We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Customized Simulated Annealing Algorithm Suitable for Primer Design in Polymerase Chain Reaction Processes

Luciana Montera[1], Maria do Carmo Nicoletti[2],
Said Sadique Adi[1] and Maria Emilia Machado Telles Walter[3]
[1]*Federal University of Mato Grosso do Sul*
[2]*Federal University of São Carlos*
[3]*University of Brasília*
Brazil

1. Introduction

The investigation of functionalities and other biological characteristics of proteins can be carried out from their corresponding gene sequence (DNA). The development of a process named PCR (Polymerase Chain Reaction) (Mullis & Faloona, 1987), based on an enzyme named DNA polymerase, was decisive for establishing a laboratorial procedure for generating thousands to millions of copies of a particular DNA sequence (amplification). Besides amplification, PCR-based techniques can also be employed in a variety of processes with different purposes, such as DNA sequencing and molecular evolution.

A basic PCR set up requires several components and reagents as described in (Sambrook & Russel, 2001). Among the components, the two most relevant for the *in silico* proposal and the experiments described in this chapter are the DNA fragment to be amplified (referred to as template) and a primer sequence pair. Primers are appropriate short nucleotide sequences added to the reaction in order to mark the limits of the target region, i.e., the region of the template to be amplified. Commonly two primers are used namely forward, for flanking the beginning of the target region, and reverse, for flanking its end. The primer pair is one of the most important parameters for a successful PCR. Its design involves several variables whose values, generally, are determined via extensive calculations. Also, primer design requires several issues; for instance, it should not be easy for primers to anneal with other primers in the mixture neither should they be biased to anneal among themselves, which would prevent them to anneal with the DNA template.

It is worth mentioning that although many different software systems are available for assisting primer design, such as those described in (Contreras-Moreira et al., 2009; Boyce et al., 2009; Mann et al., 2009; Kalendar et al., 2009; Gatto & Schretter, 2009; Piriyapongsa et al., 2009; You et al., 2008; Christensen et al., 2008; Tsai et al., 2007; Schretter & Milinkovitch, 2006; Liang et al., 2005; Boutros & Okey, 2004; Gordon & Sensen, 2004; Rose et al., 2003; Rozen & Skaletsky, 2000), the design process itself is still not well defined. This is mainly

due to the number of variables involved and the lack of consensus in relation to their adequate values. This adds an extra degree of uncertainty to the process which, by its own nature, is already prone to some uncertainty. Additionally, several available software systems have not been developed for general use; they have been designed to find primers in specific situations such as gene identification (Giegerich et al., 1996), measurement of eukaryotic gene expression (Gunnar et al., 2004), novel gene characterization (Costas et al., 2007), genetic disease diagnosis (Frech et al., 2009), detection of variations and mutations in genes (Evans & Liu, 2005; Haas et al., 2003; Ke et al., 2002) and molecular evolution (Lahr et al., 2009; Oliveira et al., 2006; Pusch et al., 2004; Antia et al., 2003; Patten et al., 1996). Besides, it must be mentioned that, undoubtedly useful, *in silico* primers must be sometimes adjusted in real experiments (Morales & Holben, 2009).

This chapter is an extension of a previous work (Montera & Nicoletti, 2008) where the design of a primer pair is approached as a search process based on a customized simulated annealing, implemented by an interactive software named SAPrimer[1]. After this Introduction the chapter is organized into four more sections as follows.

Section 2 contextualizes the application area (Molecular Biology) by introducing a few important definitions and processes relevant to the understanding of the work described in the chapter. It presents an up-to-date review of the state-of-the-art relative to PCR. PCR based methods such as real-time PCR, multiplex PCR and InterAlu-PCR will be briefly mentioned since they deal with important issues related to the process. The section also describes in detail the basic three steps composing a PCR process namely, (1) DNA denaturing, (2) primer and DNA template annealing and (3) primer extension. The three iterative steps are temperature-dependent and cyclically executed. The amplification process is responsible for producing a vast amount of copies from a small amount of DNA sequences (template). Critical to the amplification process is the adequate choice of a pair of primes. Considerations focusing on the importance of a well designed primer pair for the success of a PCR process as well as the main difficulties to design them are also presented.

Section 3 specifies and details the main variables to be considered when designing primers. Particular attention is given to the values assigned to the variables as well as their impact on the results.

Section 4 deals specifically with the use of a heuristic search method known as Simulated Annealing (SA) for solving the problem of finding an adequate pair of primes for a PCR process. As it is well known, the definition of a suitable objective (or fitness) function is a critical aspect when using SA for solving a problem. Aiming at promoting readability, Section 4 has been divided into two subsections. Subsection 4.1 presents the construction of the function in a systematic objective and didactic way, taking into account the variables and parameters listed in Section 3. It describes how they are combined into a function to be used to direct the search process conducted by the proposed customized version of SA. For a proper evaluation of a primer pair, the fitness function is defined considering how each variable value (size, composition, annealing temperature, etc.), calculated for each primer in a pair, differs from a pre-established set of values.

Subsection 4.2 focuses on the description of two releases of SAPrimer software, which implements an user-friendly computational environment to search for optimal pair of primers. Details of the main common functionalities of both releases (SAPrimer (R1) and

---

[1] http://www.facom.ufms.br/~montera/SAPrimer

SAPrimer (R2)) are presented, including a description of how a primer pair is found, evaluated and chosen (Montera & Nicoletti, 2008). Subsection 4.2 also discusses some new features incorporated to SAPrimer (R2). Particularly the search process was modified so that instead of keeping only one primer pair as a candidate solution, a list of the best primer pairs found so far during the iterative process is kept and, as final result, the list of selected pairs, ordered by their fitness values, is returned to the user. Other new functionalities included in SAPrimer (R2) are also discussed, such as finding primers for any user-defined frame size and finding degenerate primers for a given protein sequence.

Finally, Section 5 presents the final considerations and highlights the main contributions that the proposal and the available software can offer to those who need assistance for conducting PCR-related experiments.

2. The Role of Polymerase Chain Reaction (PCR) in Molecular Evolution and the Role of Primers in PCR

Directed molecular evolution (or *in vitro* evolution) is a laboratorial procedure that attempts to simulate the Darwinian evolution. One of its main goals is to obtain proteins with certain properties enhanced. The process starts with a pool of DNA (or RNA) molecules and, by implementing an iterative process of recombination, tries to construct new functional sequences from the previous ones.

The literature brings several directed molecular evolution methods such as error-prone polymerase chain reaction (epPCR) (Cadwell & Joyce, 1992), Stemmer's DNA shuffling (Stemmer, 1994a; Stemmer, 1994b), staggered extension (StEP) (Zhao et al., 1998), heteroduplex recombination (Volkov et al., 1999), degenerate homoduplex recombination (DHR) (Coco et al., 2002), assembly of designed oligonucleotides (ADO) (Zha et al., 2003) and codon shuffling (Chopra & Ranganathan, 2003). A review of some molecular evolution methods for enzymes can be found in (Lutz & Patrick, 2004). Usually, directed molecular evolution methods share the common goal of generating new sequences that encode functionally interesting proteins.

The process of *in vitro* evolution, as described in (Sun, 1999) and schematically shown in Fig. 1, starts by constructing a library of (DNA, RNA or protein) molecules using (1) random molecules of peptides or oligonucleotides or (2) variants of one or more parent molecule(s) obtained through mutagenesis, as described next. Usually option (1) is not appealing due to the vast amount of resulting molecules and their high diversity. However, a library built using a process of mutation (mutagenesis) from one or a few molecules, which are already known to have some desired property, can be more appealing, since diversity can be kept under control. The initial library (Pool 1 in Fig. 1) is then input to a process that selects potentially relevant molecules (i.e., those that can have a desired function) - generally only a small number of molecules are selected. Next a process of mutation (mutagenesis) is used in order to increase the number of the selected molecules, as well as their diversity. After mutation, the resulting molecules undergo to an amplification process, to have their numbers increased. The sequence selection, mutagenesis and amplification constitute a cycle of the *in vitro* evolution process. The cycle is repeated until molecules having the desired properties are finally selected.
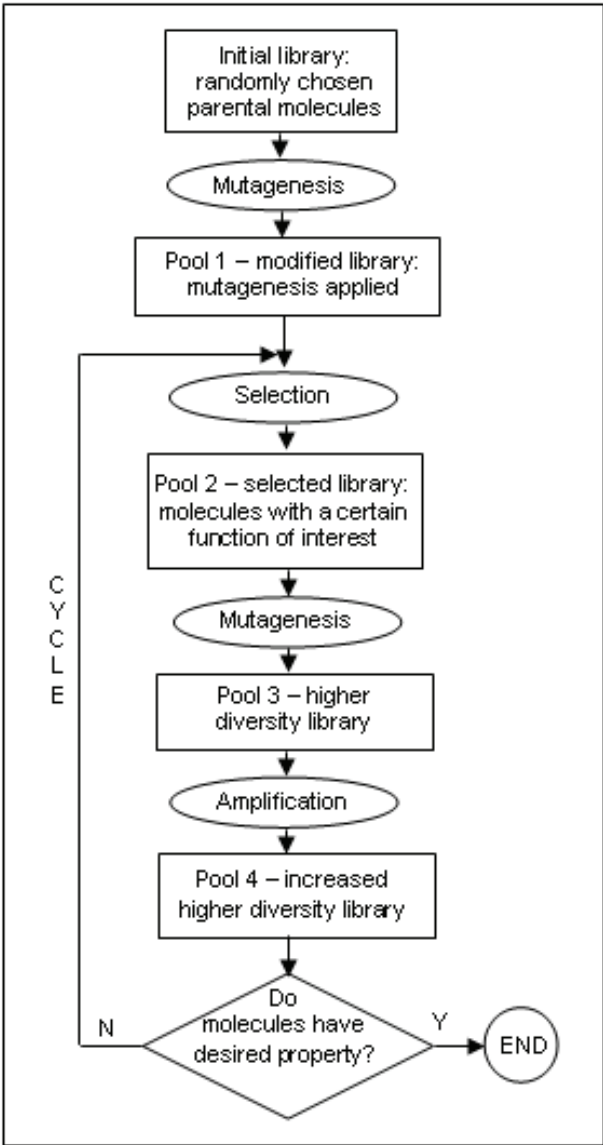
Fig. 1. General *in vitro* evolution scheme

Both mutagenesis and amplification are basic processes in *in vitro* evolution experiments. Mutagenesis can be used either for creating the initial molecular library or for increasing the molecular diversity, after a selection process took place, as shows Fig. 1. The amplification process allows the production of multiple copies of chosen target molecules and can be implemented by Polymerase Chain Reaction (PCR), that can be improved by using the primer search strategy described in the following sections of this chapter.

As defined in (Metzker & Caskey, 2001), "PCR is an elegant but simple technique for the *in vitro* amplification of target DNA using DNA polymerase and two specific oligonucleotide or primer sequences flanking the region of interest". The DNA polymerase synthesizes a new double-stranded of DNA from a single-stranded template. So that, it is necessary a $3' \rightarrow 5'$ primer (reverse) to make a complementary strand from a template in $5' \rightarrow 3'$ direction, and a $5' \rightarrow 3'$ primer (forward) to make a complementary strand from a template in $3' \rightarrow 5'$ direction. In a PCR cycle, the three temperature-controlled steps pictorially shown in Fig. 2 are:

(1) *Denaturing*: double-stranded DNA molecules are heated so that each double-stranded DNA molecule is completely separated into two single-stranded molecules;

(2) *Annealing*: the temperature is lowered such that primers anneal to their complementary single-stranded sequences;

(3) *Extension*: the temperature is raised up achieving an optimum point for the polymerase to react. DNA polymerases use the single-stranded molecules as templates to extend the primers that have been annealed to the templates in the previous step.
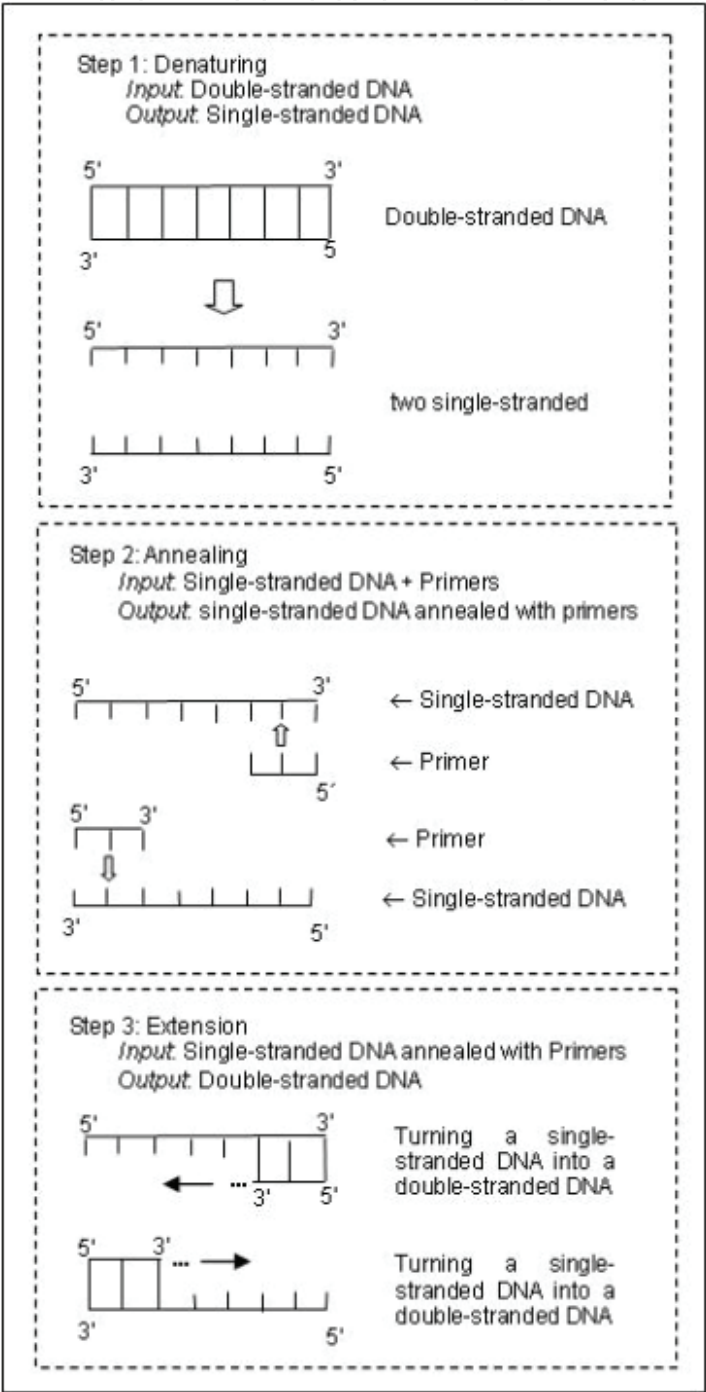
Fig. 2. The three temperature-controlled steps of a PCR process

PCR-based DNA technology can be used to perform mutation and recombination. As stated in (Tait & Horton, 1998), "The application of PCR techniques has blurred the distinctions among mutagenesis, recombination, and synthesis of genes. The product of PCR-based manipulations is really a mosaic in which sequences derived from natural sources are connected by sequences derived from the synthetic oligonucleotide primers used to direct the amplification; essentially any desired gene sequence can be constructed by combining natural, mutant, and synthetic regions".

The success of a PCR is highly dependent on the primer pair chosen and on the experimental conditions in which the reaction occurs, such as the number of cycles, the temperature and the time involved in each individual step as well as the quality and the volume of primers used in the annealing step. The denaturation and the annealing temperatures are directly dependent on the primers used. An exhaustive list of variables and parameters that interfere in a PCR experiment as well as a discussion about PCR techniques, its results, limitations and applications can be found in (Hugles & Moody, 2008; Dieffenbach & Dveksler, 2003; Kanagawa, 2003; Metzker & Caskey, 2001; He et al., 1994; Nuovo, et al., 1993; Innis & Gelfand, 1990; Allawi & SantaLucia, 1998a; Allawi & SantaLucia, 1998b; Allawi & SantaLucia, 1998c; Allawi & SantaLucia, 1997). Of particular interest for the work described in this chapter is the adequate choice of a pair of primes.

It is important to mention, as pointed out in (Hassibi et al., 2004), that "while in theory one would expect an exponential growth for the target as a function of PCR cycles (i.e., 2^n times the original DNA number, after n cycles), in practice, replication processes measured by different real-time PCR systems show varying yields, suggesting a biochemical random process. In addition to variable gains and inconsistent amplification levels within a PCR process, there is also the likelihood of creating non-specific byproducts (i.e., DNA strands different from the target) as well as of inserting mutations into the product, which further degrades the quality of the PCR product".

Besides DNA amplification and molecular evolution, PCR-based methods have been associated to a variety of processes. Studies of gene expression level (Isenbarger et al., 2008; Dixon et al., 2007), loss of allelic heterozygosity (LOH) (Vladušić et al., 2010; Chih-Ming et al., 2009; Franko et al., 2008; Saelee et al., 2008), microsatellite instability (MSI) (Eveno et al., 2010; Bertagnolli et al., 2009), microdeletions (Kolb et al., 2010; Pasmant et al., 2009), quantification of small ncRNAs (Ro et al., 2006; Berezikov et al., 2006) and detection of low-level mutations (Milbury et al., 2009) are a few examples of the popularity, success and diversity of uses of PCR. Some of the examples can be characterized as quantitative PCR or Real Time PCR (Lan et al., 2009; Roux, 2009; VanGuilder et al., 2008; Pattyn et al., 2003; Vandesompele et al., 2002) or quantitative multiplex PCR (Sasaki et al., 2010; Wang et al., 2009; Castellsagué et al., 2008), Inter-Alu PCR (Bonafè et al., 2001; Srivastava et al., 2005), or COLD-PCR (Milbury et al., 2009) and miniprimer PCR (Isenbarger et al., 2008).

3. Primers and Their Main Characteristics

The primer pair that can be used in the annealing step of a PCR is not unique, since forward and reverse primers with different sizes are possible. This poses a non-deterministic aspect to the process. The pair of primers that promotes the best results of a PCR (i.e., the pair that optimizes the amount and the specificity of the product) is named optimum pair.

Finding the optimum pair involves the simultaneous analysis of many parameters like primer size, primer contents of cytosine and guanine bases, melting temperature of the DNA strands, 3' end composition, specificity, formation of secondary structure between forward and reverse primers or between two forward or two reverse primers, and several others. Since there is not a unique value that defines the proper value for each one of these variables, generally a range of values is considered. For example, the goal could be to find a primer whose length is in the interval 18-30 bp having its cytosine and guanine composition (%CG) between 40-60%.

Frequently value restrictions imposed on some variables can conflict with values assigned to other variables. In spite of the non-existence of a consensus for the exact values of several parameters, studies can be found in the literature establishing values, or intervals of values for them (Apte & Daniel, 2009; Vikalo et al., 2006; Yu et al., 2006; Acinas et al., 2005; Abd-Elsalam, 2003; Kanagawa, 2003; Sommer & Tautz, 1989).

Each of the next four subsections groups important primer characteristics which will be used to create specific metrics to be incorporated into the objective function used by the SA (described in Section 4). The criterion used for grouping them is identified in the title of the corresponding subsection.

3.1 Repeats, runs and secondary structures

The repetition of nucleotide sequences (named *repeat*) inside of a primer sequence should be avoided since the occurrence of repeats can favor the occurrence of *misprimes*, i.e., a misplaced annealing between the template and the primer sequence that causes the amplification of a region different from the target region. The occurrence of a long repetition of one single base in the primer sequence is named *run*. Runs should be avoided because they can favor misprimes.

Another important characteristic that should be avoided during a primer design is their *self-complementarity*, which promotes the primer-primer annealing. Self-complementary primer sequences can affect the PCR efficiency by reducing the concentration of single-stranded primers since some annealed primer-primer could be extended by the polymerase, resulting in an unwanted non-specific product. Three distinct primer-primer annealing situations can occur, promoting the construction of secondary structures known as *self-dimer* (annealing between two forward or two reverse primer sequences), *hetero-dimer* (annealing between one forward and one reverse primer sequence), and *hairpins* (annealing of a primer sequence, to itself).

3.2 Specificity and primer length

A forward primer is considered specific if it anneals to the template just at the beginning of the target region. A reverse primer is considered specific if it anneals to the template just at the ending of the target region. The specificity of a primer is highly important to assure that the PCR product will correspond exactly to the target region, that is, to the region to be amplified. A way to evaluate the specificity of a primer sequence is by 'sliding' it along the length of the template, trying to detect alternative priming sites, other than the target region. Clearly, primers that promote alternative annealing sites are not a good choice. The specificity is closely related to the primer length.

The choice of primer lenght involves at least three parameters: specificity, annealing stability and cost. The longer is the primer, the smaller are the chances of existing alternative priming sites, i.e., the longer the primers, the more specific they are. Longer primers are more stable due to the greater number of hydrogen bonds they form with the template. Longer primers, however, are more biased to the formation of secondary structures and are financially more expensive to be produced. Shorter primers, in spite of their lower cost, are prone to anneal outside the target region, resulting in non-specific product, lowering the quality of the PCR product. There is no single optimum length for a primer. A rule-of-thumb suggested in (Abd-Elsalam, 2003) is "primers of 18–30 nucleotides in length are the best".

3.3 The %CG content and the 3´ end

The percentage of cytosine (C) and guanine (G) bases (%CG) in a primer sequence is very important because these numbers provide information about the annealing stability/strength. The binding between thymine (T) and adenine (A) bases occurs due to the formation of two hydrogen bonds; the binding between cytosine (C) and guanine (G) bases occurs due to the formation of three hydrogen bonds, making the latter more stable and more difficult to be formed and broken. As a consequence, the CG content of a primer directly influences the temperature in which the annealing between the primer and the template will occur. In general, primers with a CG content varying between 40% and 60% are preferred.

Mismatches can occur during the annealing between a primer and a template. They can be located anywhere (inside or at the end of the primer–template complex) and can affect the stability of the complex, causing undesirable side effects as far as the efficiency of the polymerase extension process is concerned. A mismatch located at (or near) the 3′ end of a primer (where the extension by polymerase starts) has a greater damaging effect than those located at other positions (Kwok et al., 1990). Based on this information, it can be inferred that the 3′ end of a primer should be well "stuck" to the template, so that the polymerase can start and conduct the extension process efficiently. Due to the strong binding between the C and G bases, the presence of either at the 3´ end of a primer should be preferred (over the occurrence of a T or A) since this will (potentially) assure more stability to the primer-template complex.

3.4 Melting and annealing temperatures

The melting temperature (T_m) is the temperature at which 50% of the DNA molecules are in duplex form and the other 50% are in denatured form. In a PCR, it is expected that while the template molecules denature, the primer molecules anneal to the single-stranded resulting sequences (templates). The temperature at which the annealing between the primer and the template occurs is defined as the annealing temperature (T_a). The T_m value can be defined in relation to both the product (amplified templates) and the primers; the T_a calculation is particularly dependent on both. There are several different methods to estimate the T_m value, which can be broadly classified according to the adopted methodology: *Basic* (only considers the %CG content), *Salt Adjusted* (takes into account the salt concentration at the solution) and *Thermodynamic* (uses the Nearest Neighbor model).

The most basic formula for the T_m calculation was given in (Wallace et al., 1979) and is shown in eq. (1) in Table 1, where $|C|$, $|G|$, $|A|$ and $|T|$ represent, respectively, the

number of cytosine, guanine, adenine and thymine bases present in the DNA sequence. Eq. (1) establishes that value of T_m is directly related to the length and contents of a DNA sequence. Another basic formulation, proposed in (Marmur & Doty, 1962), is given by eq. (2) in Table 1, which assumes that the reaction occurs at the standard conditions of pH = 7.0, 50mM Na^+ and 50nM of primer concentration. The salt adjusted formulation proposed in (Howley et al., 1979), considering the same values for the pH and sequence concentration as in (Marmur & Doty, 1962), is given by eq. (3), in Table 1.

$$T_m = 2*(|A|+|T|)+4*(|C|+|G|) \tag{1}$$

$$T_m = \frac{64.9+41*(|C|+|G|-16.4)}{(|A|+|T|+|C|+|G|)} \tag{2}$$

$$T_m = 100.5+41*\frac{|C|+|G|-6.4}{|A|+|T|+|C|+|G|}-\frac{820}{|A|+|T|+|C|+|G|}+16.6*\log[Na^+] \tag{3}$$

Table 1. Basic and salt adjusted T_m formulas

Formulations dependent on the Nearest Neighbor (NN) model are widely used; one of the reasons is due to "the stability of a DNA duplex appears to depend primarily on the identity of the nearest neighbor bases", as stated in (Breslauer et al., 1986). Considering the four bases, there are sixteen different pairwise nearest neighbor possibilities that can be used to predict the stability and the T_m of a duplex. The NN model establishes values for the enthalpy and entropy variation (represented by ΔH and ΔS, respectively) to each one of the sixteen pairs. Several studies propose different values for ΔH and ΔS, as those ones described in (Breslauer et al., 1986) , (Sugimoto et al., 1996) and (SantaLucia et al., 1996). The ΔH and ΔS values for a DNA sequence $X = x_1x_2 \ldots x_n$ are calculated using eq. (4).

$$\Delta H = \sum_{i=1}^{n-1}\Delta H(x_i, x_{i+1}) \qquad \Delta S = \sum_{i=1}^{n-1}\Delta S(x_i, x_{i+1}) \tag{4}$$

A commonly used formulation to calculate the T_m value considering the contribution of the NN model was proposed in (Rychlik et al., 1990) and is given by eq. (5), where R = 1.987 cal/°C*mol is the molar gas constant, γ is the primer concentration in the solution, $[Na^+]$ is the salt concentration, and ΔH and ΔS the enthalpy and entropy variation of the primer sequences, respectively.

$$T_m = \frac{\Delta H}{\Delta S + R * \ln\frac{\gamma}{4}} - 273.15 + 16.6 * \log[Na^+] \tag{5}$$

A few other proposals to calculate T_m can be found in the literature. It is important to mention, however, that all the attempts to define a proper value for T_m are only an approximation of the real melting temperature since, as commented in (Kämpke et al., 2001), "A proper computation of the primer melting temperature does not appear to exist". a comparative study of different melting temperature calculation methods as well as the influence of the different NN interaction values available for the T_m calculations is presented in (Panjkovich & Melo, 2005).

Although there have been some attempts to estimate T_a (such as in (Rychlik et al., 1990)), it seems that there is a consensus in the literature that the T_a value should be empirically determined (see (Innis & Gelfand, 1990)).

4. A Customized Simulated Annealing Algorithm for PCR Primer Design

Generally speaking, simulated annealing (Kirkpatrick et al., 1983) is a probabilistic algorithm suitable for finding a good approximation to a global optimum of a given function in a large search space. It is based on successive steps, which depend on an arbitrarily set parameter named temperature (T).

In this chapter the design of a primer pair has been approached as an optimization problem, using a customized SA to conduct a search process throughout the space of all possible primer pairs, trying to find an optimal solution (i.e., a primer pair) to a function. The SA technique is heavily dependent on an appropriate choice of the function to be optimized. For this particular domain, the function was constructed based on the primer relevant characteristics for a successful PCR when amplifying a given DNA target, as described in Section 3. The next two subsections focus, respectively, on the construction of the objective function and on the description and use of the two releases of the SAPrimer software, that implement the search for an optimal pair of primers using the objective function previously constructed.

4.1 Constructing the objective function

Before presenting and discussing the function used in the experiments, the basic metrics implemented for evaluating primer characteristics are described in Table 2, where *fp* and *rp* represent the forward and reverse primer in a primer pair, respectively.

In order to "measure" how "good" a primer pair (fp, rp) is (relative to the probability of a successful PCR), it is mandatory to evaluate its conformity to a pre-established (user-defined) range of values, as well as to check the occurrence of any of unwanted characteristics, such as runs, repeats, secondary structures and non-specificity. The range of parameter values that should be defined by the user are listed in Table 3.

In Table 3, the parameter MAX_DIF establishes the allowed maximum T_m difference ($T_m$dif) between the forward and the reverse primer. Both T_m and T_m difference are measured in Celsius degree (°C). The 3'_END is a Boolean parameter that specifies the user preference (or not) for the occurrence of base C or G at the 3' end of the primer sequences.

As there is no agreement about the best T_m formula to be used, the T_m value is estimated by the average of all T_m values calculated using all distinct formulas described in Table 1 of Section 3.4. When T_m uses the enthalpy and entropy contribution, a calculation is done for each distinct NN interaction values proposed in (Breslauer et al., 1986), (Sugimoto et al., 1996) and (SantaLucia et al., 1996).

The function to evaluate a primer pair (*fp*, *rp*) used for implementing the SA algorithm is given by eq. (6). It "measures" how well the argument pair fits the pre-established range of values (for the characteristics given in Table 3) and how "good" it is, concerning others (e.g. absence of repeats, runs, etc.), by associating a "cost"to the unwanted values of the characteristics. The highest is the function value, the less suitable is the primer pair given as its argument, since costs indicate how far the characteristics of a primer pair are from those pre-established by the user.

Metric	Type	Description
len(*fp*), len(*rp*)	integer	Gives the length of the argument sequence
CG(*fp*), CG(*rp*)	real	Gives the % of C and G bases in the argument sequence
T_m(*fp*), T_m(*rp*)	real	Gives the T_m of the argument sequence
$T_m$dif(*fp*, *rp*)	real	Gives the T_m difference between the argument sequences
3'_end(*fp*), 3'_end(*rp*)	Boolean	Checks the existence of a C or G base at the 3' end of the argument sequence
run(*fp*), run(*rp*)	Boolean	Checks the existence of runs in the argument sequence
repeat(*fp*), repeat(*rp*)	Boolean	Checks the existence of repeats in the argument sequence
spec(*fp*), spec(*rp*)	Boolean	Checks the specificity of the argument sequence
sec(*fp*, *rp*)	Boolean	Checks the existence of secondary structures in the argument sequences

Table 2. Basic metrics for evaluating a primer

Parameter	Range of values/Type
LENGTH_INTERVAL	[MIN_LEN, MAX_LEN] / both integer
%CG_CONTENT	[MIN_CG, MAX_CG] /both real
Tm	[MIN_Tm, MAX_Tm] / both real
MAX_DIF	- / real
3'_END	- / Boolean

Table 3. User-defined parameter values

fitness (*fp*, *rp*) = len_cost(*fp*) + len_cost(*rp*) + %CG_cost(*fp*) +
+ %CG_cost(*rp*) + 3*(T_m_cost(*fp*) + T_m_cost(*rp*)) +

$$+ \; T_m \text{dif\_cost}(\textit{fp}, \textit{rp}) + 3'\_\text{end\_cost}(\textit{fp}) + \tag{6}$$
$$+ \; 3'\_\text{end\_cost}(\textit{rp}) + \text{run\_cost}(\textit{fp}) + \text{run\_cost}(\textit{rp}) +$$
$$+ \; \text{repeat\_cost}(\textit{fp}) + \text{repeat\_cost}(\textit{rp}) + \text{spec\_cost}(\textit{fp}) +$$
$$+ \; \text{spec\_cost}(\textit{rp}) + \text{sec\_struc\_cost}(\textit{fp}, \textit{rp})$$

Taking sq as *fp* or *rp*, each individual cost function is defined as:

$$\text{len\_cos}\,t(\text{sq}) = \begin{cases} 0 \text{ if } \text{MIN\_LEN} \le \text{len}(\text{sq}) \le \text{MAX\_LEN} \\ \text{MIN\_LEN} - \text{len}(\text{sq}) \text{ if } \text{len}(\text{sq}) < \text{MIN\_LEN} \\ \text{len}(\text{sq}) - \text{MAX\_LEN} \text{ if } \text{len}(\text{sq}) > \text{MAX\_LEN} \end{cases}$$

$$\%\text{CG\_cos}\,t(\text{sq}) = \begin{cases} 0 \text{ if } \text{MIN\_CG} \le \text{CG}(\text{sq}) \le \text{MAX\_CG} \\ \text{MIN\_CG} - \text{CG}(\text{sq}) \text{ if } \text{CG}(\text{sq}) < \text{MIN\_CG} \\ \text{CG}(\text{sq}) - \text{MAX\_CG} \text{ if } \text{CG}(\text{sq}) > \text{MAX\_CG} \end{cases}$$

$$T_m\_\text{cos}\,t(\text{sq}) = \begin{cases} 0 \text{ if } \text{MIN\_}T_m \le T_m(\text{sq}) \le \text{MAX\_}T_m \\ \text{MIN\_}T_m - T_m(\text{sq}) \text{ if } T_m(\text{sq}) < \text{MIN\_}T_m \\ T_m(\text{sq}) - \text{MAX\_}T_m \text{ if } T_m(\text{sq}) > \text{MAX\_}T_m \end{cases}$$

$$T_m\text{dif\_cos}\,t(\text{fp},\text{rp}) = \begin{cases} 0 \text{ if } T_m\text{dif}(\text{fp},\text{rp}) \le \text{MAX\_DIF} \\ T_m\text{dif}(\text{fp},\text{rp}) - \text{MAX\_DIF} \text{ otherwise} \end{cases}$$

$$3'\_\text{end\_cos}\,t(\text{sq}) = \begin{cases} 0 \text{ if } 3'\_\text{end}(\text{sq}) = \text{true} \\ 5 \text{ otherwise} \end{cases}$$

$$\text{run\_cos}\,t(\text{sq}) = \begin{cases} 0 \text{ if } \text{run}(\text{sq}) = \text{false} \\ 5 * \text{number of runs otherwise} \end{cases}$$

$$\text{repeat\_cost}(\text{sq}) = \begin{cases} 0 \text{ if } \text{reapet}(\text{sq}) = \text{false} \\ 5 * \text{number of repeats otherwise} \end{cases}$$

$$\text{spec\_cos}\,t(\text{sq}) = \begin{cases} 0 \text{ if } \text{spec}(\text{sq}) = \text{true} \\ 5 * \text{number of alternative } pri\,ming \text{ sites otherwise} \end{cases}$$

$$\text{sec\_struc\_cos}\,t(\text{fp},\text{rp}) = \begin{cases} 0 \text{ if } \text{sec}(\text{fp},\text{rp}) = \text{false} \\ \sum_{i=1}^{5} \text{highest\_cos}\,t(G_i) \text{ otherwise} \end{cases}$$

The first three cost functions assign a cost to a primer (forward or reverse) whose values for primer length, %CG content and melting temperature, respectively, are outside the pre-established limits given in Table 3. The fourth cost function ($T_m$dif_cost) assigns a cost to a primer pair depending on its temperature difference be higher than the user-defined parameter MAX_DIF.

The inclusion of the fifth cost function for the calculation of the value of fitness is dependent on the information given by the user regarding his/her preference (or not) for a base C or G at the 3′ end of the primers. If the user has no preference, 3′_end_cost(*sq*) is not included in the fitness calculation. Otherwise, the non-existence of a C or G base at the 3′ end of a primer adds the arbitrary cost of 5 to the fitness value. The cost associated to the presence of runs and repeats is given by the functions run_cost and repeat_cost, respectively, which add a cost of 5 to each time a run or a repeat is found. The spec_cost function is similar to the two previous cost functions.

The last cost function assigns a cost to the possible secondary structures that may be formed in each of the following five groups (G_i, i = 1,...,5): hetero-dimer, self-dimer (forward), self-dimer (reverse), hairpin (forward) and hairpin (reverse). In each group, different annealing situations can happen. The sec_struc_cost function takes into consideration only the annealing situation with the highest cost per group (highest_cost(G_i)). The cost of any annealing situation is given as the sum of the numbers of A-T matches (costs 2 each) and C-G matches (costs 4 each), as suggested in (Kämpke et al., 2001).

The pseudocode of the implemented SA algorithm is given in Fig. 3. The algorithm starts by randomly choosing a pair of primers (referred to as current) *fp_cur* and *rp_cur*, such that |*fp_cur*| = m and |*rp_cur*| = n with MIN_LEN ≤ m, n ≤ MAX_LEN; the first m bases and the last n complementary bases of the target DNA sequence are the *fp_cur* and *rp_cur* respectively and their cost is evaluated.

```
procedure SAPrimer
  begin
    fp_cur = find_primer(MIN_LEN, MAX_LEN)
    rp_cur = find_primer(MIN_LEN, MAX_LEN)
    cost_cur = tot_cost(fp_cur, rp_cur)
    T = 200
    decreasing_factor = 0.999
    while (T > 0.01)
        fp_new = find_primer_neighbor (len(fp_cur))
        rp_new = find_primer_neighbor (len(rp_cur))
        cost_new = tot_cost (fp_new, rp_new)
        if (cost_new < cost_cur) then
            // change the current primer pair
            fp_cur = fp_new
            rp_cur = rp_new
            cost_cur = cost_new
        else
            num = random( )
            if (num < exp^(−ΔE / T)) then
                fp_cur = fp_new
                rp_cur = rp_new
                cost_cur = cost_new
        T = T * decreasing_factor
  end
```

Fig. 3. Pseudocode of the customized SA algorithm used by SAPrimer

At each step, the algorithm randomly chooses a new candidate-pair (fp_new, rp_new) in the neighborhood of the current pair; any primer pair such that $|fp\_cur| - 3 \leq |fp\_new| \leq |fp\_cur| + 3$ and $|rp\_cur| - 3 \leq |rp\_new| \leq |rp\_cur| + 3$ has its cost evaluated. The cost value of the new candidate solution is then compared with the cost of the current solution. The primer pair with the smaller cost becomes the current pair. Notice, however, that even if the new candidate has a bigger fitness value, it can be chosen as the current pair, depending on a probability function based on both the T parameter and the ΔE parameter (where ΔE is the difference between the new and current solution cost). The acceptance of a solution with a higher cost is an attempt to prevent local minima. The value of 200 assigned to the T parameter was empirically determined as well the decreasing_factor of 0.999.

4.2 The SAPrimer – An automatic environment for searching an optimal pair of primers

Both releases of SAPrimer were developed using Builder C++ environment and run under the Windows operating system or under the Linux operating system with an appropriated emulator. They are user-friendly and do not require in-depth knowledge of primer design or heuristic methods.

SAPrimer (R2) takes as input a DNA or a protein sequence, which can be described in fasta or text plain format, and find a primer pair that amplifies the input sequence. If not modified by the user, default values are used for all parameters involved in the primer search: minimum and maximum values for T_m, %CG, primer length, T_m difference and 3' end preference, besides those ones that control the SA algorithm: initial temperature and decreasing factor. SAPrimer (R1) takes as input only DNA sequences.

Fig. 4 shows the use of the SAPrimer (R2) when searching for a suitable primer pair to amplify the DREB1A gene from *Arabidopsis thaliana* (NCBI Reference Sequence: NC_003075.4). Genes from this family exhibit tolerance to abiotic stresses such as low temperatures and drought. The parameters in Fig. 4 are shown with their default values. SAPrimer (R2) prompts the best primer pair found (forward and reverse), showing its length, %CG, T_m and fitness. The list of the best ten primer pairs found is given by clicking on the button "See best 10 primer pairs". A graphic is plotted showing, at each iteration, the fitness of the current solution as evidence of the process convergence (or not).

Notice in Fig. 4, that the optimal primer pair found does not conform to the %CG restriction once the reverse primer has only 20.83% of bases C and G in its composition. Furthermore, the reverse primer does not include a C or G base at its 3' end. These restrictions fail due to the low GC content at the 3´ extremity of the DREB1A gene sequence (among the last 45 bases, only 4 are C and 8 are G). As explained in Subsection 4.1, an optimal primer pair is a pair that deviates the least from the user-defined parameter values; a solution will always be found, even in cases where there is no primer pair satisfying all restrictions. This decision prevents the user from not obtaining results; them results that do not entirely conform to the user's specifications are taken into account. This is not the case of a few computational systems; as an example, the online available software Primer3[2] (Rozen & Skaletsky, 2000) could not find any solution under the same entries used in the example given in Fig. 4.

---

[2] http://frodo.wi.mit.edu/primer3/

With SAPrimer (R2), the user can also perform a search taking into account only a frame of the input sequence. In order to do this, the value of parameter Product size (whose default value is the size of the input sequence) should be modified to the size of the frame.
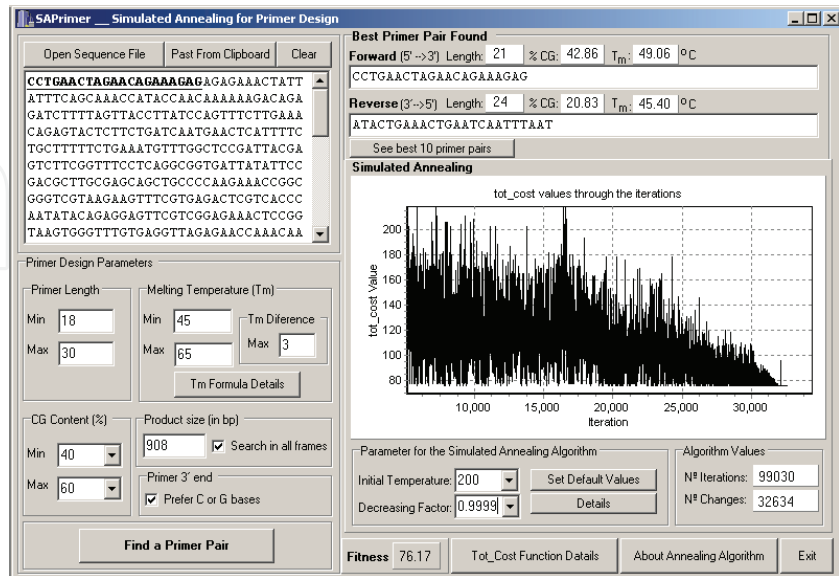

Fig. 4. SAPrimer (R2) interface

In Fig. 5 for example, the input DNA sequence has 908 bp (NC_003075.4), but a primer pair is desired for amplifying any portion of 897 bp. To provide more precise results, this search can be performed in two different ways: 1) looking for all frames having the specified size or 2) looking for a primer pair with the specified size at some frames. The user can choose the first (or second) search mode by clicking (or not) the "Search in all frames" button on the SAPrimer graphic interface. Generally, considering a sequence $S = s_0, s_1, ..., s_n$ to perform the search mode 1) for a frame of size k, k < n, the SA (as described in Fig. 3) must be executed n–k times, one for each possible frame of length k (frames starting at position 1, 2, 3 until n-k).

Fig. 5. Optimal primer pairs found for all frame sizes having 897 bp

To implement the search mode 2) the SA presented in Fig. 3 was slightly modified. It also differs from the first implementation of the search mode 1), that has a fixed frame to search. In the search mode 2), the SA randomly chooses a frame at each step. For each chosen frame, a primer pair conforming to the size restrictions (min and max) is randomly chosen and its fitness is calculated. At the end of this search mode, the results prompted to the user show the primer pair with the smaller fitness score, its size, T_m and %CG. Fig. 5 shows the results of SA, when search mode 2) is chosen, for the NC_003075.4 sequence and a frame of size 897 bp.

Another functionality implemented in SAPrimer (R2) allows the use of a protein sequence as input. In this case, the protein must be first translated into its corresponding DNA sequence. However, as it is known, an amino acid sequence can correspond to more than one DNA sequence, due to the degeneracy of the Genetic Code. As an example, consider the sequence CWY (C – cysteine, W – thryptofan, Y – tyrosine). The C amino acid has two corresponding codons: UGU and UGC, the amino acid W has only one corresponding codon: UGG and the amino acid Y has also two corresponding codons: UAU and UAC. This results in four possible translations of the sequence CWY, as shown below:

<div align="center">

C   W   Y

UGU UGG UAU

UGC UGG UAU

UGU UGG UAC

UGC UGG UAC

</div>

As can be inferred, the number of possible DNA sequences that can result from the translation of a specific protein sequence grows fast with the size of the protein and the number of codons associated to each amino acid. The total number of possible DNA sequences that corresponds to a specific protein sequence is given by the product of the number of distinct corresponding codons to each amino acid that composes the sequence[3].
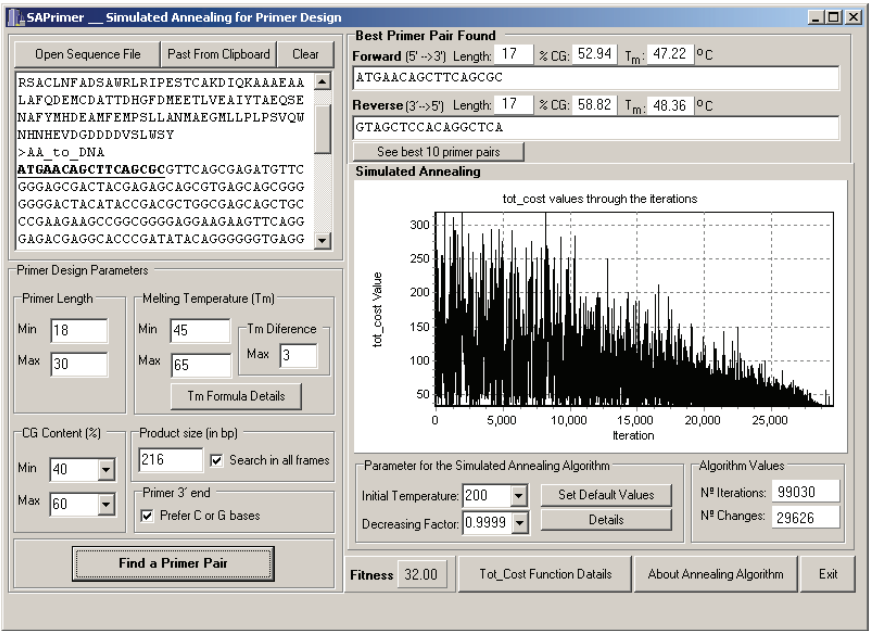

Fig. 6. Optimal primer pair found for a protein sequence given as input

---

[3] In the presented example, CWY has 2*1*2 = 4 possible corresponding DNA sequences.

Calculating all possible corresponding DNA sequence for a given protein sequence and finding the best primer pair for each one do not lead to a feasible solution to the problem because is highly time-consuming. The adopted strategy randomly chooses only one possible translation and finds the best primer pair for it. Notice that this strategy can produce different results, i.e., different primer pairs each time SAPrimer is run. So, the primers found for a particular SAPrimer run are called degenerated primers.

The DREB1A protein sequence (NCBI Reference Sequence: NP_567720.1) was used to exemplify the use of SAPrimer for proteins, and the results are shown in Fig. 6.

5. Conclusions

PCR is an important laboratorial process that can amplify a specific DNA sequence. Of crucial relevance for a successful PCR process is the identification of a suitable pair of primers. Due to the many variables and the range of values that can be assigned to them, the identification of a suitable pair of primes is a hard task, subject to a few uncertainties. A way to find out suitable primers is *via* a search process throughout the space defined by all possible pairs. By nature such spaces are quite vast preventing the use of exhaustive search methods. An efficient alternative is to use heuristic based methods such as simulated annealing.

This chapter describes the use of simulated annealing for primer design. Initially it details the relevant variables and their possible values as a first step to show how they can be combined into an objective function that "measures" the quality of a given primer pair. This function is used by the SA algorithm.

The chapter also describes the main aspects of two releases of an user-friendly software named SAPrimer, which implements the SA search, among other features. The SAPrimer software, as discussed in (Montera & Nicoletti, 2008) always finds the best possible primer pair to amplify a specific DNA sequence, even when some restrictions can not be satisfied, for example, when the given DNA sequence does not have an appropriated %CG or when the T_m value of a primer (forward or reverse) does not respect user-defined range of values. The second release, named SAPrimer (R2), is more flexible in the sense of offering to the user the possibility of finding primers to amplify any portion of the input sequence, defined by a fixed frame size and finding degenerated primer pair for a protein sequence.

The work will proceed by implementing other heuristic search strategies (starting with genetic algorithm) in an attempt to identify the most suitable type of search for dealing with the problem of primer design.

Acknowledgements

6. References

Abd-Elsalam, K.A. (2003). Bioinformatics tools and guidelines for PCR primer design. African Journal of Biotechnology, Vol. 2, No. 5, pp 91-95.

Acinas, S.G.; Sarma-Rupavtarm, R.; Klepac-Ceraj, V. & Polz, M.F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied Environmental Microbiology, Vol. 71, pp. 8966-8969.

Allawi, H.T. & SantaLucia, J.J. (1997) Thermodynamics and NMR of internal G*T mismatches in DNA. Biochemistry, Vol. 36, pp. 10581–10594.

Allawi, H.T. & SantaLucia, J.J. (1998a) Nearest neighbor thermodynamic parameters for internal G*A mismatches in DNA. Biochemistry, Vol. 37, pp. 2170–2179.

Allawi, H.T. & SantaLucia, J.J. (1998b) Nearest-neighbor thermodynamics of internal A*C mismatches in DNA: sequence dependence and pH effects. Biochemistry, Vol. 37, pp. 9435–9444.

Allawi, H.T. & SantaLucia, J.J. (1998c) Thermodynamics of internal C*T mismatches in DNA. Nucleic Acids Research, Vol. 26, pp. 2694–2701.

Antia, R.; Regoes, R.R; Koella, J.C. & Bergstrom, C.T. (2003). The role of evolution in the emergence of infectious diseases. Nature, Vol. 426, pp. 658-661.

Apte, A. & Daniel, S. (2009). PCR Primer Design. Cold Spring Harb Protoc.

Berezikov, E.; Cuppen, E. & Plasterk, R.H. (2006). Approaches to microRNA discovery. Nature Genetics, Vol. 38, S2-S7.

Bertagnolli, M.M.; Niedzwiecki, D.; Compton, C.C.; Hahn, H.P.; Hall, M.; Damas, B.; Jewell, S.D.; Mayer, R.J.; Goldberg, R.M.; Saltz, L.B.; Warren, R.S. & Redston, M. (2009). Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803. J. Clin Oncol., Vol. 27, pp. 1814-1821.

Bonafè, M.; Cardelli, M.; Marchegiani, F.; Cavallone, L.; Giovagnetti, S.; Olivieri, F.; Lisa, R.; Pieri, C. & Franceschi, C. (2001). Increase of homozygosity in centenarians revealed by a new inter-Alu PCR technique. Experimental Gerontology, Vol. 36, pp. 1063-1073.

Boutros, P.C. & Okey, A.B. (2004). PUNS: transcriptomic- and genomic-in silico PCR for enhanced primer design. Bioinformatics, Vol. 20, No. 15, pp. 2399–2400.

Boyce, R.; Chilana, P. & Rose, T.M. (2009). iCODEHOP: a new interactive program for designing COnsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. Nucleic Acids Research, Vol. 37, Web Server issue, W222-8.

Breslauer, K.J.; Frank, R.; Blocker, H. & Marky, L.A. (1986). Predicting DNA duplex stability from the base sequence, Proc. Natl. Acad. Sci. USA 83, pp. 3746-3750.

Cadwell, R.C. & Joyce, G.F. (1992). Randomization of genes by PCR. PCR Methods and Applications, Vol. 2, pp. 28-33.

Castellsagué, E.; González, S.; Nadal, M.; Campos, O.; Guinó, E.; Urioste, M.; Blanco, I.; Frebourg, T. & Capellá G. (2008). Detection of APC gene deletions using quantitative multiplex PCR of short fluorescent fragments. Clinical Chemistry, Vol. 54, pp. 1132-1140.

Chih-Ming, H.; Ming-Chieh, L.; Shih-Hung, H.; Chi-Jung, H.; Hung-Cheng, L.; Tsai-Yen, C. & Shwu-Fen, C. (2009). PTEN promoter methylation and LOH of 10q22–23 locus in PTEN expression of ovarian clear cell adenocarcinomas. Gynecologic Oncology, Vol. 112, pp. 307-313.

Chopra, S. & Ranganathan, A. (2003). Protein evolution by 'codon shuffling': a novel method for generating highly variant mutant libraries by assembly of hexamer DNA duplexes. Chem. Biol., Vol. 10, pp. 917-926.

Christensen, H.; Larsen, J. & Olsen, J. E. (2008). Bioinformatical design of oligonucleotides - design of PCR primers and hybridization probes including a summary of computer-programs. http://www.staff.kvl.dk/~hech/PrimerDesign.pdf

Coco, W.M.; Encell, L.P.; Levinson, W.E.; Crist, M.J.; Loomis, A.K.; Licato, L.L.; Arensdorf J.J.; Sica, N.; Pienkos, P.T. & Monticello, D.J. (2002). Growth factor engineering by degenerate homoduplex gene family recombination. Nature Biotechnology, Vol. 20, pp. 1246-1250

Contreras-Moreira, B.; Sachman-Ruiz, B.; Figueroa-Palacios, I. & Vinuesa, P. (2009). primers4clades: a web server that uses phylogenetic trees to design lineage-specific PCR primers for metagenomic and diversity studies. Nucleic Acids Research, Vol. 37, Web Server issue, W95-W100.

Costas, B.A.; McManus, G.; Doherty, M. & Katz, L.A. (2007). Use of species-specific primers and PCR to measure the distributions of planktonic ciliates in coastal waters. Limnol. Oceanogr. Methods, Vol. 5, pp. 163-173.

Dieffenbach, C.W. & Dveksler, G.S. (2003). PCR Primer: A Laboratory Manual, (2nd. ed.), Cold Spring Harbor Laboratory Press, New York.

Dixon, A.L.; Liang, L.; Moffatt, M.F.; Chen, W.; Heath, S.; Wong, K.C.; Taylor, J.; Burnett, E.; Gut, I.; Farrall, M.; Lathrop, G.M.; Abecasis, G.R. & Cookson, W.O. (2007). A genome-wide association study of global gene expression. Nature Genetics, Vol. 39, No. 10, pp. 1202-1207.

Evans, P. M. & Liu, C. (2005). SiteFind: a software tool for introducing a restriction site as a marker for successful site-directed mutagenesis. BMC Mol. Biol., Vol. 6, No. 22,

Eveno, C.; Nemeth, J.; Soliman, H.; Praz, F.; de The, H.; Valleur, P.; Talbot, Ian-C.& Pocard, M. (2010). Association between a high number of isolated lymph nodes in T1 to T4 N0M0 colorectal cancer and the microsatellite instability phenotype. Arch. Surg., Vol. 145, No. 1, pp. 12-17.

Franko, J.; Krasinskas, A.M.; Nikiforova, M.N.; Zarnescu, N.O.; Lee, K.K.W.; Hughes, S.J.; Bartlett, D.L.; Zeh III, H.J. & Mose, A.J. (2008). Loss of heterozygosity predicts poor survival after resection of pancreatic adenocarcinoma. J. Gastrointest. Surg., Vol. 12, pp. 1664-1673.

Frech, C.; Breuer, K.; Ronacher, B.; Kern, T.; Sohn, C. & Gebauer, G. (2009). hybseek: Pathogen primer design tool for diagnostic multi-analyte assays. Computer Methods and Programs in Biomedicine, Vol. 94, Issue 2, pp. 152-160.

Gatto, L. & Schretter, C. (2009). Designing Primer Pairs and Oligos with OligoFaktorySE. journal.embnet.org. Technical Notes. EMBnet.news 15.3, pp. 22-24.

Giegerich, R.; Meyer, F. & Schleiermacher, C. (1996). GeneFisher - software support for the detection of postulated genes. Proc. Int. Conf. Intell. Syst. Mol. Biol., Vol. 4, pp. 68-77.

Gordon, P.M.K. & Sensen, C.W. (2004). Osprey: a comprehensive tool employing novel methods for the design of oligonucleotides for DNA sequencing and microarrays. Nucleic Acids Research, Vol. 32, No. 17, e133.

Gunnar, W.; Kokocinsky, F. & Lichter, P. (2004). AutoPrime: selecting primers for expressed sequences. Genome Biology, Vol. 5, P11.

Haas, S.A.; Hild, M.; Wright, A.P.H.; Hain, T.; Talibi, D. & Vingron, M. (2003). Genome-scale design of PCR primers and long oligomers for DNA microarrays. Nucleic Acids Research, Vol. 31, No. 19, pp. 5576-5581.

Hassibi, A.; Kakavand, H. & Lee, T.H. (2004). A stochastic model and simulation algorithm for polymerase chain reaction (PCR) systems. Proc. of Workshop on Genomics Signal Processing and Statistics (GENSIPS).

He, Q.; Marjamäki, M.; Soini, H.; Mertsola, J. & Viljanen, M.K. (1994). Primers are decisive for sensitivity of PCR. BioTechniques, Vol. 17, No. 1, pp. 82-87.

Howley, P.M.; Israel, M.F.; Law, M-F. & Martin, M.A. (1979). A rapid method for detecting and mapping homology between heterologous DNAs. Journal of Biological Chemistry, Vol. 254, pp. 4876-4883.

Hugles, S. & Moody, A. (2008). PCR. Scion Publishing. Oxfordshire, England.

Innis, M.A. & Gelfand, D.H. (1990). Optimization of PCRs. In: PCR Protocols (Innis, Gelfand, Sninsky and White, eds.), Academic Press, New York.

Isenbarger, T.A.; Finney, M.; Ríos Velázquez, C.; Handelsman, J. & Ruvkun, G. (2008). Miniprimer PCR, a new lens for viewing the microbial world. Applied Environmental Microbiology, Vol. 74, No. 3, pp. 840-849.

Kalendar, R.; Lee, D. & Schulman, A.H. (2009) FastPCR Software for PCR Primer and Probe Design and Repeat Search. Genes, Genomes and Genomics. Vol. 3, No. 1, pp. 1-14.

Kämpke, T.; Kieninger, M. & Mecklenbug, M. (2001). Efficient primer design algorithms. Bioinformatics, Vol. 17, No. 3, pp. 214-225.

Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). J. Biosci. Bioeng., Vol. 96, pp. 317-323.

Ke, X; Collins, A.R. & Ye, S. (2002). PCR designer for restriction analysis of various types of sequence mutation. Bioinformatics, Vol. 18, No. 12, pp. 1688-1689.

Kirkpatrick, S.; Gelatt, C.D. & Vecchi, M.P. (1983). Optimization by simulated annealing. Science, Vol. 220, pp. 671-680.

Kolb, L.E.; Arlier, Z.; Yalcinkaya, C.; Ozturk, A.K.; Moliterno, J.A.; Erturk, O.; Bayrakli, F.; Korkmaz, B.; DiLuna, M.L.; Yasuno, K.; Bilguvar, K.; Ozcelik, T.; Tuysuz, B.; State, M.W. & Gunel, M. (2010). Novel VLDLR microdeletion identified in two Turkish siblings with pachygyria and pontocerebellar atrophy. Neurogenetics, Jan. 15, pp. 1364-6745.

Kwok, S.; Kellogg, D.E.; McKinney, N.; Spasic, D.; Goda, L.; Levenson, C. & Sninsky, J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus 1 model studies. Nucleic Acids Research, Vol. 18, pp. 999-1005.

Lahr, D.J.G. & Katz, L.A. (2009). Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. BioTechniques, Vol. 47, pp. 857-866.

Lan, C.-C.; Tang, R.; Un San Leong, I & Love, D. R. (2009). Quantitative Real-Time RT-PCR (qRT-PCR) of Zebrafish Transcripts: Optimization of RNA Extraction, Quality Control Considerations, and Data Analysis. Cold Spring Harb Protoc., Vol. 10, pp. pdb.prot5314 - pdb.prot5314.

Liang, H.-L.; Lee, C. & Wu, J.-S. (2005). Multiplex PCR primer design for gene family using genetic algorithm, GECCO'05, June 25-29, Washington DC, USA. pp. 67-74.
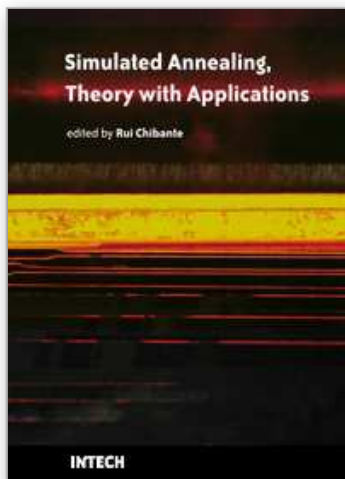
Lutz, S. & Patrick, W.M. (2004). Novel methods for directed evolution of enzymes: quality, not quantity. Current Opinion in Biotechnology, Vol. 15, pp. 291-297.

Mann T.; Humbert, R.; Dorschner, M.; Stamatoyannopoulos, J. & Noble, W.S. (2009). A thermodynamic approach to PCR primer design. Nucleic Acids Research, Vol. 37, No. 13, e95.

Marmur, J. & Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. Journal of Molecular Biology, Vol. 5, pp. 109-118.

Metzker, M.L. & Caskey, T.C. (2001). Polymerase Chain Reaction (PCR), Encyclopedia of Life Science. Nature Publishing Group.

Milbury, C.A.; Li, J. & Makrigiorogs, G.M. (2009). PCR-based methods for the enrichment of minority alleles and mutations. Clinical Chemistry, Vol. 55, No. 4, pp. 632-640.

Montera, L. & Nicoletti, M.C. (2008). The PCR primer design as a metaheuristic search process. Lecture Notes in Artificial Intelligence, Vol. 5097, pp. 963-973.

Montera. L.; Nicoletti, M.C. & Silva, F.H. (2006). Computer assisted parental sequence analysis as a previous step to DNA Shuffling process. Conference on Evolutionary Computation (CEC 2006) in IEEE Congress on Computational Intelligence - Vancouver, Canada, pp. 8079-8086.

Morales, S.E. & Holben, W.E. (2009). Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by in silico analysis. Applied and Environmental Microbiology, Vol. 75, No. 9, pp. 2677–2683.

Mullis, K.B. & Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymology, Vol. 155, pp. 335-350.

Nuovo, G.J.; Gallery, F.; Hom, R.; MacConnell, P. & Bloch, W (1993). Importance of different variables for enhancing in situ detection of PCR-amplified DNA. Genome Research, Vol. 2, pp. 305-312.

Oliveira, E.J.; Pádua, J.G.; Zucchi, M.I.; Vencovsky, R. & Vieira, M.L.C. (2006). Origin, evolution and genome distribution of microsatellites. Genetics and Molecular Biology, Vol. 29, No. 2, pp. 294-307.

Panjkovich, A. & Melo, F. (2005). Comparison of different melting temperature calculation methods for short DNA sequences. Bioinformatics, Vol. 21, No. 6, pp. 711-722.

Pasmant, E.; Sabbagh, A.; Masliah-Planchon, J.; Haddad, V.; Hamel, M-J.; Laurendeau, I.; Soulier, J.; Parfait, B.; Wolkenstein, P.; Bièche, I.; Vidaud, M. & Vidaud D. (2009). Detection and characterization of NF1 microdeletions by custom high resolution array CGH. Journal of Molecular Diagnostics, Vol. 11, No. 6, pp. 524-529.

Patten, P.A.; Gray, N.S.; Yang, P.L.; Marks, C.B.; Wedemayer, G.J.; Boniface, J.J.; Stevens, R.C. & Schultz, P.G. (1996). The immunological evolution of catalisis. Science, Vol. 271. No. 5252, pp. 1086-1091.

Pattyn, F.; Speleman, F.; Paepe, A. & Vandesompele, J. (2003). RTPrimerDB: the Real-Time PCR primer and probe database. Nucleic Acids Research. Vol. 31, No. 1, pp. 122-123.

Piriyapongsa, J.; Ngamphiw, C.; Assawamakin, A.; Wangkumhang, P.; Suwannasri, P.; Ruangrit, U.; Agavatpanitch, G. & Tongsima, S. (2009). RExPrimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP 3'-in-primer and mis-priming from structural variation. BMC Genomics, 10(Suppl 3):S4.

Pusch, C.M.; Broghammer, M.; Nicholson, G.J.; Nerlich, A.G.; Zink, A.; Kennerknecht, I.; Bachmann, L. & Blin, L. (2004). PCR-induced sequence alterations hamper the typing of prehistoric bone samples for diagnostic achondroplasia mutations. Mol. Biol. Evol., Vol. 21, No. 11, pp. 2005–2011.

Ro, S.; Park, C.; Jin, J.; Sanders, K.M. & Yan, W. (2006) A PCR-based method for detection and quantification of small RNAs. Biochem. Biophys. Res. Commun., Vol. 315, No. 3, pp. 756-763.

Rose, T.M.; Henikoff, J.G. & Henikoff, S. (2003). CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. Nucleic Acids Research, Vol. 31, No. 13, pp. 3763-3766.

Roux, K. H. (2009). Optimization and Troubleshooting in PCR. Cold Spring Harb Protoc., Vol. 4, pp. pdb.ip66 – pdb.ip66.

Rozen, S. & Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365-386.

Rychlik, W.; Spencer, W.J. & Rhoads, R.E. (1990). Optimization of the annealing temperature for DNA amplification in vitro. Nucleic Acids Research, Vol. 18, pp. 6409-6412.

Saelee, P.; Wongkham, S.; Bhudhisawasdi, V.; Sripa, B.; Chariyalertsak, S. & Petmitr, S. (2008). Allelic loss on chromosome 5q34 is associated with poor prognosis in hepatocellular carcinoma. Cancer Research Clinical Oncology, Vol. 134, No. 10, pp. 1135-1141.

Sambrook. J. & Russel, D.W. (2001). Molecular Cloning: A Laboratory Manual (3rd ed.). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

SantaLucia, J.Jr.; Allawi, H.T. & Seneviratne, P.A. (1996). Improved Nearest-Neighbor parameters for predicting DNA duplex stability. Biochemistry, Vol. 35, pp. 3555-3562.

Sasaki, T.; Tsubakishita, S.; Tanaka, Y.; Sakusabe, A.; Ohtsuka, M.; Hirotaki, S.; Kawakami, T.; Fukata, T. & Hiramatsu, K. (2010). Multiplex-PCR method for species identification of coagulase-positive staphylococci. Journal of Clinical Microbiology, Vol. 48, No. 3, pp. 765-769.

Schretter, C.; & Milinkovitch, M.C. (2006) OLIGOFAKTORY: a visual tool for interactive oligonucleotide design. Bioinformatics, Vol. 22, No. 1, pp. 115–116.

Sommer, R. & Tautz, D. (1989). Minimal homology requirements for PCR primers. Nucleic Acids Research. Vol. 17, No.16, pp. 6749.

Srivastava, T.; Seth, A.; Datta, K.; Chosdol K.; Chattopadhyay, P. & Sinha, S. (2005). Inter-alu PCR detects high frequency of genetic alterations in glioma cells exposed to sub-lethal cisplatin. Int. J. Cancer, Vol. 117, No. 4, pp. 683-689.

Stemmer, W.P.C. (1994a). DNA Shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. Proc. Natl. Acad. Sci., Vol 9, pp. 10747-10751.

Stemmer, W.P.C. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. Nature, Vol. 370, pp. 389–391.

Sugimoto, N.; Nakano, S.; Yoneyama, M. & Honda, K. (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Research, Vol. 24, pp. 4501-4505.

Sun, F. (1999). Modeling DNA Shuffling. J. Comp. Biol., Vol. 6, pp. 77-90.

Tait, R.C. & Horton, R.M. (1998). Genetic engineering with PCR. Horizon Scientific Press, Norwich.

Tsai, M.-F.; Lin, Y.-J.; Cheng, Y.-C.; Lee, K.-H.; Huang, C.-C.; Chen, Y.-T. & Yao, A. (2007). PrimerZ: streamlined primer design for promoters, exons and human SNPs. Nucleic Acids Research. pp. (Web Server issue): W63–W65.

Vandesompele, J.; Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; Paepe, A. & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology, Vol. 3, pp. research0034.1-0034.11.

VanGuilder, H.D.; Vrana, K.E. & Freeman, W.M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques, Vol. 44. No. 5, pp. 619-626.

Vikalo, H.; Hassibi, B. & Hassibi, A. (2006). On Joint Maximum-Likelihood Estimation of PCR Efficiency and Initial Amount of Target. Proc. IEEE Workshop on Genomic Signal Processing and Statistics (GENSIPS), pp. 688-695.

Vladušić, T.; Hrašćan, R.; Vrhovac, I.; Krušlin, B.; Gamulin, M.; Grgić, M.; Pećina-Šlaus, N. & Čolić, J.F. (2010). Loss of heterozygosity of selected tumor suppressor genes in human testicular germ cell tumors. Pathology – Research and Practice, Vol. 206, pp. 163-167.

Volkov, A.A.; Shao, Z. & Arnold, F. H. (1999). Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair. Nucleic Acids Research, Vol. 27, No. 18, e18.

Wallace, R.B.; Shaffer, J.; Murphy, R.F.; Bonner, J.; Hirose, T. & Itakura, K. (1979). Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. Nucleic Acids Research, Vol. 6, pp. 3543-3557.

Wang, R.; Morris, D.S.; Tomlins, S.A.; Lonigro, R.J.; Tsodikov, A.; Mehra, R.; Giordano, T.J.; Kunju, L.P.; Lee, C.T.; Weizer A.Z. & Chinnaiyan A.M. (2009). Development of a multiplex quantitative PCR signature to predict progression in non–muscle-invasive bladder cancer. Cancer Research, Vol. 69, pp. 3810-3818.

You, F.M.; Huo, N.; Gu, Y.Q.; Luo, M.C.; Ma, Y.; Hane, D.; Lazo G.R.; Dvorak, J. & Anderson, O.D. (2008). BatchPrimer3: a high throughput Web application for PCR and sequencing primer design. BMC Bioinformatics, Vol. 9, pp. 253.

Yu, W.; Rusterholtz, K.J.; Krummel, A.T. & Lehman, N. (2006). Detection of high levels of recombination generated during PCR amplification of RNA templates. BioTechniques, Vol. 40, pp. 499-507.

Zha, D.; Eipper, A. & Reetz, M.T. (2003). Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. ChemBioChem, Vol. 4, pp. 34–39.

Zhao, H.; Giver, L.; Shao, Z.; Affholter, A. & Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. Nature Biotechnology, Vol. 16, pp. 258–261.

Simulated Annealing, Theory with Applications

Edited by Rui Chibante

The book contains 15 chapters presenting recent contributions of top researchers working with Simulated Annealing (SA). Although it represents a small sample of the research activity on SA, the book will certainly serve as a valuable tool for researchers interested in getting involved in this multidisciplinary field. In fact, one of the salient features is that the book is highly multidisciplinary in terms of application areas since it assembles experts from the fields of Biology, Telecommunications, Geology, Electronics and Medicine.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds