PROTOCOL

Selection of Primers for Polymerase Chain Reaction

Wojciech Rychlik

Abstract

One of the most important factors affecting the quality of PCR is the choice of primers. In general, the longer the PCR product the more difficult it is to select efficient primers and set appropriate designing primers, and in general, the more DNA sequence information is available, the better the chance of finding an optimal primer pair. Efficient primers can be designed by avoiding the following flaws: primer-dimer formation, self-complementarity, too low T_m of the primers, and/or their incorrect internal stability profile. Tips on subcloning PCR products, calculating duplex stability (predicting dimer formation strength), and designing degenerate primers are given.

Index Entries: PCR; primer design; primer internal stability.

1. Introduction

One of the most important factors affecting the quality of polymerase chain reaction (PCR) is the choice of primers. Several rules should be observed when designing primers and, in general, the more DNA sequence information available, the better the chance of finding an "ideal" primer pair. Fortunately, not all primer selection criteria need be met in order to synthesize a clean, specific product, since the adjustment of PCR conditions (such as composition of the reaction mixture, temperature, and duration of PCR steps) may considerably improve the reaction specificity. Amplification of 200–400-bp DNA is the most efficient and, in these cases, one may design efficient primers simply by following a few simple rules described in this chapter. It is more difficult to choose primers for efficient amplification of longer DNA fragments, and use of an appropriate primer analysis software is worthwhile.

The important parameters to be considered when selecting PCR primers are the ablility of the primer to form a stable duplex with the specific site on the target DNA, and no duplex formation with another primer molecule or no hybridization at any other target site. The primer stability can be measured in the length (base pairs) of a DNA duplex, the GC/AT ratio, kcal/mol (duplex formation free energy), or in °C (melting temperature). The most accurate methods for computing helix stability are based on nearest neighbor thermodynamic parameters *(1)*. Calculation of T_m according to the nearest neighbor method is complicated, and therefore not practical for use without computer software. Similar duplex stability accuracy, however, may be achieved by calculating the free energy of duplex formation (ΔG). This calculation is simple and can be performed manually.

The Methods section describes the following: an example of ΔG calculation, needed for accurate determination of duplex stability; general rules for PCR primer selection; primer design based on a peptide sequence; and primer design for subcloning PCR products.

*Author to whom all correspondence and reprint requests should be addressed. National Biosciences Inc., 3650 Annapolis Lane North, #140, Plymouth, MN 55447-5434.

2. Methods

2.1. Calculations of DNA Duplex Stability

The method of predicting free energy of duplex formation (ΔG) for DNA oligomers, described in the following, is a simplified method of Breslauer et al. *(1)*. It is based on the equation

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

where ΔH and ΔS are the enthalpy and entropy of duplex formation, respectively, and T is the temperature in K. Table 1 lists the ΔG values of nucleotide pairs.

For simplicity, all calculations are made with T set to 298.15 K (25°C). The relative stability of base-pairing in a duplex is dependent on the neighboring bases *(1)*. Thus, for example, to calculate the ΔG of the d(ACGG/CCGT) duplex formation, add the ΔG values of the three nucleotide pairs as follows:

$$\Delta G\ (ACGG) = \Delta G\ (AC) + \Delta G\ (CG) + \Delta G\ (GG) \quad (2)$$

$$\Delta G\ (ACGG) = -(1.3 + 3.6 + 3.1) = -8.0\ (kcal/mol) \quad (3)$$

This method is especially useful for determination of primer compatibility owing to formation of 3'-terminal duplexes, discussed in the following section. Use the same approach when calculating the ΔG of a hairpin loop structure, except that the ΔG increment for loop must be added. For loops, sizes 3–8 nucleotide, I use the following values (averaged from refs. *2* and *3*): 3 nucleotide, 5.2 kcal/mol; 4 nucleotide, 4.5; 5 nucleotide, 4.4; 6 nucleotide, 4.3; 7 and 8 nucleotide, 4.1 kcal/mol. More data can be found in ref. *2*.

2.2. Selection of PCR Primers

2.2.1. General Rules

2.2.1.1. Dimer Formation

PCR primers should be free of significant complementarity at their 3' termini as this promotes the formation of primer–dimer artifacts that reduce product yield. Formation of primer–dimer artifacts may also cause more serious problems,

Table 1
Free Energy Values of a Nearest Neighbor Nucleotide[a]

First (5') nucleotide	Second nucleotide dA	dC	dG	dT
	ΔG (kcal/mol)			
dA	−1.9	−1.3	−1.6	−1.5
dC	−1.9	−3.1	−3.6	−1.6
dG	−1.6	−3.1	−3.1	−1.3
dT	−1.0	−1.6	−1.9	−1.9

[a]Calculated according to Eq. (1) in 25°C.

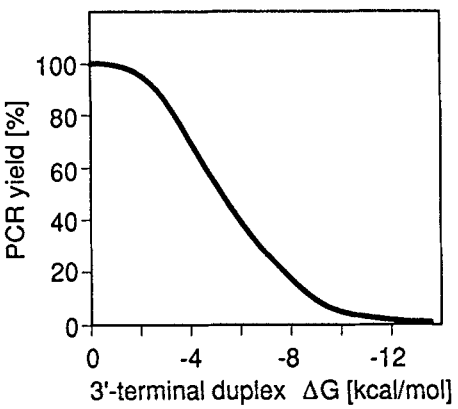

Fig. 1. Dependence of PCR yield on the ΔG of 3'-terminal primer duplexes. The ΔG values were calculated as described in Section 2.1.

such as nonspecific DNA synthesis owing to an unbalanced primer ratio (asymmetric PCRs fail more frequently than "standard" reactions). Figure 1 illustrates the PCR yield dependence on the ΔG of 3'-terminal duplexes.

These values are approximate, since the yield also depends on the annealing temperature, the specificity of primers, and other parameters not considered here. The high dependence of yield on dimer formation tendency is the result of the very high processivity of *Taq* polymerase. Duplexes need not be stable to prime DNA synthesis. Very little time is required for the enzyme to recognize a 3'-terminal duplex and start polymerization.

2.2.1.2. Self-Complementarity

In general, oligonucleotides forming intramolecular duplexes with negative ΔG should be

avoided. Although self-complementary PCR primers with hairpin loop ΔG approaching –3 kcal/mol (at 25°) are suitable in certain cases, a hairpin loop forming primer is troublesome when its 3' end is "tied up," since this can cause internal primer extension, thus eliminating a given primer from the reaction. Hairpins near the 5' end, however, do not significantly affect the PCR.

2.2.1.3. Melting Temperature: Stability

There is a widely held assumption that PCR primers should have about a 50% GC/AT ratio. This is not correct. An 81% AT-rich primer (with a second primer of a similar composition and human genomic DNA as substrates) produced a single, specific, 250-bp PCR product (70% AT-rich). Without getting into the complex calculations of product and primer T_m values, PCR primers should have a GC/AT ratio similar to or higher than that of the amplified template.

A more important factor is the T_m difference between the template and the less stable primer. PCR is efficient if this difference is minimized. Note that the T_m of DNA also depends on its length. This is the reason why researchers typically design primers that are too long and unnecessarily too stable. Longer oligos, however, are less likely to be suitable in terms of dimer formation and self-complementarity and, therefore, generally scarce in a given sequence. If the expected PCR product is ≤500 bp, select short (16–18 nucleotide) primers. For the synthesis of a 5-kb fragment, choose about 24-mers. Recently, products of over 40 kb in length have been amplified using 20–23-mers *(4)* with T_m of approx 68° or higher *(5)*. When working with longer primers, however, it is difficult to choose a compatible primer pair without the aid of primer selection software to check dimer formation, self-complementarity, and the specificity of primers. When amplifying a long DNA fragment, there is a good chance that an oligonucleotide selected "by eye" will prime from other than the intended target site, yielding nonspecific product(s). The likelihood of false priming can be significantly reduced by observing the internal stability rule, as described in the following.

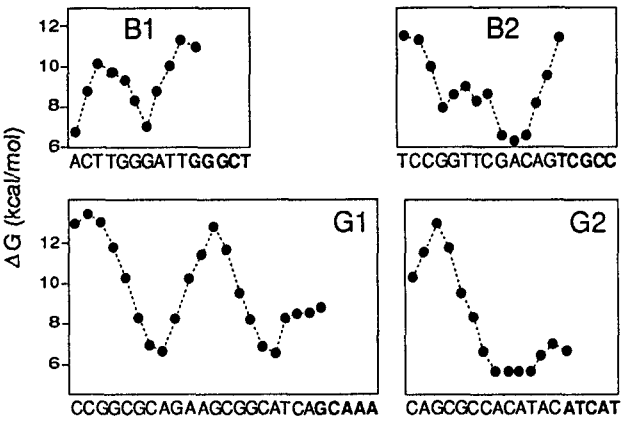

Fig. 2. Internal stability of two poorly functioning (B1, B2) and two efficient (G1, G2) sequencing primers. Primer G1 and G2 performed above average (with almost any other compatible primer) in PCR. The ΔG values were calculated for all pentamers in each primer. The last symbol in each inset represents the ΔG value of the subsequence written in bold (the 3'-terminal pentamer).

2.2.1.4. Internal Stability

Primers that are stable at their 5' termini but somewhat unstable on their 3' ends perform best in sequencing and PCR as well. This primer structure effectively eliminates false priming. These recent findings, based on primer internal stability, are supported by the experimental data presented in Fig. 2. A primer with low stability on its 3' end will function well in PCR because the base pairings near and at the 3' end with nontarget sites are not sufficiently stable to initiate synthesis (false priming). Therefore, the 5' and central parts of the primer must also form a duplex with the target DNA site in order to prime efficiently. Conversely, oligonucleotides with stable, GC-rich, 3' termini need not anneal with the target along their entire length in order to efficiently prime, resulting often in nonspecific product synthesis. Examples of efficient PCR (and sequencing) primers are presented in Fig. 2 (primers G1 and G2).

Notice the high 3'-end stability of nonspecific primers (B1 and B2) and low stability of specific primers. The optimal annealing temperature range is unusually broad when primers exhibiting low 3'-terminal stability are used. This improves the

chances of running the PCR at optimal conditions without preliminary optimization experiments. It is worth noting that the quality of the PCR product depends on the template (substrate complexity, product length, and T_m), as well as on the annealing time and temperature *(6)*. In certain conditions, primers with high 3' terminal stability perform satisfactory in PCR. Nevertheless, oligonucleotides with 3' terminal pentamers less stable than –9 kcal/mol (check Section 2.1. for calculations) are more likely to be specific primers.

2.2.1.5. Unique Primers

In order to amplify a single, specific DNA fragment, the primer's sequence should not repeat in the template *(7)*. Although it is highly unlikely that the entire primer matches perfectly at more than one site on the template, primers with 6–7 nucleotide-long nonunique 3' termini are not uncommon. This may create problems when a "false" priming site is located inside the amplified region. In these cases, a nonspecific product formation is observed (especially in later cycles), because the PCR of shorter DNA fragments is usually more efficient. Note that the more unstable the primer's 3' end, the lower the likelihood of false priming *(see* Section 2.2.1.4.). When working with mammalian genomic sequences, it is helpful to check the primer of interest for complementarity with *Alu* sequences or with other short repetitive elements. For a similar reason, homooligomers (like –AAAAAA–) and dinucleotide repeats (like –ATATAT–) should rather be avoided.

2.2.2. Specific Applications

2.2.2.1. Primer Design Based on Peptide Sequences

When designing primers from peptide sequences, the use of degenerate primers rather than "guessmers" is preferred. Although it has been reported that up to 1024-degenerate primers have been used successfully *(8)*, regions of high degeneracy should be avoided. There are many (unreported) cases in which less degenerate primers have not worked. It is generally assumed that PCR is acceptably efficient when using primers with 15–20% bp mismatches with the template. Mismatches at a primer's 3' end, however, cause more serious problems than the same mismatch ratio at the 5' end. The PCR yield using a primer with two mismatches within the last four bases is drastically reduced. Studies of Kwok et al. *(9)* indicate, however, that primers with 3'-terminal "T"-mismatches can be efficiently utilized by *Taq* polymerase when the nucleotide concentration is high. At 0.8 m*M*, most 3'-end mismatches are acceptable *(9)*, although nonspecific product formation is high, and the fidelity DNA synthesis is reduced *(10)*. There is a low level of priming from mismatched bases even at low nucleotide concentrations *(11)*, and therefore, increasing the annealing time to 3–5 min in the initial PCR cycles may yield a desired product of a better quality than when using standard annealing times and high dNTP concentrations. A total nucleotide concentration of 0.2 m*M*, or below, is recommended when unique primers are used, since high concentrations increase the misincorporation rate *(10,12)*. When degenerate oligonucleotides are used, PCRs should be run at higher primer concentrations (1–3 μ*M* instead of 0.2 μ*M*) because most oligos in the mixture will not prime specifically and only contribute to high background. More information on optimizing the reaction mixture and the use of degenerate primers can be found in Chapters 30 and 31 of ref. *13*.

2.2.2.2. Primer Design for Subcloning

The addition of a (mismatched) restriction site at the 5' terminus is the most useful method. Add a few "dummy" 5'-terminal bases beyond the recognition site, so that the restriction endonucleases can cut the DNA. Try not to extend a potential dimer structure (inherent to restriction sites) beyond the recognition site. There are no general rules as to how many nucleotides to add. A list of cleavage efficiencies of short oligonucleotides has been published *(14)*; the summary is listed in Table 2.

An alternative to incorporating a full restriction enzyme recognition site is to use oligonucleotide

Table 2
Cleavable Efficiencies of Short DNA Fragments

Enzyme	Excess bp[a]	%Cleavage after 2 h	%Cleavage after 20 h	Enzyme	Excess bp[a]	%Cleavage after 2 h	%Cleavage after 20 h
Acc I	3	0	0	*Not* I	8	25	90
Afl III	1	0	0		10	25	>90
	2	>90	>90	*Nsi* I	3	10	>90
Asc I	1	>90	>90	*Pst* I	1	0	0
BamH I	1	10	25		4	10	10
	2	>90	>90	*Pvu* I	1	0	0
Bgl II	1	0	0		2	10	25
	2	75	>90		3	0	10
	3	25	>90	*Sac* I	1	10	10
BssH II	2	0	0	*Sac* II	1	0	0
	3	50	>90		3	50	90
BstE I	1	0	10	*Sca* I	1	10	25
Cla I	1	0	0		3	75	75
	2	>90	>90	*Sma* I	0	0	10
	3	50	50		1	0	10
EcoR I	1	>90	>90		2	10	50
Hae III	1	>90	>90		3	>90	>90
Hind III	2	0	0	*Spe* I	1	10	>90
	3	10	75		2	10	>90
Kpn I	1	0	0		3	0	50
	2	>90	>90		4	0	50
Mlu I	1	0	0	*Sph* I	1	0	0
	2	25	50		3	0	25
Nco I	1	0	0		4	10	50
	4	50	75	*Stu* I	1	>90	>90
Nde I	1	0	0	*Xba* I	1	0	0
Nhe I	1	0	0		2	>90	>90
	2	10	25		3	75	>90
	3	10	50		4	75	>90
Not I	2	0	0	*Xho* I	1	0	0
	4	10	10		2	10	25
	6	10	10		3	10	75

[a]Number of base pairs added on each side of the recognition sequences.

primers with only half a palindromic recognition site at the 5' termini of each phosphorylated primer. After amplification, the PCR product should be concatamerized with ligase and then digested with the appropriate enzyme *(15)*. This is an efficient method, actually forcing a researcher to use high fidelity synthesis conditions *(10,16)*, i.e., low nucleotide concentration, low number of cycles, short extension times, and no "final extension." In these conditions, the formation of 3' overhangs, preventing efficient ligation, is minimal.

If the amplified product is to be subcloned, and the restriction site not needed, use unphosphorylated primers for the reaction and then ligate the product with a *Sma* I-digested vector in the presence of low concentrations of *Sma* I (a blunt-end cutter compatible with the ligation conditions). Again, high fidelity PCR conditions should be used, as mentioned earlier, to minimize formation of 3' overhangs.

When high fidelity synthesis is less essential, one may utilize the template-independent activity

of *Taq* polymerase to create 3'-"A" overhangs in the PCR product and use a vector with 3'-"T"-overhangs *(17,18)*. This method is very efficient when high concentration of nucleotides and long extension times are used, followed by prolonged incubation at the extension temperature after the last cycle.

References

1. Breslauer, K. J., Frank, R., Blocker, H., and Markey, L. A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA* **83**, 3746–3750.
2. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* **83**, 9373–9377.
3. Groebe, D. R. and Uhlenbeck, O. C. (1988) Characterization of RNA hairpin loop stability. *Nucleic Acids Res.* **16**, 11,725–11,735.
4. Cheng, S., Fockler, C., Barnes, W. M., and Russell, H. (1994) Effective amplification of long targets from clones inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**, 5696–5699.
5. Wu, D. Y., Ugozzoli, L., Pal, B. K., Qian, J., and Wallace, R. B. (1991) The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. *DNA Cell Biol.*, **10**, 233–238.
6. Rychlik, W., Spencer, W. J., and Rhoads, R. E. (1990) Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **18**, 6409–6412.
7. Rychlik, W. and Rhoads, R. E. (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.* **17**, 8543–8551.
8. Lee, C. C. and Caskey, C. T. (1990) cDNA cloning using degenerate primers, in *PCR Protocols* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J. eds.), Academic, New York, pp. 46–53.
9. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C., and Sninsky, J. J. (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* **18**, 999–1005.
10. Eckert, K. A. and Kunkel, T. A. (1990) High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res.* **18**, 3739–3744.
11. Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. C., Cheong, C., and Tinoco, I., Jr. (1988) Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proc. Natl. Acad. Sci. USA* **85**, 6252–6256.
12. Kawasaki, E. (1990) Amplification of RNA, in *PCR Protocols* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J. eds.), Academic, New York, pp. 21–27.
13. White, B. A., ed. (1993) *Methods in Molecular Biology, vol. 15: PCR Protocols,* Humana, Totowa, NJ.
14. New England BioLabs, 1993–1994 Catalog, "Cleavage close to the end of DNA fragments," p. 180.
15. Jung, V., Pestka, S. B., and Pestka, S. (1990) Efficient cloning of PCR generated DNA containing terminal restriction endonuclease recognition sites. *Nucleic Acids Res.* **18**, 6156.
16. Eckert, K. A. and Kunkel, T. A. (1991) The fidelity of DNA polymerase used in PCR, in *Polymerase Chain Reaction: A Practical Approach* (McPherson, M. J., Quirke, P., and Taylor, G.R. eds.), IRL, Oxford, UK, pp. 227–246.
17. Marchuk, D., Drumm, M., Saulino, A., and Collins, F. S. (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products *Nucleic Acids Res.* **19**, 1154.
18. Holton, T. A. and Graham, M. W. (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors *Nucleic Acids Res.* **19**, 1156.