

A Strategy to Optimize the Oligo-Probes for Microarray-based Detection of Viruses*

Zhuo ZHOU^{1#}, Zhi-xun DOU^{2#}, Chen ZHANG¹, Hou-qing YU², Yi-jie LIU³,
Cui-zhu ZHANG^{1,2} and You-jia CAO^{1,2,4 **}

(1. Department of Biochemistry and Molecular Biology, College of Life Sciences, Nankai University, Tianjin 300071, China; 2. The Key Laboratory of Bioactive Materials, Ministry of Education, Nankai University, Tianjin 300071, China; 3. School of Computing, National University of Singapore, 3 Science Drive 2, 117543, Singapore; 4. Tianjin Key Laboratory of Microbial Functional Genomics, College of Life Sciences, Nankai University, Tianjin 300071, China)

Abstract: DNA microarrays have been acknowledged to represent a promising approach for the detection of viral pathogens. However, the probes designed for current arrays could cover only part of the given viral variants, that could result in false-negative or ambiguous data. If all the variants are to be covered, the requirement for more probes would render much higher spot density and thus higher cost of the arrays. Here we have developed a new strategy for oligonucleotide probe design. Using type I human immunodeficiency virus (HIV-1) *tat* gene as an example, we designed the array probes and validated the optimized parameters *in silico*. Results show that the oligo number is significantly reduced comparing with the existing methods, while specificity and hybridization efficiency remain intact. The adoption of this method in reducing the oligo numbers could increase the detection capacity for DNA microarrays, and would significantly lower the manufacturing cost for making array chips.

Key words: Microarray; Oligonucleotide; Viral Detection

The detection of viral pathogens is an essential process in many areas. DNA microarray offers a high-throughput approach that has been proved to be a tool capable of detecting viral pathogens in a precise and sensitive way (17). However, some bottlenecks in the technique have been recognized, including nucleotide probe spotting/synthesis and probe design. In the production of microarrays, the density-versus-cost

issue has been a major problem. Thus, for the high-throughput viral detection, new approaches are needed to reduce the oligo numbers without compromising the detection capacity and efficiency. The design of suitable sets of oligonucleotide probes is another important step in DNA microarray experiments. For viral detection, a number of groups have designed oligos for representative strains, but the remaining

Received: 2007-01-08, Accepted: 2007-05-20

* Foundation items: Research funding for undergraduate students of Nankai University; NSFC grant (30270308); NSFC grant (30370053); Tianjin grant (05YFJZJC01301).

These authors contribute equally to the work.

** Corresponding author. Tel/Fax: +86-22-23500808. E-mail: caoyj@nankai.edu.cn

variants are beyond the capacity of the probes (2, 3, 11, 16). An alternative method involves the design of oligos within conserved regions to cover a wider range of variants (4, 8, 15). However, a significant amount of variants are still overlooked by this method, such as those with highly mutated sequences with insertions or deletions at the sites of probes (6). To avoid false-negative or ambiguous results, one can increase the number of oligos designed based on sequences in available databases to cover as many known variants as possible. However, this will lead to the higher probe numbers.

An effective oligo-probe set for viral detection should satisfy several criteria. The specificity is of the highest priority, and has been studied intensively. Using current computer programs (such as OligoArray 2.1) to design 50-mer probes for the type I human immunodeficiency virus (HIV-1) *tat* genes, we found that over 50% of the probes have the risk of cross-hybridization according to Kane's specificity criteria (7). The common problem using oligos longer than 50-mer is that the specificity is reduced with the increase in oligo length (12). Thus, a proportion of viral variants would not have specific probes if 50-mer or longer oligo designing methods were used. An alternative strategy is to design probes using the GoArrays approach. In this strategy, the oligonucleotide probe consists of the concatenation of two subsequences that are complementary to their target cDNA with an insertion of a short random linker, *e.g.* 6 bases, to facilitate the formation of the transcript loop (Fig. 1A). This strategy has been shown to broaden the range of oligo selection and increase the specificity and sensitivity (12). In addition to the specificity criteria, the oligo numbers should also be

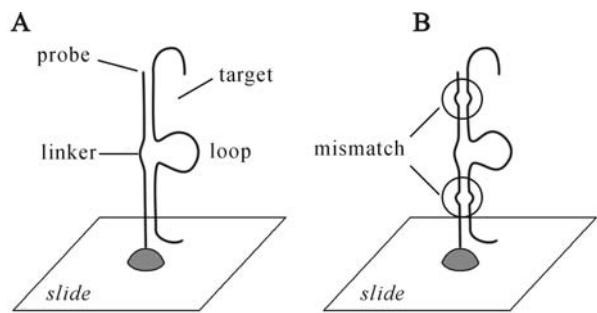


Fig. 1. Schemes of existing and newly developed strategies for oligo design. A: The GoArrays strategy (12), in which the oligo-probe is composed of two disjointed subsequences from one coding sequence (CDS). Subsequences are concatenated *via* a short linker, *e.g.*, 6 bases. The hybridization between the composite probe and the target induces the formation of a loop, as depicted. B: The modified strategy, in which the mismatch theory is adapted into the GoArrays method, *i.e.*, non-perfect match (es) is incorporated in the concatenated subsequences that were selected from conserved regions.

considered. When we tested the existing methods, such as to design oligos within the conserved regions or by the GoArrays approach, we found that all methods would require at least 1500 probes to cover the 1881 *tat* variants. Therefore, a strategy that circumvents the specificity and oligo number limitations is needed for the microarray-based viral detections.

In this paper, we describe a new strategy for oligonucleotide design aiming to cover all known variants for the viral species of interest with minimal oligo numbers. This approach was developed on the basis of the GoArrays approach (12) and the non-perfect match theory (9). Our results show that the oligo number is significantly reduced in comparison with using other published methods, while the specificity and hybridization efficiency remain intact. This approach could be generally applied to oligo-design or the development of low cost and high throughout viral detection arrays for clinical applications.

MATERIALS AND METHODS

Database and programs

HIV-1 sequences corresponding to exon 1 of the *tat* gene for 2605 strains (by August, 2005) were downloaded from the LANL HIV Sequence Database (<http://hiv-web.lanl.gov/content/hiv-db/mainpage.html>). The aligned *tat* database (in FASTA format) was obtained by searching for all *tat* exon 1 at http://hiv-web.lanl.gov/components/hiv-db/combined_search_s_tree/search.html. Redundant sequences and sequences with bases other than A, T, G, C were removed. Specially, strain AF443106 was removed because it lacks a large portion of nucleotides. After the filtering steps, *tat* exon 1 sequences of 1881 strains were collected.

The BLAST program version 2.2 was downloaded from <http://www.ncbi.nlm.nih.gov/BLAST/> download.shtml (1). The BioPerl (<http://bio.perl.org/>) Stand AloneBlast module was used to invoke, parse the result of the blastall program, and to check the specificity of probes designed. All BLAST programs were run using the parameter '-p n' to specify blast on nucleotides, and with default settings for other parameters. Unless indicated, the BLAST score of 32.2 for a 50-mer probe was used as specificity threshold. The human mRNA database used for BLAST to ensure specificity was downloaded from ftp://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/mRNA_Prot/.

OligoArray 2.1 (13) was downloaded from http://berry.engin.umich.edu/oligoarray2_1/; the required OligoArrayAux was downloaded from <http://www.bioinfo.rpi.edu/applications/hybrid/OligoArrayAux.php>. We set the oligo length 50-mer; maximum oligo number 1; other parameters were set as default.

GoArrays (12) was downloaded from <http://www.isima.fr/bioinfo/goarrays/>, and run following the

authors instructions. Maximum length for identity was set as 16; subsequence length was set to 22; other parameters were as default.

Melting temperature (Tm) calculation

The Tm value was calculated by the program downloaded from DINAMelt web server (10) (<http://www.bioinfo.rpi.edu/applications/hybrid/hybrid2.php>) and run locally as instructed. The initial concentrations of two strands were set as 10^{-7} mol/L; $[Na^+]$ was set as 0.5 mol/L; other parameters were set as default (see 'specificity evaluation' for explanation of these choices).

Free energy and hybridization structure prediction

The free energy (dG) of hybridization was calculated by Mfold web server (18) (<http://www.bioinfo.rpi.edu/applications/mfold/>). To predict the dG of structure stability, the probe sequence was fixed whereas the target sequences varied. Point mutation(s) in the matched regions was generated by replacing perfectly matched pair(s) with one of the other three non-matching nucleotides.

Determination of conserved regions

A conserved region of certain length of nucleotides is determined by the variety within each window, which shifts from the beginning to the end of the exon by a fixed offset of 5 nt (slide the window along the sequence in 5 nt steps) for each strain in the database. A window with a higher number of unrepeatable subsequences is considered to be of higher variety and less conserved. Each sequence selected from a conserved region was subjected to the BLAST program to confirm its specificity as to be used as a probe.

RESULTS

Strategies to reduce the number of oligos

We first tested whether concatenating two shorter conserved sequences by the GoArrays strategy could be a practicable way to reduce oligo numbers. We obtained several conserved regions (22-bases each) of the HIV-1 *tat* gene as targets for design, and developed a corresponding program (see ‘dynamic programming’ below) to link two subsequences selected from the conserved region pools. 660 specific probes were deduced to cover all 1881 variants, suggesting that selecting shorter conserved oligos can help to reduce oligo numbers.

Incorporating appropriate mismatches into oligo-probes could be another way to reduce probe numbers, while the feasibility of non-perfectly matched probe design has been highlighted by other studies (9). To study the capability of mismatches in reducing oligo numbers, we first chose a short conserved region (*e.g.*, the 22-base sequences beginning at position 106 in the *tat* exon 1 alignment), then calculated the amount of oligos and their inclusive strains at given number of

mismatches. As shown in Fig.2, the more mismatches that were incorporated, the fewer the number of oligos that were required while maintaining the same coverage. Choosing other regions and/or other oligo length (>20-mer) showed the same trend (data not shown). The results demonstrate that introducing non-perfectly matched probes can help to reduce the number of oligos.

The above approaches to reduce oligo numbers prompted us to develop a new strategy based on the GoArrays method and the mismatch theory in order to get the minimal number of probes. The scheme of our newly developed strategy is shown in Fig.1B. Here, designed mismatches are incorporated in the matched subsequences selected from the conserved regions. In the following results, we describe the program to optimize oligo number and specificity, and then validate the related parameters to demonstrate the feasibility of the strategy.

Optimizing oligo numbers by dynamic programming

Prior to optimizing the oligo numbers, we determined the conserved regions for subsequence pools, and then optimized the combination of linkages between every two pools using a dynamic programming algorithm in order to obtain the maximal coverage while using the minimal number of concatenated probes. Here we define a ‘linkage’ as a concatenation of two subsequences from two separate pools, whereas ‘combination’ as a set containing one or more linkages selected from the set of all possible linkages (scheme seen in Fig.3).

The first step was selection of two regions for linking. By determination of conserved regions (described above in Methods), we were able to obtain

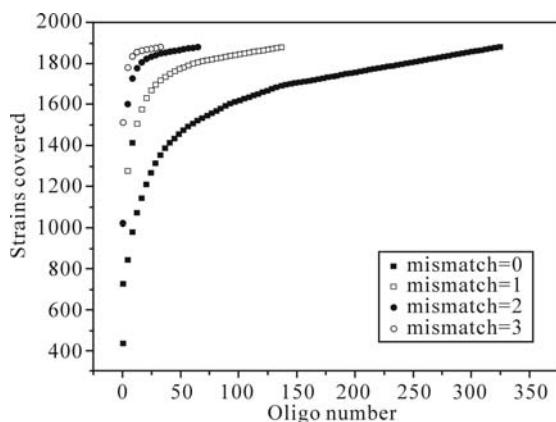


Fig. 2. The relationship between calculated strain numbers and oligo numbers with different mismatches. The region used for calculation is 22-bases, beginning at the position 106 in the *tat* exon 1 alignment. Curves representing the number of mismatches are indicated in the inset. To cover all 1881 strains, there are 326, 138, 65 and 34 oligos corresponding to 0, 1, 2, and 3 mismatches, respectively.

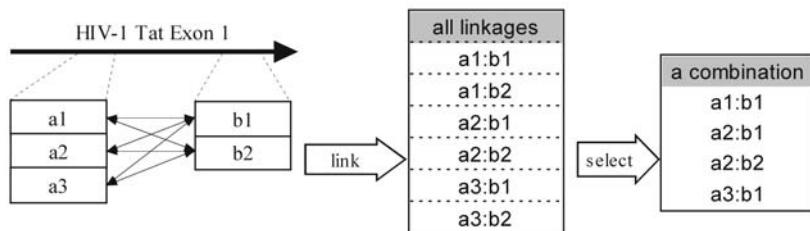


Fig. 3. Scheme of linkage and combination. (Left) The axis indicates the HIV-1 *tat* 1st exon used for design. Open boxes represent different oligos within two regions. It is important to note that there is more than one way to link the subsequences. (Middle) A list shows all ways of linkages in the given situation. (Right) An instance of a combination, which is a subset of all linkages.

several candidates (the length of each subsequence is 22 bases, see ‘specificity evaluation’ below). We then selected these regions from aligned HIV-1 *tat* database to get the subsequence pools. Next, redundant sequences in each pool were removed, and the rest were subsequently filtered to ensure specificity by the BLAST program. Thus, all eligible sequences in each pool were identified and saved.

The next step was to optimize the combination of linkages using dynamic programming in order to obtain the maximal coverage while using the minimal number of probes. If the oligos in two subsequence pools are to be linked, the length of the linkers should fall between 3 and 100 bases, according to the

restrictions by GoArrays strategy (12). Meanwhile, a probe must satisfy the condition that no more than 1 mismatch in each subsequence (see ‘specificity evaluation’ below). To get an optimized $(n+1)$ -linkage combination, we need an optimized n -linkage combination and a selected linkage, which, after joining in the n -linkage combination, makes the new $(n+1)$ -linkage combination cover the maximal number of targets (Fig.4). This dynamic algorithm ensures that the coverage at each step is maximized, and that the number of probes to cover the corresponding variants is minimized. The best resulting combination can cover 1875 out of 1881 strains with only 272 probes. The remaining 6 undetectable strains resulted from the

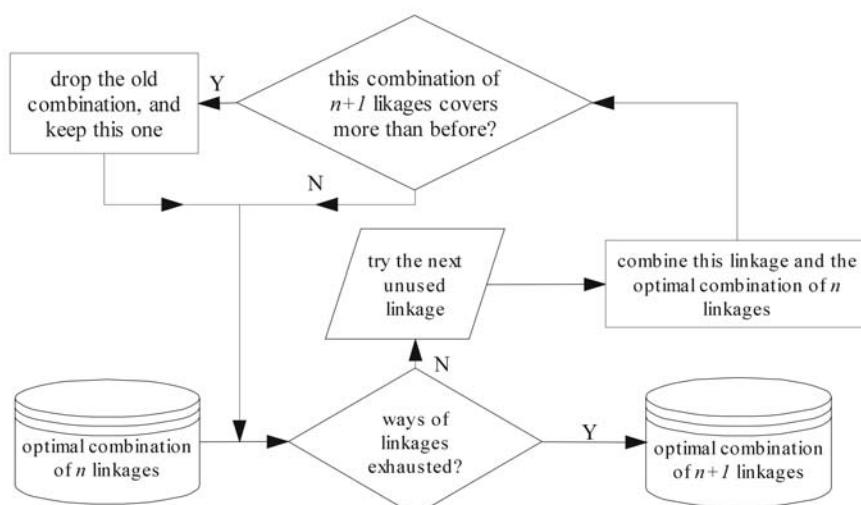


Fig. 4. Flowchart of optimizing a combination by dynamic programming (a single step of recurrence). A set of optimal n -linkage combination is input in the recurrence, and all possible ways of linkages are exhausted to find a best linkage in order to get the optimal $(n+1)$ -linkage combination. Such steps are repeated to complete the whole step. See text for more details.

preparation step were saved in a file for future use. Therefore, our strategy greatly reduces the number of oligos to 272, from over 1800 by GoArrays approach (12) or over 1500 by 50-mer or longer oligos with conserved region method (15).

A random linker of 6 bases was then concatenated between each subsequence. The specificity of every linked probe was again confirmed by the BLAST program.

***In silico* evaluation of specificity**

The melting temperature (T_m) has been widely used in simulating hybridization conditions as well as cross-hybridization (9). In this study, we used the DINAMelt web server to calculate the T_m and the melting curves (presented by hybridization rate of targets/non-targets and their probes as a function of temperature). The deviations of the model were evaluated first. We compared the calculated T_m with the experimental dissociation temperature (T_d) from two groups of sequences (55 ones in total) that contain single point mutation in different site, by Urakawa and colleagues (14). The linear correlation coefficients R of the two groups were 0.79 and 0.88. Using conditions described in ‘Materials and Methods’, as commonly used in software and experiments, the correlation coefficients were found to be better than those obtained with default parameters (R equals 0.78 and 0.84 respectively). Although the calculated T_m are about 20°C higher than the T_d data, the absolute value is less essential than the overall correlation, since we can simply reduce the 20°C for each datum. All T_m were therefore calculated by the DINAMelt with the modified parameters.

The length of the subsequence and the amount of mismatch incorporated in the oligo are directly related

to the specificity of the probe. Longer oligos and fewer mismatches contribute to the specificity, whereas shorter oligos and more mismatches have the advantage in reducing oligo numbers. To balance the relative merits of these two extremes, we sought a combination of these two requirements that maintained specificity with a manageable number of oligos. After a series of calculation and comparison (data not shown), we found an optimized parameter: 22-bases for subsequence length and no more than one mismatch for each subsequence (scheme shown in Fig.1B). This parameter ensures specificity (as measured by the hybridization rate of target- hybridization and cross-hybridizations described below) and minimized oligo numbers (as compared and calculated by dynamic programming described above).

The specificity evaluation was here presented. Among the designed probes, the probe with the lowest specificity was identified from the BLAST results. This probe hits many unrelated sequences in the specificity database. Based on the calculated melting curves of this probe with the various sequences (targets and non-targets), we then generated the melting profiles for overall hybridization (data not shown). Using these profiles, we were able to predict the hybridization rate of this probe with its targets or non-targets at given temperature. The criteria for a specific hybridization, as commonly used by previous studies (9), are that hybridization rate for target sequences is over 95%, and for non-target sequences, less than 15%. To simplify the scenario, we picked 64°C to distinguish the specific and non-specific hybridization. The calculated T_m and hybridization rate of selected sequences are shown in Table 1. The extremity of undesirable condition is represented by

the target AF325758 and the non-target Cross 5. AF325758 has the lowest target-hybridization rate among all the targets, while Cross 5 has the highest cross-hybridization rate among all the non-targets. As seen in Table 1, the hybridization rate of AF325758 and Cross 5 still satisfies the specificity criteria mentioned, indicating the specificity of the tested probes. Since the probe used in above process was of the lowest specificity, it is evident that all the other designed probes would meet the criteria as defined above.

Table 1. Calculated hybridization Tm and corresponding hybridization rate of selected sequences at 64°C.

Targets/Cross	Tm(°C)	Hybridization rate (%)
DQ011180	75	100 ^a
AY253319	74	99 ^a
AF325758	72	96 ^a
Cross 2	53	13 ^b
Cross 5	56	14 ^b
Cross 15	47	4 ^b
Cross 16	46	0 ^b

^aExpected to be positive hybridization (>95%). ^bexpected not to contribute to cross-hybridization (<15%) at assay temperature.

Probe sequence: GCTTCGTCGCTGTCTCCGCTTCgacactAA GCAAACATAAACAGTGGTAGC; lower case letters represent linker sequence; probe length 50-mer. Accession number of cross sequences is NM_000361.2, NM_022720.5, NM_024101.4 and NM_005292.2 corresponding to Cross 2,5,15 and 16, respectively.

Thermodynamic validation of hybridization stability

The stability of the hybridization structure is of the most important aspect in hybridization efficiency and sensitivity for DNA microarrays. In our design, the influential factors affecting stability are the mismatches and the internal loop. We therefore investigated these two quantities.

To examine the influence of mismatch on the

stability of hybridization, the free energy (dG) was calculated using the same probe and mismatch parameters determined above. To test a more general situation, we deliberately introduced series of non-targeting sequences with satisfied mismatch threshold (referred to below as artificial sequences). The change in dG was measured by ΔdG ($\Delta dG = dG_{mismatched} - dG_{perfect\ matched}$). For the actual targets, a single point mutation yields ΔdG 1.5~4.5 kcal/mol, whereas the maximum ΔdG can reach to 5.5 kcal/mol for artificial sequences, as shown in Fig.5. For double point mutations in the investigated probe (one point mutation in each subsequence), the ΔdG is 3~9 kcal/mol for actual targets, while it can be as high as 11 kcal/mol for artificial sequences.

To examine the influence of internal loop on the stability, we generated one artificial ‘open loop’, wherein none of the nucleic acids could match. This represents the extreme scenario in which the hybridization stability is ruined. The change in dG is still measured by ΔdG ($\Delta dG = dG_{modified} - dG_{perfect\ matched}$). We found that the greatest ΔdG could reach as high as 6.5 kcal/mol for artificial sequences. In contrast, the ΔdG was 0~4.1 kcal/mol for actual targets. The influence of loop on the stability is therefore a less important factor compared with that of mismatches in subsequences.

As far as both mismatches and loop are concerned, the dG for actual targets is -48.5 to -54.2 kcal/mol, whereas the greatest dG for artificial sequences is -40.1 kcal/mol. In the GoArrays strategy, a 20-base-subsequence probe with the dG around -40 kcal/mol can hybridize even more efficiently than the classical 50-mer probes (12). The thermodynamic data show that the dG of our probes is much lower than, or

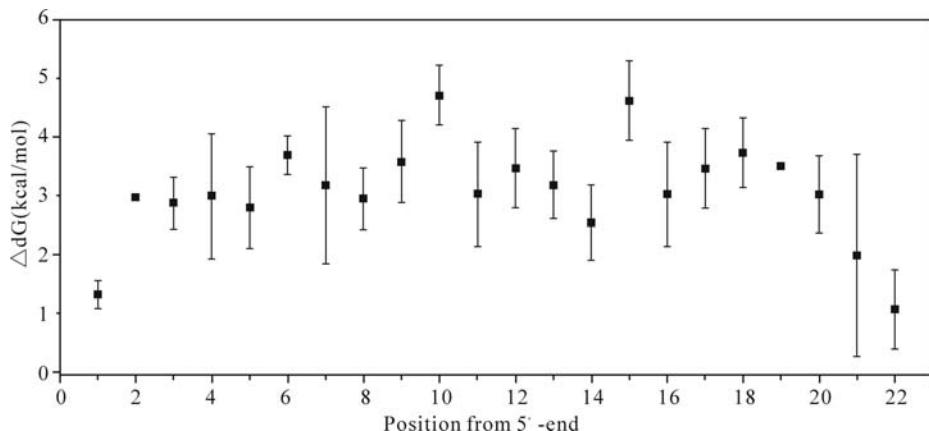


Fig.5. Calculated ΔdG for a 22-mer oligo with nucleotide mutation. Average ΔdG values of each mutation at different positions are shown with standard deviations. Probe sequence: **GCTTCGTCGCTGTCTCCGCTTC**gacactAAGCAAACTAACAGTGGTA GC; lower case letters represent linker sequence; bold letters represent the subsequence investigated in the graph. Results in another subsequence are similar to this graph (data not shown).

comparable with -40kcal/mol, indicating the intact hybridization stability of those probes designed by our strategy.

Further verification of the probe set and the guidelines for viral probe design

Other parameters affecting hybridization, such as Tm range, secondary structure and prohibited sequences, were considered (data not shown). Ineligible probes were excluded and re-designed, and 9 additional probes were added. Thus, the total probe number to sufficiently detect all 1881 HIV-1 *tat* variants should be 281 (272+9). In conclusion, the oligo number in our design strategy is significantly reduced while the specificity and sensitivity of the probes are better than, or comparable with other reported studies mentioned.

In summary, our strategy for designing viral detection probes can be summarized by the following four steps. (1) Determination of suitable 22-mer regions which are conserved and specific. Sequences in two of the selected regions will be used as subsequence candidates. (2) Pre-filtering of subse-

quence candidates for each of the selected regions. Identical sequence from different strains should appear only once; prohibited sequences and sequences with BLAST scores higher than 32.2 (for 50-mer probes) should be eliminated. (3) Linkage of subsequences into probes, using the 'dynamic programming' described above (the computer program implementing this algorithm step is freely available on request: zhangchn2004@gmail.com). (4) Examination of the designed probes for Tm consistency, secondary structures, and other parameters.

DISCUSSION

During our *in silico* study, we found that existing specificity criteria for oligonucleotides (7,13) can be further improved. More recently, He Z. and colleagues proposed an improved criteria which combines identity, stretch and free energy (5). However, we noticed that combining even these three parameters is still not sufficient to ensure specificity, because we were able to create sequences that satisfy all three conditions but still produce non-specific cross-

hybridization regardless of hybridization temperature (our unpublished data). Therefore, other thermodynamic parameters, such as Tm and the hybridization rate, should be investigated.

An requirement for DNA microarrays in virological applications is that the methodology could detect unknown variants in the given viral species as well as previously identified variants (15). However the rapid point mutations and segment reassortment of viruses have been a challenge for probe design. When we updated the database in October 2005, we found that there were another 58 new variants posted for HIV-1 *tat* in addition to the database in August that year. Promisingly, 46 of the 58 new variants can be covered in the 281 probe set as we have deduced, indicating the powerful detection capacity of the probe design.

Further considerations should also be noted. First, as to the probe number, there is still room for improvement. We found that a number of probes contribute only one target in later steps of recurrence during the dynamic programming. To synchronously perform linking combination at more than one pair of regions could be one solution. Another plausible way is to design artificial sequences that do not exist in the subsequence pool but satisfy the mismatch threshold for the targets in later steps. Second, although our calculation tools are currently most updated and the *in silico* data are statistically reliable, there are still certain deviations compared with experimental results. Finally, our strategy and the validations are based on theoretical model, therefore further experimental verifications are imperative.

Acknowledgement

We are grateful to Professor Huxin Diao for valuable suggestions. We thank Jun Tang and Jia Xie

for the critical reading of our manuscript.

References

1. Altschul S F, Madden T L, Schaffer A A, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25 (17): 3389-3402.
2. Boriskin Y S, Rice P S, Stabler R A, et al. 2004. DNA microarrays for virus detection in cases of central nervous system infection. *J Clin Microbiol*, 42 (12): 5811-5818.
3. Bystricka D, Lenz O, Mraz I, et al. 2005. Oligonucleotide-based microarray: a new improvement in microarray detection of plant viruses. *J Virol Methods*, 128 (1-2): 176-182.
4. Chizhikov V, Wagner M, Ivshina A, et al. 2002. Detection and genotyping of human group A rotaviruses by oligonucleotide microarray hybridization. *J Clin Microbiol*, 40 (7): 2398-2407.
5. He Z, Wu L, Li X, et al. 2005. Empirical establishment of oligonucleotide probe design criteria. *Appl Environ Microbiol*, 71 (7): 3753-3760.
6. Ivshina A V, Vodeiko G M, Kuznetsov V A, et al. 2004. Mapping of genomic segments of influenza B virus strains by an oligonucleotide microarray method. *J Clin Microbiol*, 42 (12): 5793-5801.
7. Kane M D, Jatkoe T A, Stumpf C R, et al. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res*, 28 (22): 4552-4557.
8. Korimboccus J, Scaramozzino N, Lacroix B, et al. 2005. DNA probe array for the simultaneous identification of herpesviruses, enteroviruses, and flaviviruses. *J Clin Microbiol*, 43 (8): 3779-3787.
9. Lee I, Dombkowski A A, Athey B D. 2004. Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a DNA microarray. *Nucleic Acids Res*, 32 (2): 681-690.
10. Markham N R, Zuker M. 2005. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res*, 33: W577-W581.
11. Oh T J, Kim C J, Woo S K, et al. 2004. Development and clinical evaluation of a highly sensitive DNA microarray for detection and genotyping of human papillomaviruses. *J Clin Microbiol*, 42 (7): 3272-3280.

12. **Rimour S, Hill D, Militon C, et al.** 2005. GoArrays: highly dynamic and efficient microarray probe design. *Bioinformatics*, 21 (7): 1094-1103.
13. **Rouillard J M, Zuker M, Gulari E.** 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res*, 31 (12): 3057-3062.
14. **Urakawa H, Fantroussi S E, Smidt H, et al.** 2003. Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Appl Environ Microbiol*, 69 (5): 2848-2856.
15. **Wang D, Coscoy L, Zylberberg M, et al.** 2002. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA*, 99 (24): 15687-15692.
16. **Wilson W J, Strout C L, DeSantis T Z, et al.** 2002. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes*, 16 (2): 119-127.
17. **Young R A.** 2000. Biomedical discovery with DNA arrays. *Cell*, 102 (1): 9-15.
18. **Zuker M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31 (13): 3406-3415.