



Operon Biotechnologies, Inc.  
2705 Artie St., Bldg. 400, Ste. 27  
Huntsville, AL 35805

[www.operon.com](http://www.operon.com)

## **Operon *Streptococcus pneumoniae* Combo Genome Array Ready Oligo Set™ (Version 1.0)**

---

Operon Biotechnologies, Inc. proudly introduces the *Streptococcus pneumoniae* Combo Genome Array Ready Oligo Set™ (AROS, version 1.0). *Streptococcus pneumoniae* Combo AROS was created by the Operon scientists using the proprietary oligonucleotide probe design platform. The Combo AROS covers the genes and intergenic regions of two *S. pneumoniae* strains R6 and TIGR4. The Combo AROS maximizes the efficiency of oligo presentation over the homologous genes and intergenic regions between two genomes with heuristic search algorithm for common oligos. The Combo AROS consists of 2,879 oligos, 2,108 common oligos covering genes and intergenic regions from the two genomes. The rest are the strain-specific oligos (314 from strain R6, 454 from strain TIGR4) and miscellaneous oligos (3). In addition, the Combo AROS is supplemented with a set of positive control oligos, randomly-generated negative controls oligos, stringency control oligos, alien spike control oligos and tracking oligos for quality control and the experimental validation.

Release date: February, 2006

### **I. The sequence sources**

The Genbank accessions NC\_003028/AE005672 (GI:15899949/GI:85720550, strain TIGR4), and NC\_003098/AE007317 (GI:15902044/GI:25307955, strain R6) were the primary sources for the gene sequences and intergenic regions in the design. The genome sequences and their annotations are accessible at the National Center for Biotechnology Information ([www.ncbi.nih.gov](http://www.ncbi.nih.gov)).

### **II. The design criteria and selection rules:**

The following selection rules were executed in the design.

- 1) The intergenic region selection: The intergenic regions with the size of equal or greater than 200 bases were included for the oligo design.

- 2) Common oligo selection: The oligo candidates from R6 strain with the alignment to the gene and intergenic sequences of TIGR6 were screened out at the identity of 97% or above for common oligo candidates. If no common oligos were available to cover the genes and the intergenic regions, the species-specific oligos were designed.
- 3) The length of oligo sequences was set at 70 bases with the exceptions for genes and regions without qualified oligo candidates.
- 4) The melting temperatures ( $T_m$ ) of  $74 \pm 5^\circ\text{C}$  was used for oligo screening.  $T_m$  calculation was done with the formula:  $T_m = 81.5 + 16.6 * \log[\text{Na}^+] + 41 * (\#G + \#C)/\text{length} - 500/\text{length}$ , where  $[\text{Na}^+] = 0.1 \text{ M}$  and  $\text{length} = \#A + \#C + \#G + \#T$ .
- 5) The oligo candidates were filtered with 40 bases away from the 3'-end of transcripts and a maximal distance of less than 1000 bases from the 3'-end.
- 6) The oligo candidates were screened against the low complexity of sequences: (a) The contiguous single nucleotide base repeat or poly (N) tract ( $\leq 8$  bases); (b) the pre-set normalized simple repeat score.
- 7) The oligo candidates were selected against the potential hairpin structure with stem length of over 9 bases.
- 8) The oligo candidates were screened against the cross-hybridization of the non-self transcripts as well as the genomic sequences from the two genomes. The non-self transcript is the transcript not represented by a specific oligo. The selection criteria were used in the selection: (a) the cross-hybridization identity score  $\leq 70\%$ ; (b) contiguous match length to any non-self transcript  $\leq 20$  bases.
- 9) Exceptions were made if no qualified oligos emerged from the selection from some genes.

### **III. The specifications and characteristics of *Streptococcus pneumoniae* Combo AROS (version 1.0)**

The Combo AROS encompasses both the gene sequences and the intergenic region sequences from the genomes of *S. pneumoniae* strain R6 and TIGR4. The Combo AROS are divided into three major components.

- 1) Common oligos: 2,108 of them are to cover the genes and the intergenic regions of both genomes. The common oligo herein Combo AROS is defined as the oligo to represent the homologous genes and regions from both R6 and TIGR4 genomes at the sequence identity of 97% or above.

- 2) Strain-specific oligos: The oligos specific to represent the genes or intergenic regions of either R6 genome or TIGR4 genome. The former consists of 314 oligos, the latter 454 oligos.
- 3) Misc oligos (3): They represent scRNA, penicillin resistant gene, box AB and box BC elements.
- 4) Control oligos:
  - Negative control oligos: They're randomly-generated oligo sequences. They were selected after filtering against the transcripts and intergenic region sequences for non cross-hybridization property.
  - Eukaryotic control oligos: Four oligos (H300006234, H300013722, H300022861, opHsV04TC000032) from human AROS (version 4.0) are included in this Combo AROS as eukaryotic control oligos for human ACTB, STAT1, ISGF3G and GAPD genes.
  - Positive control oligos: Ten oligos from this Combo AROS were selected as the positive control oligos. They represent the following genes: shikimate dehydrogenase (GI:15458871), glucose-6-phosphate 1-dehydrogenase (GI:15458749), RNA polymerase (delta subunit) (GI:15458003), glucokinase (GI:15459181), 30S ribosomal protein S3 (GI:15457740), signal peptidase I (GI:15457924), xanthine phosphoribosyltransferase (GI:15459335), D-alanine-D-alanine ligase (GI:15459177), cell division protein FtsZ (GI:15459171) and NADP-specific glutamate dehydrogenase (GI:15458813).
  - Stratagene alien spike control oligos: They are licensed from Stratagene ([www.stratagene.com](http://www.stratagene.com)). In coupling with the alien mRNA spikes in Stratagene SpotReport Alien Oligo Array Validation System, they are intended as the internal controls for the normalization and standardization of dye incorporation, microarray hybridization and data analyses.
  - Hybridization stringency control oligos: They were generated based on four oligo sequences with sequence homology of 70%, 80%, 90% identity to the original ones. They're intended as the standards to calibrate the microarray hybridization stringency. The four oligos represent the following genes: acetyl-coenzyme A carboxylase carboxyl transferase (beta subunit) (GI:15457947), 50S ribosomal protein L20 (GI:15458467), topoisomerase IV (subunit A) (GI:15458351), UDP-glucose 4-epimerase (GI:15459318).
  - Production tracking oligo: It is a randomly-generated oligo sequence with a length of 30 bases. It was filtered against the cross-hybridizations with the gene sequences used in the AROS design. The tracking oligo is randomly positioned in the 384-well and 96-well plates so that each 384-well plate has four, and each 96-well plate has one for quality assurance.

Figure 1. Tm

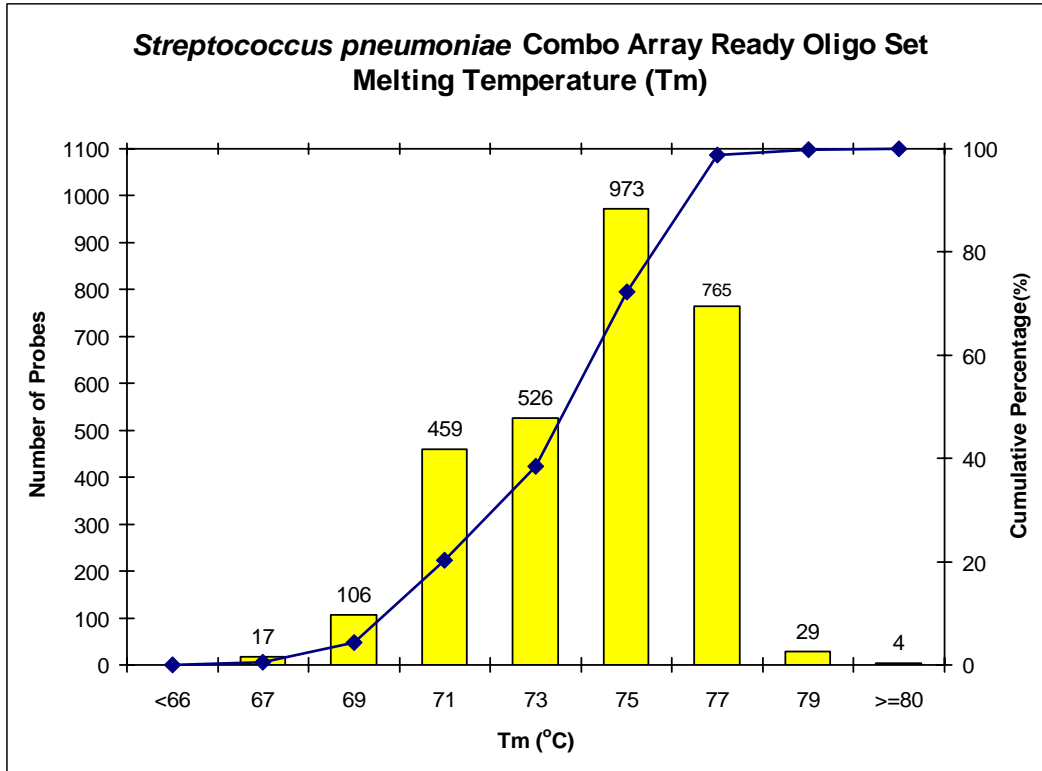


Figure 2. GC content

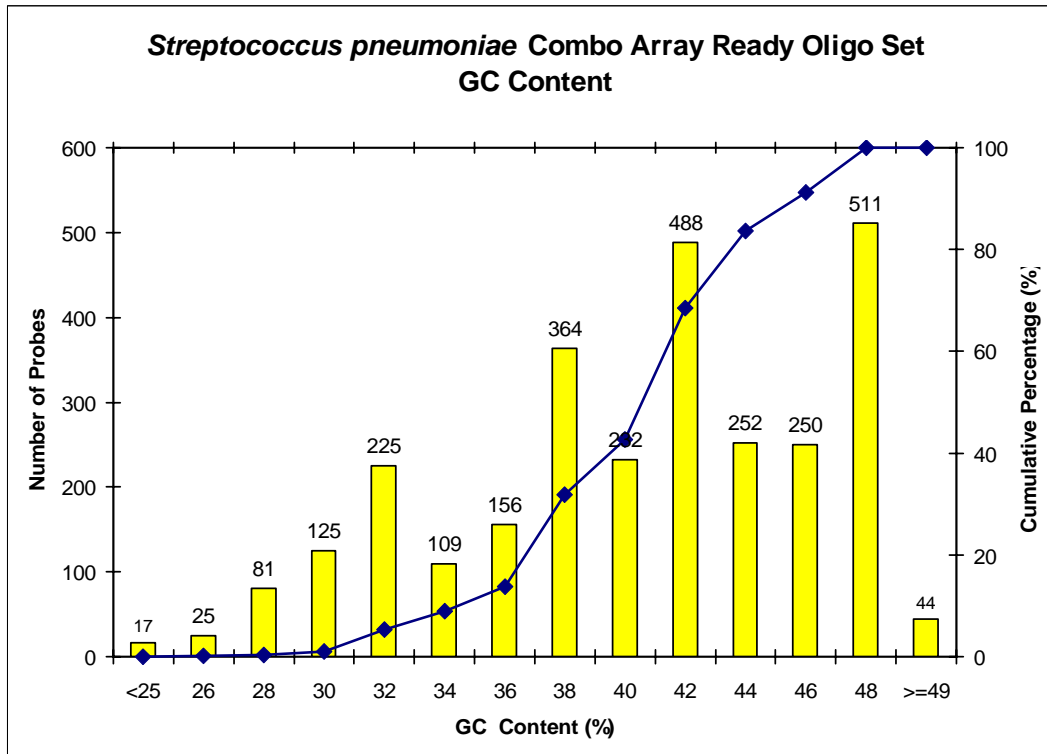


Figure 3a. Distance to the 3'-end of strain R6 transcripts

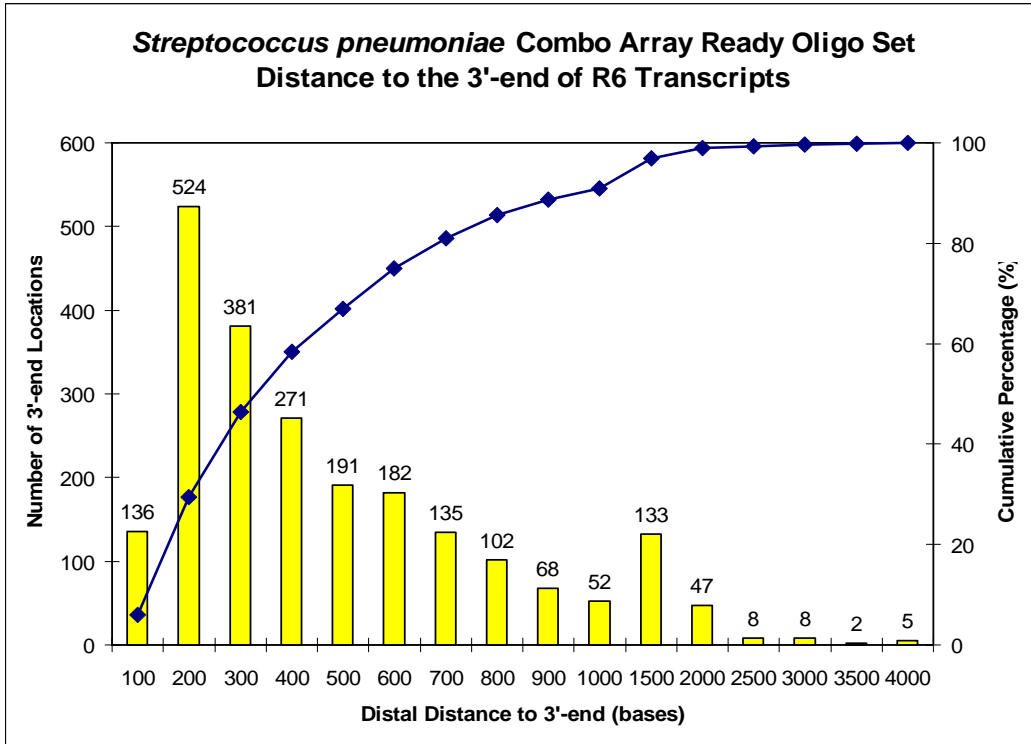


Figure 3b. Distance to the 3'-end of strain TIGR4 transcripts

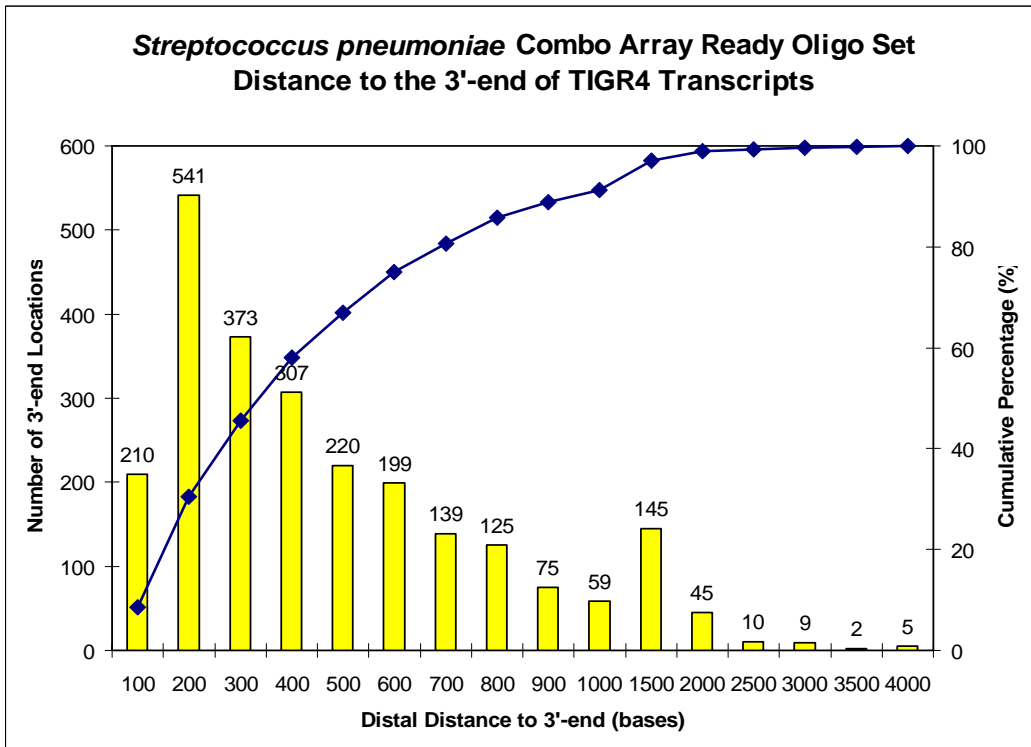


Figure 4. Hairpin stem length

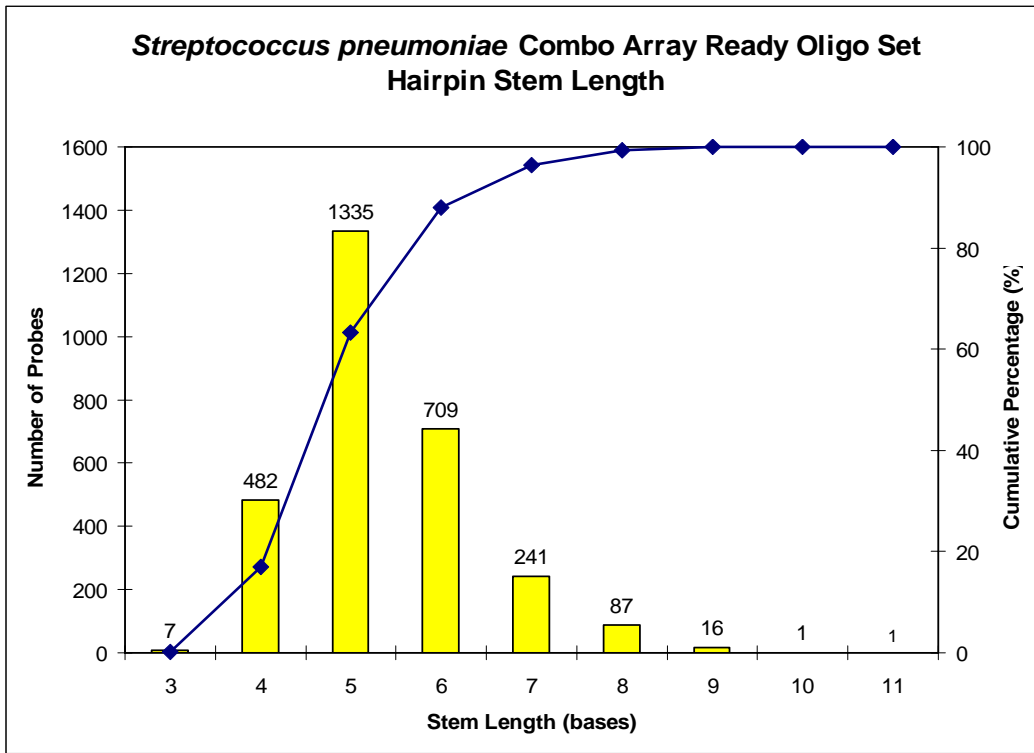


Figure 5a. Cross-hybridization to strain R6 transcripts

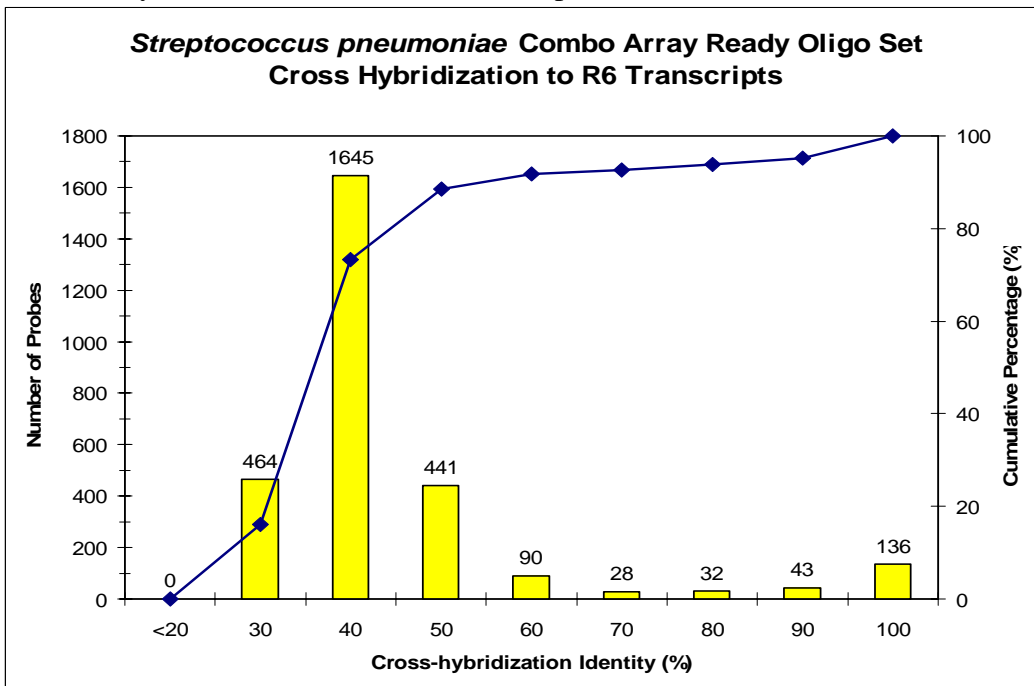


Figure 5b. Cross-hybridization to strain TIGR4 transcripts

