Primer3 on the WWW for General Users and for Biologist Programmers

Steve Rozen and Helen Skaletsky

1. Introduction

Designing PCR and sequencing primers are essential activities for molecular biologists around the world. This chapter assumes acquaintance with the principles and practice of PCR, as outlined in, for example, refs. 1–4.

Primer3 is a computer program that suggests PCR primers for a variety of applications, for example to create STSs (sequence tagged sites) for radiation hybrid mapping (5), or to amplify sequences for single nucleotide polymorphism discovery (6). Primer3 can also select single primers for sequencing reactions and can design oligonucleotide hybridization probes.

In selecting oligos for primers or hybridization probes, Primer3 can consider many factors. These include oligo melting temperature, length, GC content, 3’ stability, estimated secondary structure, the likelihood of annealing to or amplifying undesirable sequences (for example interspersed repeats), the likelihood of primer–dimer formation between two copies of the same primer, and the accuracy of the source sequence. In the design of primer pairs Primer3 can consider product size and melting temperature, the likelihood of primer–dimer formation between the two primers in the pair, the difference between primer melting temperatures, and primer location relative to particular regions of interest or to be avoided.

1.1. Primer3 Can Be Used Through its WWW Interface or as a Software Component

Most casual users will prefer Primer3’s WWW interface (Fig. 1), which is suitable for selecting primers from a few sequences. Subheading 2. discusses this interface in detail.

From: Methods in Molecular Biology, vol. 132: Bioinformatics Methods and Protocols
Edited by: S. Misener and S. A. Krawetz © Humana Press Inc., Totowa, NJ
365
Scientists who must select primers for hundreds or thousands of sequences will prefer to use Primer3 (specifically the primer3_core program) as a software component, which accepts input in a format convenient for other programs to produce and generates output in a format convenient for other programs to interpret. (We assume that no one would want to deal manually with primer picking results for hundreds or thousands of sequences.) We present examples of the use of primer3_core as a software component in Subheading 3. The underlying primer design process is identical for both the WWW interface and primer3_core, and in fact the WWW interface uses the WWW CGI protocol (9,10) layered on top of primer3_core.
In a few cases it will be preferable to modify primer3_core itself rather than simply use it as a software component. Therefore, for maximum portability and modifiability we wrote it in standard ANSI C (7) using standard POSIX calls (8) with simple and universally usable ASCII input and output. Furthermore, the distribution includes a thorough set of tests for primer3_core, which make it relatively easy to ensure that modifications do not introduce errors.

Although the WWW interface is adequately self-explanatory for many casual users, for others the background information we present here will be helpful. For potential high-volume users, customizers, and biologist programmers, this chapter introduces the use of primer3_core to streamline the particular primer-picking tasks at hand.

1.2. Where to Find Primer3

Public WWW interfaces for use by anyone with a Web browser (for example Netscape) are reachable from http://www.genome.wwi.mit.edu/cgi-bin/primer/info.cgi. You can also download the Primer3 programs from this location. The program primer3_core is available in source form only, and generating an executable program requires a C compiler; see Subheading 3.1. for details. The source code for the WWW interface is also available and can be used on computers running a Web server. The WWW interface (like primer3_core) can be modified to meet the needs of particular sets of end-users. It is written in perl, the language of choice for this sort of application (11).

1.3. What Primer3 Does not Offer

We regret that we do not have resources to distribute Primer3 in ready-to-run executable form, with “native” front ends (e.g., for Microsoft Windows or Mac), or on tape, diskette, or CD. Other primer selection software is available in widely supported commercial form (though possibly not as a customizable software component). Examples include OLIGO®, available through Molecular Biology Insights, Inc., Cascade, Colorado, http://www.mbinsights.com/, (see also ref. 12), DNASTar’s PrimerSelect module for LaserGene (http://www.dnastar.com/), and the Prime module in Genetics Computer Group’s Wisconsin Package. Primer selection programs available from academic institutions include Primer 0.5 (upon which Primer3 was based, but available as a stand-alone program and as a ready-to-run executable for Macs and PCs) (13), and OSP—oligonucleotide selection program (14).

The following tasks are not built in to Primer3:

- Automatically adding standard 5′ tails to each primer.
- Selecting nested primer pairs.
• Selecting primer pairs for multiplex amplification.
• Designing a tiling of amplicons across a sequence.
• Picking primers from a reverse-translated amino acid sequence.

(However, we have used primer3_core as a software component in conjunction with other codes to accomplish each of these tasks but the last.) The packages mentioned above perform some of these tasks.

1.4. PCR and Primer Design Applications Are Diverse

Primer design is really many different problems. Sometimes one wishes to design primers for a large number of sequences, and if for some reason it is difficult to find good primers for a particular source sequence one simply discards that source sequence. An example would be high-throughput whole genome mapping (the application for which Primer3 and its predecessors were originally designed). In this case one designs STSs from tens of thousands of sequences and then uses these STSs in hundreds of amplifications. In this application no one sequence is particularly valuable compared to the cost of primers and subsequent amplifications, so it is not worth proceeding with a sequence for which there are only dubious primer pairs.

In other applications one wishes to design primers to amplify a particular sequence if at all possible; if there are no obvious good primers one will choose several possibilities in the hope that at least one will work. Examples include designing primers to distinguish two very similar sequences or to amplify a particular exon flanked by a CpG island in which it is hard to find a good primer. In this situation the precious resource is the particular sequence to amplify, and the scientist is willing to spend considerable effort getting a clean amplicon.

There are many other variables in primer-design goals. Sometimes one wants large amplicons (for example to amplify as much of a cDNA as possible), and sometimes one wants very short amplicons (for example to flank a single nucleotide polymorphism as closely as possible). Sometimes the amplification template is complex (for example a mammalian genome), and sometimes it is simple (for example a single bacterial artificial chromosome). Some Taq formulations are less likely than others to produce primer dimers or self-priming hairpins.

Because primer design is really many different problems Primer3 gives users numerous options to specify which primers are acceptable and which primers are better than others. The number of these options can be overwhelming to new and experienced users alike, but typically for any particular application only a few need changing from default values.

This chapter cannot discuss all of Primer3’s options, but it covers those you are most likely to change. The WWW interface and the README distributed with the program document the more esoteric options.
2. Primer3 from the End-User Perspective

This section refers primarily to the WWW interface, which calls upon primer3_core to perform almost all of the work of selecting primers. Primer3 takes as input a sequence and selects single primers or PCR primer pairs. Figure 2 shows example input in the WWW interface. The user has pasted the source sequence into the large data-entry field near the top of the page, selected the RODENT Mispriming Library and entered a Sequence Id (“Example Sequence 1”) and a Target (“40, 78”).
Below the field for the source sequence are three check boxes labeled **Pick left or use left primer below. Pick hybridization probe (internal oligo) or use oligo below.** and **Pick right primer or use primer below.** These check boxes govern whether Primer3 tries to design a primer pair, a primer pair plus hybridization probe, or an individual primer (e.g., for sequencing) or hybridization probe. In Fig. 2 the **Pick left...** and **Pick right...** boxes are checked, so Primer3 will select PCR primer pairs.

Below each of these three check boxes is an input box. Placing an oligo in one of these boxes instructs Primer3 to evaluate that oligo and choose matching oligos (depending on the check boxes).

After the user clicks on any of the **Pick Primers** buttons in the input page (Fig. 2) Primer3 returns suggested primers as shown in Fig. 3. Subheading 2.2. discusses the interpretation of this output in detail; **Subheading 2.3.** suggests some strategies for proceeding when Primer3 is unable to find any acceptable primers or primer pairs.

The label for each input option is a link to documentation on the meaning of the option and how Primer3 uses it. For example, clicking on **Max End Stability** takes one to the following documentation:

Max End Stability

The **Max Mispriming** and **Pair Max Mispriming** input fields are important in many situations because the source sequence might contain one of the interspersed repeats (ALUs, LINEs, and others) that make up more than 35% of the human genome (15). The user should either replace these sequences by Ns before picking primers or should select a **Mispriming library.** (Not all mispriming is strictly speaking caused by repeats; it could be any sequence that one does not wish to inadvertently amplify.) However, the WWW interface at the [www.genome.wi.mit.edu](http://www.genome.wi.mit.edu) web site only offers repeat libraries for human and for mouse and rat (RODENT).

The maximum primer length is restricted because the nearest neighbor melting temperature model agrees well with reality only for relatively short sequences (16,17).

### 2.1. How Primer3 Picks Primers

Primer3 accepts many options that specify which primers are acceptable and which primers are better than others. In the WWW interface the user selects
Fig. 3. Output from Primer3’s WWW interface when primers have been found.

these options through text boxes, check boxes, and pull-down menus. For example in Fig. 2 these include the Mispriming Library pull down (above the sequence input field), Product Size Min, Opt, and Max input fields, and all the input fields beneath the General Primer Picking Conditions heading.

Some options specify which primers are acceptable. For example to the left of Product Length the Min and Max options set lower and upper bounds on the length of products. Such options are called constraints because they
constrain the set of acceptable primer pairs. Other options that specify constraints include Primer Tm Min and Max, Max End Stability, and Max Mispriming. (Tm is an abbreviation for “melting temperature”.)

Other options specify characteristics of the optimum output primers or primer pairs (beyond specifying those that are simply acceptable). Examples include the Product Length Opt (Optimum) and the Primer Tm Opt inputs. Roughly speaking, if Product Length Opt is specified Primer3 tries to pick a pair of primers that produce an amplicon of approximately the specified length. For some options the user need not specify the optimum value because Primer3 can take it for granted; for example there is no input field for the optimum mispriming library similarity, which Primer3 assumes to be 0.

Primer3 examines all primer pairs that satisfy the constraints and finds pairs that are closest to the optimum. How does Primer3 calculate how close a primer pair is to the optimum? By default the WWW interface tries to balance equally primer length, primer melting temperature, and product length. (For compatibility with earlier versions primer3_core by default uses only primer length and primer melting temperature.)

However, to accommodate the diversity of primer picking applications Primer3 is flexible in the formula it uses to calculate how close a primer or primer pair is to the optimum. The technical term for this formula is objective function. Thus, suppose you deem the difference in melting temperature between the two primers to be more important than their lengths, melting temperatures, and the product size. Then you can use the Objective Function Weights... sections of the input page (as partially shown in Fig. 4) to tell Primer3 to use these considerations in calculating optimality. In Fig. 5 this effect is accomplished by the values in the fields labeled Product Size Lt and Gt. Tm Difference, and Primer Penalty Weight. (Primer Penalty Weight is an adjustment factor for the entire per-oligo contribution to the objective function. More details are available in the online documentation.)

2.2. Interpreting the Output when Primers are Found

Please refer to Fig. 3. The top of the output displays the sequence id (Example Sequence 1) and a number of informational notes. The next part of the output displays the best left and right primers, and their characteristics (starting position, length, melting temperature, and so forth). Then the output displays information specific to the input sequence and the selected pair.

The next information is a quasi-graphical representation of the location of the left (>>>>>>> ...) and right (<<<<<< ...) primers in the source sequence, as well as any important features of the source sequence, in this example only the position of the target (marked by asterisks***** ...). Following the sequence is information for some number of additional primer pairs. (The user can control
specify con-

primers or

Examples

for optimi-

tized pairs

applications

is objective

temperature

Function

Fig. 4. The part of Primer3's WWW interface that allows modification of the objective function. In this example "Product Size Lt" and "Gt," "Tm Difference," and "Primer Penalty Weight" have been changed from the defaults.

cation id

teristics of

location of

sequence id

the output

d pair.

te primary sequence is controlled

t number returned by entering a different value in the Number to Return

input field.

Finally the output contains a section headed Statistics, which we discuss in detail below.

2.3. What if There Are no Acceptable Primers?

Recall from Subheading 1.3, that in some situations one wants only good primers for uniform conditions and would rather discard some source sequence than deal with dubious primers. Most of Primer3's default option values are tuned to suit these situations; constraints are strict. Given strict constraints

the specific options of the objective function are not critical because any acceptable
primer or primer pair will be reasonably good, but potentially usable primers might be rejected as unacceptable.

In other situations, however, one must design primers for a given sequence at almost any cost. Suppose you are faced with such a situation, and Primer3 cannot find acceptable primers given the default constraints. In this case Primer3 will return a screen similar to that shown in Fig. 6.

Then what? The most intuitive course is to relax the constraints that you think are least important in your particular situation and that are most likely preventing any primers or primer pairs from being acceptable. The Statistics
Primer3 Output

PRIMER PICKING RESULTS FOR Example Sequence 2
Using mispriming library rodent_ref.fasta
Using 1-based sequence positions
SEQUENCE SIZE: 180
INCLUDED REGION SIZE: 180
TARGETS (start, len): 40, 78
1 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa gg
2 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
3 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
4 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
5 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
6 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
7 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
8 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
9 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa
10 gcacaccaggaagtctgccgctggtgctggcgtcctagggcatttgaaaktktagtaa

KEYS (in order of precedence):
****** target

No Acceptable Primers Were Found

The statistics below should indicate why no acceptable primers were found. Try relaxing various parameters, including the self-complementarity parameters and mass and max oligo melting temperatures. For example, for very A-T-rich regions you might have to increase maximum primer size or decrease minimum melting temperature.

Statistics

<table>
<thead>
<tr>
<th></th>
<th>3' lib</th>
<th>poly</th>
<th>high end</th>
<th>high poly</th>
<th>high</th>
<th>high any</th>
<th>high 3'</th>
<th>tm high</th>
<th>tm too high</th>
<th>no</th>
<th>bad</th>
<th>DCC clamp</th>
<th>GQ%</th>
<th>reg</th>
<th>get</th>
<th>N</th>
<th>many</th>
<th>sid</th>
<th>tar</th>
<th>in</th>
<th>in</th>
<th>excl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>175</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>168</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Right</td>
<td>465</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>150</td>
<td>117</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pair Stats:</td>
<td>considered</td>
<td>0</td>
<td>ok</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Output from Primer3’s WWW interface when no primers were found.

Section at the bottom of the output indicates reasons that individual primers are unacceptable. In the example above it is clear that the main problem is that all acceptable left primers have too high a melting temperature (as indicated by the column headed tm too high).

A word of caution: Primer3 never considers a primer that is unacceptable because of its position. Thus, if a primer is outside of the included region or can never be acceptable given the length of the sequence, the position of any specified “included region” and targets and the range of allowable product sizes then it is not counted in the considered column. If it seems as though very few primers are even being considered, you might want to modify your maximum
and minimum product size options, or expand the included region. An example of such a situation is the following, in which no left primers are considered:

<table>
<thead>
<tr>
<th>Statistics</th>
<th>consid</th>
<th>too many</th>
<th>in tar</th>
<th>in excl</th>
<th>bad</th>
<th>no GC</th>
<th>tm too</th>
<th>tm high</th>
<th>high</th>
<th>high 3'</th>
<th>poly</th>
<th>high end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Right</td>
<td>1401</td>
<td>25</td>
<td>0</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>484</td>
<td>191</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The **Pair Stats**: section indicates reasons that pairs of primers (as opposed to single primers or oligos) are rejected. For example, there might only be a few acceptable primers, all of which when paired would create a product with too low or too high a melting temperature.

Examining the **Statistics** (and less commonly) the **Pair Stats**: sections should suggest constraints that if relaxed would allow primers to be chosen.

In some cases (especially when relaxing several constraints at once) it might be desirable to also modify the objective function to reflect specific primer design objectives. In the sequence in Fig. 6, the temperatures of all possible left primers are too high. One way to proceed is to incrementally relax what seem to be the limiting constraints, for example increasing the **Primer Tm Max** option until an acceptable left primer is found. Alternatively, it can be more expeditious to simply relax the limiting constraint totally, as in Fig. 5, in which the **Primer Tm Min** and **Max** are set to 0 and 100°C, respectively. The primer pair selected with these relaxed constraints is:

<table>
<thead>
<tr>
<th>OLIGO</th>
<th>start</th>
<th>len</th>
<th>tm</th>
<th>gc%</th>
<th>any</th>
<th>3'</th>
<th>rep</th>
<th>seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT PRIMER</td>
<td>10</td>
<td>19</td>
<td>68.72</td>
<td>61.16</td>
<td>6.00</td>
<td>1.00</td>
<td>10.00</td>
<td>aagcagtctgcgcctggt</td>
</tr>
<tr>
<td>RIGHT PRIMER</td>
<td>175</td>
<td>20</td>
<td>60.42</td>
<td>55.00</td>
<td>2.00</td>
<td>0.00</td>
<td>12.00</td>
<td>gggctcaaggggaataacac</td>
</tr>
</tbody>
</table>

And the Statistics are

<table>
<thead>
<tr>
<th>Statistics</th>
<th>consid</th>
<th>too many</th>
<th>in tar</th>
<th>in excl</th>
<th>bad</th>
<th>no GC</th>
<th>tm too</th>
<th>tm high</th>
<th>high</th>
<th>high 3'</th>
<th>high lib</th>
<th>poly</th>
<th>high end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Right</td>
<td>397</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>378</td>
</tr>
</tbody>
</table>

Now there are acceptable primer pairs but a large difference in melting temperatures between the left and right primers. To reduce this difference one can include it as part of the objective function, as shown in Fig. 4. After this adjustment, **Primer3** selects the following primer pair:

<table>
<thead>
<tr>
<th>OLIGO</th>
<th>start</th>
<th>len</th>
<th>tm</th>
<th>gc%</th>
<th>any</th>
<th>3'</th>
<th>rep</th>
<th>seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT PRIMER</td>
<td>11</td>
<td>18</td>
<td>67.83</td>
<td>66.67</td>
<td>4.00</td>
<td>1.00</td>
<td>10.00</td>
<td>aagcagtctgcgcctggt</td>
</tr>
<tr>
<td>RIGHT PRIMER</td>
<td>180</td>
<td>20</td>
<td>66.95</td>
<td>60.00</td>
<td>2.00</td>
<td>2.00</td>
<td>10.00</td>
<td>tagcggcgcgtcagaatag</td>
</tr>
</tbody>
</table>
3. Primer3 for Biologist Programmers

3.1. Installation Instructions

The source distribution is available as a UNIX "tar" archive, which can be managed by the UNIX tar utility, by the Windows / Windows NT WinZip utility (Nico Mak Computing; http://www. winzip.com/winzip.htm) or by the Mac DropStuff with Expander Enhancer utility www.aladdinsys.com. To run primer3_core you will first need to compile it using an ANSI C compiler with POSIX libraries and run the tests supplied with the distribution as documented in the README.

3.2. Examples of How to Use primer3_core as a Software Component

In this section we present two examples of using primer3_core as a software component. The code for these examples is available in the Primer3 distribution.

3.2.1. Using primer3_core with UNIX Pipes

The first example is a relatively lightweight application of the kind that requires only a minimum of perl scripting experience.

This example shows how to postprocess primer3_core's output to complete an oligo design task. The example task is the specification of "overgos" (John D. McPherson, pers. comm.), in which a 36-mer double-stranded hybridization probe is constructed from annealing overlapping 22-mers and filling in the singled stranded tails:

```
ACTGTGCCCTGCATTTCAGAGA
       |||||
ACGTCCTCTCCATTAATTCATT
```

Specifically, we will show code that takes an existing primer pair and then designs an overgo that will hybridize to the site amplified by that primer pair. Here is the UNIX command line one would use:

```
prompt> ./primer3_core < input | ./overgo.pl
```

In this command line primer3_core runs first, taking its input from the file "input", whereas its output is sent directly to the perl program overgo.pl via a
UNIX pipe (specified by the vertical bar, ‘|’, on the command line). The input could be prepared by hand in a text editor or (more likely) produced by another program. It has the form of tag=value pairs, a format dubbed Boulder-IO (18):

```
PRIMER_SEQUENCE_ID=Overgo Example
PRIMER_PICK_INTERNAL_Oligo=1
PRIMER_INTERNAL_Oligo_MAX_MISHYB=36
PRIMER_INTERNAL_Oligo_MIN_SIZE=36
PRIMER_INTERNAL_Oligo_MAX_SIZE=36
PRIMER_INTERNAL_Oligo_OPT_SIZE=36
PRIMER_INTERNAL_Oligo_MIN_TM=10
PRIMER_INTERNAL_Oligo_MAX_TM=90
PRIMER_INTERNAL_Oligo_OPT_TM=70
PRIMER_INTERNAL_Oligo_SELF_ANY=30
PRIMER_INTERNAL_Oligo_SELF_END=30
PRIMER_INTERNAL_Oligo_MISHYB_LIBRARY=humrep
PRIMER_PRODUCT_SIZE_RANGE=70-1000
PRIMER_EXPLAIN_FLAG=1
PRIMER_PAIR_WT_IO_QUALITY=1
PRIMER_PAIR_WT_PR_QUALITY=0
PRIMER_IO_WT_REP_SIM=1
PRIMER_IO_WT_TM_GT=0
PRIMER_IO_WT_TM_LT=0
PRIMER_IO_WT_SIZE_GT=0
PRIMER_IO_WT_SIZE_LT=0
PRIMER_NUM_RETURN=1
PRIMER_LEFT_INPUT=GAAATGTGTCCCTCCCCAGA
PRIMER_RIGHT_INPUT=GAGTTCAACCATAACGACCTCA
SEQUENCE=GGATCACAACGTTTTTTTGAACACACCCCTCTATAATGATGTATT...
```

Boulder-IO is a format for moving semistructured data between programs. Primer3 receives its input and (by default) produces its output in a simple subset of Boulder-IO. The README in the Primer3 distribution describes the meanings of all these tag=value pairs in the input, as well as those in the output. The output from primer3_core given the input above is:

```
PRIMER_SEQUENCE_ID=Overgo Example
PRIMER_PICK_INTERNAL_Oligo=1
PRIMER_INTERNAL_Oligo_MAX_MISHYB=36
```
The input

```
PRIMER_INTERNAL_OLOGO_MIN_SIZE=36
PRIMER_INTERNAL_OLOGO_MAX_SIZE=36
PRIMER_INTERNAL_OLOGO_OPT_SIZE=36
PRIMER_INTERNAL_OLOGO_MIN_TM=10
PRIMER_INTERNAL_OLOGO_MAX_TM=90
PRIMER_INTERNAL_OLOGO_OPT_TM=70
PRIMER_INTERNAL_OLOGO_SELF_ANY=30
PRIMER_INTERNAL_OLOGO_SELF_END=30
PRIMER_INTERNAL_OLOGO_MISHYB_LIBRARY=humrep
PRIMER_PRODUCT_SIZE_RANGE=70-1000
PRIMER_EXPLAIN_FLAG=1
PRIMER_PAIR_WT_IO_QUALITY=1
PRIMER_PAIR_WT_PR_QUALITY=0
PRIMER_IO_WT_REP_SIM=1
PRIMER_IO_WT_TM_GT=0
PRIMER_IO_WT_TM_LT=0
PRIMER_IO_WT_SIZE_GT=0
PRIMER_IO_WT_SIZE_LT=0
PRIMER_NUM_RETURN=1
PRIMER_LEFT_INPUT=GAAATGTGTCTTCCCCACAGA
PRIMER_RIGHT_INPUT=GAGTTTCACCCTACGACCTCA
SEQUENCE=GGATCACAAGTTTTTTTGGACACACCATATGATGTATT...
PRIMER_LEFT_EXPLAIN=considered 1, ok 1
PRIMER_RIGHT_EXPLAIN=considered 1, ok 1
PRIMER_INTERNAL_OLOGO_EXPLAIN=considered 224, long poly-x seq 12, ok 212
PRIMER_PAIR_EXPLAIN=considered 1, ok 1
PRIMER_PAIR_QUALITY=15.0000
PRIMER_LEFT_SEQUENCE=GAAATGTGTCTTCCCCACAGA
PRIMER_RIGHT_SEQUENCE=GAGTTTCACCCTACGACCTCA
PRIMER_INTERNAL_OLOGO_SEQUENCE=ACTGTGCCTGCCATTTGCA...
PRIMER_LEFT=99,20
PRIMER_RIGHT=198,21
PRIMER_INTERNAL_OLOGO=140,36
PRIMER_LEFT_TM=59.903
PRIMER_RIGHT_TM=59.981
PRIMER_INTERNAL_OLOGO_TM=72.885
PRIMER_LEFT_SELF_ANY=3.00
PRIMER_RIGHT_SELF_ANY=4.00
PRIMER_INTERNAL_OLOGO_SELF_ANY=8.00
```

ATT...

programs.

imple subscribes the in the out-
The second program, overgo.pl, takes the sequence of the 36-mer hybridization probe from this output and produces the overlapping 22-mers that constitute the overgo:

```perl
#!/usr/local/bin/perl5 -w
$/ = "\n=\n"; # Set the record terminator.
while (<>) {
    %rec = split /[=\n]/; # A DANGEROUS approach
    # to parsing the sequence.
    for (keys %rec) { $rec{$_} =~ s/\n// }
    $seq = $rec{'PRIMER_INTERNAL_Oligo_SEQUENCE'};
    next unless $seq;
    print "MARKER\t\t$rec{PRIMER_SEQUENCE_ID}\n";
    $left = substr($seq,0,22); # Get left oligo.
    $r = substr($seq, 14); # Get the right oligo,
    $right = reverse($r); # reverse it, and
    $right =~ tr/GATC/CTAG/; # complement it.
    print "LEFT_MID_Oligo\t$left\n";
    print "RIGHT_MID_Oligo\t$right\n";
    print "MAX_SCORE\t"
    $rec{'PRIMER_INTERNAL_Oligo_MISHYB_SCORE'}\n"
    $gc = ($seq =~ tr/GC/GC/); # GC content
    printf "GC_content\t%d%%%\n", $gc * 100 / 36;
}
```

The statement `$/ = "\n=\n"`; tells perl that each record is terminated by an “=” sign on a line by itself (which is the standard record terminator for
Boulder-IO). The statement \%rec = split /\[=\n\]/; parses Boulder-IO record into the perl hash \%rec. This method of parsing the output requires that we know that "=" will not appear in the value part of any Boulder-IO tag=value pair. For situations in which more robustness is required, use Lincoln Stein's perl Boulder module (available at http://www.genome.wi.mit.edu/genome_software/other/boulder.html). Using this module overgo.pl would be rewritten as:

```perl
#!/usr/local/bin/perl5 -w

use Boulder::Stream;

$in = new Boulder::Stream;

while ($rec = $in->read_record()) {

    $seq = $rec->get('PRIMER_INTERNAL_Oligo_SEQUENCE');
    next unless $seq;
    print "MARKER\t\t",
    $rec->get('PRIMER_SEQUENCE_ID'), "\n";

    $left = substr($seq,0,22); # Get the left # oligo.
    $r = substr($seq, 14); # Get the right # oligo,
    $right = reverse($r); # reverse it, and
    $right =~ tr/GATC/CTAG/; # complement it.

    print "LEFT_MID_Oligo\t\t$left\n"
    print "RIGHT_MID_Oligo\t\t$right\n"
    print "MAX_SCORE\t",
    $rec->get('PRIMER_INTERNAL_Oligo_MishyB_SCORE'), "\n";

    $gc = ($seq =~ tr/GC/GC/);
    printf "GC_content\t%d\%\n\n", $gc * 100 / 36;
}
```

Using the Boulder module is preferable because it is more robust. It will run correctly even if someone puts an "=" in, for example, the value for PRIMER_SEQUENCE_ID. The only disadvantage is that you need to get the Boulder module before you can try it. The output for the input above is

<table>
<thead>
<tr>
<th>MARKER</th>
<th>Overgo Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT_MID_Oligo</td>
<td>ACTGTGCTGCTATTTGCAGAGA</td>
</tr>
<tr>
<td>RIGHT_MID_Oligo</td>
<td>TTACCTTAATTACCTCTCTGCA</td>
</tr>
<tr>
<td>MAX_SCORE</td>
<td>15.00, MLT1b (MLT1b subfamily)</td>
</tr>
</tbody>
</table>

... consensus sequence
3.2.2 Calling primer3_core from perl

The second example is Primer3's WWW interface itself. This code fragment is adapted from the CGI script, primer3.www_results.cgi, which implements part of that interface. The CGI module is available from http://www.genome.wi.mit.edu/ftp/distribution/software/WWW/Primer3.www.cgi calls primer3_core, with a flag requesting formatted output (-format_output), and then grabs primer3_core's output and tweaks it a bit:

```perl
#!/usr/local/bin/perl5 -w
...
use FileHandle;  # Standard part of perl distribution
use IPC::Open3;   # Standard part of perl distribution
use CGI;
...
$query = new CGI;
  # $query now contains the parameters
  # to the cgi script
...
my @names = $query->param;
...
for (@names) {
  next if ... # Some cgi parameters do not get
  # sent to primer3_core
  ...
  $line = "$_==$v\n";
  push @input, $line;  # Save a Boulder-IO line for
                    # primer3_core's eventual consumption.
}
my $cmd = "./primer3_core -format_output -strict_tags"
my $primer3_pid;
my ($childin, $childout) = (FileHandle->new, FileHandle->new);
{
  local $^W = 0;
  $primer3_pid = open3($childin, $childout, $childout, $cmd);
}
```
if (!$primer3_pid) {
    print "Cannot excecure $cmd:<br>
    exit;
}
print "<pre>
",
print $chldin@input;
$chldin->close;
my $cline;
while ($cline = $chldout->getline) {
    if ($cline =~ /(.*start len tm gc% any 3\' seq/) { # Grap a particular line and # add hyperlinks to it:
        $cline = $1
        . "<a href="$DOC_URL#PRIMER_START">start</a>"
        . "<a href="$DOC_URL#PRIMER_LEN">len</a>"
        . "<a href="$DOC_URL#PRIMER_TM">tm</a>"
        . "gc% any 3\' seq"
    }
    print $cline;
}
print "</pre>"
waitpid $primer3_pid, 0;
if ($? != 0 && $? != 64512) { # 64512 == -4 ...
    # primer3_core exited with
    # an error code; alert the browser.
}

Of course the -formatted_output flag in $cmd is not an essential part of the paradigm at work in this example. The script could have parsed Boulder-IO output and then formatted or processed the information in some other way.

3.2.3 Other Uses of primer3_core as a Software Component

The two examples above show how to use primer3_core as a component in a lightweight Unix pipeline (the overgo design example) and how to use perl’s open3 command to start an execution of primer3_core and then grab its output for further processing. An intermediate approach that is simpler to program
than using open3 is to simply use the perl open command and then return primer3_core’s output unmodified, e.g.

```perl
if (!open(PRIMER, "| $cmd")) {
    print "Cannot execute <pre>$cmd\n</pre>\n$wrapup";
    return;
}
print PRIMER @input;
close PRIMER;  # primer3_core’s output is the same
    # as this script’s output.
if ($? != 0 && $? != 64512) {  # 64512 == -4
    ...  # $cmd exited with an error code.
}
```

At the Whitehead institute we have used primer3_core as part of an industrial-strength primer design pipeline that includes vector clipping (identification and electronic “removal” of vector arms), microsatellite repeat identification, and automatic screening for vector contaminants. We have also used it in pipelines that add constant 5’ tails to each primer and in pipelines that find a tiling of amplicons across a sequence. For this last application we set PRIMER_FILE_FLAG=1 in primer3_core’s input, which directs primer3_core to create files containing all acceptable left and right primers. A different program then selects primers from these lists to produce the tiling.

### 3.3. Efficiency Considerations

The running time of Primer3 is seldom an issue for users of the WWW interface. However, users of primer3_core for high volume applications should be aware of the factors that determine running time. The most expensive operation in selecting individual primers is a check against a mispriming or mishyb library (the actual time needed for each oligo is a linear function of the size of the library). The next most expensive operations are checks for oligo self-complementarity, and, if Primer3 examines a large number of primer pairs, checking oligo pairs for self-complementarity.

Primer3’s running time depends also on the size of the sequence in which to select primers. Selecting a single primer pair anywhere within a 10-kb sequence will take approx 10 times as long as selecting a single primer pair anywhere within a 1-kb sequence (all other options being equal).

The following are additional determinants of Primer3’s running time:

- Strict as opposed to liberal constraints on oligos. Primer3 excludes primers based on cheap computations (e.g., oligo melting temperature) before examining more
en return

apup

e same

in indus-
lentifica-
e repeat
have also
lines that
don we set
directs
primers. A

tiling.

WWW
is should
xpensive
iming or
on of the
for oligo
ner pairs,
which to
sequence
anywhere
me:
ters based
ing more

expensive-to-compute characteristics (e.g., similarity to mispriming library entries) so relaxing cheap-to-compute constraints entails evaluation of a larger number of expensive-to-compute constraints.

- Acceptable locations for primers (considering also constraints on product size). This item is similar to the preceding one. Primer3 does not perform expensive operations to characterize primers which, because of their location, can never be part of an acceptable primer pair.

- The PRIMER_FILE input flag. This flag causes Primer3 to compute every characteristic, including mispriming similarity and self-complementarity, of every possibly acceptable primer.

- Cost of computing the objective function. There are two subcases.
  - The objective function depends on expensive-to-compute characteristics of oligos or primers, such as similarity to mispriming or mishyb libraries or complementarity between primers in a pair. In this case Primer3 must perform these expensive computations on essentially all acceptable primers.
  - The objective function depends on characteristics of primer pairs per se, such as product melting temperature or product size. In this case Primer3 must calculate whether each individual primer is acceptable, which usually requires some expensive computation to determine acceptability.

(When the objective function depends neither on expensive characteristics of individual primers nor on characteristics of primer pairs then Primer3 organizes its search so that it only checks expensive constraints on the best primers.)

Acknowledgments

The development of Primer3 and the Primer3 WWW interface were funded by Howard Hughes Medical Institute and by the National Institutes of Health, National Human Genome Research Institute, under grants R01-HG00257 (to David C. Page) and P50-HG00098 (to Eric S. Lander).

We gratefully acknowledge the support of Digital Equipment Corporation, which provided the Alphas that we used for much of the development of Primer3, and of Centerline Software, Inc., whose TestCenter memory error, memory leak, and test coverage checker helped us discover and correct a number of otherwise latent errors in Primer3.

Primer3 is the most recent of a number of primer-picking programs implemented at Whitehead Institute, starting with primer 0.5 (13). Primer3 started as a reimplementaion of Primer 0.5 as software component; the design of Primer3 draws heavily on the design of Primer 0.5 and Primer v2 (Richard Resnick) and the WWW interface designed by Richard Resnick for Primer v2.

Thanks to Alex Bortvin, Mark Daly, Nathan Siemers, and William J. Van Etten for reviewing drafts of this chapter.
References