Molecular Microbiology (2003) **48**(1), 25–51

Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: differences in molecular structure and physiological function

Rui Alves[1,2,3] and Michael A. Savageau[1*†]

[1]*Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II Ann Arbor, MI 48109–0620, USA.*

[2]*Grupo de Bioquimica e Biologia Teoricas, Instituto Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250 Lisboa, Portugal.*

[3]*Programa Gulbenkian de Doutoramentos em Biologia e Medicina, Departamento de Ensino, Instituto Gulbenkian de Ciencia, Rua da Quinta Grande 6, 1800 Oeiras, Portugal.*

Summary

Signal transduction by a traditional two-component system involves a sensor protein that recognizes a physiological signal, autophosphorylates and transfers its phosphate, and a response regulator protein that receives the phosphate, alters its affinity toward specific target proteins or DNA sequences and causes change in metabolic activity or gene expression. In some cases the sensor protein, when unphosphorylated, has a positive effect upon the rate of dephosphorylation of the regulator protein (bifunctional sensor), whereas in other cases it has no such effect (monofunctional sensor). In this work we identify structural and functional differences between these two designs. In the first part of the paper we use sequence data for two-component systems from several organisms and homology modelling techniques to determine structural features for response regulators and for sensors. Our results indicate that each type of reference sensor (bifunctional and monofunctional) has a distinctive structural feature, which we use to make predictions regarding the functionality of other sensors. In the second part of the paper we use mathematical models to analyse and compare the physiological function of systems that differ in the type of sensor and are otherwise equivalent. Our results show that a bifunctional sensor is better than a monofunctional sensor both at amplifying changes in the phosphorylation level of the regulator caused by signals from the sensor and at attenuating changes caused by signals from small phosphodonors. Cross-talk to or from other two-component systems is better suppressed if the transmitting sensor is monofunctional, which is the more appropriate design when such cross-talk represents pathological noise. Cross-talk to or from other two-component systems is better amplified if the transmitting sensor is bifunctional, which is the more appropriate design when such cross-talk represents a physiological signal. These results provide a functional rationale for the selection of each design that is consistent with available experimental evidence for several two-component systems.

Introduction

Two-component systems (TCS) are signal transduction modules that exist mainly in bacteria but have also been found in fungi and plants (recently reviewed in Hellingwerf *et al.*, 1998; see also Parkinson, 1993; Hoch and Silhavy, 1995; Perego and Hoch, 1996; Schaller, 1997; Posas *et al.*, 1998). Two-component systems differ from eukaryotic phosphorylation cascades like the MAP kinase cascade. In the former, ATP is only consumed in the first step of the cascade to provide for autophosphorylation of a histidine residue in the sensor, whereas in the latter, ATP is consumed at each step in the cascade to provide for phosphorylation of the protein at that step. More than 180 different TCS have been identified in bacteria (Kadner, 1996; see also http://www.genome.ad.jp/kegg/regulation.html) and over one thousand putative sensors and response regulators in over one hundred organisms have had their gene sequence determined (e.g. http://www.expasy.ch/srs5/ http://www-fp.mcs.anl.gov/~gaasterland/genomes.html). Two-component systems are often involved in complex circuits that exhibit enormous variation in design. The physiological importance of

Accepted 28 October, 2002. *For correspondence. E-mail masavageau@ucdavis.edu; Tel. (+1) 530 752 1033; Fax (+1) 530 754 5739. †Present address: Department of Biomedical Engineering, One Shields Avenue, University of California, Davis, CA 95616, USA.

TCS provides strong motivation to search for underlying design principles that would allow one to rationalize the variations in design. The search is in its infancy and there are undoubtedly many design principles that remain to be discovered.

Here we focus on a specific example of a simple design principle involving TCS composed of a sensor protein and a response regulator protein. These traditional TCS can be considered the prototype for a motif that has been elaborated on and modified in more complex cases that will not be considered here. [For example, we will not address cases in which there are specific aspartate phosphatases (e.g. Perego and Hoch, 1996; Blat *et al*., 1998) that are independent of the sensor protein.] In some prototype cases the sensor protein, when unphosphorylated, has a positive effect upon the rate of dephosphorylation of the regulator protein (bifunctional sensor) whereas in other cases it has no such effect (monofunctional sensor). For example, the TCS involved in the regulation of chemotaxis via the CheA/CheY cascade in *Escherichia coli* has a monofunctional sensor (reviewed in Eisenbach, 1996), whereas the TCS involved in the regulation of osmotic pressure via the EnvZ/OmpR cascade has a bifunctional sensor (reviewed in Pratt *et al*., 1996). These bifunctional and monofunctional sensor proteins exhibit specific differences in molecular structure that, by homology modelling, are found to reoccur in many other TCS.

In the first part of this paper we use sequence data for other TCS from several organisms, together with homology modelling techniques, to predict structural features and in turn functionality of their sensor proteins. These predictions are confirmed in a few cases for which there is independent biochemical or genetic evidence for functionality. Our results suggest an association between a sensor's mode of action (monofunctional or bifunctional) and the structure of its ATP-binding domain. In the second part of this paper we analyse and compare the functional effectiveness of TCS with either bifunctional or monofunctional sensors by using a technique known as Mathematically Controlled Comparison. This technique, which is analogous to a well-controlled experiment, determines the differences in the physiology between alternative designs that are otherwise equivalent. This approach reveals qualitative differences (independent of specific values for the parameters of the systems) as well as quantitative differences (statistical tendencies when a large sample of numerical values for the parameters is examined). We also obtain results that discriminate between different types of cross-talk to and from TCS. Our results suggest physiological situations that favour each of the two sensor designs. Hence, these results provide a functional rationale for the selection of each design. Experimental evidence for several TCS is discussed in the light of this rationale.

Results

Abstractions of the alternative designs for a prototype TCS are represented in Fig. 1. The unphosphorylated sensor protein (S,X_3) becomes phosphorylated (S^*,X_1) in response to changes in a physiological signal (Q_1,X_5). Autophosphorylation of the sensor is achieved with consumption of ATP. Once phosphorylated, the sensor loses its phosphate group, either by auto-dephosphorylation (although this occurs on a time scale that is not of interest here) or by transfer of the phosphate to an aspartate residue in the regulator protein (R^*,X_2) (e.g. Weiss and Magasanik, 1988; Sanders *et al*., 1989; Hsing *et al*., 1998; Jung and Altendorf, 1998; Jiang and Ninfa, 1999). This covalent modification causes changes in the response regulator, thereby altering its affinity toward specific target proteins or DNA sequences (T), and this in turn leads to changes in metabolic activity or gene expression. The dephosphorylation of response regulators (e.g. Perego and Hoch, 1996) is, in some cases, enhanced by the unphosphorylated form of the sensor protein (bifunctional sensor, Fig. 1A). In other cases, the sensor protein lacks this ability (monofunctional sensor, Fig. 1B). In either case, the regulator protein also can be phosphorylated by other mechanisms (Q_2,X_6), e.g. by small phosphodonors like acetyl-phosphate, but not by ATP (e.g. Lukat *et al*., 1992; McCleary, 1996; Mayover *et al*., 1999). Cross-talk from the sensor proteins of other two-component systems has been demonstrated with purified proteins *in vitro* (Ninfa *et al*., 1988) and with over-expressed proteins *in vivo* (Grob *et al*., 1994), but the physiological significance of these results has been questioned (Wanner, 1992).

Structure of regulators

Structures of some response regulators have been determined and can be used as templates to predict the 3D structure of other regulators by homology modelling. The known structures of response regulators have receiver domains composed of alternating α-helices and β-sheets. The 3D shape resembles a barrel composed of five α/β sequential segments, with parallel β-sheets (blue) in the middle, surrounded by the α-helices (red) shown in Fig. 2. In the protein classification database SCOP (Murzin *et al*., 1995) response regulators are classified within the flavodoxin-like folds. The aspartate residue that is phosphorylated is at the tip of one of the internal β-sheets, almost in the loop that connects it to the next α-helix (Fig. 2A). These structural features appear to be at least partially conserved in the receiver domains of all the response regulators we have been able to model (Fig. 2B, Table 1). Comparison of the experimentally determined structures of

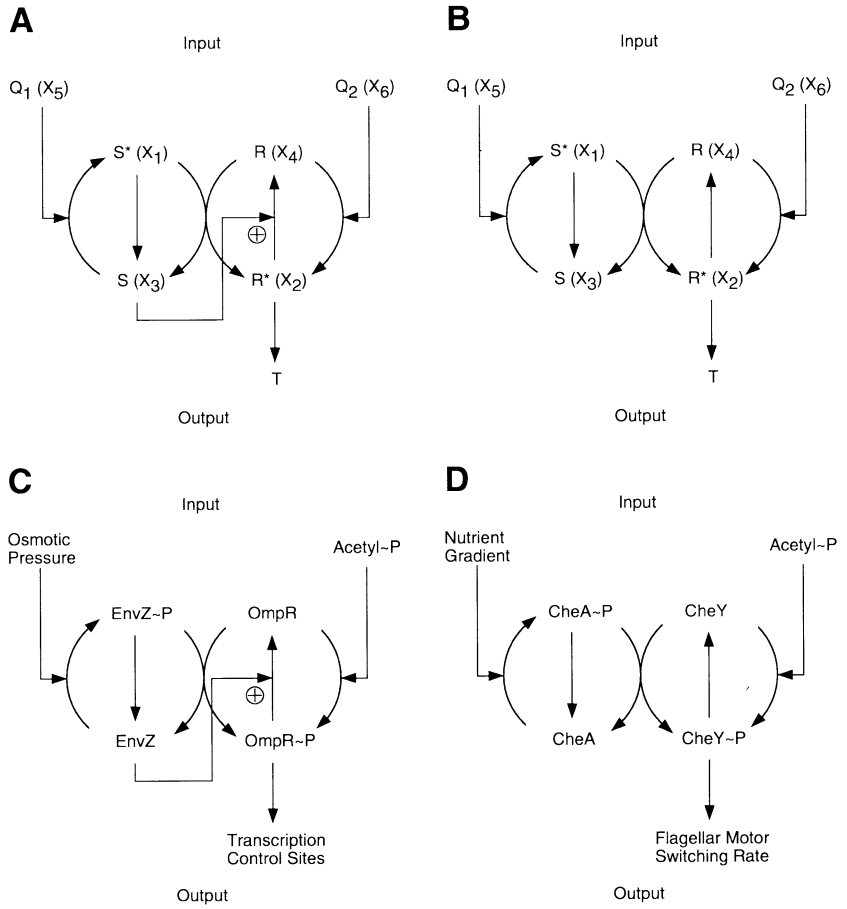

Fig. 1. Schematic representation of two-component system modules with either a bifunctional or a monofunctional sensor.
A. Module with a bifunctional sensor in which *S** phosphorylates *R and S* increases the rate of dephosphorylation of *R**.
B. Module with a monofunctional sensor in which *S** phosphorylates *R and S* and has no direct effect on the rate of dephosphorylation of *R**.
C. Example of a TCS module with a bifunctional sensor. EnvZ is the sensor protein of the module. It is a membrane protein that responds to changes in the osmotic pressure by changing its phosphorylation state. In its phosphorylated state (EnvZ~P) it transfers phosphate to the response regulator OmpR. Phosphorylated OmpR represses expression of proteins that form large pores and induces expression of proteins that form small pores. Unphosphorylated EnvZ increases the rate of OmpR~P dephosphorylation, repressing expression of small membrane pores and derepressing expression of larger membrane pores. OmpR can also be phosphorylated by other phospho-donors, for example acetyl-P.
D. Example of a TCS module with a monofunctional sensor. CheA is a membrane protein that is attached to a complex sensorial system. The system senses gradients of nutrients in the medium. When a nutritional gradient is sensed, CheA autophosphorylates and transfers its phosphate to the response regulator CheY. CheY can also be phosphorylated by other phospho-donors, for example acetyl-P. Phosphorylated CheY changes the tumbling movement of the cells into a directed motion towards the higher concentrations of nutrients by affecting flagellar proteins. Unphosphorylated CheA has no effect on CheY.
Symbols: *S* (X_3) – unphosphorylated sensor protein; *S** (X_1) – phosphorylated sensor protein; *R* (X_4) – unphosphorylated response regulator protein; *R** (X_2) – phosphorylated regulator protein; Q_1 (X_5) – signal that modulates the fraction of phosphorylated sensor; Q_2 (X_6) – signal that modulates the fraction of phosphorylated regulator independent of signalling by the cognate sensor; *T* – target for the response regulator.

response regulators did not reveal structural features that might indicate any distinction between response regulators of TCS with either bifunctional sensors or monofunctional sensors. Similar results also were obtained when we examined the modelled structures for response regulators. Although the phosphorylation domain of all response regulators whose structure has been determined or modelled so far have similar structures this does not allow us to conclude that the hundreds to thousands more that are as of yet unresolved all share the same structure. There are known cases where protein domains with very similar sequence have very different folds and cases where protein domains with very different sequence have very similar folds. Table 1 may be used as a guide to indicate response regulators that are likely to have similar structures. In the effort to explore the fold space of response regulators, it will probably be more informative to concentrate on resolving structures of response regulators that are not present in Table 1.

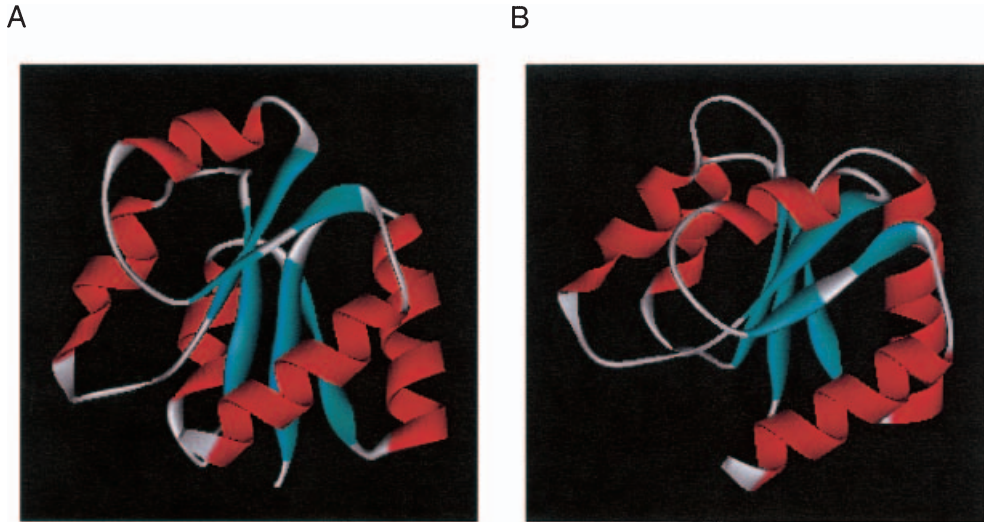

Fig. 2. Three-dimensional structure for the receiver domain of response regulators.
A. Structure determined by X-ray crystallography for the Spo0F regulator from *Bacillus subtilis* (PDB reference identifier 1NAT).
B. Structure predicted by homology modelling for the ArcA regulator from *Haemophilus influenza*.

Structure of sensors

Partial structures for both types of sensors also have been determined and can be used as templates to predict the 3D structure of other sensors by homology modelling. Structures of the soluble domains of the CheA sensor (monofunctional, Fig. 3A) as well as the EnvZ sensor (bifunctional, Fig. 4A) have been determined. The CheA transmitter domain is of the type HPt, reminiscent of the PTS phosphotransferase systems. The EnvZ sensor has an HK transmitter domain. In the protein classification database SCOP (Murzin *et al*., 1995) sensor histidine kinases are classified within the ROP-like folds ('four helices; dimers of identical alpha-hairpin subunits; bundle, closed, left-handed twist'). Even though the domains of the EnvZ sensor responsible for its bifunctionality are not defined, an HK transmitter domain appears to be essential for the increase in the rate of dephosphorylation of the response regulator (Zhu *et al*., 2000). In the same work, Zhu *et al*. (2000) also showed that this increase is much less for the bifunctional sensors if their catalytic ATP-binding domain is truncated from their transmitter domain.

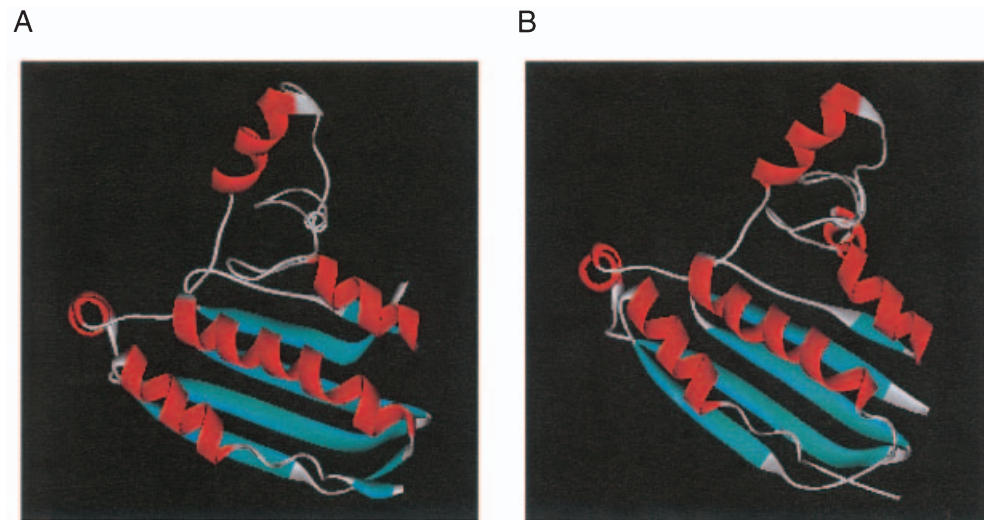

Fig. 3. Three-dimensional structure for the catalytic domain of monofunctional sensors.
A. Structure determined by X-ray crystallography for the CheA sensor from *Thermotoga maritima* (PDB reference identifier 1B3Q).
B. Structure predicted by homology modelling for the CheA sensor from *Escherichia coli*.

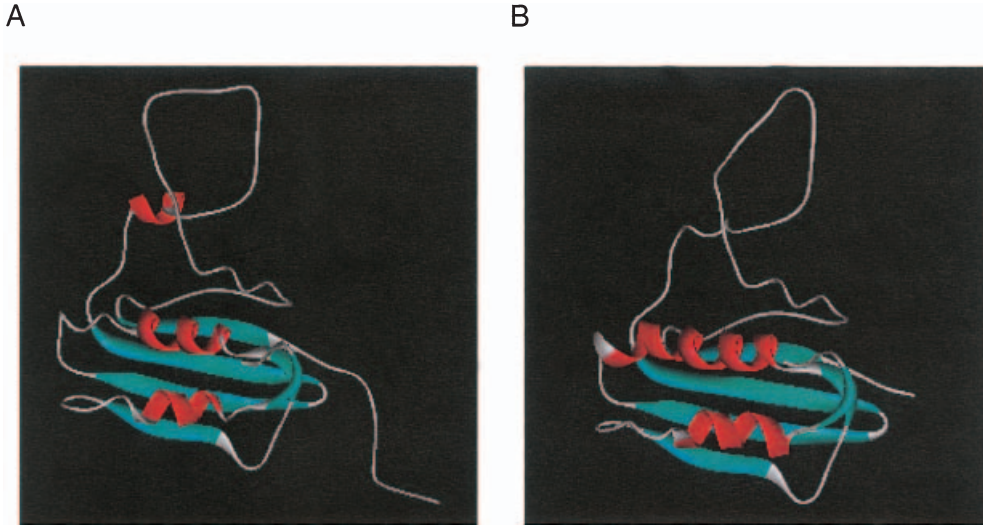

Fig. 4. Three-dimensional structure for the catalytic domain of bifunctional sensors.
A. Structure determined by X-ray crystallography for the EnvZ sensor from *Escherichia coli* (PDB reference identifier 1BXD).
B. Structure predicted by homology modelling for the putative O66656 sensor from *Aquifex aeolicus*.

Table 1. Putative response regulator proteins that yielded a predicted structure by homology modelling techniques.

Organism	SWISSPROT entry no.	Name	Organism	SWISSPROT entry no.	Name
Aeromas jandaei	P97015	P97015	*Bacillus megaterium*	P39486	YGD1
	P97018	P97018	*Bacillus subtillis*	P45709	CCDB
Agrobacterium	Q07783	CHVI		Q05522	CHEB
tumefaciens	P07545	VIRG		P37583	CHEY
	Q44430	NTRR		P14204	CMA1
Alcaligenes eutrophus	Q44006	CZCR		O34534	CITT
	P29267	HOXA		P13800	DEGU
	P94153	P94153		Q02115	LYTR
Anabaena sp.	P39048	PatA		O34804	PHOP
	O87395	OrrA		P35163	RESD
Aquifex aeolicus	O66551	NtrC1		P06534	SP0A
	AQ1792	NtrC2		P06535	SP0B
	O66596	NtrC3		P06628	SP0F
	O66657	PhoB		Q9R9J8	YBDG
Arabidopsis thaliana	Q9ZWS8	Q9ZWS8		O31432	YBDJ
	Q9ZWS7	Q9ZWS7		P40759	YCBB
	Q9ZWK0	Q9ZWK0		P42244	YCBL
	Q9ZWJ9	Q9ZWJ9		P70955	YCCH
	Q9ZWS9	Q9ZWS9		P94413	YCLJ
	O80365	O80365		O31517	YESN
	O80366	O80366		P94439	YFIK
	Q9ZQJ8	Q9ZQJ8		O07528	YHCZ
	O82798	O82798		O34903	YKOG
Archeoglobus fulgidus	O29221	CheB		O34723	YOCG
	O28035	O28035		P54443	YRKP
	O28381	O28381		O34951	YTSA
	O28799	O28799		O05251	YUFM
	O28887	O28887		O06978	YVCP
	O29012	O29012		Q9L4F3	YVFU
	O29199	O29199		O32192	YVQA
	O29220	O29220		O32197	YVQC
	O29800	O29800		P94504	YVRH
	O30252	O30252		P42421	YXDJ
Azorhizobium	P26487	FixJ		P55184	YXJL
caulonidans	Q04849	NtrX		P37478	YYCF
Azospirillum brasiliense	P45671	NtrC	*Bacillus thurigiensis*	P52935	Spo0A
Azobacter vinelandii	O87478	O87478		P52942	Spo0F
Bacillus brevis	Q44929	GTCR	*Bacteroides*	Q45788	RTEA
	P54662	DEGU	*thetaiotamicron*	Q00936	RTEB
	P52929	SP0A	*Bordetella pertussis*	P16574	BvgA

Table 1. *Cont.*

Organism	SWISSPROT entry no.	Name	Organism	SWISSPROT entry no.	Name
	O88117	RisA		P07027	UVRY
Borrelia burgdorferi	O51380	O51380		P97172	YEDW
	O51501	O51501		P33356	YEHT
	O51517	O51517		P52076	YGIX
	O51615	O51615		P77380	YLCA
	O51704	O51704	*Eubacterium acidaminophilum*	O86183	NtrC
Bradyrhizobium japonicum	P72489	RegR			
	P10576	NTRc	*Fermiella diplosiphon*	O32610	O32610
	Q45272	NwsB		Q01473	RCAC
	O52845	PhoB	*Guillardia theta*	O78425	YC27
	P31908	HoxA	*Haemophilus influenza*	P44918	ARCA
	P23221	FixJ		P76777	CPXR
Brucella abortus	Q9ZHS1	CtrA		P44845	NARP
	O67996	BvrR		P45189	PHOB
Brucella melitensis	Q9XDD4	Q9XDD4		P45337	YGIX
Brucella suis	O31409	FEUP	*Helicobacter pylori*	O25337	CHEV
Campylobacter jejuni	P71129	CheY		P71403	CHEY
	O68795	O68795		HP0703	HP0703
Campobacterium divergens	Q9ZEG3	Q9ZEG3		HP1365	HP1365
Caulobacter crescentus	Q45976	DviK		O25408	O25408
	Q46020	PleD		O25684	O25684
	Q45994	SokA		O25918	O25918
Chlamydia trachomatis	O84474	O84474		O24973	OMPR
Clostridium difficile	P52938	Spo0A	*Hyphomicrobium methylovorum*	P56644	SgaR
Clostridium innocuum	P52939	Spo0A			
Clostridium pasteurianum	P52940	Spo0A	*Klebsiela pneumoniae*	P03029	NTRC
Clostridium perfringens	Q9XDT7	VirI		P45605	PHOB
Cyanidium caldarium	P48259	P48259	*Lactobacillus sake*	Q9ZI91	Q9ZI91
	P48359	P48359		Q9ZI93	Q9ZI93
	P28257	YC27		Q9ZI95	Q9ZI95
Cyanophora paradoxa	P48259	Yc27		Q9ZI97	Q9ZI97
Desulfovibrio vulgaris	P33394	RRF1		Q9ZIA0	Q9ZIA0
Enterococcus faecalis	Q06239	VanRA	*Lactococcus lactis*	Q9ZI77	Q9ZI77
	Q47744	VanRB	*Listeria monocytogenes*	Q48767	Q48767
Erwinia amylovora	Q9X3S9	HrpY	*Methylobacterium extorquens*	O30796	MXAB
Erwinia carotovora	O86197	EXPM		Q49121	Q49121
	O32556	O32556[a]	*Mycobacterium leprae*	Q50136	PHOP
	Q9ZIN9	Q9ZIN9		Q4Z5G8	Q4Z5G8
	O08235	RPFA		Q50136	YV17
Escherichia coli	P03026	ARCA	*Mycobacterium tuberculosis*	Q10531	COPR
	Q06065	ATOC		Q50447	MTRA
	B2381	B2381		MTV43	MTV43
	P30846	BAER		MTV44	MTV44
	P30843	BASR		MTV25	MTV25
	P78070	CHEB		O50447	O50447
	P06143	CHEY		O50806	O50806
	P76777	CPXR		O50447	O50447
	P08368	CREB		O50806	O50806
	P76794	DCUR		O53856	O53856
	Q54149	DPIA		O53894	O53894
	P30854	EVGA		O69730	O69730
	P21502	FIMZ		P71814	PHOP
	P37055	HNR		Q50825	Q50825
	P14375	HYDG		Q11156	RGX3
	P77512	YFHA		O07776	TCRA
	P76822	KDPE		Q50806	TRCR
	P10957	NARL	*Neisseria meningitides*	NMB0114	NMB0114
	P31802	NARP		NMB0595	NMB0595
	P06713	NTRC	*Nostoc punctiforme*	Q51309	Q51309
	P03025	OMPR	*Plectonema boryuanum*	P51586	YSO1
	P08402	PHOB	*Porphyra purpurea*	P51358	YC27
	P23836	PHOP		P51343	YC29
	P14374	RCSB	*Porphyrium aerugineum*	P28835	YC27
	P35163	RESD	*Proteus vulgaris*	P28787	NTRC
	P52108	RSTA	*Providencia stuartii*	O85058	O85058
	P56644	SGAR		O85059	O85059
	P77344	TORR	*Pseudoalteromonas sp.*	O68498	O68498
	P10940	UHPA	*Pseudomonas aeruginosa*	P29369	AGMR

Table 1. *Cont.*

Organism	SWISSPROT entry no.	Name	Organism	SWISSPROT entry no.	Name
	P23747	ALGB	*Streptomyces coelicolor*	O50496	O50496
	P26275	ALGR	*Streptomyces hygroscopicus*	Q54292	Q54292
	Q51455	CHEY	*Streptomyces lividans*	P72471	P72471
	Q51454	FLER		Q9X6J0	Q9X6J0
	Q51373	GACA	*Streptomyces peucetius*	Q54821	Q54821
	P72150	GLTR	*Streptomyces violaceoruber*	Q9ZA47	Q9ZA47
	O33493	O33493	*Synechococcus sp.*	O68523	NBLR
	O54039	O54039		P72561	P72561
	Q04803	PFER		Q56180	PHOB
	P23620	PHOB		P39663	SPHR
	P46384	PILG		Q56003	SRRB
	P43501	PILH	*Synechocystis sp.*	P72781	P72781
	Q00934	PILR		P72790	P72790
Pseudomonas aureofaciens	Q9XD07	GACA		P72936	P72936
Pseudomonas pseudomallei	O31395	IRLR		P72948	P72948
Pseudomonas putida	Q52201	PPRA		P73006	P73006
Pseudomonas solanacearum	O07846	O07846		P73036	P73036
	O07847	O07847		P73078	P73078
	Q52582	VIRR		P73104	P73104
	Q45417	VSRD		P73150	P73150
Pseudomonas syringae	Q02540	COPR		P73175	P73175
	Q52406	CORP		P73176	P73176
	Q52376	GACA		P73243	P73243
Pseudomonas tolaasii	O34175	PHEN[a]		P73404	P73404
Rhizobium leguminosarum	Q9X574	CELR		P73927	P73927
	P10046	DCTD		P73686	P73686
	Q52852	POPP		P73864	P73864
Rizhobium meliloti	P50350	CHVI		P73928	P73928
	P13632	DCTD		P74014	P74014
	O54063	EXSF		P74138	P74138
	P10577	NTRC		P74139	P74139
	Q52990	PHOB		P74288	P74288
	Q52913	Q52913		P74294	P74294
	P55701	Y4XI		P74298	P74298
Rhodobacter sphaeroides	O33553	CHER[a]		P74314	P74314
	O30741	DMSO		P74541	P74541
	O32479	DMSR		P74626	P74626
	P95652	P95652		Q55733	Q55733
	Q53228	REGA		Q55808	Q55808
Rhodobacter sulfidophilus	O82868	RegA		Q55890	Q55890
Rhodococcus erythropolis	Q9ZNJ4	Q9ZNJ4		Q55919	Q55919
Rickettsia prowazekii	O05971	OmpR		Q55933	Q55933
Saccaromyces cereviseae	P38889	SNK7		Q55942	Q55942
Salmonella dublin	O85302	CopR		slr0474	RCP1
Salmonella typhimurium	P36556	BASR		S11797	S11797
	P04042	CHEB	*Thermotoga maritima*	Q9WY30	Q9WY30
	P06657	CHEY		Q9WYN0	Q9WYN0
	P26319	FIMZ		Q9WYT9	Q9WYT9
	P25852	HYDG		Q9WZY6	Q9WZY6
	Q56065	MVIA		Q9X181	Q9X181
	P41789	NTRC		Q56312	CHEY
	P41405	OMPR		P74922	P74922
	P06184	PGTA		Q9WXY0	Q9WXY0
	P14146	PHOP		Q9WXZ0	Q9WXZ0
	Q56127	RCSB	*Thiocystis violacea*	P45365	P45365
	Q56128	RCSC[a]	*Vibrio cholerae*	O30664	O30664
	P96058	SIRA		O68318	O68318
	P22104	TCTD		O85089	O85089
	P27667	UHPA		Q9X2S6	Q9X2S6
Shigella dysenteriae	P45606	PhoB		Q9ZH73	Q9ZH73
Shigella flexneri	P45607	PhoB	*Vibrio harveyi*	P54299	LUXO
Staphylococcus aureus	P96456	LYTS	*Wollinela succinogenes*	P74968	P74968
	Q9XCM7	YYCF	*Xanthomas campestris*	Q04527	Q04527
Streptococcus pneumoniae	Q54954	CIAR	*Yersinia pseudotuberculosis*	P74991	PHOP
	O54138	KDPE			
	Q9X4S8	PNPR			

a. These response regulators are fused to sensors as part of a composite TCS consisting of a sensor-regulator protein, a sensor-regulator-sensor protein, or a sensor-regulator-sensor-regulator protein.

We have made predictions regarding the monofunctional or bifunctional character of the sensor for many TCS (Table 2). In most cases, TCS have been identified by sequence similarity, and definitive biochemical evidence regarding the monofunctional or bifunctional character of their sensor is lacking. Our predicted molecular models for these sensors reveal a consistent difference in the folding of their catalytic domain: the fold resembles either that of the monofunctional template or that of the bifunctional template.

We also have tested our predictions with the TCS for which biochemical and genetic evidence regarding the monofunctional or bifunctional character of their sensor is available and homology modelling was possible (Table 3). The sensor protein has an ATP-binding domain that contains a small sequence of amino acids giving rise to a 3D structure known as the ATP lid. This structure is highly organized and its position shifts, enclosing the nucleotide or releasing it from its binding site. The secondary structure of the ATP lid is mostly α-helical. If the sensor is known to be bifunctional, we find that its ATP lid resembles that of EnvZ (Fig. 3B). Approximately in the middle of the amino acid sequence, the α-helix fold is interrupted by a small non-α-helical T-loop. If the sensor protein is known to be monofunctional, we find a different folding of the ATP lid in which the three-dimensional structure resembles the ATP-binding domain of CheA (Fig. 4B). In this case, the α-helical structure of the ATP lid is interrupted by two

Table 2. Putative sensor proteins that yielded a predicted structure by homology modelling techniques.

Organism	SWISSPROT entry no.	Name	Prediction	Organism	SWISSPROT entry no.	Name	Prediction
Agrobacterium tumefaciens	Q07737	ChvG	Bi	*Calothrix viguieri*	O52937	O52937	Bi
				Candida albicans	O74171	O74171	Lid[d]
Aquifex aeolicus	O66656	O66656	Bi		O42696	Canik1	Lid[d]
	O66597	O66597	Bi		O42695	CasIn1	Lid[d]
Arabidopsis thaliana	P49333	ETR1	Lid[d]		O59892	HK1	Lid[d]
	O22267	O22267	Lid[d]	*Caulobacter crescentus*	Q9X688	CckA	Lid[b]
	O48834	O48834	Lid[d]		Q03228	DivJ	Bi
	O82798	O82798	Lid[d]		P37894	PleC	Lid[b]
Archeoglobus fulgidus	O28171	O28171	Lid[b]	*Colletotrichum gloeosporioides*	P79086	CHK1	Lid[b]
	O28653	O28653	Lid[b]				
	O28800	O28800	Lid[b]	*Deinococcus* radiodurans	DR1175	DR1175	Bi[c]
	O28789	O28789	Lid[b]		DR1606	DR1606	Lid[b]
	O29222	O29222	Mono		DR0892	DR0892	Bi[a]
	O30214	O30214	Lid[b]		DR2416	DR2416	Lid[b]
Azorhizobium caulonidans	P26489	FixL	Lid[b]		DR205	DR205	Lid[d]
	Q04850	NtrY	Bi[a]		DR744	DR744	Bi
Azospirillum brasiliense	P45670	NtrB	Mono[e]	*Dictyostelium discoideum*	O15763	DhkB	Lid[d]
					Q23901	HkA	Lid[d]
Bacillus subtillis	P29072	CheA	Mono		O15783	O15783	Lid[d]
	P37599	CheV	Lid[d]		O15784	HkC	Lid[d]
	Q03069	DegM	Lid[b]	*Enterococcus faecalis*	Q47745	VanSB	Bi
	P16497	KinA	Lid[b]				
	Q08430	KinB	Bi[a]	*Escherichia coli*	P22763	ARCB	Mono
	P39764	KinC	Lid[b]		Q06067	ATOS	Bi
	P23545	PhoR	Bi		P30847	BAES	Bi[a]
	P35164	ResE	Lid[b]		P26607	BARA	Lid[d]
	Q9Z2K7	SpaF	Lid[b]		P30844	BASS	Bi[a]
	P33113	SpaK	Bi		P08336	CPXA	Bi
	P94414	YclK	Bi		P08401	CREC	Bi[a]
	BG11896	YxjM	Bi		P77644	EVGS	Lid[d]
	O34638	YkoH	Bi		P14377	HYDH	Bi
	O32193	IvqB	Bi		P21865	KDPD	Lid[b]
	O34989	YvrG	Lid[b]		P06712	NTRB	Mono[e]
	P42422	YxdK	Bi		P23837	PHOQ	Bi[a]
	Q45614	YycG	Bi		P76457	RCSC	Lid[d]
Bordetella bronchiseptica	P26762	BvgS	Bi[c]		P35164	RESE	Lid[b]
					P18392	RSTB	Bi[a]
Bordetella parapertusis	P40330	BvgS	Bi[c]		P75887	TORS	Bi[a]
					P77485	YBCZ	Bi[a]
Bordetella pertussis	P16575	BvgS	Bi[c]		P76339	YEDV	Bi[a]
Borrelia burgdorferi	Q44737	CheA	Mono		P76587	YFHK	Bi[a]
	O51381	O51381	Lid[d]	*Fusarium solani*	O94094	FikS	Lid[d]
	Q44875	Q44875	Mono	*Haemophilus* influenza	P44578	ARCB	Bi[a]
Bradyrhizobium japonicum	P23222	FixL	Lid[b]		P71380	PHOR	Mono
	P10578	NtrB	Mono[e]		P45336	YGIY	Bi

Table 2. *Cont.*

Organism	SWISSPROT entry no.	Name	Prediction
Halobacterium salinarium	Q48297	CheA like	Mono
Helicobacter pylori	Q9ZKE9	Q9ZKE9	Mono
	O25153	O25153	Mono
Herbaspirillum seropedicae	O86056	O86056	Lid[b]
Klebsiela pneumoniae	P06218	NTRB	Mono[e]
	P45608	PHOR	Lid[b]
Lactobacillus sake	Q9ZI92	Q9ZI92	Bi
	Q9ZI94	Q9ZI94	Lid[b]
Lactococcus lactis	Q48675	NisK	Bi
	O07382	O07382	Mono[e]
	O07383	O07383	Bi[a]
	O07384	O07384	Bi[a]
	O07386	O07386	Bi[a]
Listeria monocytogenes	Q48768	CheA	Mono
Mastigocladus laminosus	O07114	O07114	Mono
Methanobacterium thermoautotrophicum	MT902	MT902	Lid[d]
	MT901	MT901	Lid[d]
	O26540	O26540	Lid[d]
	O26544	O26544	Lid[b]
	O26545	O26545	Lid[d]
	O26546	O26546	Lid[d]
	O26547	O26547	Lid[d]
	O26557	O26557	Lid[d]
	O26648	O26648	Lid[d]
	O26879	O26879	Lid[d]
	O26913	O26913	Lid[d]
	O26987	O26987	Lid[d]
	O26988	O26988	Lid[d]
	O27644	O27644	Lid[d]
	O27792	O27792	Lid[d]
Methanococcus jannaschii	MJECL24	Spo0J	Lid[b]
Methylobacterium extorquens	Q49120	Q49120	Lid[b]
Mycobacterium leprae	P54883	YV16	Bi
Mycobacterium tuberculosis	P96372	KDPD	Bi
	MT1302	MT1302	Lid[b]
	MT3764	MT3764	Lid[b]
	MT490	MT490	Lid[b]
	MT902	MT902	Bi
	MT982	MT982	Lid[b]
	MT600	MT600	Bi
	O53895	O53895	Lid[b]
	P71815	P71815	Bi
	mtu:Rv0758	PhoR	Bi
	Q10560	Q10560	Bi
	Q11155	Sex3	Lid[b]
	Y902	Y902	Bi
Neisseria meningitides	Q9K1K2	NtrX	Lid[d]
	NMB114	NMB114	Bi
	NMB1792	NMB1792	Bi
	NMB594	NMB594	Bi
Neurospora crassa	O93851	NIK1	Lid[d]
	Q01318	OS1P	Lid[d]
Nostoc punctiforme	Q9Z693	Q9Z693	Lid[b]
Proteus mirabilis	O85662	RcsB	Lid[d]
Pseudomonas aeruginosa	Q51453	FLES	Lid[b]
	O34206	KINB	Bi
	Q04804	PFES	Bi[a]
	P33639	PILS	Bi[a]
	O68597	PIRS	Bi
	Q51453	PSEA	Bi[a]
	Q51420	Q51420	Lid[b]
	O34206	KINB	Bi[a]
	O68596	O68596	Lid[d]
Pseudomonas pseudomallei	O31396	IrlS	Bi
Pseudomonas putida	O07831	BziP	Lid[b]
Pseudomonas solanacearum	O07845	O07845	Bi[a]
Pseudomonas syringae	O06437	CVGS	Lid[d]
	P48027	GACS	Lid[b]
	Q9WWH5	Q9WWH5	Lid[d]
	Q9WWH9	Q9WWH9	Lid[d]
Pseudomonas tolaasii	Q9ZNQ1	RtpA	Lid[b]
Pyrococcus horikoshii	O58192	CheA	Mono
Rathayibacter rathayi	O34971	KdpD	Bi
Rhizobium leguminosarum	P10047	DctB	Bi[a]
	P41503	NtrB	Mono[e]
Rizhobium meliloti	Q52880	CHEA	Mono
	P13633	DCTB	Bi[a]
	P10955	FIXL	Lid[b]
	Q52912	Q52912	Bi
Rhodobacter capsulatus	P09431	NTRB	Mono[e]
	P37739	DCTS	Bi[a]
Rhodobacter sphaeroides	Q53135	CheA	Mono
	Q53163	HupS	Lid[d]
	O33554	O33554	Mono
	Q53162	Q53162	Bi[a]
Rhodospirillum centenum	Q9W2W8	Q9W2W8	Lid[d]
Rickettsia prowazekii	Q9ZDU5	BarA	Bi
Riftia pachyptila symbiont	O33541	RSSA	Lid[d]
	O33542	RSSB	Lid[d]
Saccaromyces cereviseae	P39928	SLN1	Lid[d]
Salmonella dublin	O85303	CopS	Bi[a]
Salmonella typhimurium	P36557	BASS	Bi
	P09384	CHEA	Mono
	P41406	ENVZ	Bi
	P37461	HYDH	Lid[b]
	Q9ZHD4	Q9ZHD4	Bi[a]
Scizosaccaromyces pombe	O14002	MCS4	Lid[d]
	O74539	O74539	Lid[d]
Shigella dysenteriae	P45609	PHOR	Lid[b]
Shigella flexneri	O31140	EnvZ	Bi
Staphylococcus aureus	Q9XCM6	YYCG	Bi
Streptococcus pneumoniae	Q54955	Q54955	Bi
Streptomyces hygroscopicus	Q54293	Q54293	Bi
Synechococcus sp.	Q56181	PHOR	Bi[a]
	P72560	P72560	Lid[b]
	P20169	P20169	Bi
	P72728	P72728	Lid[b]
	P72791	P72791	Lid[b]
	P73025	P73025	Bi[a]
	P73035	P73035	Bi[a]
	P73184	P73184	Lid[b]
	P73337	P73337	Lid[b]
	P73687	P73687	Lid[b]
	P73828	P73828	Lid[d]
	P73865	P73865	Bi[a]
	P73926	P73926	Lid[d]
	P73932	P73932	Lid[d]
	P74004	P74004	Lid[b]
	P74015	P74015	Lid[b]
Synechocystis sp.	P74622	P74622	Bi[a]
	Q55475	Q55475	Lid[b]
	Q55693	Q55693	Lid[b]

Table 2. *Cont.*

Organism	SWISSPROT entry no.	Name	Prediction	Organism	SWISSPROT entry no.	Name	Prediction
	Q55718	Q55718	Lid[b]	*Thermotoga maritima*	Q56310	CheA	Mono
	Q55783	Q55783	Lid[b]		P74923	P74923	Lid[b]
	Q55838	Q55838	Lid[b]		Q9WYN1	Q9WYN1	Bi
	Q55918	Q55918	Bi[a]		Q9WZV7	Q9WZV7	Bi
	Q55932	Q55932	Bi[a]		Q9X180	Q9X180	Lid[b]
	Q55941	Q55941	Bi[a]	*Vibrio alginolyticus*	P19906	NtrB	Mono[e]
	Q73276	Q73276	Lid[b]	*Vibrio cholerae*	O30663	FlrB	Lid[b]
	Q74111	Q74111	Lid[b]		O68317	O68317	Bi
	Q74137	Q74137	Bi		O85090	PhoR	Bi
	Q9ZAL8	Q9ZAL8	Lid[d]	*Vibrio harveyi*	P54302	LUXQ	Lid[d]
	slr 2099	slr 2099	Lid[b]		P54301	LUXN	Lid[d]
	P39664	P39664	Lid[b]	*Xanthomas campestris*	P49246	RpfC	Lid[d]
	Y437	Y437	Bi[a]				

a. These are partially modelled structures that closely resemble the ATP-binding domain of sensor EnvZ. The resulting model has the ATP lid and a portion, but not all, of the remaining ATP binding domain of the sensor EnvZ.
b.These are partially modelled structures that only slightly resemble the ATP-binding domain of sensor EnvZ. The resulting model has an ATP lid similar to that of EnvZ, but lacks the remainder of the EnvZ-like ATP binding domain.
c. These are partially modelled structures obtained from a composite TCS. The resulting model has the ATP lid and a portion, but not all, of the remaining ATP binding domain of the sensor EnvZ, as in a.
d. These are partially modelled structures obtained from a composite TCS. The resulting model has an ATP lid similar to that of EnvZ, but lacks the remainder of the EnvZ-like ATP binding domain, as in b.
e. This is a case in which the sensor protein, NRI, is monofunctional unless it is bound with a second protein, PII, in which case it becomes bifunctional.

small non- α-helical loops, with the α-helix stretch in between the two loops laying almost perpendicular to the other two α-helical stretches.

Qualitative functional differences

The concentrations of the input signal molecules (X_5 primary input, such as a chemical gradient of nutrients in the CheA case or osmotic pressure changes in the EnvZ case, and X_6 secondary input, such as acetyl phosphate) and the total concentrations of sensor protein ($X_7 = X_1 + X_3$) and regulator protein ($X_8 = X_2 + X_4$) are determined by external influences that are independent of changes within the system. Thus, these are defined as *independent variables*. By contrast, the concentrations of the sensor proteins (X_1 phosphorylated and X_3 unphosphorylated) and of the regulator proteins (X_2 phosphorylated and X_4 unphosphorylated) are determined by the values of the independent variables and by the system's internal dynamical behaviour. These variables are defined as *dependent state variables* in our models.

Table 3. Sensor proteins observed and predicted, on the basis of homology modelling techniques, to be either monofunctional or bifunctional.

Sensor name	Functionality		Reference
	Predicted	Observed	
CheA (*E. coli*)	Mono	Mono	Ninfa *et al.* (1991)
EnvZ (*E. coli*)	Bi	Bi	Kanamaru *et al.* (1989)
KdpD (*E. coli*)	Bi	Bi	Jung *et al.* (1997)
VanSA (*E. faecalis*)	Bi	Bi [a]	Wright *et al.* (1993)
PhoR (*B. subtilis*)	Bi	Bi [b]	Shi *et al.* (1999)

a. Inferred from gene expression results.
b. Shows an increase in the rate of dephosphorylation only for the soluble domain of PhoR in the presence of ADP or ATP.

Protein synthesis and degradation occur on a timescale that is much slower than that of phosphorylation, phosphotransfer and dephosphorylation in the TCS modules. Thus, on the time scale of interest here, one can ignore gene regulation and consider the total amount of sensor protein (S_{Total}, X_7) and the total amount of regulator protein (R_{Total}, X_8) to be conserved quantities (i.e. $X_1 + X_3 = X_7 =$ constant *and* $X_2 + X_4 = X_8 =$ constant).

Thus, when considering changes in the dependent variables, in this and the following section, we need only emphasize the phosphorylated forms of the sensor protein X_1 and the regulator protein X_2. The corresponding changes in the unphosphorylated forms, X_3 and X_4, have the same magnitude but opposite sign. If for any reason, one of the X species in one of the conserved pairs increases by some amount, then the other X species of the pair must decrease by exactly the same amount. For example, amplification of signals at the level of the unphosphorylated proteins is equal to the negative of that at the level of the phosphorylated proteins. This amplification can be measured by the logarithmic gain $L(X_k,X_j)$ (see *Experimental procedures* section), which is defined as the percentage change in a dependent variable X_k in

response to a one per cent change in an independent variable X_j. Thus, $L(X_3,X_j) = -L(X_1,X_j)$ and $L(X_4,X_j) = -L(X_2,X_j)$, where X_j is any independent variable of the model. Similarly, the parameter sensitivities in the levels of the unphosphorylated proteins in response to parameter fluctuations are equal to the negative of those in the levels of the phosphorylated proteins. These parameter sensitivities can be measured by the expression $S(X_k,p_j)$ (see *Experimental procedures* section), which is defined as the percentage change in a dependent variable X_k in response to a one per cent change in a parameter p_j. Thus, $S(X_3,p_j) = -S(X_1,p_j)$ and, $S(X_4,p_j) = -S(X_2,p_j)$ where p_j is any parameter of the model.

The kinetic behaviour of the models in Fig. 1 can be described by a system of differential equations [Equations (1) (2) and (2')], as outlined in the *Experimental procedures* section. Solving these equations allows us to quantify and to compare the systemic properties of the alternative designs shown in Fig. 1 based on the following functional considerations.

A signal transduction cascade should have a set of large logarithmic gains to amplify physiological signals and another set of small logarithmic gains that attenuate pathological noise. The cascade should be robust, i.e. it should function reproducibly despite perturbations in the values of the parameters that define the structure of the system. This is, by definition, equivalent to saying that the parameter sensitivities should, in general, be as low as possible. The steady state of the system should be stable and have a sufficient margin of stability, such that it will not become unstable when subjected to random fluctuations in the parameters of the system. If the margin of stability is small, a small change in a parameter of the system (e.g. ionic strength of the medium) may destroy the possibility of a stable steady state and make the system dysfunctional. Finally, the system should respond quickly to changes in its environment because otherwise the system is unlikely to be competitive in rapidly changing environments.

These properties, for each of the alternative models, are quantified and then compared by taking the ratio of the value for a property in the reference system (bifunctional sensor) to the corresponding value in the alternative system (monofunctional sensor). As we have analytical expressions for the steady-state properties, we can determine in some cases whether the ratio is always equal to one, less than one, or greater than one, independent of parameter values. With this approach, we have obtained the following qualitative results for steady-state concentrations and the amplification factors.

The concentrations (and rates of change) of the corresponding state variables in the two models can always be the same (see *Calculating the constraints for external equivalence*). Changes in the secondary signal X_6 are amplified less in each of the corresponding state variables [i.e. each $L(X_i,X_6)$ for $i = 1, 2, 3$ and 4] with the bifunctional design. Similarly, these changes are amplified less in the flux through each of the sensor pools [i.e. $L(V_i,X_6)$ for $i = 1$ and 3] with the bifunctional design. This is shown in Table 4 by the ratio E, which is always less than 1. On the other hand, amplification of the phosphorylated regulator X_2 in response to a percentage change in the primary input signal X_5 [i.e. $L(X_2,X_5)$] is always greater with the bifunctional design. This is shown in Table 4 by the ratio B, which is always greater than 1.0. In all other cases, the differences between the amplification factors of the alternative designs are dependent upon the specific numerical values of the parameters and can not be determined analyticaly.

Quantitative functional differences

Although some functional differences between TCS with the alternative sensor designs are analyticaly indeterminate, numerical comparisons can be used to determine quantitative differences in function, and thus to establish statistical tendencies in the differences. With this approach we have obtained the following quantitative results for signal amplification, robustness, margin of stability, response time, and phosphorylation/dephosphorylation ratio.

Numerical results for signal amplification of concentrations are shown in Fig. 5 and described in detail in the next paragraph. The data are represented as Density of Ratios plots for moving medians (as described in *Experimental procedures*). The moving median of the ratio for the bifunctional design (reference) to monofunctional design (alternative), shown on the vertical axis, is a function of the moving median of the amplification for the bifunctional design, shown on the horizontal axis. [In symbolic terms, $<L(X_i,X_j)_{Bi}/L(X_i,X_j)_{Mono}>$ is plotted as a function of $<L(X_i,X_j)_{Bi}>$, where X_j indicates an independent variable, X_i a dependent variable and the angular brackets indicate averages.

The pattern of responses exhibited by the sensor protein is as follows. Overall, values for the amplification of the phosphorylated sensor signal X_1 in response to a percentage change in the primary input signal X_5 [$L(X_1,X_5)$] are similar in both designs. This is shown by the curve in Fig. 5A, which remains about 1.0. Values for the amplification of the phosphorylated sensor signal X_1 in response to a percentage change in the secondary input signal X_6 [$L(X_1,X_6)$] are also smaller in the bifunctional design (the curve remains below 1.0 in Fig. 5B). On the other hand, values for the amplification of the phosphorylated sensor signal X_1 in response to a percentage change in either the total concentration of sensor protein X_7 [$L(X_1,X_7)$] or the total concentration of regulator protein X_8 [$L(X_1,X_8)$] are similar for the two designs (curves

Table 4. Qualitative ratios of corresponding logarithmic gains for the reference system relative to the alternative system.

Systemic property	Dependent variable of the system			
	X1	X_2	V_1	V_2
$L(\bullet, X_5)$	A	B > 1	C	D
$L(\bullet, X_6)$	E < 1	E < 1	E < 1	F
$L(\bullet, X_7)$	G	H	J	K
$L(\bullet, X_8)$	L	M	N	P

$$Ratio = L(\bullet,\bullet)_{Bi} \,/\, L(\bullet,\bullet)_{Mono}$$

$$A = 1 + \frac{-g_{26}h_{12}h_{21}}{g_{15}g_{21}(g_{22}-h_{22}) + g_{26}\left((g_{11}-h_{11})(h_{22}-g_{22}) + h_{21}h_{12}\right)} <> 1$$

$$B = 1 + \frac{g_{26}h_{21}(g_{11}-h_{11})}{g_{21}\left(g_{15}(g_{21}-h_{21}) - g_{26}(g_{11}-h_{11})\right)} > 1$$

$$C = 1 + \frac{g_{26}g_{11}h_{12}h_{21}}{\begin{pmatrix}(g_{21}h_{12}-g_{22}h_{11})(g_{15}g_{21}-g_{26}(g_{11}-h_{11})) - g_{26}h_{11}h_{12}h_{21}\\ -g_{15}g_{21}(h_{12}h_{21}-h_{11}h_{22}) - h_{11}h_{22}(g_{26}g_{11}-h_{11})\end{pmatrix}} <> 1$$

$$D = 1 + \frac{g_{26}h_{21}(g_{22}(g_{11}-h_{11}) + g_{21}h_{12})}{g_{21}(h_{22}(g_{15}g_{21}-g_{26}(g_{11}-h_{11})) - h_{21}(g_{15}g_{22}+g_{26}h_{12}))} <> 1$$

$$E = 1 + \frac{g_{15}h_{21}}{g_{15}(g_{21}-h_{21}) - g_{26}(g_{11}-h_{11})} < 1$$

$$F = 1 + \frac{g_{15}h_{21}(g_{22}(g_{11}-h_{11}) + g_{21}h_{12})}{h_{21}(h_{11}-g_{11})(g_{15}g_{22}+g_{26}h_{12}) - h_{11}h_{22}(g_{15}g_{21}+g_{26}h_{11}) + g_{11}h_{22}(g_{15}g_{21}-g_{26}(g_{11}-2h_{11}))} <> 1$$

$$G = 1 + \frac{-h_{12}(g_{17}g_{26}h_{21} + h_{27}(g_{15}g_{21}+g_{26}(h_{11}-g_{11})))}{g_{17}(g_{15}g_{21}(g_{22}-h_{22}) + g_{26}((h_{11}-g_{11})(g_{22}-h_{22}) + h_{12}h_{21}))} <> 1$$

$$H = 1 + \frac{g_{17}g_{26}h_{21}(g_{11}-h_{11}) + h_{27}(g_{11}(g_{15}g_{21}-g_{11}g_{26}) + h_{11}(-g_{15}g_{21}+g_{26}(2g_{11}-h_{11})))}{g_{17}g_{21}(g_{15}g_{21}-g_{11}g_{26}+g_{26}h_{11}-g_{15}h_{21})} <> 1$$

$$J = 1 + \frac{\left(g_{17}g_{26}h_{21}g_{11}h_{12} + g_{11}h_{12}h_{27}(g_{15}g_{21}-g_{26}(g_{11}-h_{11}))\right)}{\begin{pmatrix}g_{17}((g_{21}h_{12}-g_{22}h_{11})(g_{15}g_{21}-g_{26}(g_{11}-h_{11}))) - g_{17}g_{26}(h_{21}h_{11}h_{12}\\ -g_{15}g_{17}g_{21}(h_{11}h_{22}-h_{12}h_{21}) + h_{11}h_{22}(g_{11}-h_{11}))\end{pmatrix}} <> 1$$

$$K = 1 + \frac{\left(\left(g_{11}g_{22}-g_{22}h_{11}+g_{21}h_{12}\right)\left(g_{17}g_{26}h_{21}+\left(g_{15}g_{21}+g_{26}(h_{11}-g_{11})\right)\right)\right)}{g_{17}g_{21}(g_{15}g_{21}h_{22}+g_{26}h_{11}h_{22}-g_{15}g_{22}h_{21}-g_{26}h_{12}h_{21}-g_{11}g_{26}h_{22})} <> 1$$

$$L = 1 + \frac{-h_{12}h_{21}(g_{15}g_{28}+g_{26}h_{18})}{\begin{pmatrix}g_{15}(g_{21}g_{22}h_{18}-g_{21}g_{28}h_{12}+g_{28}h_{12}h_{21}-g_{21}h_{18}h_{22})\\ +g_{26}(g_{11}g_{28}h_{12}-g_{28}h_{11}h_{12}-g_{11}g_{22}h_{18}+g_{22}h_{11}h_{18})\\ +g_{26}(h_{12}h_{18}h_{21}+g_{11}h_{18}h_{22}-h_{11}h_{18}h_{22})\end{pmatrix}} <> 1$$

$$M = 1 + \frac{h_{21}\left((g_{15}g_{28}+g_{26}h_{18})(g_{11}-h_{11})\right)}{(g_{28}(g_{11}-h_{11})+g_{21}h_{18})(g_{15}(g_{21}-h_{21})-g_{26}(g_{11}-h_{11}))} <> 1$$

$$N = 1 + \frac{g_{11}h_{12}h_{21}(g_{15}g_{28}+g_{26}h_{18})}{\begin{pmatrix}(g_{28}h_{12}g_{11}+g_{11}h_{18}h_{22})(g_{15}g_{21}-g_{26}(g_{11}-h_{11}))\\ -g_{21}h_{18}g_{11}g_{15}g_{22}-g_{28}h_{21}g_{15}g_{11}h_{12}\\ +g_{26}h_{18}(g_{11}^2g_{22}-h_{11}g_{11}g_{22})-h_{18}h_{21}g_{11}g_{26}h_{12}\end{pmatrix}} <> 1$$

$$P = 1 + \frac{h_{21}\left((g_{15}g_{28}+g_{26}h_{18})(g_{22}h_{11}-g_{11}g_{22}-g_{21}h_{12})\right)}{\left(g_{11}g_{28}-g_{28}h_{11}+g_{21}h_{18}\left\{\begin{matrix}g_{15}g_{22}h_{21}+g_{26}h_{12}h_{21}-\\ g_{15}g_{21}h_{22}+g_{11}g_{26}h_{22}-g_{26}h_{11}h_{22}\end{matrix}\right\}\right)} <> 1$$

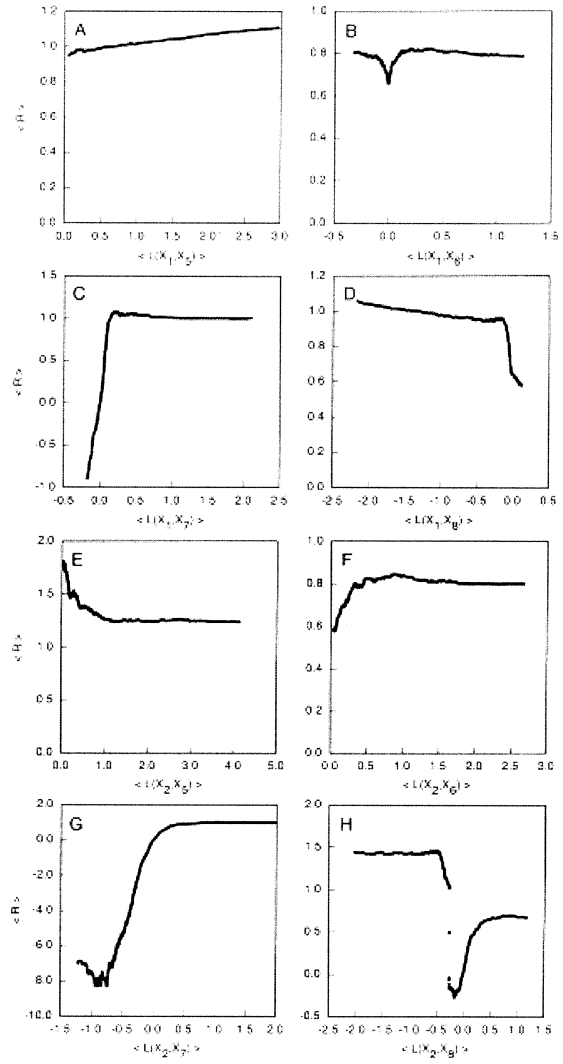

Fig. 5. Comparison of signal amplification factors for concentrations in two-component systems with either a bifunctional or a monofunctional sensor. The logarithmic gains are compared in a density of ratios plot using moving medians (see *Numerical analysis*). On the *x*-axis are average values of a given logarithmic gain for the bifunctional design. On the *y*-axis are average values for the ratio of that logarithmic gain in the bifunctional design (Fig. 1A) over the corresponding logarithmic gain in the monofunctional design (Fig. 1B).
A. Logarithmic gain in phosphorylated sensor protein X_1 with respect to changes in the primary input signal X_5.
B. Logarithmic gain in X_1 with respect to changes in the secondary input signal X_6.
C. Logarithmic gain in X_1 with respect to changes in the total amount of sensor, X_7.
D. Logarithmic gain in X_1 with respect to changes in the total amount of regulator, X_8.
E. Logarithmic gain in phosphorylated regulator protein X_2 with respect to changes in the primary input signal X_5.
F. Logarithmic gain in X_2 with respect to changes in the secondary input signal X_6.
G. Logarithmic gain in X_2 with respect to changes in the total amount of sensor, X_7.
H. Logarithmic gain in X_2 with respect to changes in the total amount of regulator, X_8.

remain very close to the *y*-axis value of 1.0 in Fig. 5C and D), except when signal amplification values are close to zero.

The regulator protein exhibits a different pattern of responses. Values for the amplification of the phosphorylated regulator signal X_2 in response to a percentage change in the primary input signal X_5 [$L(X_2,X_5)$] are greater for the bifunctional design as can be seen by the curve in Fig. 5E, which is always above 1. For low median values of the gain $L(X_2,X_5)$ in the bifunctional design, the differences in gain can be as large as 100%. For high median values, the differences are around 30%, with the gain in the bifunctional design being higher. This means that, for example, the bifunctional design will provide for larger changes in gene expression than would a monofunctional design for the same amount of osmotic pressure change in the EnvZ/OmpR system. Values for the amplification of the phosphorylated regulator signal X_2 in response to a percentage change in the secondary input signal X_6 [$L(X_2,X_6)$] are smaller in the bifunctional design (Fig. 5F). On the other hand, values for the amplification of the phosphorylated regulator signal X_2 in response to a percentage change in the total concentration of sensor protein X_7 [$L(X_2,X_7)$] can be larger in either alternative depending on the parameter values (Fig. 5G). If the logarithmic gain $L(X_2,X_7)$ in the bifunctional design is negative, then the amplification in the bifunctional design is smaller (in absolute value) as can be seen in Fig. 5G. If the logarithmic gain in the bifunctional design is positive, then the amplification in the monofunctional design is smaller. Values for the amplification of the phosphorylated regulator signal X_2 in response to a percentage change in the total concentration of regulator protein X_8 [$L(X_2,X_8)$] also can be larger in either alternative depending on the parameter values (Fig. 5H). If the logarithmic gain $L(X_2,X_8)$ in the bifunctional design is negative, then the amplification in the bifunctional design is larger (in absolute value); if the logarithmic gain in the bifunctional design is positive, then the amplification in the monofunctional design is larger. Even though it seems that this shift occurs at values of about -0.3 for $L(X_2,X_8)$ in this case, the actual values are about zero. Because of the moving averaging technique there are residual ratios that keep the median above 1.0. When the values for $L(X_i,X_i)$ change sign, the curves tend to exhibit a dip, which is a consequence of the moving average technique when positive and negative values are being averaged.

Numerical results for signal amplification in flux are shown in Fig. 6. The pattern of responses exhibited by flux through the pools of sensor protein is as follows. Values for the amplification of the flux V_1 in response to a percentage change in the primary input signal X_5 [$L(V_1,X_5)$] can be larger in either alternative depending on the parameter values (Fig. 6A). If the logarithmic gain $L(V_1,X_5)$

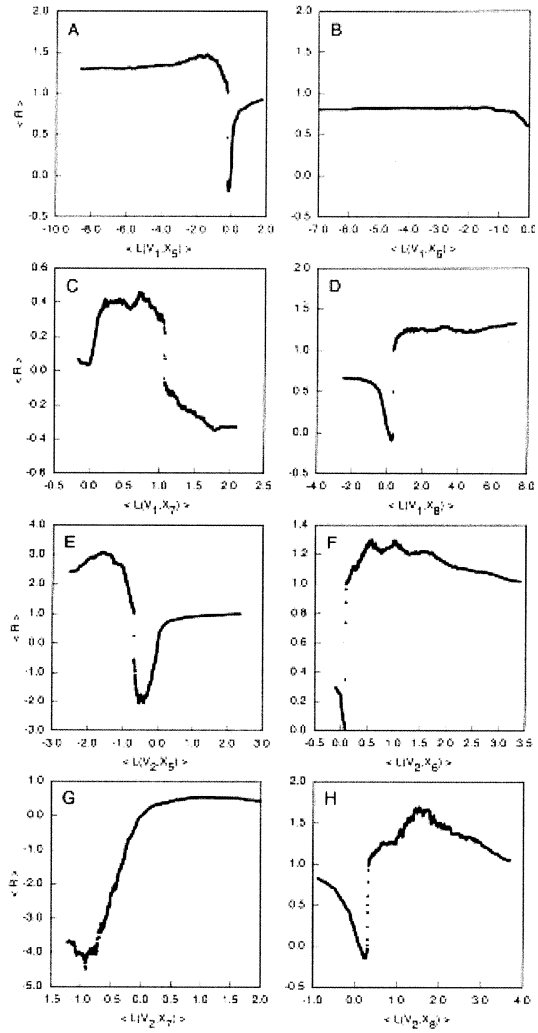

Fig. 6. Comparison of signal amplification factors for fluxes in two-component systems with either a bifunctional or a monofunctional sensor. The logarithmic gains are compared in a density of ratios plot using moving medians (see *Numerical analysis*). On the *x*-axis are average values of a given logarithmic gain in the bifunctional design. On the *y*-axis are average values for the ratio of that logarithmic gain in the bifunctional design (Fig. 1A) over the corresponding logarithmic gain in the monofunctional design (Fig. 1B).
A. Logarithmic gain in flux through the pool of phosphorylated sensor protein V_1 with respect to changes in the primary input signal X_5.
B. Logarithmic gain in V_1 with respect to changes in the secondary input signal X_6.
C. Logarithmic gain in V_1 with respect to changes in the total amount of sensor, X_7.
D. Logarithmic gain in V_1 with respect to changes in the total amount of regulator, X_8.
E. Logarithmic gain in flux through the pool of phosphorylated regulator protein V_2 with respect to changes in the primary input signal X_5.
F. Logarithmic gain in V_2 with respect to changes in the secondary input signal X_6.
G. Logarithmic gain in V_2 with respect to changes in the total amount of sensor, X_7.
H. Logarithmic gain in V_2 with respect to changes in the total amount of regulator, X_8.

in the bifunctional design is negative, then the amplification in the bifunctional design is larger (in absolute value); if the logarithmic gain in the bifunctional design is positive, then the amplification in the monofunctional design is larger. Values for the amplification of the flux V_1 in response to a percentage change in the secondary input signal X_6 [$L(V_1,X_6)$] are smaller in the bifunctional design (Fig. 6B). Values for the amplification of the flux V_1 in response to a percentage change in the total concentration of sensor protein X_7 [$L(V_1,X_7)$] in the monofunctional design are always larger in absolute value (Fig. 6C). Values for the amplification of the flux V_1 in response to a percentage change in the total concentration of regulator protein X_8 [$L(V_1,X_8)$] can be larger in either alternative depending on the parameter values (Fig. 6D). If the logarithmic gain $L(V_1,X_8)$ in the bifunctional design is negative, then the amplification in the monofunctional design is larger (in absolute value); if the logarithmic gain in the bifunctional design is positive, then the amplification in the bifunctional design is larger.

The pattern of responses exhibited by the flux through the pools of regulator protein is as follows. Values for the amplification of the flux V_2 in response to a percentage change in the primary input signal X_5 [$L(V_2,X_5)$] can be larger in either alternative depending on the parameter values (Fig. 6E). If the logarithmic gain $L(V_2,X_5)$ in the bifunctional design is negative, then the amplification in the bifunctional design is larger (in absolute value); if the logarithmic gain in the bifunctional design is positive, then the amplification in the monofunctional design is larger. Even though it seems that this shift occurs at values of about −0.8 for $L(V_2,X_5)$ in this case, the actual values are about zero. Because of the moving averaging technique there are residual ratios that keep the median above 1.0. Values for the amplification of the flux V_2 in response to a percentage change in the secondary input signal X_6 [$L(V_2,X_6)$] can be larger in either alternative depending on the parameter values (Fig. 6F). If the logarithmic gain $L(V_2,X_6)$ in the bifunctional design is negative, then the amplification in the monofunctional design is larger (in absolute value); if the logarithmic gain in the bifunctional design is positive, then the amplification in the bifunctional design is larger. Values for the amplification of the flux V_2 in response to a percentage change in the total concentration of sensor protein X_7 [$L(V_2,X_7)$] is on average larger in the monofunctional design (Fig. 6G). Values for the amplification of the flux V_2 in response to a percentage change in the total concentration of regulator protein X_8 [$L(V_2,X_8)$] can be larger in either alternative depending on the parameter values (Fig. 6H). If the logarithmic gain $L(V_2,X_8)$ in the bifunctional design is negative, then the amplification in the monofunctional design is larger (in absolute value); if the logarithmic gain in the bifunctional design is positive, then the amplification in the bifunctional

design is larger. Even though it seems that this shift occurs at values of about 0.5 for $L(V_2,X_8)$ in this case, the actual values are about zero. Again, because of the moving averaging technique there are residual ratios that keep the median above 1.0. When the values for $L(V_i,X_j)$ change sign, the curves tend to exhibit a dip, which is a consequence of the moving average technique when positive and negative values are being averaged. The dip for the logarithmic gains in flux tends to be more pronounced than for the logarithmic gains in concentration.

Numerical results for robustness of the alternative designs are shown in Fig. 7. The data for aggregate parameter sensitivities are represented as Density of Ratios plots for moving medians. The moving median of the ratio for the bifunctional design (reference) to monofunctional design (alternative) is on the vertical axis, and the moving median of aggregate parameter sensitivity for the bifunctional design is on the horizontal axis. [In symbolic terms, $<S(X_i)_{Bi}/S(X_i)_{Mono}>$ as a function of $<S(X_i)_{Bi}>$, or $<S(V_i)_{Bi}/S(V_i)_{Mono}>$ as a function of $<S(V_i)_{Bi}>$, where X_i indicates a dependent concentration, V_i a dependent flux and the angular brackets indicate averages.] On average, the aggregate sensitivities of the sensor signals (Fig. 7A and C) are less in the bifunctional design than in the monofunctional design when the values for the parameter sensitivities are low, whereas the aggregate sensitivities of the regulator signals (Fig. 7B and D) are greater in the bifunctional design than in the monofunctional design under these conditions. The aggregate sensitivities of all the signals are lower in the monofunctional design when the values for the parameter sensitivities are high, which is the less physiologically relevant case. However, in all cases, the average differences in corresponding sensitivities between the monofunctional and the bifunctional system are very small. In the case of the sensor signals, this difference is negligible and the median ratio is for all practical purposes 1.0.

Numerical results for the stability margins of the alternative designs are shown in Fig. 8. These magnitudes, which correspond to the two critical Routh Criteria for stability, provide a measurement for the amount of perturbation that the system will tolerate before the steady state becomes unstable (see the *Experimental procedure* section). The bifunctional design has a larger margin of stability with respect to the first of the critical Routh criteria when this margin is small, which is when this margin is most important. When this margin becomes large, and its value becomes less important, the two designs have essentially the same value (Fig. 8A). The bifunctional design also has a larger margin of stability with respect to the second of the critical Routh criteria (Fig. 8B).

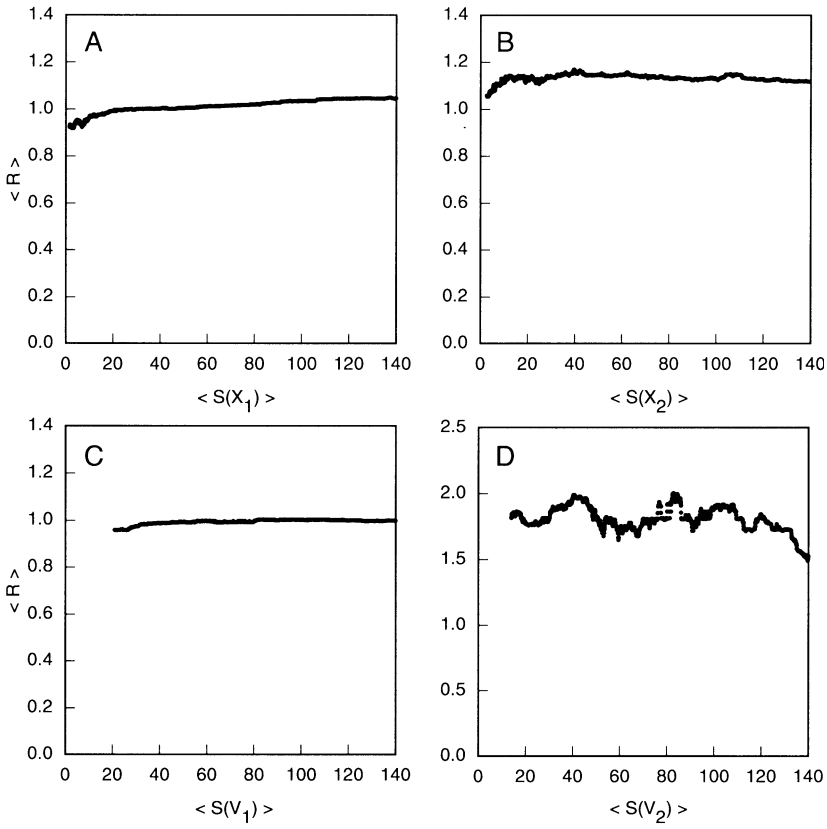

Fig. 7. Comparison of robustness for two-component systems with either a bifunctional or a monofunctional sensor. The aggregate parameter sensitivities (see *Steady-state solution and key systemic properties*) are compared in a density of ratios plot using moving medians. On the *x*-axis are average values of a given aggregate sensitivity in the bifunctional design. On the *y*-axis are average values for the ratio of that aggregate sensitivity in the bifunctional design (Fig. 1A) over the corresponding aggregate sensitivity in the monofunctional design (Fig. 1B).
A. Aggregate sensitivity of phosphorylated sensor protein X_1.
B. Aggregate sensitivity of phosphorylated regulator protein X_2.
C. Aggregate sensitivity of flux through the pool of phosphorylated sensor protein V_1.
D. Aggregate sensitivity of flux through the pool of phosphorylated regulator protein V_2.

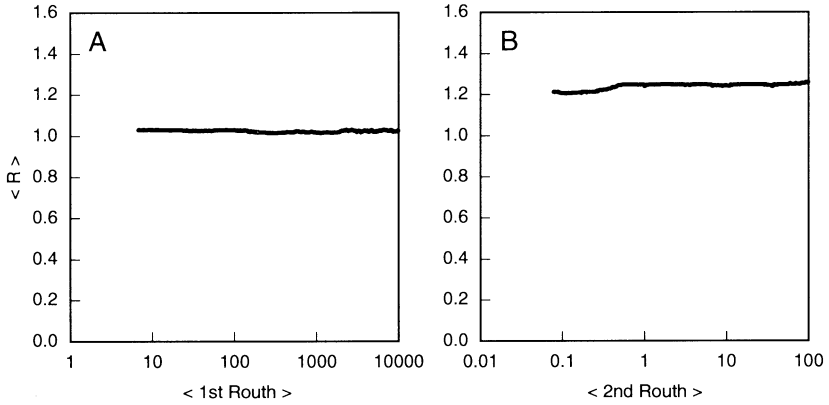

Fig. 8. Comparison of stability margins for two-component systems with either a bifunctional or a monofunctional sensor. The stability margins determined by the two critical Routh criteria for local stability (see *Steady-state solution and key systemic properties*) are compared in a density of ratios plot using moving medians. On the *x*-axis are average values of a given stability margin in the bifunctional design. On the *y*-axis are average values for the ratio of that stability margin in the bifunctional design (Fig. 1A) over the corresponding stability margin in the monofunctional design (Fig. 1B).
A. First critical Routh criterion.
B. Second critical Routh criterion.

Numerical results for the temporal responsiveness of the alternative designs are shown in Fig. 9. The response time, τ, is defined as time required for the return to a steady state following a perturbation (see the *Experimental procedures* section). The data are represented as Density of Ratios plots for the raw data (Fig. 9A) and for moving medians (Fig. 9B). The ratio for the bifunctional design (reference) to monofunctional design (alternative) is on the vertical axis, and the response time for the bifunctional design is on the horizontal axis. [In symbolic terms, $<\tau_{Bi}/\tau_{mono}>$ as a function of $<\tau_{Bi}>$, where τ is the response time and the angular brackets indicate averages.] On average, the response time is slightly less for the bifunctional design when response times are small, but the differences between designs become insignificant as the response times become larger.

In this work we allowed the sensor and regulator each to have steady-state operating levels of phosphorylation between 0% and 100%. In practice, the random values for the ratios $f_{31} = -X_{10}/X_{30}$ and $f_{42} = -X_{20}/X_{40}$ in our ensemble range between -0.0000001 and -999. The amplification properties of the TCS with a bifunctional design are not influenced to any large extent by the steady-state operating value for phosphorylation of either the sensor or the regulator (data not shown), although more significant changes can be seen in some cases when the operating value for phosphorylation drops to near 0%.

Discussion

In this work we examined alternative designs for the sensor proteins of prototype TCS. A bifunctional sensor is characterized by two functions: (i) when phosphorylated, the sensor transfers its phosphate group to the response regulator; (ii) when unphosphorylated the sensor increases the dephosphorylation rate of the response regulator. A monofunctional sensor has only the first of these functions. Our results have identified both structural and functional attributes of these alternative designs.

Structural differences

The results of our homology modelling show a highly conserved structural feature ('ATP lid') that appears to be distinctive for each of the alternative designs (Figs 3 and 4). Whether or not the characteristic ATP lid is responsible

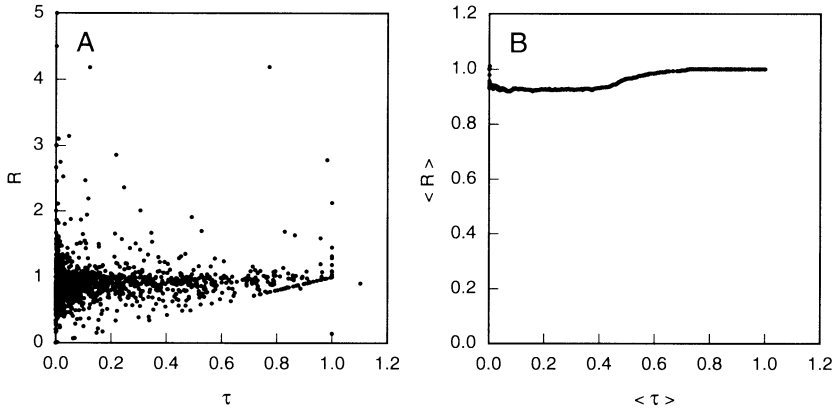

Fig. 9. Comparison of response times for two-component systems with either a bifunctional or a monofunctional sensor. The response times (see *Steady-state solution and key systemic properties*), τ, are compared in a density of ratios plot. On the *x*-axis are values of response times for the bifunctional design. On the *y*-axis are values for the ratio of response times for the bifunctional design (Fig. 1A) over the response times for the monofunctional design (Fig. 1B).
A. Raw data.
B. Average values in a moving median plot.

for the bifunctionality of a sensor is unknown, but recent experiments suggest that this part of the sensor has a role in enhancing the rate of dephosphorylation of the response regulator for EnvZ/OmpR (Zhu *et al*., 2000) and for NRII-PII/NRI (Pioszak and Ninfa, 2003). An experiment that exchanges the ATP lid of the EnvZ and CheA sensor proteins and assays the resulting proteins for bifunctional behaviour would help to resolve this issue. Whether this structural feature is responsible for the bifunctionality of a sensor or not, our rigorous comparative analysis of physiological function demonstrates that there is a clear basis for selection of monofunctional and bifunctional sensor proteins.

Functional differences

Protein levels for the majority of TCS are regulated on a slow time-scale by mechanisms affecting transcription. Regulation by these mechanisms, which have been studied elsewhere (Hlavacek and Savageau, 1995, 1996, 1997), is beyond the scope of this article, which focuses on the more rapid time scale of regulation within the TCS. Nevertheless, the results for the logarithmic gains in concentrations (Fig. 5C, D, G and H) and fluxes (Fig. 6C, D, G and H) with respect to changes in the total concentration of sensor (X_7) and regulator (X_8) provide some insight regarding the influence of changing protein levels.

The mathematically controlled comparisons in our study show that monofunctional and bifunctional designs for signalling within TCS differ on the basis of several criteria for functional effectiveness. The model with a bifunctional sensor has higher signal amplification in response to changes in the primary signal (Fig. 5E), X_5, as represented by $L(X_2,X_5)$. It has lower signal amplification in response to changes in the secondary signal (Fig. 5F), X_6, as represented by $L(X_2,X_6)$. If X_5 is considered to be the physiological signal for a TCS, then this implies that the model with the bifunctional sensor is more effective in responding to this signal. Robustness, except in the case of V_2, tends to be similar when the values for the parameter sensitivities are low (Fig. 7). Although, robustness for all the variables is greater in the bifunctional design when the values for the parameter sensitivities are high, systems with high parameter sensitivities are less likely to be biologically significant. The margins of stability for the steady state, as measured by both Routh criteria, are larger for the bifunctional design (Fig. 8). Response times are similar for the alternative designs (Fig. 9). These functional differences have implications for cross-talk among TCS.

Cross-talk to and from a TCS

The specificity of sensors and regulators is not absolute. Regulator proteins can respond to signals other than those transmitted by their cognate sensor, and sensor proteins can transmit signals to destinations other than their cognate regulator. In some cases, this cross-talk might be undesirable noise that should be minimized. Sensors that are homologous to the cognate sensor but are involved in distinct physiological responses may represent such a case. In other cases, this cross-talk might represent the physiological co-ordination of several processes that needs to be enhanced. Chemotaxis represents such a case, where the state of cellular metabolism needs to be taken into account before the cell migrates towards nutrient sources. It is less important for a cell that is already well fed to spend energy in migrating towards nutrients than if the cell is starving. The results in the previous sections allow us to identify designs that are appropriate for dealing with cross-talk in each of these contexts.

Cross-talk to a TCS module is represented by any secondary input signal (Q_2) coming from sources other than the regulator's cognate sensor that causes a change in phosphorylation level of the module's response regulator (R^*). The schematic diagrams in Fig. 1 explicitly represent the case in which the secondary signal is a small phosphodonor like acetyl-phosphate, but not ATP. Cross-talk in this case is less amplified by the module with a bifunctional sensor (Fig. 1A) than by the module with the monofunctional sensor (Fig. 1B). Thus, the design with a bifunctional sensor is better at attenuating the cross-talk to the module, and this is appropriate when the cross-talk is physiologically undesirable. Conversely, the design with a monofunctional sensor is better at amplifying the cross-talk to the module, and this is appropriate when the cross-talk is a relevant physiological signal. A TCS with a design that could change from bifunctional, when only its primary signal conveyed physiologically relevant information, to monofunctional, when other signals should also be considered, would have an advantage in dealing with more complex situations in which there is a changing requirement for suppression or integration of secondary signals. (The NRI/NRII system of *E. coli*, which will be discussed below, may be one such example.)

Cross-talk to the module also can result from secondary input signals originating from other sensors. There are several formal possibilities shown schematically in Fig. 10. The analysis of these possibilities yields results similar to those already described (data not shown). A controlled comparison of the alternatives in Fig. 10A and B shows that a bifunctional design for the cognate sensor (S_1) is better at enhancing amplification of the regulator (R) response to the primary input signal (Q_1) while it is better at suppressing noise represented by the secondary input signal (Q_2). A controlled comparison of the alternatives in Fig. 10C and D shows a similar result. Thus, regardless

of the design for the non-cognate sensor, a bifunctional design for the cognate sensor results in better amplification of the primary input signal and better suppression of the secondary input signal. A controlled comparison of the alternatives in Fig. 10B and D shows that a bifunctional design for the non-cognate sensor (S_2) results in better amplification of the regulator (R) response to the secondary input signal (Q_2). Taken together, these results suggest that the design in Fig. 10C is preferred for enhancing amplification of the primary input signal while promoting attenuation of the secondary input signal. However, the design in Fig. 10A is preferred for enhanced cross-talk to achieve a more balanced integration of the two input signals.

Cross-talk from the module occurs when the cognate sensor transmits its signal to a non-cognate regulator protein. Again, there are several formal possibilities, shown schematically in Fig. 11, that we have analysed (data not shown). A controlled comparison of the alternatives in Fig. 11A and B shows that the design with a sensor that is bifunctional with respect to its cognate regulator (R_1) is better in two respects. It is better at amplifying the signal that is transmitted from the input Q to the primary output (T_1) and better at suppressing the signal that is transmitted to the secondary output (T_2). A controlled comparison of the alternatives in Fig. 11C and D shows a similar result. Thus, regardless of the design with respect to the non-cognate regulator, the design with a sensor that is bifunctional with respect to its cognate regulator is better at amplifying the signal transmitted to the primary output and at suppressing that to the secondary output. A controlled comparison of the alternatives in Fig. 11B and D shows that the design with a sensor that is bifunctional with respect to the non-cognate regulator (R_2) is better at amplifying the signal transmitted from the input Q to the secondary output (T_2). Taken together, these results suggest that the design in Fig. 11C is more effective in suppressing cross-talk from the module to the response regulators of other TCS. However, the design in Fig. 11A is more effective in enhanced cross-talk to achieve a more balanced set of response in both regulators.

Examples

The results in the previous sections suggest a rationale for selection of the two alternative sensor designs based on the physiology of the system in which the TCS is embedded. Assume that the output of the module is the phosphorylation level of the response regulator (X_2). Phosphorylation levels change as a response to changes in the input signals X_5 and X_6. Usually, X_5 is thought of as the input signal and X_6 is not considered. However, X_6 is also an input signal because it changes the phosphorylation level of the response regulator. Thus, we can consider the TCS in Fig. 1 as integrators of two signals, X_5 and X_6.

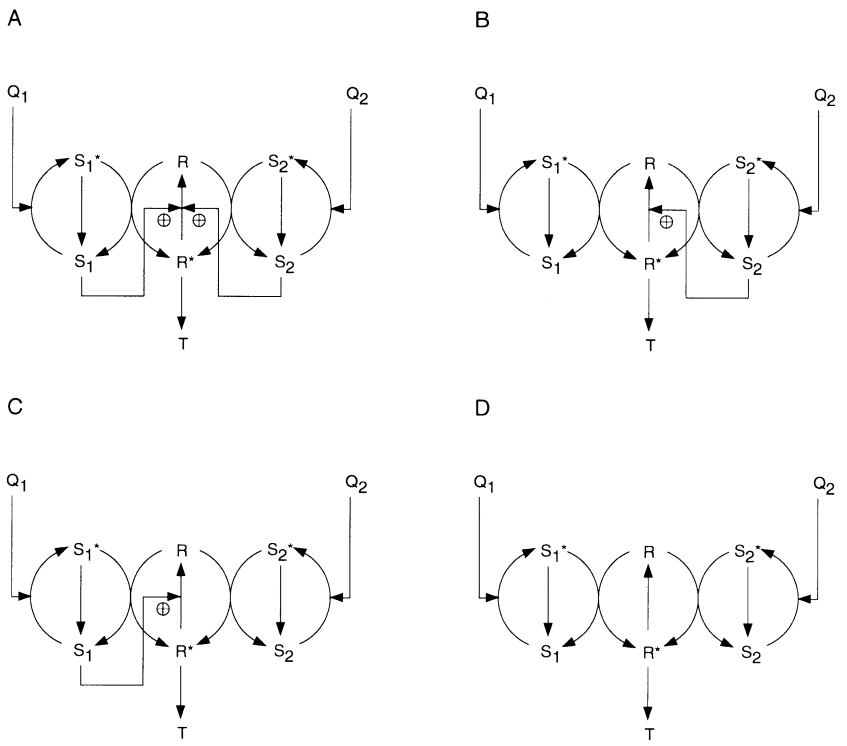

Fig. 10. Cross-talk to a common response regulator from two distinct sensor proteins.
A. Both sensors are bifunctional with respect to the common response regulator.
B and C. One sensor is bifunctional and the other is monofunctional with respect to the common response regulator.
D. Both sensors are monofunctional with respect to the common response regulator. See text for discussion.

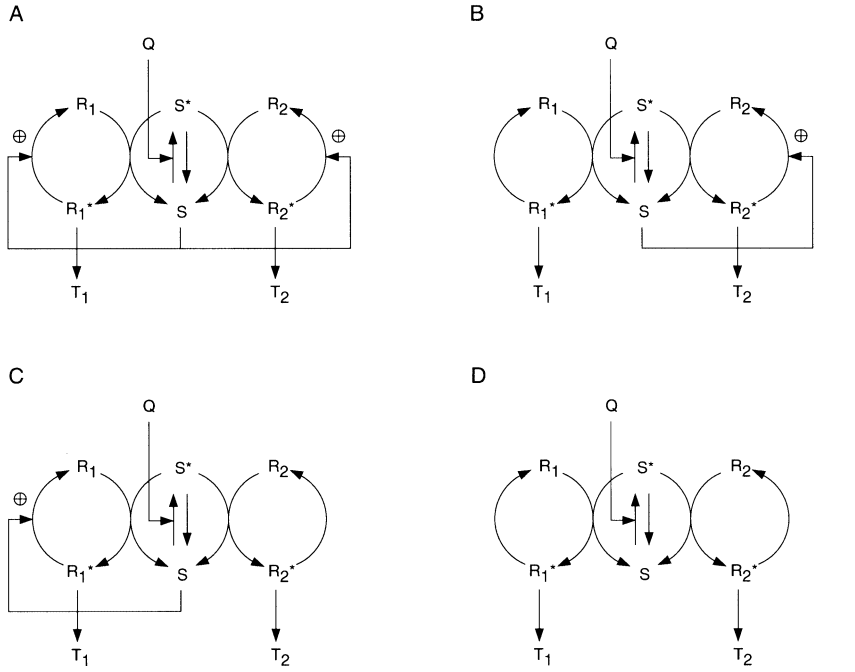

Fig. 11. Cross-talk from a common sensor protein to two distinct response regulator proteins.
A. The sensor is bifunctional with respect to both response regulators.
B and C. The sensor is bifunctional with respect to one response regulator and monofunctional with respect to the other.
D. The sensor is monofunctional with respect to both response regulators. See text for discussion.

When compared to the TCS with a monofunctional sensor, those with a bifunctional sensor maximize amplification of the signal X_5 and minimize amplification of the signal X_6. On the other hand, the TCS with a monofunctional sensor maximize X_6 amplification and minimize X_5 amplification when compared to otherwise equivalent TCS with a bifunctional sensor. Thus, in systems for which X_5 is the major signal and the influence of other signals (represented by X_6) needs to be minimized, the design with a bifunctional sensor should be selected; in systems for which the other signals need to be taken into account and integrated, the design with a monofunctional sensor should be selected. We will clarify these notions with a few examples.

The regulation of pore size in bacteria by changes in the osmolarity of the medium is mediated by a TCS with a bifunctional sensor. The pores in their cell membrane are composed of two different proteins. Subunit OmpF forms large pores, whereas OmpC forms smaller pores. EnvZ (X_1) is a membrane protein and the sensor for a TCS module. Changes in the osmolarity of the medium lead to changes in the EnvZ protein. In a high osmolarity medium, EnvZ increases its autophosphorylation rate. This in turn leads to a transfer of phosphate from EnvZ to the response regulator of the TCS, OmpR (X_2). OmpR is a transcription factor that binds DNA either in its phosphorylated (high affinity) or unphosphorylated (low affinity) form. In its phosphorylated form, OmpR dimerizes to increase its interaction with DNA. This leads to increased expression of OmpC and to decreased expression of OmpF (Pratt *et al*., 1996 for a review). The sensor in this system, EnvZ, is bifunctional (Igo *et al*., 1989), which can be rationalized in terms of our results as follows. Pore size should be determined exclusively by differences in osmolarity between the intracellular and the extracellular medium. If pore size were to be affected by other signals, such as changes in the levels of small phosphodonors, then osmotic balance could not be maintained and cell viability would be diminished. Thus, the bifunctional sensor design used in this TCS module is the one that maximizes amplification of changes in the osmotic pressure (X_5) and minimizes amplification of changes in other spurious signalling processes (X_6).

The VanS/VanR module in *Enterococus faecalis* is another example of a TCS with a bifunctional sensor (Wright *et al*., 1993, Arthur *et al*., 1997). This TCS regulates synthesis of proteins responsible for the organism's resistance to antibiotics. Again, as the function of the proteins is to confer resistance to antibiotics, it should not be activated by other signals, because this would unnecessarily increase the protein burden of the cell. Thus, selection of the design with a bifunctional sensor is to be expected.

Chemotaxis in *E. coli* is mediated by a TCS module with a monofunctional sensor (see Eisenbach, 1996; for a review). In this case, the CheA sensor protein (X_1) transfers its phosphate to either of two response regulators, CheY or CheB (X_2). CheB is responsible for desensitizing the cell to chemical gradients, whereas CheY is responsible for changing the rate of cell tumbling so as to pro-

mote movement towards favourable concentrations. CheA is a monofunctional sensor, which means that CheY and CheB can be more effectively phosphorylated by other sources (X_6), either by phospho-donors or by other sensors, than they could if CheA were a bifunctional sensor. Thus, the internal metabolism of the bacterium regulating the levels of these phospho-donors is more likely to be involved in determining whether the cell will search for nutrients than it would be if CheA were a bifunctional sensor. [It must be emphasized that the chemotaxis system in *E. coli* has a phosphatase protein, CheZ, that acts downstream of both response regulators. Another case where this situation also occurs is in the Spo0 phosphorelay in *B. subtilis*. These phosphatases are also the subject of regulation. However, in this work our goal has been to evaluate the effect of alternative sensor design on signal transmission *within* the prototype module and downstream aspects of design will not be considered here.]

FlbE/FlbD in *Caulobacter crescentus* is a presumptive TCS module involved in flagellum assembly and cell cycle regulation (e.g. Wingrove and Gober, 1996). We are not aware of studies showing whether FlbE is a bifunctional or monofunctional sensor, but based on its involvement in the cell cycle we would predict that it would have a monofunctional sensor as several signals must be integrated to co-ordinate the timing of cell cycle events. Similar arguments apply to the co-ordination of sporulation events in *Bacillus subtilis*. The *Spo* phosphorelay system appears to have a monofunctional sensor and to include different phosphatases that are specific for the different components of the relay (Perego and Hoch, 1996 for a review).

PhoR/PhoP is a TCS module with a monofunctional sensor involved in regulating expression of genes responsible for the transport of phosphate in *B. subtilis*. This module transduces signals generated by phosphate starvation (X_5). There is also cross-regulation between this module and PTS sugar systems (X_6) (Hulett, 1996; for a review). Thus, it is important for this TCS to sense other signals, beside the one coming from PhoR, and the design with a monofunctional sensor should be favoured. In fact, it has been shown that membrane-bound PhoR does not seem to influence the dephosphorylation rate of PhoB significantly (Shi *et al.*, 1999) and thus it borders on the design of a monofunctional sensor. However, from Table 3, the PhoR sensor is predicted to be bifunctional which seem to contradict this result. A careful analysis of Shi *et al.* (1999) explains the apparent contradiction. A soluble version of PhoR (without the membrane spanning domains) enhances the dephosphorylation rate of PhoP only slightly in the absence of ATP or ADP but much more significantly in the presence of either of these molecules. Thus, the design of soluble PhoR is that of a bifunctional sensor as predicted from the homology modelling; however, this bifunctionality is effectively inhibited and transformed into monofunctionality by locating PhoR in the membrane.

If one accepts the rule we have suggested for the selection of monofunctional and bifunctional sensors, and the supportive evidence in the above cases where the physiological context can be interpreted in a fairly straightforward fashion, then one can go on to apply this rule in the interpretation of more complex systems. Two such cases in *E. coli* are considered below, the NRI/NRII system involved in nitrogen fixation and the NarX/NarL and NarQ/NarP systems involved in nitrate and nitrite dependent gene expression.

The NRI/NRII system in *E. coli* regulates nitrogen fixation and glutamine production. NRII is the sensor protein that phosphorylates the response regulator NRI. Under conditions of low nitrogen availability NRI upregulates the expression of glutamine synthase. This enzyme condenses nitrogen and glutamate to form glutamine. Glutamine increases the affinity of a third protein, PII, towards the sensor protein NRII, inhibiting NRII phosphorylation and creating a complex that binds phosphorylated NRI and increases NRI′s rate of dephosphorylation. Alone, NRII has little or no effect upon the rate of NRI dephosphorylation (Keener and Kustu, 1988). Thus, under normal nitrogen conditions (i.e. with normal levels of glutamine), this TCS is bifunctional and nitrogen fixation is more sensitive to regulation by glutamine levels than it would be if the module were monofunctional. Under nitrogen depletion (causing a decrease in the concentration of glutamine) PII does not bind NRII, which then becomes monofunctional. This causes the module to integrate the signals coming from glutamine/glutamate levels with those coming from other parts of metabolism, via the changes in the concentration of acetyl-phosphate, more efficiently than it would if NRII were bifunctional. Intuitively, one might think that, under nitrogen depletion, NRII should be bifunctional, in order to more efficiently respond to the nitrogen fixation needs of the cell and buffer against regulation of this fixation by other parts of metabolism. However, probably as a result of the central role of glutamate in amino acid biosynthesis, this is not so. Glutamate is a ubiquitous amino acid that is needed for the biosynthesis of all other amino acids and not just glutamine. Under nitrogen depleting conditions, it is important that the use of glutamate be co-ordinately regulated by the concentration of all amino acids, in order not to deplete the cell of some of them, by its overuse to produce glutamine. It has been reported that cell growth on glucose minimal medium containing arginine, a poor nitrogen source, is greatly decreased in mutants lacking either NRII or phosphate acetyl-transferase (E.C. 2.3.1.8) (Feng *et al.*, 1992). This implies that phosphorylation by both NRII and acetyl-phosphate is important under these conditions, which agrees with a strong regulatory role for

secondary signals in nitrogen fixation that get integrated through the NRI/NRII module. It is clear that the NR system is complex and that a more detailed model would provide additional insight.

Two TCS modules are involved in nitrate and nitrite dependent gene expression in *E. coli*, NarX/NarL and NarQ/NarP. NarX and NarQ are sensor proteins (X_1) that independently recognize nitrate in the medium. Each of these sensors can transfer phosphate to both regulators NarL and NarP (X_2) which have different specificities (Schroder *et al*., 1994; Chiang *et al*., 1997). Whereas NarL regulates expression of genes encoding nitrate and fumarate reductases as well as nitrite exporter proteins, both NarL and NarP regulate expression of genes encoding nitrite reductase and anaerobically expressed proteins. NarQ and NarX are both sensors for NarL. NarQ has a higher affinity than NarX for phosphorylating NarL. However, NarX is more effective in enhancing the rate of NarL dephosphorylation (Schroder *et al*., 1994). In light of our results this can be interpreted as a design that maximizes the signals transduced via NarX/NarL with respect to those transduced via NarQ/NarL, because NarX is bifunctional with respect to NarL whereas NarQ is not. It would be interesting to determine whether either of the sensors is bifunctional with respect to NarP. If not, this would indicate that NarP, in contrast to NarL, is designed to integrate signals coming from both NarX and NarQ.

It is clear from the examples above, and others, that two-component systems exhibit a diversity of design issues that are not well understood. Our analysis has contributed to the elucidation of just one of the variations in design, monofunctional versus bifunctional sensors. Based on our results, we have proposed the following rule: Relative to the transduction of cognate signals, bifunctional sensors enhance suppression of signals from non-cognate sources whereas monofunctional sensors enhance their integration. The experimental evidence we have discussed is consistent with this rule. This leads us to suggest that this rule may be useful in understanding whether other TCS are acting simply as transducers of the sensor signal or as integrators of signals coming from different sources.

Experimental procedures

Structures for TCS

Crystal or NMR structures have been determined for a limited number of TCS sensor and regulator proteins. In the PDB database one can find structures for a prototype bifunctional sensor (EnvZ: PDB reference identifiers 1BXD and 1JOY) and for a prototype monofunctional sensor (CheA: PDB reference identifiers 1A0O, 1B3Q and 1FWP). Partial structures have been determined for other sensor proteins (ArcB: PDB reference identifiers 1A0B and 2A0B; FixL: PDB reference identifiers 1D06, 1DRM, 1DRQ and 1EW0) and other response regulators (NarL: PDB reference identifiers 1A04 and 1RNL; CheY: PDB reference identifiers 1A0O, 1BDJ, 1C4W, 1D4Z, 1DJM, 1EAY, 1EHC, 1VLZ, 2CHE, 2CHF, 2CHY, 5CHY and 6CHY; Spo0A: PDB reference identifiers 1DZ3, 1FC3 and 1QMP; Spo0F: PDB reference identifiers 1FSP, 1SRR and 2FSP; Spo0B: PDB reference identifiers 1IXM and NAT; OmpR: PDB reference identifiers 1ODD and 1OPC (DNA binding domain; no structure for the phosphorylation domain); CheB: PDB reference identifiers 1A2O and 1CHD; DrrD: PDB reference identifier 1KGS; NtrC: PDB reference identifiers 1NTRC, 1NTC, 1DC7 and 1DC8; Etr1: PDB reference identifier 1DCF; FixJ: PDB reference identifiers 1D5W, 1DBW, 1DCK and 1DCM; PhoB: PDB reference identifiers 1B00, 1GXP, 1GXQ and 1QQI; Rcp1: PDB reference identifiers 1I3C and 1JLK). This is a very small portion of the total number of TCS with identified proteins. However, the known structures can be used in combination with sequence data and homology modelling techniques to predict the structures for these other proteins of TCS.

Sequence data for proteins of TCS

We searched GeneBank and Swissprot for Two Component Systems proteins. We also looked for proteins of TCS in the databases for sequenced microbial genomes (listed in MAGPIE at http://www-fp.mcs.anl.gov/~gaasterland/genomes.html). Whenever necessary, we translated the cDNA sequence into its corresponding amino acid sequence. The protein sequences that we extracted from the databases were then catalogued by organism. We did homology modelling of each protein by submitting the sequence to SWISS-MODEL (Guex and Peitsch, 1997). Those that were successfully modelled (at least for some domain of the protein) by this program are presented in Tables 1 and 2.

Homology modelling

The protein sequences were submitted via internet to the SwissModel server (http://swissmodel.expasy.org/; Guex and Peitsch, 1997). As templates for the homology modelling we have used files from the PDB database (http://www.rcsb.org/pdb/ Berman *et al*., 2000). The server attempted to generate a three-dimensional model for each sequence and returned the results via E-mail. In many cases there was not enough homology to any known structure to create a working model. In the remaining cases we have modelled (at least partially) the catalytic domain of the sensor as well as the receiver and DNA-binding domains of the regulator. In some cases we were also able to model the transmitter or the signal-sensing domain of the sensors. We then visualized and manipulated the models using SWISSPROTVIEWER (Guex and Peitsch, 1997).

Mathematical representations

We can describe the dynamical behaviour of a system exhibiting variations about any of its steady states by using a well-established power-law formalism for modelling biochemical systems (Savageau, 1969; Shiraishi and Savageau, 1992).

For the models in Fig. 1, the independent variables of our models are the concentrations of the input signal molecules (X_5 primary and X_6 secondary) and the concentration totals for sensor protein ($X_7 = X_1 + X_3$) and regulator protein ($X_8 = X_2 + X_4$). The dependent state variables, which are dependent upon the values of the independent variables and the internal dynamics of the system, are the concentrations of the sensor proteins (X_1 phosphorylated and X_3 unphosphorylated) and of the regulator proteins (X_2 phosphorylated and X_4 unphosphorylated).

There is one equation for each of the four dependent variables. For each dependent variable X_i its change in concentration with time (dX_i / dt) is given by the difference between the corresponding aggregate rate of production (V_i) and the aggregate rate of consumption (V_{-i}). Each of these aggregate rates is represented by a mathematical rate law whose exact form is unknown, but whose arguments consist of all the variables that have a direct influence on the aggregate rate in question. For example, the aggregate rate law for consumption of the phosphorylated sensor protein in Fig. 1 is a function of its concentration X_1 and of the concentration of the unphosphorylated regulator protein X_4 acting as a co-substrate.

Each aggregate rate law can be represented by a product of power-law functions, one for each argument in the rate law. This representation is guaranteed to be an accurate representation for some region of operation about a nominal steady state (e.g. see Shiraishi and Savageau, 1992). Thus, the equations for the first two dependent variables of the TCS with the bifunctional sensor (Fig. 1A) are the following (the dependent variables X_3 and X_4 will be treated below).

$$dX_1 / dt = V_1 - V_{-1} = \alpha_1 X_3^{g_{13}} X_5^{g_{15}} - \beta_1 X_1^{h_{11}} X_4^{h_{14}} \quad (1)$$

$$dX_2 / dt = V_2 - V_{-2} = \alpha_2 X_1^{g_{21}} X_4^{g_{24}} X_6^{g_{26}} - \beta_2 X_2^{h_{22}} X_3^{h_{23}} \quad (2)$$

The corresponding equations for the TCS with the monofunctional sensor (Fig. 1B) are the same, except for the last term in Equation (2):

$$dX_2 / dt = V_2 - V_{-2} = \alpha_2 X_1^{g_{21}} X_4^{g_{24}} X_6^{g_{26}} - \beta_2' X_2^{h'_{22}} \quad (2')$$

The primed parameters [in Equation (2')], whose significance will become evident below, have values that in general are different from the corresponding unprimed parameters in Equation (2).

The multiplicative parameters (rate constants), the α's for production and the β's for consumption, influence the time scales of the reactions and are always non-negative. The subscript of a rate constant refers to the molecular species that is being produced or consumed. For example, α_1 is the rate constant for the aggregate rate of production of X_1 (phosphorylated sensor kinase). The exponential parameters (kinetic orders), *g's* for production and *h's* for consumption, represent the direct influence of each variable on each aggregate rate law. These parameters need not have integer values, but can assume real values (positive, negative or zero). The first subscript of a kinetic order refers to the molecular species that is being produced or consumed; the second refers to the variable that has a direct influence on the aggregate rate of production or consumption. For example, the influences on V_{-1}, the aggregate rate of consumption of X_1 (phosphorylated sensor protein), are represented by the kinetic orders h_{11}, which is the kinetic order representing the influence of X_1 (phosphorylated sensor protein) acting as substrate in the aggregate rate of its own consumption, and h_{14}, which is the kinetic order representing the influence of X_4 (unphosphorylated response regulator) acting as a co-substrate in the aggregate rate of consumption of X_1.

It may seem more natural to represent the aggregate rate of consumption of X_1 by two separate terms such as $\beta_1^a X_1^{h_{11}^a} + \beta_1^b X_1^{h_{11}^b} X_4^{h_{14}^b}$. However, it has been demonstrated that this is less accurate than the simple product of power-law functions described above for small variations about the steady state; this also tends to be the case for large variations, although this is not necessarily true in general (Voit and Savageau, 1987).

Protein synthesis and degradation occur on a time-scale that is in much slower than that of the catalytic activities within the TCS modules. Thus, on the time scale of interest here, the total amount of sensor protein (S_{Total}, X_7) and the total amount of regulator protein (R_{Total}, X_8) are considered to be conserved quantities. This permits one dependent variable to be expressed in terms of the conserved total minus the other dependent variable. Accordingly, one can write:

$$X_3 = X_7 - X_1 \quad (3)$$

$$X_4 = X_8 - X_2 \quad (4)$$

Because of these conservation relationships [Equations (3) and (4)], the rate of change in the concentration X_3 is equal in amount and opposite in sign to that in the concentration X_1 and the rate of change in the concentration X_4 is equal in amount and opposite in sign to that in the concentration X_2.

At any given steady state, one can represent the differences in Equations (3) and (4) by the following products of power-laws (see Sorribas and Savageau, 1989; for a more detailed explanation of the procedure)

$$X_3 = \gamma_3 X_7^{f_{37}} X_1^{f_{31}} \quad (5)$$

$$X_4 = \gamma_4 X_8^{f_{48}} X_2^{f_{42}} \quad (6)$$

The new parameters in Equations (5) and (6) are defined as follows: $f_{37} = (X_7 / X_{30})(dX_{30} / dX_7) = X_7 / X_{30}$, $f_{31} = (X_{10} / X_{30})(dX_{30} / dX_{10}) = - X_{10} / X_{30}$, $\gamma_3 = X_{30} / (X_7^{f_{37}} X_{10}^{f_{31}})$, $f_{48} = (X_8 / X_{40})(dX_{40} / dX_8) = X_8 / X_{40}$, $f_{42} = (X_{20} / X_{40})(dX_{40} / dX_{20}) = - X_{20} / X_{40}$, and $\gamma_4 = X_{40} / (X_8^{f_{48}} X_{20}^{f_{42}})$. The additional subscript $_0$ indicates the operating point at which the representation is made (in this case, the steady state). It must be emphasized that this representation is exact at the operating point. The multiplicative and exponential parameters in Equations (5) and (6) are analogous to the rate-constant and kinetic-order parameters in Equations (1), (2) and (2'). Thus, the γ's are non-negative and the *f's* are real (positive, negative or zero). The subscript of a γ parameter, and the first subscript of an f parameter, refers to the dependent variable that is being represented in terms of its paired variable and their sum. The second subscript of an f parameter refers to either the paired variable or the variable representing their sum.

The final form of the equations that will be used here is obtained by substituting the expressions in Equations (5) and (6) into Equations (1), (2) and (2'), and then redefining terms. Thus, one can write

$$dX_1 / dt = \alpha_{10} X_1^{g_{11}} X_5^{g_{15}} X_7^{g_{17}} - \beta_{10} X_1^{h_{11}} X_2^{h_{12}} X_8^{h_{18}} \quad (7)$$

$$dX_2 / dt = \alpha_{20} X_1^{g_{21}} X_2^{g_{22}} X_6^{g_{26}} X_8^{g_{28}} - \beta_{20} X_1^{h_{21}} X_2^{h_{22}} X_7^{h_{27}} \quad (8)$$

and

$$dX_2 / dt = \alpha_{20} X_1^{g_{21}} X_2^{g_{22}} X_6^{g_{26}} X_8^{g_{28}} - \beta'_{20} X_2^{h'_{22}} \quad (8')$$

where the new parameters in Equations (7), (8), and (8′) are related to the original parameters in Equations (1), (2), (2'), (5) and (6) as follows: $\alpha_{10} = \alpha_1 \gamma_3^{g_{13}}$, $g_{11} = g_{13} f_{31}$, $g_{17} = g_{13} f_{37}$, $\beta_{10} = \beta_1 \gamma_4^{h_{14}}$, $h_{12} = h_{14} f_{42}$, $h_{18} = h_{14} f_{48}$, $\alpha_{20} = \alpha_2 \gamma_4^{g_{24}}$, $g_{22} = g_{24} f_{42}$, $g_{28} = g_{24} f_{48}$, $\beta_{20} = \beta_2 \gamma_3^{h_{23}}$, $h_{21} = h_{23} f_{31}$, and $h_{27} = h_{23} f_{37}$.

All of the parameters in Equations (7), (8) and (8′) are positive except for g_{11}, h_{12}, g_{22} and h_{21}, which are negative. A negative exponent implies that an increase in the argument results in a decrease in the value of the power-law function. For example, an increase in X_2 results in a decrease in the rate of consumption of X_1. This apparent inhibition of V_{-1} by X_2 is a result of the conservation relation among the different forms of the response regulator. Thus, an increase in X_2 corresponds to an equivalent decrease in X_4, which actually causes a de-activation of V_{-1}. Similar explanations account for the behaviour associated with the other negative kinetic orders.

As a practical matter, the results for the monofunctional design can be obtained directly from those for the bifunctional design simply by making the following exchanges: $h_{21} \rightarrow 0$, $h_{27} \rightarrow 0$, and $\beta_{20} \rightarrow \beta'_{20}$, and $h_{22} \rightarrow h'_{22}$. Hence, in the following sections we shall focus on the bifunctional design, make these exchanges to obtain the corresponding results for the monofunctional design, and then make the appropriate comparisons.

Steady-state solution and key systemic properties

The equations describing the dynamic behaviour of the model in Fig. 1A. [Equations (7) and (8)] can be solved analytically for the steady state (Savageau, 1969; 1971a), where the rates of aggregate production and aggregate consumption for each metabolite are the same. By equating these aggregate rates, taking logarithms of both sides of the resulting equations and rearranging terms, one can write the steady-state equations as follows:

$$a_{11} Y_1 - h_{12} Y_2 = b_1 - g_{15} Y_5 - g_{17} Y_7 + h_{18} Y_8 \quad (9)$$

$$a_{21} Y_1 + a_{22} Y_2 = b_2 - g_{26} Y_6 - h_{27} Y_7 - g_{28} Y_8 \quad (10)$$

where $Y_i = log\ X_i$, $b_i = log(\beta_{i0} / \alpha_{i0})$, and $a_{ij} = g_{ij} - h_{ij}$. Equations (9) and (10) are single linear algebraic equations that can be solved for the dependent variables Y_1 and Y_2 in terms of the parameters of the system, the input variables Y_5 and Y_6, and the total concentrations of sensor and regulator proteins Y_7 and Y_8. Thus,

$$Y_1 = \frac{(a_{22} b_1 + h_{12} b_2) - a_{22} g_{15} Y_5 - h_{12} g_{26} Y_6 - (a_{22} g_{17} - h_{12} h_{27}) Y_7 + (a_{22} h_{18} - h_{12} g_{28}) Y_8}{a_{11} a_{22} + h_{12} a_{21}} \quad (11)$$

$$Y_2 = \frac{(a_{11} b_2 - a_{21} b_1) + a_{21} g_{15} Y_5 - a_{11} g_{26} Y_6 + (a_{21} g_{17} + a_{11} h_{27}) Y_7 - (a_{21} h_{18} + a_{11} g_{28}) Y_8}{a_{11} a_{22} + h_{12} a_{21}} \quad (12)$$

Two types of coefficients, logarithmic gains and parameter sensitivities, can be used to characterize further the steady state of such models. Because steady-state solutions exist in explicit form [Equations (11) and (12)], we can calculate each of the two types of coefficients simply by taking the appropriate derivatives. Although the mathematical operations involved are the same in each case, it is important to keep in mind that the biological significance of the two types of coefficients is very different.

Logarithmic gains provide important information concerning the amplification or attenuation of signals as they are propagated through the system (Savageau, 1971a; Shiraishi and Savageau, 1992). For example,

$$L(X_i, X_5) = \frac{d\,log\,X_i}{d\,log\,X_5} = \frac{dY_i}{dY_5} \quad \text{or} \quad L(V_i, X_5) = \frac{d\,log\,V_i}{d\,log\,X_5} \quad (13)$$

measures the percent change in the value of the dependent concentration variable X_i (or in V_i the flux through the pool of X_i) caused by a percentage change in the concentration of the input signal X_5. A positive sign indicates that the changes are in the same direction (both increase or both decrease); a negative sign indicates that the changes are in the opposite direction (one increases while the other decreases).

Parameter sensitivities provide important information about system robustness, i.e. how sensitive the system is to perturbations in the structural determinants of the system (Savageau, 1971b; Shiraishi and Savageau, 1992). For example,

$$S(X_i, p_j) = \frac{d\,log\,X_i}{d\,log\,p_j} = p_j \frac{dY_i}{dp_j} \quad \text{or} \quad S(V_i, p_j) = \frac{d\,log\,V_i}{d\,log\,p_j} \quad (14)$$

measures the per cent change in the value of the dependent concentration variable X_i (or in V_i the flux through the pool of X_i) caused by a percentage change in the value of the parameter p_j. Again, a positive sign indicates that the changes are in the same direction (both increase or both decrease); a negative sign indicates that the changes are in the opposite direction (one increases while the other decreases). The aggregate sensitivity of a given variable is defined as the Euclidean norm of the vector whose components are the individual parametric sensitivities for that variable. That is,

$$S(X_i) = \sqrt{\sum_j S(X_i, p_j)^2} \quad \text{or} \quad S(V_i) = \sqrt{\sum_j S(V_i, p_j)^2} \quad (15)$$

These systems should be stable in the face of perturbations in their dependent state variables. That is, following a perturbation, the systems should return to their predisturbance

state. The local stability of the steady state can be determined by applying the Routh criteria (Dorf, 1992). The magnitude of the two critical Routh conditions can be used to quantify the margin of stability (Savageau, 1976). The two critical Routh conditions are given by

$$F_1 a_{11} + F_2 a_{22} < 0 \tag{16}$$

$$F_1 F_2 (a_{11} a_{22} + h_{12} a_{21}) > 0 \tag{17}$$

where $F_1 = V_{10} / X_{10}$ and $F_2 = V_{20} / X_{20}$ are the reciprocal of the turnover times for the X_1 and X_2 pools respectively.

Systems should respond quickly to changes in their environment (Savageau, 1975). Thus, another key property of the systems is their temporal response, which was determined by computer solution of the dynamic equations [Equations (7) (8) and (8')]. At time zero, each intermediate concentration was set to a value 20% less than its steady-state value. The concentrations were then followed as a function of time from this initial condition, and the time for all the concentrations to settle within 1% of their final steady-state value was calculated and denoted by the symbol τ.

Mathematically controlled comparisons

To determine the differences in systemic behaviour between the reference model (bifunctional sensor, Fig. 1A) and the alternative model (monofunctional sensor, Fig. 1B) we use a technique known as mathematically controlled comparison (Savageau, 1972, 1976; Irvine and Savageau, 1985; Alves and Savageau, 2000a). This technique introduces mathematical controls to ensure that the differences observed in the systemic behaviour of alternative models are a result of the specific differences in the design and not some accidental difference. The parameters of the alternative model are fixed relative to those of the reference model by introducing constraints to ensure that the two models are as nearly equivalent as possible from both an internal and an external perspective.

Only the step that accounts for the dephosphorylation of the regulator is allowed to differ between the reference model and the alternative model. Therefore, to establish internal equivalence (Savageau, 1972, 1976; Irvine, 1991) between the two designs, we require the values for the corresponding parameters of all other steps in the two models to be the same. The step that accounts for the dephosphorylation of the regulator is then the only step that differs between the reference model and the alternative model. If we reason that loss or gain of an activation site on the regulator protein comes about by mutation, and that this mutation can cause changes in all the parameters of the process, then a mutation that converts a bifunctional sensor to a monofunctional sensor would change the parameters h_{22}, h_{23}, and β_2 in Equation (2) to h'_{22}, $h'_{23} = 0$, and β'_2 in Equation (2'). Alternatively, the parameters, h_{21}, h_{22}, h_{27}, and β_{20} in Equation (8) would change to $h'_{21} = 0$, h'_{22}, $h'_{27} = 0$, and β'_{20} in Equation (8').

Because we wish to determine the differences that are due solely to changes in the structure of the system, we need to specify values for h'_{22} and β'_2 that minimize all other differences. This is accomplished by deriving the mathematical expression for a given steady-state property in each of the two models, equating these expressions to produce a constraint, and then solving the constraint equation for the value of a primed parameter (see *Calculating the constraints for external equivalence*). The process we have just described determines the maximal degree of external equivalence (Savageau, 1972, 1976; Irvine, 1991) between the two models. Once the two primed parameters have been determined in this manner, there are no more 'free' parameters that can be adjusted to reduce the differences, and all remaining differences can be attributed to the change in system structure from one with a bifunctional sensor to one with a monofunctional sensor. Having established the conditions on the parameters for maximal equivalence, we can proceed to analyse the two models and determine their remaining differences.

Calculating the constraints for external equivalence

Each constraint is established by requiring the value of a relevant systemic property to be the same in both designs. Because we have two primed parameters that need to be constrained, we require two different systemic properties to be invariant between the designs.

First, the value of h'_{22} is fixed by requiring the total gain of the system, $L(X_2,X_5) + L(X_2,X_6)$, to be the same in both designs. This total gain is determined from the explicit solution for X_2 [$Y_2 = \log X_2$] in Equation (12)] by calculating the logarithmic gains as defined by the derivatives in Equation (13). Equating the results for each of the alternative designs allows the value of h'_{22} to be calculated as a function of the kinetic-order parameters in the reference model, which is taken to be the bifunctional case.

$$h'_{22} = \frac{g_{15}g_{21}h_{22} + g_{26}h_{11}h_{22} - g_{15}g_{22}h_{21} - g_{26}h_{12}h_{21} - g_{11}g_{26}h_{22}}{g_{15}g_{21} + g_{26}h_{11} - g_{11}g_{26} - g_{15}h_{21}} \tag{18}$$

By using this constraint we ensure that the output signal of the response regulator X_2 is the same for both designs in responses to the aggregate of signals X_5 and X_6 that change its phosphorylation state. We also have constrained the alternative designs to have equal responses to the primary input signal X_5 [$L(X_2,X_5)$],or to the secondary input signal X_6 [$L(X_2,X_6)$]. The results of the comparative analysis in these other cases are qualitatively the same as those of the comparative analysis reported here (data not shown).

Second, the value of β'_{20} is fixed by requiring the steady-state concentrations of the corresponding variables to be the same in both designs. By equating the explicit solution for X_2 [$Y_2 = \log X_2$ in Equation (12)] in each of the designs, and by utilizing the result in Equation (18), one is able to express the value of β'_{20} in terms of the independent variables and parameters of the reference model.

$$\log \beta'_{20} = \frac{\begin{pmatrix} g_{15}h_{21}\log\alpha_{20} - (g_{15}g_{21} - g_{26}(g_{11} - h_{11}))\log\beta_{20} \\ +g_{26}h_{21}\log(\beta_{10}/\alpha_{10}) + g_{15}g_{26}h_{21}\log(X_6/X_5) \\ +(g_{26}(h_{27}(g_{11} - h_{11})) - g_{15}g_{21}h_{27} - g_{17}g_{26}h_{21})\log X_7 \\ +h_{21}(g_{15}g_{28} + g_{26}h_{18})\log X_8 \end{pmatrix}}{g_{26}(g_{11} - h_{11}) + g_{15}(h_{21} - g_{21})} \tag{19}$$

In addition to making the steady-state value of X_2 equal in the alternative designs, this value of β'_{20} also makes the steady-state value of X_1 [and X_3 (= $X_7 - X_1$) and X_4 (= $X_8 - X_2$)], and the steady-state values of the corresponding fluxes V_2 (= V_4) and V_1 (= V_3) equal in the alternative designs.

Numerical analysis

The analytical results give qualitative information that characterizes the effect of bifunctional versus monofunctional sensors in the models of Fig. 1. To obtain quantitative information about the comparisons, one must introduce specific values for the parameters and compare models (Alves and Savageau, 2000a). For this purpose we have randomly generated a large ensemble of parameter sets and selected 5 000 of these sets that define models consistent with various physical and biochemical constraints. These constraints include conservation of mass considerations, a requirement for positive signal amplification, and stability margins large enough to ensure local stability of the systems. A detailed description of these methods can be found in Alves and Savageau (2000b).

When applied to the current comparisons, the procedure is as follows. We select random values for all the unprimed parameters and for all the independent concentration variables (X_5 through X_8) in Equations (7) (8) and (8'). The values of the primed parameters in Equation (8') are then fixed by the relationships in Equations (18) and (19). The analytical solution in Equations (11) and (12) determines the steady-state values for dependent state variables (X_1 and X_2), complementary variables (X_3 and X_4), fluxes, logarithmic gains and parameter sensitivities. This information in turn determines the values for all the parameters following Equations (6) and (8). Taken together, this information determines the values for all the parameters in the original equations [Equations (1) (2) and (2')]. Mathematica™ (Wolfram, 1997) was used for all the numerical procedures.

To interpret the ratios that result from our comparative analysis we use Density of Ratios plots as defined in Alves and Savageau (2000c). The primary density plots from the raw data have the magnitude for some property of the reference model on the *x*-axis and the corresponding ratio of magnitudes (reference model to alternative model) on the *y*-axis. Secondary density plots are constructed from the primary plots by the use of moving quantile techniques with a window size of 500. The slope in the secondary plot measures the degree of correlation between the quantities plotted on the *x*- and y-axes.

Acknowledgements

This work was supported in part by a joint PhD fellowship PRAXIS XXI/BD/9803/96 granted by PRAXIS XXI through Programa Gulbenkian de Doutoramentos em Biologia e Medicina (R.A), U.S. Public Health Service Grant RO1-GM30054 from the National Institutes of Health (M.A.S.), and U.S. Department of Defense Grant N00014-97-1-0364 from the Office of Naval Research (M.A.S.). We thank Drs Susana Nery, Armindo Salvador, Vic DiRita and Alex Ninfa for critically reading early versions of this manuscript and making useful comments. We also thank an anonymous referee for a very careful and constructive review of the original manuscript. This work was completed while MAS was a guest at the Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France. He thanks Drs A. Carbonme. M. Gromov and F. Képès for providing resources, intellectual stimulation and generous hospitality.

References

Alves, R., and Savageau, M.A. (2000a) Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics* **16:** 786–798.

Alves, R., and Savageau, M.A. (2000b) Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics* **16:** 534–547.

Alves, R., and Savageau, M.A. (2000c) Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics* **16:** 527–533.

Arthur, M., Depardieu, F., Gerbaud, G., Galimand, M., Leclercq, R., and Courvalin, P. (1997) The VanS sensor negatively controls VanR-mediated transcriptional activation of glycopeptide resistance genes of Tn1546 and related elements in the absence of induction. *J Bacteriol* **179:** 97–106.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res* **28:** 235–242.

Blat, Y., Gillespie, B., Bren, A., Dahlquist, F.W., and Eisenbach, M. (1998) Regulation of phosphatase activity in bacterial chemotaxis. *J Mol Biol* **284:** 1191–1199.

Chiang, R.C., Cavicchioli, R., and Gunsalus, R.P. (1997) Locked on and locked off signal transduction mutations in the periplasmic domain of the *Escherichia coli* NarQ and NarX sensors affect nitrate and nitrite dependent regulation by NarL and NarP. *Mol Microbiol* **24:** 1049–1060.

Dorf, R.C. (1992) *Modern Control Systems*, 6th edn. Reading, MA: Addison-Wesley.

Eisenbach, M. (1996) Control of bacterial chemotaxis. *Mol Microbiol* **20:** 903–910.

Feng, J., Atkinson, M.R., McCleary, W., Stock, J.B., Wanner, B.L., and Ninfa, A.J. (1992) Role of phosphorylated metabolic intermediates in the regulation of glutamine synthetase synthesis. *J Bacteriol* **174:** 6061–6070.

Grob, P., Hennecke, H., and Gottfert, M. (1994) Cross talk between the two component regulatory systems NodVW and NwsAB of *Bradyrhizobium japonicum*. *FEMS Microbiol Lett* **120:** 349–353.

Guex, N., and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* **18:** 2714–2723.

Hellingwerf, K.J., Crielaard, W.C., and de Mattos, M.J.T. (1998) Current topics in signal transduction in bacteria. *Anton Leeuw Int J G* **74:** 211–227.

Hlavacek, W.S., and Savageau, M.A. (1995) Subunit structure of regulator proteins influences the design of gene circuitry: Analysis of perfectly coupled and completely uncoupled circuits. *J Mol Biol* **248:** 739–755.

Hlavacek, W.S., and Savageau, M.A. (1996) Rules for cou-

pled expression of regulator and effector genes in inducible circuits. *J Mol Biol* **255:** 121–139.
Hlavacek, W.S., and Savageau, M.A. (1997) Completely uncoupled and perfectly coupled gene expression in repressible systems. *J Mol Biol* **266:** 538–558.
Hoch, J.E., and Silhavy, T.J. (1995) *Two-Component Signal Transduction*. Washington D.C.: American Society for Microbiology Press.
Hsing, W.H., Russo, F.D., Bernd, K.K., and Silhavy, T.J. (1998) Mutations that alter the kinase and phosphatase activities of the two-component sensor EnvZ. *J Bacteriol* **180:** 4538–4546.
Hulett, F.M. (1996) The signal-transduction network for Pho regulation in *Bacillus subtilis*. *Mol Microbiol* **19:** 933–939.
Igo, M.M., Ninfa, A.J., Stock, J.B., and Silhavy, T.J. (1989) Phosphorylation and dephosphorylation of a bacterial transcriptional activator by a transmembrane receptor. *Gene Devel* **3:** 1725–1734.
Irvine, D.H. (1991) The method of controlled mathematical comparisons. In *Canonical Nonlinear Modeling: S-Systems Approach to Understanding Complexity.* Voit, E.O., (ed.) New York: Van Nostrand Reinhold, pp. 90.
Irvine, D.H., and Savageau, M.A. (1985) Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J Immunol* **134:** 99–109.
Jiang, P., and Ninfa, A.J. (1999) Regulation of autophosphorylation of *Escherichia coli* Nitrogen regulator II by the PII signal transduction protein. *J Bacteriol* **181:** 1906–1911.
Jung, K., and Altendorf, K. (1998) Individual substitutions of clustered arginine residues of the sensor kinase KdpD of *Escherichia coli* modulate the ratio of kinase to phosphatase activity. *J Biol Chem* **273:** 26415–26420.
Jung, K., Tjaden, B., and Altendorf, K. (1997) Purification, reconstitution, and characterization of KdpD, the turgor sensor of *Escherichia coli*. *J Biol Chem* **272:** 10847–10852.
Kadner, R.J. (1996) Cytoplasmic membrane. In *Escherichia Coli and Salmonella: Cellular and Molecular Biology*. Neidhardt, F.C., (ed). Washington D.C.: American Society for Microbiology Press, pp. 58–87.
Kanamaru, K., Aiba, H., Mizushima, S., and Mizuno, T. (1989) Signal transduction and osmoregulation in *Escherichia coli*. A single amino acid change in the protein kinase, EnvZ, results in loss of its phosphorylation and dephosphorylation abilities with respect to the activator protein, OmpR. *J Biol Chem* **264:** 21633–21637.
Keener, J., and Kustu, S. (1988) Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. *Proc Natl Acad Sci USA* **85:** 4976–4980.
Lukat, G.S., McCleary, W.R., Stock, A.M., and Stock, J.B. (1992) Phosphorylation of bacterial response regulator proteins by low molecular weight phosphodonors. *Proc Natl Acad Sci USA* **89:** 718–722.
Mayover, T.L., Halkides, C.J., and Stewart, R.C. (1999) Kinetic characterization of CheY phosphorylation reactions: Comparison of P-CheA and small-molecule phosphodonors. *Biochemistry* **38:** 2259–2271.
McCleary, W.R. (1996) The activation of PhoB by Acetylphosphate. *Mol Microbiol* **20:** 1155–1163.
Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995) SCOP. a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247:** 536–540.
Ninfa, A.J., Ninfa, E.G., Lupas, A.N., Stock, A., Magasanik, B., and Stock, J. (1988) Crosstalk between bacterial chemotaxis signal transduction proteins and regulators of transcription of the NTR regulon – Evidence that nitrogen assimilation and chemotaxis are controlled by a common phosphotransfer mechanism. *Proc Natl Acad Sci USA* **84:** 5492–5496.
Ninfa, E.G., Stock, A., Mowbray, S., and Stock, J. (1991) Reconstitution of the bacterial chemotaxis signal transduction system from purified components. *J Biol Chem* **266:** 9764–9770.
Parkinson, J.S. (1993) Signal transduction schemes of bacteria. *Cell* **73:** 857–871.
Perego, M., and Hoch, J.A. (1996) Protein aspartate phosphatases control the output of two-component signal transduction systems. *Trends Genetics* **12:** 97–101.
Pioszak, A.A., and Ninfa, A.J. (2003) Genetic and biochemical analysis of the phosphatase activity of *Escherichia coli* NRII (NtrB) and its regulation by the PII signal transduction protein. *J. Bacteriol* **185:** 1299–1315.
Posas, F., Takekawa, M., and Saito, H. (1998) Signal transduction by MAP kinase cascades in budding yeast. *Curr Op Microbiol* **1:** 175–182.
Pratt, L.A., Hsing, W.H., Gibson, K.E., and Silhavy, T.J. (1996) From acids to osmZ: Multiple factors influence synthesis of the OmpF and OmpC porins in *Escherichia coli*. *Mol Microbiol* **20:** 911–917.
Sanders, D.A., Gillecastro, B.L., Stoch, A.M., Burlingame, A.L., and Koshland, D.E. (1989) Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. *J Biol Chem* **264:** 21770–21778.
Savageau, M.A. (1969) Biochemical Systems Analysis II: The steady-state solution for an n-pool system using a power-law approximation. *J Theor Biol* **25:** 370–379.
Savageau, M.A. (1971a) Concepts relating the behaviour of biochemical systems to their underlying molecular properties. *Arch Biochem Biophys* **145:** 612–621.
Savageau, M.A. (1971b) Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature* **229:** 542–544.
Savageau, M.A. (1972) The behavior of intact biochemical control systems. *Curr Top Cell Reg* **6:** 63–130.
Savageau, M.A. (1975) Optimal design of feedback control by inhibition: Dynamical considerations. *J Mol Evol* **5:** 199–222.
Savageau, M.A. (1976) *Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology.* Reading, MA: Addison-Wesley.
Schaller, G.E. (1997) Ethylene and cytokinin signalling in plants: the role of two-component systems. *Essays Biochem* **32:** 101–111.
Schroder, I., Wolin, C.D., Cavicchioli, R., and Gunsalus, R.P. (1994) Phosphorylation and dephosphorylation of the NarQ, NarX, and NarL proteins of the nitrate dependent

two-component regulatory system of *Escherichia coli*. *J Bacteriol* **176:** 4985–4991.

Shi, L., Liu, W., and Hulett, F.M. (1999) Decay of activated *Bacillus subtilis* pho response regulator, PhoP-P, involves the PhoR-P intermediate. *Biochemistry* **37:** 14575–14584.

Shiraishi, F., and Savageau, M.A. (1992) The tricarboxylic acid cycle in *Dictyostelium discoideum* II. Evaluation of model consistency and robustness. *J Biol Chem* **267:** 22919–22925.

Sorribas, A., and Savageau, M.A. (1989) A comparison of variant theories of intact biochemical systems 1. Enzyme–enzyme interactions and biochemical systems theory. *Math Biosci* **94:** 161–193.

Voit, E.O., and Savageau, M.A. (1987) Accuracy of alternative representations for integrated biochemical systems. *Biochemistry* **26:** 6869–6880.

Wanner, B.L. (1992) Is cross regulation by phosphorylation of two-component response regulator proteins important in bacteria? *J Bacteriol* **174:** 2053–2058.

Weiss, V.V., and Magasanik, B. (1988) Phosphorylation of nitrogen regulator I (NRI) of *Escherichia coli*. *Proc Natl Acad Sci USA* **85:** 5492–5496.

Wingrove, J.A., and Gober, J.W. (1996) Identification of an asymmetrically localized sensor histidine kinase responsible for temporally and spatially regulated transcription. *Science* **274:** 597–601.

Wolfram, S. (1997) *Mathematica^{tm}: a System for Doing Mathematics by Computer*. Menlo Park, CA: Addison-Wesley.

Wright, G.D., Holman, T.R., and Walsh, C.T. (1993) Purification and characterization of VanR and the cytosolic domain of VanS – A two component regulatory system required for vancomycin resistance in *Enterococcus faecium* BM4147. *Biochemistry* **32:** 5057–5063.

Zhu, Y., Qin, L., Yoshida, T., and Inouye, M. (2000) Phosphatase activity of histidine kinase EnvZ without kinase catalytic domain. *Proc Natl Acad Sci USA* **97:** 7808–7813.