Dear Collaborators,

This document describes shortly all the informations related to the *Lactobacillus sakei* Oligoset. A rather brief description from operon is given in file:

CustomerArrayReadyOligoSetDescription_ByOperon.pdf

1) Explanations on the genes selected and the design of the 70-mers oligos:

All genes from *L. sakei* 23K strain chromosome were selected. These are caracterized by a locus_tag LSAXXXX and their specific description (product and gene name) is available in file :

DescriptionOfGenesInOligoset.pdf

The *L. sakei* 23K genome contains 1,883 protein/peptide encoding genes, however, some of them are duplicated such as IS copies. Therefore, the number of *L. sakei* 23K **unique genes** chosen for oligo design was of 1870. To these were added the 3 rRNA genes (16S, 23S & 5S) giving a total of **1873** genes.

It should be noted that some genes (~ 40) were quite small (between 70 and 100 nucleotides) and encode hypothetical small peptides such as bacteriocins precursors. Apparently there has not been any problem for oligo design. However, I would like to emphasize on 11 small genes (noted possible peptides in the above mentioned file [from LSA0075 to LSA1826]. These genes were shortly smaller than 70 nucleotides and our HMM models (gene detector) gave high confidence as being 'real' hypothetical genes. To allow design of a 70-mers oligo for them, I included part of their RBS signal (which in principal should be transcribed) hoping it will not give too much cross hybridisation with the whole genome or other oligos since their RBS signal was quite degenerate). Fortunatly it was the case so in principal they should not give any problem. You should be carefull when analysing the signal they give however.

They were 11 grey holes in the genome (zones ~ 500bp) without predicted genes (noted with a locus_tag GHOXXXX). Sense and antisense paired probes were designed in these regions in case they may encode misc_RNAs.

There was 105 oligos left to be designed. I decided **not to** include the 63 tRNAs, but rather include genes from other strains and available publicly. I have not changed their gene name but gave them a locus_tag to distinguish them from each other and from genes from strain 23K. They can be used, at least, as negative control when strain 23K is used to set up hybridisation conditions.

Locus_tag	Origin	Accession No.
TNLXXXX	Strain 332: Lactose PTS transposon	(unpublished)
PRVXXXX	Strain 332: plasmid pRV500	(AF438419)
SKXXXXX	Strain LSA5: SakacinXP cluster	(AY206863)
SKPXXXX	Strains lb674/lb790: additional Sakacin P genes	(Z48542 & AJ626710)
SKAXXXX	Strain lb706: Sakacin A cluster	(Z46867)
SKGXXXX	Strain 2512: Sakacin G cluster	(AF395533)
SKQXXXX	Strain LTH673: Sakacin Q cluster	(AJ844595)
LCSXXXX	Strain L45: Lactocin S cluster	(Z54312)
GSUXXXX	Strain KG15: Glucan sucrase	(AY697434)
TETXXXX	Unknown strain: plasmidic tetracycline res.	(Many ref. available...)

Finally, for 22 genes (largest genes in the genome, including 16S & 23S) as well as for the positive control LSA1606 (LDH), two sense oligos were designed (one at each end of the gene), whereas other oligos were designed in central regions for all the other genes. Thus a total of **2,000 oligos** were designed.

2) Descriptions of oligos and Problems:

All oligos are described in file : DescriptionsOfOligos.pdf

There were few minor problems summarized below (and at the end of file mentioned above), but if I exclude these problems:

Average TM of oligos: 74.7°C
with a minimum of 70.6°C and a maximum of 77.7°C. So differences are < 7°C.
Average GC-content: 41.25
with a minimumof 30 and a maximumof 48.57. So differences are < 19
The average cross-hybridization: 37.5% identity (O.K. should be <= 70%)
with a maximum of 51%
The average contiguous bases common to other ORFs: 14.22 (O.K. should be <=20)
with a maximum of 20.

Problems

1) Some oligos display lower TM because of the low GC-content of the corresponding genes, many of which are plasmidic genes:
Chromosomal genes: LSA0214, LSA0362, LSA0789, LSA1335, LSA1823 and TNL0019, SKX0005, SKX0010, SKX0013.
The lower TM amonsgt these oligos is 68.9°C and GC%=27.1 (LSA0214)

Plasmidic genes: LCS0010, LCS0014, LCS0015, LCS0016, LCS0018, PRV0006, PRV0011, PRV0012, SKA0002, SKA0004, SKA0007, SKA0008, SKP0001, SKP0007, SKQ0002.
The lower TM amongst these oligos is 68.3°C and GC%=25.1 (SKP0001: sakacin P inducing peptide, and LCS0012: hypothetical protein. Oligos from these two show 50% cross-hybridisation). According to Operon, it was not possible to get any better design.

2) Some oligos display quite high cross-hybridisation to other ORFs:

LSA0262 with LSA0474 (IS150, orfA) 90% cross-hybridisation
LSA0261 with LSA0473 (IS150, orfB) 88% cross-hybridisation

Although only one set of oligo were designed for each IS (since copies are identical), one copy of IS150 has slightly derived from the two others (LSA0261/0262) and the gene content was only 90% identical. I forgot about it before sending the file to Operon. Well, it could help to check hybridisation quality if spotted at different zone of the slides.... (always look to the bright side of life!)

SKG0002 with SKG0003 (alpha and beta subunit of sakacin G) 81% cross-hybridisation

The two subunits are quite similar and according to operon it was not possible to get any better design since the genes are quite small.

SKP0008 (noted hypothetical peptide) and SKQ0001 (sakacin Q precursor) 100 to 92% identity.

My mistake ! I first included a small gene located at 3'end of sakacinP cluster from strain lb674 (which I annotated hypothetical peptide) and did not noticed it was almost the same gene deposited recently by Lars as SakacinQ precursor. Thus same remark as for IS150...

LSA1159 and LSA1165 (putative aggregation factors) 80 to 87% cross-hybridisation.

These are two paralogous proteins whose genes are almost identical at the nucleotidic level. It was impossible to get any better design.

Finally, LSA1483 with GHO0008_rev 62% cross-hybridisation. (should be OK however for hybridisation)

In conclusion, to the exception of the aggregation factors, most are minor problems (low gc-content and TM slighly lower than average) or redundancy of oligos for some almost identical genes. One should check during our first hybridisations experiments whether they do not give problems.

3) Last comment:

Cross-hybridisation was also filtered against E.coli database. Although I told Operon it was not the smartest strategy to use since other genomes closely related to L. sakei were available. Apparently, the argument did not make much sense to them. Probably they were lazzy to set up new databases. So, to speed up the synthesis, I did carried out some filtering against *L. plantarum, E. faecalis, L. monocytogenes, L. johnsoni* which took me some work and time.... No cross-hybridisation above 50% was found excepted TNL0020 (IS30-like) showing 100% identity to ISLpl1 from *L. plantarum.*

Best regards and successful DNA-array experiments,

Stéphane Chaillou,
FLEC laboratory, I.N.R.A Jouy-en-Josas.
July, 8th 2005