International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:3, No:5, 2009

Primer Design with Specific PCR Product using Particle Swarm Optimization

Cheng-Hong Yang, Yu-Huei Cheng, Hsueh-Wei Chang, Li-Yeh Chuang

***Abstract*—**Before performing polymerase chain reactions (PCR), a feasible primer set is required. Many primer design methods have been proposed for design a feasible primer set. However, the majority of these methods require a relatively long time to obtain an optimal solution since large quantities of template DNA need to be analyzed. Furthermore, the designed primer sets usually do not provide a specific PCR product. In recent years, evolutionary computation has been applied to PCR primer design and yielded promising results. In this paper, a particle swarm optimization (PSO) algorithm is proposed to solve primer design problems associated with providing a specific product for PCR experiments. A test set of the gene CYP1A1, associated with a heightened lung cancer risk was analyzed and the comparison of accuracy and running time with the genetic algorithm (GA) and memetic algorithm (MA) was performed. A comparison of results indicated that the proposed PSO method for primer design finds optimal or near-optimal primer sets and effective PCR products in a relatively short time.

***Keywords*—**polymerase chain reaction (PCR), primer design, evolutionary computation, particle swarm optimization (PSO).

I. Introduction

POLYMERASE chain reaction (PCR) is a common technology applied in the biomedical and biotechnology field. This technology is used for fast mass duplication of DNA sequences [1]. By repeating some cycles of the denaturation, annealing and extension reactions, it allows a small amount of DNA to be amplified exponentially. Typically 25–45 of these cycles are performed [2]. Amplification of specific regions of the genome is determined by specific primer sets with forward and reverse orientation. In the past, feasible primer sets were usually found manually through trial and error, but this method is time consuming, because many constraints have to be satisfied at the same time. Typically considered primer design constraints are the length of the primer, the GC content, the melting temperature and GC clamp. The length of primers should be within 16 bps~28 bps, while the difference of the primer pair length should not exceed 3 bps. The GC content of primers should be within 40%~60%. The melting temperature of primers should be within the range of 50°C~62°C, while the difference of the melting temperature of a primer pair should not exceed more than 5°C. Finally, the 3' end of primers should be G or C.

In recent years, secondary structures were applied to primer design, such as dimer, hairpin and specificity of a primer pair. To date, various approaches for primer design have been proposed. Kämpke et *al.* (2001) use dynamic programming to design primers [2]. This allows for designing multiple primers for multiple target DNA sequences. Chen et *al.* (2003) used a GA to develop a web-based tool (PDA) for primer design [3]. Hsieh et *al.* (2003) developed an efficient algorithm using automatic variable fixing and automatic redundant constraint elimination to tackle the binary integer programming problem associated with the minimal primer set (MPS) selection problem [4]. Wang et *al.* employed a greedy algorithm to generate the MPS specifically annealed to all open reading frames (ORFs) in a given microbial genome to improve the hybridization signals of microarray experiments [5]. Wu et *al.* (2004) proposed a genetic algorithm (GA) imitating nature's process of evolution and genetic operations on chromosomes in order to achieve optimal solutions [6]. Miura et *al.* (2005) developed an algorithm identifies the specificity-determining subsequence (SDSS) of each primer and examines its uniqueness in the target genome [7]. Qu et *al.* (2009) developed MFEprimer program for evaluating the specificity of PCR primers based on multiple factors [8].

In this paper, an evolutionary algorithm for primer design with specific PCR product is proposed. The proposed algorithm is a particle swarm optimization (PSO). PSO is a population-based stochastic optimization technique, which was developed by Kennedy and Eberhart in 1995 [9]. PSO simulates the social behavior of organisms, such as birds in a flock or fish in a school, and describes an automatically evolving system. In PSO, each single candidate solution can be considered "an individual bird of the flock", that is, a particle in the search space. Each particle makes use of its own memory, as well as knowledge gained by the swarm as a whole to find the best (optimal) solution. All of the particles have fitness values, which are evaluated by an optimized fitness function. They also have velocities which direct the movement of the particles. During movement, each particle adjusts its position according to its own experience, and according to the experience of a neighboring particle, thus making use of the

Cheng-Hong Yang is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan (e-mail: chyang@cc.kuas.edu.tw).

Yu-Huei Cheng is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan (e-mail: yuhuei.cheng@gmail.com).

Hsueh-Wei Chang is with the Faculty of Biomedical Science and Environmental Biology, Kaohsiung Medical University, Taiwan (e-mail: changhw@kmu.edu.tw).

Li-Yeh Chuang is with the Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan (email: chuang@isu.edu.tw).

best position encountered by itself and its neighbor. The particles move through the problem space by following a current of optimum particles. The process is then reiterated a predefined number of times or until a minimum error is achieved [10]. PSO was originally introduced as an optimization technique for real-number spaces. PSO has been successfully applied in many areas: function optimization, artificial neural network training, fuzzy system control, and other application problems. A comprehensive survey of the PSO algorithms and their applications can be found in Kennedy *et al.* [11].

Here, primer constraints, such as primer length, difference of primer pair length, GC proportion, PCR product size, melting temperature (Tm), difference of melting temperature (Tm-diff), GC clamp, dimer of primers (including cross-dimer and self-dimer), hairpin and specificity are used to design optimal primer sets. Different PCR product and different methods of calculating the melting temperature are performed to compare with genetic algorithm (GA) and memetic algorithm (MA) primer design algorithms using the gene CYP1A1, which has been associated with a heightened lung cancer risk, and the accession number is NM_000499, which is defined in NCBI as "Homo sapiens cytochrome P450, family 1, subfamily A, polypeptide 1 (CYP1A1), mRNA". Experimental results indicated the proposed PSO for the primer design correctly and quickly identifies an optimal primer pair required for a specific PCR experiment.

II. PROBLEM DEFINITION

In this section, we define the primer design problem. Let T_D be the template DNA sequence, which is made up of base-nucleic acid codes of the DNA. T_D is defined as follows:

$$T_D = \{B_i \mid \forall B_i \in \{'A','T','C','G'\}, i \in Z^+\} \quad (1)$$

where B is the base-nucleic acid sequence, which is made up by 'A', 'T', 'C', or 'G'; i is the index of the position on T_D and is a positive integer (Z^+).

The primer design problem consists of finding a pair of sub-sequences of corresponding constraints from T_D. One sub-sequence is called the forward primer and the other is called the reverse primer. The forward primer and the reverse primer are defined as follows:

$$P_f = \{B_i \mid \forall B_i \in \{'A','T','C','G'\}, F_s \le i \le F_e \le T_D, i \in Z^+\} \quad (2)$$

$$P_r = \{\overline{B_i} \mid \forall B_i \in \{'A','T','C','G'\}, R_s \le i \le R_e \le T_D, i \in Z^+\} \quad (3)$$

where P_f is the forward primer, and F_s and F_e denote the start index and the end index of P_f in T_D. P_r is the reverse primer, R_s and R_e denote the start index and the end index of P_r in T_D. P_f and P_r are called a primer pair. $\overline{B}$ represents an the anti-sense sequence of the original base-nucleic acid sequence. For the sequence, B = "ACGTCGAACGGT", for example, the complement sequence is "TGCAGCTTGCCA". Since the complement of 'A' is 'T' and the complement of 'C' is 'G'. The anti-sense sequence is the reverse of the complement sequence; therefore the anti-sense sequence is $\overline{B}$ = "ACCGTTCGACGT".

In Fig. 1, the length of the template DNA is ι, the minimum PCR product length is ρ, the maximum PCR product length is σ, the start position of the forward primer is F_s, the length of the forward primer is F_l, the PCR product length between the forward primer and the reverse primer is P_l, the length of the reverse primer is R_l, the random range of F_s is η, and the length from F_s to the template DNA end is δ. Now, a vector with F_s, F_l, P_l and R_l can determine a primer pair. We define this vector as:

$$P_v = (F_s, F_l, P_l, R_l) \quad (4)$$

With P_v, we can calculate the reverse primer start index as:

$$R_s = F_s + P_l - R_l \quad (5)$$

Consequently, the forward primer and the reverse primer can be obtained from P_v. P_v is the prototype of a particle in the PSO, and later sections will use P_v to perform evolutionary computations. Table 1 gives a summary of all parameters in Fig. 1.

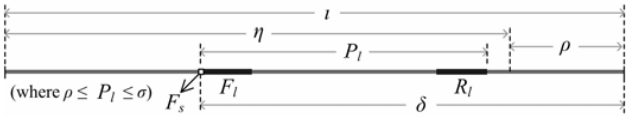

Fig. 1 Parameters of the template DNA and primer set.

TABLE I
LIST OF PARAMETERS IN FIG. 1

Parameter	Description
F_s	Start position of the forward primer
F_l	Length of the forward primer
P_l	PCR product length between forward primer and reverse primer
R_l	Length of the reverse primer
η	Random range of F_s
ρ	Minimum PCR product length
σ	Maximum PCR product length
δ	Length from F_s to the template DNA end
ι	Length of template DNA
F_s	Start position of the forward primer
F_l	Length of the forward primer

III. PARTICLE SWARM OPTIMIZATION FOR PRIMER DESIGN

The flowchart of the proposed algorithm is shown in Fig. 2. The processes of initial particle swarm, fitness evaluation, pbest and gbest finding, particle updating, and stopping criteria are described below.

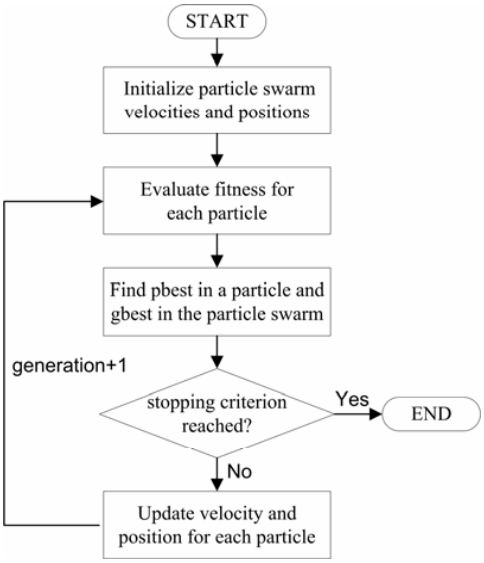

Fig. 2 PSO primer design flowchart

A. Initial particle swarm

Initially, one hundred particles $P_v = (F_s, F_l, P_l, R_l)$ are randomly generated as an initial particle swarm without duplicates. F_s is randomly generated between 1 and ($\iota \ \rho \ 1$). F_l is randomly generated between the minimum length of the primer and the maximum length of the primer. In the present study, the minimum length of the primer was set to 16 bps and the maximum length of the primer was set to 28 bps. In order to limit the PCR product length, we did not select $P_l = R_e \ F_s \ 1$, but randomly generated P_l between ρ and σ. R_l was randomly generated in the same way as F_l. Each particle is given a velocity (v). The velocity of each particle is randomly generated within 0~1.

B. Fitness evaluation

PSO requires a fitness function to evaluate the fitness of each particle in order to check whether the primers satisfy the design constraints. We use the primer design constraints as values for the fitness function and the fitness value is minimized.

A primer length of 16 bps to 28 bps is considered feasible for a PCR experiment. If a primer is longer, its specificity is higher, but a relatively high Tm is also required. On the other hand, a relatively short primer will decrease the specificity. Hence, neither a primer that is too long nor too short is suitable. We do not include the length constraint in the fitness function, because F_l and R_l are always limited between the minimum length of the primer and the maximum length of the primer in the constraints. The fitness value is provided by the following fitness functions, which are made up of $Len_{diff}(P_v)$, $Tm(P_v)$, $Tm_{diff}(P_v)$, $GC_{proportion}(P_v)$, $GC_{clamp}(P_v)$, $dimer(P_v)$, $hairpin(P_v)$ and $specificity(P_v)$, each of which is described below:

$$Fitness(P_v) = 3 * (Len_{diff}(P_v) + GC_{proportion}(P_v) + GC_{clamp}(P_v)) + 10 * (Tm(P_v) + Tm_{diff}(P_v) + dimer(P_v) + hairpin(P_v) + 50* specificity(P_v)) \quad (6)$$

$Len_{diff}(P_v)$ is used to check whether the length difference of a primer pair exceeds 3 bps. A length difference no more than 3 bps for the forward primer and the reverse primer is considered optimal in a PCR experiment.

The $Tm(P_v)$ function is used to check whether the melting temperature T_m of a primer pair is between 50°C and 62°C, and $Tm_{diff}(P_v)$ checks whether the difference of the melting temperature exceeds 5°C.

In the paper, the melting temperatures of primers are calculated by two formulas. The first one is Wallace's formula [12] and the other formula was proposed by Bolton and McCarthy [13]. The two computational formulas are:

(a) Wallace's formula:

$$Tm_W(P) = (\#G + \# C) * 4 + (\#A + \#T) * 2 \quad (7)$$

(b) Formula proposed by Bolton and McCarthy:

$$Tm_B(P) = 81.5 + 16.6 * (\log_{10}[Na^+]) + 0.41 * (GC\ content) - 675 / |P| \quad (8)$$

where P represents a primer and $|P|$ represents the length of P.

The GC proportion in a primer is denoted $GC\%(P)$, a value that indicates the ratio of the nucleotides G and C that appear in a primer. The $GC_{proportion}(P_v)$ function is used to check whether the $GC\%(P)$ of the forward and reverse primer is to lie in between a specific region. An appropriate GC proportion of a primer is in the range of 40-60%.

The function $GC_{clamp}(P_v)$ is used to check whether the 3' terminal end of a primer is G or C.

Annealing of two primers (called a dimer) will influence the results of a PCR experiment. Consequently, the combination of two primers should be avoided. Dimers include cross-dimers and self-dimers. A cross-dimer is formed when P_f and P_r anneal to each other, and a self-dimer is formed when P_f and P_f, or P_r and P_r anneal to each other. The function $dimer(P_v)$ is used to check whether the forward primer and the reverse primer anneal to each other or anneal to themselves.

A primer also should avoid annealing to itself. When a primer anneals to itself, it form a hairpin. The function $hairpin(P_v)$ is used to check for this condition in a primer.

Finally, the $specificity(P_v)$ function is used to judge whether the primer repeats itself in the template DNA sequence or not; it thus ensures the specificity of the primer. The PCR experiment might fail if the primers are not annealed to specific region and appears more than once in the DNA sequence. We used the number of times of P_f and P_r appear in T_D to adjust the fitness value; $specificity(P_v)$ is thus defined as the number of times of P_f and P_r reappear in T_D.

In the proposed fitness function, weights were used to discriminate between good and bad design constraint functions. Three weights represent different degrees of importance for the design constraint functions; these weights were set to 3, 10 and 50, respectively. The weights 3, 10 and 50 were chosen according to the PCR experiment requirements; these weights can be adjusted by biologists and researchers based on their

own experimental requirements. A weight of 50 represents a design constraint function of major importance, and a weight of 10 represents a design constraint function of a secondly importance. A weight of 3 represents a minor design constraint function. For example, consider P_{v1} and P_{v2} with the fitness function described above, for the $Len_{diff}(P_{v1}) = 0$, $GC_{proportion}(P_{v1}) = 0$, $GC_{clamp}(P_{v1}) = 0$, $Tm(P_{v1}) = 0$, $Tm_{diff}(P_{v1}) = 0$, $dimer(P_{v1}) = 0$, $hairpin(P_{v1}) = 0$ and $specificity(P_{v1}) = 1$. The fitness value is thus $Fitness(P_{v1}) = 50$. For $Len_{diff}(P_{v2}) = 1$, $GC_{proportion}(P_{v2}) = 2$, $GC_{clamp}(P_{v2}) = 2$, $Tm(P_{v2}) = 0$, $Tm_{diff}(P_{v2}) = 0$, $dimer(P_{v2}) = 0$, $hairpin(P_{v2}) = 0$ and $specificity(P_{v2}) = 0$, the fitness value is $Fitness(P_{v2}) = 15$. Although P_{v1} is only deficient in $specificity(P_{v1})$ and P_{v2} is deficient in the three parameters $Len_{diff}(P_{v2})$, $GC_{proportion}(P_{v2})$ and $GC_{clamp}(P_{v2})$, $Fitness(P_{v2})$ is better than $Fitness(P_{v1})$. Thus, the one deficiency in P_{v1} affects the PCR experiment to a greater extent P_{v2}, and P_{v2} is therefore the more feasible primer set for the PCR experiment.

C. Pbest and gbest finding

One of the characteristics of PSO is that each particle has a memory of its own best experience. Each particle can find its personal best position and velocity (called as *pbest*) and the global best position and velocity (called as *gbest*) when moving. If the fitness of a particle P_v is better than the fitness of *pbest* in the previous generation, *pbest* will be updated as P_v in the current generation. And then if the fitness of a particle P_v is better than *gbest* in the previous generation and is the best one in the current generation, *gbest* will be updated as P_v. Through *pbest* and *gbest*, each particle will adjust its appropriate direction in the next generation.

D. Particle updating

Each generation, the particles will change their position and velocity. Equation (9) and (10) are the updating formulas for each particle.

$$v_i^{next} = w \times v_i^{now} + c_1 \times rand() \times (s_i^p - s_i^{now}) + c_2 \times rand() \times (s_i^g - s_i^{now}) \quad (9)$$

$$s_i^{next} = s_i^{now} + v_i^{next} \quad (10)$$

In equation (9) and (10), v_i^{next} is the updated velocity of the *i*th particle; v_i^{now} is the current velocity of the *i*th particle; c_1 and c_2 are the constriction factors; w is the inertial weight; *rand( )* is a number which is randomly generated within 0~1; s_i^p is the personal best position of the *i*th particle; s_i^g is the global best position of the particles; s_i^{now} is the current position of the *i*th particle; s_i^{next} is the updated position of the *i*th particle. In order to avoid the particle outrunning the limits of F_s, F_l, P_l and R_l when updating, we use a random process to reset the particle according to primer constraints.

E. Stopping criteria

The algorithm is terminated when the particles have achieved the best position, i.e. their fitness value is 0, or the number of generations has reached.

IV. RESULTS AND DISCUSSION

In the recent years, primer design has become an important issue. The qualities of primers always influence whether a PCR experiment is successful or not. In this paper, we proposed a PSO algorithm to design an optimal primer set. The sequence of NM_000499 which is defined in NCBI as “Homo sapiens cytochrome P450, family 1, subfamily A, polypeptide 1 (CYP1A1), mRNA” was tested with the proposed algorithm. The gene CYP1A1 has been associated with a heightened lung cancer risk. We use Pentium 4 CPU 3.4 GHz and RAM 1GB with Microsoft Windows XP SP3 as test environment.

Five main parameters were set for the PSO; they are the number of iterations (generations), the number of particles, the inertial weight w, the constriction factors c_1 and c_2. Their values are set to 200, 10, 0.8, 2 and 2, respectively. We compare the proposed approach to GA and MA primer design. Four main parameters were set for the GA and MA; these parameters are the number of iterations, the population size, the probability of crossover and the probability of mutation. The respective values are 500, 100, 1.0 and 0.01 for GA and 100, 100, 1.0 and 0.01 for MA. Owing to the local search mechanism of MA, it took more time than GA for primer design. For a fair competition, the parameters for GA and MA only have a difference of the number of iterations. The processing time under these circumstances was almost identical for GA and MA, and thus the accuracy results are directly comparable.

Five hundred runs were performed with the three primer design methods, with PCR product length between 150~300 bps, 500~800 bps and 800~1000 bps, and a Tm calculated by Wallace formula and Bolton and McCarthy formula (Table II). The average accuracy was perfect when using PSO with the Wallace formula for primer design, and as high as 98.33% when using MA with the Wallace formula for primer design. However, the average accuracy only got up to 74.93% when using the GA to design primer pairs with the same Tm formula. We also performed Tm calculations with the Bolton and McCarthy formula with the same primer restrictions for PSO, MA and GA primer design methods. When using Bolton and McCarthy formula, the average accuracy was 94.93% for PSO primer design, and 88.93% for MA primer design. However, the average accuracy obtained with the GA primer designed is even worse.

From Table II, we can find that no matter accuracies or running times of the proposed approach for primer design with different product lengths are all better than MA and GA primer design. In the paper, we allow GA more iterations and running time approximate MA. We found the accuracies of using GA for primer design with specific PCR product still greatly lower than MA’s and PSO’s. The results indicate that primer design with specific PCR product using PSO and MA are excellent

methods and far superior to a GA. However, the best one is PSO.

GA was developed based on the Darwinian principle of the 'survival of the fittest' and the natural process of evolution through reproduction [14]. The performance of GA has been shown to outperform SFS (sequential forward search), PTA (plus and take away) and SFFS (sequential forward floating search) [15]. Although GA is used in many applications, they don't necessarily result in optimal solutions for all problems. The GA used in this study employs the fitness function mentioned above. GA is commonly employed in primer design, whereas the resulting accuracy is usually insufficient. Average accuracy for the Wallace formula and the Bolton and McCarthy formula were 74.93% and 32.40%, respectively. In conclusion, it can be said that GA lead to inferior solutions for primer design problems, especially when the Bolton and McCarthy formula is used to calculate Tm.

MA can be described as GA that focuses on local search. A local search is performed on each population member to improve its experience when the initial population is created and the offsprings of crossover and mutation are subjected to a local search process. The local search is conducted by adding or subtracting an incremental value from every gene, and then testing the chromosome's performance. In primer design, MA yielded higher accuracies. The average accuracies were 98.33% and 88.93% when using the Wallace formula and the Bolton and McCarthy formula to calculate Tm, respectively. MA results in near-optimum solutions when applied to primer design, and are generally superior to GA when the processing time is in the same range. However, it is still not generally used in primer design problems.

PSO is based on the idea of collaborative behavior and swarming in biological populations. PSO shares many similarities with evolutionary computation techniques like GA. PSO, MA and GA are population-based search approaches that depend on information sharing among their population members to enhance the search processes by using a combination of deterministic and probabilistic rules. However, PSO does not include genetic operators such as crossover and mutation. The recognition and social model of interaction between particles is similar to crossover and the random parameters will affect the speed of a particle, similarly to mutation in a GA or MA. In fact, the only different among them is that crossover and mutation in a GA or MA is probabilistic (crossover rate and mutation rate), but the renewed particle in PSO should be processed at each iteration without any probability. Compared with GA and MA, the information sharing mechanism in PSO is considerably different. In GA, the evolution is generated by using crossover and mutation in the same population. Chromosomes share information with each other, so the whole population moves like one group towards an optimal area. In the problem space, this model is similar to a search for only one area. Therefore, the drawback of this model is that it can easily trap into a local optimum. Although mutation is used, the probability usually is lower, limiting the performance. In MA, a local search mechanism is applied to avoid trap into a local optimum. It is better than GA to find the optimal solution, but it must take more time than GA. In PSO, particles are uniformly distributed in the problem space, and only *gbest* gives out information to other particles. It is a one-way information sharing mechanism. Evolution only looks for the best solution. In most cases all the particles tend to converge to the best solution quickly, even in the local version.

Compared to GA, PSO has a more profound intelligent background and can be performed more easily [16]. Computation time used in PSO is shorter than in GA [17], needless to say MA. The performance of PSO is affected by the parameter settings, inertia weight w, and the acceleration factors c_1 and c_2. However, if the proper parameter values are set, the results can easily be optimized. Proper adjustment of the inertia weight w and the acceleration factors c_1 and c_2 is very important. If the parameter adjustment is too small, the particle movement is too small. This scenario will also result in useful data, but is a lot more time-consuming. If the adjustment is excessive, particle movement will also be excessive, causing the algorithm to weaken early, so that a useful feature set can not be obtained. Hence, suitable parameter adjustment enables particle swarm optimization to increase the efficiency of optimal primer design.

The average accuracies were 100.00% and 94.93% when using the Wallace formula and the Bolton and McCarthy formula to calculate Tm with design different PCR product size, respectively. PSO produce near-optimum solutions when applied to primer design, and are superior to GA and MA when the processing time greatly less than GA and MA. Nevertheless, it is still covered up in primer design problems.

TABLE II
THE ACCURACY AND RUNNING TIME FOR GA, MA AND PSO PRIMER DESIGN USING WALLACE FORMULA AND BOLTON AND MCCARTHY FORMULA WITH PCR PRODUCT LENGTH BETWEEN 150 ~ 300 BPS, 500~800 BPS AND 800~1000 BPS. A, ACCURACY (%); T, RUNNING TIME (MS). BOLDFACE INDICATES HIGHEST VALUES

Tm formula and primer design methods	Wallace's formula						Bolton and McCarthy formula					
	GA		MA		PSO		GA		MA		PSO	
PCR product length	a (%)	t (ms)	a (%)	t (ms)	a (%)	t (ms)	a (%)	t (ms)	a (%)	t (ms)	a (%)	t (ms)
150~300 bps	76.60	1179390	99.60	1167562	**100.00**	90032	30.80	1654391	88.20	1152250	**95.40**	778797
500~800 bps	72.39	1167531	97.80	1180594	**100.00**	86609	33.20	1746156	89.40	1201156	**93.60**	844063
800~1000 bps	75.80	1171125	97.60	1168875	**100.00**	90718	33.20	1653312	89.20	1242563	**95.80**	781938
average	74.93	1172682	98.33	1172344	**100.00**	89120	32.40	1684619	88.93	1198656	**94.93**	801599

In this paper, PSO is proposed to design optimal primer sets, which can be correctly and efficiently applied to PCR experiments. The above results demonstrate that proposed approach is indeed design feasible primers in a dry dock analysis.

V. CONCLUSION

Primer design is important. It is the pre-action of PCR. The qualities of primers always influences whether a PCR experiment is successful or not. To date, many primer design approaches have been developed, but most of they are inefficient or do not design optimal primers for use in the PCR experiments. In this study, we designed optimal primer pairs using PSO, and performed primer constraints to appraise the fitness values, such as primer length, difference of primer pair length, GC proportion, PCR product length, melting temperature (Tm), difference of melting temperature (Tm-diff), GC clamp, dimer of primer pair (including cross-dimer and self-dimer), hairpin and specificity. Based on their significance, each constraint was given a corresponding weight. Through the design of a fitness function, feasible primer sets could always be found using the PSO algorithm.

We use a test set of the gene CYP1A1, associated with a heightened lung cancer risk to design primers with different PCR product and Tm formulas using three evolutionary computation approaches which are the proposed approach PSO, GA and MA, respectively. The accuracy and running time of our proposed approach was compared with GA and MA primer design. A comparison of results indicated that the proposed PSO method for primer design finds optimal or near-optimal primer sets and effective PCR products in a relatively short time. And the primer design results show that different Tm calculation methods affect the size of the primer length and the melting temperature. Using Wallace formula for calculating Tm acquires a shorter primer length and lower temperature value, but using Bolton and McCarthy formula for calculating Tm yields longer primer lengths and a higher temperature value.

In conclusion, this study applied PSO to design primers with specific PCR product lengths. The proposed approach is effective and takes relatively short time to design optimal primers. It can help biologists and researchers to effectively complete PCR experiments.

APPENDIX

The results of PSO, MA and GA primer design for NM_000499 with 500 runs in Pentium 4 CPU 3.4 GHz and RAM 1GB with Microsoft Windows XP SP3 and parameters setting described in this paper can be download at ftp://pd@bio.kuas.edu.tw/PSO/PSO-primer-results.rar.

ACKNOWLEDGMENT

This work is partly supported by the National Science Council in Taiwan under grants 97-2622-E-151-008-CC2 and 96-2221-E-214-050-MY3.

REFERENCES

[1] K. B. Mullis and F. A. Faloona, "Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction," *Methods Enzymol,* vol. 155, pp. 335-50, 1987.
[2] T. Kämpke, M. Kieninger, and M. Mecklenburg, "Efficient primer design algorithms," *Bioinformatics,* vol. 17, pp. 214-25, Mar 2001.
[3] S. H. Chen, C. Y. Lin, C. S. Cho, C. Z. Lo, and C. A. Hsiung, "Primer Design Assistant (PDA): A web-based primer design tool," *Nucleic Acids Res,* vol. 31, pp. 3751-4, Jul 1 2003.
[4] M. H. Hsieh, W. C. Hsu, S. K. Chiu, and C. M. Tzeng, "An efficient algorithm for minimal primer set selection," *Bioinformatics,* vol. 19, pp. 285-6, Jan 22 2003.
[5] J. Wang, K. B. Li, and W. K. Sung, "G-PRIMER: greedy algorithm for selecting minimal primer set," *Bioinformatics,* vol. 20, pp. 2473-5, Oct 12 2004.
[6] J. S. Wu, C. Lee, C. C. Wu, and Y. L. Shiue, "Primer design using genetic algorithm," *Bioinformatics,* vol. 20, pp. 1710-7, Jul 22 2004.
[7] F. Miura, C. Uematsu, Y. Sakaki, and T. Ito, "A novel strategy to design highly specific PCR primers based on the stability and uniqueness of 3'-end subsequences," *Bioinformatics,* vol. 21, pp. 4363-70, Dec 15 2005.
[8] W. Qu, Z. Shen, D. Zhao, Y. Yang, and C. Zhang, "MFEprimer: multiple factor evaluation of the specificity of PCR primers," *Bioinformatics,* vol. 25, pp. 276-8, Jan 15 2009.
[9] J. Kennedy and R. Eberhart, "Particle swarm optimization." vol. 4, 1995.
[10] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm." vol. 5, 1997.
[11] J. F. Kennedy, R. C. Eberhart, Y. Shi, and ScienceDirect, *Swarm intelligence*: Springer, 2001.
[12] R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura, "Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch," *Nucleic Acids Res,* vol. 6, pp. 3543-57, Aug 10 1979.
[13] E. T. Bolton and B. J. McCarthy, "A General Method for the Isolation of RNA Complementary to DNA." vol. 48: National Acad Sciences, 1962, pp. 1390-1397.
[14] E. Elbeltagi, T. Hegazy, and D. Grierson, "Comparison among five evolutionary-based optimization algorithms." vol. 19: Elsevier, 2005, pp. 43-53.
[15] I. S. Oh, J. S. Lee, and B. R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans Pattern Anal Mach Intell,* vol. 26, pp. 1424-37, Nov 2004.
[16] X. H. Shi, Y. C. Liang, H. P. Lee, C. Lu, and L. M. Wang, "An improved GA and a novel PSO-GA-based hybrid algorithm." vol. 93: Elsevier, 2005, pp. 255-261.
[17] Y. Rahmat-Samii, "Genetic algorithm (GA) and particle swarm optimization (PSO) in engineering electromagnetics," 2003, pp. 1-5.