

### Array-Ready Oligo Set™ for the *H. influenzae* Genome

We are pleased to announce the release of our *H. influenzae* Genome Oligo Set containing 1714 70mer probes representing 1714 *Haemophilus influenzae* Rd open reading frames (ORFs). For our probe design we use state-of-the-art methods and proprietary software. An amino linker is attached to the 5' end of each oligo.

#### ORF Sequence Source and Selection

The 1714 probes for *H. influenzae* are designed from the NCBI *refseq* NC\_000907 using the database file NC\_000907.ffn (updated on March 2002; source: the National Center of Biotechnology Information [NCBI] at [ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Haemophilus\\_influenzae/](ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Haemophilus_influenzae/)). All ORFs in NC\_000907.ffn correspond to the *H. influenzae* Rd strain and more information is available at the following web site: <http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/efetch?db=Genome&gi=25>.

#### Probe Design and Selection Rules

A probe is selected with an optimal set of parameters. Large numbers of 70mer candidate probes for each ORF are selected using the following criteria for the *H. influenzae* Genome Oligo Set.

1) All oligos are within  $76 \pm 5^\circ\text{C}$  using the following formula:

$$T_m = 81.5 + 16.6 \times \log[\text{Na}^+] + 41 \times (\#G + \#C)/\text{length} - 500/\text{length} \text{ where } [\text{Na}^+] = 0.1 \text{ M and length} = \#A + \#C + \#G + \#T$$

2) To accommodate both random and gene-specific priming, each oligo is located more than 25 bases away from both the 5' and 3' end of the ORF. In other words, the 5' end of the oligo is more than 25 bases from the 5' end of the ORF and the 3' end of the oligo is at least 25 bases downstream from the 3' end of the ORF.

3) An oligo cannot have a contiguous single nucleotide base repeat or poly (N) tract longer than 7 bases.

4) An oligo cannot have a potential hairpin structure with a stem length longer than 8 bases.

5) Each oligo has less than or equal to 70% identity to all other ORFs. Each oligo is aligned against all 1714 ORF sequences in NC\_000907.ffn. Using the alignment with the candidate oligo versus the highest scoring non-self sequence, a BLAST percent identity score is computed. The highest scoring non-self sequence is defined as the sequence that yields the most matched bases in an alignment. This BLAST percent identity is also referred to as cross-hybridization identity of the oligo.

This calculated percent identity score is dependent on the size of the sequence database used to BLAST against, oligo sequence, and use of either gapped or no gap alignment method.

6) Each oligo of any length cannot have greater than 20 contiguous bases common to any other gene.

Once oligo candidates have been selected satisfying all the selection rules mentioned above, each oligo is ranked based on BLAST percent identity as computed in Step 5. One final oligo for each gene is selected with the minimum cross-hybridization identity.

Note that for 29 (1.6%) ORFs that did not yield oligos satisfying all the above criteria, certain rules were relaxed. For those ORFs, the oligo is located less than 25 bases to either the 3' or 5' end or the cross-hybridization identity is greater than 70%.

SUMMARY

Oligo selection criteria	Value	Number of oligos in genome set satisfying these criteria
Length Melting temperature Location from 5' and 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other genes Contiguous base match to any other gene	70mer 76°C ± 5°C > 25 ≤ 7 ≤ 8 ≤ 70% ≤ 20	1685
Location from 5' and 3' end	≤ 25	4
Cross-hybridization identity to all other genes	>70%	23
Melting temperature	69°C–71°C	2
<b>Total</b>		1714

The following illustrations show the distribution of all 1714 oligos for melting temperature, GC content, location from 3' end of gene sequence, length of maximum stem length, and cross-hybridization identity.

Figure 1. Melting Temperature

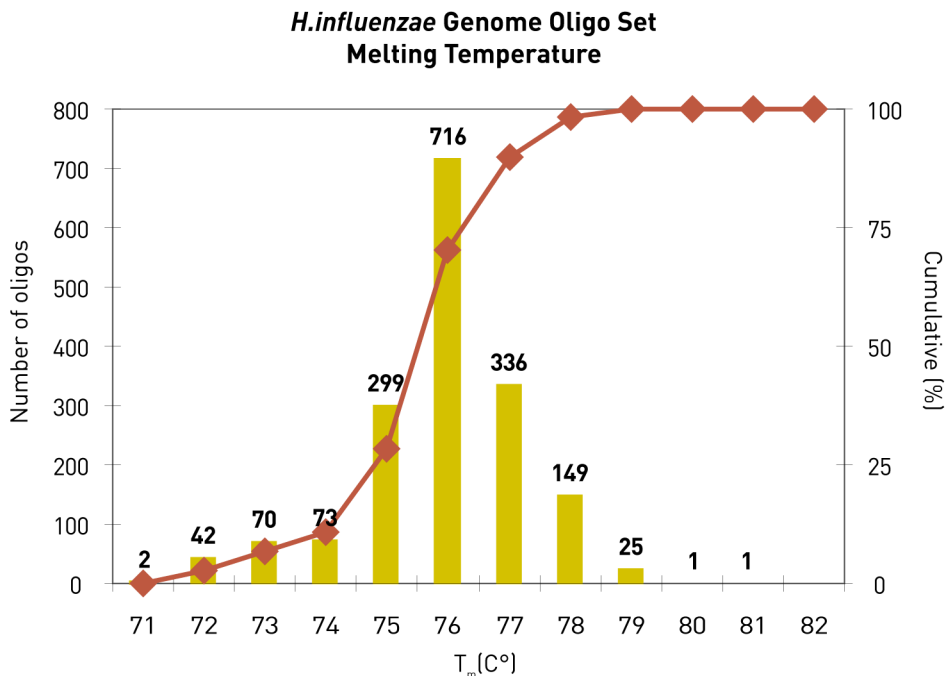


Figure 2. GC Content

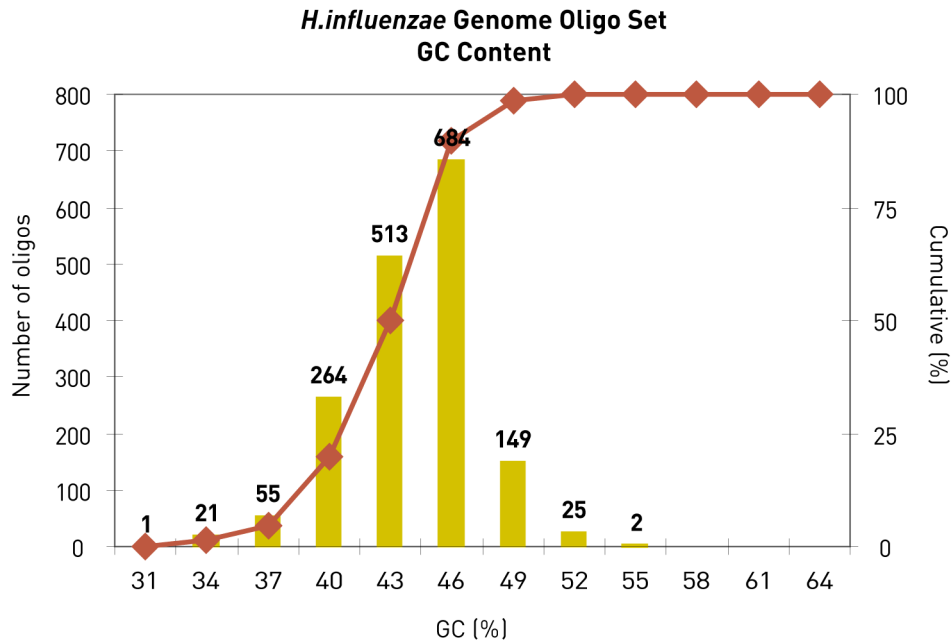


Figure 3. Location from 3' End

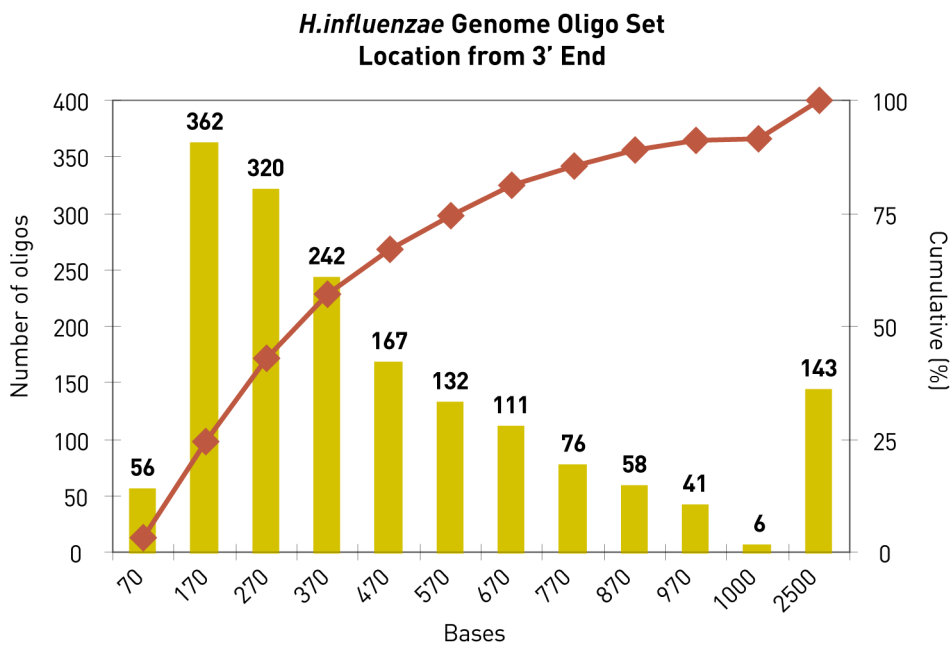


Figure 4. Length of the Longest Hairpin Stem

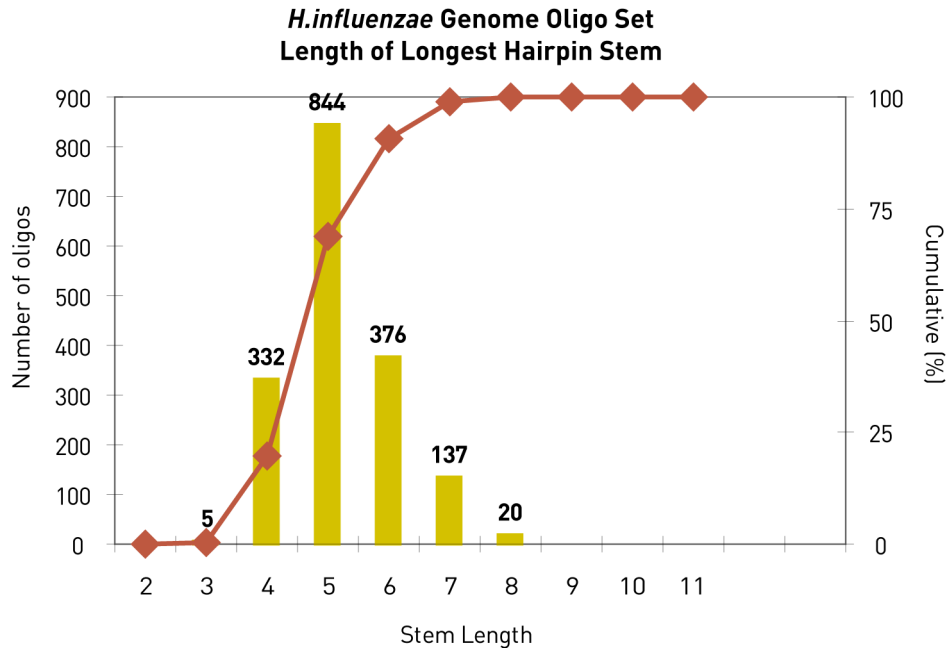
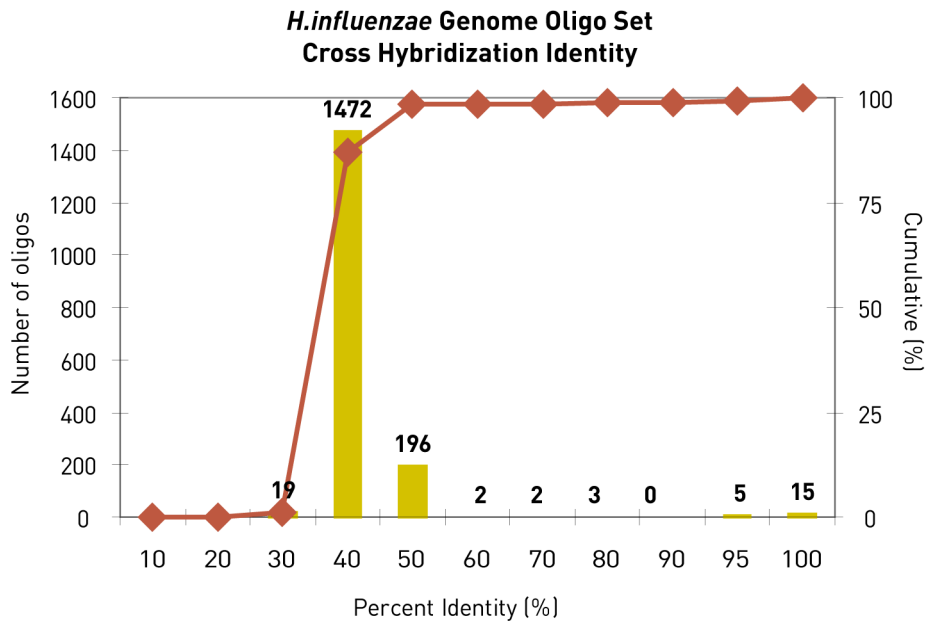


Figure 5. Cross-Hybridization Identity



### Quality Check of Probe Design Specifications

Once the final oligo has been selected to represent a gene, each oligo undergoes design specifications quality control where we use an independent method to confirm that all oligos have met the specified design specifications. The table below summarizes data from our quality check for probe design specifications for all 1714 oligos in the set.

Probe design specification	Expected value	Verified range	Number of oligos
Melting temperature [C°]	76 ± 5 °C	71.2–80.1	1712
Melting temperature [C°]	69°C –71°C	69.5–70.6	2
Cross-hybridization similarity (%)	≤ 70	27–67	1691
Cross-hybridization similarity (%)	71–100	71–100	23