A simulation of subtractive hybridization

Tae-Ju Cho* and Sang-Soo Park

Department of Biochemistry, College of Natural Sciences, Chungbuk National University, Cheongju 361-763, Korea

Received November 21, 1997; Revised and Accepted January 6, 1998

ABSTRACT

Various strategies employed in genomic DNA cloning by subtractive hybridization have been examined by computer simulations, with the comparison between the predictions and the published results. The results show that the efficiency of target sequence enrichment and the sensitivity to experimental conditions depend strongly on the enrichment strategy employed. The strategy selecting only tester/tester after hybridization can be very efficient to enrich targets. For successful target enrichment, however, the strategy requires a highly efficient subtraction method and proper hybridization conditions. The strategy also requires that the selected DNA be amplified by polymerase chain reaction (PCR) after each or each alternate subtraction. By contrast, the strategy selecting tester/tester plus single-stranded tester is less sensitive to various experimental factors, compared with the strategy selecting only tester/tester. However, it is not as efficient. With this strategy, the tester DNA selected may or may not be amplified by PCR before the next round. In the case of the strategy selecting single-stranded tester, the target DNA can be successfully enriched only when the selected DNA is directly used without PCR amplification in the next round. The strong features of existing methods can be combined to develop a protocol that is more efficient and more reliable.

INTRODUCTION

Subtractive hybridization has been successfully applied to clone mRNA sequences that are more abundant in one mRNA population than in another, or to clone a DNA that is deleted in a mutant genome (1). In subtractive hybridization, the target sequence in one nucleic acid population (‘tester’) is enriched by hybridizing with an excess amount of another population (‘driver’) containing either less or no target sequence, and then by removing the common sequences from the tester. Milner et al. (2) used a kinetic model for subtractive hybridization to simulate the outcome of cDNA or genomic subtraction. They tested the model by comparing its predictions with the published results, and reported that the predictions were generally in good agreement with the published data. However, they also noted that there are discrepancies between the predictions and the experimental results in some cases, particularly when target sequence is of low abundance. The apparent discrepancies might have arisen because the efficiency of the subtractive enrichment procedure was not taken into consideration in the kinetic model.

Removal of common sequence is achieved by various strategies. Commonly used protocols employ biotinylated driver DNA. In this protocol, the tester/driver hybrids are removed using streptavidin or avidin (3–6). The hybrids are also removed by various other means, including hydroxyapatite chromatography (7–9), chemical cross-linking (10) or enzymatic degradation (11). In some cDNA subtractions, cDNA covalently linked to a latex particle is used as driver, and the hybrid is removed by simple centrifugation (12,13). All of the enrichment strategies are designed to remove the tester hybridized with driver, however, the removal is often incomplete for various reasons (1). If the tester/driver hybrid is not completely removed, it can seriously affect the target enrichment. The efficiency of subtractive hybridization is also expected to be influenced by other key factors, such as driver:tester ratio, DNA concentration and salt concentration. The efficiency depends also on the enrichment strategy employed. Although it is evident that the efficiency of subtractive hybridization is varied depending on a variety of factors, it has not been thoroughly examined exactly how much the efficiency of target enrichment is influenced by each factor.

In this study, simulations have been carried out using a model genome system, in an attempt to examine various hybridization conditions and enrichment strategies employed in subtractive enrichment. Simulations were also performed for the genomic subtraction experiments done by Lisitsyn et al. (14), Wieland et al. (7), Straus and Ausubel (5) and Sun et al. (6) to compare the predictions with the published results. In the simulations, the efficiencies of subtraction process and biotinylation reaction were incorporated as input variables to accommodate various experimental situations. Simulation of genomic subtraction has the advantage that more thorough examination can be done since the information on a genome, such as genomic size, abundance class data and the size of target DNA, is in general available.

MATERIALS AND METHODS

Kinetics and computer program

When a DNA sample is composed of several abundance classes, the absolute concentration of each class depends on the fraction of the class among total DNA. Assuming that DNA reassociation follows the ideal second-order kinetics (15), the time when reassociation is half complete, of a particular class i is determined by the equation:
\[ t'_{1/2} = \frac{1}{k \cdot C_0^i} \]

where \( C_0^i \) is the concentration of a particular class \( i \) in nucleotides per liter. The rate constant \( k \) can be calculated by the following equation formulated by Milner et al. (2):

\[ k = Z \cdot 1.18 \times 10^6 M^{-1} s^{-1} \cdot (L/450)^{1/2} \]

where \( Z \) is the enhancement factor that depends on salt concentration (15), \( G_i \) is the complexity of a specific class \( i \) in bp and \( L \) is the length of the shorter of the two reacting complementary strands in bp. The rate constant is proportional to the square root of \( L \) (16).

The fraction of single-stranded DNA (Fss) for a class after hybridization time \( t \) can be calculated by the equation:

\[ F_{ss}^i = \frac{1}{1 + (t/t'_{1/2})} = \frac{t'_{1/2}}{1 + t} \]

The amount of single-stranded tester DNA of class \( i \) (ssT) after time \( t \) can then be calculated by the equation:

\[ ssT^i = (\text{Tester}^i) \times F_{ss}^i \]

where \( \text{Tester}^i \) is the absolute amount of class \( i \) tester DNA.

The fraction of double-stranded DNA is \((1 - Fss)\). The double-stranded tester after hybridization would be in one of two forms: tester/tester (TT) and tester/driver (TD). The amount of TT of class \( i \) is calculated by the equation:

\[ TT^i = (\text{Tester}^i) \times (1 - F_{ss}^i) \times \frac{\text{Tester}^i}{\text{Driver}^i + \text{Tester}^i} \]

where \( \text{Driver}^i \) is the absolute amount of class \( i \) driver DNA. Likewise, the amount of tester DNA of class \( i \) in TD form is calculated by the equation:

\[ TD^i = (\text{Tester}^i) \times (1 - F_{ss}^i) \times \frac{\text{Driver}^i}{\text{Driver}^i + \text{Tester}^i} \]

The model system

The model genome used for simulation is shown in Table 1. It was assumed that there are three abundance classes, since there are typically three reassociating components (slow, intermediate and fast components) in the eukaryote genome (17). Class 1 corresponds to single-copy sequences. Classes 2 and 3 represent repetitive DNAs whose copy numbers are 100 and 5000, respectively. Class 4 represents the target sequence, which is a single-copy sequence of 5 kb and present only in tester. The genomic size of the model system is 2.5 \( \times \) 10⁷ bp. The abundance class data on human and Arabidopsis thaliana genomes are the same as used by Milner et al. (2), except that the snapback/highly repetitive sequences in the human genome are included in the fast reassociating component. It was assumed that the genome of yeast is 2 \( \times \) 10³ bp in size with no repetitive sequences. In the simulation, the DNA concentrations of tester and driver DNA were 0.1 and 10 mg/ml, respectively, unless stated otherwise. The standard \( Z \) value was 7, which is equivalent to 1.0 M Na⁺, and the standard hybridization time was 20 h.

Table 1. Model genome

<table>
<thead>
<tr>
<th>Class</th>
<th>No. of species</th>
<th>% in Driver</th>
<th>% in Tester</th>
<th>% of each species in Tester</th>
<th>Copy no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 000 000</td>
<td>50</td>
<td>50</td>
<td>0.00001</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>30 000</td>
<td>30</td>
<td>30</td>
<td>0.001</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>20</td>
<td>20</td>
<td>0.05</td>
<td>5000</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0.00002</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

aClasses 1, 2 and 3 represent common sequences. The target sequence represented by class 4, which is 5 kb in size, is single-copy and present only in tester.

bAverage length of the DNA is 250 bp.

RESULTS

Strategies in subtractive hybridization

The subtractive hybridization procedure for modeling is depicted in Figure 1. After hybridization, the sample DNA would be in one of five different forms: single-stranded tester (ssT), double-stranded tester/tester (TT), double-stranded tester/driver (TD), single-stranded driver (ssD), and double-stranded driver/driver (DD). Based on the types of the selected tester DNA after subtraction, existing methods can be categorized into three selection strategies. One of the frequently used methods to remove common sequence in tester is that the driver DNA is biotin-labeled and the hybrid tester (TD) is removed using streptavidin (3,4). In this case, both ssT and TT are selected; therefore, this procedure is categorized as ‘ssT+TT selection’ strategy. In many cDNA subtractions, single-stranded tester is hybridized with single-stranded driver. In these protocols, first strand cDNA tester is usually hybridized with excess driver mRNA or cDNA. Alternatively, single-stranded phagemids with directional inserts are used to provide driver and tester (18,19). These methods using single-stranded tester and driver are basically the same as ‘ssT+TT selection’ in that kinetically unhybridized common sequences are not removed.

Lisitsyn et al. (14) developed an elaborate procedure to select and amplify only tester/tester after hybridization. In this ‘TT
The results are shown in Figure 2. In this simulation, the enrichment employing PAAS was performed with the model each subtraction (PAAS). The simulation of the subtractive subtraction, and then used as a new tester in the subsequent round. In this strategy, the tester DNA is amplified by PCR after each subtraction round. The experimental conditions are kept the same as in the initial conditions. The efficiency of TD removal was assumed to be 99%. It was also assumed that 99% of TT and ssT was recovered in each subtraction round. The remaining DNA after a subtraction round is directly used in the next round without PCR amplification.

Figure 1. Subtractive enrichment strategy. Tester DNAs are ligated with linkers at their 5' and 3' ends, and hybridized with excess driver DNAs which have either different or no linkers. After hybridization, depending on the strategy employed, single-stranded tester, tester/tester or single-stranded tester plus testefaster is selected. The tester DNA selected may be amplified by polymerase chain reaction (PCR) before being used in the next round. The selected tester DNA may be amplified by PCR after each or every alternate subtraction, and then used in the subsequent round. Or, the remaining DNA after a subtraction round is directly used in the next round. Depending on the enrichment strategy, the outcome of the subtractive hybridization can be substantially different.

**Enrichment strategy with PCR amplification after each subtraction**

In this strategy, the tester DNA is amplified by PCR after each subtraction, and then used as a new tester in the subsequent round. This enrichment strategy is here termed PCR amplification after each subtraction (PAAS). The simulation of the subtractive enrichment employing PAAS was performed with the model genome. The results are shown in Figure 2. In this simulation, the efficiency of TD removal was assumed to be 99%. It was also assumed that 99% of TT and/or ssT was recovered in each subtraction round. The experimental conditions are kept the same throughout throughout the subtraction rounds in the initial conditions. The target sequence is enriched to >90% of total DNA after round 9 when ssT+TT selection strategy was employed (Fig. 2A). When TT selection strategy was employed, similar enrichment level can be achieved after round 6 (Fig. 2B). In the case of ssT selection, the target sequence is enriched ∼100 times after round 5, then little enrichment is achieved afterwards (Fig. 2C). The maximum enrichment level achieved with the ssT selection strategy is largely determined by driver:tester (D:T) ratio: when D:T ratio is 1000:1, then the target sequence is enriched up to 1000 times (data not shown). The simulation shows that TT selection strategy is far better than the other two selection strategies. However, as will be described shortly, the TT selection strategy is very sensitive to subtraction efficiency. Furthermore, presence of unbiotinylated driver DNA, where applicable, can greatly affect the target enrichment.

Listsyn et al. (14) used an elaborate TT selection protocol employing the PAAS strategy to enrich single-copy target sequences in the human genome. The TT selection involves kinetic enrichment. As discussed by the authors, if target sequences are present at 100 times the amount of non-target sequences after one or two subtractive rounds, then the self-reassociation rate of the targets to form double-stranded tester/tester can be 10 000 times higher than that of non-targets. Exploiting this kinetic difference, target DNA can be greatly enriched by TT selection. They also employed a novel procedure to reduce the complexity of the DNA sample. They used combinations of restriction endonucleases to generate different sets of DNA fragments. Then they generated subgenomes using PCR by taking advantage of the fact that small DNA fragments are preferentially amplified. Using this genomic sampling they succeeded in cloning single-copy target sequences. In the absence of genomic sampling, however, they were unable to isolate targets.

Their experiment using λ or adenovirus DNA as a target was simulated. Unlike the model genomic subtraction described earlier, the D:T ratio was not kept constant in their procedure: the amount of the tester DNA used was gradually decreased as subtraction rounds proceeded. To simulate the situation, the following values were used for the amount of tester DNA: 100 µg/ml (round 1), 15 µg/ml (round 2), 0.04 µg/ml (round 3) and 0.001 µg/ml (round 4). To simulate the genomic sampling process, the number of different DNA fragments in each abundance class of the human genome was reduced to either 1/10 or 1/50 of the original value. The outcome of the simulation seems to closely reflect the actual result obtained in the experiment, when assuming that the efficiency of subtraction is 99.5%: the target sequences are enriched to major difference products after two or three rounds of subtractive hybridization.

The simulation also shows that, if the tester DNA in the form of TD and ssT is completely removed, then the target sequence can be enriched to >90% of total DNA after the fourth round, even without genomic sampling. However, if the subtraction efficiency is decreased just slightly to 99.9%, target DNA enrichment does not occur: the target sequences are enriched by ∼50-fold after three rounds, and little enrichment occurs thereafter. That the enrichment procedure is extremely sensitive to subtraction efficiency explains why the protocol needed genomic sampling for successful target enrichment. The outcome, however, is also dependent on the experimental scheme. If the tester DNA concentration had been kept constant at 100 µg/ml, the target enrichment would have been less sensitive to subtraction efficiency. Under the circumstances, the target sequence can be enriched to >90% even without genomic sampling, as long as the subtraction efficiency is >96%. With lower subtraction efficiency, genomic sampling is a prerequisite for successful target enrichment.
Although they designed a deliberate protocol to select and amplify only TT, there is a possibility that other forms of DNA are amplified too. In their protocol, the tester DNA of TD form and ssT DNA would have an adaptor sequence at only one end after hybridization and fill-in reaction. In subsequent PCR, however, a fraction of the tester of TD and ssT can pick up another ‘missing’ adaptor sequence when annealed with other tester DNA molecules. This might happen during the repeated cycles of denaturation, annealing and extension, especially in the later rounds of PCR where the DNA concentration is rather high. Once they have adaptor sequences at both ends, they can also be amplified by PCR, leading to a decrease in subtraction efficiency. Although only a minute fraction of TD and ssT is expected to be amplified by PCR, a slight decrease in subtraction efficiency can greatly affect target sequence enrichment, as mentioned before.

**Effect of various experimental factors**

The outcome of subtractive hybridization can be very different depending on various experimental factors, as exemplified by the experiment of Lisitsyn *et al.* (14). Simulation was performed to further examine the effects of TD removal efficiency, D:T ratio, DNA concentration and salt concentration on target enrichment. Since biotinylated DNA is frequently used, the impact of unbiotinylated DNA was also investigated. The simulation was carried out using the model genome for the subtractive hybridization employing the PAAS strategy. The ssT selection strategy is not considered here, since the strategy in PAAS is inappropriate.

Figure 3 shows the result of the simulation where percent TD removal efficiency was varied. The efficiencies of ssT removal and TT recovery were all assumed to be 99%. In TT selection strategy, it is shown that the target sequence enrichment is greatly affected by the efficiency of TD removal. It also shows that there is a situation where enrichment is not achieved. In the model genomic subtraction with TT selection strategy, the target sequence is not enriched but decreases as subtraction rounds proceed when the efficiency of TD removal is 94% or less. In the case of ssT+TT selection strategy, the efficiency of target sequence enrichment is not considerably affected by the low TD removal efficiency. The impact of subtraction efficiency also depends on the abundance of target sequence. Suppose that the target sequence of the model genome constitutes 0.002% of total tester (10 times more abundant compared with the model genomic system). Then the target sequence is enriched by the TT selection strategy even if the TD removal is less efficient: the
target sequence is enriched to >90% after seven rounds even when TD removal efficiency is as low as 90% (data not shown). Previously, genomic sampling was shown to be a key to the success of target sequence enrichment in the experiment of Lisitsyn et al. The genomic sampling procedure in effect resulted in increase of the abundance of a particular target sequence in accordance with the reduction of complexity.

Target sequences are enriched by removal of common sequences from tester. In the hybridization, the more driver DNA is used, the more common sequence of tester is trapped in the form of TD. Therefore, it is expected that the target enrichment becomes more effective as the D:T ratio increases. Figure 4 shows that this is generally true in the case of ssT+TT selection. In the case of TT selection, however, there is an optimum D:T ratio. The inefficiency at the very high D:T ratio in TT selection is due to the small amount of target DNA. When the amount of target DNA is limiting, most of the target DNA is in the form of ssT after hybridization. Under this circumstance, only a tiny fraction of target DNA is selected, thereby lowering the enrichment level. At high D:T ratio in the TT selection strategy, furthermore, the enrichment procedure is very sensitive to other experimental factors. For example, at a ratio of 1000:1, the target sequence is not enriched even with 98% of TD removal efficiency (data not shown).

DNA and salt concentrations are also an important factor determining the enrichment efficiency. The simulation shows that the enrichment efficiency drops as the DNA concentration is decreased. It also shows that the TT selection strategy is more sensitive to the DNA or salt concentration compared with the ssT+TT selection strategy, and there is a critical DNA or salt concentration below which no enrichment occurs. For example, at 1 mg/ml of driver DNA with 100:1 of D:T ratio, the target sequence is not enriched but decreases when employing TT selection strategy. In the case of ssT+TT selection strategy, the target sequence is enriched to >90% at the concentration of 1 mg/ml, though the enrichment is very inefficient (data not shown). The hybridization rate can be accelerated by phenol emulsion reassociation technique (20,21) or in the presence of dextran sulfate (3). If dextran sulfate is included, the target sequence can be successfully enriched by TT selection strategy even at a DNA concentration as low as 1 mg/ml. In the presence of 11% dextran sulfate and 1.5 M NaCl, where Z is 91 (3), the enrichment efficiency is dramatically increased such that the target sequence is enriched to >90% after five rounds (data not shown).

In many subtractive hybridization experiments, biotinylated DNA is used as a driver. However, the biotinylation is rarely complete. Barr and Emanuel (3) reported that, when samples of 3H-labeled DNA were biotinylated by two cycles of photobiotinylation, ~1% of the 3H label was not removed, even by three cycles of streptavidin binding and organic extraction. This might indicate that 1% of the DNA is unbiotinylated. However, the fraction for the unbiotinylated DNA might have been underestimated, considering that the double-stranded DNA with only one strand biotinylated can be captured by streptavidin. Thus, 1% of unbiotinylated DNA at the double-strand level can be as much as 10% at single-strand level. Under this circumstance, as much as 10% of tester/driver hybrids would not be removed by streptavidin or avidin. Under the standard conditions with TT selection strategy, the presence of unbiotinylated driver DNA can significantly reduce the efficiency of target sequence enrichment, as shown in Figure 5. When the unbiotinylated DNA is 5%, then no enrichment is achieved. The impact of unbiotinylated DNA is greater at high D:T ratio. For example, at 1000:1, the targets are not enriched even with 0.5% unbiotinylated DNA. It must be stressed that the percent values here represent the unbiotinylated single-stranded DNA. In the case of ssT+TT selection strategy, the impact of unbiotinylated DNA is not so serious.

**Enrichment strategy employing PCR amplification after each alternate subtraction**

The tester DNA can be amplified after each alternate subtraction. Although this strategy has been employed in cDNA subtractions (11,22), it can also be applied to genomic subtraction. This strategy is termed as PCR amplification after consecutive subtraction (PAACS). Figure 6 shows the result of the simulation of the subtractive enrichment employing this strategy. Under standard conditions of subtractive hybridization with PCR amplification of tester DNA after every two cycles of subtraction, the target sequences are enriched to >90% after five rounds (10 subtraction cycles) when ssT+TT selection strategy is employed (Fig. 6A). This result is similar to that obtained with PAAS. The effects of TD removal efficiency, D:T ratio, DNA and

---

**References**

Figure 5. Effect of unbiotinylated DNA on target enrichment. The subtractive hybridization employing PAAS was simulated with different levels of unbiotinylated DNA, using either ssT+TT or TT selection strategy. The unbiotinylated DNA represents the unbiotinylated single-stranded driver DNA. The efficiency of the subtractive hybridization is shown as the number of rounds to achieve target sequence enrichment to >90%. The input parameters for experimental conditions and assumptions, except the unbiotinylated DNA, are the same as described in Figure 2.

salt concentrations, and unbiotinylated driver are also similar as in the case of PAAS. One major difference could be that the PAACS strategy is relatively insensitive to the D:T ratio. Similar levels of enrichment (>90%) can be obtained after five rounds at the D:T ratios ranging from 10:1 to 10 000:1 (data not shown).

In PAACS strategy, target sequence can also be enriched by TT selection, as shown in Figure 6B. In the TT selection strategy, an optimum D:T ratio exists as in the case of PAAS. However, the dependency is even greater. Furthermore, this strategy is highly sensitive to TD removal efficiency and presence of unbiotinylated driver (data not shown). Previously, the TT selection strategy in PAAS was shown to be more sensitive to TD removal efficiency and presence of unbiotinylated DNA at higher D:T ratios. In PAACS, the D:T ratio becomes higher in the second subtraction cycle, since a significant amount of tester DNA is removed in the first cycle. Thus, it is not surprising that the PAACS procedure with TT selection is more sensitive to TD removal efficiency and presence of unbiotinylated driver, compared with the PAAS procedure. It is then predicted that the TT selection in the PAACS strategy would be more efficient and more reliable at higher tester concentration. Indeed, when the tester concentration is increased to 1 mg/ml (10 times higher than under standard conditions), the target DNA is more efficiently enriched: only three rounds are needed to enrich the targets to >90%. Moreover, at this higher tester DNA concentration, the enrichment procedure is less sensitive to the subtraction efficiency and presence of unbiotinylated driver (data not shown).

In the ssT selection strategy, the target sequences are enriched to 0.074% (~370-fold increase) after the third round, and little enrichment is achieved afterwards. This result is similar to that obtained with PAAS, although the efficiency of target sequence enrichment is a little higher. Again, the outcome depends strongly on the D:T ratio.

Enrichment strategy employing consecutive subtraction without PCR amplification

In a strategy where subtractions are performed without intermittent PCR amplification, the remaining DNA after a subtraction round is directly used in the next round. This strategy is known as consecutive subtraction (CS). In this case, PCR amplification of tester DNA is performed only after a final subtraction round. In genomic subtraction, this CS strategy was employed by Wieland et al. (7), Straus and Ausubel (5) and Sun et al. (6).

In PAAS, the initial D:T ratio can be maintained throughout the subtractive hybridization rounds. In CS strategy, the D:T ratio inevitably changes. Simulation was performed using the model genome for subtractive hybridization employing the CS strategy. The results are shown in Figure 7. The simulation was carried out with the assumption that the subtraction efficiency was 99% as in the case of PAAS. When ssT+TT selection strategy is employed, the target DNA can be enriched to >90% after nine rounds of subtraction (see Fig. 7A), which is comparable with the result obtained by the PAAS strategy shown in Figure 2. Different results were obtained in the case of TT or ssT selection strategy. In TT selection strategy, the target sequences are not enriched. This is because the target DNA is in low concentration and only a small fraction of the target DNA can self-anneal to form TT. In subsequent rounds, the target DNA will have less and less chance to form TT, consequently it fades away as subtraction rounds proceed. The results with ssT selection strategy also show a sharp
contrast to the results obtained with PAAS strategy. The target DNA can be enriched to >90% after eight rounds of subtraction, as shown in Figure 7 B.

In CS strategy, the unbiotinylated driver, if any, is added in each subtraction cycle. Should the unbiotinylated driver accumulate, this will lead to a decrease in enrichment efficiency. Simulation shows, however, that the additive accumulation of unbiotinylated DNA does not occur. Suppose that 10% of driver is unbiotinylated at the single-stranded DNA level. The situation was simulated for the model genome with the assumption of complete removal of biotinylated DNA, that is, 100% subtraction efficiency and 100% recovery. The result shows that, after the first cycle of subtractive hybridization and addition of fresh driver, the fraction of unbiotinylated DNA increases from 10% to 12.2% (class 1), 11.0% (class 2) and 11.0% (class 3). After the second cycle, the fraction increases to 12.8% (class 1), 11.1% (class 2) and 11.1% (class 3). However, the fraction of unbiotinylated DNA does not increase very much afterwards, that is, the system has reached a steady state where input and output of unbiotinylated DNA are nearly equal. Similar results were obtained when it was assumed that biotinylated driver DNA was not completely removed (data not shown). Nevertheless, the increased proportion of unbiotinylated DNA reduces the efficiency of target enrichment.

That 10% of driver is unbiotinylated does not mean that 10% of the driver accumulates in each subtraction round. If the unbiotinylated DNAs were in single-stranded form, they would not be removed by streptavidin. If the unbiotinylated DNA forms driver/driver double-stranded DNA with biotinylated DNA, however, streptavidin can capture the DNA. Under this circumstance, only 1% of total DD is unbiotinylated on both strands. Therefore, only a fraction of unbiotinylated driver DNA remains in the subtracted DNA. Even for the unremoved unbiotinylated DNA, there is a chance of forming 'biotinylated' double-stranded DNA in the next round of hybridization.

Wieland et al. (7) employed a ssT selection strategy using biotinylated tester DNA to enrich target DNA in human genome. In the experiment, the double-stranded DNA including TD was removed by hydroxyapatite chromatography. After three rounds of subtraction, they obtained between 100- and 700-fold enrichment of target sequences. Assuming 99% of subtraction and recovery efficiencies, the simulation of the experiment using bacteriophage λDNA as a target shows that the target DNA is enriched by ~400-fold after three rounds, generally consistent with the published result. The simulation also shows that the target DNA can be enriched to >90% if seven rounds of subtraction are performed. In this case, the presence of unbiotinylated DNA has negligible effect, since the TD removal procedure using hydroxyapatite chromatography has nothing to do with biotinylated DNA. The only effect of unbiotinylated tester is that the yield of the target DNA after the final selection step using avidin can be lower. However, the low recovery would not be a big problem, considering that PCR is performed after the avidin selection. Thus, the enrichment efficiency is mainly determined by the efficiency of the method separating single-stranded and double-stranded DNA. If the efficiency of the subtraction using hydroxyapatite is 90%, nine subtraction rounds are required to achieve target DNA enrichment to >90%. If the subtraction efficiency drops to 80%, 11 rounds are needed (data not shown). Therefore, it is important to remove as much double-stranded DNA as possible, even if some single-stranded DNA is lost during the process.

Straus and Ausubel (5) employed a CS strategy to clone a DNA corresponding to a deletion mutation in yeast. After repeated rounds of subtraction using biotinylated driver (ssT+TT selection), only the tester/tester DNA was selected in a final step. When their experiment was simulated, the target DNA was shown to be enriched to 46% and 99% after the second and the third rounds, respectively, assuming that the overall efficiency of the procedure is 99%, i.e., 99% efficiency of subtraction, recovery and biotinilation. The prediction based on the assumption of 99% efficiency is not consistent with the experimental results, indicating some inefficiency in the subtractive enrichment procedure. When assuming that TD removal efficiency is 90% and that unbiotinylated single-stranded DNA was 10% of total driver DNA, the target sequences would be enriched to 8.6% after the second and 70.8% after the third round, which might reflect the experimental results. It should be mentioned here that the target DNA was significantly enriched by the final TT selection.

Figure 7. Simulation of the subtractive hybridization without PCR amplification after subtraction. In this simulation, the remaining DNA from a subtraction round is directly used in the next round of subtractive hybridization without PCR amplification. Thus, the driver:tester ratio increases as subtraction rounds proceed. The other experimental conditions and assumptions are the same as described in Figure 2. The outcomes of simulation employing ssT+TT (A) and ssT (B) selection strategies are represented.
procedure. Without the TT selection, the target sequences constitute only 0.5% and 2.5%, respectively, after the second and third rounds.

Sun et al. (6) extended the method of Straus and Ausubel to clone a deleted DNA in Arabidopsis. In the subtractive cloning of a 5 kb target in Arabidopsis, they found that the target sequences were enriched to only ~5% after five cycles. Moreover, little enrichment occurred after two additional subtraction cycles. When the experiment was simulated with 99% subtraction efficiency, the target sequences were found to be enriched to >90% after three cycles. The discrepancy can be explained when assuming that the overall efficiency is far lower than 99%. However, even in this case, the target fraction would increase considerably by the two additional subtraction cycles. This rather perplexing result can be partly explained by contaminant in tester, whose presence was examined and confirmed by the authors. The situation was simulated with the assumption that contaminating DNA of 5 \times 10^6 bp in complexity constitutes 0.5% of tester DNA. Under this circumstance, the target DNA was shown to be enriched to 4.7% with the contaminating DNA being enriched to 95.3% after five cycles, if the subtraction efficiency was 90% and if 10% of single-stranded driver DNA were unbiotinylated. In this case, the target DNA fraction does not change significantly after two additional subtraction cycles.

**Combination of selection strategies**

The outcome of subtractive hybridization is greatly influenced by the genomic size or the abundance of target sequence. Reduction of complexity, hence increase of target sequence abundance, can be accomplished by genomic sampling, as manifested by Lisitsyn et al. (14). Once the abundance of target sequence is increased, the enrichment procedure becomes more efficient and less sensitive to subtraction efficiency and presence of unbiotinylated DNA. Another strategy to make the enrichment procedure more efficient and more reliable might be employing a combination of selection strategies, as used by Straus and Ausubel (5) and Sun et al. (6). Figure 8 shows the enrichment achieved either by ssT+TT or by TT selection strategy at each subtraction round in the model genomic subtraction employing PAAS, with 99% subtraction efficiency and 1% unbiotinylated driver DNA. It shows that ssT+TT selection is more efficient at the first two subtraction rounds. From the third round, however, TT selection is far better than the ssT+TT selection. This is because the target DNA is sufficiently enriched after two rounds of subtraction and PCR, so that kinetic enrichment can be greatly enhanced.

Considering this data, it would be better to perform initial subtractions by ssT+TT selection strategy, and then perform subtractions by TT selection strategy. Assuming that the subtraction efficiency is 99%, and unbiotinylated driver is 5% (at single-stranded DNA level), the target sequence in the model genome is enriched to >90% after 10 rounds by ssT+TT selection alone in PAAS. When TT selection strategy is employed, the target DNA is not enriched. However, if ssT+TT selection strategy is employed in the first two rounds of subtraction, the enrichment of target DNA to >90% can be achieved by five rounds of TT selection (seven rounds in total).

The same strategy can be applied to PAACS, resulting in enhancement of target sequence enrichment. In PAACS with PCR amplification after every two cycles of subtraction, the target sequence in the model genome is enriched to >90% after five rounds (10 subtraction cycles) of ssT+TT selection, assuming that the subtraction efficiency is 99%, and unbiotinylated single-stranded DNA is 5%. TT selection gives no enrichment. When TT selection strategy is employed after one round of ssT+TT selection and PCR amplification, however, the target sequence is enriched to >90% after three rounds (four rounds in total). The strategy can also be applied to CS protocols. However, in this case, the strategy may result in initial increase and then decrease of target DNA fraction. For example, when the subtraction is performed with TT selection after two rounds of ssT+TT selection, then the target sequence is enriched to 35% after five cycles of TT selection, and decreases thereafter.

**DISCUSSION**

Successful genomic subtraction requires effective removal of tester/driver hybrids and, where applicable, efficient biotinylation. DNA and salt concentrations must also be high enough to ensure good hybridization. However, precautions should be taken since hybridization may be hampered at high salt concentration due to high Tm (3), and hybridization rate may be reduced by increased viscosity at high DNA concentration (15).

Successful target sequence enrichment also depends on the enrichment strategy employed. The simulation showed that the ssT+TT selection strategy can be widely used in the experiments employing PAAS, PAACS and CS strategies and that it is relatively less sensitive to various experimental factors, though it is not as efficient as the TT selection strategy. The TT selection strategy can be highly efficient, however, it is very sensitive to various experimental conditions. With TT selection strategy, only a fraction of target DNA would be saved. In contrast, almost all of the target DNA can be saved in the ssT+TT selection protocol. Therefore, with the same subtraction efficiency (same amount of unremoved TD), the proportion of the non-target DNA in the subtracted DNA is far greater when only TT is selected. This explains why TT selection strategy is so sensitive to subtraction efficiency. In TT selection, the efficiency of ssT removal is also an important factor. Previously, it was shown that the target enrichment does not occur with 94% TD removal efficiency in the model genomic subtraction employing TT selection (Fig. 3). In
the simulation, the efficiency of ssT removal was assumed to be 99%. Interestingly, the target can be enriched to >90% after nine rounds if the subtraction efficiencies for TD and ssT are all 94%. This shows that an effort to remove ssT as much as possible without ensuring high TD removal efficiency can have a detrimental effect on target enrichment.

The protocol employing ssT+TT selection strategy is subtractive in nature, whereas the protocol employing TT selection strategy has a kinetic enrichment component as well. Incorporating kinetic enrichment in subtractive hybridization can greatly accelerate target enrichment, though it is difficult. To maximize kinetic enrichment component, the tester DNA concentration must be deliberately lowered to an appropriate level, as in the experiment of Lisitsyn et al. (14). However, the consequent high D:T ratio can reduce overall efficiency of target DNA enrichment (Fig. 4), as far as the subtraction efficiency is not 100%. The risk of high D:T ratio is also manifested in the TT selection protocol employing CS strategy, where target sequence is not enriched at all.

The TT selection strategy employed by Lisitsyn et al. is elaborately designed to select TT with high efficiency. The only disadvantage would be that the procedure is very complicated. An alternative method for TT selection could be the EDS method developed by Zeng et al. (11). In this protocol, however, removal of unhybridized single-stranded DNA may not be efficient due to the intramolecular or intermolecular annealing between linkers. To remove single-stranded DNA effectively, excess of a primer which has a sequence complementary to the linker at the 5′ end of tester DNA can be added just after hybridization. Alternatively, S1 nuclease or mung bean nuclease treatment may be included to ensure that single-stranded DNA is removed to completion.

However, the very fact that single-stranded tester DNA is not efficiently removed without appropriate measures can be exploited to develop a more efficient protocol. If necessary, the DNA mixture after hybridization may be slowly cooled to room temperature so that the linkers of single-stranded testers can be annealed with each other. This protocol may be applied in the early rounds of subtractive hybridization to save as much target DNA as possible. In the later rounds, a strict protocol selecting only tester/tester DNA can be used to enhance target sequence enrichment. As mentioned before, this strategy can greatly improve the efficiency and reliability of a subtractive enrichment procedure.

TT selection strategy is effective in eliminating contaminating sequences, provided that the frequency of the contaminating sequence is lower than that of the target sequence. As discussed in the case of the experiment of Sun et al., the contaminating sequence present in tester can limit the target sequence enrichment.

If the tester DNA sample of the model genome is contaminated with bacteria of 5 × 10^6 bp in complexity in the proportion of one microorganism per five sample cells, the 5 kb target and the contaminating DNA constitute 0.0002% and 0.02% of tester, respectively. In ssT+TT selection procedure, both the target and the contaminating sequence are equally saved and amplified. Thus, the 1:100 ratio is maintained throughout the enrichment rounds and the target DNA will never be enriched to >1%. When TT selection strategy is employed, however, the target sequence is successfully enriched to >90% without any noticeable effect from the contaminating sequence. This is because self-reassociation is strongly in favor of the target DNA under the circumstance that the frequency of target sequence is 10 times that of the contaminating sequence.

If tester is contaminated, the above-mentioned strategy using a combination of ssT+TT and TT selection strategies can pose a problem, since the initial ssT+TT selection can also enrich the contaminating sequence to a point where target enrichment may no longer be favored in the TT selection strategy. Yet, it is still possible to eliminate the contaminating sequences after a subtraction experiment, provided that the contaminating organism(s) are identified: another subtraction experiment using the DNA of the contaminating organism (or closely related organism) as driver can eliminate the contaminating DNA from the enriched DNA. If the frequency of contaminating sequence is higher than that of target sequence, the target sequence diminishes as subtraction proceeds since enrichment of the contaminating sequence is more favored. Therefore, it is important that the sample used for preparation of tester be germ-free. One way to alleviate the problem with contaminating sequences could be for the DNA of potential pathogens or common inhabitants to be added to driver DNA in advance.

ACKNOWLEDGEMENTS

We would like to thank Dr Nam-Jeong Cho for his helpful comments. This work was partly supported by grants from the Korea Science and Engineering Foundation.

REFERENCES