

Appendix 6.

Putative strong promoters *versus* experimentally verified promoters of *Escherichia coli* genome

<i>CI</i>										
gene	functional promoter identifier (*)	score (**)	strand	BACTRANS ² TSS	RegulonDB TSS	Δ_{BR}	PromEC TSS	Δ_{BP}	Δ_{RP}	
manX		1.4	forward	1899939	1899957	17	see RegulonDB	see RegulonDB	0	
otsB		1.4	reverse	1980463	1980489	25	see RegulonDB	see RegulonDB	0	
argI		1.4	reverse	4475920	4476367	447	4475907	13	460	
oxyR		1.4	forward	4156021	4156480	459	4156036	15	444	
kdpA		1.4	reverse	728127	728072	54	-	-	-	
entC		6.4	forward	624073	624054	18	-	-	-	
flgB		11.4	forward	1130089	1130216	127	-	-	-	
clpB		1.0	reverse	2732221	-	-	2732225	4	-	
infC		1.0	reverse	1798841	-	-	1799370	529	-	
phnC		1.0	reverse	4322794	-	-	4322784	10	-	
ahpC		1.4	forward	638106	-	-	638144	38	-	
ansB	P2	1.4	reverse	3098821	-	-	3098770	51	-	
ansB		1.4	" "	" "	-	-	3098809	12	-	
acnA	P1	1.4	forward	1333811	-	-	1333448	363	-	
acnA	P2	1.4	" "	" "	-	-	1333805	6	-	
frdA		11.4	reverse	4380059	-	-	4379990	69	-	
<i>CII</i>										
gene	functional promoter identifier (*)	(***)	score(**)	strand	BACTRANS ² TSS	RegulonDB TSS	Δ_{BR}	PromEC TSS	Δ_{BP}	Δ_{RP}
manX		x					17			
otsB		x					25			
argI		x							13	
oxyR		x							15	
gltB			1.8	forward	3352072	3352531	459	3352144	72	387
kdpA		x					54	-	-	-
entC		d	1.8	forward	624080	624054	26	-	-	-
flgB		d	1.6	forward	1130068	1130216	148	-	-	-
rhaS			1.6	forward	4095267	4095734	467	-	-	-
zntA			1.6	forward	3604064	3604445	381	-	-	-
cpxR			1.6	reverse	4103389	4103709	320	-	-	-
clpB		x					-	-	4	
infC		x					-	-	529	
phnC		x					-	-	10	
ahpC		x					-	-	38	
ansB		x					-	-	12	
acnA	P2	x					-	-	6	
frdA		d	1.8	reverse	4379979	-	-	4379990	11	-
clpA	P1		1.2	forward	922303	-	-	922303	0	-
clpA	P2		" "	" "	" "	-	-	922315	12	-
clpA	P3		" "	" "	" "	-	-	922433	130	-
carAB	P1		1.2	forward	29553	-	-	29551	2	-
carAB	P2		" "	" "	" "	-	-	29619	66	-
pfkB		1.8	forward	1804300	-	-	1804376	76	-	-

Table 6.1. Comparison of the transcription start site (TSS) relative to a putative strong $\sigma 70$ promoter identified by BACTRANS² in *E. coli* genome, with respect to the TSS(s) of the functional $\sigma 70$ promoter(s) harboured in the same gene. The functional promoters of *E. coli* are annotated in RegulonDB and PromEC databases. (*) When several functional promoters are known, they are denoted *P1*, *P2* and so on; (**) BACTRANS² score (see Subsection "Scoring function used"); (***) x indicates when the putative strong promoters identified under *CI* and *CII* constraints are identical; d indicates a difference between them; Δ_{BR} : bp distance between BACTRANS² TSS and RegulonDB TSS; Δ_{BP} : bp distance between BACTRANS² TSS and PromEC TSS; Δ_{RP} : bp distance between RegulonDB TSS and PromEC TSS; - indicates that the gene is not referred to in the corresponding database; low location differences (≤ 20) are highlighted in boldface characters.

Regarding *CI* constraints, 14 out of the 97 genes harbouring potentially strong $\sigma 70$ promoters are present in RegulonDB or PromEC databases. For no gene does the TSS relative to the putative promoter exactly coincide with a functional promoter referred to in RegulonDB or PromEC. Genes *manX* and *otsB* each have the same two TSSs for the functional promoter in RegulonDB and the functional promoter in EcoProm database. On the contrary, *argI* and *oxyR* TSSs differ in RegulonDB and PromEC databases. Three genes (*kdpA*, *entC*, *flgB*) are referred to in RegulonDB only. Seven genes (*clpB*, *infC*, *phnC*, *ahpC*, *ansB*, *acnA*, *frdA*) are only mentioned in PromEC database. For some genes, the TSSs relative to putative strong promoter and functional promoter are close (4 (*clpB*), 6 (*acnA*), 10 (*phnC*), 12 (*ansB*), 13 (*argI*), 15 (*oxyR*)), or relatively close (17 (*manX*), 18 (*entC*)). Two genes show a bp distance between the putative strong promoter and the functional promoter over 100: *flgB*, *infC*.

When constraints are relaxed (*CII*), 21 out of the 255 genes identified with putative strong promoters are present in RegulonDB or PromEC databases. Again, no putative strong promoter exactly coincides with a functional promoter mentioned in RegulonDB or PromEC. In addition to *manX*, *otsB*, *argI* and *oxyR* already identified under *CI* constraints, gene *gltB* is also referred to in both databases. As for *argI* and *oxyR*, the TSSs for the functional promoters of *gltB* greatly differ between RegulonDB and PromEC. In addition to *kdpA*, *entC* and *flgB* identified under *CI* conditions, three more genes identified by BACTRANS² also harbour a functional promoter only referred to in RegulonDB: *rhaS*, *zntA* and *cpxR*. Note that *entC* and *flgB* putative promoter scores are improved under the more relaxed constraints *CII*. In addition to the seven genes identified under *CI* constraints (*clpB*, *infC*, *phnC*, *ahpC*, *ansB*, *acnA* and *frdA*), three more genes are identified as harbouring a functional promoter mentioned in PromEC only: *clpAP*, *carAB* and *pfkB*. Note that *frdA* putative strong promoter has a better score under *CII* constraints than under *CI* constraints. On the other hand, this time, the bp distance between the TSS of the putative strong promoter and that of the functional promoter may be very low. For instance, this distance is null when promoter *P1* of gene *clpA* is considered (This gene has three functional promoters identified.). The distance is 2 for promoter *P1* of gene *carA*. Proximity of the putative strong promoter and the functional promoter is verified for *frdA* (11) and *entC* (26). Eleven distances are below 20. Five genes show a bp distance between the putative strong promoter and the functional promoter over 100: *flgB*, *rhaS*, *zntA*, *cpxR* and *infC*.

To recapitulate, we note that PromEC significantly contributes to the comparisons with 7 functional promoters not referred to in RegulonDB under *CI* constraints (respectively 10 under *CII* constraints). Moreover, when a gene is referred to by both databases, the two functional promoters are not always identical and may be located at remote positions. Comparing the distances between putative strong promoters and their nearest functional promoters, we observe that 8 distances out of 14 are below 20, under *CI* conditions. Eleven distances out of 21 are below 20 under *CII* conditions, including a superposition of the TSSs in one case.

Finally, none of the putative strongest promoters harbouring an UP element identified by BACTRANS² is present in RegulonDB database or PromEC database. This remark holds either under *CI* or *CII* constraint.