Greedy Algorithms for Finding a Small Set of Primers Satisfying Cover and Length Resolution Conditions in PCR Experiments

Koichiro Doi **Hiroshi Imai**
doi@is.s.u-tokyo.ac.jp imai@is.s.u-tokyo.ac.jp
Department of Information Science, Faculty of Science, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Abstract

Selecting a good collection of primers is very important for polymerase chain reaction (PCR) experiments. Most existing algorithms for primer selection are concerned with computing a primer pair for each DNA sequence. In generalizing the arbitrarily primed PCR, etc., to the case that all DNA sequences of target objects are already known, like about 6000 ORFs of yeast, we may design a small set of primers so that all the targets are PCR amplified and resolved electrophoretically in a series of experiments. This is quite useful because deceasing the number of primers greatly reduces the cost of experiments. Pearson et al. [7, 8] consider finding a minimum set of primers covering all given DNA sequences, but their method does not meet necessary biological conditions such as primer amplification and electrophoresis resolution.

In this paper, based on the modeling and computational complexity analysis by Doi [2], we propose algorithms for this primer selection problem. These algorithms do not necessarily minimize the number of primers, but, since basic versions of these problems are shown to be computationally intractable, especially even for approximability with the length resolution condition, this is inevitable. In the algorithms, the amplification condition by a primer pair and the length resolution condition by electrophoresis are incorporated. These algorithms are based on the theoretically well-founded greedy algorithm for the set cover in computer science. Preliminary computational results are presented to show the validity of this approach. The number of computed primers is much less than a half of the number of targets, and hence is less than one forth of the number needed in the multiplex PCR.

1 Introduction

The polymerase chain reaction (PCR) amplifies DNA sequences efficiently and this experiment has been used for various purposes. Its use will be further enhanced as many DNA sequence data become available.

Among many applications of the PCR experiment, genome typing, DNA fingerprints, etc., require PCR amplifications of many different target objects, and producing these amplifications in a series of experiments by grouping them can save experimental costs greatly. This saving can be achieved by designing a good collection of primers for PCR experiments.

Designing such a good primer set is a challenging problem, and this paper proposes greedy algorithms for finding a small collection of primers satisfying cover and length resolution conditions in PCR experiments. Some promising results of preliminary computational experiments are given.

In the field of genome informatics, this primer selection problem might be rather new, and we need more explanations to state our results in more detail. Hence, from here in this introduction, we try to explain the backgrounds and requirements of this problem, in the following subsections, and then state our results.

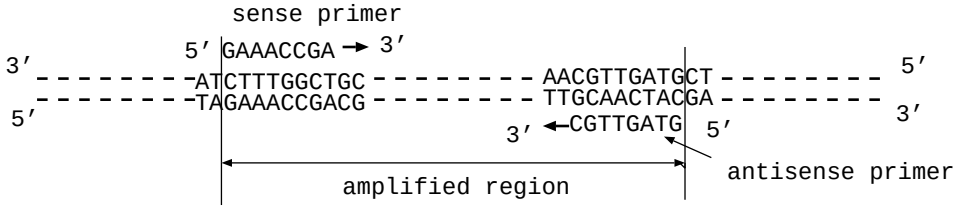

Figure 1: DNA sequences and primers

1.1 PCR

PCR amplifies the part of the DNA sequence between these primers. One primer is called a sense primer, and the other primer is called an antisense primer. Primers have an orientation, as depicted in Figure 1.

PCR is usually used to amplify a particular region in a target sequence. Therefore a primer pair that matches one position of a target sequence is selected. If selected primer matches many positions of DNA sequences, many subsequences may be amplified. The length of primers is usually set to be > 17 mer for human genome, because the random permutation of 17 mer ($4^{17} \approx 1.6 \times 10^{10}$) will match less than once for total human genome (3×10^9 bp).

1.2 PCR with Many Primers

As stated in the beginning of this introduction, it is often necessary to produce PCR amplifications of many different target objects. There have been proposed many methods for this.

The multiplex PCR amplifies multiple genes with multiple sets of primers in a single reaction (Chamberlain, Gibbs, Ranier, Nguyen and Caskey [1]), where a pair of primers is designed for each amplified region in this PCR and all the primers are distinct to one another. Hence, in this case, the length of primers is set 17 to 20. The number of primers is exactly twice as many as the number of target objects, and is fixed.

On the other hand, Welsh and McClelland [10] proposed arbitrarily primed PCR (AP-PCR, in short). Short (8 to 12 nucleotides) arbitrary primers are used in this PCR. These short primers match at several places and produces more fragments, while the number of necessary primers to amplify each target may be much smaller than in the multiplex PCR.

In both cases, PCR products are visualized by agarose gel electrophoresis and these are utilized for DNA fingerprinting, etc. For example, Ito and Sakaki [3] proposed Fluorescent Differential Display (FDD) by using arbitrarily primed RT-PCR. The number of applications is really enormous.

1.3 Primer Selection: Necessity of Reducing the Number of Primers

The following primer selection problem was communicated with T. Ito (see the acknowledgment).

If no prior sequence information is required, AP-PCR may only be applicable. But now, all sequences are known for yeast, *E. coli*, etc., and AP-PCR may be extended further. If arbitrarily primers for AP-PCR are replaced with primers obtained from the known sequence data, PCR products can be computed, which can be compared with PCR experimental results to identify missing objects, etc.

Thus, the primer selection design is getting more important, especially the design for multiple target objects when iterations of amplifications by primers exist. If the minimum set of primers to amplify all sequences in a distinguishable way is obtained, the costs for experiments are decreased. For example, yeast has about 6000 ORFs. This number is too large to obtain distinguishable amplified regions for all ORFs whose lengths are different from one another in a single PCR experiment under

the current technology of electrophoresis, but a series of multiple experiments can resolve this difficulty. Minimizing the primers in this series of experiments directly affect the experiment cost.

This paper basically considers a case that many ORFs are amplified by applying multiple primers simultaneously in a single experiment. This is the most fundamental case, and can be extended to the case of series of experiments, whose details will be described elsewhere.

1.4 Related Works on Primer Selection

Some computer programs for primer selection are designed in [5, 9], etc. These programs are used for designing a pair of primers for a single target gene. These find pairs obeying given biological constraints and displays information about the amplified product.

Nicodème and Steyaert [6] designed a program for the multiplex PCR. As mentioned above, in the multiplex PCR, the number of primers is fixed to the double of the number of targets, and hence minimizing the number of primers is not considered. This program designs the sets of primers which obtain amplified regions for all genes in multiple experiments so as to minimize the number of experiments, which is a main part of the proposed algorithm. The lengths of amplified regions in a experiments are different each other in each experiment in the series.

Pearson et al. [7, 8] formulated a minimum primer set which covers all DNA sequences, and analyzed it, which was a much simpler version than in our paper. Their paper does not consider orientations of primers and length constraint for electrophoresis analysis. From the viewpoint of computational complexity, they analyze these formulated problem, prove approximability of the problem using directly the set cover reduction, and construct an exact branch and bound algorithm and an approximation algorithm with $\ln n$ approximation ratio, where n is a number of DNA sequences.

1.5 Our Results

We have formulated minimizing the number of primers for PCR and proved complexity properties in Doi [2]. The formulations are much more rigorous than Pearson et al. [7, 8]. Doi [2] mainly shows complexity properties, and mentions potential applicability of algorithms for the set cover problem in computer science in a different manner with [7, 8].

In this paper, based on these modeling and computational complexity analysis in [2], we propose greedy algorithms for this primer selection problem. These algorithms do not necessarily minimize the number of primers, but, since basic versions of these problems are shown to be computationally intractable, and in fact, the problem becomes much harder with the length resolution constraint, which is shown by recent approximability results [2], this would be inevitable.

In the algorithms, the amplification condition by a primer pair and the length resolution condition by electrophoresis are incorporated. These algorithms are based on the theoretically well-founded greedy algorithm for the set cover in computer science. Preliminary computational results are presented, which are quite promising. The number of computed primers is much less than a half of the number of targets, and hence is less than one forth of the number in the multiplex PCR.

The paper proceeds as follows. In Section 2, biological constraints which are considered this paper and problem formulation and properties for these problems in are summarized. The approximate property of the simple greedy algorithm was proved in [2]. In Section 3, we construct a simple greedy algorithm for primer selection as a basis, and modify this algorithm for various biological constraints. Experimental results for these algorithms are described in Section 4.

2 Biological Constraints and Problem Formulation

2.1 Biological Constraints for PCR Primer Selection

In this paper, the following biological constraints on primers are considered.

GC content primers with 40-60% GC content are widely used

Lengths of primers In this paper, a primer can match multiple positions in DNA sequences. We can set short length (8-12) for primers

Complementarity sequence The set of primers must not contain self-complementary sequence and complementary sequences. (For example, 5'-GCCTAGGC-3' is a self-complementary sequence, 5'-GACAATGC-3' and 5'-GCATTGTC-3' are complementary sequences.)

Length of amplified regions PCR products which are between 50 and 500 base pairs are desired. In this paper, the lower bounds for amplified regions are considered.

Difference of lengths of amplified segments PCR products should have different lengths with each other to analyze products by electrophoresis. (In practical, it is preferable that the difference of the lengths of the PCR products is 5.)

Details of the biological constraints for designing primer are described in [4]. We will consider the other biological constraints in a full version.

2.2 Formulations and Properties

For various biological constrains, Doi [2] formulated some problems and proved these complexity properties. In this subsection, these results are summarized.

We are given n DNA sequences ($1 \leq j \leq n$). The length of DNA sequence j is denoted by m_j. (j,p) means position p of DNA sequence j. A primer S_i is a DNA sequence of constant length (much shorter than m_j). $|S_i|$ denotes the length of primer S_i. Consider a primer set $S = \{S_1, ..., S_i, ..., S_l\}$. $(j,p,\mathrm{r/l}) \in S_i$ means primer i match position p of DNA sequence j (direction is r(right) or l(left)), that is, for $k = 1$ to $|S_i|$ the $(p+k-1)$th character (resp. the $(p-k+1)$th character) of the sequence j is identical to the kth character of the primer i in the case of r (resp. l).

Definition 1 (Primer Selection Problem (PSP)) *Primer Selection Problem (PSP) is finding minimum primer subset $S' \subseteq S$ which covers all DNA sequences and putting DNA subsequence between sense primer and antisense primer for all DNA sequences, that is,*

$$\forall j\ \exists S_i, S_{i'} \in S',\ (j, p_1, \mathrm{r}) \in S_i, (j, p_2, \mathrm{l}) \in S_{i'},\ p_2 - p_1 + 1 > 0.$$

(The length of amplified segment is $p_2 - p_1 + 1$.)

Theorem 1 ([2]) *PSP has a polynomial approximation greedy algorithm with a factor* $\ln mn$ *($m = \max(m_1, ..., m_n)$).*

Theorem 2 ([2]) *PSP cannot be approximated in polynomial time within a factor* $(1-\epsilon)\ln n$ *for any* $\epsilon > 0$ *unless* $NP \subset TIME(n^{O(\log\log n)})$.

To consider a length of DNA amplification, we must handle the next constraint.

Definition 2 (Amplification Constraint between sense and antisense primers) *If a DNA sequence is amplified by primers (j, p_1, r) and (j, p_2, l) in PCR experiments, primers which match with position p ($p_1 \leq p \leq p_2$) of DNA sequence j cannot exist.*

In Figure 2, the part between S_1 and S_2 is amplified and part between S_1 and S_3 is not amplified (in practice the latter part is slightly amplified in PCR experiments. But the majority is dominated by the former and hence the latter is neglected.)

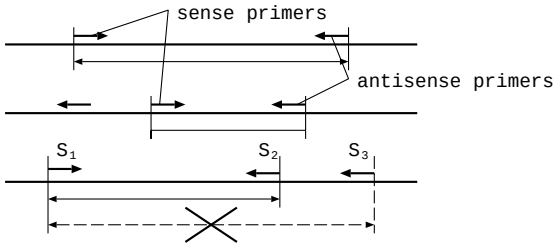

Figure 2: Amplification Constraint

Definition 3 (Primer Selection Problem with Minimum Length Constraint (PSPMLC)) *Primer Selection Problem with Minimum Length Constraint (PSPMLC) is, for a given k, the finding a minimum primer subset $S' \subseteq S$ which covers all DNA sequences, as in PSP, and putting DNA subsequence at least length k between sense and antisense primer for all DNA sequences. ($p_2 - p_1 + 1 \geq k$ instead of $p_2 - p_1 + 1 > 0$ in PSP)*

This problem, PSPMLC, models a fact that amplified parts by a pair of primers S_i and $S_{i'}$ have length at least 50.

If we ignore the amplification constraint, the next theorem is obtained.

Theorem 3 ([2]) *PSPMLC has a polynomial approximation greedy algorithm with a factor* $\ln mn$ *($m = \max(m_1, ..., m_n)$).*

Theorem 4 ([2]) *PSPMLC cannot be approximated in polynomial time within a factor* $(1 - \epsilon) \ln n$ *for any* $\epsilon > 0$ *unless* $NP \subset TIME(n^{O(\log \log n)})$.

Definition 4 (Primer Selection Problem with distinguishable Length Constraint (PSPLC)) *The Primer Selection Problem with distinguishable Length Constraint (PSPLC) is a primer selection problem and the lengths of DNA subsequences between sense and antisense primers are all different, that is,*

1. $\forall j\ \exists S_i, S_{i'} \in S',\ \exists (j, p_1, \mathrm{r}) \in S_i, (j, p_2, \mathrm{l}) \in S_{i'},\ p_2 - p_1 + 1 > 0$,

2. $\forall p'\ (p_1 \leq p' \leq p_2)\ (j, p', \mathrm{r}), (j, p', \mathrm{l})$ *are not in sets which are elements of* S'. *(amplification constraint),*

3. $\forall j' (\neq j)(j', p_3, \mathrm{r}), (j', p_4, \mathrm{l})$ *which satisfy the amplification constraint (2) and* $p_4 - p_3 + 1 \neq p_2 - p_1 + 1$

This problem, PSPLC, models a requirement that may be DNA sequences distinguished with one another by the lengths of their amplified segments using the electrophoresis. If one amplified segment of different length exists for each DNA sequence, DNA sequences can be identified using the electrophoresis.

Theorem 5 ([2]) *PSPLC cannot be approximated in polynomial time within a factor* $n^{\frac{1}{2}-\epsilon}$ *for any* $\epsilon > 0$ *unless* $NP = coRP$.

Theorem 6 ([2]) *PSPLC cannot be approximated in polynomial time within a factor* $n^{\frac{1}{8}-\epsilon}$ *for any* $\epsilon > 0$ *unless* $P = NP$.

3 Algorithms

In previous section, variations of primer selection problems are showed. We solve these problems by some greedy algorithms which are used for proof of Theorems 1, 3. Besides, we modify these algorithm to satisfy constraints.

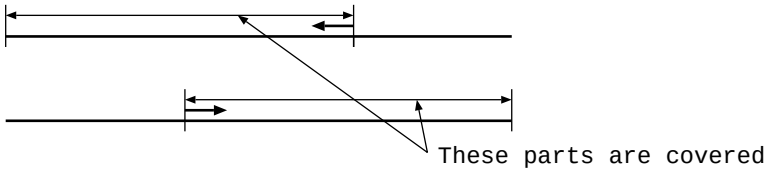

Figure 3: Covered elements for a primer

Basic idea for algorithms is that the primer selection problem is regarded as the set cover problem. The position p of the DNA sequence j (j,p) is regarded as the element of underground set. Primers are regarded as subsets. Figure 3 shows covered elements for a primer.

To be exact, (j,p) $(1 \leq j \leq n, 1 \leq p \leq m_j + 1)$ are elements in the underground set for the set cover problem. (We add $(j,p+1)$ for all DNA sequence j.) If $(j,p,\mathrm{r}) \in S_i$, primer S_i covers $(j,p+1),(j,p+2),...,(j,m_j),(j,m_j+1)$. If $(j,p,\mathrm{l}) \in S_i$, primer S_i covers $(j,1),(j,2),...,(j,p-1),(j,p)$.

All (j,p) $(1 \leq j \leq n, 1 \leq p \leq m_j + 1)$ are covered if and only if amplified sequences exist in all DNA sequences.

We apply the greedy algorithm to this problem.

1. Scan DNA sequences and pick up primers which satisfy biological conditions. Make the list of candidates of primers.

2. Select the primer S_i that covers the largest number of uncovered elements. Delete S_i from the candidate set. Add S_i to the solution.

3. Repeat 2 until all elements are covered.

We call this algorithm *basic algorithm*. Algorithms in this paper are extension of this greedy algorithm.

Algorithms consider the constraints on GC content and lengths of primers only in step 1 where candidates of primers are obtained. Modification to the *basic algorithm* is necessary for the other constraints. Basic idea of modification is constructing prohibitive conditions in step 2 (except the case in Section 3.2.1). Candidates of primers which satisfy prohibitive conditions can not be selected. Hence, the step 2 is replaced with the following $2'$

$2'$. Select the primer S_i that covers the largest number of uncovered elements does not satisfy prohibit conditions. Add S_i to solution.

As prohibitive conditions are added, solutions which cover all DNA sequences may not be obtained. If candidates of primers do not exist, algorithms must end, and we use the following $3'$ instead of 3.

$3'$. Repeat $2'$ until all elements are covered or candidates do not exist.

We consider various biological constraint and modify the *basic algorithm* in the following subsection.

3.1 Complementary Sequence Property

Self complementary sequences are not appropriate candidates for primers. We remove the sequences from candidates of primers in the step 1.

We modify the *basic algorithm* to that the sequence whose complementary sequence was selected can not be selected. For example, if 5'-GACAATGC-3' was selected for primer, 5'-GCATTGTC-3' can not be selected.

This is first prohibitive condition for the *basic algorithm*. We call this condition the *complementary condition*.

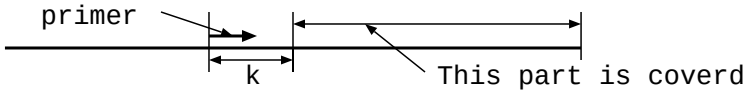

Figure 4: Modifying covered elements for a primer

3.2 Minimum Length Constraint

We consider two types of modifications for minimum length constraint.

3.2.1 Without Amplification Constraint

If the amplified region has not length enough, this is useless. We must modify the algorithm to obtain amplified regions whose lengths are long enough. (Given k for minimum length of amplified segment.)

First we explain the algorithm for the proof of Theorem 3. This algorithm is that covering elements for primers are modified in the *basic algorithm*. If $(j, p,\mathrm{r}) \in S_i$, primer S_i covered $(j, p+1), (j, p+2), ..., (j, m_j), (j, m_j+1)$ in the *basic algorithm*. But in this algorithm, if $(j, p,\mathrm{r}) \in S_i$, primer S_i covered $(j, p+k+1), (j, p+k+2), ..., (j, m_j+k), (j, m_j+k+1)$.

This modification (is called the *covered elements modification*) do not consider amplification constraint, and hence short segments may be amplified.

3.2.2 Satisfying Amplification Constraint

The following prohibitive condition is added to satisfy minimum length constraint. We consider selecting primer which makes short amplified sequences is prohibited. (Given t for minimum length of amplified segment.) To be exact, if primer S_i ($(j, p,\mathrm{r}) \in S_i$) was selected, S'_i ($(j, p',\mathrm{r/l}) \in S'_i$, $p \leq p' < p+t$) can not be selected. If primer S_i ($(j, p,\mathrm{l}) \in S_i$) was selected, S'_i ($(j, p',\mathrm{r/l}) \in S'_i$, $p - t < p' \leq p$) can not be selected.

We call this prohibitive condition the *short length condition*.

3.3 Difference of lengths of amplified segments

If the following prohibitive condition is added, the amplified segments are different from each other in their lengths. If lengths of segments which were amplified by the set of a candidate sequence and primers which was selected by algorithm are not different from each other, the candidate segments can not be selected this time.

We call this prohibitive condition the *resolution condition*.

4 Experimental Results

We apply these greedy algorithms to randomly selected 500 sequences out of about 6000 ORFs of yeast from GenBank. We set GC content is 45-55% in a primer and length of primer is 8 or 15. We consider 50 and 100 lower bounds. We pay attention to amplified segments whose lengths are between 50 and 500.

In the following tables, symbols mean these:
k: parameter for the *covered elements modification*
t: parameter for the *short length condition*
n_d: number of DNA sequences which have different lengths of amplified segments from the other amplified segments
$\bar{n}$: number of DNA sequences which have not amplified segment

Table 1: Effects for the *complementary condition* (Lengths of primers = 8)

	basic algorithm	+ *complementary condition*
#primers	102	106
#positions that primers match	3949	3965
#amplified segments	1269	1053
#amplified segments (lengths < 50)	984	456
#distinct amplified segments by lengths	154	170
n_d	135	135
$n_d(50, 500)$	111	116
$\bar{n}$	0	0
$n(1, 50)$	269	124
$n(1, 100)$	323	189

Table 2: Effects for the *covered elements modification* and the *short length condition* (Lengths of primers = 8)

k	0	50	50	100	100
t	0	0	50	0	100
#primers	106	117	186	129	183
#positions that primers match	3965	4348	3274	4601	2561
#amplified segments	1053	1154	834	1224	602
#amplified segments (lengths < 50)	465	513	0	535	0
#distinct amplified segments by lengths	170	143	174	157	206
n_d	135	121	146	127	167
$n_d(50, 500)$	111	95	115	118	120
$\bar{n}$	0	0	15	0	98
$n(1, 50)$	124	111	0	100	0
$n(1, 100)$	189	201	94	169	0

$n(i, j)$: number of DNA sequences which have only amplified segments whose lengths are between i and j
$n_d(i, j)$: number of DNA sequences which have different lengths of amplified segments from the other amplified segments and these lengths are between i and j

Table 1 shows the effects for the *complementary condition*. The *basic algorithm* can cover 500 sequences by about 100 primers. By modify this algorithm by the *complementary condition*, the number of primers is almost same. If self-complementary sequences and complementary sequences are removed, short amplified segments are decreased. After this, all algorithms in this paper adopt the *complementary condition*.

Table 2 shows the effects for the *covered elements modification* and the *short length condition*. Short amplified sequences are not decreased by the *covered elements modification*. The *short length condition* is necessary to satisfy lower bounds for amplified regions. By the *short length condition*, short length amplified sequences are removed, but some DNA sequences remain uncovered.

The case of primer length 15 is also investigated. Table 3 shows this case. In this case, effects for the *covered elements modification* can be seen. Amplified segments whose lengths less than 50 become half. This is because the number of positions which primers match in sequences is decreased. In the case of the *short length condition*, almost all DNA sequences are also covered. The number of the distinct DNA sequences by lengths are about 270, but about 150 sequences are distinct by long

Table 3: Effects for the *covered elements modification* and the *short length condition* (Lengths of primers = 15)

k	0	50	50	100	100
t	0	0	50	0	100
#primers	804	843	823	847	833
#positions that primers match	1045	1098	1034	1093	1039
#amplified segments	509	511	507	510	505
#amplified segments (lengths < 50)	105	50	0	42	0
#distinct amplified segments by lengths	268	275	286	273	281
n_d	266	271	283	270	278
$n_d(50, 500)$	118	117	122	111	113
$\bar{n}$	0	0	0	0	1
$n(1, 50)$	100	46	0	39	0
$n(1, 100)$	145	100	63	48	0

Table 4: Effects for the *resolution condition* (Lengths of primers = 8)

k	50		100	
t	50		100	
resolution condition	no	yes	no	yes
#primers	186	174	183	50
#positions that primers match	3274	2414	2561	1299
#amplified segments	834	453	602	243
#amplified segments (lengths < 50)	0	0	0	0
#distinct amplified segments by lengths	174	422	167	228
n_d	146	317	167	199
$n_d(50, 500)$	115	263	120	133
$\bar{n}$	15	170	98	199
$n(1, 50)$	0	0	0	0
$n(1, 100)$	94	30	0	0

amplified segments.

Table 4 shows the effects for the *resolution condition*. By the *resolution condition*, all amplified segments have different lengths. If $k = t = 50$, by the *resolution condition*, 263 DNA sequences are distinct by lengths of amplified segments between 50 and 500, and 263 pairs of primers are not necessary, only 186 primers are necessary. (263 pairs of primers are necessary for multiplex PCR.) If $k = t = 100$, by the *resolution condition*, 133 sequences are distinct. Because $t = 100$ in the *short length condition* is too strong, no more than 50 primers can be selected.

As the *basic algorithm* is modified, better solutions for biological constraints in 2.1 are obtained. By the *short length condition*, short amplified segments are removed. By the *resolution condition*, many amplified segments are distinct in length. Because short primers match many positions of DNA sequences, the number of primers for covering can be decreased. On the other hand, many short amplified segments are obtained. As the *short length condition* and the *resolution condition* are added, covering all sequences become difficult. The modified greedy algorithms can not obtain solutions which satisfy constraints and cover all sequences. But 500 sequences can not distinct between 50 and 500 in one experiments. We consider it is enough for our aim that about 250 sequences are distinct by lengths of amplification segments from 50 to 500 in one experiment.

5 Conclusions and Future Works

We have considered minimizing the number of primers for PCR experiments. As the greedy algorithms are modified, solutions which almost satisfy biological constraints are obtained. But these algorithms do not minimize the number of primers. It is necessary to analyze approximate properties of these algorithms from the theoretical and practical viewpoints. We will consider improving solutions for these algorithms by meta heuristics or other methods.

We will also consider take account of the other biological constraints and design the sets of primers for multiple experiments.

Acknowledgments

The authors would like to thank Dr. T. Ito of Human Genome Center, University of Tokyo for introducing them the primer selection problem and giving various comments on the problem. Part of this work by the second author was supported by the Grant-in-Aid for Priority Areas 'Genome Science' of the Ministry of Education, Science, Sports and Culture of Japan.

References

[1] J.S. Chamberlain, R.A. Gibbs, J.E. Ranier, P.N. Nguyen and C.T. Caskey, Deletion Screening of the Duchenne Muscular Dystrophy Locus via Multiplex DNA Amplification, Nucleic Acids Research 16, pp.11141–11156, 1988.

[2] K. Doi, Complexity of Primer Selection Problem in Genome Informatics, Preprint, Department of Information Science, University of Tokyo, May 1997.

[3] T. Ito and Y. Sakaki, Fluorescent Differential Display Method for High-speed Scanning of Tissue or Cell-specific Transcripts, Methods Mol. Genet. 8, pp.229–245, 1996.

[4] M. Mitsuhashi, Technical Report: Part 2. Basic requirements for designing Optimal PCR Primers, Journal of Clinical Laboratory Analysis 10, pp.285–293, 1996.

[5] M. Mitsuhashi, A. Cooper, M. Ogura, T. Shinagawa, K. Yano and T. Hosokawa, Oligonucleotide Probe Design – a New Approach, Nature 367, pp.759–761, 1994.

[6] P. Nicodème and J. Steyaert, Selecting Optimal Oligonucleotide Primers for Multiplex PCR, Proc. Fifth International Conference on Intelligent Systems for Molecular Biology, AAAI press, pp.210–213, 1997.

[7] W.R. Pearson, G. Robins, D.E. Wrege and T. Zhang, A New Approach to Primer Selection in Polymerase Chain Reaction Experiments, Proc. Third International Conference on Intelligent Systems for Molecular Biology, AAAI Press, pp.285–291, 1995.

[8] W.R. Pearson, G. Robins, D.E. Wrege and T. Zhang, On the Primer Selection Problem in Polymerase Chain Reaction Experiments. Discrete Appl. Math. 71, pp.231–246, 1996.

[9] T. Rowe, J. Sharefkin, S.Q. Yan and C.W. Diffenbach, A Computer Program for Selection of Oligonucleotide Primers for Polymerase Chain Reactions, Nucleic Acids Research 18, pp.1757–1761, 1990.

[10] J. Welsh and M. McClelland, Fingerprinting Genomes Using PCR with Arbitrary Primers, Nucleic Acids Research 18, pp.7213–7218, 1990.