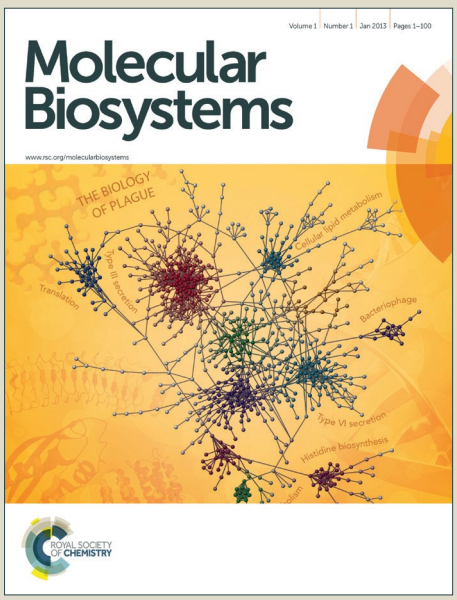
Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Phosphorelay in non-orthodox two component systems *in vivo* functions through a bi-molecular mechanism: the case of ArcB

Goran Jovanovic[1,*], Xia Sheng[1,2,*], Angelique Ale[1,2], Elisenda Feliu[3], Heather A. Harrington[1,2], Paul Kirk[1,2], Carsten Wiuf[3], Martin Buck[1], Michael P.H. Stumpf[1,2,4,5]

[1]Department of Life Sciences, Imperial College London, UK
[2]Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, UK
[3]Department of Mathematics, University of Copenhagen, Denmark
[4]Institute of Chemical Biology, Imperial College London, UK
[5]Department of Mathematics, University of Melbourne, Australia

[*]*These authors contributed equally.*

Subject Area:
Biology/biochemistry/signalling
computational biology

Authors for correspondence:
Michael P.H. Stumpf
m.stumpf@imperial.ac.uk
Martin Buck
m.buck@imperial.ac.uk

Molecular BioSystems Accepted Manuscript

Summary

Two-component systems play a central part in bacterial signal transduction. Phosphorelay mechanisms have been linked to more robust and ultra-sensitive signalling dynamices. The molecular machinery that faciliates such signalling is, however, only understood in outline. In particular the functional relevance of the dimerization of a non-orthodox or hybrid histidine kinase along which the phosphorelay takes place has been subject of debate. We use a combination of molecular and genetic approaches, coupled to mathematical and statistical modelling, to demonstrate that the different possible intra- and inter-molecular mechanisms of phosphotransfer are formally non-identifiable in *Escherichia coli* expressing the ArcB non-orthodox histidine kinase used in anoxic redox control. In order to resolve this issue we use further analysis of the mathematical model in order to identify discriminatory experiments, which are then performed to address *cis*- and *trans*-phosphorelay mechanisms. The results suggest that exclusive *cis*- and *trans*- mechanisms will not be operating, instead the functional phosphorelay is likely to build around a sequence of allosteric interactions among the domain pairs in the histidine kinase. This is the first detailed mechanistic analysis of the molecular processes involved in non-orthodox two-component signalling and our results suggest strongly that dimerization facilitates more discriminatory proof-reading of external signals, via these allosteric reactions, prior to them being further processed.

Keywords: two-component system; ArcB; phosphorelay; algebraic statistics; model selection; allostery

Introduction

Cells experience their environment through a set of complex molecular machines. In bacteria sensing of external signals and the environmental state relies primarily on two-component systems (TCSs)[1-4]. These typically consist of a membrane bound protein – referred to as the histidine kinase (HK)[5] – that includes a sensor and one or more protein phosphorylation domains, and a second protein – the response regulator (RR)[6] – that generally acts as a transcriptional regulator upon its activation through phosphorylation. These TCSs pick up the signal — for example through binding of a small molecule to an extracellular receptor

Molecular BioSystems Accepted Manuscript

domain of the kinase — and marshall an appropriate adaptive response by activating the down-stream targets of the RR.

The complements of TCSs found in different organisms vary considerably in both size and composition[4,7,8]. *Escherichia coli*, for example, has 30 TCSs, whereas *Mycobacterium tuberculosis* relies on some 70 TCSs to control its physiological state, with only a small overlap among their respective sets [9]. The TCSs present in a species reflect the bacterial life-styles and prevailing environmental conditions in which bacteria live[10,11]. In *E. coli* for example, they regulate the switch between aerobic and anaerobic metabolic states, response to environmental stresses (such as pH, nitrogen starvation and osmotic stress) damage to the cell's membrane, and control the chemotactic behaviour and modes of cellular motility.

The extant sets of TCSs found in bacterial species are powerful testament to the flexibility but also the frugality of the evolutionary process: the constituent parts of different TCSs show high levels of sequence similarity (within and across species) and the domain architectures of histidine kinases and response regulators tend to follow tightly prescribed patterns (see Fig. 1A). Response regulators always contain an aspartate (D) domain, which upon phosphorylation typically leads to activation of the DNA binding functionality[11]. HKs usually come in two different forms[5,12]. The predominant form, hereafter referred to as the *orthodox* HK contains a single HK domain (H) which also shows phosphatase activity and which regulates the RR activity by controlling the phosphorylation state of the RR's D domain (Fig. 1Ai). A second, *non-orthodox*/*hybrid* HK architecture is characterized by a sequence of three phosphorylation domains (H1-D1-H2) along which the signal is propagated in a so-called phosphorelay before the cognate RR is phosphorylated (see Fig. 1Aii,iii, Bi).

The functional implications of the different HK architectures have been attracting growing interest. In *E. coli* 26 HKs are of the orthodox type and only 4 non-orthodox HKs are present; this ratio is also found in closely related enterobacterial species[9]. Quite generally, orthodox HKs outnumber their non-orthodox alternatives by nearly a factor of 10 in most bacterial species.

Functional differences between orthodox and non-orthodox HKs have been put forward as potential causes for this pattern that appears to extend across all bacterial genomes sequenced to date. The phosphorelay architecture has been shown to imply different dynamical

Molecular BioSystems Accepted Manuscript

behaviour compared to what can be achieved for the orthodox HK architecture (Fig. 1A). In particular signalling via non-orthodox HKs using a phosphorelay is generally more robust to noisy inputs and exhibits ultra-sensitivity in its response characteristics[13-16]. These differences in the behaviour of orthodox and non-orthodox signalling may explain their relative abundances in bacterial genomes; more generally, this will also have implications for applications of microbes in biotechnology and synthetic biology[14] where certain response characteristics to given environments may be either desirable or design objectives.

Despite their importance many of the molecular details of TCS signalling dynamics – especially as occurs *in vivo* and in the physiological context of bacterial cells – are still only known in outline. While it is known, for example, that in order to function two HK monomers have to form a dimer, the role that the dimerization plays is not known. Signalling along a monomer does not appear to work *in vitro*[1,12,17,18], but *in vivo* direct observations of the mechanism have unsurprisingly proved challenging. For phosphorelays, in particular, it is not known if the phosphoryl group moves along the domains of a single monomer (in a *cis*-like manner) or jumps between the two monomers (in a *trans*- mechanism) making up the functional dimer (Fig. 1). Because only dimers appear to be capable of propagating the signal the nature of the mechanism (i.e. *cis* vs. *trans*) is not just of fundamental biophysical interest but will also have profound implications for our ability to manipulate non-orthodox HKs and understand their function and molecular evolution. Potential reasons for dimerization of individual HKs into functional homo-dimer units may include increased specificity due to dimerization, the prevention of extensive cross-talk or shielding of molecular/electrostatic interactions that may interfere with the faithfulness of the signal transduction process. As with many other membrane-bound proteins (in particular high molecular weight ones like the typical non-orthodox HKs) a lack of reliable molecular structures[19] limits our mechanistic understanding.

The Arc (Anoxic redox control) is a complex two component system that in general mediates regulation of operons implicated in respiratory metabolism and enables facultative aerobic bacteria to sense and respond to different respiratory conditions[20-22]. The Arc system comprises membrane-bound tripartite non-orthodox sensor kinase ArcB and the cognate response regulator ArcA (Fig. 1B). The transcription of *arcB* is constitutive under all growth conditions. Under anoxic conditions *arcA* transcription is significantly increased in an Fnr

Molecular BioSystems Accepted Manuscript

(global regulator that together with Arc system regulates transcription in response to O_2)-dependent manner, ArcB autophosphorylates and through His^{292}-Asp^{576}-His^{717}–Asp^{54} phosphorelay phosphorylates ArcA (Fig. 1B)[23]. Phosphorylated ArcA then represses many operons involved in aerobic respiration and up-regulates genes involved in anaerobic respiration (e.g. *cydAB*) and fermentation (e.g., *pfl*). It has been shown that the ArcBA system activation is of particular importance under microaerobic growth conditions. ArcB, as reported for many other sensor kinases, act as homodimer. Under aerobic growth conditions ArcB kinase activity is inhibited by quinone electron carriers that promote the oxidation of redox-active cysteine residues (C180, C241) implicated in formation of an inactive ArcB dimer through intermolecular disulfide bond formation[17] (Fig. 1B). ArcA is then dephosphorylated by ArcB-dependent reverse phosphorelay Asp^{54}-His^{717}-Asp^{576}-P_i (Fig. 1B). It is likely that the ArcB-Asp^{576}-P dephosphorylates by the intrinsic lability of the phospho-aspartyl bond[17]. The phosphatase activity of ArcB abolishes the non-specific acetyl-phosphate -dependent phosphorylation of ArcA.

Here we use a combination of mathematical modelling, targeted molecular investigations, and statistical analysis to elucidate the ArcB phosphorelay mechanism under physiologycal *in vivo* conditions. The mathematical models rigorously (by incorporation of the known biomolecular and biophysical constraints on the dynamics) force us to explicitly state our assumptions and follow them – using simulation whenever exact solutions cannot be obtained – until the point where testable predictions are possible. Here we have used this approach to develop an experimental strategy that allows us to discern between different modes of phosphorelay action — in particular *cis* vs. *trans* phosphorelay mechanisms — and, coupled with statistical analysis determine a model that best captures experimental observations [24]. As we will show below neither of the two initial models for *cis* or *trans* relay is capable of producing the predicted behaviour; in light of this we were able to derive a different model which requires an "allosteric" interaction, where phosphorylation at e.g. H1 is required to open up, or make receptive the D1 domains for phosphorylation, and where this is only possible if both monomers have functional H1 and D1 domains; similarly for D1 needs to be phosphorylated before the H2 domains become available (and again both D1 and both H2 domains need to be functional); in the *cis* and *trans* mechanisms, the domains are assumed to be always open but will receive the phosphoryl group from the adjacent sites. This protection

Molecular BioSystems Accepted Manuscript

of the D1 and H2 sites, unless H1 or D1 are phosphorylated will, for example, protect the phosphorelay from interference by acetyl-phosphate, which can lead to spurious activation. Based on this integrative analysis we propose that the phosphorelay mechanisms exhibits the hall-marks of allostery[25,26], which requires cooperativity between the individual molecules making up the HK dimer.

Results

Models of *cis* and *trans* signalling are mathematically equivalent. While there have been a number of models of TCSs and phosphorelays they have only paid scant attention to the precise mechanisms by which phosphoryl group is moved along the non-orthodox HK homo-dimer. Both *cis* and *trans* mechanisms are straightforwardly phrased in terms of biomolecular reactions. These are then easily translated into deterministic or stochastic mathematical models. Additionally, we consider a model which requires an interaction among the two monomers in order for the phosphorelay to work: in particular functional phosphorylation sites are required at all steps and on both monomers and the two monomers act cooperatively in the phosphorelay; this model will be referred to as the bi-molecular model. The reactions of all three models are provided in Table 1 and in the electronic supplementary material (ESM). Other models, e.g. combiations of *cis* and *trans* are also possible.

As is apparent from the reaction schemes in Table 1, but also more explicitly from the descriptions in the ESM, we cannot distinguish between the three different models. Their main functional difference is in the precise way in which the phosphoryl group group moves along (*cis*) or between (*trans*) the two monomers constituting functional HK homodimers. This can be expressed in terms of a simple effective model which subsumes the three different mechanisms into a single effective model, which can be written as

$$HK_0 \rightarrow HK_1 \qquad HK_1 \rightarrow HK_2 \qquad HK_2 \rightleftharpoons HK_3 \qquad HK_3 + RR \rightleftharpoons HK + RR_p \qquad HK_2 \longrightarrow HK_0 \quad (1)$$

Mathematically this means that the mode of operation at the molecular level of the phosphorelay is non-identifiable from data collected from the wild-type (WT) system. Such non-identifiability is irrespective of whether such data are time-resolved (as opposed to steady state); moreover, more flexible models that allow for multiply phosphorylated HKs show the same behaviour and can be expressed in terms of a simplified model that ignores the

Molecular BioSystems Accepted Manuscript

dimeric structure (see ESM) but which produces behaviour that is indistinguishable from the dynamics considered here[27]. Thus dimerisation does not in itself confer a dynamical advantage for HKs (such as increased speed of transmission, system robustnes or susceptibility to noise in the input, for which all our models here are equivalent); instead it might be that the precise structural organisation of the dimer has other advantages such as rotation of the subunits leading to the "on-off" switch as proposed by Kwon *et al*.[28], or indeed enables the phosphorelay mechanism in the first place. This is not covered in the modelling frameworks considered for TCSs and the mathematical model of the dynamics do not allow us to distinguish between different molecular mechanisms.

***In vivo* analysis of phosphorelay mechanisms.** The detailed molecular mechanism by which phosphoryl group groups are transferred along a phosphorelay are, however, of practical importance as different mechanisms, *i.e., cis* vs. *trans* vs. allosteric, and combinations thereof, may exhibit different tendencies *in vivo* to cross-talk or otherwise interact with neighbouring (orthodox or non-orthodox) HKs[29,30]. To elucidate the mechanism *in vivo* we therefore used the structure of the mathematical model in order to propose experimental set-ups that allow us to distinguish the molecular processes facilitating phosphorelay.

In order to induce differences into these models that render the mechanisms identifiable[31] we generated mutants in which phosphorylation sites have been selectively ablated (see *Materials and Methods*). In a system with such mutant homodimers it is still not possible to separate the *cis*-phosphorelay, *trans*-phosphorelay and bi-molecular phosphorelay models, as any output will be fully blocked because of ablation of phosphorylation sites. We therefore consider heterodimer arrangements of all mutant (i.e. of the H1, D1 and H2 domains) and WT combinations. We verified that binding between the different variant monomers is possible through bacterial two-hybrid (BATCH) assays. All mutant monomers were expressed from two different plasmids (see *Materials and Methods*) with different copy numbers in *E. coli* for quantitative assays of the downstream cognate RR ArcA activity as detailed below and in the legend of Fig. 2. Suppose we consider a system where one plasmid expresses WT ArcB and the other the H1 mutant, and that the frequencies of the two plasmids are f_1 and f_2, respectively. Then a fraction of f_1^2 ArcB dimers will be WT homodimers; $2f_1f_2$ will be WT–H1 heterodimers; and f_2^2 will be H1 homodimers. We can use these fractions to predict patterns of ArcA activity (measured as outlined in *Materials and Methods*) under each of the

Molecular BioSystems Accepted Manuscript

different mechanisms. In particular, for different mutant combinations we are able to use differences between observed and predicted activity levels to distinguish between models. For example, along an H1–D1 heterodimer the *cis*-phosphorelay mechanism should be completely blocked, while the *trans*-mechanism should be functioning at reduced level (*a priori* under a naive model for the *trans*-phosphorelay we might expect half the activity compared to a WT homo-dimer; under the allosteric process activity should be heavily reduced). By assessing the activity levels for the combinations of different mutants we will thus be able to rule out or invalidate these simple models.

Activities of the constitutively active ArcB* phosphorelay mutants. To assess the activities of the non-orthodox sensors in phosphorelay we constructed mutants of constitutively active ArcB* [32] (see Fig. 2A, B and Table 2) and used a Δ*arcB* strain carrying the chromosomal ϕ(*cydA-lacZ*) transcriptional fusion specifically regulated only by ArcA with the *cydA* Fnr regulatory sequences deleted[22] (see Table 2). The ArcB* has C180 and C241 residues substituted with alanine so behaves as an "on" state active kinase but since its activity does not depend on upstream signalling in principle should have abolished dephosphorylation/deactivation of ArcA. To avoid ArcA expression level as the limiting factor, the experiments were carried out in microaerobiosis where expression of ArcA is induced by Fnr. In microaerobiosis the SixA protein, which dephosphorylates ArcB at H2, is not functioning since this protein is only active in anaerobiosis.

Initially, we confirmed that the constitutively active form of ArcB, denoted ArcB*, activates transcription of ϕ(*cydA-lacZ*) fusion independently of growth conditions, bypassing all upstream signalling needed for the activation of the kinase activity (Fig. 2A). We then determined that WT and mutant forms of ArcB* proteins were similarly expressed when placed on either pCA24N or pAPT110, respectively (Fig. 2B); in addition their expression is in the range of chromosomally expressed ArcB (ESM Fig. S1). Since dimerisation of the sensor is a prerequisite for its kinase activity, ArcB* and its phosphorelay mutants were tested for their ability to interact in heterodimers using the BACTH system (see *Materials and Methods* and Fig. 2C); the results confirmed that mutations introduced into ArcB* do not interfere with formation of dimers. As a control experiment, ArcB* or its corresponding phosphorelay mutants were expressed from either pCA24N or pAPT110 vectors in a Δ*arcB* strain, and efficiency of ArcA phosphorylation was assessed by measuring β-Gal activity of

Molecular BioSystems Accepted Manuscript

ϕ(*cydA-lacZ*) fusion following ArcA-dependent activation of transcription (Fig. 2D). As expected, the results showed that any disruption of phosphorelay present in both monomers of the sensor kinase homodimer abolishes activation of ArcA. Notably, some residual activation above the control basal level (vector alone) was detected when ArcB*^{D576A} or ArcB*^{H717A} mutants were expressed from a low copy number vector (Fig. 2D). To support these data, we did the same experiment using a wild type ArcB/phosphorelay mutants that is activated by upstream signalling in a canonical fashion (ESM Fig. S2A). These results were in agreement with data we obtained with ArcB* and its mutants.

Combinations of either ArcB* or different phosphorelay mutants of ArcB* were co-expressed in a Δ*arcB* ϕ(*cydA-lacZ*) strain and phosphorylation efficiency of ArcA is determined as above (Fig. 2E). The results showed that:

- Co-expression of ArcB* with any of the sensor kinase phosphorelay mutants significantly diminished activation of ArcA, with the ArcB*^{H292A} mutation having the most detrimental effect;
- Co-expression of different ArcB* mutants that disrupt the sensor kinase phosphorelay abolished activation of ArcA, with the most pronounced effect of the ArcB*^{H292A} mutation evident in its combinations; one possible explanation for the reduced activity in heterodimers containing the H1 mutant is that this mutant is likely to maintain phosphatase activity (which is largely ablated in D1 and H2 mutants).
- A below basal level activation of ϕ(*cydA-lacZ*) is obtained when at least one monomer of ArcB* carried intact D576 and H717 and inactivated H292 residues, suggesting an unexpected reverse phosphorelay activity of the corresponding ArcB* mutant monomer and so a phosphatase activity of ArcB* mutant sensor kinase and hence dephosphorylation of ArcA (that might be activated by cross-talk with *e.g.*, acetyl-P, see Fig. 1B for a diagram of ArcB phosphorelay and de-phosphorelay).

In order to address whether non-specifically activated ArcA might contribute to the basal level expression (only vectors) of ϕ(*cydA−lacZ*) fusion in Δ*arcB* strain, we compared this result to the basal level activity obtained in a Δ*arcB*Δ*arcA* strain lacking ArcA response regulator (Fig. 2F). Apparently, the presence of ArcA contributes to the basal level expression seen in Δ*arcB* strain co-expressing both vectors alone.

Molecular BioSystems Accepted Manuscript

Significantly, the combinatorial experiment using wild type ArcB and most of its phosphorelay mutants (ESM Fig. S2B) showed that as for the ArcB* and its mutants (Fig. 2E), co-expression of ArcB with any of its phosphorelay mutants significantly diminishes activation of ArcA and co-expression of different ArcB phosphorelay mutants mainly abolishes activation of ArcA. However, no below basal level activation of ϕ(*cydA-lacZ*) was obtained. This is expected since the wild type ArcB phosphatase acting upon ArcA is sensitive to growth conditions and upstream signalling and should be pronounced in aerobiosis. Interestingly, even though the ArcB* variant is in a generaly active state which does not depend on the upstream signalling, when specific phosphorelay domains are inactivated it might regain the ability to dephosphorylate ArcA.

An allosteric mechanism for the phosphorelay. There is evidence that autophosphorylation in ArcB is transferred intramolecularly[33], but beyond this the mechanisms of phosphotransfer are unknown. Our results clearly indicate that neither the *cis* (intra-molecular) nor the *trans* (inter-molecular) model are capable of reproducing the observed *in vivo* activities. Instead functional versions of both monomers appear to be required in order for the phosphorelay to work appropriately. One mechanism that could explain this type of behaviour involves a bi-molecular mechanisms exhibiting the hallmarks of allostery, where binding of phosphoryl group to a H1 domain is required to open up the phosphorylation sites at the D1 domains and where functional H1 and D1 domains are required at both monomers. In this case it would be the binding of the phosphoryl group to one of the two H1 domains (to explain our data we do not require phosphorylation at both H1 domains) which would open up the two D1 domains to be receptive for phosphorylation. For phosphotransfer from D1 to H2 we would again require all domains/phosphorylation sites to be functional and for one D1 domain to be phosphorylated before the H2 domain can accept the activating phosphoryl group. The domains of the two monomers thus act cooperatively to faciliate the phosphorelay.

In the light of our models, allostery would explain, for example, why ablating combinations of domains that would lead to clear signalling along or between monomers in a *cis* or *trans* manner fail to capture the experimental data. We therefore propose a new model, which incorporates the essential hallmarks of allosteric behaviour. This model, while in behaviour indistinguishable in the WT from the other two models (see Table 1 and ESM), behaves very differently in its predictions of the behavioural outcomes of mutant hetero-

Molecular BioSystems Accepted Manuscript

dimer ArcB configurations, where the requirement of phosphorylation at H1 and D1 for the phosphorylation sites at D2 and H2 to "open up" introduces different qualitative and quantitative behaviour.

In part E of Fig. 2 we observe that the patterns of inferred ArcA-P activity are in fact much reduced in response to all ArcB heterodimers tested, in a manner which is hard to reconcile with the simple *cis*- and *trans* mechanisms; this observation did, indeed, motivate the development of the bi-molecular allosteric model in the first place. All three models are mathematically indistinguishable if only WT HK data are available; in the ESM we discuss this further (see also ESM Fig. S3). From the same analysis it emerges that the mutant data described above are sufficiently quantitative and discriminatory to facilitate a model based analysis, permitting statistical model selection approaches to be used. In particular this set-up allows us to assess which mechanism can explain the data best.

We calibrate the mathematical models (see ESM) against the data using a maximum-likelihood procedure. We used both an unconstrained model where parameters for each step in the phosphorelay are allowed to vary freely (see Fig. 3A), and a model where the respective phosphorylation and dephosphorylation reaction rate constants along a cascade are restricted to be identical (see ESM Fig. S4). For both approaches we find that the allosteric model always results in a markedly improved fit to experimental data compared to the other two models.

In order to choose among the different models we need to balance model complexity with a model's explanatory power (or its ability to fit the data). While the *trans* and *cis* models have the same number of parameters (in each case 17), the allosteric model has more parameters (in total 27); however, biophysical considerations allow us to reduce the number of parameters to seven distinct parameters for all three models (see *Materials and Methods*). We use the BIC criterion (and the corresponding Schwarz weight) to distinguish formally between the three competing mechanistic models (see *Materials and Methods*). The BIC values of the *trans*-, *cis*- and allosteric phosphorelay models are -20.25, -14.04 and -31.14. The Schwarz weight of allosteric model is 99.55%, much higher than for the other two models (as shown in Fig. 3B). For the restricted model we obtain Schwartz weights of -25.63, -28.53 and -40.28 for the *trans*-, *cis*- and allosteric phosphorelay models, respectively; but here the fits are noticeably reduced in quality compared to the unrestricted model. Based on

Molecular BioSystems Accepted Manuscript

these results, the bi-molecular allosteric model is clearly preferable to the other two mechanisms; neither the *cis* nor the *trans* model achieve more than negligible statistical support by comparison.

We explored the models further by evaluating their sensitivity and robustness with respect to variation in parameters around their respective maximum-likelihood estimates (ESM Fig. S4 and S5). All three models, reassuringly, show very similar sensitivity profiles; in each case the most sensitive parameter is the rate of autophosphorylation of the H1 domain (Fig. 3C). This is also borne out by a detailed robustness analysis (see ESM S5-S8 for results) of the mathematical models.

Discussion

In the TCSs the HKs require dimerization to be active as a functioning signalling unit. In HK signalling, successfull sensing of a specific signal is followed by autophosphorylation of the conserved histidine residue in the HisKA domain. Studies of the autophosphorylation process have shown that most HKs autophosphorylate in an intermolecular manner, as in CheA, EnvZ and NtrB. The intramolecular autophosphorylation mechanism appears to occur less frequently but have been reported in some HKs (HK853 and PhoR), including ArcB[9,34].

ArcB belongs to a relatively rare type of HK, the non-orthodox HK, and employs a 3-step phosphorelay instead of the one-step phosphorylation of the more common simple or orthodox HKs. Given the fact that HKs are active as dimers, the same question for autophosphorylation is also of interest to phosphorelay: does the phosphorelay of a certain non-orthodox HK happen in an intramolecular (*cis*) and/or an intermolecular (*trans*) way?

Here we have used a combined experimental and theoretical approach to provide insights into the phospho-transfer in ArcB, and by extension also other non-orthodox HKs, in particular those that also show intramolecular autophosphorylation (see above). The principal challenge was to identify a set of experiments with the power to discriminate amongst different potential phosphorelay mechanisms; in particular, we have fitted different mechanistic models to experimental data, and compared them using statistical model selection approaches. First, however, we had to design experiments that can generate data that allows us to detect differences between our mechanistic models. Then mathematical models representing different phosphorelay mechanisms were constructed. These models showed

Molecular BioSystems Accepted Manuscript

clearly that the mechanisms are indistinguishable – mathematically non-identifiable even – in studies of the WT ArcB in its phosphorylation of RR ArcA. Based on this outcome and mathematical insights, we then developed a set of ArcB mutants in which the phosphorylation activity at individual phosphorylation sites was ablated by site-directed mutations. By investigating these signal on ArcB* (and ArcB) variant mutants in combination we were able to study their *in vivo* phosphorelay activities in a manner that allows us to apply statistical model selection tools that enabled us to distinguish between the different mechanisms.

We generated the data for our model analysis from experiments performed with the native ArcB or constitutively active ArcB* variants co-transformed into cells using plasmids with different copy numbers as shown in Fig. 2. It is formaly possible that the H and D substitutions lock the ArcB or ArcB* into an inactive conformation and so change the structure of subunit(s) and the activity of the ArcB or ArcB* dimer. However, we directly show that subunits interactions are intact (Fig. 2C) and that some phospho-transfer activities, specifically with ArcB WT/mutant and mutant/mutant combinations are evident (ESM Fig. S2). In addition, the H292 mutations having the most detrimental effect in our experiments with ArcB*, do not change the structure of ArcB [33]. Also, it is important to note that intermolecular, *trans*- phosphorylation was observed *in vitro* with ArcB variants having enhanced kinase activity when truncated for its N-terminal region required for membrane binding[33]. Differences between *in vitro* actions of ArcB variant lacking its natural membrane interactions and *in vivo* activities of native membrane-bound ArcB could well be related to an increased level of freedom of ArcB subunits movement otherwise limited in an ArcB dimer tightly associated with the membrane leading to the indigenous redox response and phosphorelay[28]. In agreement, no *trans*- phosphorilation is observed in *in vivo* complementation experiments using different H and D mutants of native BvgS[35] suggesting intramolecular H1-D1-H2 phosphorelay requires a dimeric HK, and the subsequent H2-D2 transmission then acivates the RR[17]. This is in concert with with our result and the reason that some mutants combinations in our work (e.g. H292A) have outputs lower than the negative control is probably due to synergistic effect of increased phosphatase activity and the lack of phosphorylation activity of some ArcB* variants. There are some remaining quantitative differences even for our most successful model, the bi-molecular mechanism. Improving this model, by *e.g.* considering the allosteric interactions between the two monomers, or the

Molecular BioSystems Accepted Manuscript

details of the dephosphorylation process more explicitly, may help to resolve such residual disagreements.

In ESM Fig. S9 we provide a graphical depiction of more general models of phosphorelay activity and their relationships in terms of simple molecular reactions. All of these models are mathematically indistinguishable for wild-type data. Designing mutant combinations as done here does allow us to distinguishe between models. In light of such extended analysis it remains clear, however, that only models that require functional sites on both monomers are capable of explaining the data generated here.

It had been suggested that dimerization may act as a mechanism for increasing the specificity of the HK-RR interaction. Our results are very much in line with, and further strengthen this type of argument: the bi-molecular mechanism found here allows for careful transmission of the signal along the dimer. Rather than straightforward transmission in a *cis* or *trans* manner, the requirement of an allosteric set of interactions that facilitate the phosphotransfer reactions between H1 and D1, and D1 and H2 domains does allow for richer dynamics of the signal transduction along the dimer. More importantly, perhaps, it also acts as an additional safe-guard against unwanted cross-talk with other TCSs or triggering of RRs by acetyl phosphate, which is effectively ruled out by such an allosteric process. This seems sensible if the conventional reasoning that non-orthodox TCSs are more robust against noisy environmental signals: if there is an evolutionary advantage to employing phosphorelays instead of orthodox TCSs, then it appears also prudent and evolutionary favourable to protect against noise eminanting from within the cell. The cooperative action of the bimolecular relay model would ensure this insulation from many forms of cross-talk. What is needed to make progress in this area are mathematical models that are more closely informed by structural considerations. The dynamical models considered here (as well as for other modelling studies of non-orthodox TCSs)

We conclude by stressing the need for this type of integrative approach in order to understand how biological processes occur *in vivo*. For probably most biological processes and systems of interest, direct elucidation of mechanisms is easier *in vitro* than *in vivo*. In such isolation, however, the molecular processes will generally be influenced by the lack of the crowded cellular environment and the contributions of temporal and spatial effects. In order to gain mechanistic insights in an *in vivo* setting, experiments combined with

Molecular BioSystems Accepted Manuscript

mathematical analysis and statistical modelling can also be used to infer molecular mechanisms under physiological conditions. And moreover mathematical analysis can lead the way in the design of new, more discriminative experiments.

Materials and Methods

Bacterial strains, media and growth conditions. The bacterial strains used in this study are shown in Table 2. Strains were constructed by transduction using the $P1_{vir}$ bacteriophage (see Table 2) and were grown in Luria-Bertani (LB) broth or on LB agar plates at 37°C. To eliminate the Kan cassette from a Δ*arcA*::Kan mutant and construct the marker-less Δ*arcA* variant (strain MVA113), we used plasmid pCP20 and the method described by Cherepanov and Wackernagel[36] (see Table 2). The colony PCR and *creD*/*yjjY* pair of primers was used to verify that Δ*arcA* mutant in MVA113 strain had the correct structure. For *in vivo* bacterial two-hybrid (BACTH) system assays, strains were grown in LB at 30°C. For aerobic growth, overnight cultures of cells were diluted 100-fold into 5 ml of LB in a universal tube with loose fitting caps and shaken at 200 rpm. The growth in microaerobic conditions is achieved when overnight cultures are diluted 100-fold and cells are incubated in universal tubes with tight caps and shaken at 100rpm at 37°C (this is exactly the same growth condition used and verified in reference 32). For induction of the T5/lacUV5 promoters, 0.1 mM isopropyl-β-D-galactopyranoside (IPTG) was added for 1 hour. For scoring the $lacZ^+$ colonies, indicator plates containing 40 μ*l* of 20 mg/ml stock solution of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and 0.5 mM IPTG were used. Antibiotics were routinely used at the following concentrations: ampicillin (Amp; 100 $\mu g{:}ml^{-1}$), kanamycin (Kan; 25 (or 50 for BACTH) $\mu g{:}ml^{-1}$), and chloramphenicol (Cam; 30 $\mu g{:}ml^{-1}$), and spectinomycin (Spc; 50 $\mu g{:}ml^{-1}$).

DNA manipulations. Plasmids used in this study are listed in Table 2. Plasmid pGJ73 was constructed by amplifying wild type *arcB* from pJW5536(-) using PCR and primers that introduce *Xba*I-*Kpn*I restriction sites and cloning into pGEM-T Easy (Promega), and than subcloning into pAPT110 digested with *Xba*I-*Kpn*I. Using PCR-based site-specific

Molecular BioSystems Accepted Manuscript

mutagenesis (Quickchange mutagenesis kit, Stratagene) of the plasmid templates pJW5536(-), pGJ23, pGJ73 and pGJ74, we constructed plasmids pGJ71, pGJ72, pGJ74/75/77 and pGJ76/78/80, respectively (see Table 2). All constructs were verified by DNA sequencing. For the *in vivo* BACTH system experiments, coding sequences for ArcB* and its phosphorelay mutants were fused to either T25 (plasmid pKT25) or T18 (pUT18C) Cya domains as described in Jovanovic et al.[32]. The *arcB* mutants genes were PCR amplified from corresponding plasmids using primers that introduce *Xba*I-*Kpn*I restriction sites, cloned in pGEM-T Easy (Promega), and than subcloned in-frame into multiple cloning sites (MCS) of pKT25 and pUT18C, respectively. All constructs were verified by DNA sequencing. Transformation of bacteria was performed as described by Miller[37].

Western blot analysis. Bacterial cells were harvested at mid-log phase (OD600<0.5) and re-suspended in a mix of 30 μ*l* 4% SDS and 30 μ*l* Laemelli buffer (Sigma). Samples used for Western blotting were normalized according to cell growth measured at OD600. Samples were separated on 12.5% (SDS)-PAGE and transferred onto PVDF membrane using a semidry transblot system (Bio-Rad). Western blotting was performed as described[38] using antibodies to ArcB provided by D. Georgellis) (1:15000 with anti-rabbit). The proteins were detected using the ECL plus Western Blotting Detection Kit according to manufacturers guidelines (GE Healthcare). Images were captured in a FujiFilm intelligent Dark Box by an image analyser with a charge-coupled device camera (LAS-3000). Densitometry analysis was performed with MultiGauge 3.0 software (FujiFilm USA Inc., Valhalla, NY) and quantification (results expressed in arbitrary units) performed using the AIDA software.

***In vivo* bacterial two-hybrid (BACTH) system.** The adenylate cyclase (Cya)-based BACTH system allows detection of protein-protein interactions in vivo and is particularly appropriate for studying interactions among membrane proteins[39]. The BACTH assay was used here to study *in vivo* protein-protein interactions of ArcB* and its phosphorelay mutants. Protein fusions (see Fig. 1C) were assayed as described. As a negative control we used pKT25 and pUT18C vectors in the absence of fusion proteins, as a positive control we used pKT25-zip and pUT18C-zip plasmids carrying fused GCN4 leucine-zipper sequence[39]. After co-transformation of the BTH101 strain with the two plasmids expressing the fusion proteins, selection plates (Kan, Amp, X-gal and IPTG) were incubated at room temperature for 72 h. The levels of the interactions were quantified by β-Galactosidase activity in liquid cultures

Molecular BioSystems Accepted Manuscript

(see below). Chromosomal LacZ expression 3-fold above the negative control (vectors alone) value was scored as a positive interaction signal.

β-Galactosidase (β-Gal) assays. Activity from a single copy chromosomal ϕ(*cydA-lacZ*) transcriptional fusion exclusively regulated by ArcA[26] was assayed to gauge the level of *cydA* promoter activation. Cells were grown overnight at 37°C in LB broth containing the appropriate antibiotic (s) and then diluted 100-fold (initial OD600nm<0.025) into the same medium (5 ml). Following incubation to OD600nm 0.2-0.3, cultures were induced with IPTG for 1 h and then assayed for β-Gal activity as described by Miller[37]. The β-Gal activity from a chromosomal *lacZ* gene in BTH101 strain was assayed to estimate the protein-protein interactions in BACTH assay. Bacteria were grown in LB medium containing 100 $\mu g{:}ml^{-1}$ Amp and 50 $\mu g{:}ml^{-1}$ Kan at 30 for 16 h, then cultures were diluted 1:25 and grown until the OD600 nm<0.3, then 0.5 mM IPTG was added and the cells incubated for a further 1 h at 30. For all β-Gal assays, mean values of six independent assays taken from technical duplicates of three independently grown cultures of each strain were used to calculate activity. The data are shown as a mean values with SD error bars.

Simulation of phosphorelay models. The ODE models are simulated using the scipy module in Python[40]. The parameters we used are as described in ESM Table S1, but we also explore parameter space more globally by sampling parameters using Latin Hypercube sampling[41] . We model HK dimers composed of wildtype or mutated monomers.

In models of trans-phosphorelay or cis-phosphorelay, reaction rates are set to 0 if either of the substrate or product can not exist because of the site-directed mutation. For instance, if site 2 (D1) of the right dimer is mutated (D576A), k_{l1r2}, k_{r2l3}, k_{l3r2}, k_{r20} in the trans-phosphorelay model, and k_{r1r2}, k_{r2r3}, k_{r3r2}, k_{r20} in the cis-phosphorelay model will all be set to 0. In the allosteric model the forward phospho transfer reaction rate of site X is set to 0 if either of the monomers involves a mutation at site X. For instance, if site 2 of the right dimer is mutated, k_{l1r2}, k_{r2l3}, k_{r1l2}, k_{l2r3}, k_{r2r3}, k_{l2l3}, k_{r1r2}, k_{l1l2} *etc.* will all be set to 0. However, reverse phosphotransfer is blocked only when both phosphorylation sites are mutated.

In the simulations, as in the experiments, the system will have plasmids producing different monomers with different copy numbers. We refer to them as “left" monomer or

Molecular BioSystems Accepted Manuscript

"right" monomer in this work. Suppose the system contains *nl* copies of HK *U* and *nr* copies of HK *V* (*U*,*V* can be wildtype ArcB, H292A, D576A or H717A). Then the proportion of *UU* and *VV* homodimers and *UV* heterodimer among the total of HK dimers would be $[nl/(nl+nr)]^2$, $[nr/(nl+nr)]^2$ and $2nl \times nr/(nl+n3)^2$. The output of the system is calculated as the sum of the contributions from the different homo- and hetero-dimers, weighted by their stochiometries.

In addition to the full model we also consider a simplified model where we assume that similar reactions share the same reaction rate. In this more constrained configuration there are only 4 parameters for each of the three models, k_f, k_r, k_{t1} and k_{t2}. They denote the reaction rates of the forward phosphorylation, reverse phosphorylation, phsphotransfer from HK to RR and phosphotransfer from RR to HK, with $k_f \in \{k_{0x1}, k_{x1y2}, k_{x2y3}\}$, $kr \in \{k_{x20}, k_{x3y2}\}$, $k_{t1}=k_{x30}$, $k_{t2}=k_{0x3}$, $x \in \{l,r\}$, $y \in \{r,l\}$. In another model where the parameters are more flexible, and thus probably more realistic, we have 7 distinct parameters in the model: $k_{f0}=k_{0x1}$, $k_{f1}=k_{x1y2}$, $k_{f2}=k_{x2y3}$, $k_{r3}=k_{x3y2}$, $k_{r2}=k_{x20}$, $k_{t1}=k_{x30}$, $k_{t2}=k_{0x3}$, $x \in \{l,r\}$, $y \in \{l,r\}$. In both configurations, the spontaneous dephosphorylation rates of HK1 are set to be 10% of the dephosphorylation rate of D1 (k_r or k_{r2}) [13].

Parameter optimisation and model evaluation. We apply the parameters listed in Table S1 to the models, and optimise the parameters to obtain the minimum difference between the models' output and the observed experimental data. We use the fmin_l_bfgs_b function in the python module scipy to optimise the parameters. To keep the parameters in a reasonable biological range, we set the boundaries for the parameters to be in line with the biophysical constraints operating inside *E. coli*. For each parameter the lower boundary is set to be 0.1% of the unoptimised parameter, while the upper boundary is set to be 1000 folds of the original parameter. The output of each optimised model is compared with experimental data and the Bayesian Information Criteria (BIC)[24] is calculated for each model as

$$BIC = n \cdot \ln(\sigma_e^2) + k \cdot \ln(n).$$

Molecular BioSystems Accepted Manuscript

Here n is the number of datapoints, k is the number of free parameters, σ_e^2 is the error variance. The minimum value of the BIC across all models is denoted by BIC_{min}, and subsequently, the Schwarz weight of each model is calculated as,

$$\Delta_i = BIC_i - BIC_{min}$$

$$w_i = \frac{exp(\frac{-\Delta_i}{2})}{\sum_{j=1} exp(\frac{-\Delta_j}{2})}.$$

In the above equations, $i,j \in \{trans, cis, \text{allosteric}\}$. The Schwarz weight equals the probability that, a model is the correct model (given the data and the panel of candidate models considered). It provides an approximation to the marginal likelihood, the central object in Bayesian model selection.

Sensitivity assays. We performed sensitivity analysis on the model parameters using the matlab package StochSens[42]. This package calculates the Fisher Information Matrix (FIM) for each model and calculates their sensitivities from the FIMs[43].

Acknowledgements

We thank P. Mehta for technical help, C. Georgelis for ArcB antibodies, J. Beckwith for plasmids, M. Bekker for strains, and D. Ladant for providing strains and plasmids. We gratefully acknowlegde financial support from the BBSRC, MRC, the Kwok Foundation, the Wellcome Trust and the Leverhulme Trust. Finally, we thank two anonymous referees for their comments which have helped in improving the manuscript.

Author Contribution

GJ, XS, MB and MPHS conceived and designed the study. GJ performed the experiments; XS, AA, and PK analyzed the data; XS, AA, EF, PK, HAH, CW and MPHS developed and

Molecular BioSystems Accepted Manuscript

analyzed mathematical models. HJ, XS, MB and MPHS wrote the manuscript.

Conflict of Interest

The authors have no conflict of interest.

References

1. A. M. Stock, V. L. Robinson, and P. N. Goudreau, *Annu. Rev. Biochem.*, 2000, **69**, 183–215.
2. R. Gao and A. M. Stock, *Annu. Rev. Microbiol.*, 2009, **63**, 133–154.
3. P. Casino, V. Rubio, and A. Marina, *Curr. Opin. Struct. Biol.*, 2010, **20**, 763–771.
4. E. J. Capra and M. T. Laub, *Annu. Rev. Microbiol.*, 2012, **66**, 325–347.
5. Z. Zhang and W. A. Hendrickson, *J. Mol. Biol.*, 2010, **400**, 335–353.
6. M. Y. Galperin, *Curr. Opin. Microbiol.*, 2010, **13**, 150–159.
7. K. Wuichet, B. J. Cantwell, and I. B. Zhulin, *Curr. Opin. Microbiol.*, 2010, **13**, 219–225.
8. K. Wuichet and I. B. Zhulin, *Sci Signal*, 2010, **3**, ra50–ra50.
9. X. Sheng, M. Huvet, J. W. Pinney, and M. P. H. Stumpf, *Adv. Exp. Med. Biol.*, 2012, **751**, 121–137.
10. E. Alm, K. Huang, and A. Arkin, *PLoS Comput. Biol.*, 2006, **2**, e143.
11. M. Galperin, R. Higdon, and E. Kolker, 2010, **6**, 721–728.
12. J. Cheung and W. A. Hendrickson, *Curr. Opin. Microbiol.*, 2010, **13**, 116–123.
13. J.-R. Kim and K.-H. Cho, *Comput Biol Chem*, 2006, **30**, 438–444.
14. C. P. Barnes, D. Silk, X. Sheng, and M. P. H. Stumpf, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, **108**, 15190–15195.
15. A. Csikász-Nagy, L. Cardelli, and O. S. Soyer, *J R Soc Interface*, 2011, **8**, 480–488.
16. M. Knudsen, E. Feliu, and C. Wiuf, *J. Theor. Biol.*, 2012, **300**, 7–18.
17. R. Malpica, G. R. P. Sandoval, C. Rodríguez, B. Franco, and D. Georgellis, *Antioxid. Redox Signal.*, 2006, **8**, 781–795.
18. A. I. Podgornaia and M. T. Laub, *Curr. Opin. Microbiol.*, 2013, **16**, 156–162.
19. P. Casino, V. Rubio, and A. Marina, *Cell*, 2009, **139**, 325–336.
20. D. Georgellis, O. Kwon, and E. C. Lin, *Science*, 2001, **292**, 2314–2316.
21. M. D. Rolfe, A. Ter Beek, A. I. Graham, E. W. Trotter, H. M. S. Asif, G. Sanguinetti, J. T. de Mattos, R. K. Poole, and J. Green, *J. Biol. Chem.*, 2011, **286**, 10147–10154.
22. M. Bekker, S. Alexeeva, W. Laan, G. Sawers, J. Teixeira de Mattos, and K. Hellingwerf, *J. Bacteriol.*, 2010, **192**, 746–754.
23. G. Pena-Sandoval, O. Kwon, and D. Georgellis, *J. Bacteriol.*, 2005, **187**, 3267–3272.
24. P. Kirk, T. W. Thorne, and M. P. H. Stumpf, *Curr. Opin. Biotechnol.*, 2013, **24**, 767–774.
25. J. Monod, J. P. Changeux, and F. Jacob, *J. Mol. Biol.*, 1963, **6**, 306–329.
26. H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser, *Nature*, 2014, **508**, 331–339.
27. H. A. Harrington, K. L. Ho, T. W. Thorne, and M. P. H. Stumpf, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, 15746–15751.
28. O. Kwon, D. Georgellis, and E. C. C. Lin, *J. Biol. Chem.*, 2003, **278**, 13192–13195.
29. M. T. Laub and M. Goulian, *Annu. Rev. Genet.*, 2007, **41**, 121–145.
30. J. Skerker, M. S. Prasol, B. S. Perchuk, E. G. Biondi, and M. T. Laub, *PLoS Biol.*, 2005, **3**, e334.
31. J. Liepe, S. Filippi, M. Komorowski, and M. P. H. Stumpf, *PLoS Comput. Biol.*, 2013, **9**, e1002888.
32. G. Jovanovic, C. Engl, and M. Buck, 2009, **74**, 16–28.

Molecular BioSystems Accepted Manuscript

33. G. Pena-Sandoval and D. Georgellis, *J. Bacteriol.*, 2010, **192**, 1735–1739.
34. J. A. Hoch, 2000, **3**, 165–170.
35. M. A. Uhl and J. F. Miller, *EMBO J.*, 1996, **15**, 1028–1036.
36. P. P. Cherepanov and W. Wackernagel, *Gene*, 1995, **158**, 9–14.
37. J. H. Miller, *A Short Course in Bacterial Genetics*, CSHL Press, 1992.
38. S. Elderkin, S. Jones, J. Schumacher, D. Studholme, and M. Buck, *J. Mol. Biol.*, 2002, **320**, 23–37.
39. G. Karimova, J. Pidoux, A. Ullmann, and D. Ladant, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, **95**, 5752–5756.
40. B. G. Olivier, J. M. Rohwer, and J. Hofmeyr, *Mol. Biol. Rep.*, 2002, **29**, 249–254.
41. A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis*, John Wiley & Sons, 2008.
42. M. Komorowski, J. Zurauskiene, and M. P. H. Stumpf, *Bioinformatics*, 2012, **28**, 731–733.
43. M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, **108**, 8645–8650.
44. T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori, *Molecular Systems Biology*, 2006, **2**, 2006.0008.
45. M. Kitagawa, T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga, and H. Mori, *DNA Res.*, 2005, **12**, 291–299.

Supporting information

The supporting information contains additional experimental and modelling results in addition to the results of the sensitivity analysis.

Molecular BioSystems Accepted Manuscript

cis	**trans**	**allosteric**	
$HK_0 \longrightarrow HK_{L1}$	$HK_0 \longrightarrow HK_{L1}$	$HK_0 \longrightarrow HK_{L1}\phi(L_1, R_1)$	autophosphorylation of $L1$ domain on L monomer upon signal
$HK_{L1} \longrightarrow HK_{L2}$	$HK_{L1} \longrightarrow HK_{R2}$	$HK_{L1} \longrightarrow HK_{L2}\phi(L_1, R_1, L_2, R_2)$ $HK_{L1} \longrightarrow HK_{R2}\phi(L_1, R_1, L_2, R_2)$	
$HK_{L2} \rightleftharpoons HK_{L3}$	$HK_{R2} \rightleftharpoons HK_{L3}$	$HK_{L2} \rightleftharpoons HK_{L3}\phi(L_2, R_2, L_3, R_3)$ $HK_{R2} \rightleftharpoons HK_{L3}\phi(L_2, R_2, L_3, R_3)$	
$HK_{L2} \longrightarrow HK_0$	$HK_{R2} \longrightarrow HK_0$	$HK_{L2} \longrightarrow HK_0$	dephosphorylation
$HK_{L3} + RR \rightleftharpoons HK_0 + RR_p$	$HK_{L3} + RR \rightleftharpoons HK_0 + RR_p$	$HK_{L3} + RR \rightleftharpoons HK_0 + RR_p$	phosphorylation of RR
$HK_0 \longrightarrow HK_{R1}$	$HK_0 \longrightarrow HK_{R1}$	$HK_0 \longrightarrow HK_{R1}\phi(L_1, R_1)$	autophosphorylation of $R1$ domain on R monomer upon signal
$HK_{R1} \longrightarrow HK_{R2}$	$HK_{R1} \longrightarrow HK_{L2}$	$HK_{R1} \longrightarrow HK_{R2}\phi(L_1, R_1, L_2, R_2)$ $HK_{R1} \longrightarrow HK_{L2}\phi(L_1, R_1, L_2, R_2)$	
$HK_{R2} \rightleftharpoons HK_{R3}$	$HK_{L2} \rightleftharpoons HK_{R3}$	$HK_{R2} \rightleftharpoons HK_{R3}\phi(L_2, R_2, L_3, R_3)$ $HK_{L2} \rightleftharpoons HK_{R3}\phi(L_2, R_2, L_3, R_3)$	
$HK_{R2} \longrightarrow HK_0$	$HK_{L2} \longrightarrow HK$	$HK_{R2} \longrightarrow HK_0$	dephosphorylation
$HK_{R3} + RR \rightleftharpoons HK_0 + RR_p$	$HK_{R3} + RR \rightleftharpoons HK_0 + RR_p$	$HK_{R3} + RR \rightleftharpoons HK_0 + RR_p$	phosphorylation of RR

Table 1: Biomolecular reactions of the three models considered here. Here L and R subscripts denote the two monomers of the dimers; the numerical subscripts denote the three domains; that is phosphorylation at the -H1, D1, and –H2 domains are indicated by the subscripts 1, 2 and 3. For the allostertic model we have a indicator function $\phi(x, y, \ldots)$, which is equal to one if all the arguments (i.e. the corresponding domains) are present and zero otherwise. E.g. for the phosphorylation of D1 to be enabled, both monomers must carry functional H1 and D1 domains. Only then can the phosphoryl group move from the H1 (or H2 in the case of phosphatase activity of ArcB) to D1.

Molecular BioSystems Accepted Manuscript

Strain or plasmid	Relevant characteristics	Reference
Strain		
MG1655	Wild type	Laboratory collection
ASA12	MC4100 *recA*, λRSS2[ϕ(‘*cydA-lacZ*)] (kan^r)	[22] A gift from M. Bekker
MG1655 $P_{cydA\text{-}lacZ}$	MG1655 [ϕ(*cydA-lacZ*)] (kan^r)	This work MG1655xASA12
MVA92	Δ*arcB59*	[32]
MVA104	MVA92 [ϕ(*cydA-lacZ*)] (kan^r)	This work MVA92×ASA12
JWK4364	BW25113 Δ*arcA*::Kan (kan^r)	[44]
MVA113	Δ*arcB59* Δ*arcA*	This work MVA92×JWK44364×pCP20
MVA114	MVA113 [ϕ(*cydA-lacZ*)] (kan^r)	This work MVA113×ASA12
BTH101	cya^-, lac^+	A gift from D. Ladant
XL1-Blue	tet^r	Laboratory collection
Plasmid		
pCA24N	Expression vector, $P_{T5/lac}$ promoter, $lacI^q$, *ori* pMB1, (cam^r)	[45]
pJW5536(-)	$P_{T5/lac}$-*6xhis-arcB* (*arcB* cloned into pCA24N, encoding ArcB wild type), $lacI^q$ (cam^r)	[45]
pGJ33	pJW5536(-) encodes $ArcB_{H292A}$ (cam^r)	[32]
pGJ71	pJW5536(-) encodes $ArcB_{D576A}$ (cam^r)	This work
pGJ21	pJW5536(-) encodes $ArcB_{H717A}$ (cam^r)	[32]
pGJ23	pJW5536(-) encodes $ArcB_{C180A/C241A}$ (ArcB*) (cam^r)	[32]
pGJ30	pGJ23 encodes $ArcB^*_{H292A}$ (cam^r)	[32]
pGJ72	pGJ23 encodes $ArcB^*_{D576A}$ (cam^r)	This work
pGJ27	pGJ23 encodes $ArcB^*_{H717A}$ (cam^r)	[32]

Molecular BioSystems Accepted Manuscript

pAPT110	Expression vector, P_{lacUV5} promoter, *lacI*^{q}, *ori* p15A, (spc^r)	*A gift from J. Beckwith*
pGJ73	P_{lacUV5}-*6xhis-arcB* (*arcB* cloned into pAPT110 *Xba*I-*Kpn*I, encoding ArcB wild type (spc^r)	This work
pGJ77	pGJ73 encodes $ArcB_{D576A}$ (spc^r)	This work
pGJ75	pGJ73 encodes $ArcB_{H717A}$ (spc^r)	This work
pGJ74	pGJ73 encodes ArcB* (spc^r)	This work
pGJ80	pGJ74 encodes $ArcB^*_{H292A}$ (spc^r)	This work
pGJ78	pGJ74 encodes $ArcB^*_{D576A}$ (spc^r)	This work
pGJ76	pGJ74 encodes $ArcB^*_{H717A}$ (spc^r)	This work
pKT25	IPTG-inducible vector containing the T25 domain of Cya upstream of the MCS (kan^r)	A gift from D. Ladant
pUT18C	IPTG-inducible vector containing the T18 domain of Cya upstream of the MCS (amp^r)	A gift from D. Ladant
pKT25-zip	GCN4 leucine zipper fusion to the C-terminus of the T25 domain of Cya in pKT25 (kan^r)	A gift from D. Ladant
pUT18C-zip	GCN4 leucine zipper fusion to the C-terminus of the T18 domain of Cya in pUT18C (amp^r)	A gift from D. Ladant
pCP20	*FLP*$^+$, λ *c*I857$^+$, λ p_R Rep^{ts}, (amp^r, cam^r)	[36]
pGEM-T Easy	Cloning vector (amp^r)	Promega

Table 2. *E. coli* K-12 strains and plasmids used in this study.

Figure Legends

Figure 1: Diagrams of the models and constructs used in this paper. (A): (i), orthodox two-component system. (ii), non-orthodox two component system with an intermediate Hpt (His-containing phosphotransfer) protein. (iii), non-orthodox two-component system with no intermediate molecule. (iv)-(vi), different possible phosphorelay mechanisms of the non-orthodox two-component systems. (B): Diagram of ArcB mutants. i, the diagram of phosphorelay and the mutation sites.

Molecular BioSystems Accepted Manuscript

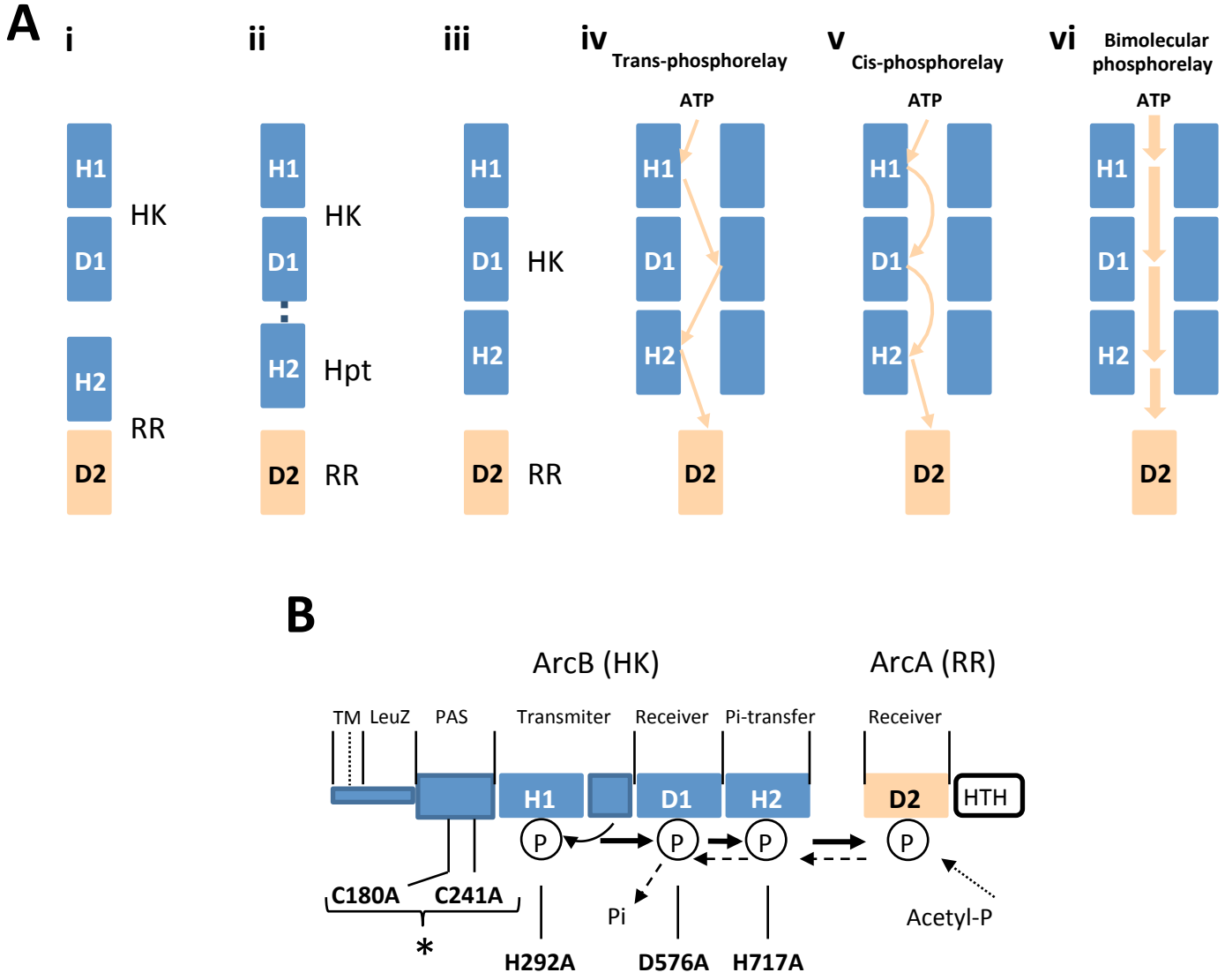
Figure 2: Phosphorelay of constitutively active ArcB* variants lacking the key phosphorelay residues. (A): Expression of a *cydA−lacZ* chromosomal transcription fusion under uniquecontrol of the ArcBA system was measured using a β-Gal assay (see*Materials and Methods*) in cells grown in aerobiosis (i) or microaerobiosis (ii) in the presence of 0.1 mM IPTG. *arcB*$^{+}$ (MG1655 P*cydA-lacZ*); Δ*arcB* (MVA104); Δ*arcB* (MVA104) carried vector pCA24N or pAPT110 alone, expressed plasmid (pCA24N- or pAPT110-based) borne ArcB wild type [pKW5536(-) or pGJ73] or constitutively active ArcB* (pGJ23 or pGJ74). (B): The level of ArcB* and its variant proteins (see Fig. 1B) expression from either pCA24N or pAPT110 based plasmids in a Δ*arcB* (MVA104) strain grown in microaerobiosis (control was vector pCA24N alone) was assessed using Western blotting and antibodies against ArcB (α-ArcB) (see*Materials and Methods*). The level of expressionwas presented in arbitrary units after quantification using AIDA software. (C): The *in vivo* BACTH system was used to detect protein-protein interactions between the ArcB* and its variant proteins fused to T25 or T18 subunit of adenylate cyclase and expressed in the presence of 0.5 mM IPTG in a *cya*$^{-}$ BTH101 strain. For construction of fusion proteins, growth conditions and calculation of mean values and SD see*Materials and Methods*. Negative control: BTH101/pKT25+pUT18C vectors alone; Positive control: BTH101/pKT25-zip+pUT18C-zip; the levels of the interactions were quantified by β-Gal activity in liquid cultures grown microaerobically. (D): Expression of a *cydA−lacZ* chromosomal transcription fusion was measured using a β-Gal assay in Δ*arcB* (MVA104) cells grown in microaerobiosis in the presence of 0.1 mM IPTG co-expressing ArcB* or its variants from either pCA24N- (left hand part) or pAPT110-based (right hand part) plasmids (see schematic presentation below graph). (E) As in sub-Fig. D, except cells expressed different combinations of the ArcB* and/or its variants from co-transformed pCA24N- and pAPT100-based plasmids (see schematic presentation below graph; the last two columns represent ArcB* WT replicate experiments). (F) Expression of a *cydA−lacZ* chromosomal transcription fusion was measured using a β-Gal assay in either Δ*arcB* (MVA104) or Δ*arcB*Δ*arcA* (MVA114) cells grown in microaerobiosis in the presence of 0.1 mM IPTG carrying either pCA24N or pAPT110 alone, both vectors alone, or expressing ArcB* from either pCA24N- or pAPT110-based plasmids.

Molecular BioSystems Accepted Manuscript

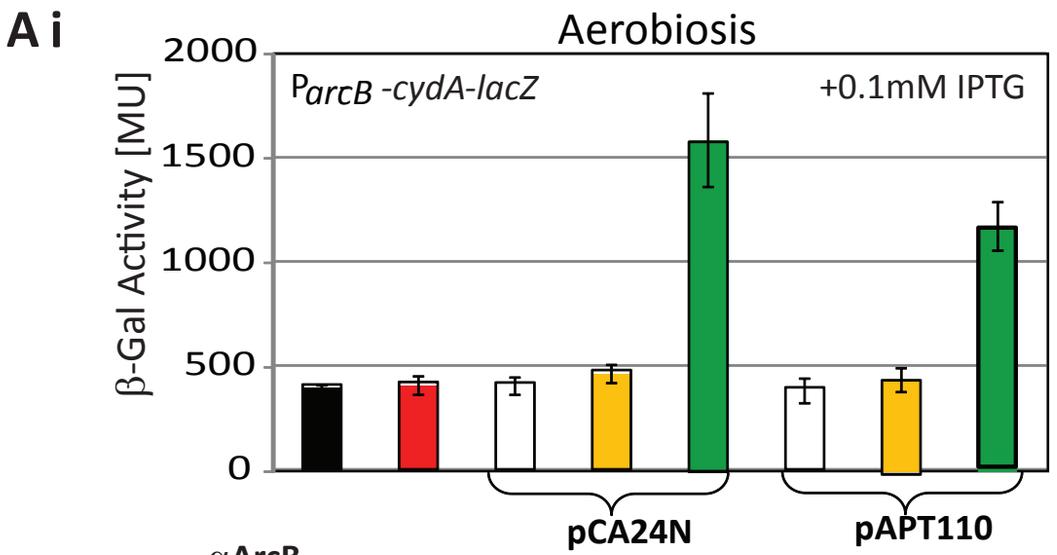
Figure 3: (A): Simulation of three phosphorelay models with more flexibale parameter settings. In these simulations, there are 7 parameters for each of the three models, kf_0,kf_1,kf_2,kr_3,kr_2,$kt_1$and kt_2, as described in the*Materials and methods* section. Normalised [RRp] is used as output for each model. At the top of part A we show the combinations of ArcB* mutants in the two co-transformed plasmids corresponding to the measured and simulated activity levels in the three panels below. The most left lane in each panel corresponds to the negative control where only empty plasmids are co-transformed. In all the other lanes, the left bar stands for the constructs based on the higher copy number plasmids pCA24N, while the right bar denotes the constructs based on the lower copy number plasmids pAPT110. In each sub-Fig., the bars with the darker colors show the levels of the normalised [RRp] from model simulation, while the lighter coloured bars show the levels of normalised β-Gal activities in the experimental data (as shown in Fig. 2E). (B): Schwarz Weight of three phosphorelay models with optimised parameters under more flexible parameter configuration. The probabilities are: trans-phosphorelay model, 0.43%; cis-phosphorelay model, 0.02%; allosteric, 99.55%. (C): Sensitivity of each phosphorelay model on the seven biophysical reaction rate parameters. Each model is most sensitive to variation in the rate of autophosphorylation, *kf0*. This is also borne out by the detailed sensitivity analyses in the ESM (ESM Figs. S5 – S8).

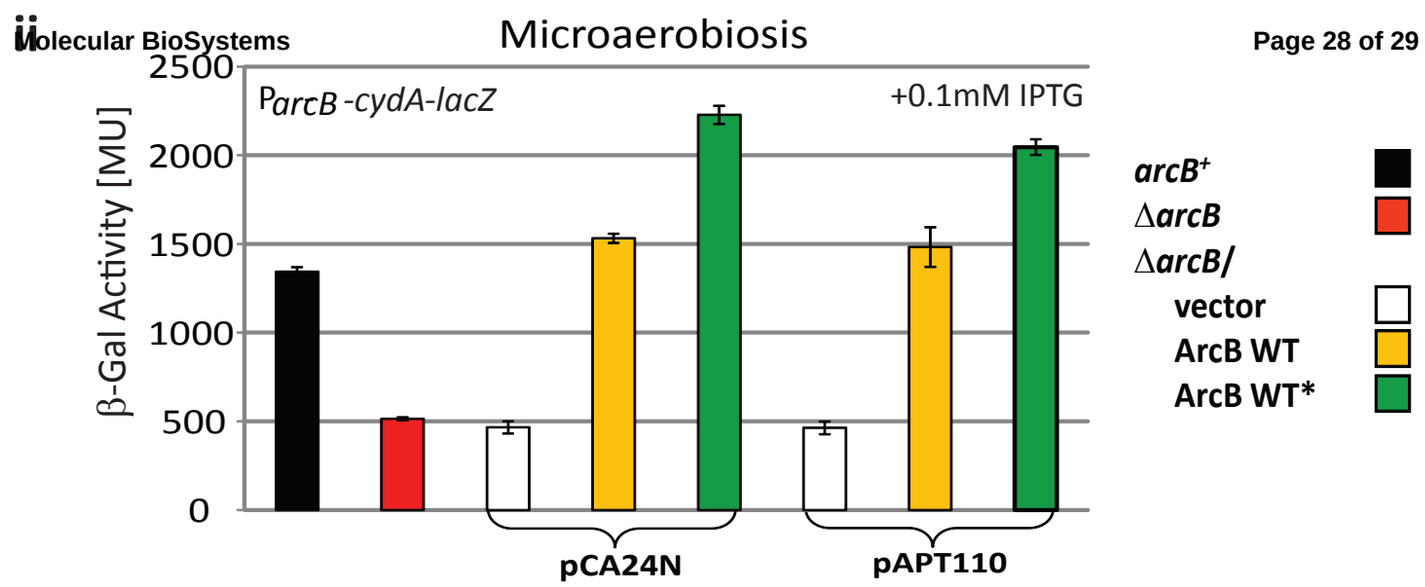
Molecular BioSystems Accepted Manuscript

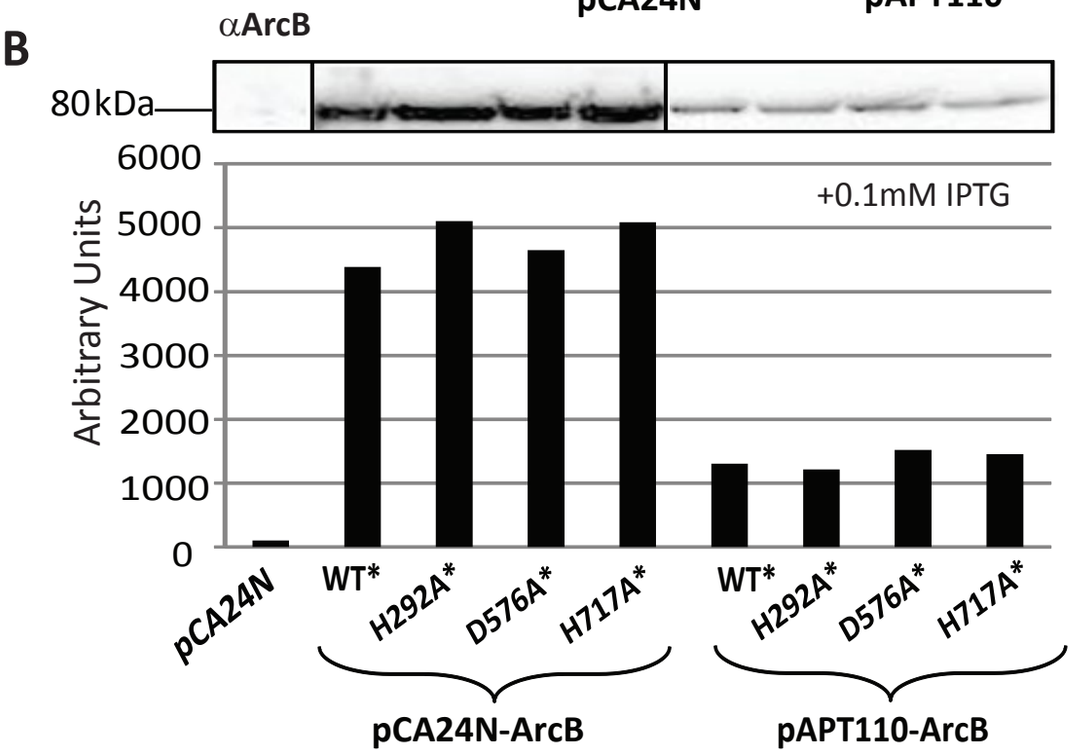
Fig. 1

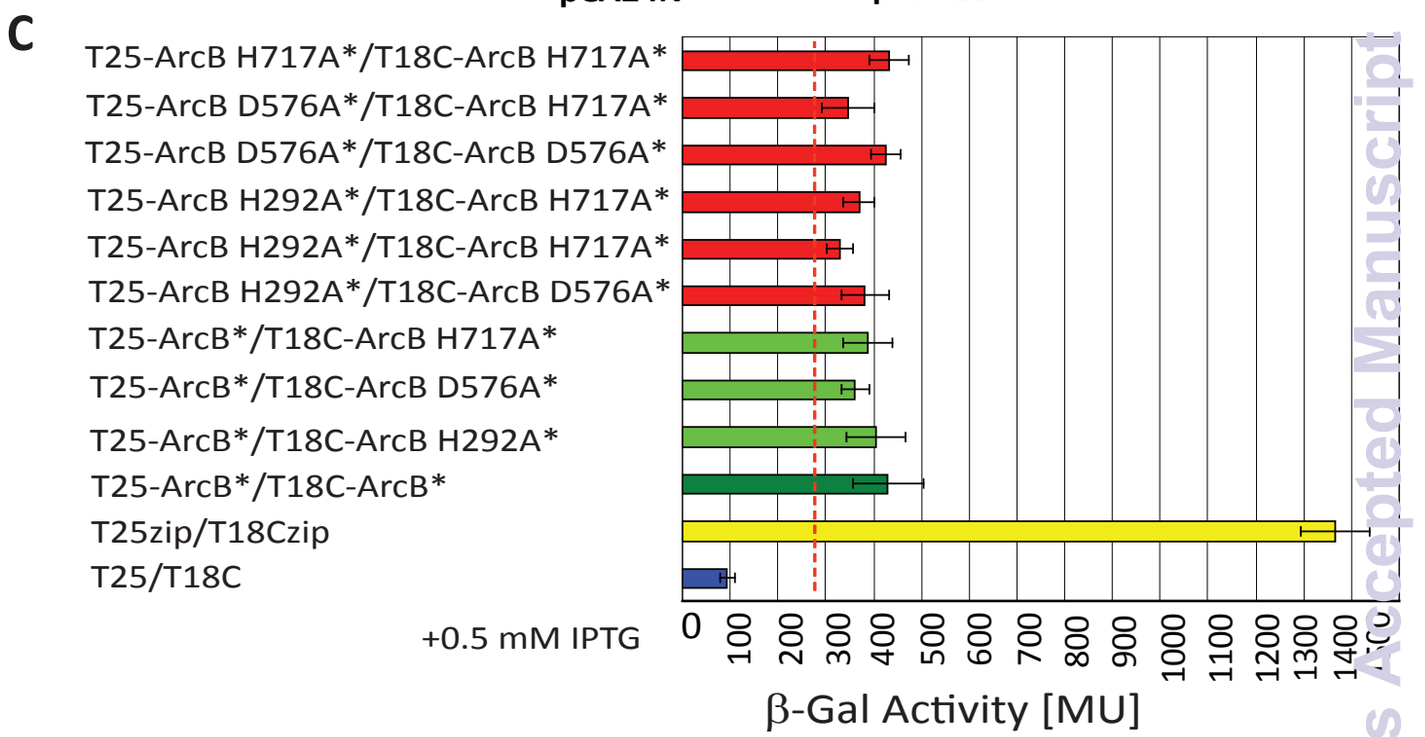


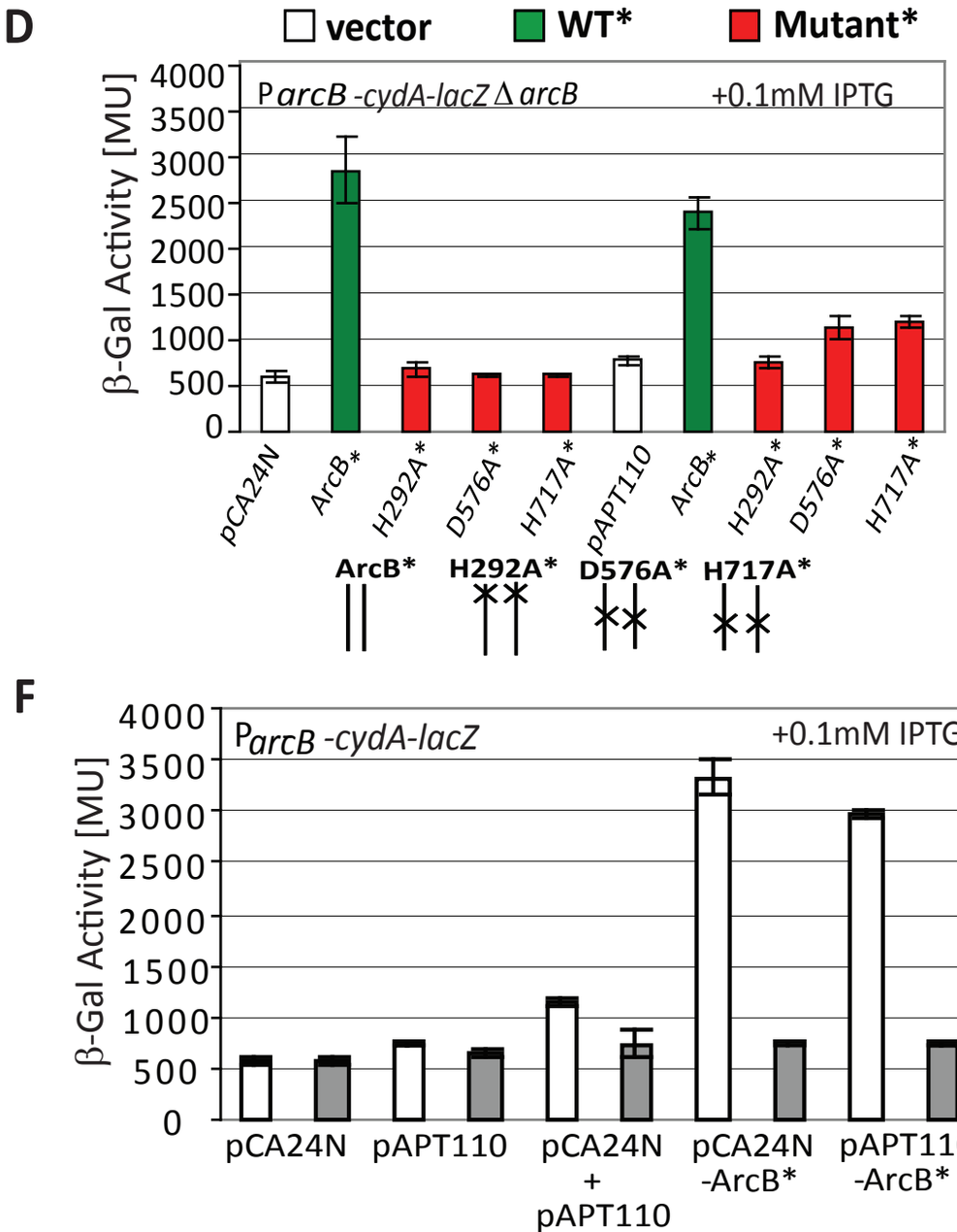
Molecular BioSystems Accepted Manuscript

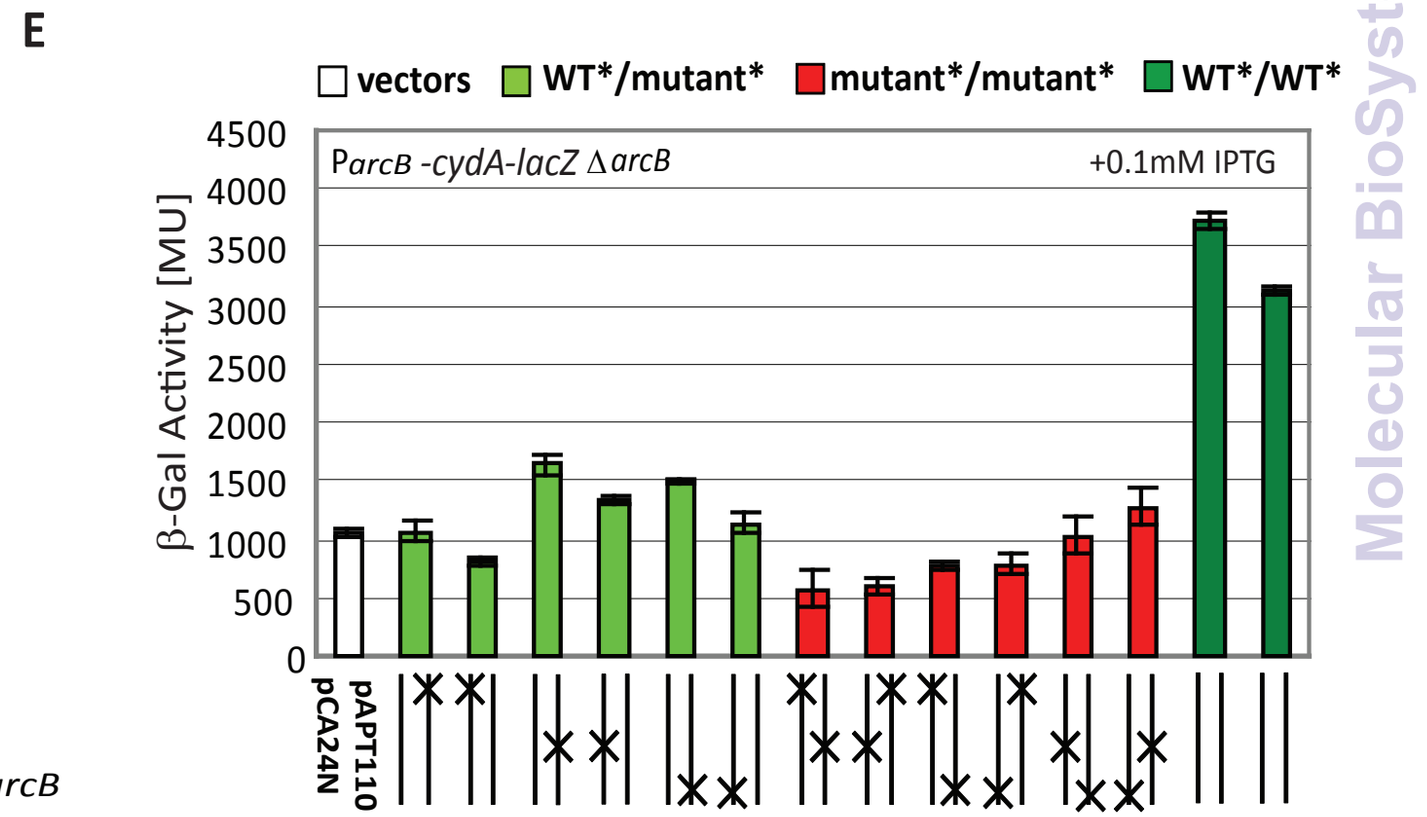


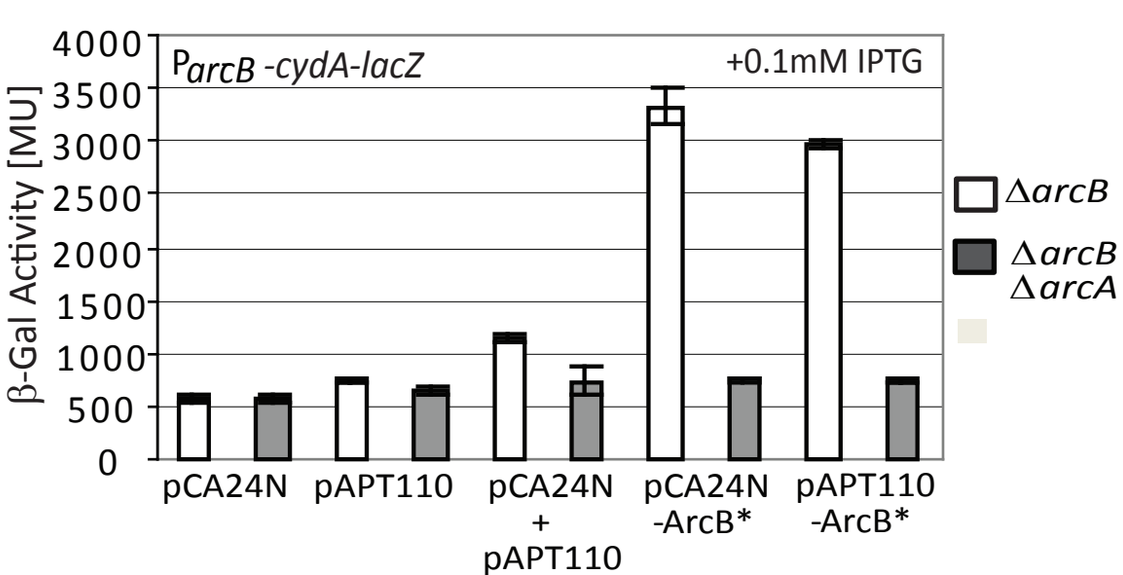




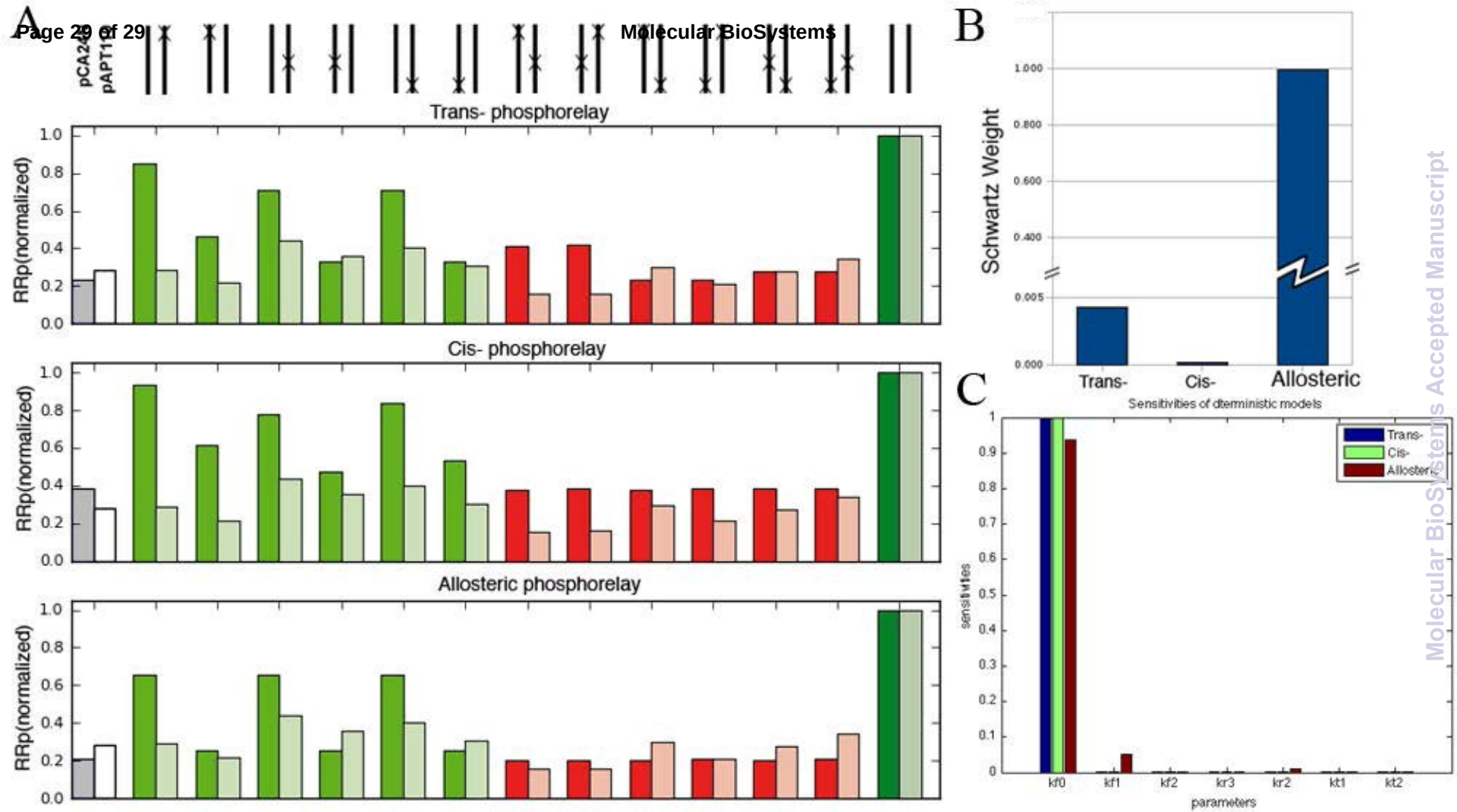








Molecular BioSystems Accepted Manuscript



Molecular BioSystems Accepted Manuscript