	2.1	2.3	2.0	

^a Without hybrid Evg.

ordering. And the ratios HK:RR for all cases, discarding the hybrid EvgS, seem to be very close to 2:1. From this simple statistics we already note that *E. coli* follows the same pattern in gene length than the average microbe.

We may further analyse the representativeness of *E. coli* by plotting the length of histidine-kinases versus their cognate response regulators' length, as shown in Fig. 5.1. As it is noticeable from the data in Table 5.1, but more clearly illustrated in Fig. 5.1, response regulators form three well differentiated clusters, whereas histidine-kinases group roughly into one cluster. Data from *E. coli* is shown for the three organisation classes. As expected, they fall into the main groups. Both orderings show one dominant cluster with RRs of approximately 700 bp. There are two additional minor clusters, albeit much less significant expect for the cluster at approximately 1400 bp for the HK-RR ordering. The differentiation in response regulators length seems to be determined by the RR family. All *E. coli* genes found in the lower cluster belong to the family OmpR and NarL, whereas the upper cluster consists of members of the NtrC family*. The first two have DNA binding domains (HTH), whereas the longer NtrC family has an AAA-FIS DNA binding domain (Galperin, 2006). All *E. coli* response regulator found in the upper RR cluster belong to the NtrC family and with ordering HK-RR. Likewise, in *E. coli* no pairs with RR-HK are found in this cluster. Other genes with equal ordering, but belonging to the other families are found in the lower major cluster.

Histidine-kinases follow a two-clustering pattern, although these are less differentiated. On the one hand, HK-RR ordering the main cluster is broad, and

*data extracted from KEGG.

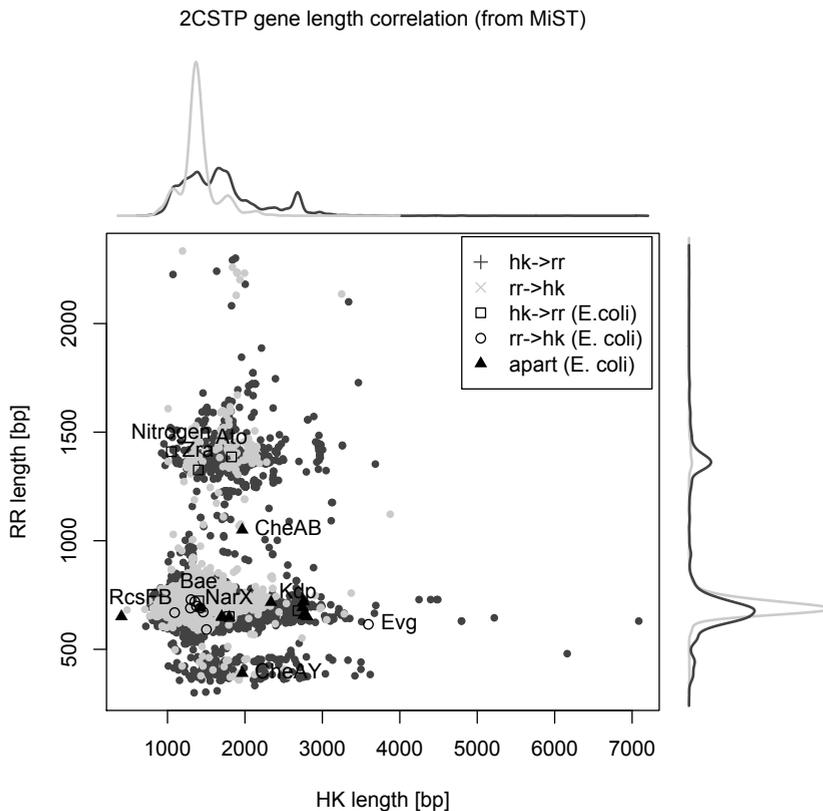


Figure 5.1: Distribution of gene pairs for microbial two-component signal transduction pathways from MiST.

has a secondary cluster around 2700 bp. On the other, RR-HK ordering contains primarily one cluster, although we may notice two smaller cluster on either side of the main one. These are difficult to discern, but for those pairs with long RR, we may notice the two sub-clusters. As we already argued before and summarised in the statistics Table 5.2, HKs have a wider distribution of lengths. Considering the location of the sensor domain, members of periplasmatic and cytoplasmatic kinases in *E. coli* where found in both orderings, so we rule out this being a factor to determine a particular ordering.

It is interesting to note that for histidine-kinases, the only extreme values, higher than 4000 bp, are only found for the HK-RR ordering. Whereas extreme values for RR, longer than 2000 bp exists in both orderings.

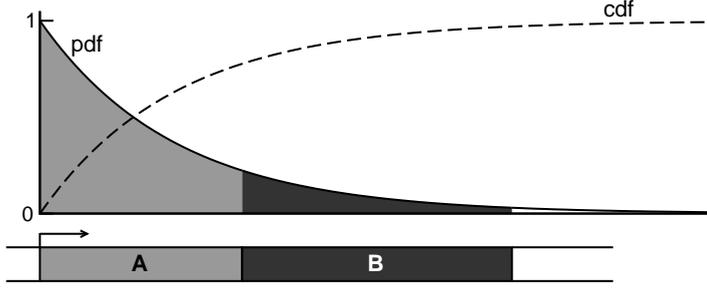


Figure 5.2: Generic representation of a dicistronic gene in which *A* is read before *B* and the probability for an RNA polymerase, or ribosome, to fall off. This also represent a polycistronic mRNA with one ribosome binding site before the first cistron.

5.2 Modelling Expression of HK and RR

To analyse the probability of expression of each gene, we make use of a probabilistic model that accounts for the stochastic effect of premature termination. In this model we assume we have a dicistronic system with gene *A* followed by and adjacent to gene *B* with length l_A and l_B nucleotides (abbreviated n , or bp) respectively (as illustrated in Fig. 5.2). An RNAP starts transcribing from the first position of *A*, and proceeds towards the end of *B*. We assume that premature termination, an RNAP unbinding the DNA, may occur at a random position during transcription with probability p on any position of either gene. If the RNAP does not unbind in *A* or *B*, then both genes are successfully transcribed. RNAP *processivity* is then p^{-1} .

We define the ratio of the number of complete mRNA genes as *A/B* (or *A:B*). This ratio is equivalent to the probability of transcribing each gene. Thus

$$\frac{A}{B} = \frac{Prob(A)}{Prob(B)}. \quad (5.1)$$

Gene *A* is transcribed whenever the RNAP falls after *A*. Thus, $Prob(A)$ includes those transcripts with only *A* successfully transcribed and those where both genes were completed. Gene *B* is only completed when RNAP falls after *B*.

The probability of an RNAP to successfully transcribe n bases follows a Bernoulli process, where for *A* we had at least l_A success trials, and for *B* at least l_{A+B} success trials. Therefore, we may write

$$Prob(A) = \sum_{n=l_A}^{\infty} (1-p)^n p = (1-p)^{l_A}, \quad (5.2)$$

$$Prob(B) = \sum_{n=l_{A+B}}^{\infty} (1-p)^n p = (1-p)^{l_{A+B}}. \quad (5.3)$$

5. PROCESSIVITY EFFECTS IN GENE EXPRESSION

For large values of n we approximate this discrete distribution by its continuous counterpart: the negative exponential. Then substituting Eq. 5.3 into Eq. 5.1 and taking their continuous equivalent we obtain

$$\frac{A}{B} = (1 - p)^{l_B} \approx \exp(-pl_B). \quad (5.4)$$

Thus the ratio of the number of completed mRNA A:B does not depend on the length of the first gene, but only on the length of the second. Eqn. 5.4 states that the longer the second gene is, the higher the ratio A:B may reach because the probability of failing during B is higher while A has been already transcribed. Note that the model describes the final steady state of a large number of transcription events. We shall model the dynamic behaviour in the next section.

The Bernoulli process also applies for translation, where a ribosome may also terminate prematurely. We shall see, however, that translation has some specific features that need further assumptions.

5.2.1 Transcription

In Fig. 5.3 we show the ratio A/B given by Eq. 5.4 for a range of processivities. In prokaryotes a typical value is in the order of 10^4 nucleotides, which is one order of magnitude longer than an average histidine-kinase (Hartl and Jones, 2005). The shaded regions in the plot indicate the ranges found for each ordering obtained from Table 5.2 for all microbial organisms. Both regions do not overlap, however due to the typical high processivity (solid line) the ratios do not differ much, 1.09 and 1.15 for HK-RR and RR-HK ordering respectively. Halving the typical processivity would increase slightly the difference, but the ratios remain between 1.1 and 1.5. As summarised in Table 5.3, for the data for all microbes in the MiST database, the probability of full transcription of the operon is high and differs little between the two orderings. Consequently the probability of successfully transcribing only the first gene are small (7-12%), and similar to that of failing to synthesize any functional mRNA (16-7%).

Table 5.3: *Probability of transcribing mRNA with both genes AB, only A or none (\emptyset) depending on the order of histidine-kinase and response regulator in a dicistronic mRNA. A is the first one, and B the second. Processivity error every is taken as $p = 1/10000$ nucleotides (Hartl and Jones, 2005).*

order	AB	A	\emptyset
HK-RR	0.78	0.07	0.16
RR-HK	0.80	0.12	0.07

We should keep in mind that these ratios are averages of a large ensemble of transcription processes. Therefore these are more representative of a population of cells than rather just one bacterium. When transcription events are

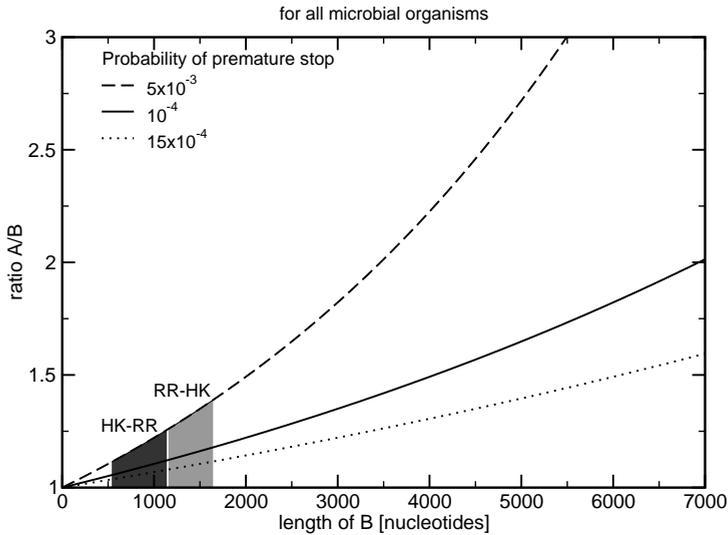


Figure 5.3: *Expression ratio of a dicistronic operon (transcription) for a range of premature stopping probabilities.*

rare, however, we should take into consideration additional parameters such as the discrete nature of the fully transcribed mRNA, the lifetime of the mRNA and the degradation processes that regulate mRNA's lifetime. Roughly 80% of the times one complete mRNA is transcribed, resulting in a ratio 1:1. Although this ratio may facilitate an optimal ratio for histidine-kinases and transcription factors in two-component signalling systems, as studied previously in section 4.3, the final protein ratio depends ultimately on the translation process, discussed in the next section. However, if a transcription ends prematurely during B, approx. 10% chances, and assuming that the previous transcript is already degraded, the resulting mRNA ratio is 1:0. Under this rare event, the number of proteins synthesized from the first cistron would increase. The absence of proteins from the second gene may be relevant only if transcription initiations are very infrequent as to allow for those proteins to be degraded. This may be the case if the expressed proteins have long lifetimes. The infrequent expression should keep the number of proteins in balance; otherwise, we could expect and accumulation of proteins. Half-lifetime of mRNA related to two-component signalling systems is found to be short, approx. 0.6 min, compared to an average mRNA in *E. coli* which averages 2–4 min, and may occasionally extend up to 50 min (Aiso and Ohki, 2003). This allows only for the completion of some translation processes of the first cistron before a prematurely ended mRNA is

quickly degraded (Hartl and Jones, 2005).

5.2.2 Translation

The second step in the synthesis of histidine-kinases and response regulators is translation of a nascent mRNA by a sequence of ribosomes. Unlike in transcription, translation initiation is not a rare event but a rather frequent one. In this section we analyse translation of dicistronic mRNA under two assumptions: that there is only one ribosome binding site before the first gene, and that there is a strong translational coupling, so the same ribosome continues to translate the second gene. These assumptions allow us to use the same probabilistic model shown Eqn. 5.4, in order to determine the ratio of the protein HK:RR. Note that the following results only concern translation, not the combined effects with transcription.

Typical rates of translational premature termination (or ribosome processivity) range from one in every 2600 to one every 4000 codons (Jørgensen and Kurland, 1990 ; Manley, 1978, and references therein). We shall note that this processivity is similar to that of transcription since a codon consists of three nucleotides. Thus, a premature stop in translation means that the error occurred in three nucleotides. Hence, the similarity between RNAP and ribosome processivity ($p = 3/10000 = 1/3333$). The mRNA template follows the diagram in Fig. 5.2, a simplified model which does not include the untranslated regions of a polycistronic mRNA (i.e. inter-cistron space). Because non-translated regions are short compared to that of proteins (few tens of nucleotides), we find that the length of a cistron in codons is, to a good degree of approximation, a third of its gene length.

From Fig. 5.4 we observe again that the ratios between the two orderings, with typical translation processivity, remain similar for the range of length of the second cistron (solid line). In principle, this plot is equivalent to that in Fig. 5.3, because of the above-mentioned equivalence between length and processivity. The expression ratio ranges are shown for *E. coli* data from Table 5.2, where despite observing an overlapping region for typical ribosome processivity there are no significant differences. Only for the longest cistrons (approximately 800 codons) and with a hypothetical lower processivity of $p = 10^{-3}$ codons per error the A:B ratio would reach at most 2:1.

In Table 5.4 we show that the probability that a ribosome translates both cistrons is between 71-83%, and only between a 7-12% of translation initiations synthesize the first cistron. These values, as calculated for the The EcoCyc *E. coli* data, are similar or slightly smaller than those found for the synthesis of mRNA due to a slightly lower ribosome processivity ($p = 1/2600$) (shown in Table 5.3).

For translation, nonetheless, the rare initiation events' effects characteristic of the transcription process are now no longer present due to a higher ribosome initiation frequency. Combining a low transcription initiation rate with a frequent translation initiation, which occurs on average once per second, one mRNA produces in its short lifetime a burst of proteins.

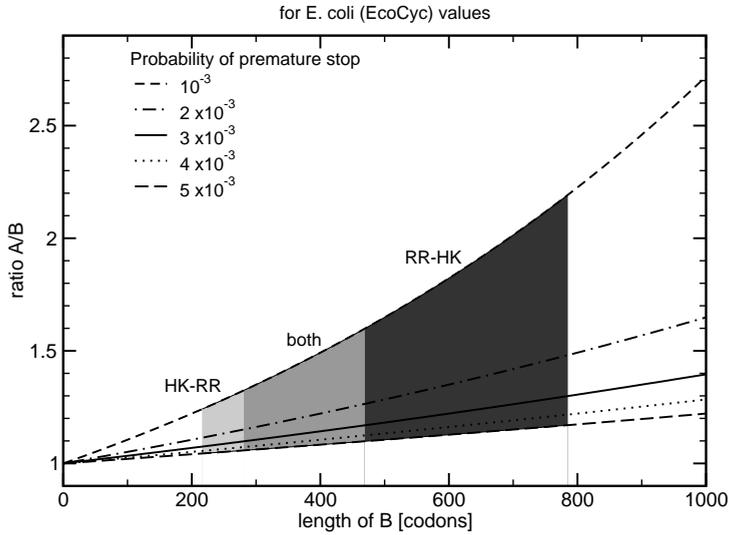


Figure 5.4: Expression ratio of dicistronic genes (translation) for a range of error probability per codon.

Table 5.4: Probability of translation of both cistrons (AB), only A or none (\emptyset), depending on the order of histidine-kinase and response regulator in a dicistronic mRNA. A is the first one, and B the second. The range is calculated for a probability processivity of one error every 1/2600–1/4000 codons.

order	AB	A	\emptyset
HK-RR	0.71-0.80	0.10-0.07	0.20-0.13
RR-HK	0.75-0.83	0.17-0.12	0.08-0.05

5.3 Expression Ratios and Optimality

The range of ratios obtained appear to be near the optimal ratio of HK:TF described earlier in the analysis of two-component signalling systems in section 4.3. We shall distinguish the implication of this ratio in view of the two possible orderings, HK-RR and RR-HK, described earlier in section 5.1. For the RR-HK ordering we find that the ratio is higher than for the reverse ordering. In view of the bias of the optimality curve in Figs. 4.4 and 4.6, under the constraints assumed for translation, the higher average A:B is between the 1:1 and 3:2 optimal range. In this situation shifting the ratio towards the right side of the plot is not detrimental, unless it goes too far. But this seems to be prevented by the strong correlation in the numbers of both proteins. For the reverse ordering, HK-RR, in view of the likely higher response times due to a shift in the ratio towards the left side of the curve, the system with this ordering is better off to stay closer to the semi-optimal ratio 1:1.

We should note that the ratios combining the transcription and translation processes would result in a slight increase of the ratio A:B. For such analysis, the dynamics and interacting between both processes should be taken into account. The static analysis done so far does not account for number of proteins, only their ratios. This property has not been investigated in this work. Nevertheless, in the next section we present a model for the dynamics of translation.

5.4 Dynamics of Translation

From the static model derived in section 5.2 we may not obtain information in the number of polypeptide chains being translated. Additionally, the static analysis is equivalent to a steady state analysis, that is, of a large ensemble and after the system has relaxed. We already pointed out that translation is limited by the lifetime of an mRNA, which for HK-RR is relatively unstable and short. The number of proteins successfully translated depends on three factors: the speed ribosomes walk along the mRNA, the initiation rate (separation between ribosomes) and the lifetime of the mRNA.

We develop a model to describe the rate of a polypeptide chain synthesis based on translation dynamics with one binding ribosome at the beginning of the first cistron. Our aim is to model only the average rate of synthesis, or output flux F_O , taking into account the processivity and the initiation rates of the process, or input flux F_I . As we have seen before, only a fraction of the particles that initiated the process finish successfully walking till the end of gene A or B (as shown in Fig. 5.2). On average, the number of fully translated cistrons is proportional to the number of ribosomes that start the translation of the cistron. Thus, the output flux of the first gene (A) corresponds to Eqn. 5.5 and for the second (B) is Eqn. 5.6.

$$F_O^A = F_I \exp(-p \cdot l_A)(t - \Delta_A), \tag{5.5}$$

$$F_O^{AB} = F_I \exp(-p \cdot l_A) \exp(-p \cdot l_B)(t - \Delta_{A+B}), \tag{5.6}$$

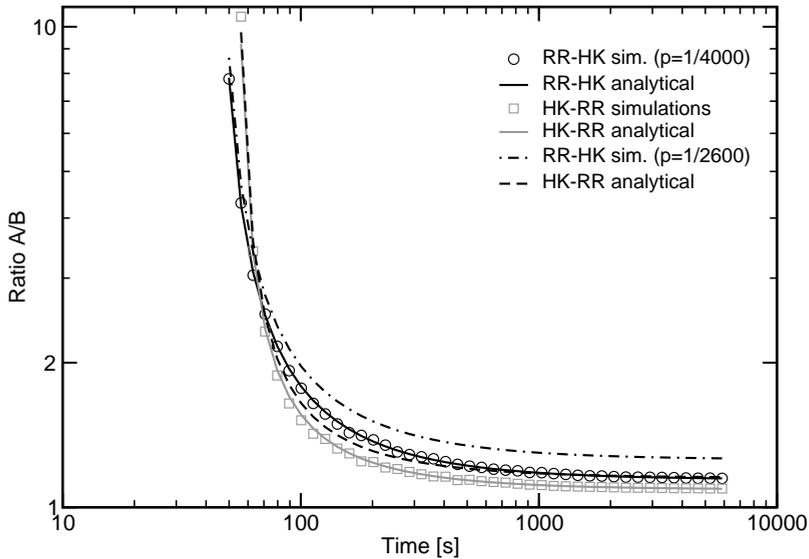


Figure 5.5: *Dynamic evolution of the ration A/B . The length of genes is take from The EcoCyc database. The processivity is $1/4000$ codons, initiation rate $1/17$ codons and Ribosome speed 17 codons/s.*

where Δ_A is the approximate time required for the first particle to traverse the whole length of gene A (or both A+B). This delay affects the timing of the process, especially for the short time events. Because of the high processivity of the process we have approximated this delay by the average time required by one particles to process a complete cistron which equals to $\Delta_A = l_A/v$, where v is the speed of the ribosome.

Translation initiation occurs at a rate of $0.8\text{--}1.1$ initiations $\cdot\text{s}^{-1}$ (Mitarai et al., 2008). Ribosomes process an average of $15\text{--}20$ codons $\cdot\text{s}^{-1}$ Spirin (1999), similar to RNA polymerase (50 nt)(Young and Bremer, 1976). Given these rates, one ribosome takes an average of $43\text{--}57$ s to translate both HK and RR cistrons.

In Fig. 5.5 the dynamic evolution of the A:B ratio described by model in Eqns. 5.5 and 5.6 shows that a steady state is reached within approx. 2 min minutes. As expected, the steady estate of the model, which is given by

$$\lim_{t \rightarrow \infty} A/B = \exp(l_B p),$$

corresponds to that summarised in Table 5.2.

Table 5.5: Number of finished polypeptides when mRNA start degrading from the ribosome binding site at t_{mRNA} . Length are taken from the MiST dataset in Table 5.2, frequency of initiation $1 s^{-1}$, ribosome speed 17 codons/s

t_{mRNA} [s]	HK-RR		RR-HK	
	A	B	A	B
10	8	7	9	7
20	16	15	18	15
40	33	30	36	31
60	49	45	55	47
80	66	60	73	63

mRNA Degradation

However, the average life of a two-component-related polycistronic mRNA has a typical half-lifetime of 36 s (Aiso and Ohki, 2003). There are several mechanism for degrading mRNA, but we shall only consider here mRNA degradation starting from the ribosome binding site which prevents new initiations. With this conditions the ribosome that already started translation may proceed undisturbed. The last polypeptide synthesized by each each is produced at time:

$$t_{last} = t_{mRNA} + l_X/v, \tag{5.7}$$

where t_{mRNA} is the start time of degradation, l_X is the length of cistron A (l_A) or the length of both cistrons l_{A+B} , and v the speed of a ribosome. Adding this degradation type to the model causes the ratio to converge faster to the steady state value (not shown). In fact, we find that the steady state is independent on the lifetime of an mRNA. The total number of completed polypeptide chains is found by substituting t in Eqn. 5.5 and 5.6 by t_{last} .

Table 5.5 shows the number of completed polypeptide chains for each gene, for both types of ordering and for a number of mRNA start degradation times. The short lifetime of mRNA ensures that the number of completed polypeptide chains remains low. The number of successfully folded proteins might be slightly slower, since protein folding is also prone to fail by a number of different processes.

In Fig. 5.6 we compare how the ordering affects the signalling response time and the number of polypeptide chains synthesized. We take two cases of gene lengths, and their reverse order. For the translation model with mRNA degradation we observe that the RR-HK ordering produces shorter response times because it is able to synthesize more particles. In these cases the bias in response time curve has little effect because the expression ratios are close to 1:1.

Incrementing the processivity of translation, for example to $p = 30000$, reduces the response time by approximately 10% (see Table 5.6). This reduction

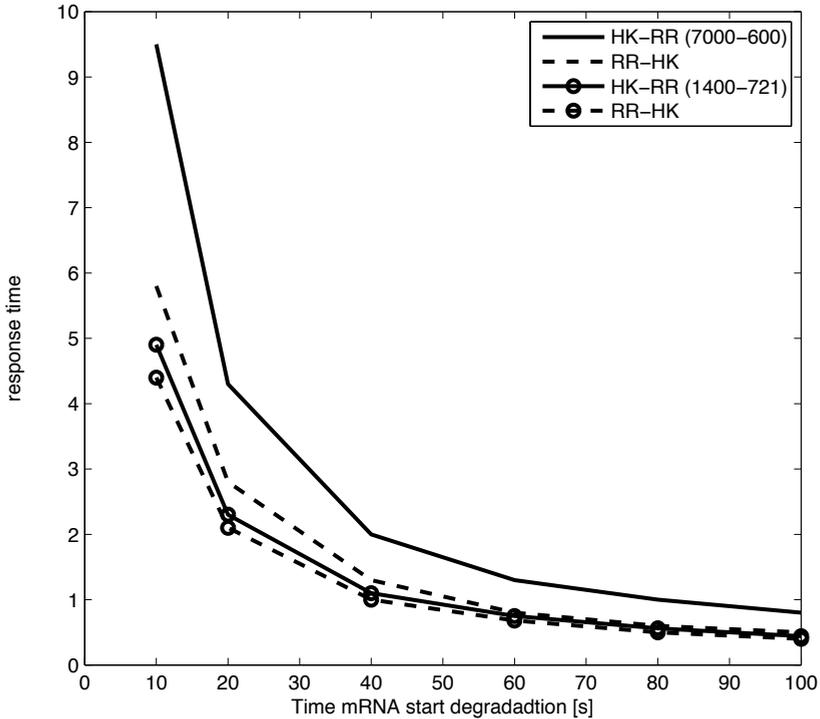


Figure 5.6: Comparison of the number of completed polypeptide for an average size dicistronic mRNA ($HK=1400$ bp, $RR=600$ bp) and an extreme one ($HK=7000$, $RR=600$). Given a starting time for mRNA degradation, $RR-HK$ synthesizes more proteins (Table 5.5) and consequently reduces the signalling response time (Eqn. 4.9). Processivity $p = 3000$.

is quite small compared to the time needed for translation (approx. 1 min). Additionally, the number of completed polypeptide chains is higher, approx. 13%. Lowering the processivity to $p = 300$ has a more significant impact in the response time, over 600%. This occurs because the number of completed polypeptide chains is much lower, and the signalling response is roughly inversely proportional to the number of molecules. A typical processivity of $p = 3000$ seems to offer a quick response time with a moderate use of resources.

Table 5.6: Number of completed polypeptide chains for a range of mRNA starting degradation times and processivities. The typical processivity found in bacteria is $p = 3000$, and half-lifetime of mRNA related to two-component signalling systems is 36 s.

$t_{\text{mRNA}}[\text{s}]$	p=300		p=3000		p=30000	
	#molec	$\tau_{\text{act}}[\text{s}]$	#molec	$\tau_{\text{act}}[\text{s}]$	#molec	τ_{act}
10	3.19	32.56	17.13	4.5	19.62	4.01
20	6.39	13.53	34.26	2.1	39.25	1.87
40	12.78	6.26	68.52	1.0	78.50	0.90
60	19.17	4.07	102.78	0.7	117.75	0.60
80	25.56	3.02	137.31	0.5	157.0	0.45
100	31.95	1.88	171.31	0.4	196.3	0.36

5.5 Discussion

We have studied the statistical properties of genes involved in two-component signalling systems with emphasis on *E. coli*, which are widespread in microbes. Often histidine-kinase and response regulator genes are located in the same operon and adjacent. The operon organisation is a mechanism that guarantees that related genes are expressed together, and the number of proteins are correlated (Hartl and Jones, 2005). Additionally, we have assumed that there is a strong translational coupling so that the expression of the second cistron depends on the translation of the first. When the processes involved in gene expression are prone to failure the gene order on the operon affects the final expression protein levels. In this study we have considered the processivity effects in transcription and translation. Unlike the single initiation site of an operon, a polycistronic mRNA may have one initiation site, ribosome binding site (RBS), before the first gene, or one RBS before each cistron. The former case has equivalent dynamics to that of transcription.

The models and simulations show that with typical processivities for both transcription and translation the ratio between the first and second gene are always larger than one and similar (1.1:1–1.3:1), and that this ratio only depends on the length of the second gene. This range of ratios approximates the ratio proposed in chapter 4 where the optimal ratio for signalling (and scattered sensors) was RR:HK = 1.5:1, favouring more response regulators (or transcription factors). Thus, the RR–HK ordering seems to agree with that result, as it lies closer to the optimal (minimum) point. In fact, the correlation in the number of proteins due to the effects of a single RBS suggest that the ratio will stay close to the central 1:1, avoiding shifting the ratio towards higher ratios and therefore working on an inefficient ratio.

However, we note that the HK–RR ordering, which in principle lies on the left half of the optimality curve (see Figs. 4.7 and 4.6) seems to compensate

this disadvantage by having a lower ratio (see Fig. 5.4), avoiding to left-most region that has a higher penalty in signalling time. As with the reverse ordering, by strongly correlating the number of proteins the system works in the central region of the optimality curve, which as we saw in the chapter 4 its central region is flat and small deviation from the optimal point have little effects in incrementing the response time.

We should be aware that the number of expressed polypeptides chain is slightly larger than the number of successfully folded proteins, as this process is also prone to error. However, we have refrained to include this effect in this study. Additionally, when calculating the response times for these ratios we are implicitly assuming that all histidine-kinases are activated by the signal and that all response regulators are available to transport the signal. Under that assumption we conclude that the RR-HK ordering is the optimal to reduce the response time. However, as we argued in chapter 4, if, for instance, a signal only activates typically a fraction of the sensors available, because it is present in small concentrations or not strong enough to activate all sensors. Under this circumstance the system might obtain faster responses by having an HK-RR ordering, which would create more HK to guarantee a weak signal activates (and induces autophosphorylation) enough HK. In this way, the mapping between the expression rates reported and this chapter and those reported in chapter 4 would need to be adjusted to include this effect. To find out about these properties we need very detailed kinetic *in vivo* data for each particular system. This data is, to our knowledge, not generally available yet. But further studies could make use of single-molecule single-cell techniques to elucidate and validate the phenomena theorised here.

In the case of one ribosome binding site per cistron we may expect qualitatively different results in the expression ratios. Then also the translational coupling needs to be taken into consideration. For instance, is there is a strong translational coupling and *de novo* initiation of the second cistron, the ratio A:B may be lower than 1. But these depend also on the initiation rate, which may differ since the inter-cistron RBS may be often inaccessible due to secondary structure of mRNA, or because it is occupied by a ribosome in the re-initiation phase. Thus the ratio range may be found to be broader.

We find that an unusual short lifetime of two-component-related mRNA and a typical processivity of one error per 3000 codons, seems to offer a reasonable low number of polypeptide chains yet still short responses time may be achieved (Table 5.6). Since the degradation used in our model only prevents more initiation events without affecting the mRNA strand, there are no alterations in the final protein ratios. However, it is easy to see that if premature degradation affects the whole mRNA quickly, without letting the ribosomes to finish, then the final ratio depends on the time translation has been running. Additionally, differences in response time (τ_{act}) attributed to the ordering seem not to be significant for typical gene length, when compared to the time required for gene expression to synthesize the first polypeptide chain (1 min approx). The flatness of the optimality curve is the indication of the robustness of the system for relatively small biases of the optimal ratio.

Orphan genes seem to have evolved to produce independently regulated proteins, which may also have larger variation in their ratios. For example, in the chemotaxis pathway, the response regulators have a considerable 30 to 50 times difference in concentration: CheY 6300–8200 molecules, and CheB 240–270 (Li and Hazelbauer, 2004). For CheA, the receptor, the numbers of molecules are in the range 6700–7800 (containing both long and short forms as described in Kofoid and Parkinson (1991))

5.6 Conclusions

In the statistical study of gene length and clustering the intrinsic structure of gene families arises clearly, but more importantly we find indications that the ordering of genes favour higher A:B ratios for RR-HK, as suggested in the previous chapter. However, we also argue that this results need to be further studied in a more case-by-case basis and under *in vivo* conditions since there are other factors that might affect the number of participating particles in the two-component signalling system.

We have modelled the origins of protein levels bias focusing only on the first two processes of the gene expression mechanism. Translation might be accounted for as the principal mechanism for the bias. However, failures in the infrequent transcription process have a potential larger impact in the protein levels, since only the first cistrone may be completed.

To the best of our knowledge, this is the first study that analyses the implications of gene order on the efficiency of two-component signalling pathways in bacteria. The initial modelling has helped to lay future directions in how to proceed and validate our initial hypothesis. Due to the computational modelling nature of this research, there are many details of the real system omitted in this work than may play a crucial role in favouring a particular gene order and that require a more experimental, modelling research cycle to fully understand the implications of gene order.

Acknowledgements

We are grateful to Luke Ulrich for providing the gene's statistical data from the microbial signal transduction database (MiST, <http://genomics.ornl.gov/mist/>, accessed November 2008) .