A critical review of PCR primer design algorithms and cross-hybridization case study

F. John Burpo

Department of Chemical Engineering, Stanford University, CA 94305

Submitted August 11, 2001

ABSTRACT

The success of the polymerase chain reaction (PCR) is highly dependent on primer design. Commonly used primer design programs rely upon a core set of parameters such as melting temperature, primer length, GC content, and complementarity to optimize the PCR product, but weight those parameters to differing degrees, as well as include other parameters for PCR specific tasks. An analysis of these design algorithms ,and other available PCR primer analysis software, was conducted to find the best means of predicting non-specific PCR products in a laboratory environment using *Saccharomyces cerevisiae* deletion strains. Results here show that there is no web-based program that is well-suited to the task of post-design PCR analysis for non-specific annealing and secondary structure in the context of the whole genome. A *brute force* technique was employed using the National Center for Biotechnology Information's (NCBI's) Megablast and Zuker's mfold to correlate primer sequence similarities and secondary structure predictions in a full genome context, with inconclusive results. The ultimate conclusion is that there is a need for a user-friendly, post-design PCR analysis program to accurately predict non-specific hybridization events that impair PCR amplification.

INTRODUCTION

The ability to reproduce a target section of a DNA sequence through the use of the polymerase chain reaction has facilitated a wide array of amplification techniques. Whether the objective is shotgun sequencing, or target specific, the success of the PCR strategy is highly dependent on the small synthetic oligonucleotides that hybridize to the complementary DNA sequence. These short nucleotides function in pairs known as the forward and reverse primers, which amplify a specific DNA sequence and, more importantly, anneal exclusively to that DNA target locus (Lexa, 2001).

The primer pairs are designed and selected so that they extend toward each other, polymerizing the complementary DNA sequence to the extent that the target region is covered in each cycle of the PCR. Each cycle begins at a high temperature (~95°C) to denature the double-stranded DNA into two single strands. This is followed by a lower temperature step (45-65°C) in which the primers anneal to their respective complementary DNA sequence. The temperature is then increased (~75°C) to enable the primers to extend by polymerizing nucleotides complemetary to the target DNA as shown in Figure 1. This process is then repeated for 25-45 cycles (Glick and Pasternack, 1998).

After recently joining the NASA Functional Genomics Group at the Stanford

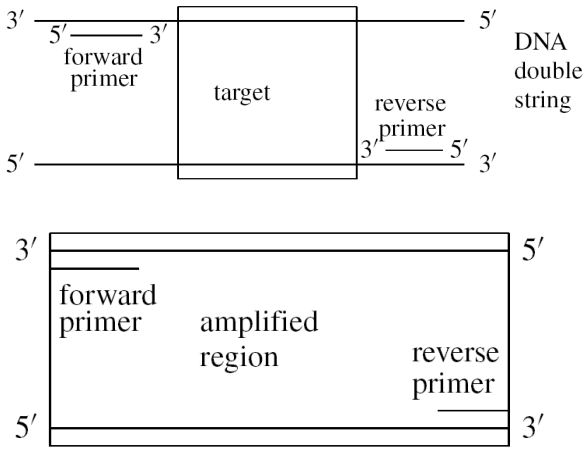

Figure 1 (Kampke, et al, 2001)

Genome Center, I became involved in a project to clone all of the deletion cassettes in the 6,000(+) *Saccharomyces cerevisiae* deletion strains in order to then sequence the amplification products and verify the inserted molecular bar codes. A lab colleague had optimized two 96-well test plates with two corresponding sets of primers (each well contained a separate deletion strain matched to a corresponding well in two plates of primers). The optimized process resulted in a number of seemingly non-specific PCR products for many of the wells. I then became interested in analyzing available primer design algorithms and determining the best means to predict and avoid similar non-specific products for the remaining deletion strains.

The objective of this paper is to analyze the algorithms and parameter weightings of commonly used primer design programs, such as PRIDE, PRIME+, DOPRIMER, PRIMO, Primer Master, and MEDUSA, and then apply these tools to the primer sets used in the NASA Functional Genomics lab. The longer-range objective is to develop a predictive model program for non-specific hybridization products.

REVIEW OF PCR PRIMER DESIGN

Designing the optimal primer pair entails a tradeoff of a variety of parameters. Over the course of the last ten years, a set of parameters has evolved to establish the core of many available programs. These include melting temperature, string-based alignment scores for complementarity, primer length, and GC content. Most programs establish real values for these primer criteria and involve trade-offs to find the optimal primer for a particular use (Kampke, et al, 2001). Only the PRIDE program was found to use qualitative user inputs and a fuzzy logic in its algorithm (Haas, Vingron, et al, 1998). Many programs include additional parameter objectives such as minimizing the total number of primers for a project (and therefore cost), excluding various target sections (repeat rich regions, or GC content <20 or >80%), target length, and so forth to improve primer quality (Haas, Vingorn, et al, 1998). Once the parameters and weightings are set by the user, these algorithms iteratively cycle through all primer candidates to identify the optimal primer set with the highest objective score.

Physical Parameters

Melting Temperature (Thermodynamic Stability). Regardless of the type of PCR to be conducted, the melting temperature of the primers used must be similar to ensure as consistent performance as possible between forward and reverse primer pairs. Almost all the programs analyzed here enable user defined melting temperature ranges for both primers and PCR products. Many programs use either the original nearest neighbor method (Breslauer, et al, 1986), or the same method with empirically determined thermodynamic values (Sugimoto, et al, 1996) to determine primer melting temperature (T_m) with the DNA target according to the equation,

$$T_m(\text{prim}) = \frac{\Delta H}{[\Delta S + R \ln(c/4)]} - 273.15^\circ C + 16.6 \log_{10}[K^+]$$

where the terms ΔH, ΔS, R, c, and $[K^+]$ represent the enthalpy and entropy of helix formation, molar gas constant, DNA concentration, and salt concentration

respectively. Various incarnations of this formula are used in different programs, where the critical aspects are the thermodynamic datasets used and the approximation for salt content. A more intuitive approximation of primer melting temperature assigns 2°C for each A-T pair and 4°C for G-C pairs (Suggs, et a, 1981).

The PCR product melting temperature is determined by,

$$T_m(prod) = 0.41 \times (\%GC) + 16.6 \times \log[K^+] - 675 / length + 81.5$$

where length is the number of nucleotides in the PCR product (Rychlik, et al, 1990). For primer pairs, the annealing temperature is calculated using the values from above with the formula,

$$T_a = 0.3 \times T_m(prim) + 0.7 \times T_m(prod) - 14.9$$

(Rychlik, et al, 1990). The primer melting temperature is a straightforward estimation of a DNA-DNA hybrid stability and critical in determining the annealing temperature. A T_a too high will result in insufficient primer-template hybridization and, therefore, low PCR product yield. While a T_a too low might possibly lead to non-specific products caused by a higher number of base pair mismatches (Rychlik, et al, 1990), where mismatch tolerance has been found to have the strongest influence on PCR specificity (Rubin and Levy, 1996).

G/C Content. A general rule followed by most primer design programs is to bracket the G/C content of primers to between 40 and 60% (Lowe, et al, 1990). A G-C pairing involves three hydrogen bonds versus two for an A-T pair (Alberts, et al., 1994), where an optimal balance of GC content enables stable specific binding, yet efficient melting at the same time (Li, et al, 1997). While program default settings and user input value ranges lie within some middle limits, such as the Genetics Computer Group's PRIME+ with a G/C content limit of 40-55% (Edelman, 1999), some programs allow for a lower or higher G/C content. This may be appropriate for an application such as differential display PCR (DD-PCR), where primers are designed to anneal to a 3' untranslated region where the A/T content as high as 60-80% (Graf, et al, 1997).

GC Clamp. Some programs enable the requirement to have a GC-type nucleotide pair at the 3' end of primers. These pairs include CC, GG, CG, or GC and are believed to create a more stable hybridization, or clamp-like effect, at the point of polymerization with Taq polymerase (Lowe, et al, 1990). Programs like PRIME+ enable the specification of any combination of clamps through the use of nucleotide ambiguity codes (Edelman, 1999). Prescribing the two 3' bases may be restrictive and not generate primer solutions for large-scale projects.

Self-Complementarity. Self-complementarity of a primer enables either the formation of secondary structure in the single-stranded oligonucleotide, or binding to another copy of itself such that it forms a primer dimer that is able to extend during polymerization. Either case will prevent the primer from annealing to the target DNA. A straightforward approach, used in programs such as PRIDE and DOPRIMER, is to conduct a pairwise comparison of a primer to a reverse copy of itself to identify the primer dimer with the highest number of complementary matches where there is a 3' terminal base and 5' overhang as shown in Figure 2. A lower weighting results for primers that form primer dimers (Haas, Vingron, et al, 1998, Kampke, et al, 2001).

Figure 2

A similar complementarity comparison and weighting test is conducted for forward and reverse primer pairs. As in the case with self-complementarity, forward-reverse primer annealing creates a primer dimer that has the possibility of extending during the polymerization phase of PCR, or simply prevents the primer pair from hybridizing with the target sequence (Kampke, 2001). Programs such as PRIMO also enable the user to define acceptable primer self-complementarity, and acceptable 3' end primer self-complementarity Li, 1996).

Primer Length. Primer length is a function of the competing criteria of uniqueness, hybridization stability, and cost-minimization by seeking the shortest oligonucleotide (Li, et al, 1997). Early algorithms, such as Lowe's Turbo Pascal program (Lowe, et al, 1990) to GCG's PRIME+, limit primer size from 18-22 nucleotides (Edelman, 1999). While short primers, 8-11mers, may yield several products, increasing primer size counter-intuitively does not indefinitely increase specificity according to real PCR data. It has been speculated that increasing primer length may also increase nucleotide mismatch tolerance (Rubin and Levy, 1996).

Another aspect to primer length is cost, where oligonucleotide cost is measured in terms of expense per nucleotide. Short 8-12mer oligonucleotides, which have multiple annealing sites, are used in a Greedy algorithm to minimize the total number of primers needed for applications where all the target sequences are known (Doi and Imai, 1999).

Other Parameters. Many primer design programs enable the user to define other evaluation criteria such as salt and DNA concentrations; number of bases to skip after each acceptable primer; PCR product size range; total number of primers or primer pairs; penalties for ambiguous bases in the target sequence; and excluding regions with non-random sequences or poor base quality. These additional user defined parameters vary depending on whether the PCR seeks to amplify defined target regions with a primer pair for each, or seeks a smaller number of primers to amplify all sequences. PRIMER3 was found to have the most user input controls over the design process (Rozen, 1998).

Algorithms

With the weighted parameters defined by the user, primer design programs iteratively cycle through an algorithm to generate the primer candidates with the highest score. The Primer Master algorithm outlines a typical algorithm where the first step is to identify all the primer candidates with the following properties:

- No repeat structure in the sequence (e.g. actgactgactgactg)
- No large GC rich or deficient regions
- No nucleotide stretches (e.g. agctTTTTTT)
- No hairpin secondary structure
- No dimer potential
- Annealing temperature within user set limits

The resulting set of primer candidates is then used to search for target sequence similarities, to identify those with the most energetically favorable annealing properties. These primers are then coupled in pairs if they meet the following criteria:

- Annealing temperature difference between the primers does not exceed user defined value
- The target sequence flanked by the primers is within the user defined limits
- Primer pairs do not form dimers
- The 3' end of the primers cannot bind to any site of the other primer
- Primers cannot form a hairpin at the annealing temperature of the other primer
- All PCR products do not differ in size from each other more than user defined value (Proutski and Holmes, 1996)

Genome-Wide Similarity Searches. One major drawback in most current PCR primer design programs is the default setting to screen primers against only the DNA target sequence for possible non-specific hybridization. None of the programs analyzed here enabled the user to screen primer candidates against the whole genome in a user-friendly manner. To do so would require manually attaching a full-genome database to the target sequence input field.

Mismatch Tolerance. What primer design programs also do not do is to correlate primer length to mismatch tolerance under various deviations from the predicted annealing temperature as mathematically modeled (Rubin and Levy, 1996). Additionally, 3'-terminal nucleotide primer mismatches automatically exclude primers as possible candidates to amplify the target sequence; however, based on results that indicate 3' C-T mismatch extensions are possible at a 10^{-2} rate in PCR (Huang, et al, 1992), mismatch extensions should be included as part of a non-specific PCR product search, albeit with low weight.

RNA Complementarities. Searching for possible RNA-DNA hybridization is not part of primer design algorithms. This may be the case for a number of reasons. 1) RNA databases are generally not complete and well-documented for many species. Although the 3-dimensional structure is available for many RNA transcripts, full, searchable databases of open reading frames (ORF's) with introns removed are not generally available. 2) There seems to be an implicit disregard for RNA non-specific hybridization in many laboratories. This may be justified for protocols requiring purified DNA for PCR, but for protocols involving in-situ or whole-cell PCR, RNA products are present in the reaction mixture. 3) Secondary and tertiary structure present in RNA products are subject to a similar denaturing that DNA undergoes at ~95˚C, leading to a variable extent of refolding during the PCR annealing phase, whereby RNA complementary sequences may, or may not be exposed to primers. Further work is needed in this area, whereby primers could be specifically designed to anneal to both RNA and DNA with varying degrees of nucleotide mismatches to find the ratio of PCR product between the two template types. In this manner it could be empirically justified to either include or discard a RNA database search as a weighted parameter in PCR primer design.

MATERIALS AND METHODS

*Polymerase Chain Reaction***.** The PCR conducted used whole yeast cells to eliminate a DNA purification step in order to expedite amplifying and sequencing all existing yeast deletion strains. Two 96 well plates (Plate 1 and Plate 2) are considered here, where there is a different yeast deletion strain in each well. For each plate of yeast cells, there is a plate of forward (A) and plate of reverse (D) primers. The optimized reaction mixture per well for this whole-cell PCR consists of:

1.7 _l H_2O
1.5 _l 10X Buffer (Perkin Elmer)
1.5 _l 10 mM dNTP's (Amersham Pharmacia)
2.0 _l 25 mM $MgCl_2$ (Perkin Elmer)
2.0 _l 20 mM Primers (1 _l forward and reverse), (OPERON, Illumina)
0.3 _l *Taq* polymerase
7.0 _l *S. cerevisiae* deletion strain cells (non-standard concentrations depending on growth times – standardized growth curves were in development at the time of writing)

YPD broth was removed from yeast cells and 100 _l of H_2O added to each well. Cells were washed by shaking and centrifugation. H_2O was removed from the wells and 40 _l H_2O added back in to roughly *standardize* yeast cell concentration. Cells, primers, and then master mix were added together. PCR was conducted with a hot start, with the following settings:

94°C	5'	Initial Denature
94°C	30"	Denaturing Segment
55°C	1'	Annealing Segment
72°C	3'	Elongation Segment

72°C	7’	Final Extension
4°C	∞	

After the PCR, 6 _l of product are added to 3 _l of buffer and run on a 1% agarose gel, with 5 _l of ethidium bromide / 100 ml H_2O, for approximately 30 minutes at 130 V and then imaged.

This final protocol was optimized by Ammon Corl.

Web Searches. Web search programs, techniques, and settings are described in Results.

Note: **The source primer design program and user defined settings for the PCR primers used in the protocol outlined above were unavailable. All of the following searches were conducted without prior knowledge of the user defined primer design characteristics.**

RESULTS

Two representative PCR product images are shown in Figures 4 and 5. These images were selected due to the potential non-specific annealing indicators such as smears and multiple bands that are still present in the optimized protocol. This protocol used a forward and reverse primer to amplify the entire yeast deletion cassette, which includes up/down molecular bar codes, and a kanamycin resistance gene. Based on the inconsistent product quality analyzed here, the PCR strategy was later changed to amplifying deletion cassettes with four primers to produce two shorter products that do not include the kan^r sequence. Although a higher percentage of PCR product resulted (>90%), smears and multiple bands were still present. This follow-on protocol was not analyzed pending the availability of the additional primer data sets. The analysis, here, of the initially optimized protocol is taken to be representative of non-specific annealing problems as the A and D primers are used in the follow-on protocol, as well (Figure 3).

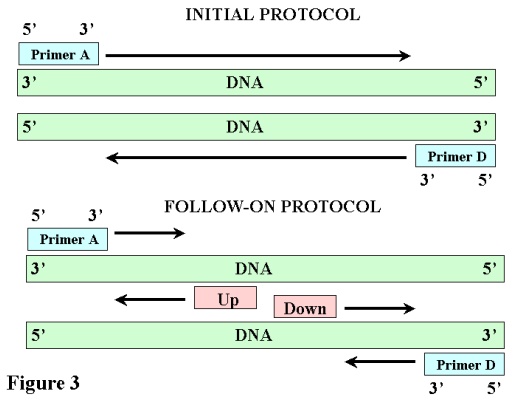

Figure 3

The hand-written annotations in Figure 4 indicate plate well correspondence to the image bands. For instance, each image consists of four labeled rows indicating the initial position of the PCR product before electrophoresis. The wells in the image rows alternate between the PCR product plate rows (E.G. ABABAB . . .) with PCR product plate columns numerically marked on Figures 4 and 5.

Plate 1 (OPERON)

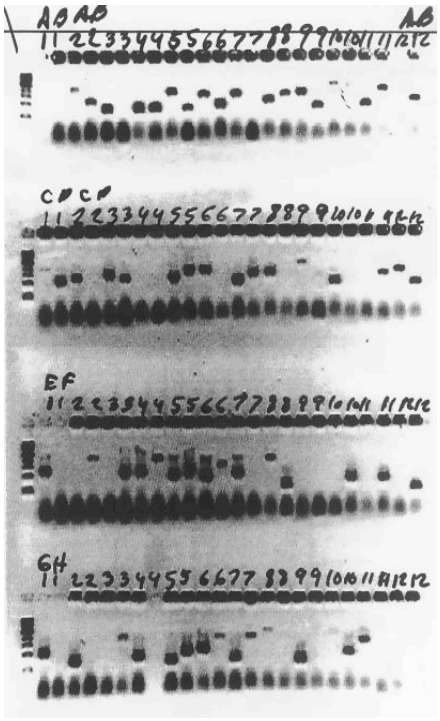

Figure 4

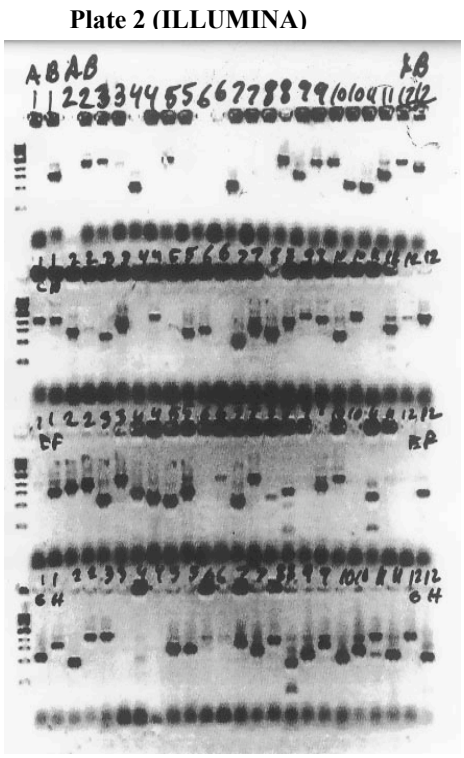

Figure 5

In a Microsoft Excel spreadsheet data file containing the primer sequences and target ORF, a column was added to characterize possible non-specific hybridization (no band, faint band, smear, 2 bands, 3 bands, etc). For instance, in Figure 5, well F5 is characterized as 'smear' and well B3 is characterized as 'faint band, ' and so forth for both plates. In this manner all primer sequences that qualitatively did not provide a singular PCR product as desired were identified.

Without knowing the algorithm or user defined settings used to generate the primer sets, a sequence similarity search and secondary structure analysis were determined to be the most appropriate techniques for predicting the non-specific PCR product results in Figures 4 and 5.

Sequence Similarity Search

As discussed in the first half of this paper, a variety of programs are available for primer design and analysis. The analysis function these programs, however, involves only the candidate primers used in the process of generating the primer sets, rather than enabling a post-production review. For instance, in this case, when inheriting or using a primer set from an unidentified design source, one may want to analyze the primers before use, or may want to determine if on-hand primers will function for a different PCR objective. Another goal may be to use a second algorithm as an independent check on primer quality, or to use additional features not available on the original program.

For this purpose, PCR primer design programs in general are woefully deficient in post-production primer analysis. PRIMER3, PRIME+, *Saccharomyces cerevisiae* Genome Database Web Primer, DOPRIMER, XPRIMER, CASSANDRA, and PRIMER DESIGN were all investigated for their ability to input a batch primer data file and analyze it for secondary hybridization sites and secondary structure in a whole genome context. This was met without success. The implicit underlying assumption of these programs seems to be that if the primer set was developed using that program, then the analysis has already been performed and indicated as part of the design output.

One program that seems promising is Virtual PCR, which accepts user-given primers and then conducts a similarity search using NCBI Blast to identify sequences in public databases that are complementary to any two primers. Input is limited to only two primers at a time, however, making batch analysis difficult (Lexa, 2001). While the algorithm was analyzed, the actual program was not exercised due to an unfamiliarity with the Perl scripts required to run the publicly available program. A hands-on evaluation of this program is recommended as future work in developing a comprehensive post-design/production primer analysis tool.

To overcome the available web-based program limitations, the Operon and Illumina primer sets were reformatted in Microsoft Word to facilitate a NCBI Nucleotide Megabast

Search (Madden, 1996). Four files were created, one for each primer plate. The sequence similarity desired was any continuous complementarity beginning at the terminal 3' position of the primer, where this type of complemetarity is believed to be most likely polymerized for non-specific annealing.

The Megablast search parameters used were,

- Database: Yeast
- Expect: 10
- Word Size: 12
- % Identity: 75
- Gap Opening Penalty: -G 10
- Gap Extension Penalty: -E 2

The –G 10 –E 2 input in the *Other Advanced* field was used to emulate an ungapped search as recommended by the NCBI Blast Help page; however, occasional single or double nucleotide gaps resulted. Adjusting the settings to –G15 –E 10, -G20 –E5, etc did not completely eliminate the occasional gaps. The expectation, word size, and % identity values were chosen to facilitate permissive, and therefore, long stretches of sequence similarity. In this manner, every primer was tested against the entire yeast genome for sequence similarity.

Two output formats were used to display results. First, the *query-anchored without identities* output was used to enable a visual means to quickly identify sequence similarities beginning at the 3' terminal position. For instance, the very first ORF tested, YAL068C, from the Plate 1, Operon primer set, generated the following results:

```
1       tcacgaacaccgtcattgatcaaata 26
11396   tcacgaacaccgtcattgatcaaata 11421
782428  tcacgaacaccgtcattgatcaaata 782403
12930   tcacgaacaccgtcattgatcaaata 12955
1063428 tcacgaacactgtcattgatcaaata 1063403
1651    tcacgaacaccgtcattgatcaaata 1676
8614    tcacgaacaccgtcattgatcaaata 8639
7452    tcacgaacaccgtcattgatcaaata 7477
8631    tcacgaacaccgtcattgatcaaata 8656
412708   caaggacaccgtcattga         412725
368896               tcattgatcaaat  368908
236011             gtaattgatcaaata  235997
1648    tcacgaacaccgtcattgatcaaata 1673
1523769 tcacgaacactgtcattgatcaaata 1523744
890067   cacgaacaccgtcat            890081
289047      gagcaccgtcattgatc       289031
44445           accgtcattgattaa     44459
353823              tcattgatcaaata  353836
196008        acaccgtcattgat        196021
305853              tcattgatcaaat   305865
90305         accgtcattgatc         90317
```

The first line, highlighted in green, represents the test primer sequence in 5' to 3' order. The output then spatially aligns genome sequences underneath the test set, listing the nucleotide start and stop positions to the flanks. Chromosome sequence links are also listed, but not shown here. Ten sequences, highlighted in yellow, have 100% similarity with the test primer sequence. Two additional sequences highlighted in blue have partial similarity starting at the 3' position, with similarities of 15 and 14 nucleotides. These multiple similarity sites may explain why no PCR product is seen for this primer in Figure 4, well A1.

While this data format is visually useful, its ability to be parsed in an Excel format was not achieved. Additionally, Megablast does not easily lend itself to mandatory similarity starts at the 3' pimer position. Therefore, the *Hit Table* format was also generated.

Hit Table

Although continual stretches of sequence similarities starting at the 3' position are most likely to yield competing PCR products, any relatively long sequence similarity may cause non-target-specific annealing, with the possible effects of no product, smeared or multiple product bands.

Considering that Doi and Imai's Greedy algorithm for primer generation uses 8mer's (Doi and Imai, 1999), and other work has shown PCR product formation with as few as 13 base pairs shared between a 20mer primer and DNA template (Rychlik, 1995), a parsing technique was used that simply listed a raw count of continuous sequence similarity at varying thresholds for each reaction well. The output results listed in the Hit Table, like those listed here,

Fields: % identity alignment length . . .
YAL068C 100.00 26 . . .

were sorted into categories of ≥20 bp, ≥15 bp, and ≥10 bp continuous sequence similarities regardless of starting position. These values were recorded in the Microsoft Excel columns next to the non-specific annealing characterization (no band, faint band, smear, 2 bands, 3 bands, etc) to determine a correlation between the sequence similarities and the non-specific annealing effects empirically observed. Once the number of similarity hits were recorded for each base pair threshold, the data was divided between single PCR products and non-specific annealing events. Once separated, the number of cumulative hits for ≥20 bp, ≥15 bp, and ≥10 bp was averaged for the number of wells. The results are shown below,

Avg Hits/Well	≥20 bp	≥15 bp	≥10 bp
Single PCR Product	1.4166667	3.4722222	8.3888889
Non-Specific Annealing	1.4833333	3.3666667	6.2666667

The data reveals a very finite difference in the number of hits and the total similarity of those hits. Counter-intuitively, the average number of hits for cumulative continuous base pair similarities over 10 and 15 base pairs is higher for reaction wells producing only a single PCR product, the desired effect. Plate 1, well C12, for instance, had 66 sequence similarities greater than 10 bp, with 15 over 15 bp, and yet only a single PCR product was formed.

On the other hand, empirically observed non-specific hybridization events had a higher number of hits over 20 bp, indicating that for the reaction conditions specified in Materials and Methods, continuous sequence similarities greater than 20 bp may lead to non-specific annealing. This was seen in the case of Primer Plate 1, well B7 where 38 total sequence similarities of 10 base pairs or greater were found. Six of them were over 20 bp.

With a more efficient parsing system, the sequence similarities starting at the 3' terminus could be used to refine this data by eliminating hits that do not initiate similarity at the more critical 3' position.

Other patterns not examined and correlated here include primer pair annealing temperature variations and their effect on base pair mismatch thresholds. Given the high number of continuous sequence similarities present in the yeast genome for most primers, varying the mismatch threshold in Megablast was not deemed necessary.

Primer Secondary Structure

In order to determine the possible effect of secondary structure on PCR product inhibition, all of the primer sequences for the non-specific hybridization events that resulted in *no band* for the Operon Primer Plate 1 were put in a batch file and submitted to the mfold website with a temperature setting of 55°C to match the annealing temperature of the PCR protocol (Zucker, 1999). Of the 33 samples submitted, all 33 had at least one predicted secondary structure, and some had as many as 10 structures. The 36 primer sequences for the same plate that produced only one distinct PCR product were tested using the same program. Again, all test sequences resulted in secondary structure, with as many as 8 folded structures per primer. This technique is seemingly inconclusive and may be complicated by the many other molecular compounds released during cell lysis as part of the whole-cell PCR protocol.

DISCUSSION

The work presented here may serve as a foundation for further examining the relationship between full genome sequence similarity searches and empirical PCR data in order to develop a predictive model for non-specific hybridization events that hamper amplification. Exploring a primer's mismatch/annealing threshold as a function of length, base composition, and annealing temperature should yield a model on which to

base user defined inputs for post-design PCR batch analysis. In order for this to happen, however, a more efficient search and sort algorithm is required. The *brute force* Megablast technique employed in Results provides a good indication of non-specific annealing possibilities, but does not serve as the optimal method for further work

My intent is to program a search algorithm to that enables the sorting analysis required to correlate non-specific hybridization events as a predictive model for mis-primed PCR experiments. Once this is accomplished, the next step will be to verify these models in the laboratory before finally developing a web-based interface that enables a comprehensive PCR primer analysis for batch file inputs. This program should offer the standard functions of most PCR primer design programs such as primer length, molecular weight, and annealing temperature, but also include a predictive output of possible non-specific amplification products based on PCR conditions such as annealing temperature relative to primer T_a, and time of the annealing phase. Although such a comprehensive, full genome search will undoubtedly consume more computation time, producing more effective primers should save both the time and cost of trouble-shooting mis-primed PCR experiments.

AKNOWLEDGMENTS

All PCR labwork was performed under the direction of Viktor Stolc, research scientist at the NASA Ames Research Center. Actual bench work was performed at the Stanford Genome Center with the guidance and assistance of Ammon Corl, Department of Biology, Stanford University. This project was undertaken independently to incorporate pre-existing lab data with a critical review of PCR primer design programs in order to benefit future PCR and sequencing efforts with a predictive model of non-specific hybridization.

REFERENCES

Alberts, B., et al. (1994) *Molecular Biology of the Cell*. Garland Publishing, New York, NY.

Breslauer KJ, Frank R, Blocker H, Marky LA. (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A*. **83(11),** 3746-3750.

Doi, K. and Imai, H. (1999) A Greedy Algorithm for Minimizing the Number of Primers in Multiple PCR Experiments. *Genome Inform Ser Workshop Genome Inform*. **10**, 73-82.

Edelman, I. (1999) Primer+. *Genetics Computer Group, Inc*. http://pmgm.stanford.edu/gcg-bin/seqweb.cgi.

Graf, D., Fisher, A., and Merkenschlager, M. (1997) Rational primer design greatly improves differential display-PCR (DD-PCR). *Nucleic Acids Research*. **25(11),** 2239-2240.

Glick, B. and Pasternack, J. (1998) *Molecular Biotechnology: Principles and Applications of Recombinant DNA*. ASM Press, Washington, D.C.

Haas, S., Vingron, M., et al. (1998) Primer Design for large scale sequencing. *Nucleic Acids Research*. **26(12),** 3006-3012.

Huang, M., Arnheim, N., and Goodman, M. (1992) Extension of base pairs by Taq polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Research*. **20(17),** 4567-4573.

Kampke, T, Kieninger, M and Mecklenbug, M. (2001) Efficient primer design algorithms. *Bioinformatics*. **17(3)**, 214-225.

Lexa, M., Horak, J., and Brzobohaty, B. (2001) Virtual PCR. *Bioinformatics*. **17(2),** 192-193.

Li, P., et al.. (1997) PRIMO: A Primer Design Program That Applies Base Quality Statistics for Automated Large-Scale DNA Sequencing. Genomics. **40**, 476-485.

Lowe, T. et al. (1990) A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nucleic Acids Research*. **18(7),** 1757-1761.

Madden, T.L., Tatusov, R.L. & Zhang, J. (1996) Applications of network BLAST server. *Meth. Enzymol*. **266**, 131-141.

Proutski, V. and Holmes, E. (1996) Primer Master: a new program for the design and analysis of PCR primers.

Computer Applications in the Biosciences. **12(3),** 253-255.

Rozen, S. and Skaletsky, H. (1998) Primer3. Code available at http://www.genome.wi.mit.edu /genome_software/other/primer3.htm.

Rubin, E. and Levy, A. (1996) A mathematical model and a computerized simulation of PCR using complex templates. *Nucleic Acids Research*. **24(18),** 3538-3545.

Rychlik, W., Spencer, W.J., and Rhoads, R.E. (1990) Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Research*. **18(21),** 6409-6412.

Rychlik, W. (1995) *Biotechniques*. **18**, 84-90.

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

Suggs, S., et al. (1981) *ICN-UCLA Symposia on Developmental Biology Using Purified Genes*. Academic Press Inc., New York, NY, Vol 23, pp683-693.

Zuker, M., Mathews, D. and Turner, D. (1999) Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. *RNA Biochemistry and Biotechnology*, 11-43.