



Fig. 1. Complete oligonucleotide selection pipeline (A) and algorithm for estimating non-target melting temperature (T_m) (B).

EST and known exon positions of its similarity-assigned gene. The non-overlapping output oligonucleotides are spread over as many exons as possible with minimum intron–exon junctions and lowest possible penalties. The complete pipeline is depicted in Figure 1A.

We have developed a novel non-target T_m algorithm that improves reduction of cross-hybridization by combining sequence similarity with thermodynamics. Using blastn, the sequence of a given oligonucleotide is aligned with the source ESTs (Fig. 1B, step 1). Source ESTs that have the same Gene ID assigned as the input EST are classified as target ESTs, while those source ESTs having a different Gene ID assigned are classified as non-target ESTs (Fig. 1B, step 2). Melting temperatures (T_m s) between oligonucleotide and non-target ESTs are computed (Fig. 1B, step 3). The maximum T_m is defined as the non-target T_m (Fig. 1B, step 4). SEPON was tested on 337 randomly picked rat ESTs that were assigned a Gene ID by their similarity to transcripts from NCBI's human RefSeq (see Supplementary information). Only 25% of the oligonucleotides with highest non-target similarity also had maximum non-target T_m leading us to introduce a user-defined number of considered non-targets. Most importantly, even for oligonucleotides with identical non-target similarity, the non-target T_m s varied 23°C demonstrating that the oligonucleotide selection pipeline is improved by the inclusion of thermodynamics in predicting cross-hybridization.

REFERENCES

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and

PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Res.*, **25**, 3389–3402.

Bozdech,Z., Zhu,J., Joachimiak,M.P., Cohen,F.E., Pulliam,B. and DeRisi,J.L. (2002) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol.*, **4**, R9.

Chen,H. and Sharp,B.M. (2002) Oliz, a suite of Perl scripts that assist in the design of microarrays using 50mer oligonucleotides from the 3' untranslated region. *BMC Bioinformatics*, **3**, 27.

Kane,M.D., Jatkoa,T.A., Stumpf,C.R., Lu,J., Thomas,J.D. and Madore,S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acid Res.*, **28**, 4552–4557.

Li,F. and Stormo,G.D. (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics*, **17**, 1067–1076.

Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. and Brown,E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.

Novère,N.L. (2001) MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics*, **17**, 1226–1227.

Rouillard,J.-M., Herbert,C.J. and Zuker,M. (2002) OligoArray: genome-scale oligonucleotide design for microarrays. *Bioinformatics*, **18**, 486–487.

SantaLucia,J. Jr. (1998) Unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci., USA*, **95**, 1460–1465.

Zucker,M., Mathews,D.H. and Turner,D.H. (1999) *Algorithms and Thermodynamics for RNA Secondary Structure Prediction*. A Practical Guide, NATO ASI Series, Kluwer, Dordrecht.