

# The $\sigma^{54}$ regulon (sigmulon) of *Pseudomonas putida*

Ildefonso Cases,<sup>1</sup> David W. Ussery<sup>2</sup> and Víctor de Lorenzo<sup>1\*</sup>

<sup>1</sup>Centro Nacional de Biotecnología, CSIC Campus de Cantoblanco, 28049 Madrid, Spain.

<sup>2</sup>Center for Biological Sequence Analysis, Biozentrum-DTU, DK-2800 Lyngby, Denmark.

## Summary

$\sigma^{54}$  is unique among the bacterial sigma factors. Besides not being related in sequence with the rest of such factors, its mechanism of transcription initiation is completely different and requires the participation of a transcription activator. In addition, whereas the rest of the alternative sigma factors use to be involved in transcription of somehow related biological functions, this is not the case for  $\sigma^{54}$  and many different and unrelated genes have been shown to be transcribed from  $\sigma^{54}$ -dependent promoters, ranging from flagellation, to utilization of several different carbon and nitrogen sources, or alginate biosynthesis. These genes have been characterized in many different bacterial species and, only until recently with the arrival of complete genome sequences, we have been able to look at the  $\sigma^{54}$  functional role from a genomic perspective. Aided by computational methods, the  $\sigma^{54}$  regulon has been studied both in *Escherichia coli*, *Salmonella typhimurium* and several species of the Rhizobiaceae. Here we present the analysis of the  $\sigma^{54}$  regulon (sigmulon) in the complete genome of *Pseudomonas putida* KT2440. We have developed an improved method for the prediction of  $\sigma^{54}$ -dependent promoters which combines the scores of  $\sigma^{54}$ -RNAP target sequences and those of activator binding sites. In combination with other evidence obtained from the chromosomal context and the similarity with closely related bacteria, we have been able to predict more than 80% of the  $\sigma^{54}$ -dependent promoters of *P. putida* with high confidence. Our analysis has revealed new functions for  $\sigma^{54}$  and, by means of comparative analysis with the previous studies, we have drawn a potential mechanism for the evolution of this regulatory system.

## Introduction

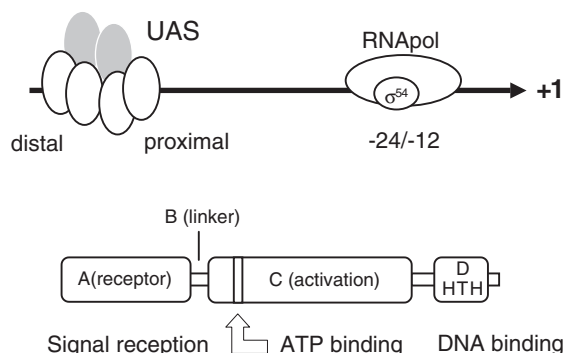
Bacterial RNA polymerase requires transient association with a sigma factor for the initiation of transcription (Helmann and Chamberlin, 1988). Often several sigma factors are present in bacterial genomes, each of them providing a different specificity to the RNA polymerase, and thus the ability to initiate transcription in different sets of promoters. For instance, 22 different sigma factors are present in the genome of *P. putida* (Martínez-Bueno *et al.*, 2002). Whereas most of the sigma factors belong to the same broad family ( $\sigma^{70}$ ) based on sequence similarity,  $\sigma^{54}$  does not belong to this group. Additionally, there are also differences in the molecular mechanism involved in transcription initiation among  $\sigma^{54}$  and the rest of the sigma factors. Functioning of  $\sigma^{54}$  strictly requires the concurrence of an activator protein (Buck *et al.*, 2000), which binds specific sequences located in a relatively remote position from the transcription start site (upstream activating sequences or UAS; Fig. 1). Because of the reminiscence to factors which bind eukaryotic promoters,  $\sigma^{54}$ -dependent activators are frequently referred to as prokaryotic EBPs (enhancer-binding proteins). These EBPs are often described as well as members of the NtrC class of transcriptional regulators, after the name of one of the best known examples of the family. Transcription activation by these regulators involves a complex mechanism which includes the hydrolysis of ATP as its key step (Zhang *et al.*, 2002). In addition, the intervening DNA region between the activator binding site and the  $\sigma^{54}$  binding site has to be bent in order to place the two protein complexes in contact. In some cases, this bending can be assisted by a nucleoid-associated protein, such as IHF or HU (Hoover *et al.*, 1990; Pérez-Martín and de Lorenzo, 1995).

Although most of the alternative sigma factors participate in the transcription of related genes this is not the case for  $\sigma^{54}$ . Despite being initially identified as a nitrogen specific sigma factor (Beynon *et al.*, 1983; Merrick, 1983; Hirschman *et al.*, 1985), now a number of different and disparate genes are known to be transcribed from  $\sigma^{54}$ -dependent promoters. Among these diverse functions there are genes involved in the transport and metabolism of different nitrogen and carbon sources, plant pathogenicity, synthesis of several cofactors, alginate production and flagellation (Studholme and Buck, 2000).

While more and more promoters are added to this list, the arrival of complete genome sequences have allowed

Received 3 July, 2003; accepted 27 August, 2003. \*For correspondence. E-mail vdlorenzo@cnb.uam.es; Tel. (+34) 91 585 4536; Fax (+34) 91 585 4506.

## 2 I. Cases, D. W. Ussery and V. de Lorenzo



**Fig. 1.** General organization of *rpoN*-dependent promoters and cognate activators. The distribution of relevant DNA sequences is shown on top. These include the upstream activating sequences (UAS) which are the target of the prokaryotic enhancer binding proteins (EBPs, activators working in concert with  $\sigma^{54}$ -RNAP) and the -12/-24 region bound by  $\sigma^{54}$ -RNAP itself. Protein sizes are symbolic. The number of activator monomers at the UAS is uncertain, it may range 4–8 units (most probably three dimers). Bottom, domain organization of with  $\sigma^{54}$ -related EBPs. Relevant portions of the proteins include the signal reception, N-terminal A domain, the central (C) module involved in NTP binding, and the D domain at the C-terminus, with a helix–turn–helix (HTH) motif for DNA binding. The A and C modules are connected by a linker B domain.

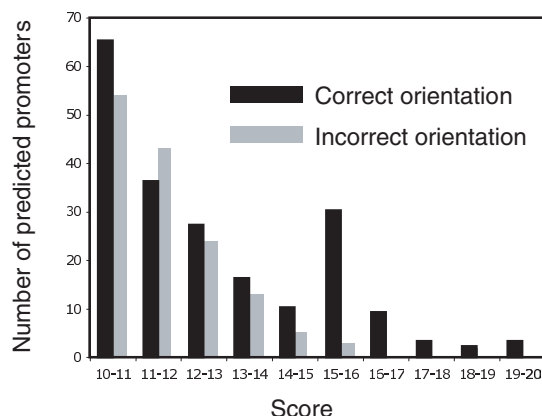
analysis of the complete regulon of a single species. With the help of computational methods, very detailed analyses have been carried out in *E. coli*, *Salmonella typhimurium* and several species of Rhizobiaceae (Reitzer and Schneider, 2001; Dombrecht *et al.*, 2002; Studholme, 2002). This analysis has broadened the range of  $\sigma^{54}$  roles and in the case of *E. coli*, has provided an unified explanation for the apparent disparity of genes regulated by  $\sigma^{54}$ , based on the physiological conditions derived from nitrogen deprivation. However this hypothesis is only applicable to *E. coli* and fails to explain many other functions present in other species such as, for example, its involvement in the catabolism of aromatic compounds (Dixon, 1986). Here we have addressed the study of the complement of genes dependent on  $\sigma^{54}$  in the genome of *P. putida* (Nelson *et al.*, 2002). This species presents impressive metabolic capabilities, regarding both the range of carbon and nitrogen sources that this bacterium is able to use, and some of this capabilities have been shown to depend on the presence of a functional  $\sigma^{54}$  (Kohler *et al.*, 1989). In addition, *P. putida*, as a soil inhabitant, can be anticipated to have a more complex regulation network than the previously characterized enteric bacteria (Cases *et al.*, 2003), and at the same time, less biased in the  $\sigma^{54}$  function than the nitrogen fixing rhizobia, in which most of the  $\sigma^{54}$ -dependent promoters are involved in nitrogen fixation and the establishment of the symbiotic nodule (Dombrecht *et al.*, 2002). With these two considerations in mind, we expected that our analysis would reveal new aspects of the  $\sigma^{54}$  biological role.

To this end we have developed a new algorithm that allowed us to increase the sensitivity of traditional methods. Our method takes advantage of the well-established requirement of an activator protein in this kind of promoters, and of the sequence features of the binding sites for these factors. Thus, we have been able to predict 46  $\sigma^{54}$  promoters in the *P. putida* genome, and estimate the total number of  $\sigma^{54}$  promoters to be around 55. A comparison of our results with the previous analysis of the complement of  $\sigma^{54}$ -dependent genes of *E. coli* and *P. putida* is also presented. In addition, these comparisons shed some light on the possible evolutionary mechanism operating on this kind of regulatory systems.

## Results

### Promoter detection and validation: estimating the number of $\sigma^{54}$ promoters

In order to predict  $\sigma^{54}$  promoters, we developed GenomeMatScan, a program to search instances of DNA patterns according to a position-weight matrix. We built a matrix for  $\sigma^{54}$  promoters based on the compilation by Barrios *et al.* (1999). In order to test the program and the matrix, we ran GenomeMatScan on the *E. coli* intergenic regions. All of the 17 experimentally determined  $\sigma^{54}$  promoters (Reitzer and Schneider, 2001) were detected (not shown). Then we set to run GenomeMatScan on the *Pseudomonas putida* genome. First, the intergenic regions extracted according to the ORFs included in the NCBI version of the *P. putida* genome (Nelson *et al.*, 2002). Sequences (4365) were obtained. These sequences were scanned with GenomeMatScan. In order to select the most appropriate cut-off value, we performed two different tests. First, we compared the score distribution among those predicted promoters which are correctly oriented, that is in the same direction that the downstream gene, and those which are not (Fig. 2). The reasoning of this analysis is that false positives should be homogeneously distributed between both types of promoters, whereas true positives must be correctly oriented. Figure 2 shows that there are no incorrectly oriented promoters scoring over 16, and only four scored over 15, suggesting that 15 is a reasonable threshold, and 16 would avoid almost any false positive. Fifty-one promoters scored over 15, four were in the incorrect direction, and a similar number of false positives could be expected to occur in the correct orientation. Thus we estimate about 15% false positive rate (8/51). Finding out the false negative ratio is more difficult, given the limited number of  $\sigma^{54}$  promoters characterized in the *Pseudomonas putida* genome. If we look at the results obtained in *E. coli*, four out of 17 known promoters scored less than 15, suggesting a false negative ratio around 25%. Assuming a similar ratio in *P. putida*, we can estimate that the total

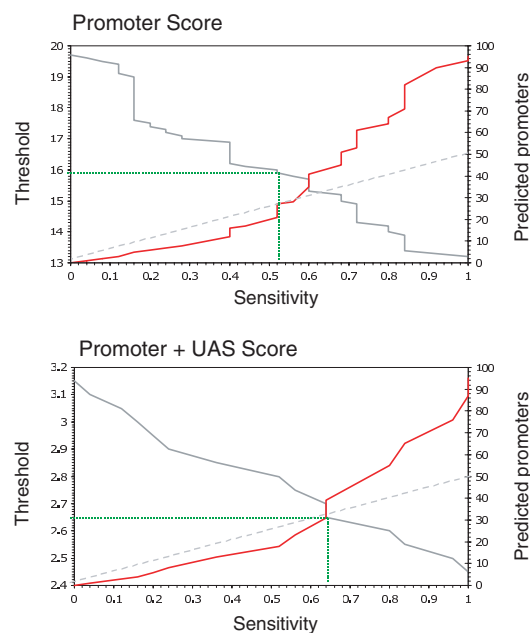


**Fig. 2.** Distribution of scores in correctly oriented and incorrectly oriented promoters predicted in the *P. putida* intergenic regions. Note that the frequencies are almost identical at scores below 15. Only four incorrectly oriented promoters score over 15, and none over 16.

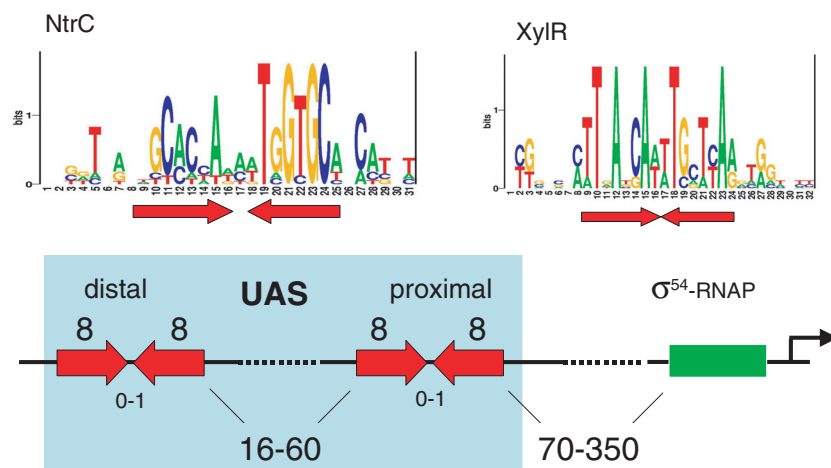
number of  $\sigma^{54}$  promoters should be between 50 and 60. The difficulty to ascertain low-score false negatives is exposed in the case of the promoter driving expression of the *dnaA* (PP1668) gene of *P. putida* (Ingmer and Atlung, 1992). Whereas this *dnaA* promoter has been experimentally characterized as being dependent on  $\sigma^{54}$ , our criterion assigned it a score of only 12.06. In order to better determine an precise threshold which would minimize the number of false positives, we performed an analysis based on the sensitivity of our prediction methods. Sensitivity is defined as the proportion of the real targets that are detected at a given threshold. When the number of predicted targets for different threshold values is plotted against the sensitivity achieved at each value, typically a curve with two regions of different slope is obtained (Ussery *et al.*, 2001). The point where the slope changes corresponds to the point where false positives start to accumulate, and thus is an indication of an adequate threshold. As only a few  $\sigma^{54}$  (mostly extra-chromosomal) promoters have been characterized in detail in *P. putida*, matching the sensitivity of the prediction with experimental data is not trivial. In order to estimate the sensitivity of our method we thus resorted again to a numerical method. To this end, we selected 24 promoters scoring more than 13 as our *true positive* set, based on what is known in close relative species and from other indirect evidences in *P. putida*. It is important to note that the set of promoters to estimate sensitivity has no influence whatsoever in the search method, only on the threshold selected, so minor errors in this step do not influence the promoters found by our method. The plot obtained is shown in Fig. 3. Based on this plot we estimated that a threshold of 15.95 practically guarantees the absence of false positives. Twenty-one predicted promoters scored above this figure. The highest scoring promoter in the incorrect orientation was 15.89.

#### UAS detection and validation: improvements on promoter-only prediction

According to our previous estimation of the number of *P. putida*  $\sigma^{54}$  promoters, the 21 sequences that we predicted represent between 35 and 40% of total number of promoters. We investigated improving our promoter prediction method. One approach that has been successfully employed in the prediction of eukaryotic promoters is the combined detection of the promoter together with transcription factor binding sites. This approach cannot be applied directly to our case given the variety of different regulators able to activate  $\sigma^{54}$ -dependent promoters, for most of which the binding DNA sequence is unknown. However, when the known binding sequences are studied in detail, a number of common features can be observed (Reitzer and Schneider, 2001). First, as it is well known, they all bind the DNA at a relatively distant position from the transcription start site. Frequently two or more binding-sites are present in the promoter, each one consisting in an imperfect palindromic sequence. Although different lengths have been reported, most of known sites can be represented as a palindrome of two hemi-boxes eight nucleotide long, separated by at most one nucleotide (Fig. 4).



**Fig. 3.** Selecting the threshold value based on sensitivity. In order to select the proper threshold value we plotted sensitivity versus the number of predicted promoters (red line). The selected threshold corresponds to a change in the slope of the graph. The projection of this point on the score axis (Y) renders the threshold. The projection of the same point on the sensitivity axis (X) gives the sensitivity achieved (dashed green line). The projection of the number of promoters over the total number of predicted promoters is an estimation of the total number of promoters (grey dashed line). Top panel: sensitivity plot for the Promoter Score. Bottom panel: sensitivity plot for the combined score of the promoter and UAS.



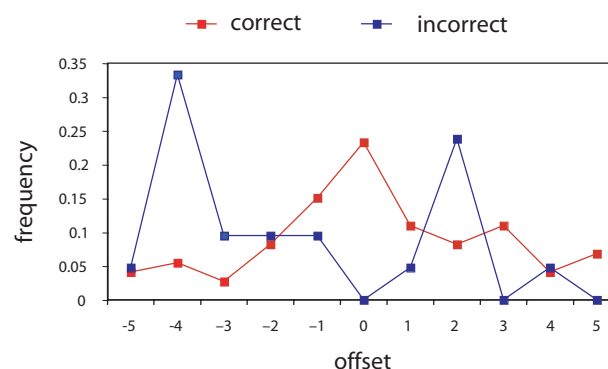
**Fig. 4.** Model of the upstream activating sequences for  $\sigma^{54}$ -dependent promoters. Top panels, LOGO representation (Schneider and Stephens, 1990) of the NtrC and XylR binding sites. Note the symmetry and the spacing: while in the case of NtrC the 8-base inverted repeats are separated by one nucleotide, the half-sites lie next to each other in the UAS for XylR. Bottom panel, representation of the model used for finding UAS. The algorithm searches two repetitions of 8 base long palindromes, in which the hemi-boxes are separated 1 or 0 bases, with the two repetitions located between 16 and 60 bases apart, the search being made in a window spanning 70–350 bases from the transcription start site.

We developed a program to scan the sequence upstream of the predicted binding site for structures with the properties described above. In order to score, we measured the information content of the site considering the four hemi-boxes (Schneider *et al.*, 1986). It should be noted that no sequence similarity is used here, so the program was designed to find unknown binding sites. In order to evaluate the performance of our program, we tried to detect the UAS of those 10 genes in *E. coli* for which binding sites has been characterized. By adjusting the search to combinations of –12/–24 motifs and UAS within genomic DNA segments of 200 bp, we were able to detect eight of them. The other two (those of the *astC* and *hypA* promoters) happened to have their UAS more than 200 nucleotides upstream of the transcription start site, and thus lied outside of the searched region (Reitzer and Schneider, 2001). We then set out to detect UAS for the predicted  $\sigma^{54}$  promoters that score 13 or higher. Such an analysis of the predicted UAS produced an encouraging result. The values for the distance between the two repetitions are not homogeneously distributed, but are maximal at values that are multiple of 10.5 (Fig. 5). In other words they are placed in the same face of the DNA. This is a known property of experimentally determined UAS (Pérez-Martín and de Lorenzo, 1996), but was not included in the algorithm used. We then combined both promoter score and UAS score in a unified score, calculated as  $S_p/12 + S_u$  where  $S_p$  is the score of the promoter (in which 12 bases are considered) and  $S_u$  is the score of the UAS. In order to set the appropriate threshold and to test the predictive power of this new score, we used the sensitivity analysis, and the results are shown in Fig. 3. If compared with the scores obtained based only on the promoter sequence, it can be noted that prediction quality improved significantly, increasing the sensitivity from 52% to 64% (measured over our artificial positive set). A threshold of 2.65 was selected, which allowed us now to

predict up to 31 promoters with minimal ratio of false positives. Again the highest scoring promoter in the incorrect orientation scored 2.60. The final list of our predicted promoters is summarized in Table 1. These promoters will be discussed in the following sections in the broader context of their biological functions.

#### Flagella and motility

A *P. putida* KT2440 mutant strain lacking  $\sigma^{54}$  is known to be unable to produce flagella (Kohler *et al.*, 1989). Here, three high scoring promoters were found in front of three flagellar related genes: PP4391 (the *flgBCDE* operon), PP4370 (the *fliEFGHIJ* operon) and PP4372 (the *fleSR* operon). In addition to these three, sequence strongly resembling  $\sigma^{54}$  promoters (score >13) were found also in



**Fig. 5.** Phasing of UAS in canonical  $\sigma^{54}$ -dependent promoters. The activator-binding sites tend to be separated by a exact number of helix turns. We calculated the offset from an exact number of helix turns (assuming 10.5 bases per turn) for the distance between the two repetitions of the predicted UAS. Note that whereas 65% of the UAS in the correctly oriented promoters have an offset smaller than two bases with an almost normal distribution centred at 0, this is not the case in the incorrectly oriented promoters.



**Table 1.** Genes downstream of the predicted promoters scoring over the selected threshold.

Gene	Score	Description
PP5046	3.086	glutamine synthetase, type I
PP4867	3.007	branched-chain amino acid ABC transporter, periplasmic binding protein
PP0556	3.001	acetoin catabolism protein
PP4391	2.972	flagellar basal-body rod protein FlgB
PP2094	2.931	nitrate-binding protein NasS, putative
PP1833	2.916	conserved hypothetical protein
PP1703	2.856	assimilatory nitrate reductase/sulfite reductase, putative
PP3931	2.853	transporter, sodium/sulphate symporter family
PP4370	2.829	flagellar hook-basal body complex protein FlIE
PP3184	2.825	hypothetical protein
PP2106	2.822	ammonium transporter, putative
PP3463	2.816	aldehyde dehydrogenase family protein
PP4309	2.811	transporter, NCS1 nucleoside transporter family
PP1188	2.799	C4-dicarboxylate transport protein
PP3655	2.794	cytosine/purine/uracil/thiamine/allantoin permease family protein
PP1297	2.788	general amino acid ABC transporter, periplasmic binding protein
PP3420	2.777	sensor histidine kinase
PP0545	2.754	aldehyde dehydrogenase family protein
PP0052	2.725	metallo-beta-lactamase family protein
PP3009	2.723	hypothetical protein
PP3950	2.728	methyl-accepting chemotaxis transducer
PP2092	2.715	nitrate transporter
PP3120	2.709	oxidoreductase, aldo/keto reductase family
PP4372	2.696	sensory box histidine kinase FleS
PP3947	2.689	isoquinoline 1-oxidoreductase, putative
PP1071	2.686	amino acid ABC transporter, periplasmic amino acid-binding protein
PP2810	2.684	conserved hypothetical protein
PP1726	2.672	ABC transporter, periplasmic binding protein
PP2764	2.669	NADH-dependent FMN reductase MsuE
PP2774	2.663	glycine betaine/L-proline ABC transporter, ATPase/permease fusion protein
PP3196	2.659	nodulation protein d1

front of PP4394 (*flgA*), PP4344 (*flhA*) and PP4386 (the *flgFGHIJKL* operon), and weak matches (score >10 and <13) in front of PP4352 (*flhB*), PP4361 (the *fliKLMNOPQR* operon) and PP4376 (*fliD*). In *Pseudomonas aeruginosa* all these genes are known to be regulated by FleQ, a  $\sigma^{54}$ -dependent regulator, either directly (Totten *et al.*, 1990; Arora *et al.*, 1997, 1998; Jyot *et al.*, 2002) or through the action of FleR, also a  $\sigma^{54}$ -dependent regulator. The activity of FleQ is controlled by the FleN (Dasgupta and Ramphal, 2001) whereas FleR is believed to be controlled by FleS. Orthologues to FleQ and FleR are present in the *P. putida* genome, and also of the regulatory proteins FleN and FleS.

#### Nitrogen metabolism

We have found potential  $\sigma^{54}$ -dependent promoters upstream of several genes related to nitrogen metabolism. The highest scoring promoter in our screen was PP5046 (*glnA*) coding for the glutamine synthetase. This gene is one of the paradigms in  $\sigma^{54}$ -dependent transcription. This gene is in the same operon with the *ntrB* (PP4057) and *ntrC* (PP5048) genes and in *E. coli* its transcription is regulated by the NtrC protein. The genes regulated by

NtrC have been recently determined in *E. coli*, by the comparison of the transcription profiles of two strains, one lacking the *ntrC* gene, and other carrying an constitutive allele of the regulator (Zimmer *et al.*, 2000). In order to establish the extent of the NtrC regulon in *P. putida*, we tried to identify orthologues to those of the NtrC regulon in *E. coli*. We could not identify clear orthologues for several of them. For those that we could, we found  $\sigma^{54}$ -dependent promoters in front of the *yhdWXYZ* orthologues (PP1297-PP1300), and the *glnK-amtB* operon (PP5234-PP5233), although in this later case with a score below our selected threshold. Additional weak promoters (score >10 and <13) were found also in front of *argT* (PP4486, score 12.33), the *dppABCDBE* operon (PP0885, score 10.33) and the *yeaGH* operon, although in this case only the *yeaH* orthologue (PP0396, score 12.76) seem to be transcribed from a  $\sigma^{54}$ -dependent promoter. An interesting case is the *gltIJKL* operon, which codes for a glutamate transporter. The *P. putida* orthologues (PP1071-PP1068) are very close to, and possibly co-transcribed with, a sensor-histidine kinase and a  $\sigma^{54}$ -dependent regulator (PP1067, PP1068), suggesting that in *P. putida* this operon might be regulated by a mechanism not involving NtrC. The case of the orthologues for the *potFGHI* operon (PP5181-PP5177) is also noteworthy.

## 6 I. Cases, D. W. Ussery and V. de Lorenzo

These genes, which code for a putative putrescine transporter, could be co-transcribed with the PP5182 and PP5183 genes, which code for a putative aminotransferase and putative glutamine synthetase. Interestingly a weak  $\sigma^{54}$  promoter could be found in front of the first gene of the cluster PP5183 (score 10.83). Finally we could not detect any  $\sigma^{54}$ -dependent promoter in the upstream regions of the orthologues for the *astCADBE* operon (PP4481-PP4475).

Three genes involved in nitrate metabolism were also identified as  $\sigma^{54}$ -dependent: the PP2092-PP2090 operon, coding for a nitrate transporter, a serine/threonine kinase and a uroporphyrin III-methyl-transferase, the PP2094-PP2093 coding for a nitrate binding protein and a transcription antiterminator, and PP1703-PP1706 encoding a nitrate reductase and the two subunits of the nitrite reductase. Significantly,  $\sigma^{54}$  lacking mutants of *P. putida* are unable to grow on nitrate as the only nitrogen source (Kohler *et al.*, 1989). Nitrate assimilation is known to be  $\sigma^{54}$ -dependent in several *Pseudomonas* (Totten *et al.*, 1990; Hartig and Zumft, 1998) and *Ralstonia* (Zumft, 2002) species.

Several other promoters scoring above the selected threshold were found in front of nitrogen related genes. These include: the operon PP4867-PP4863 which encodes the five subunits of a high affinity transporter for branched amino acids; the PP3655 gene, encoding a protein of the cytosine/purine/uracil/thiamine/allantoin permease family; the PP4309 gene, a member of the NCS1 nucleoside transporter family, which is probably co-transcribed with PP4310, an hydantoin racemase; PP2106 determining a putative ammonia transporter; and finally, the PP2774-PP2276 cluster, encoding two subunits of a glycine betaine/L-proline ABC transporter (perhaps involved in osmotic adaptation) and one protein belonging to the family of homocysteine S-methyltransferases. In addition, we found a strong match in front of PP3196. This gene determines a transcriptional regulator of the LysR family with strong similarity with the Nodulation protein 1. This protein is involved in the regulation of genes related with the symbiotic nitrogen fixation of rhizobia, and can bind flavonoids as inducers (Honma *et al.*, 1990). Its function in *P. putida*, a non-nodulating bacteria, is unknown, but it could have a role in the interaction of this organisms with the plant rhizosphere (Espinosa-Urgel *et al.*, 2002).

### Carbon metabolism

We found  $\sigma^{54}$ -dependent promoters in front of several dehydrogenases and oxidoreductases, like PP3463 or PP0545, both belonging to the aldehyde dehydrogenase family; the PP3120, which belongs to the oxidoreductase, aldo/keto reductase family; or the PP2764-PP2770 operon that contains the genes for one NADH-dependent

FMN reductase, one sulphonate monooxygenase, one protein related to the long-chain-fatty-acid-CoA ligases and the three subunits of a putative branched-chain amino acid ABC transporter. The first two proteins of this operon are very similar to the *sfnE* and *sfnC* genes of *P. putida* strain DS1, which are responsible for the catabolism of dimethyl sulphide in this strain. In *P. putida* strain DS1, the *sfnE* and *sfnC* genes are co-transcribed along with *sfnR*, the gene encoding the cognate  $\sigma^{54}$ -dependent activator of the whole operon (Endoh *et al.*, 2003). This is not identical to the situation in *P. putida* KT2440: our prediction identifies the presence of a  $\sigma^{54}$  promoter in front of the operon that contains the *sfnE* and *sfnC*, but these genes lack an adjacent *sfnR*-like regulator.

Another gene cluster driven by a predicted  $\sigma^{54}$  promoter is the PP3947-PP3948 operon. This includes a putative isoquinoline 1-oxidoreductase alpha subunit and a putative fusion protein containing an isoquinoline 1-oxidoreductase beta module and an alcohol dehydrogenase protein. In addition an orthologue of the dicarboxylate transporter DctA (PP1188) was also found downstream of a predicted  $\sigma^{54}$  promoter. The expression of this gene is known to be  $\sigma^{54}$ -dependent in several other bacteria such as *Rhizobium meliloti* or *Rhizobium leguminosarum* (Ledebur *et al.*, 1990). Finally an operon for acetoin metabolism was also found (PP0556-PP0550), almost identical to the one defined in *P. putida* PpG2 (Huang *et al.*, 1994). This operon is transcribed from a  $\sigma^{54}$ -dependent promoter in *Ralstonia eutropha* (Kruger and Steinbuchel, 1992) and *B. subtilis* (Ali *et al.*, 2001).

### Other functions

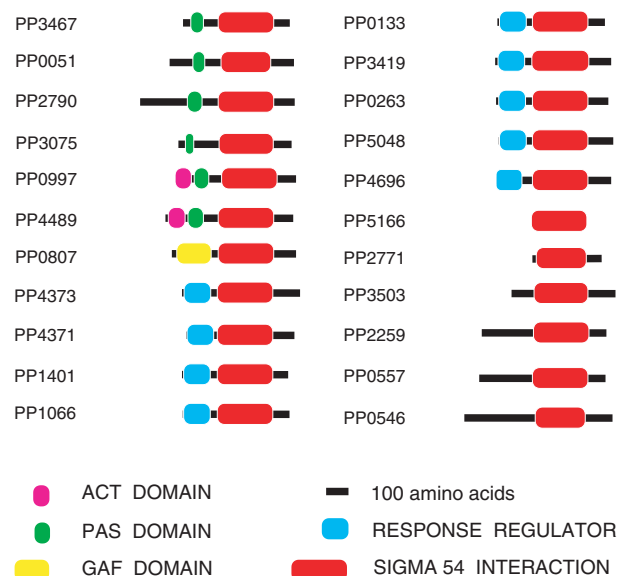
We also found  $\sigma^{54}$  promoters in front of genes of which the function is very difficult to assess. These include the operon formed by PP1726-PP1721 including four subunits of an ABC transporter of unknown substrate, one hypothetical protein and a member of the haloacid dehalogenase-like family which includes L-2-haloacid dehalogenases, epoxide hydrolases and phosphatases. Also in this group is the PP3931 gene which codes for a member of the sodium/sulphate symporter family and the PP0052, a member of the metallo-beta-lactamase family which apart from the beta-lactamases also includes thiolesterases and members of the glyoxalase II family. Two genes with regulatory potential, a sensor histidine kinase (PP3420), and methyl-accepting chemotaxis transducer (PP3950), were also found to be transcribed from  $\sigma^{54}$ -dependent promoters. Four additional promoters were found upstream of genes without similarity to any functionally characterized protein. The PP2810-PP2813 includes three hypothetical proteins and a putative transporter. The PP1833, PP3009 and PP3184 genes seem to be transcribed independently.

$\sigma^{54}$ -dependent systems of *P. putida* 7 $\sigma^{54}$  promoters of the TOL plasmid pWW0

Despite its extrachromosomal location, the TOL plasmid pWW0, encoding among others a suite of genes (the *xyl* genes) determining enzymes for biodegradation of toluene, *m*-xylene and *p*-xylene (Assinder and Williams, 1990) can be considered as related to the *P. putida* KT2440 genome. This is because the KT2440 strain is in fact a derivative of the natural isolate *P. putida* mt-2 cured of the pWW0 plasmid (Greated *et al.*, 2002). Furthermore, pWW0 plasmid bears two well characterized  $\sigma^{54}$ -dependent promoters: *Pu*, driving expression of the operon led by *xylU*, and *Ps*, transcribing *xylS* (Marques and Ramos, 1993). The recent completion of the plasmid sequence (Greated *et al.*, 2002) thus offers an opportunity to match our predictive method with experimental data. By applying a permissive score of 12 to the search for potential  $\sigma^{54}$ -RNAP binding sites, we found seven genes with suitable upstream regions: *rep* (12.03), ORF35 (13.66), ORF44 (12.62), *xylR* (12.53), *xylS* (12.76), *xylU* (18.93) and ORF145 (12.42). Yet, only two of them (ORF35 and *xylU*) passed the score  $\geq 13$  threshold that we employed in the genome-wide search above, whereas *xylS* was close but below the limit. Search of UAS through the upstream regions of ORF35, *xylS* and *xylU* somewhat refined the global scores for each promoter. ORF35 had a score of 1.38 for the best UAS and the same (1.38) for the second best, whereas *xylS* and *xylU* had 1.51/1.28 and 1.41/1.36 for the same figures. The best composed promoter/UAS scores were then 2.52 for ORF35, 2.57 for *xylS* (*Ps*) and 2.99 for *xylU* (*Pu*). It is interesting to note that the best score for the UAS of *Ps* matched those sequences determined experimentally, whereas those of *Pu* corresponded to the second best score in the prediction. As the threshold for the genome search was 2.65, a naïve analysis of the sequence would have qualified *Pu* as a clear  $\sigma^{54}$ -promoter but would not have picked *Ps* as a leading choice. This suggest that our search criteria is quite stringent and, if anything, may be inclined to disregard false negatives.

Prokaryotic enhancer binding proteins (EBPs) encoded in the *P. putida* genome

One of the characteristics of the  $\sigma^{54}$ -dependent promoters is the strict requirement of an activating transcription factor. Thus, identifying these proteins is also important for the proper description of the  $\sigma^{54}$  regulon in the context of a genome. Using Hidden Markov Models, we were able to identify 22 proteins with  $\sigma^{54}$  activator domains in the chromosome of *P. putida* (Fig. 6). This is a high number compared with the 13 found in *E. coli* and identical to the 22 found in *P. aeruginosa* (Studholme and Dixon, 2003). Most members of the  $\sigma^{54}$ -dependent activators family



**Fig. 6.** Domain structure of the  $\sigma^{54}$ -dependent activators in the *P. putida* genome. The type of domains is indicated. All proteins are drawn at the scale indicated.

share a common domain structure (Fig. 1), that includes three domains (Morett and Segovia, 1993): a C-terminal domain or DNA binding, a central domain, very conserved along the whole family, and responsible for the interaction with the  $\sigma^{54}$  sigma factor, and an N-terminal domain, or A module, that normally regulates the activity of the central domain (Shingler, 1996). The A domain is the most variable part, and the different  $\sigma^{54}$ -dependent regulators can be classified according to their A domain (Studholme and Dixon, 2003). We found mainly four different domain arrangements in the *P. putida* genome: those with a PAS domain, either alone or in combination with a ACT domain; a GAF domain; those that also have a two-component-response-regulator (RR) domain; and those without known domains in the N-terminal of the protein.

PAS domains are present in many signalling proteins where they are used as signal sensor domains. They appear in archaeobacteria, bacteria and eukaryotes, and some are known to detect their signal by way of an associated cofactor such as haeme or flavin chromophores (Taylor and Zhulin, 1999). Six  $\sigma^{54}$ -dependent activators presented PAS domains in *P. putida*: PP3467, PP0051, PP2790, PP3075, PP0997 and PP4489. The last two also contains an ACT domain, a module linked to a wide range of metabolic enzymes that are regulated by amino acid concentration (Aravind and Koonin, 1999). Interestingly, PP0997 and PP4489 are very similar to the TyrR protein of *E. coli*. This is an unusual member of the EBP family, as it activates several  $\sigma^{70}$  promoters involved in the degradation of aromatic amino acids, but in a fashion which does not depend on ATP hydrolysis (Pittard and Davidson,

## 8 I. Cases, D. W. Ussery and V. de Lorenzo

1991). But the same TyrR also represses transcription in an ATP-dependent manner. The TyrR protein of *E. coli* is unresponsive to  $\sigma^{54}$  because it lacks the GAFTGA motif required for the interaction with this sigma (Pittard and Davidson, 1991). Despite its close similarity to TyrR, PP0997 has an intact GAFTGA motif, which suggest it is competent for  $\sigma^{54}$  activation. PP4489 presents a partially conserved GAFEGA motif, the function of which is difficult to assess. Its closest homologue is the PhhR protein from *P. aeruginosa*, which activates the divergently transcribed gene for the phenylalanine hydroxylase (Song and Jensen, 1996). In *P. putida*, a similar chromosomal context is found, and indeed a gene encoding a phenylalanine hydroxylase is located divergently from the PP4489 gene. However we could not detect any sequence resembling a  $\sigma^{54}$ -dependent promoter in this intergenic region.

GAF domains are present in several  $\sigma^{54}$ -dependent activator, such as NifA in *Azotobacter vinelandii* or FlhA in *E. coli* (Studholme and Dixon, 2003). We found only one  $\sigma^{54}$ -dependent activator in *P. putida* harbouring a GAF domain, PP0807. This protein is very similar to the *E. coli* YgaA protein, responsible for the activation of a nitric oxide detoxification system (Hutchings *et al.*, 2002). Interestingly, there is a gene (PP0808) coding for a flavohaemoprotein very similar to some nitric oxide dioxygenases, which is divergently transcribed from PP0807. A weak  $\sigma^{54}$  promoter (score 10.79) can be detected just upstream of this gene. A similar system for the reduction of NO has been found in *Ralstonia eutropha*, and the genes involved are expressed from  $\sigma^{54}$ -dependent promoters (Pohlmann *et al.*, 2000).

We found that nine  $\sigma^{54}$ -dependent activators in *P. putida* also contain a two-component-response-regulator domain. These proteins are activated by phosphorylation of a conserved aspartate residue by a histidine kinase protein in response to a signal (Hoch, 2000). This is probably the most common A domain in  $\sigma^{54}$ -dependent activators, and it is present in the paradigmatic NtrC activator (Keener and Kustu, 1988). Seven out of nine of these proteins were found to be co-transcribed with a sensor histidine kinase, suggesting that there is a strong selective pressure to maintain these pairs of proteins under the same regulation. Among this kind of activators we can find PP4373, the *P. putida* orthologue of the FleQ regulator. Although a response regulator domain can be detected in this protein, it cannot be phosphorylated, as it lacks the conserved aspartate residue, but instead its activity is regulated by the FleN protein (Dasgupta and Ramphal, 2001). PP4373 is indeed one of the two response-regulator  $\sigma^{54}$  activating proteins that are not co-transcribed with a sensor histidine kinase.

Finally six  $\sigma^{54}$ -dependent activators do not show any recognizable A domain. Two cases, PP5166 and PP2771, actually only have around 20 residues upstream the cen-

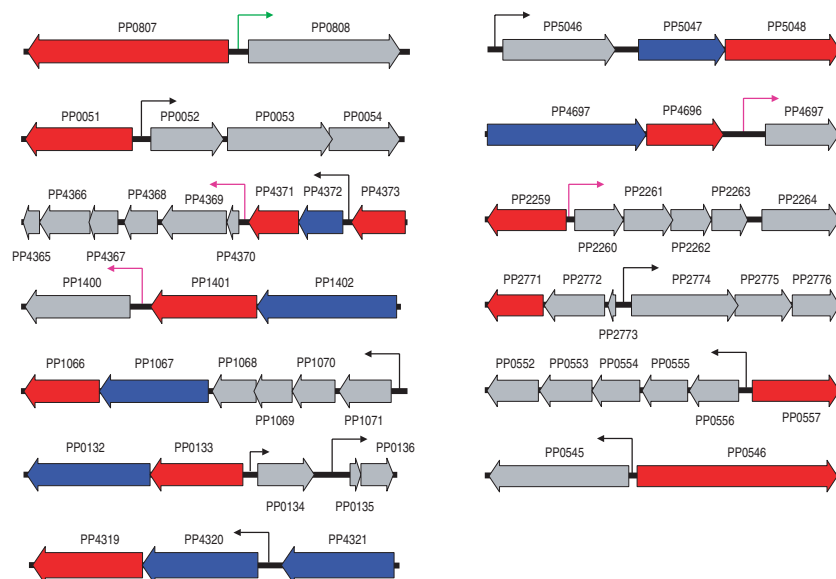
tral domain, in a similar way to the *E. coli* PspF (Jovanovic *et al.*, 1997) or the *P. syringae* HrpR/S proteins (Grimm and Panopoulos, 1989). These two regulators are negatively controlled by the PspA (Elderkin *et al.*, 2002) and HrpV (Preston *et al.*, 1998) proteins, respectively, but they do not share any sequence similarity (Studholme and Dixon, 2003). PP5116 is also unique in that it lacks a DNA binding domain.

There are several other alternative A domains in  $\sigma^{54}$ -dependent activators, such as the V4R domain, present in DmpR and XylR proteins (Shingler, 1996), or CBS domains present in the PrdR protein of *Clostridium sticklandii* (Kabisch *et al.*, 1999). In addition, PTS (phosphoenolpyruvate-dependent phosphotransferase system) regulation domains are also found associated with  $\sigma^{54}$  interaction domains (Studholme, 2002), like in *B. subtilis* LevR protein (Martin-Verstraete *et al.*, 1998). None of these variants were found in the *P. putida* genome.

Another interesting property of the  $\sigma^{54}$ -dependent activators is their tendency to map close to their target promoters (Studholme, 2002). That is indeed the case for a majority of the *P. putida* regulators. Fourteen of the 22 are associated with a  $\sigma^{54}$ -dependent promoter (Fig. 7), although not all of these promoters passed our very restrictive threshold. Among these, we have found three typical organizations. The most frequent one is to find a  $\sigma^{54}$ -dependent promoter just upstream of the activator and pointing in the opposite direction, like in the case of PP0807, PP0051, PP0133, PP2259, PP0557 and PP0546. This type of arrangement is also present in several other  $\sigma^{54}$ -dependent activators, like the case of the XylR protein of the pWW0 plasmid of *P. putida* (Pérez-Martín and de Lorenzo, 1996). In this case the same activator binding sites are responsible of activate one of the target promoter, *Ps*, and repress its own transcription (Marques and Ramos, 1993; Bertoni *et al.*, 1998). A variant of this structure is the case of PP2771, in which the activator is included in an operon that contains two other genes, and the target  $\sigma^{54}$ -dependent promoter is placed divergently to this one.

A second distinctive organization that of the cases in which the activator is included in the same operon that it activates, namely PP1066, PP4319 and PP5048. Finally, a third possibility is that the target promoter is located immediately downstream of the activator, like in the case of PP4373, PP4371, PP1401 and PP4696. Thanks to the chromosomal association of activators and promoters we can confidently add four additional promoters to our prediction, despite of their low score. These are located in front of the PP0808 gene (score 10.79) coding for a flavohaemoprotein and discussed above, the PP1400 gene (score 15.73), coding for a dicarboxylate transporter, the PP4697 gene (score 15.89), which codes for a putative poly A polymerase very similar to the *pcnB* gene in *E. coli*,





**Fig. 7.** Chromosomal context of  $\sigma^{54}$ -dependent activators in the *P. putida* genome. The chromosomal context of those  $\sigma^{54}$  activators which have an associated  $\sigma^{54}$  promoter is shown here. Genes encoding activators are coloured in red, whereas sensor histidine kinases are represented in blue. Promoters are represented as small arrows and coloured black if they score over the threshold, magenta if they score between 13 and 15 and green if they scored between 10 and 13. Each region is drawn at scale (which may vary between the chromosomal segments).

and the PP2260 (score 14.58) which is in fact part of an operon coding for the five subunits of a putative sugar ABC transporter.

## Discussion

A updated compilation of experimentally determined and putative  $\sigma^{54}$  promoters of various bacterial species (based mostly on matches to the  $-12/-24$  region) can be accessed in the URL <http://www.promscan.uklinux.net>. In this work we have characterized the  $\sigma^{54}$  regulon (sigmulon) of *P. putida*. Using a computational prediction that combines  $-12/-24$  sequence motifs and activator binding sites we have been able to detect with high confidence 31  $\sigma^{54}$ -dependent promoters. We advocate that combining searches for various DNA motifs which bear functional significance does increase significantly the quality of promoter predictions. In fact, there is still room for the improvement of our algorithm. Although we have only scored UAS-like sequences in the upstream region of  $\sigma^{54}$  promoters, there are known cases in which they are located in the 5'-non-translated regions of the  $\sigma^{54}$  transcribed genes (Jyot *et al.*, 2002). Moreover, given the molecular mechanism underlying  $\sigma^{54}$ -dependent transcription, which implies the looping of the intervening DNA between the RNA polymerase and the activator binding sites, many  $\sigma^{54}$  promoters require the participation of the IHF factor. As not all the  $\sigma^{54}$  promoters require IHF, we anticipate that searching for IHF consensus sequences around  $\sigma^{54}$  predicted promoter could increase the sensitivity of the searches, but at the same time would discard some true positives. However, it could be used as a reassuring indication for low scoring true positives. Even in the cases in which IHF is not required, specific properties in

the intervening DNA related to its *bendability*, and even in some cases static DNA bends can be observed (Pérez-Martín and de Lorenzo, 1997). Computer algorithms to recognize these properties are available (Shpigelman *et al.*, 1993; Brukner *et al.*, 1995) and could also be used for improving predictions of  $\sigma^{54}$ -dependent systems. One uncertainty of this class of promoters is that, in some cases,  $\sigma^{54}$ -RNAP binding sites are not intended to promote transcription, but on the contrary, to repress overlapping  $\sigma^{70}$  or  $\sigma^E$  promoters. Typical examples include the promoter of the *rpoN* gene (that encodes the  $\sigma^{54}$  protein itself), or the *algD* gene in *P. aeruginosa* (Kohler *et al.*, 1989; Boucher *et al.*, 2000). We found the known  $\sigma^{54}$  binding site in front of the *rpoN* gene, although it scored under our selected threshold. It is possible that some of our predicted promoters are in fact binding sites for  $\sigma^{54}$  in its repressor facet.

Chromosomal context and the similarity with other species allowed us to include another 15 promoters as highly confident predictions (Table 2), reaching a grand total of 46, or approximately 83% of the around 55 estimated total. The predicted promoters cover a variety of biological functions. Nine of them are related with flagella synthesis in agreement with the observation that an  $\sigma^{54}$  mutant was unable to form flagella, and resembling the situation in the close relative *P. aeruginosa*, where the regulation of flagella synthesis has been studied to a much greater extend. We also found many  $\sigma^{54}$ -dependent promoters leading to the expression of nitrogen metabolism, 18 in total. Significantly, many of those were involved in transport of amino acids, ammonia and polyamines, a situation similar to *E. coli* (Reitzer and Schneider, 2001). Nitrogen related functions also include the assimilatory nitrate reductase, nitrite reductase and a transport system for

**Table 2.** Genes scoring less than the selected threshold but with another supporting evidences.

Gene	Description	Evidence <sup>a</sup>
PP4394	flagella basal body P-ring formation protein FlgA	H
PP4344	flagellar biosynthetic protein FlhA	H
PP4386	flagellar basal-body rod protein FlgF	H
PP4352	flagellar biosynthetic protein FlhB	H
PP4361	flagellar hook-length control protein FliK	H
PP4376	flagellar cap protein FliD	H
PP5234	nitrogen regulatory protein P-II	H
PP4486	basic amino acid ABC transporter, periplasmic basic amino acid-binding protein	H
PP0885	dipeptide ABC transporter, periplasmic peptide-binding protein	H
PP0396	conserved hypothetical protein (YeaH orthologue)	H
PP5183	glutamine synthetase, putative	H
PP0808	flavo-haemoprotein	R
PP1400	dicarboxylate MFS transporter	R
PP4697	poly(A) polymerase	R
PP2260	sugar ABC transporter, ATP-binding protein	R

a. H: based on homology to known  $\sigma^{54}$ -dependent genes in closely related species. R: promoter located close to a  $\sigma^{54}$ -dependent activator.

nitrate. Again, this is consistent with the inability of  $\sigma^{54}$  mutants to grow on nitrate as the only nitrogen sources (Kohler *et al.*, 1989) and similar to that observed in other species as *Ralstonia* sp. (Zumft, 2002). Nine predicted promoters also drive the transcription of carbon metabolism related functions, and can be divided into those related with transport, particularly the transport of dicarboxylates, and a number of aldehyde dehydrogenases and oxidoreductases. The 10 remaining promoters are in front of genes either not included in the previous categories or for which the sequence do not offer enough information for a reliable functional assignment. Included here, there are four hypothetical proteins that have no functionally characterized homologue. This range of biological processes regulated by  $\sigma^{54}$  seems to be very similar to those reported in other global analyses performed in *E. coli* and *Salmonella typhimurium* with the exception of the flagellar machinery, and the abundance of carbon metabolism related genes (Reitzer and Schneider, 2001; Studholme, 2002), possibly reflecting the metabolic versatility of pseudomonas species.

We also performed an analysis on the  $\sigma^{54}$ -dependent regulators. Twenty-two  $\sigma^{54}$ -dependent regulators were found, a number identical to those found in *P. aeruginosa* and much larger than those found in *E. coli* or *S. typhimurium* (Studholme, 2002). Despite this difference, the ratio between regulators and promoters seems to be very similar. We have estimated around 55 promoters and found 22 regulators, that is 2.5 promoters per regulator, whereas in *E. coli* around 30 promoters have been estimated and 13 regulators found (Studholme and Dixon, 2003), i.e. 2.3 promoters per regulator. In the case of *E. coli*, the influence of  $\sigma^{54}$ -dependent activator is amplified by the cascade establish between NtrC and the transcription activator Nac, resulting in effects on as many as 75 genes distributed in 25 operons (Zimmer *et al.*, 2000). Although we could not find an *nac* orthologue in the *P. putida* genome,

we have found at least one transcriptional regulator that is transcribed from a  $\sigma^{54}$ -dependent promoter, NodD1, suggesting that the role of  $\sigma^{54}$  in *P. putida* could be amplified in a similar manner. But given that the target promoters of NodD1 are unknown this is difficult to assess.

Interestingly 17 of the 22 activators were associated in the chromosome to a  $\sigma^{54}$ -dependent promoter, something similar to that previously observed in *S. typhimurium* (Studholme, 2002). As we discussed before, some of these regulators activates more promoters than the one they are associated with, meaning that close proximity is not needed for the activation to take place, so an alternative explanation is needed. One possibility can be found in the chromosomal arrangements present in the *P. putida* genome. These arrangements will allow the coordinated expression of the regulators and the regulated genes. However two different modes of activation can be distinguished. Whereas in the case of divergent promoters, the binding of the regulator to its activating sites also represses its transcription, and thus keeps its level to a minimum enough for activation to take place, in the case of regulators that are in the same operon they activate, a positive loop is established leading to a amplified response. The maintenance of these *modes* could be the selective force that keeps this association.

Once the promoter/regulator association is guaranteed, regulators can be engaged in the activation of other promoters in the genome. How has this process occurred? One possibility is that originally every  $\sigma^{54}$  promoter had an associated activator, but as a result of spurious unspecific activation by other regulators some genes coding similar biological functions fell under the control of a single regulator. That could explain the differences in regulation between the *E. coli* *gltJJKL* operon and its *P. putida* equivalent. Whereas in *E. coli* this operon seems to be regulated by NtrC, in *P. putida* a regulator and its corresponding sensor histidine kinase are included in the

operon, suggesting that NtrC could have no role. That would imply that in the *E. coli* lineage, NtrC took control of the operon and the local regulator was lost whereas in the *P. putida* that did not happen. Apart from the disparities in physiology, this mechanism of transcription regulation evolution could explain some of the differences between the  $\sigma^{54}$  regulon in *P. putida* and *E. coli* and *S. typhimurium*.

## Experimental procedures

### Source sequence and general methods

The complete genome sequence of *Pseudomonas putida* KT2440 (Nelson *et al.*, 2002) and the pWW0 plasmid (Greated *et al.*, 2002) were obtained from the NCBI site, where are stored with the accession code NC\_002947 and NC\_003350 respectively. Intergenic sequences were extracted using customized Perl scripts.

### Promoter prediction and UAS prediction

$\sigma^{54}$  promoters were detected using a variation of the relative entropy method (Stormo, 2000). A position weight-matrix was constructed base on the compilation made by Barrios *et al.* (1999) containing 186 promoters from several bacterial species. Log likelihood ratios were the calculated for every position considering the *Pseudomonas putida* single-nucleotide background distribution calculated for the whole chromosome. The final log likelihood matrix was then used with our GenomeMatScan to find  $\sigma^{54}$  promoters in the intergenic regions of the *Pseudomonas putida* genome. This program uses the Log Likelihood matrix to score all the words of the size of the matrix, and reports all the words scoring above a given threshold. The output includes the score, the word, the orientation and the position in the scanned sequence. For predicting the upstream activating sites (UAS) of  $\sigma^{54}$ -dependent regulators we created the UASScan program. The program uses several features of these binding sites to find the most likely UAS in the vicinity of a predicted  $\sigma^{54}$  promoter. We used UASScan to search the region between 70 and 350 nucleotides upstream the predicted promoter for two repetitions of a palindromic sequence of 16 or 17 nucleotides in length. The two repetitions were required to be between 16 and 60 nucleotides apart. Instances are scored based on the information content (Schneider *et al.*, 1986) of the four hemipalindromes. For each predicted promoter the best scoring UAS was selected. Both GenomeMatScan and UASScan are written in Perl and can be accessed in an user-friendly form at the web pages <http://caravaggio.cnb.uam.es/gmscan> and <http://caravaggio.cnb.uam.es/uasscan> respectively.

### Detection of EBPs

Putative prokaryotic enhancer-binding proteins (EBPs) were discovered using The PF00158 Pfam motif and the HMMSEARCH program [<http://hmmer.wustl.edu/>], on all the ORFs of the *Pseudomonas putida* genome, using  $10^{-5}$  and threshold. Blast searches against SWISSPROT v.40 (Boeckmann *et al.*, 2003) were used for further characterization of

the proteins. Hmmpfam [<http://hmmer.wustl.edu/>] searches were also used for detection of other motifs in the putative EBP. Unless stated otherwise, functional assignments were taken from the NCBI GenBank file.

## Acknowledgements

Authors thank Chris Workman, Lars Juhl Jensen and Thomas Schou Larsen from the CBS (Copenhagen, Denmark) for useful comment and technical advice and David Studholme (Sanger Center, UK) for the critical reading of the manuscript during the development of this work. This work was supported by EU contracts QLK3-CT-2002-01933, QLK3-CT-2002-01923, and INCO-CT-2002-1001, by grant BIO2001-2274 of the Spanish *Comisión Interministerial de Ciencia y Tecnología* (CICYT) and by the Strategic Research Groups Program of the Autonomous Community of Madrid. I.C. is the recipient of an EMBO Long-term Fellowship.

## References

- Ali, N.O., Bignon, J., Rapoport, G., and Debarbouille, M. (2001) Regulation of the acetoin catabolic pathway is controlled by sigma L in *Bacillus subtilis*. *J Bacteriol* **183**: 2497–2504.
- Aravind, L., and Koonin, E.V. (1999) Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J Mol Biol* **287**: 1023–1040.
- Arora, S.K., Ritchings, B.W., Almira, E.C., Lory, S., and Ramphal, R. (1997) A transcriptional activator, FleQ, regulates mucin adhesion and flagellar gene expression in *Pseudomonas aeruginosa*. a cascade manner. *J Bacteriol* **179**: 5574–5581.
- Arora, S.K., Ritchings, B.W., Almira, E.C., Lory, S., and Ramphal, R. (1998) The *Pseudomonas aeruginosa* flagellar cap protein, FliD, is responsible for mucin adhesion. *Infect Immun* **66**: 1000–1007.
- Assinder, S.J., and Williams, P.A. (1990) The TOL plasmids: determinants of the catabolism of toluene and xylenes. *Adv Microbiol Physiol* **31**: 1–69.
- Barrios, H., Valderrama, B., and Morett, E. (1999) Compilation and analysis of sigma 54-dependent promoter sequences. *Nucleic Acids Res* **27**: 4305–4313.
- Bertoni, G., Marques, S., and de Lorenzo, V. (1998) Activation of the toluene-responsive regulator XylR causes a transcriptional switch between sigma 54 and sigma70 promoters at the divergent Pr/Ps region of the TOL plasmid. *Mol Microbiol* **27**: 651–659.
- Beynon, J., Cannon, M., Buchanan-Wollaston, V., and Cannon, F. (1983) The *nif* promoters of *Klebsiella pneumoniae* have a characteristic primary structure. *Cell* **34**: 665–671.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**: 365–370.
- Boucher, J.C., Schurr, M.J., and Deretic, V. (2000) Dual regulation of mucoidy in *Pseudomonas aeruginosa* and sigma factor antagonism. *Mol Microbiol* **36**: 341–351.
- Brukner, I., Sanchez, R., Suck, D., and Pongor, S. (1995)

## 12 I. Cases, D. W. Ussery and V. de Lorenzo

- Trinucleotide models for DNA bending propensity. comparison of models based on DNaseI digestion and nucleosome packaging data. *J Biomol Struct Dyn* **13**: 309–317.
- Buck, M., Gallegos, M.T., Studholme, D.J., Guo, Y., and Gralla, J.D. (2000) The bacterial enhancer-dependent sigma 54 transcription factor. *J Bacteriol* **182**: 4129–4136.
- Cases, I., de Lorenzo, V., and Ouzounis, C.A. (2003) Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol* **11**: 248–253.
- Dasgupta, N., and Ramphal, R. (2001) Interaction of the antiactivator FleN with the transcriptional activator FleQ regulates flagellar number in *Pseudomonas aeruginosa*. *J Bacteriol* **183**: 6636–6644.
- Dixon, R. (1986) The *xylABC* promoter from the *Pseudomonas putida* TOL plasmid is activated by nitrogen regulatory genes in *Escherichia coli*. *Mol Gen Genet* **203**: 129–136.
- Dombrecht, B., Marchal, K., Vanderleyden, J., and Michiels, J. (2002) Prediction and overview of the RpoN-regulon in closely related species of the Rhizobiales. *Genome Biol* **3**: 0076.
- Elderkin, S., Jones, S., Schumacher, J., Studholme, D., and Buck, M. (2002) Mechanism of action of the *Escherichia coli* phage shock protein PspA in repression of the AAA family transcription factor PspF. *J Mol Biol* **320**: 23–37.
- Endoh, T., Habe, H., Yoshida, T., Nojiri, H., and Omori, T. (2003) A CysB-regulated and sigma54-dependent regulator, SfnR, is essential for dimethyl sulfone metabolism of *Pseudomonas putida* strain DS1. *Microbiology* **149**: 991–1000.
- Espinosa-Urgel, M., Kolter, R., and Ramos, J.L. (2002) Root colonization by *Pseudomonas putida*: love at first sight. *Microbiology* **148**: 341–343.
- Greated, A., Lambertsen, L., Williams, P.A., and Thomas, C.M. (2002) Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*. *Environ Microbiol* **4**: 856–871.
- Grimm, C., and Panopoulos, N.J. (1989) The predicted protein product of a pathogenicity locus from *Pseudomonas syringae* pv. *phaseolicola* is homologous to a highly conserved domain of several procaryotic regulatory proteins. *J Bacteriol* **171**: 5031–5038.
- Hartig, E., and Zumft, W.G. (1998) The requirement of RpoN (sigma factor sigma 54) in denitrification by *Pseudomonas stutzeri* is indirect and restricted to the reduction of nitrite and nitric oxide. *Appl Environ Microbiol* **64**: 3092–3095.
- Helmann, J.D., and Chamberlin, M.J. (1988) Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**: 839–872.
- Hirschman, J., Wong, P.K., Sei, K., Keener, J., and Kustu, S. (1985) Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product is a sigma factor. *Proc Natl Acad Sci USA* **82**: 7525–7529.
- Hoch, J.A. (2000) Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* **3**: 165–170.
- Honma, M.A., Asomaning, M., and Ausubel, F.M. (1990) *Rhizobium meliloti* nodD genes mediate host-specific activation of NodABC. *J Bacteriol* **172**: 901–911.
- Hoover, T.R., Santero, E., Porter, S., and Kustu, S. (1990) The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. *Cell* **63**: 11–22.
- Huang, M., Oppermann, F.B., and Steinbuchel, A. (1994) Molecular characterization of the *Pseudomonas putida* 2,3-butanediol catabolic pathway. *FEMS Microbiol Lett* **124**: 141–150.
- Hutchings, M.I., Mandhana, N., and Spiro, S. (2002) The NorR protein of *Escherichia coli* activates expression of the flavorubredoxin gene *norV* in response to reactive nitrogen species. *J Bacteriol* **184**: 4640–4643.
- Ingmer, H., and Atlung, T. (1992) Expression and regulation of a *dnaA* homologue isolated from *Pseudomonas putida*. *Mol Gen Genet* **232**: 431–439.
- Jovanovic, G., Dworkin, J., and Model, P. (1997) Autogenous control of PspF, a constitutively active enhancer-binding protein of *Escherichia coli*. *J Bacteriol* **179**: 5232–5237.
- Jyot, J., Dasgupta, N., and Ramphal, R. (2002) FleQ, the major flagellar gene regulator in *Pseudomonas aeruginosa*, binds to enhancer sites located either upstream or atypically downstream of the RpoN binding site. *J Bacteriol* **184**: 5251–5260.
- Kabisch, U.C., Grantzdorffer, A., Schierhorn, A., Rucknagel, K.P., Andreesen, J.R., and Pich, A. (1999) Identification of D-proline reductase from *Clostridium sticklandii* as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. *J Biol Chem* **274**: 8445–8454.
- 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.
- Kohler, T., Harayama, S., Ramos, J.L., and Timmis, K.N. (1989) Involvement of *Pseudomonas putida* RpoN sigma factor in regulation of various metabolic functions. *J Bacteriol* **171**: 4326–4333.
- Kruger, N., and Steinbuchel, A. (1992) Identification of *acoR*, a regulatory gene for the expression of genes essential for acetoin catabolism in *Alcaligenes eutrophus* H16. *J Bacteriol* **174**: 4391–4400.
- Ledebur, H., Gu, B., and Sojda, J. (1990) 3rd, and Nixon, B.T. *Rhizobium meliloti* and *Rhizobium leguminosarum* *dctD* gene products bind to tandem sites in an activation sequence located upstream of sigma 54-dependent *dctA* promoters. *J Bacteriol* **172**: 3888–3897.
- Marques, S., and Ramos, J.L. (1993) Transcriptional control of the *Pseudomonas putida* TOL plasmid catabolic pathways. *Mol Microbiol* **9**: 923–929.
- Martínez-Bueno, M.A., Tobes, R., Rey, M., and Ramos, J.L. (2002) Detection of multiple extracytoplasmic function (ECF) sigma factors in the genome of *Pseudomonas putida* KT2440 and their counterparts in *Pseudomonas aeruginosa* PA01. *Environ Microbiol* **4**: 842–855.
- Martin-Verstraete, I., Charrier, V., Stulke, J., Galinier, A., Erni, B., Rapoport, G., and Deutscher, J. (1998) Antagonistic effects of dual PTS-catalysed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. *Mol Microbiol* **28**: 293–303.
- Merrick, M.J. (1983) Nitrogen control of the *nif* regulon in *Klebsiella pneumoniae*. involvement of the *ntrA* gene and analogies between *ntrC* and *nifA*. *EMBO J* **2**: 39–44.



- Morett, E., and Segovia, L. (1993) The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol* **175**: 6067–6074.
- Nelson, K.E., Weinell, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A., *et al.* (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**: 799–808.
- Pérez-Martin, J., and de Lorenzo, V. (1995) The sigma 54-dependent promoter *Ps* of the TOL plasmid of *Pseudomonas putida* requires HU for transcriptional activation *in vivo* by XylR. *J Bacteriol* **177**: 3758–3763.
- Pérez-Martín, J., and de Lorenzo, V. (1996) Physical and functional analysis of the prokaryotic enhancer of the sigma 54-promoters of the TOL plasmid of *Pseudomonas putida*. *J Mol Biol* **258**: 562–574.
- Pérez-Martín, J., and de Lorenzo, V. (1997) Clues and consequences of DNA bending in transcription. *Annu Rev Microbiol* **51**: 593–628.
- Pittard, A.J., and Davidson, B.E. (1991) TyrR protein of *Escherichia coli* and its role as repressor and activator. *Mol Microbiol* **5**: 1585–1592.
- Pohlmann, A., Cramm, R., Schmelz, K., and Friedrich, B. (2000) A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha*. *Mol Microbiol* **38**: 626–638.
- Preston, G., Deng, W.L., Huang, H.C., and Collmer, A. (1998) Negative regulation of *hrp* genes in *Pseudomonas syringae* by HrpV. *J Bacteriol* **180**: 4532–4537.
- Reitzer, L., and Schneider, B.L. (2001) Metabolic context and possible physiological themes of sigma 54-dependent genes in *Escherichia coli*. *Microbiol Mol Biol Rev* **65**: 422–444. table of contents.
- Schneider, T.D., and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**: 6097–6100.
- Schneider, T.D., Stormo, G.D., Gold, L., and Ehrenfeucht, A. (1986) Information content of binding sites on nucleotide sequences. *J Mol Biol* **188**: 415–431.
- Shingler, V. (1996) Signal sensing by sigma 54-dependent regulators: derepression as a control mechanism. *Mol Microbiol* **19**: 409–416.
- Shpigelman, E.S., Trifonov, E.N., and Bolshoy, A. (1993) CURVATURE software for the analysis of curved DNA. *Comput Appl Biosci* **9**: 435–440.
- Song, J., and Jensen, R.A. (1996) PhhR, a divergently transcribed activator of the phenylalanine hydroxylase gene cluster of *Pseudomonas aeruginosa*. *Mol Microbiol* **22**: 497–507.
- Stormo, G.D. (2000) DNA binding sites: representation and discovery. *Bioinformatics* **16**: 16–23.
- Studholme, D.J. (2002) Enhancer-dependent transcription in *Salmonella enterica typhimurium*: new members of the sigma N regulon inferred from protein sequence homology and predicted promoter sites. *J Mol Microbiol Biotechnol* **4**: 367–374.
- Studholme, D.J., and Buck, M. (2000) The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. *FEMS Microbiol Lett* **186**: 1–9.
- Studholme, D.J., and Dixon, R. (2003) Domain architectures of sigma 54-dependent transcriptional activators. *J Bacteriol* **185**: 1757–1767.
- Taylor, B.L., and Zhulin, I.B. (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* **63**: 479–506.
- Totten, P.A., Lara, J.C., and Lory, S. (1990) The *rpoN* gene product of *Pseudomonas aeruginosa* is required for expression of diverse genes, including the flagellin gene. *J Bacteriol* **172**: 389–396.
- Ussery, D., Larsen, T.S., Wilkes, K.T., Friis, C., Worning, P., Krogh, A., and Brunak, S. (2001) Genome organisation and chromatin structure in *Escherichia coli*. *Biochimie* **83**: 201–212.
- Zhang, X., Chaney, M., Wigneshweraraj, S.R., Schumacher, J., Bordes, P., Cannon, W., and Buck, M. (2002) Mechanochemical ATPases and transcriptional activation. *Mol Microbiol* **45**: 895–903.
- Zimmer, D.P., Soupene, E., Lee, H.L., Wendisch, V.F., Khodursky, A.B., Peter, B.J. *et al.* (2000) Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci USA* **97**: 14674–14679.
- Zumft, W.G. (2002) Nitric oxide signaling and NO dependent transcriptional control in bacterial denitrification by members of the FNR-CRP regulator family. *J Mol Microbiol Biotechnol* **4**: 277–286.