

Analysis of enteric bacteria SOS operator sequences and description of potential DNA damage-inducible genes*

Ana T. Vasconcelos¹, Rosane S. Machado² and Darcy F. de Almeida²

ABSTRACT

An analysis of enteric bacteria nucleotide sequences was carried out using a computer-assisted similarity search for potential SOS boxes in *Escherichia coli* and *Salmonella typhimurium*. The strategy consisted of searching the GenBank, initially for extragenic sequences containing the consensus CTG (TA)₅CAG with up to six mismatches in the central (TA)₅ decamer. Selected sequences were individually examined for their heterology index (λE) score, which indicates their relative affinity to the repressor protein, and for the presence of associated regulatory features (-10 and -35 promoter boxes, ribosome-binding sites, and a downstream open-reading frame (ORF)) suitably placed in the neighboring regions. A total of 10 newly identified sequences were selected, seven from the *E. coli* and three from the *S. typhimurium* genome; five had $\lambda E < 7.0$ and five had $15.5 > \lambda E > 10.0$. Three selected sequences (two in *E. coli*, one in *S. typhimurium*) were associated with recognizable ORFs. Others are positioned upstream of regions coding for antibiotic resistance and enterohemolysin production, while still others are ambiguously positioned relative to known coding regions, particularly between divergently transcribed genes. The results show that a search for prokaryote regulatory sequences may be useful for the identification of genes belonging to specific regulons.

INTRODUCTION

The SOS system of DNA damage repair, widely studied in *Escherichia coli*, is a set of more than twenty genes coordinately regulated by the proteins RecA and LexA (Walker, 1985, 1987; Peterson *et al.*, 1988). Central to the regulation of the SOS system is the presence of an operator site specifically recognized by the LexA

repressor protein. The LexA-binding operator DNA consensus sequence is a perfect palindrome, TACTGTATATATACAGTA (Berg, 1988). In practice, the trimer CTG and its inverted complement CAG are highly conserved (Berg, 1988; Schnarr *et al.*, 1991) while extensive variation can be found in the central (TA)₅ motif and in the two bp at the flanking ends (Wertman and Mount, 1985). Individual bp in the consensus sequence do not contribute equally to repressor recognition; thus, the affinity of the LexA protein to different operator sites varies as a function of the operator sequence. Different affinities can be quantitatively expressed by the heterology index, λE (Berg and von Hippel, 1987; Berg, 1988), which reflects the relative contribution of the operator individual bp to specify a LexA-binding site. Associated with the

* This paper is dedicated to Prof. Carlos Chagas Filho, founder of the Institute of Biophysics, on the occasion of its 50th anniversary.

¹ Laboratório Nacional de Computação Científica, CNPq 22290-160 Rio de Janeiro, RJ, Brasil.

² Laboratório de Fisiologia Celular, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, CCS, Bloco G, Cidade Universitária, 21941-590 Rio de Janeiro, RJ, Brasil. Tel.: (021) 260-6963. Fax: (021) 280-8193 ou 274-5655. Send correspondence to D.F.A.

presence of regulatory sequence features, the heterology index may help in the identification of the SOS system operators (Payne and Sancar, 1989).

Although our knowledge of the SOS regulon derives largely from the study of *E. coli*, several reports indicate that this system is widespread among Enterobacteriaceae. Prominent among these is *Salmonella typhimurium*, a close genetic relative of *E. coli* (Lodevick *et al.*, 1990; Thomas *et al.*, 1990; Smith *et al.*, 1990, 1991; Mustard *et al.*, 1992). Such facts and the current availability of DNA sequence databases prompted us to investigate the extension of that particular DNA damage-inducible system among enteric bacteria, focusing initially on *E. coli* and *S. typhimurium*. For this purpose, we used a computer-assisted similarity search of potential LexA repressor-binding sequences and associated regulatory features similar to the one recently described (Lewis *et al.*, 1994). The results reported here confirm, complement and extend their findings. Several potential SOS operator sites were identified both in *E. coli* and *S. typhimurium*. These sites are now amenable to experimental investigation in order to establish their *in vivo* functional status. A preliminary report of this work was presented at a meeting of the Brazilian Academy of Sciences (Vasconcelos *et al.*, 1994).

MATERIAL AND METHODS

Databases

GenBank version 73.0 was used as database for most of the work; a partial updating was made more recently using version 80.0. The EcoSeq6 collection of sequenced genes (kindly provided by Ken Rudd) was used to search the *E. coli* genome.

Search strategy and software

The strategy adopted for the final selection of potential LexA-binding sites was similar to the one described by Lewis *et al.* (1994), with some modifications: (i) A program was developed to search for interrupted palindromes of general structure CTG(TA)₅CAG, containing a maximum of six sites non-identical to the (TA)₅ consensus in the central decamer. Our search was restricted to the extragenic regions; peptide- and stable RNAs-coding regions were excluded (Vasconcelos, 1991). (ii) Potential LexA-binding sequences localized upstream of a recognizable open-reading frame (ORF) in either DNA strand were selected and analyzed for their position

relative to putative -10 and -35 promoter boxes using a specially devised program (Vasconcelos and de Almeida, 1994). (iii) The presence of a ribosome-binding site (RBS) and its position relative to the other sites were annotated. (iv) Individual measurements of the λE and the cut-off score were determined using the set of known LexA operators described by Berg (1988). Since the SOS-like box near *ruvAB* ($\lambda E = 16.2$) does not qualify as a functionally active site (Shinagawa *et al.*, 1988; Payne and Sancar, 1989), and the top score known for a LexA-binding site is that of *lux* ($\lambda E = 15.94$; Shadel *et al.*, 1990), sequences scoring below 15.5 were selected. Sequences found in plasmids, Tn, IS or those coding for antibiotic resistance (with a single exception, see Discussion) were excluded. Sequences are identified by their GenBank accession number and entry designation. DNA and protein homology searches were made using the GCG package program, FASTA program (Pearson and Lipman, 1988) and FRAMES.

RESULTS

SOS box-like sequences in extragenic DNA

Out of 8623 entries in the Bacteria division of GenBank version 73.0 (15.1×10^6 bp) a set of 20382 extragenic DNA sequences (4.2×10^6 bp; mean sequence length \pm standard deviation = 206 ± 478 bp) was obtained. These figures indicate a considerable dispersion of the length of sequences, which consist mostly of the short fragments found up- and downstream of the coding regions that usually constitute the sequencing target. This set was searched for the presence of sequences bearing the general structure CTG(TA)₅CAG, containing up to six mismatches in the central (TA)₅ motif. The output was 289 SOS operator-like sequences. Search of the EcoSeq6 database (442 entries; 1.82×10^6 bp) produced 1656 extragenic sequences (0.4×10^6 bp; mean sequence length 251 ± 356 bp).

Potential LexA-binding operators

Potential SOS boxes (Table I) are listed according to the number of non-identical base pairs (Nlbp) found in the central decamer motif, as compared with the consensus (TA)₅ sequence. An internal positive control of search strategy reliability was the obligatory retrieval of the operator sites upstream of the region coding for genes known to belong to the SOS regulon. All known LexA-binding sites were recovered between the 1 and 5 Nlbp levels, except for the *E. coli* *sosE* locus,

Table I - SOS box-like sequences found in the *Escherichia coli* and *Salmonella typhimurium* genomes.

No.	Ac. no. (ent.)*	NI**	Sequence***	λE ****	Organism
1	X03928 (STYPYRC)	1	AACTGTATATATATCCAGTA	1.70	<i>S. typhimurium</i>
2	U00009 (ECOHU43)	2	TACTGTATATAAAAACAGTA	1.96	<i>E. coli</i>
3	L10328 (ECOUW82)	2	TACTGTTTATTATACAGTA	2.81	<i>E. coli</i>
4	M60066 (STYRFC)	2	CACTGTATATAAACACAGTA	3.86	<i>S. typhimurium</i>
5	L20897 (ECOGARA)	4	TACTGTATAAAATCACAGTT	6.66	<i>E. coli</i>
6	L08613 (STYPEFABCD)	4	CACTGGTTATGGATACAGTT	10.18	<i>S. typhimurium</i>
7	Y00452 (ECAPH)	4	GTCTGCTTACATAAACAGTA	12.10	<i>E. coli</i>
8	D13169 (ECORBAB)	5	TACTGTCTACCAAAAACAGAG	14.49	<i>E. coli</i>
9	X70047 (ECEHLYL)	6	GACTGGGCACGTAACAGGA	12.97	<i>E. coli</i>
10	M13608 (ECOPHOM) & M97495 (ECOROBPHOM)	6	TGCTGTTTACATTACAGTG	13.81	<i>E. coli</i>

* GenBank accession number and entry designation (within parentheses).

** Number of non-identical base pairs in the central decamer motif, compared with the consensus (TA)₅ sequence.

*** The conserved trimer palindrome (CTG/CAG) is shown in bold face.

****Heterology index (Berg, 1988).

which was not retrieved (Lewis *et al.*, 1994; see Discussion) and the *Vibrio fischeri lux* operon box (Shadel *et al.*, 1990), which was recovered at the 6 N1bp level. No case of 100% identity with the consensus sequence was found, as noted by Schnarr *et al.* (1991). The searches revealed cases of redundancy in the databases: identical sequences were found in independent entries X55791 (ECOMRPMET) and U00007 (ECOHU47), while entries M13608 (ECOPHOM) and M97495 (ECOROBPHOM) have the region containing the potential SOS box independently sequences from different ends (Table I).

A total of 10 new potential sequences (seven for *E. coli* and three for *S. typhimurium*) were found (Table I).

Selected sequences (Table I) were classified into two groups, based on the λE scores: the first, with $\lambda E < 7.0$, includes five sequences (three from *E. coli* and two from *S. typhimurium*); the other, with $15.5 > \lambda E > 10.0$, includes the remaining five (four from *E. coli* and one from *S. typhimurium*).

High score ($\lambda E < 7.0$) boxes

In the first group, the highest scoring sequence ($\lambda E = 1.70$, Table I) was found in a small (202 bp) DNA

fragment downstream of the *S. typhimurium pyrC* gene, under entry X03928 (STYPYRC). It is homologous to the *E. coli dinI* locus (ECOPYRC; Lewis *et al.*, 1994), not only on account of their similar positions in the genome, but mainly because FASTA showed a 79.8% homology to ECOPYRC over 109 bp at the 5' end of the *dinI* locus, downstream of the termination codon of *E. coli pyrC* gene (Figure 1). Both sequences have recognizable regulatory features (-10 and -35 boxes, ribosome-binding sites (RBS)). The STYPYRC sequence extends 34 nt further down than the sequenced ECOPYRC and putative initiation codons could be detected (Figure 1). The precise characterization of a peptide-coding sequence in this region requires additional sequencing.

The sequence in U0009 (ECOHU43) had the highest λE score among *E. coli* entries. It is found upstream (from nt c1153 to 1134) of the region coding for the hypothetical 157-aa protein YeeB (c1119-646) and overlaps the putative -10 promoter box (Figure 2).

Down the list there follow three high score sequences, L10328 (ECOUW82), M60066 (STYRFC), and L20897 (ECOGARA) (Table I). The first (ECOUW82) contains a potential box from position nt 42693 to 42712, between two divergently transcribed regions, the *ilvBN* operon (starting with the *ilv* leader peptide, from position c42382 to 42284) and the *spoR*

		-35
ECOPYRC	1	AAATCGTAGCTT-CCTGTTGTCAATTA
STYPYRC	1	AAAATGCGCTTACCTGTTGCTATAT
		-10
ECOPYRC	25	GGTATTTTACCTGTATAAATAACC
STYPYRC	25	ACTGGATTTA <u>ACTGTATATATCC</u>
		SOS-BOX
		RBS
ECOPYRC	51	AGTATATTCAACAGGGGGCTATTAT
STYPYRC	51	<u>AGTATATTCAACAGGGGGCCATTAT</u>
ECOPYRC	76	GCGAATTGAAGTCACCATAGCGAAA
STYPYRC	76	<u>GCGTATTGAAGTCACTATAGCCAAA</u>
ECOPYRC	101	ACTTCTCCATTGCCAG
STYPYRC	101	ACGTCACCTCTGCCTGCCGGTGCGA
STYPYRC	126	TTGATGCGCTGGCGGGCGAACTCGT
STYPYRC	151	CCGCGGTATTAGCCATCATTTTCC
STYPYRC	176	GGAGAATTTGGGTAA

Figure 1 - Alignment of the entries D00002 (ECOPYRC) and X03928 (STYPYRC) immediately downstream of gene *pyrC* in *Escherichia coli* and *Salmonella typhimurium*, respectively. Sequences were aligned by the SOS box-like sites. Promoter boxes (-10 and -35), ribosome-binding sites (RBS), and putative initiation codons (ATG) are indicated under or above the sequences and apply to both, except for the second ATG in the *S. typhimurium* sequence, whose corresponding nucleotides in *E. coli* are not yet known.

gene (from position 43310 to 44500). The distances from the SOS box to the initiation codons are unusually long (311 and 598 nt, for *ilvBN* and *spoR*, respectively), and do not include detectable ORFs, a fact that adds to the ambiguity of the positional relationships. The sequence in M60066 (STYRFC) is upstream of the gene coding for an O-antigen polymerase of *S. typhimurium* (Collins and Hackett, 1991). Besides being rather distant (270 bp) from the polymerase gene initiation codon, it is also located a mere 170 bp down from the start of the entry sequence; this further limits the site analysis. Since a putative initiation codon is found at nt c153-151, a sequence homology search was made over the c150-1 region, with no significant results. Such a finding does not exclude its possible association with a coding region in the opposite strand, in view of the heterogeneous character of the SOS regulon genes (see Discussion).

Entry L20897 (ECOGARA) has an SOS box-like sequence from nt c1540 to 1521, with characteristics indicative of a functionally active operator: a high score heterology index ($\lambda E = 6.66$) overlapping a well-characterized promoter -10 box (TAAAAT), followed by an RBS (Figure 2). This sequence is located upstream of a putative *yebG* locus (from nt c1503 to 1213) coding for a 96-aa protein.

Low score ($15.5 > \lambda E > 10.0$) boxes

The second group ($15.5 > \lambda E > 10.0$) includes four sequences from the *E. coli* genome and one from *S. typhimurium*. There are two adjacent potential SOS boxes in the *S. typhimurium* entry L08613 (STYPEFABCD): one was not considered, due to its rather poor score (6 Nlbp; $\lambda E = 18.82$); the other (4 Nlbp; $\lambda E = 10.18$) was selected (Figure 2). Since it is positioned 1176 bp upstream of the *pefB* gene, an ORF search was carried out in this region, revealing a putative 231-bp ORF (*orf76*, from nt 1214 to 1444).

Operator-like sequences found upstream of antibiotic resistance-coding regions are being studied separately (Vasconcelos and de Almeida, unpublished results) and were not considered here. A representative of that category, the sequence in the *E. coli* entry Y00452 (ECAPH), was exceptionally selected due to the relevance of its characteristics for the purposes of the present study (Figure 2). The entry describes the sequence coding for the aminoglycoside 3'-phosphotransferase (Vakulenko *et al.*, 1987).

There are no references to relevant features in the sequenced region downstream of the SOS box-like sequence found from nt 6618 to 6637 in D13169 (ECORBAB). A homology search of this region identified a 384-nt ORF which is 75.8% homologous to *Serratia liquefaciens orfX* (Givskov and Molin, 1992). However, a rather long (653 nt) fragment separates the ECORBAB SOS-like box from *orfX*, and within this interval a 537-nt ORF was identified, corresponding to a putative *orf178* (Figure 2).

Another putative SOS box was found upstream of the gene coding for *E. coli* enterohemolysin 1 (X70047; ECEHLYL). The box ($\lambda E = 12.97$) is 86 nt upstream of the transcription initiation site, partially overlapping the -35 promoter box. An RBS was also detectable (Figure 2).

The sequence at the bottom of Table I was found both under entries M13608 (ECOPHOM) and M97495 (ECOROBPHOM). Placed between two divergently transcribed genes it could not be unequivocally associated to either the *rob* or the *phoM* gene.

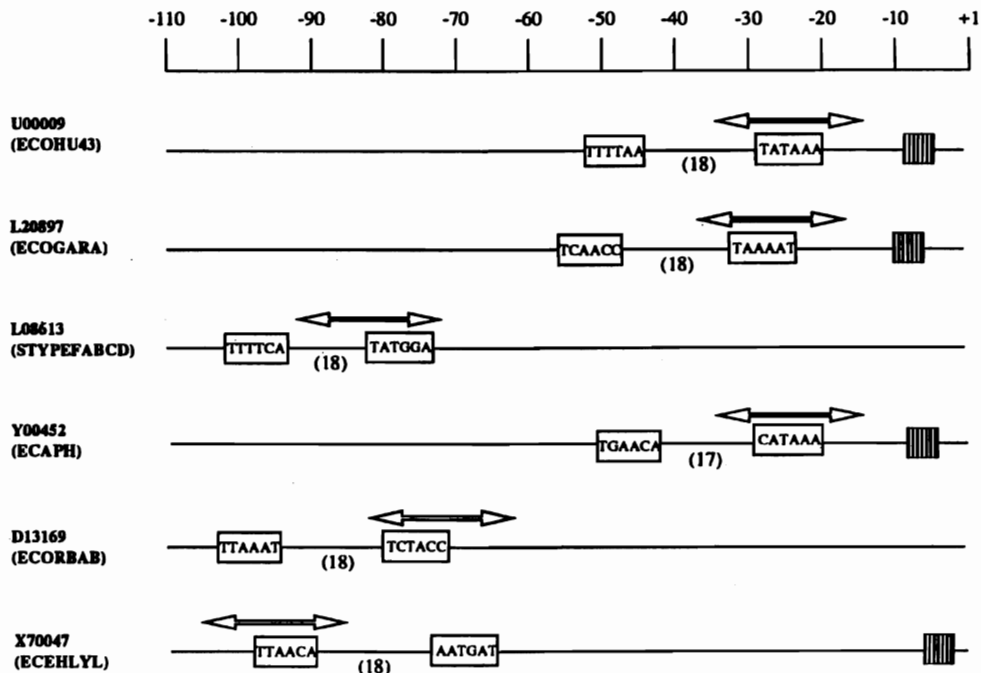


Figure 2 - Scheme of the relative positions of regulatory sequences upstream of the transcription initiation site (+1). Accession numbers and designations are shown at the left. SOS box-like sequences (double-headed arrows), -10 and -35 promoter regions (boxes) and ribosome-binding sites (hatched squares) are indicated. Figures within parentheses denote the number of bp in the spacer regions.

DISCUSSION

A new *E. coli* locus, *yeeB*, detected by similarity search (Vasconcelos, 1995), contains an operator sequence identical to that of the *umuDC* operon. Sequence analysis also led to the association of ORFs with three new SOS box-like sites described here: in *E. coli*, *yebG* (ECOGARA) and *orf178* (ECORBAB), and in *S. typhimurium*, *orf76* (STYPEFABCD) (Figure 2).

Ambiguous cases were found in *E. coli* L10328 (ECOUW82) and M13608 (ECOPHOM) due to the SOS box-like sequence position between divergently transcribed genes. These are now open for experimental investigation.

A common feature of SOS boxes described here is their close association with the promoter region, which they usually overlap. That seems to be an essential condition for the fulfillment of the regulatory function, which involves a physical constraint on the RNA polymerase binding to the promoter site (Collado-Vides *et al.*, 1991).

There is a relationship of an SOS regulatory sequence with the gene coding for aminoglycoside 3'-phosphotransferase, in Y00452 (ECAPH). Such an association seems to be a common finding in genes coding for antibiotic-resistance (Vasconcelos and de Almeida, unpublished results). There is also an

association between the SOS regulatory system and the enterohemolysin-coding gene, in X70047 (ECEHLYL). It is already known that the production of several extracellular enzymes is dependent on the SOS system (Ball *et al.*, 1990), hence the possibility that extracellular products acting as virulence factors may also be under SOS regulation.

Some observations can be made on previously reported DNA damage-inducible genes: (i) the *dinD* locus was recently shown to correspond to *orfY* upstream of *pyrE* (Lundegaard and Jensen, 1994) confirming Lewis *et al.* (1994) assignment; (ii) locus *sosD* seems to be associated to the operon *istAB*, found 227 bp downstream; it codes for a 390-aa protein in *E. coli* which is homologous to a 280 aa in *Shigella sonnei*; (iii) the *sosF* locus is placed 9 bp upstream of gene *yehF*, coding for a hypothetical 274-aa protein.

Though the heterology index brings a quantitative factor into the process, operator sequence identification still benefits from a close examination of neighboring regions. Thus, the sequence (C)GCCTGTCTGAACAAACAGTA in entry J01719 (ECOTYRS), identified as locus *sosE* (Lewis *et al.*, 1994), has escaped our search (limited to extragenic regions; Vasconcelos, Machado and de Almeida, unpublished results; Vasconcelos, 1995) because it is part of the *E. coli* tyrosyl-tRNA synthetase structural gene (Barker *et al.*,

1982). In evaluating an operator-like sequence as a functionally active site, it is thus useful to add to the heterology index determination: the *in vitro* LexA-binding assay; a search for features indicative of a regulatory site in the flanking regions; finally, and more decisively, an evaluation of its functional role *in vivo*.

About half the *E. coli* genome has been sequenced. Assuming that the SOS regulon genes are distributed with no preferential clustering on the genome, it is feasible that its total number will double current figures when genome sequencing is completed. The SOS genes distribution pattern in *S. typhimurium* tends to follow its close genetic relative. The relatively scarce data available for other enteric bacteria do not permit a precise perspective to be drawn on the distribution of the SOS regulon genes in Enterobacteriaceae.

The present report and a recent work (Lewis *et al.*, 1994) demonstrate that the search of potential operator sequences may be successfully used for the identification of new genetic loci of prokaryote regulons.

ACKNOWLEDGMENTS

The authors wish to thank Ken Rudd for the EcoSeq database and the Núcleo de Computação Eletrônica (NCE/UFRJ) for access to the GCG package program. We thank Luis Carlos S. Ferreira and Carlos A. Coimbra for helpful suggestions and discussions. D.F.A. thanks Luiz Levy and Suely Spiguel for stimulating discussions. This work was supported by grants from CNPq, FINEP, and CPEG/ and Forum de Ciência e Cultura/UFRJ.

RESUMO

Uma análise computacional de possíveis caixas SOS em *Escherichia coli* e *Salmonella typhimurium*, com base em similaridade, foi realizada no GenBank. A estratégia consistiu em selecionar inicialmente sequências extragênicas contendo a região de consenso CTG(TA)₅CAG, com até seis bases não-idênticas no decâmero (TA)₅ central. As sequências selecionadas foram então examinadas individualmente, determinando-se seu índice heterológico (λE), que indica sua afinidade relativa pelo repressor, além da presença de características reguladoras associadas (regiões -10 e -35 de promotores, sítios ligadores de ribossomos, uma ORF a jusante) situadas nas regiões adjacentes. Foram selecionadas sete sequências de *E. coli* e três de *S. typhimurium*; cinco tinham valores de $\lambda E < 7,0$ e cinco tinham $15,5 > \lambda E > 10,0$. Um locus homólogo ao gene *dinI* de *E. coli* foi detectado em *S. typhimurium*. Três sequências selecionadas (2 em *E. coli*, 1 em *S. typhimurium*) estão associadas a ORFs reconhecíveis; outras se encontram a montante de regiões codificantes de resistência

a antibióticos e produção de enterohemolisina, enquanto outras têm posição ambígua em relação a regiões codificantes, particularmente por se situarem entre genes de transcrição divergente. Os resultados mostram que a busca de sequências similares a operadores conhecidos em procariotos pode ser usada para a identificação de genes pertencentes a regulons específicos.

REFERENCES

- Ball, T.K., Wasmuth, C.R., Braunagel, S.C. and Benedik, M.J. (1990). Expression of the *Serratia marcescens* extracellular proteins requires *recA*. *J. Bacteriol.* 172: 342-349.
- Barker, D.G., Bruton, C.J. and Winter, G. (1982). The tyrosyl-tRNA synthetase from *Escherichia coli*: complete nucleotide sequence of the structural gene. *FEBS Lett.* 150: 419-423.
- Berg, O. (1988). Selection of DNA binding sites by regulatory proteins: the LexA protein and the arginine repressor use different strategies for functional specificity. *Nucleic Acids Res.* 16: 5089-5105.
- Berg, O.G. and von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins: statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* 193: 723-750.
- Collado-Vides, J., Magasanik, B. and Gralla, J.D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* 55: 371-394.
- Collins, L.V. and Hackett, J. (1991). Molecular cloning, characterization, and nucleotide sequence of the *rfc* gene, which encodes an O-antigen polymerase of *Salmonella typhimurium*. *J. Bacteriol.* 173: 2521-2529.
- Givskov, M. and Molin, S. (1992). Expression of extracellular phospholipase from *Serratia liquefaciens* is growth-phase-dependent, catabolite-repressed and regulated by anaerobiosis. *Mol. Microbiol.* 6: 1363-1374.
- Lewis, L.K., Harlow, G.R., Gregg-Jolly, L.A. and Mount, D.W. (1994). Identification of high affinity binding sites for LexA which define new DNA damage-inducible genes in *Escherichia coli*. *J. Mol. Biol.* 241: 507-523.
- Lodevick, D., Owen, D. and Strike, P. (1990). DNA sequence analysis of the *imp* UV protection and mutation operon of the plasmid TP110: identification of a third gene. *Nucleic Acids Res.* 18: 5045-5050.
- Lundegaard, C. and Jensen, K.F. (1994). The DNA damage-inducible *dinD* gene of *Escherichia coli* is equivalent to *orfY* upstream of *pyrE*. *J. Bacteriol.* 176: 3383-3385.
- Mustard, J.A., Thliveris, A.T. and Mount, D.W. (1992). Sequence of the *Salmonella typhimurium* LT2 *lexA* gene and its regulatory region. *Nucleic Acids Res.* 20: 1813.
- Payne, N.S. and Sancar, A. (1989). The LexA protein does not bind specifically to the two SOS-like sequences immediately 5' to the *phr* gene. *Mut. Res.* 218: 207-210.
- Pearson, W. and Lipman, D. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85: 2444-2448.
- Peterson, K.R., Ossana, N., Thliveris, A.T., Ennis, D.G. and Mount, D.W. (1988). Derepression of specific genes promotes DNA repair and mutagenesis in *Escherichia coli*. *J. Bacteriol.* 170: 1-4.

- Schnarr, M., Oertel-Buchheit, P., Kazmaier, M. and Granger-Schnarr, M. (1991). DNA binding properties of the LexA repressor. *Biochimie* 73: 423-431.
- Shadel, G.S., Devine, H.H. and Baldwin, T.O. (1990). Control of the *lux* regulon of *Vibrio fischeri*. *J. Biolumin. Chemilumin.* 5: 99-106.
- Shinagawa, H., Makino, K., Amemura, M., Kimura, S., Iwasaki, H. and Nakata, A. (1988). Structure and regulation of the *Escherichia coli* *ruv* operon involved in DNA repair and recombination. *J. Bacteriol.* 170: 4322-4329.
- Smith, C.M., Koch, W.H., Franklin, S.B., Foster, P.I., Cebula, T.A. and Eisenstadt, E. (1990). Sequence analysis and mapping of the *Salmonella typhimurium* LT2 *umuDC* operon. *J. Bacteriol.* 172: 4964-4978.
- Smith, C.M., Arany, Z., Orrego, C. and Eisenstadt, E. (1991). DNA damage-inducible loci in *Salmonella typhimurium*. *J. Bacteriol.* 173: 3587-3590.
- Thomas, S.M., Crowne, H.M., Pidsley, S.C. and Sedgwick, S.G. (1990). Structural characterization of the *Salmonella typhimurium* LT2 *umu* operon. *J. Bacteriol.* 172: 4979-4989.
- Vakulenko, S., Kalman, M., Horvath, B. and Simoncsits, A. (1987). The nucleotide sequence of an aminoglycoside 3'-phosphotransferase gene from *Escherichia coli*. *Nucleic Acids Res.* 15: 8111.
- Vasconcelos, A.T. (1991). Sistema gerenciador de um banco de dados genético. *Relatório de Software/LNCC No. 3*. LNCC, Rio de Janeiro, pp. 31.
- Vasconcelos, A.T. (1995). Análise de seqüências de nucleotídeos no ADN extragênico de procariotos: estudo de palíndromos interrompidos, com aplicação ao sistema SOS. Master's thesis. Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil. pp. 180.
- Vasconcelos, A.T. and de Almeida, D.F. (1994). Análise de um banco de dados genético: aplicação ao estudo do sistema SOS de procariotos. *Anais do XVII Congresso Nacional de Matemática Aplicada à Computação (CNMAC)*, pp. 59-62.
- Vasconcelos, A.T., Coimbra, C.A., Pfeffer, A.V., Martins, L.C., Correa, G. and de Almeida, D.F. (1994). Prokaryote DNA damage-inducible SOS system genes detected by nucleotide sequence analysis. *An. Acad. Bras. Ci.* 66: 127.
- Walker, G.C. (1985). Inducible DNA repair systems. *Ann. Rev. Biochem.* 54: 425-457.
- Walker, G.C. (1987). The SOS response of *Escherichia coli*. In: *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umberger, H.E., eds.). American Society for Microbiology, Washington, DC, pp. 1346-1356.
- Wertman, K.F. and Mount, D.W. (1985). Nucleotide sequence binding specificity of the LexA repressor of *Escherichia coli* K12. *J. Bacteriol.* 163: 376-384.

(Received July 12, 1995)