World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:6, No:3, 2012

Probe selection for pathway-specific microarray probe design minimizing melting temperature variance

Fabian Horn and Reinhard Guthke

Open Science Index, Biotechnology and Bioengineering Vol:6, No:3, 2012 waset.org/Publication/4421

Abstract—In molecular biology, microarray technology is widely and successfully utilized to efficiently measure gene activity. If working with less studied organisms, methods to design custom-made microarray probes are available. One design criterion is to select probes with minimal melting temperature variances thus ensuring similar hybridization properties. If the microarray application focuses on the investigation of metabolic pathways, it is not necessary to cover the whole genome. It is more efficient to cover each metabolic pathway with a limited number of genes. Firstly, an approach is presented which minimizes the overall melting temperature variance of selected probes for all genes of interest. Secondly, the approach is extended to include the additional constraints of covering all pathways with a limited number of genes while minimizing the overall variance. The new optimization problem is solved by a bottom-up programming approach which reduces the complexity to make it computationally feasible. The new method is exemplary applied for the selection of microarray probes in order to cover all fungal secondary metabolite gene clusters for *Aspergillus terreus*.

Keywords—bottom-up approach, gene clusters, melting temperature, metabolic pathway, microarray probe design, probe selection

I. Introduction

MICROARRAYS are a widely used technology to measure the composition of the transcriptome in an organism. Its applications range from gene expression profiling, over functional gene annotations, to medical diagnosis. With the help of standardized experimental protocols, microarrays detect gene activity reliably. In molecular biology, this technology is prominent for transcriptome analysis because it is fast and cost-efficient. In order to fully facilitate and interpret large amounts of data, the transcriptome data may be integrated with extra information from other technologies. For instance, full genomic metabolic network reconstructions are becoming more and more available and high efforts are currently made to integrate high-throughput data into these metabolic models [1]. Besides assisting the interpretation of the underlying transcriptomic data, the integration helps to reduce the complexity of metabolic networks [2].

Many prefabricated commercial and academic microarray technology platforms are available for commonly used species. If the research is focused on less studied model organisms, it is possible to design and spot oligonucleotides for custom-made technical solutions.

Fabian Horn and Reinhard Guthke are with Research Group Systems Biology/Bioinformatics,Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knöll-Institute (HKI), D-07745 Jena, Germany, E-mail: fabian.horn@hki-jena.de

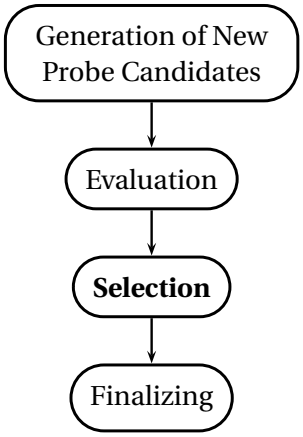

Fig. 1. **Workflow of microarray probe design** The workflow can be subdivided into four tasks. After the generation of the probe candidates, they are evaluated if they fulfill all applied probe design criteria. Afterwards, probes closer to an optimum are chosen and prepared for the final probe design. The presented approach deals with the probe selection step.

Generally, a microarray probe design task can be divided into four tasks (see figure 1). New sequences for probe candidates are generated with the help of available tools like ArrrayOligoSelector, Picky, or OligoWiz (reviewed in [3]). Furthermore, the probe candidates are evaluated whether they fulfill additional probe design criteria. In this step, specific requirements of the experimental protocol can be applied. Only probes that pass all design criteria are further processed. If more probes than the desired number of probe candidates per gene are available, probes closer to a given optimum are selected subsequently. In a last step, the selected probes are finalized, i.e. they are randomly positioned in a microarray grid.

The main design objective is the reduction of systematic errors in order to obtain highly specific and sensitive probes. It is highly desirable that all probes show a similar physicochemical behavior in the binding process. This ensures that the measured signal intensity derives from changes in the gene activity and not from different binding properties. Therefore, probes should have a high uniformity regarding their melting temperatures. The melting temperatures of the probe sequences can be calculated with the nearest-neighbor model which takes into account the base-stacking energies of certain nucleotide combinations [4]. The nearest-neighbor model is implemented in the freely available tool MELTING [5].

Open Science Index, Biotechnology and Bioengineering Vol:6, No:3, 2012 waset.org/Publication/4421

Typically, the production of custom-made microarrays is financially constrained which makes it necessary to confine the number of oligonucleotides which should be synthesized and eventually spotted onto the array. If more candidates than the desired number of probes per gene are available, an optimal subset of valid probe candidates has to be chosen. The probes are usually scored according to certain probe design criteria, e.g. uniqueness scores [6], non-overlapping sequences [7], or the positioning towards the 3'-end of the transcript [8], [9]. Different probe design criteria can also be combined using a weighting function [10], [11].

One particular utilization of microarrays is to instantly and reliably check if genetic modifications or environmental influences have a significant influence on gene expression. For instance, after gene knock-out experiments it is tested if certain metabolic pathways are extensively regulated. Due to budget limits and simplicity, metabolic pathways only need to be covered with a limited number of marker genes. The increased degree of freedom of choosing the representative marker genes generates a new optimization problem which is addressed in this paper. Even though our showcase is the coverage of metabolic pathways, the same approach can be extended to other applications, like gene-regulatory pathways or fungal gene clusters.

In this paper, we present a probe selection method which i) covers a set of metabolic pathways with a given number of genes and where simultaneously ii) only probes for these genes are selected which minimize the melting temperature variance.

II. Methods

The presented approach deals with the question of selecting optimal probesets from a given set of validated probes, meaning that each probe candidate passed all probe design criteria and the corresponding melting temperature for each probe candidate has been pre-calculated using MELTING [5].

In the first section, a probe selection approach is presented which chooses probe candidates for each gene by minimizing the overall variance of the probe melting temperatures.

In the second section, the probe selection criterion is extended to cover metabolic pathways with a limited number of marker genes. An efficient bottom-up programming approach is presented to solve this problem.

A. Selection of probes with uniform melting temperature

Problem 1 *Given a set of genes G and a set of validated probe candidates C. For each $g \in G$ there exists an associated subset of probe candidates $C_g \subset C$. Each gene g should be covered with a probeset PS which contains a given number of k probe candidates $c_i \in C_g$ while all selected probe candidates $c_i \subseteq C$ have a minimal variance in the melting temperature T_m.*

To illustrate the problem, the following example is given:

Input
Gene1: Probe 1a (53.0 °C), Probe 1b (53.5 °C), Probe 1c (55.0 °C), ...
Gene2: Probe 2a (54.8 °C), Probe 2b (55.2 °C), Probe 2c (57.0 °C), ...
...
Output (*number of probes per gene k*: 2)
T_m: 54.5 °C
minimal variance: 2.3
Gene1: Probe 1b, Probe 1c
Gene2: Probe 2a, Probe 2b
...

Definition 1 *In this paper, the term* internal variance D_{PS} *is defined as the mean absolute deviation of the melting temperatures and a certain reference melting temperature $\overline{T_m}$:*

$$D_{PS} = \frac{\sum_{i=1}^{k} |T_m^{c_i} - \overline{T_m}|}{k}$$

where k is the size a candidate probeset PS and $T_m^{c_i}$ is the melting temperature of probe candidate $c_i \in PS$.

Definition 2 *A probeset PS is called* gap-free *if it contains all possible probe candidates which fall within the melting temperature interval which is spanned by the maximum and minimum melting temperature.*

$$\forall c_i \in PS: \quad \min_{c_i \in PS} T_m^{c_i} \leq T_m^{c_i} \leq \max_{c_i \in PS} T_m^{c_i} \quad \wedge$$

$$\neg \exists c_j \notin PS: \min_{c_i \in PS} T_m^{c_i} \leq T_m^{c_j} \leq \max_{c_i \in PS} T_m^{c_i}$$

.

Problem 1 can be solved by determining the probeset with the lowest internal variance for each possible reference melting temperature. Due to the fact that not every possible reference melting temperature can be considered, we look for melting temperature intervals. Therefore, the following propositions address the question of identifying the optimal probesets for melting temperature intervals.

Proposition 1 *Probe candidates can be ordered by their melting temperature. Probesets which are gap-free with respect to the melting temperature range minimize the internal variance D_{PS}. Probesets containing gaps will always be suboptimal for any given reference melting temperature $\overline{T_m}$ because a gap-free probeset with lower internal variance can be found.*

Proof (outline): A gap-free probeset spans a narrower melting temperature interval $[\min_{c_i \in PS} T_m^{c_i}, \max_{c_i \in PS} T_m^{c_i}]$ than a probeset PS with gaps. These interval borders are used to calculate the internal variance D_{PS} with the formula given in definition 1. For any given reference melting temperature $\overline{T_m}$, there exists a narrower melting temperature interval where either a higher $\min_{c_i \in PS} T_m^{c_i}$ or lower $\max_{c_i \in PS} T_m^{c_i}$ lead to a lower internal variance D_{PS}.

As an example, let us consider figure 2. Probes are ordered according to their melting temperatures T_m. Three probesets consisting of four probes are built. Boxes are drawn for each probeset if the corresponding probe is part of the probeset.

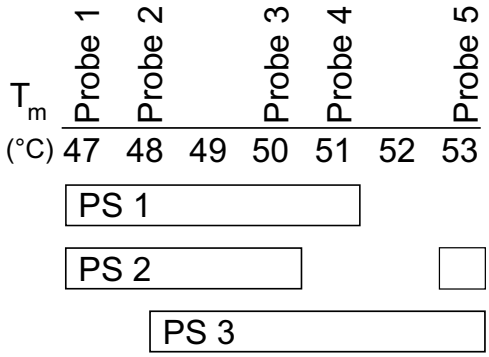

Fig. 2. **Comparison of the internal variance of gap-free probesets (PS1, PS3) and a probeset with gaps (PS2).**

Exemplary, probeset 1 consists of probe 1 (47 °C), probe 2 (48 °C), probe 3 (50 °C), and probe 4 (51 °C). Probeset 1 and probeset 3 are gap-free, meaning that all probes span a continuous melting temperature range. Probeset 2 contains a gap because the interval does not contain probe 4 (51 °C). Probesets 1 and 3 have a lower internal variance than probeset 2 because one of them is more compact regarding any specified reference melting temperature.

Consequently, only ordered probesets that span a continuous gap-free melting temperature interval are considered. This makes it easier to illustrated the problem. In figures 3 and 4, probes are displayed in the order of their corresponding melting temperature. Because only gap-free probesets are considered, the probesets can be visualized with boxes that cover a certain T_m range. These boxes can be shaded to indicate the optimal melting temperature interval for which the particular probeset has the lowest internal variance.

Proposition 2 *The mean of the melting temperatures of the two outermost probes of two adjoining melting temperature intervals determines the border between the optimal melting temperature intervals of the corresponding probesets.*

Proof (outline): Without loss of generality:

$$\frac{\sum_{i=1}^{k} |T_m^{c_i} - \overline{T_m}|}{k} \leq \frac{\sum_{i=2}^{k+1} |T_m^{c_i} - \overline{T_m}|}{k}$$

$$|T_m^{c_1} - \overline{T_m}| \leq |T_m^{c_{k+1}} - \overline{T_m}|$$

$$(T_m^{c_1} - \overline{T_m})^2 \leq (T_m^{c_{k+1}} - \overline{T_m})^2$$

$$\overline{T_m} \leq \frac{T_m^{c_1} + T_m^{c_{k+1}}}{2}$$

Proposition 3 *Probesets which flank the overall melting temperature range are optimal for the bordering intervals.*

Proof (outline): With the help of proposition 2, the border between the optimal melting temperature intervals of the first (last) interval and its successive (preceding) probeset can be determined. Since no other previous (next) probeset exists, the first (last) probeset is optimal for the overall flanking interval.

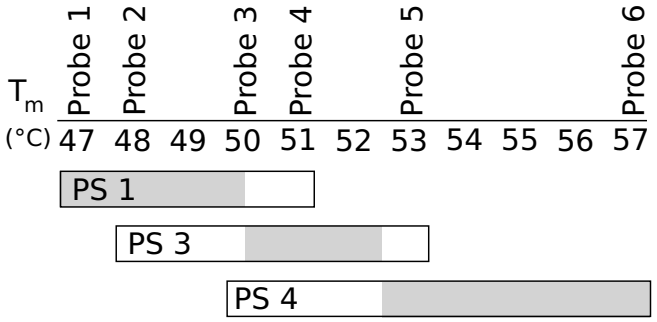

Fig. 3. **Determination of the optimal melting temperature intervals where the probeset has the lowest internal variance.** The borders between the optimal intervals are determined by the mean of the two outermost probes which are not shared by two adjoining probesets. Exemplary, probes 1 and 5 are the two outermost probes of probesets 1 and 3. Their mean melting temperature (50 °C) determines the border between the optimal intervals of these probesets.

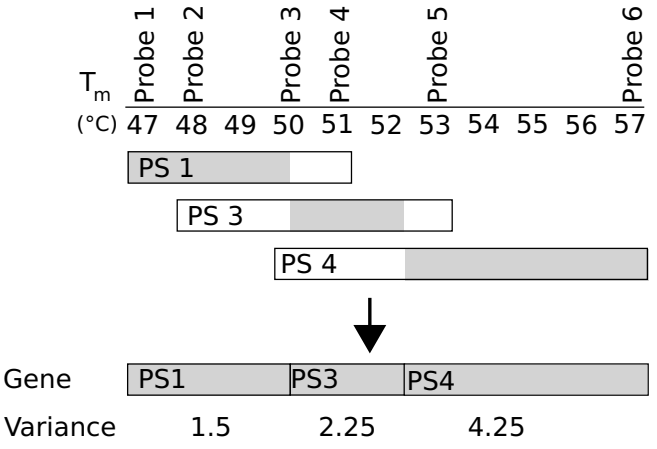

Fig. 4. **Transformation of the optimality information into a gene-centric data structure.**

For example, probesets 1, 3, and 4 consist of four probes in figure 3. Neighboring probesets share all but the two outermost probes. Precisely, probesets 1 and 3 do not share probes 1 (47 °C) and 5 (53 °C) and probeset 3 and 4 do not share probes 2 (48 °C) and 6 (57 °C). These probes determine the border of the optimal intervals between the two probesets. The mean of the melting temperature of the two outermost probes constrains the optimal temperature interval. In this example, the mean melting temperature between probe 1 and probe 5 is 50 °C and between probe 2 and probe 6 it is 52.5 °C. Because there is no previous or next probeset available in figure 3, the optimality interval of the flanking probesets can be extended to the minimum or maximum of the overall melting temperature range. The optimal intervals of the flanking probesets are extended to the overall limits of the melting temperature.

The propositions above make it possible to determine which probeset is optimal for certain melting temperature ranges. It is favorable to build a data structure that holds each optimal melting temperature interval and the corresponding probeset for each gene. In the lower part of figure 4, the information

Open Science Index, Biotechnology and Bioengineering Vol:6, No:3, 2012 waset.org/Publication/4421

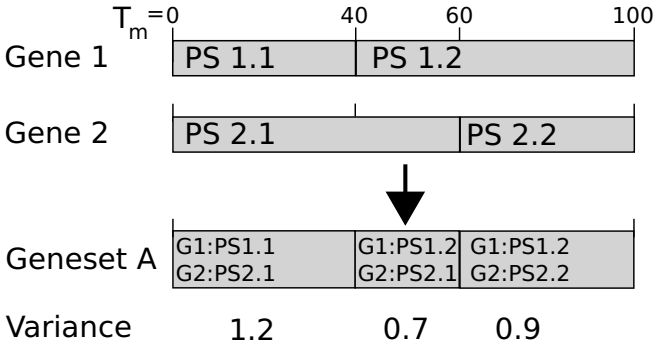

Fig. 5. **Finding the optimal probeset combination with minimal internal melting temperature variance.**

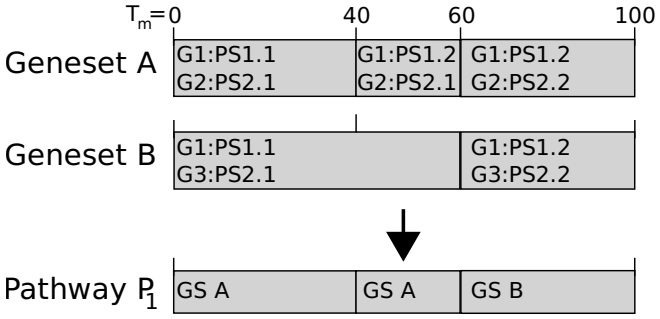

Fig. 6. **Determination of optimal genesets (GS) for each pathway.**

for the gene is merged into one data structure. Eventually, for each gene the optimal probeset and its internal variance of each melting temperature interval is known.

The data structure can be used to find the melting temperature with the lowest internal variance (see figure 5). The interval limits of the covered genes determine the solution space for the optimization problem. The emerging T_m-ranges only depend on the actual composition of the probesets. In the example of figure 5, the interval limits of 40 °C and 60 °C originate from the data structures of gene 1 and gene 2, respectively. In order to find the minimal internal variance, the variances are calculated for each melting temperature interval of interest. The resulting variances are minimal because the underlying probesets are optimal. In a final step, the overall minimum is selected and the corresponding probesets, which cover all included genes, are backtracked. In the example, the optimal solution is the interval between T_m 40 °C and 60 °C which has an internal variance of 0.7. Consequently, gene 1 is covered with probeset 1.2 whereas gene 2 is covered with probeset 2.1 in the optimal solution.

The algorithm above solves the problem that for each gene a certain number of probe candidates is selected which minimize the overall melting temperature variance.

B. Selection of optimized marker genes and their corresponding probes

Problem 2 *Given a set of pathways P and a set of associated genes G_p for each pathway $p \in P$. Given a set C_g of validated probe candidates for each gene $g \in G_p$. For each pathway $p \in P$, find a subset of genes $g_i \subseteq G_p$ of size n that are covered by k probe candidates $c_i \subseteq C_g$ which minimize the overall variance in the melting temperature T_m.*

As an example, each metabolic pathway should be covered by three genes which itself are covered by five different probes. The overall probe selection should be optimized regarding the melting temperature uniformity.

Generally, the selection of genes is not independent of the selection of probes. If different genes are selected, then a different set of probesets is optimized. Besides naive approaches (see discussion), the problem can be solved with a three-step bottom-up programming approach.

In a first step, the optimal probeset for each melting temperature range is calculated for every combination of genes in every single pathway (subsequently called genesets). Given that G_p genes are assigned to a pathway p_j and n genes should be selected, there are $\binom{G_p}{n}$ genesets for this pathway. The optimal probesets for each geneset can be calculated by the method established in section II-A (see figure 5). For each interval, the corresponding probesets are saved and the internal variance regarding these ranges are recalculated. Because each probeset is guaranteed to be optimal for the corresponding gene and temperature interval, the resulting selection for the geneset is also optimal.

In the second step, the best genesets are chosen for each pathway (see figure 6). For each geneset (GS), optimal probesets (PS) for each melting temperature interval are known from the procedure described in section II-A. For each emerging range, the minimal internal variance for each geneset is calculated. For this task, not every probeset needs to be tested again because the optimal probesets are already known. The geneset with the lowest internal variance regarding the corresponding melting temperature interval is subsequently chosen to cover the pathway. Eventually, for each pathway and each possible melting temperature the optimal geneset and its corresponding optimal probesets are known.

In a last step, it is possible to traverse this resulting information for the melting temperature range that on average contains the lowest internal variance over all pathways (similar to figure 5). This melting temperature interval can be used to backtrack the corresponding optimal genesets and their respective probeset.

III. Results and Discussion

We applied the presented method to select probe candidates for all predicted gene clusters of secondary metabolite pathways of *Aspergillus terreus*. In this context, a gene cluster is a group of genes which are closely located on the chromosome and generally encode one secondary metabolite pathway [12]. The secondary metabolite clusters were predicted based on their genomic content using the tool SMURF [13]. For probe generation and evaluation, we applied the method formerly presented [8]. In total, 81 gene clusters are covered with three different genes each. To obtain reliable results, each gene is covered with five different probes. Up to 15 genes were associated with each pathway. Additionally, five regulatory

Open Science Index, Biotechnology and Bioengineering Vol:6, No:3, 2012 waset.org/Publication/4421

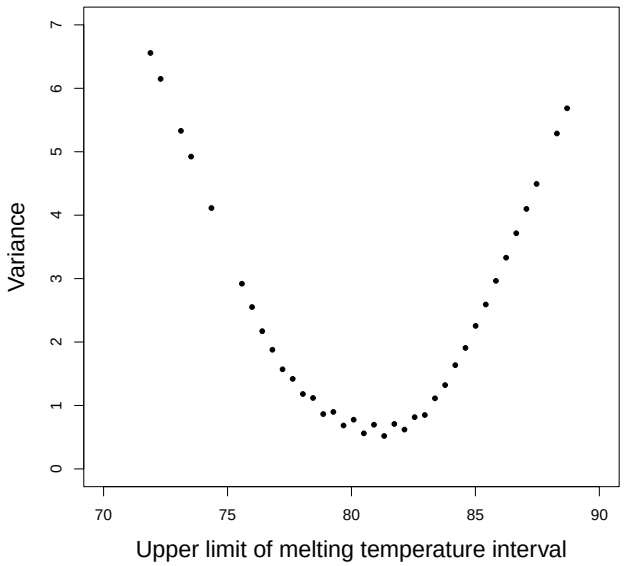

Fig. 7. **Minimal internal variance of probe selection as a function of the melting temperature for the gene cluster coverage for *A. terreus*.**

genes and seven house-keeping genes were added as control probesets to the microarray probe design.

For this application, the presented novel method selected the optimal probesets very efficiently (2 minutes using a currently standard CPU (2x2.33GHz) and 4GB RAM). The results of the optimization are visualized in figure 7 where the relationship between variances and the corresponding melting temperature intervals is shown. The graph shows that there are many melting temperature ranges which yield a low internal variance. Generally, the melting temperature should be selected as low as possible in order to ensure that hybridization takes place at low temperatures [14]. Hence, the result can be used to find a good compromise between high T_m uniformity and a low melting temperature.

In our application, an additional requirement was to include "key genes". They encode for key enzymes of secondary metabolite pathways and are well studied. It was straightforward to extend the method to take these genes into account during the selection process. Only genesets containing these genes are considered for further processing.

The proposed method is designed to find a fixed number of optimal probe candidates for a given number of genes for each pathway while minimizing the variance of the melting temperature. This optimization problem could also be addressed with naive approaches. One suboptimal approach is to determine the probeset selection for all genes at once - regardless of the pathway coverage. In a subsequent step, pathways are covered with those genes that show the lowest variance for the global optimization. However, this method is not optimal because the relationship between gene selection and probe selection is not considered.

Another possibility to get the optimal solution is to generate all possible geneset combinations and use the approach from section II-A. Afterwards, the minimal internal variance is known for each possible combination of genesets and the combination with the lowest variance is finally chosen. Even for a small number of pathways and associated genes, this approach is not computationally feasible. For this naive approach, the number of possible combinations is $\prod_{p \in P} \binom{G_p}{n}$, where p is the index for possible pathways within P, G_p is the number of available genes for this pathway, and n is the number of genes that cover each pathway. Our bottom-up approach presented in section II-B only needs $\sum_{p \in P} \binom{G_p}{n}$ combinations for the calculation of the optimal probeset and their corresponding melting temperature ranges. The complexity of the optimization problem is significantly decreased using the bottom-up programming approach because it does not calculate all possible probeset combinations.

Generally, there are many probe candidates which have the same predicted melting temperature. The presented method above does not explicitly account for the fact that several combinations of different probe candidates can span the same melting temperature range. Nevertheless, the optimality criterion is valid regarding the melting temperature interval examined. In the last backtracking step, the method allows to specifically select all probe candidate combinations for the selected range. If several probesets offer an optimal solution, additional selection criteria can be applied. For instance, the probeset which is located closest to the 3'-end of the transcript may be preferred.

In the presented application, the method was limited to secondary metabolite pathways, even though, the concept can be generalized to any classification of genes, e.g. regulatory pathways or diseases-associations.

IV. Conclusion

We established a novel efficient method for microarray probe selection. Due to limited financial resources, the objective is to cover metabolic pathways with a limited number of genes and a limited number of probes per gene. The optimization considers the constraints regarding the number of probes while simultaneously minimizing the melting temperature variance of all selected probes. The problem is solved with a bottom-up programming approach which effectively reduces the complexity of the optimization problem and makes it computationally feasible. The resulting set of selected probe candidates is optimal regarding a certain melting temperature interval for which the variance of the set is minimal. The selected probes have the advantage of similar physicochemical properties which guarantee highly reliable signal intensities. An exemplary application of the probe selection is presented for the coverage of all fungal gene clusters involved in the secondary metabolism of *Aspergillus terreus*. The method was implemented using Perl. It is available at www.sysbio.hki-jena.de/software/.

Acknowledgment

This work was supported by the International Leibniz Research School for Microbial and Biomolecular Interactions

Open Science Index, Biotechnology and Bioengineering Vol:6, No:3, 2012 waset.org/Publication/4421

(ILRS Jena) as part of the excellence graduate school Jena School for Microbial Communication (JSMC), supported by the Deutsche Forschungsgemeinschaft. Thanks to Markus Greßler for providing the gene cluster list for *Aspergillus terreus* and thanks to Nina Kottenhagen for proofreading.

REFERENCES

[1] P. A. Jensen and J. A. Papin, "Functional integration of a metabolic network model and expression data without arbitrary thresholding." *Bioinformatics*, vol. 27, no. 4, pp. 541–547, Feb 2011. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/btq702

[2] J.-M. Schwartz, C. Gaugain, J. C. Nacher, A. de Daruvar, and M. Kanehisa, "Observing metabolic functions at the genome scale." *Genome Biol*, vol. 8, no. 6, p. R123, 2007. [Online]. Available: http://dx.doi.org/10.1186/gb-2007-8-6-r123

[3] S. Lemoine, F. Combes, and S. L. Crom, "An evaluation of custom microarray applications: the oligonucleotide design challenge." *Nucleic Acids Res*, vol. 37, no. 6, pp. 1726–1739, Apr 2009. [Online]. Available: http://dx.doi.org/10.1093/nar/gkp053

[4] J. SantaLucia and D. H. Turner, "Measuring the thermodynamics of RNA secondary structure formation." *Biopolymers*, vol. 44, no. 3, pp. 309–319, 1997. [Online]. Available: http://dx.doi.org/3.0.CO;2-Z

[5] N. L. Novère, "MELTING, computing the melting temperature of nucleic acid duplex." *Bioinformatics*, vol. 17, no. 12, pp. 1226–1227, Dec 2001.

[6] S. Gräf, F. G. G. Nielsen, S. Kurtz, M. A. Huynen, E. Birney, H. Stunnenberg, and P. Flicek, "Optimized design and assessment of whole genome tiling arrays." *Bioinformatics*, vol. 23, no. 13, pp. i195–i204, Jul 2007. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/btm200

[7] L. Jourdren, A. Duclos, C. Brion, T. Portnoy, H. Mathis, A. Margeot, and S. L. Crom, "Teolenn: an efficient and customizable workflow to design high-quality probes for microarray experiments." *Nucleic Acids Res*, vol. 38, no. 10, p. e117, Jun 2010. [Online]. Available: http://dx.doi.org/10.1093/nar/gkq110

[8] F. Horn, H.-W. Nützmann, V. Schroeckh, R. Guthke, and C. Hummert, "Optimization of a microarray probe design focusing on the minimization of cross-hybridization," in *Proceedings of the International Conference on Bioinformatics and Computational Biology (BIOCOMP'11)*, H. R. Arabnia and Quoc-Nam, Eds., 2011, ISBN: 1-60132-172-4.

[9] H.-W. Nützmann, Y. Reyes-Dominguez, K. Scherlach, V. Schroeckh, F. Horn, A. Gacek, J. Schümann, C. Hertweck, J. Strauss, and A. A. Brakhage, "Bacteria-induced natural product formation in the fungus Aspergillus nidulans requires Saga/Ada-mediated histone acetylation." *Proc Natl Acad Sci U S A*, vol. 108, no. 34, pp. 14 282–14 287, Aug 2011. [Online]. Available: http://dx.doi.org/10.1073/pnas.1103523108

[10] F. Bidard, S. Imbeaud, N. Reymond, O. Lespinet, P. Silar, C. Clavé, H. Delacroix, V. Berteaux-Lecellier, and R. Debuchy, "A general framework for optimization of probes for gene expression microarray and its application to the fungus Podospora anserina." *BMC Res Notes*, vol. 3, p. 171, 2010. [Online]. Available: http://dx.doi.org/10.1186/1756-0500-3-171

[11] R. Wernersson and H. B. Nielsen, "OligoWiz 2.0–integrating sequence feature annotation into the design of microarray probes." *Nucleic Acids Res*, vol. 33, no. Web Server issue, pp. W611–W615, Jul 2005. [Online]. Available: http://dx.doi.org/10.1093/nar/gki399

[12] A. A. Brakhage and V. Schroeckh, "Fungal secondary metabolites - strategies to activate silent gene clusters." *Fungal Genet Biol*, vol. 48, no. 1, pp. 15–22, Jan 2011. [Online]. Available: http://dx.doi.org/10.1016/j.fgb.2010.04.004

[13] N. Khaldi, F. T. Seifuddin, G. Turner, D. Haft, W. C. Nierman, K. H. Wolfe, and N. D. Fedorova, "SMURF: Genomic mapping of fungal secondary metabolite clusters." *Fungal Genet Biol*, vol. 47, no. 9, pp. 736–741, Sep 2010. [Online]. Available: http://dx.doi.org/10.1016/j.fgb.2010.06.003

[14] G. Bhanot, Y. Louzoun, J. Zhu, and C. DeLisi, "The importance of thermodynamic equilibrium for high throughput gene expression arrays." *Biophys J*, vol. 84, no. 1, pp. 124–135, Jan 2003. [Online]. Available: http://dx.doi.org/10.1016/S0006-3495(03)74837-1

Open Science Index, Biotechnology and Bioengineering Vol:6, No:3, 2012 waset.org/Publication/4421