



Operon Biotechnologies, Inc.
2211 Seminole Drive
Huntsville, AL 35805
www.operon.com

***Danio rerio* (zebrafish) Genome Array Ready Oligo Set™ (Version 4.0)**

Danio rerio (zebrafish) V4.0 AROS is designed using state-of-the-art methodology and proprietary software for detection of alternative splice variants from Ensembl *D. rerio* release 42.6b transcripts. This version of the Ensembl gene and transcript predictions and annotations were based on from the Zv6, whole genome shotgun assembly (Mar 2006) provided by the Sanger Institute. The Ensembl genebuild of Zv6 predicts 36,190 gene transcripts and 3,520 RNA genes. In this AROS set, 30,280 transcripts are covered by at least one AROS oligo probe providing essentially full coverage of the zebrafish nuclear and mitochondrial genomes.

In this *D. rerio* AROS V4 oligo probes provide coverage for 404 miRNA, 85 miscellaneous RNA, 6 mitochondrial RNA, 22,708 probes, 117 rRNA, 469 snoRNA, and 797 snRNA for a total coverage of 24,586. From the original Ensembl dataset, 341 genes of all types were eliminated because they were duplicated and an additional 285 protein genes were eliminated probably for Tm.

Probe design targets each oligo to be fully contained within a single exon and makes this probe set also applicable for CGH (comparative genome hybridization) experiments. The *D. rerio* AROS V4.0 contains 22,845 probes including 70 Stratagene SpotReport™ Alien™ control oligos, 5 Negative controls, 42 Positive and Positive Stringency control oligos. Additionally, 242 Production control oligos used by Operon for internal checks and also useful to researchers for plate layout checks are included free of charge. These AROS control probes supplement with the set to facilitate experimental validation and quality control.

Release date: October, 2007

I. The sequence sources

Danio rerio genome sequence and its associated annotations are from Ensembl *D. rerio* Ensembl 42.6b release. More information on the underlying data is available at http://www.ensembl.org/Danio_rerio/index.html and at http://www.sanger.ac.uk/Projects/D_rerio/Zv6_assembly_information.shtml

II. Design criteria and selection rules:

The following selection rules were executed in the design.

- 1) Target length of oligo probe sequences was 70 bases. Exceptions for genes and regions without qualified oligo candidates at this length are routinely made in AROS probe sets in order to maintain tight control of hybridization characteristics to allow for highly stringent experimental conditions. These exceptions are based on the inherent characteristics of the *D. rerio* genome and not a limitation of the design algorithm.
- 2) The target T_m was taken to be 74 °C to allow oligos with higher T_m to be shortened to minimize the T_m range 74 ± 6 °C. T_m was calculated using a 'nearest-neighbor' method (Novère, 2001).
- 3) Probe candidates were processed to shorten those with $T_m > 80$ although none was allowed to be shorter than 60 bps. The resulting oligos were processed to remove probes representing gene copies and then to replace or concatenate records with cross-hybridization greater than 85% or with more than 20 contiguous bases in common (He *et al*, 2005) as described below.
- 4) Probe candidates were screened against the low complexity of sequences: (a) The contiguous single nucleotide base repeat or poly (N) tract (≤ 8 bases); (b) the pre-set normalized simple repeat score.
- 5) Probe candidates were selected against the potential hairpin structure with stem length of over 9 bases.
- 6) Probe candidates were screened against the cross-hybridization of the non-self transcripts. 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 ≤ 20 bases.
- 7) Limited exceptions were made if no qualified oligos were available from the selection. Information covering these exceptions is detailed in the *D. rerio* (zebrafish) AROS V4.0 genelist.
- 8) Probe design targets each oligo to be fully contained within a single exon and makes this probe set also applicable for CGH (comparative genome hybridization) experiments. The probes contained within a single exon are detailed in the

III. Annotation notes:

Oligos were identified by using NCBI BLAST to map each oligo to the sequences in the databases described above. Identifiers are according to the BLAST score (score $>97\%$ identity).

The oligos are classified according what transcripts and genes they represent as shown in Table 1. The design process attempts to find an oligo for every unique combination of transcripts for each gene.

The Ensembl *D. rerio* data contains many copies of some genes and the oligos were processed to remove redundant entries.¹ The resulting set was further processed to remove or concatenate entries with unacceptable cross-hybridization.

In some cases it was possible to identify an oligo from another region of the transcript which met cross-hybridization requirements. In other cases the transcripts were too similar to find unique regions. Many of these were associated with annotated Ensembl gene families with similar biological function and these oligos are marked with type 'F'. In other cases the genes appear unrelated and these are marked with type 'M'. All of the 'M' type oligos are at the end of the list to enable the user to ignore them if they wish.

Oligo Type	Definition
I	Oligo represents one transcript of an Ensembl gene with multiple transcripts
CI	Oligo represents the only transcript of an Ensembl gene
C	Oligo represents all transcripts of an Ensembl gene
P	Oligo represents a subset of transcripts of an Ensembl gene
F	Oligo represents multiple Ensembl genes that belong to a gene family
M	Oligo represents multiple Ensembl genes that cannot be identified as belonging to a family

Table 1 Oligo Type Definitions

It is possible that an oligo probe maps to more than one Ensembl gene or transcript. When the oligo represents more than one gene, the genes it represents are separated by 'pipe' symbols (|) in the same order the genes appear. When the oligo represents more than one transcript, the transcripts it represents are separated by semi-colon symbols (;) in the same order the transcripts appear. If the oligo represents multiple transcripts from multiple genes the transcripts are separated by semi-colons for a particular gene and by 'pipe' symbols for different genes.

Oligos based on Ensembl data include the following ancillary data from the core, compara, and external databases: gene symbol and description; human and rat homologies; gene ontology terms; and gene family information. The latter is useful for determining the impact of oligos that map to more than one gene.

Oligos are further described according to whether they span more than one exon. Oligos that are contained in a single exon have the potential to be used in CGH experiments.

Exon Type	Definition
E	Oligo is fully contained in one exon
T	Oligo spans more than one exon

Table 2 Exon Type Definitions

¹ Ensembl does not formally define gene copies. A copy here is defined as any gene entry with identical transcripts.

The oligos are also mapped with NCBI BLAST to the genome represented by the Ensembl reference chromosome sequences to determine their potential for comparative genome hybridization (CGH) experiments. An oligo that is unique is indicated as 'Predicted_CGH_Oligo' and an oligo whose maximum cross-hybridization match is less than 85% identical is indicated by 'Predicted_CGH_Oligo_Potential'. The maximum cross-hybridization score is provided.

Several microarray data analysis platforms utilize Genbank identifiers to incorporate gene annotation data. All BLAST hits greater than 97% identity are included.

IV. Characteristics of *Danio rerio* (zebrafish) AROS V4.0 Control Probes

1) Control Oligos:

- Negative Control Oligo Probes: Randomly-generated oligo sequences which were selected after filtering against the transcripts.
- Positive Control Oligo Probes: 6 oligo probes were selected as positive controls based on a literature search for genes which are constitutively expressed. The selected genes and the reference information is provided in Table 3.
- The Positive and Negative Control Oligo Probes and the corresponding Hybridization Stringency Control Oligos (see below) are randomly scattered throughout the AROS set to be used to check a variety of potential issues including wash stringency and variation of hybridization stringency in different areas of the array.

Table 3 Summary of Positive Control Selection

Control Oligo ID	Base Oligo ID	Ensembl ID	Gene Symbol	Reference
DaRePC100006	DaRe10011851	ENSDARG00000043503	Hspa8	1
DaRePC100012	DaRe10004590	ENSDARG00000019230	Rpl7a	1
DaRePC100024	DaRe10017779	ENSDARG00000056160	Cpn60	2
DaRePC100030	DaRe10006348	ENSDARG00000027552	Mapk1	3
DaRePC100036	DaRe10007368	ENSDARG00000032103	Mapk6	3
DaRePC100042	DaRe10000065	ENSDARG00000000857	Mapk14a	3

- Hybridization Stringency Control Oligos: Generated based on several positive control oligo sequences with sequence homology of 50% 60% 70%, 80%, 90% and antisense identity to the original oligo probe sequences. They're intended as the standards to calibrate the microarray hybridization conditions and stringency.
- Stratagene SpotReport™ Alien™ Spike Control Oligos: Licensed from Stratagene (www.stratagene.com). Coupled 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.

- Production Tracking Oligo opHsV04NC000001: It is a randomly-generated oligo sequence with a length of 30 bases. 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 the purpose of quality control in production. The CyDye labelled antisense probe for this oligo can be used in a QC assay to ensure the correct placement/order of the probe printing plates on the spotting robot.

References

- He, Z., Wu, L., Li, X., Fields, M.W., Zhou, J. (2005). "Empirical Establishment of Oligonucleotide Probe Design Criteria", *Applied and Environmental Microbiology*, **71**:3753-3760.
- Novère, N. (2001). "MELTING Computing the melting temperature of a nucleic acid duplex", *Bioinformatics*, 17:1226-1227.
- Wang, H., Kensinger, J.W., Zhou, Q., Wren, J.D., Martin, G., Turner, S., Young, T., Frank, M.B., Centola, M. (2007). "Identification and characterization of zebrafish ocular formation genes", Submitted
- Martin, C.C., Tsang, C.H., Beiko, R.G., Krone, P.H.. (2002). "Expression and genomic organization of the zebrafish chaperonin complex", *Genome* **45**:804-811.
- Kreus, S.F.G., He, S., Spaink, H.P., Snarr-Jagalska, B.E.. (2006). "Characterization and expression patterns of the MAPK family in zebrafish", *Gene Expression Patterns*, **6**:1019-1026.