The XylS/AraC family of regulators

María-Trinidad Gallegos, Carmen Michán and Juan L. Ramos*

Consejo Superior de Investigaciones Científicas, Estación Experimental del Zaidín, Department of Plant Biochemistry, Profesor Albareda 1, E-18008 Granada, Spain

Received January 14, 1993; Accepted January 22, 1993

ABSTRACT

At least twenty-seven proteins belong to the XylS/AraC family of prokaryote transcriptional regulators. All members of this family except CelD and TetD are positive transcriptional factors. Three subgroups were distinguished within the family in accordance with the Needleman and Wunsch algorithm. Multiple alignment of these proteins revealed that they shared a high degree of sequence homology at their C-terminal end, where a characteristic conserved motif, whose consensus sequence is I-DIA—GF-S—YF—F—G-TPS—R (where — means any aminoacid), was found. Within the homologous C-terminal region, but outside the above consensus motif, a putative DNA-binding domain organized as a helix-turn-helix motif was located in all regulators. For regulators recognizing chemical signals, the non-homologous N-terminal region of these regulators is presumed to contain binding sites for activator molecules that confer specificity.

INTRODUCTION

Searches for homology among protein sequences have facilitated the identification of well conserved motifs, which indicate common ancestry relatedness and reflect conservation of functions, eg cofactor binding domain, transient peptides, DNA binding domain, etc. (1).

Among prokaryote transcriptional regulators a series of families have been identified according to their primary structure homology. In most of the regulators, e.g. those grouped in the LysR, GalR-LacI, the sensor and the activator functions seem to reside in the same polypeptide. In contrast, in the NtrB/NtrC family (5) as well as in some members of the XylS-AraC (6–7) and LuxR-UhpA families (8), the sensor and the activator functions were located in two different proteins. In the NtrB/NtrC family, one member of the pair acts as the environmental or cellular sensor and signal transmitter element, while the other performs as the receptor and regulatory element, the latter being the one that interacts with regulatory DNA sequences (5, 9).

A series of regulators were identified as members of the XylS/AraC family, namely, AraC from *Escherichia coli* (10, 11) and *Erwinia carotovora* (12); MelR (13), RhaS (14), RhaR (14, 15) and Rns (16) from *E. coli*; VirF from *Yersinia enterocolitica*

(17) and XylS from the TOL plasmid of *Pseudomonas putida* (18–20). These regulators were involved in the catabolism of certain C-sources—eg rhamnose (RhaR, RhaS) and melibiose (MelR) in *E. coli*, arabinose (AraC) in *E. coli* and *E. carotovora* and alkylbenzoates (XylS) in *P. putida*—and virulence factors in *E. coli* (Rns) and *Y. enterocolitica* (VirF). The following features were shared by most of these regulators: i) all of them are positive transcriptional factors; ii) their DNA binding domain was predicted to be organized as a helix-turn-helix motif, and was located at their C-terminal (6, 7, 21). In contrast, it was proposed that critical residues for effector binding in regulators involved in C metabolism were located at their N-terminal (6, 22); and iii) in most cases transcription of the regulatory gene was divergent from the gene (operon) they regulate.

RESULTS AND DISCUSSION

To find out new proteins belonging to the XylS/AraC family, the eight proteins previously grouped in the family, namely, AraC from *E. coli* and *E. carotovora*, MelR, RhaR, RhaS and Rns from *E. coli*, XylS from *P. putida* and VirF from *Y. enterocolitica* were compared against all the entries the Swiss Protein Library (Release 23, August 1992), a number of published sequences and sequences provided by other colleagues, by using the Pearson and Lipman algorithm (23) as described in the FastA programme provided by the GCG computer package. These analyses revealed that certain proteins exhibited some degree of homology with one or more of the members of the previously defined XylS/AraC family of positive regulators, namely, from *E. coli*: AppY and the almost identical M5 protein (24, 25), SoxS (26, 27), CelD (28), CfaD (29), EnvY (30) and FapR (31); from Tn10, TetD (32); from *Salmonella typhimurium* AraC (33) and RhaS (34); from *Citrobacter freundii*, AraC (35) from *Bacillus subtilis*, AdaA (36); from *Yersinia pestis*, LcrF (37); from *Shigella flexneri* VirF (38) and from *Pseudomonas aeruginosa* MmsR (39) and ExsA (40); from *Vibrio cholerae*, Tcpn (also called ToxT) (41, 42) and XylS1, and XylS3 from *Pseudomonas putida* pWW53 and XylS from *Pseudomonas putida* pDK1 (P. Williams, personal communication). pDK1 and pWW53 are TOL plasmids, whose physical maps are clearly distinct from the map of the archetypal TOL plasmid pWWO (43). Most of these proteins are about 300 amino acids long, except TetD and SoxS which contain 138 and

* To whom correspondence should be addressed

Table 1. Proteins belonging to the XylS/AraC family of regulators.

Protein	Microorganism	SWISS-PROT accession number	Reference
AdaA	<i>Bacillus subtilis</i>	P19219	36
AppY	<i>Escherichia coli</i>	P05052	24, 25
AraC	<i>Citrobacter freundii</i>	P11765	35
AraC	<i>Escherichia coli</i>	P03021	10, 11
AraC	<i>Erwinia chrysanthemi</i>	P07642	12
AraC	<i>Salmonella typhimurium</i>	P03022	33
CelD	<i>Escherichia coli</i>	P17410	28
CfaD	<i>Escherichia coli</i>	P25393	29
EnvY	<i>Escherichia coli</i>	P10805	30
ExsA	<i>Pseudomonas aeruginosa</i>	P26993	40
FapR	<i>Escherichia coli</i>	P23774	31
LcrF	<i>Yersinia pestis</i>	P28808	37
MelR	<i>Escherichia coli</i>	P10411	13
MmsR	<i>Pseudomonas aeruginosa</i>	P28809	39
RhaR	<i>Escherichia coli</i>	P09378	14
RhaS	<i>Escherichia coli</i>	P09377	14, 15
RhaS	<i>Salmonella typhimurium</i>	P27029	34
Rns	<i>Escherichia coli</i>	P16114	16
SoxS	<i>Escherichia coli</i>	P22539	26, 27
TcpN	<i>Vibrio cholerae</i>	unannoted	41, 42
TetD	<i>Escherichia coli</i>	P28816	32
VirF	<i>Shigella flexneri</i>	unannoted	38
VirF	<i>Yersinia enterocolitica</i>	P13225	17
XylS	<i>Pseudomonas putida</i> pWW0	P07859	18–20
XylS1	<i>Pseudomonas putida</i> pWW53	unannoted	P.Williams
XylS3	<i>Pseudomonas putida</i> pWW53	unannoted	P.Williams
XylS-DK1	<i>Pseudomonas putida</i> pDK1	unannoted	P.Williams

107 amino acids, respectively. Table 1 lists the proteins found to belong to this family and their SWISS-PROT accession number.

Many of the new members of the family are regulators involved in metabolism of C-sources and pathogenesis. In fact, the two different XylS proteins from pWW53 and pDK1 are likely to be involved in the metabolism of alkylbenzoates. The *Pseudomonas* MmsR is the positive regulator for the methylmalonate semialdehyde dehydrogenase operon (39). The *E.coli* CfaD protein is highly homologous to Rns and is also involved in the expression of adhesins CS1 and CS2 (29). FapR controls expression of the fimbrial proteins in enterotoxigenic *E.coli* strains (31). The ExsA regulator induces expression of exoenzyme S, one of the proteins implied in pathogenicity by *P.aeruginosa* (40) and LcrF of *Y.pestis*, Tcpn of *V.cholerae* and VirF of *S.flexneri* are involved in virulence (37, 38, 41, 42). However, other functions among the new members of the family were found, eg AdaA from *B.subtilis* is involved in response to alkylating agents; the *E.coli* EnvY protein controls temperature-dependent expression of *E.coli* envelope proteins, as OmpF, OmpC and LamB (30) and SoxS and AppY proteins from *E.coli* are involved in protection against superoxide and synthesis of acid phosphatase, respectively. The *E.coli* CelD protein functions as a repressor of the *celABCF* operon (28).

For further comparison of these proteins each one was compared to the others using the Dayhoff MATRIX 78-DJD according to the algorithm of Needleman and Wunsch (44). The algorithm enables the alignment score (A) to be calculated, which is the number of standard deviations by which the maximum score of the real sequences exceeds the mean of the maximum score for permutations (see legend for Figure 1). An A value higher than 3.0 is considered significant for homology. According to

the results obtained, three groups were distinguished (See Figure 1).

The first group includes 18 proteins of about 300 amino acids long, which show a high degree of homology with an alignment score of $A \geq 3.0$ for any combination considered (See Figure 1A for examples). This group includes AraC from *E.coli*, *C.freundii*, *S.typhimurium* and *E.carotovora*; RhaS from *E.coli* and *S.typhimurium*; MelR, RhaR, CelD and EnvY from *E.coli*; VirF from *Y.enterocolitica*, XylS from pWW0; XylS1 and XylS3 from pWW53; XylS from pDK1 and MmsR and ExsA from *Pseudomonas*.

The second group includes proteins which exhibited significant homology $A \geq 3.0$, with most of the proteins of the family but not with all of them. This group includes FapR, AppY, Rns and CfaD from *E.coli*; VirF from *S.flexneri*; AdaA from *B.subtilis*; TcpN of *V.cholerae* and LcrF from *Y.pestis*. Comparison of these eight proteins with each other also gave a large variety of A values (Figure 1B).

The third group includes SoxS and TetD, proteins of 107 and 138 residues in size, respectively. Therefore, these proteins are significantly smaller than the rest of the proteins of the family. Comparison among them gave alignment scores $A > 3.0$ (Figure 1C), but they exhibited alignment score values $A < 3.0$ with most of the proteins of the family (See Figure 1C for examples). This is due to the fact that the algorithm of Needleman and Wunsch does not take into account the size of the compared proteins. In any case, the fact that clear subgroups can be distinguished within the family probably reflects that the members of the family have evolved from a common ancestor.

A multiple sequence alignment of the proteins belonging to the family was carried out by using the Pileup programme supplied in the GCG package, which is based on the algorithm

A

	CelD	EnvY	VirFye	MmsR	ExsA	XylS3	XylS	AraC
CelD	52.5							
EnvY	4.8	54.4						
VirFye	4.1	4.1	42.1					
MmsR	6.7	5.3	3.7	44.9				
ExsA	4.7	5.2	39.0	7.9	73.7			
XylS3	3.6	3.0	4.6	7.8	4.1	51.3		
XylS	3.0	3.9	3.8	7.3	4.3	58.5	75.5	
AraC	5.0	4.5	4.6	10.7	7.9	7.8	3.8	98.9

B

	FapR	LcrF	M5	Rns	AdaA	VirFsf	TcpN
FapR	61.5						
LcrF	7.3	54.0					
M5	14.6	3.1	49.0				
Rns	13.9	3.9	10.8	69.0			
AdaA	4.5	3.5	2.8	2.2	48.2		
VirFsf	27.6	3.1	10.4	10.5	2.2	66.6	
TcpN	7.5	3.6	6.5	7.7	0.6	4.5	
XylS	2.6	0.9	0.4	1.0	1.0	1.3	2.8
AraC	3.3	3.9	4.5	2.0	4.7	3.8	3.0
MmsR	1.2	1.4	3.7	3.4	6.5	3.7	3.3

C

	SoxS	TetD
SoxS	22.4	
TetD	18.9	39.6
XylS	2.1	1.8
AraC	3.7	0.7
MmsR	2.8	1.5

Figure 1. Relatedness among the members of the three subgroups of the XylS/AraC family of regulators. Each protein was compared to the other proteins with the Dayhoff MATRIX 78-DJD using the algorithm of Needleman and Wunsch. The maximum score (S) obtained by the alignment of two sequences is compared with the distribution of maximum scores obtained on 100 randomized sequences of the same length and composition. The mean (Sr) and standard deviation (SDr) of these randomized comparisons are calculated. The given alignment score (A) is the number of standard deviations by which the maximum score of the real sequences exceeds the mean of the maximum score for permutations, according to the following equation $A = (S - Sr) / SDr$. An A value higher than 3.0 is considered significant for homology. A, B and C correspond to the three subgroups defined in this study.

of Feng and Doolittle (45) and carries out a progressive pair wise alignment. To avoid bias, we have chosen one sequence from sets of obviously related proteins, eg we chose AraC and RhaS from *E. coli* among the different AraC and RhaS sequences; since LcrF from *Y. pestis* and *Y. enterocolitica* on one hand, and the *E. coli* Rns and CfaD proteins on the other, are highly homologous in their respective sequences, only VirF from *Y. enterocolitica* and Rns were considered in the alignment. Among XylS proteins, XylS from pWVO was considered in the multiple alignment, since it is highly homologous to XylS1 and XylS3 from pWW53 and XylS from pDK1.

The multiple alignment revealed that sequence conservation was restricted to the 1/3 of the C-terminal region of the regulators being about 300 residues long, while it expanded all along the sequence of the regulators being 107 and 138 amino acids long. Figure 2 shows the alignment of the conserved region, where a consensus sequence was derived at each position. An amino acid was chosen for the consensus if it was present in at least 50% of the proteins. The consensus sequence revealed that the highest degree of homology within the family lies in the zone

VirF-Sf	156	RLSDISNNLN	LSEIAVRKRL	ESE.KLTFTQ	ILLDIRMHAA	AKLLLSNS..	202
RnS-Ec	179	TLGIADAFN	ASEITIRKRL	ESE.NTNFNQ	ILMLRMSKA	ALLLLENS..	225
FapR-Ec	169	KLSDIAEEMH	ISEISVRKRL	EQE.CLNFNQ	LILDVRMNAQ	AKFIIRSD..	215
M5-Ec	154	HLKDIAELIY	TSESLIKKRL	RDE.GTSTFE	ILRDRMRYA	KKLITSNS..	200
EnvY-Ec	164	NLRIVASSLC	LSPSLKKKL	KNE.NTSYSQ	IVTECRMRYA	VQMLLMON..	212
TcpN-Vch	187	RWADICGELR	TNRMLKKEL	ESR.GVKFRE	LINSIRISYS	ISLMKTGE..	233
AdaA-Bs	117	TLESADICH	GSPYHMRFT	KKIKGITLVE	YIQQVRHAA	KKYLIQTN..	167
CeID-Ec	191	ALENMVALSA	KSQEYLTRAT	QRYYGKTPMQ	IINEIRINFA	KKQLEMTN..	238
MmsR-Pa	216	ELERLAFCN	LSKFHFVSRY	KAITGRTPIQ	HFLHLKIEYA	CQLDSSD..	263
MeIR-Ec	209	TINDVAEHVK	LNANYAMGIF	QRVMQLTMKQ	YITAMRINH	RALLSDTD..	256
SoxS-Ec	23	NIDVVAKSG	YSKWYLQRMF	RTVHTQLTG	YIRQRLLLA	AVELRTTE..	70
TetD-Tn10	46	LLDDVANKAG	YTKWYQRLF	KKVTGVTLAS	YIRARRLTKA	AVELRLTK..	93
ExsA-Pa	206	KLSDFSREFG	MGLTTFKELF	GSVYGVSPRA	WISERRILYA	HQLLNSD..	253
VirF-Ye	182	KLKFAREFG	MGLTTFKELF	GTVYGVISPR	WISERRILYA	HQLLNGK..	229
RhaR-Ec	224	ALDKFCDEAS	CSEKRLRQFQ	RQQTGMTIQT	YLRQVRVCHA	QYLLQHSR..	271
RhaS-Ec	189	NMDAVADQFS	LSRLTLRHQL	QKQGTGLTPQ	YLNRLRLMKA	RHLRHSE..	236
AraC-Ec	196	DIASVAQHVC	LSPSRLSHLF	RQQLGISVLS	WREDQIRISA	KLLLSSTR..	243
XyIS-Pp	229	SLERLAELAM	MSPRSLYNLF	EKHAGTTPKN	YIRNRKLESI	RACLNPSAN	278

Consensus	-L---A---S-----F---G-T---I---R---A---L-----
	13 13 13 9 12 10 10 16 15 11 15

VirF-Sf	203	.SYINDVSRL	IGISSPSYFI	RKFNEYGYIT	PKKFYLHKK	241
RnS-Ec	226	.YQISQISNM	IGISSASYFI	RIFNKHYGVT	PKQFTTYFKG	264
FapR-Ec	216	.HQIGMIASL	VGYSVSFYFI	KTFKFEYGV	PKKFEIGIKE	254
M5-Ec	201	.YSINNVQAK	CGYNSTSYFI	CAFQDYGVGT	PSHYFEIKIG	239
EnvY-Ec	213	.KNITQVAQL	CGYSSTSYFI	SVFKAFYGLT	PLNYLAKORQ	249
TcpN-Vch	234	.FKIKQIAYQ	SGFASVSFNS	TVFKSTNNVA	PSEYLFMLTG	272
AdaA-Bs	168	.KAIGDIAIC	VGIANAPYFI	TLFKKGTGT	PARFRQSKM	203
CeID-Ec	239	.YSVTDIAFE	AGYSSPSLFI	KTFKLTSTF	PKSYRKLTE	277
MmsR-Pa	264	.QSVARVGQA	VGYYDSYFVS	RLFSKVMGLS	PSAYRQRVRQ	302
MeIR-Ec	257	.KSILDIALT	AGFRSSRFY	STFGKYVGM	PQYRKLQSQ	295
SoxS-Ec	71	.RPIFDIAMD	LGYSQQTFS	RVFRQFQRT	PSDYRHRL..	107
TetD-Tn10	94	.KTILEIATK	YQFDSQSF	RRFYIFKVT	PSYRRNKLW	132
ExsA-Pa	254	.MSIVDIAME	AGFSQSYFT	QSYRRFGCT	PSRSRQKDE	292
VirF-Ye	230	.MSIVDIAME	AGFSQSYFT	QSYRRFGCT	PSQARLTIA	268
RhaR-Ec	272	.LLISDISTE	CGFEDSNYFS	VVFRETGMT	PSQMRHLMSQ	310
RhaS-Ec	237	.ASVTDIAYR	CGFSDSNHFS	TLFRREFNMS	PRDIRQGE..	272
AraC-Ec	244	.MPIATVGRN	VGFDQQLYFS	RVFKCTGTAS	PSEFRAGCEE	282
XyIS-Pp	279	VRSITEIALD	YGFLHLGRFA	ENYRSAFGEL	PSDTRLQCKK	318

Consensus	---I-DIA---GF-S---YF---F---G-T-PS--R-----
	15 9 13 17 12 11 15 13 18 10
	13 9 18 12 10

Figure 2. Multiple alignment of the C-terminal end of the proteins belonging to the XylS/AraC family. Multiple alignments were carried out using the Pileup program provided by the GCG package. The following abbreviations were used: Bs, *Bacillus subtilis*; Ec, *Escherichia coli*; Pa, *Pseudomonas aeruginosa*; Pp, *Pseudomonas putida*; Sf, *Shigella flexneri*; Tn10, transposon Tn10; Vch, *Vibrio cholerae*; and Ye, *Yersinia enterocolitica*. The region marked by a line indicates the putative DNA binding site of these proteins based on the location of the AraC (21) and XylS (20, 47) DNA binding sites. Numbers underneath the amino acids in the consensus sequence represent their order of appearance.

between residues 290 and 313 of XylS. A phenylalanine and a proline residue are conserved in 18 out of 18 aligned proteins (see Figure 2, these residues correspond to F-297 and P-309 in XylS). It should also be noted that the glycine at position 290 in the XylS protein is also a conserved residue in most of the piled proteins of the family (17 out of 18). A clear motif was derived in this region, whose consensus sequence is I-DIA---GF-S---YF---F---G-T-PS---R (where --- means any residue). According to the algorithm of Garnier et al. (46) this region is predicted to be rich in α -helix, although its role remains to be elucidated.

Previous biochemical and genetics works located the DNA-binding domain of the *E. coli* and *S. typhimurium* AraC proteins (21, 22), the *E. coli* RhaR protein (15) and the *P. putida* XylS regulator (47) within the region of homology (i.e., between residues 230 and 250 in TOL pWVO encoded XylS). The algorithm of Garnier et al. (46) predicts this domain to be organized as a helix-turn-helix motif. Analysis of secondary structure of the corresponding aligned sequences in the proteins

of the family suggested that the DNA-binding domain of these regulators is likely to be organized as a helix-turn-helix motif. If it is indeed the case, certain peculiarities of the helix-turn-helix motif of the family should be noted; first most of these regulators do not exhibit a Gly at position 9, which is usually a conserved residue in proteins with this type of structure (48, 49); second most proteins exhibit a leucine or a phenylalanine residue at position 15 instead of an isoleucine or valine residue; third the first amino acid of the first helix is a leucine or isoleucine residue (15 out of 18) and fourth the first amino acid of the second helix is a proline for the *Pseudomonas* XylS proteins, as it also is for EnvY and AraC from *E.coli* and AdaA of *B.subtilis*. Among the conserved features of the helix-turn-helix motif is that most (13 out of 18) proteins of the family exhibit an alanine residue at position 5.

In the stretch between the proposed DNA binding domain and the family consensus region certain residues were also highly conserved e.g. in 18 out of 18 proteins at position 263 and in 17 out of 18 proteins at position 271 in the XylS numbering was found arginine or lysine and leucine or isoleucine, respectively.

The alignment of sequences revealed that, on one hand, LcrF from *Y.pestis*, VirF from *Y.enterocolitica* and ExsA from *P.aeruginosa* and on the other, the *E.coli* M5 and EnvY proteins exhibit a high degree of sequence conservation at their C-terminal end (see Fig. 2).

The non-homologous N-terminal and central regions of the regulators recognizing chemical signals are presumed to contain binding sites for activator molecules that confer specificity (6). In fact, mutations at the unconserved N-terminal region in the XylS (50, 51) and AraC (52) regulators led to an alteration in effector specificity. It has been proposed that the effector-regulator complex constitutes the active form of the regulator able to mediate transcription activation from its respective regulated promoters (50, 51). Stimulation of transcription by CfaD (53) and SoxS (54) may involve overproduction mediated by an unknown factor and SoxR respectively, in response to appropriate signals. The XylS regulator can also mediate transcription stimulation from the TOL *meta*-cleavage pathway operon promoter in the absence of effectors, after its physiological overproduction mediated by the XylR regulator, which in turn is activated by hydrocarbon effectors (55).

ACKNOWLEDGEMENTS

This work was supported by a grant from the Comision Interministerial de Ciencias y Tecnología (BIO91-0659). We thank Luis Pezzi for his invaluable assistance and advice in using the GCG's programmes. We thank P.Williams and J.Sokatch for providing sequences before publication. We thank C.J.Dorman for providing the *S.flexneri* VirF sequence. We thank Amos Bairoch for comments on the MS.

REFERENCES

- Argos, P., Hanei, M., Wilson, J.M. and Kelley, W.N. (1985) *J. Biol. Chem.* **258**, 6450–6457.
- David, M., Davaeran, M.L., Batut, J., Dedieu, A., Domergue, J., Ghai, J., Hertig, C., Boistaard, P. and Hahn, D. (1988) *Cell* **54**, 671–683.
- Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6602–6606.
- Weickert, M.J. and Adhya S. (1992) *J. Biol. Chem.* **267**, 15869–15874.
- Ninfa, A.J., Ninfa, E.G., Lupas, A.N., Stock, A., Magasanik, B. and Stock, J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5492–5496.
- Ramos, J.L., Rojo, F., Zhou, L. and Timmis, K.N. (1990) *Nucl. Acids Res.* **18**, 2149–2152.
- Henikoff, S., Wallace, J.C. and Brown J.P. (1990) *Meth. Enzymol.* **183**, 111–132.
- Weston, L.A. and Kadner, R.J. (1988) *J. Bacteriol.* **170**, 3375–3383.
- Wanner, B.L. (1992) *J. Bacteriol.* **174**, 2053–2058.
- Miyada, C.G., Horwith, A.H., Cass, L.G., Timko, J. and Wilcox, G. (1990) *Nucl. Acids Res.* **8**, 5267–5274.
- Stoner, C.M. and Schleif, R.F. (1982) *J. Mol. Biol.* **154**, 649–652.
- Lei, S.P., Lin, H.-C., Heffernan, L. and Wilcox, G. (1985) *J. Bacteriol.* **164**, 717–722.
- Webster, C., Gardner, L. and Busby, S. (1989) *Gene* **83**, 207–213.
- Tobin, J.F. and Schleif, R.F. (1987) *J. Mol. Biol.* **169**, 789–799.
- Tobin, J.F. and Schleif R.F. (1990) *J. Mol. Biol.* **211**, 75–89.
- Caron, J., Coffield, M. and Scott, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 963–967.
- Cornelis, G., Sluiters, C., de Rouvroit, C.L. and Michielis, T. (1989) *J. Bacteriol.* **171**, 254–262.
- Spooner, A.R., Lindsay, K. and Francklin, F.C.H. (1986) *J. Gen. Microbiol.* **132**, 1347–1358.
- Inouye, S., Nakazawa, A. and Nakazawa, T. (1986) *Gene* **44**, 235–242.
- Mermoud, N., Ramos, J.L., Bairoch, A. and Timmis, K.N. (1987) *Mol. Gen. Genet.* **207**, 349–354.
- Brunelle, A. and Schleif, R.F. (1989) *J. Mol. Biol.* **209**, 607–622.
- Lauble, H., Georgalis, Y. and Heinemann, V. (1989) *Eur. J. Biochem.* **185**, 319–325.
- Pearson, W. and Lipman, D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
- Atlung, T., Nielsen, A. and Hansen, F.G. (1989) *J. Bacteriol.* **171**, 1683–1691.
- Kemp, E.H., Minton, N.P. and Mann, N.H. (1987) *Nucl. Acids Res.* **15**, 3924.
- Wu, J. and Weiss, B. (1991) *J. Bacteriol.* **173**, 2864–2871.
- Amabile-Cuevas, C.F. and Dempe, B. (1991) *Nucl. Acids Res.* **19**, 4479–4484.
- Parker, L.L. and Hall, B.G. (1990) *Genetics* **124**, 455–471.
- Savelkoul, P.H.M., Willshaw, G.A., McConnell, M.M., Smith, H.R., Harmes, A.M., Der Zeijst, B.A.M. and Gaastra, W. (1990) *Microbial Pathog.* **8**, 91–99.
- Lundrigan, M.D., Friedrich, M.J. and Kadner, R.J. (1989) *Nucl. Acids Res.* **17**, 800.
- Klaassen, P. and de Graaf, F.K. (1990) *Mol. Microbiol.* **4**, 1779–1783.
- Braus, G., Argast, M. and Beck, C.F. (1984) *J. Bacteriol.* **160**, 504–509.
- Clarke, P., Lin, H.-C. and Wilcox, G. (1982) *Gene* **18**, 157–163.
- Nishitani, J. and Wilcox, G. (1991) *Gene* **105**, 37–42.
- Burke, K.A. and Wilcox, G. (1987) *Gene* **61**, 243–252.
- Morohoshi, F., Hayashi, K. and Munakata, N. (1990) *Nucl. Acids Res.* **18**, 5473–5480.
- Hoe, N.P., Minion, F.C. and Goguen, J.D. (1992) *J. Bacteriol.* **174**, 4275–4286.
- Dorman, C.J. (1992) *Mol. Microbiol.* **6**, 1575–1576.
- Steele, M.I., Lorenz, D., Hatter, K., Park, A. and Sokatch, J.R. (1992) *J. Biol. Chem.* **267**, 13585–13592.
- Frank, D.W. and Iglewski, B.H. (1991) *J. Bacteriol.* **173**, 6460–6468.
- Higgings, D.E., Nazareno, E. and DiRita, V.J. (1992) *J. Bacteriol.* **174**, 6974–6980.
- Ogierman, M.A. and Manning, P.A. (1992) *Gene* **116**, 93–97.
- Assinder, S.J. and Williams, P. (1990) *Adv. Microbiol. Physiol.* **31**, 1–70.
- Needleman, S.B. and Wunsch, C.D. (1970) *J. Mol. Biol.* **48**, 443–453.
- Feng, D.F. and Doolittle, R.F. (1987) *J. Mol. Evol.* **35**, 351–360.
- Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
- Zhou, L., Timmis, K.N. and Ramos, J.L. (1990) *J. Bacteriol.* **172**, 3707–3710.
- Pabo, C.A. and Sauer, R.T. (1984) *Annu. Rev. Biochem.* **53**, 293–297.
- Harrison, S.C. and Aggarwal, A.K. (1990) *Annu. Rev. Biochem.* **59**, 933–969.
- Ramos, J.L., Michán, C., Rojo, F., Dwyer, D. and Timmis, K.N. (1990) *J. Mol. Biol.* **211**, 373–382.
- Michán, C., Zhou, L., Gallegos, M.T., Timmis, K.N. and Ramos, J.L. (1992) *J. Biol. Chem.*, **267**, 22897–22901.
- Cass, L.G. and Wilcox, G. (1986) *J. Bacteriol.* **166**, 892–900.
- Jordi, B.J.A.M., Dagber, B., Haan, L.A.M., Hamers, A.M. Der Zeijst, B.A.M., Gaastra, W. and Uhlin, B.E. (1992) *EMBO J.* **11**, 2627–2632.
- Nunoshiba, T., Hidalgo, E., Amabile-Cuevas, C.A. and Demple, B. (1992) *J. Bacteriol.* **174**, 6054–6060.
- Ramos, J.L., Mermoud, N. and Timmis, K.N. (1987) *Mol. Microbiol.* **1**, 291–300.