Computational Principles of Primer Design for Site Directed Mutagenesis

Alexey Novoradovsky*, Vivian Zhang**, Madhushree Ghosh**,
Holly Hogrefe**, Joseph A. Sorge**, Terry Gaasterland***

*The Rockefeller University, New York, NY 10021
**Stratagene Cloning Systems, La Jolla, CA 92037
**Scripps Genome Center, SIO/UCSD, La Jolla, CA 92093-0202

ABSTRACT

Site directed mutagenesis (SDM) is a method to alter DNA molecules with intentionally designed mutations. SDM creates mutated copies of the original sequence as follows: an oligonucleotide primer hybridizes to the DNA template to form a duplex with base mismatches or single-stranded DNA loops. Then, DNA polymerase extends the primer to create a mutated copy of the initial DNA molecule. In this work, we address the problem of designing optimal primers to introduce single or multiple site mutations, deletions, or insertions. Our approach is based on three principles. First, for mutations that introduce amino-acid substitutions, preference in codon replacement is given to changes that minimize the number of nucleotide substitutions. Second, codons vary in the frequency with which they are incorporated into successful mutants, so highly successful codons are favored. We conducted a series of SDM experiments to determine success rates for all codons. These rates are used in primer design. Third, primer length and primer position relative to the mutated site is chosen to minimize duplex free energy, and thus maximize duplex stability.

Keywords: Biotechnology, DNA Hybridization, Computational Methods and Numerics

1 INTRODUCTION

Hybridization of short oligonucleotides with long nucleic acid molecules is used in many molecular biological applications including microarray technologies, polymerase chain reaction (PCR), and site directed mutagenesis (SDM). A general goal in the field of nanobiotechnology is the manipulation of biomolecular structures to engineer DNA molecules with structure alterations. Today, SDM is the primary method to alter DNA or their protein derivatives.

Both PCR and SDM rely on an enzymatic extension of a deoxy-oligonucleotide 3'-end hybridized to a DNA target. The long target DNA molecule acts as a template, and the oligonucleotide acts as a reaction primer. A number of computational methods exist to design optimal PCR primers and probes [1,2]. Although the principles of SDM primer design are well established [3,4], to date, computational methods to design SDM oligonucleotides do not exist. Primer design approaches for SDM are similar but not identical to those for PCR. An SDM primer-template duplex contains intentional base-mispairing or single-stranded loops or bulges that can significantly destabilize it, thus compromising the reaction efficiency. The computational challenge is to balance the need for mutation with reaction efficiency.

SDM experiments tend to be more successful when using higher stability primer-template duplexes. This paper describes a novel computational approach to the SDM primer design in which oligonucleotide sequences are selected to form the most stable duplexes with the template despite introduced mismatches. If the desired mutations are aimed to alter a protein sequence, the method selects optimal nucleotide changes for the desired amino acid substitutions. Primer-template duplex stability is predicted by calculating the duplex free energy using the nearest-neighbor method [5,6]. The primers minimally increasing duplex energy due to the mismatches are recommended for the SDM experiments.

2 METHODS

2.1 Codon Selection

When multiple codons can introduce a desired amino acid change, codons that involve fewer base changes are preferred. For example, if threonine encoded by ACG is to be mutated to arginine then the desired change can be achieved by six different ways: AGA, AGG, CGA, CGC, CGG, or CGT. In this case, AGG is the most parsimonious choice, because the change from ACG to AGG involves just a single base substitution.

2.2 Primer Design

Primers are designed with the goal of minimizing free energy of the mismatched primer-template duplex. Free energy is calculated using the nearest-neighbor energy principles for nucleic acid duplexes [5,6]. Mismatches in a duplex are treated as internal loops or bulges flanked by perfect duplex branches (Figure 1). Therefore, the total free energy of a primer-template duplex includes a sum of stacking energies of nearest neighbors plus the energies of internal loops, bulges, and terminal "dangling ends". The nearest-neighbor free energy parameters for the temperatures from 37°C to 85°C from the program *mfold*, version 2.3 [7,8] are used as an initial data for energy calculations. The data are adapted for DNA hybridization as described earlier [5]. The data in the energy files contain free energies for all possible 256 complete internal or terminal dinucleotide duplexes, 128 possible single-nucleotide dangles, and 1-30 nucleotide internal or terminal loops and bulges. The temperature dependence of the nearest-neighbor and loop/bulge free energy parameters within 37-85°C range is close to linear, and can be approximated by the equation:

$$E_T = a * T + b$$

where T is temperature (°C), a is a slope, and b is an intercept, respectively. Parameters a and b for all nearest-neighbor and loop objects listed above are organized as a PostgreSQL-7.4 database which is queried for the calculation of the free energy of the SDM primer-template duplexes.

The primer design algorithm for amino acid substitution mutagenesis has three steps. First, optimal substitution codons are selected, as described above. Second, the template regions upstream and downstream of the mutated codon(s) are used to design complementary sequences that become parts of the SDM primer. The primer sequences immediately upstream and downstream of the mutation(s) are designed to satisfy common requirements listed in SDM manuals [4].

Additionally, primer sequences on both sides of mutation are selected to have similar GC content, and thus similar melting properties. Third, the free energy of the mismatch-containing duplex is calculated and compared to the perfect duplex energy. The difference between two energies is interpreted as an "energy cost" (Figure 1), which is to be minimized for successful SDM reaction priming. If DNA mutagenesis does not involve amino acid changes, the codon selection step is omitted.

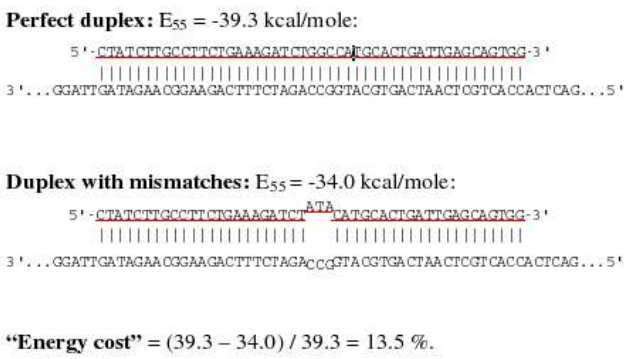

Figure 1: Calculation of "energy cost".

2.2 Web implementation

The codon selection and primer design computational algorithms are implemented as a web service at Scripps Genome Center, SIO, UCSD [9]. The process is accessed through HTML forms submission that generate and pass variables via PHP scripts. Four applications are available: 1) optimization of codon selection; 2) design of single or multiple primers with up to five single-nucleotide changes; 3) design of multiple mutagenesis primers for simultaneous introduction of up to five amino acid changes in one experiment; and 4) design of primers to generate deletions or insertions in the target DNA sequence. Further information can be obtained online through the help hyperlinks available at http://sgc.ucsd.edu/~anovo/SDM/SDM_Help.php

3 RESULTS AND DISCUSSION

SDM experiments with degenerate primers provided empirical data for codon selection principles. In these experiments, primers contained a single degenerated codon "NNK" (which includes 32 possible triplets, encoding for 19 amino acids: ACDEFGHILMNPQRSTVWY). The numbers of mutant DNA sequences were counted among 607 mutant clones. Codon biases were calculated as a deviation of observed numbers of resulting codons (NO) from expected 1/32 of the total number of analyzed mutants (NE). Codon biases ranged from -12 to +23 (Table 1).

Codon	NO	NE	Codon bias
TCT	7	18.97	-11.97
CTT	7	18.97	-11.97
TGT	10	18.97	-8.97
CCT	10	18.97	-8.97
CCG	11	18.97	-7.97
CTG	12	18.97	-6.97
ACT	12	18.97	-6.97
TTG	13	18.97	-5.97
CGT	13	18.97	-5.97
GTT	13	18.97	-5.97
AGT	14	18.97	-4.97
TTT	15	18.97	-3.97
CAG	15	18.97	-3.97
GGT	15	18.97	-3.97
TAT	16	18.97	-2.97
TCG	18	18.97	-0.97
TGG	18	18.97	-0.97
GGG	19	18.97	0.03
GAT	20	18.97	1.03
ATG	20	18.97	1.03
GTG	21	18.97	2.03
CAT	21	18.97	2.03
AAT	22	18.97	3.03
CGG	22	18.97	3.03
GCT	23	18.97	4.03
GCG	23	18.97	4.03
ATT	24	18.97	5.03
ACG	25	18.97	6.03
TAG	30	18.97	11.03
AGG	32	18.97	13.03
AAG	41	18.97	22.03
GAG	42	18.97	23.03

Table 1. Codon biases revealed by the SDM experiments with degenerate primers. NO, observed numbers; NE, expected numbers.

Codons with highest positive biases contain significantly higher proportions of the dinucleotides AG, AT, AA, TA, AC, and GA, while the dinucleotides CT, TG, TT, CC, and GT were rarely observed among successfully mutagenized clones. Correspondingly, the codon selection algorithm favors codons with high positive biases.

Duplex stability was assessed using the nearest-neighbor energy calculation, using an approach similar to the calculation of energy in RNA or single stranded DNA secondary structure [7,8]. This approach can be used in any molecular biological application that involves imperfect nucleic acid hybridization, including primer-allele-specific amplification [10] and assessment of non-specific microarray hybridization [11].

4 CONCLUSIONS

Primer design for site directed mutagenesis was addressed by integrating primer-template duplex energy calculations with empirical data on codon biases from SDM experiments that used degenerate primers.

A PHP-enabled web interface has been created to provide user access to database-driven server-side applications.

5 REFERENCES

[1] Rozen, S., Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers." In S.

Krawetz and S. Misener, eds. Bioinformatics Methods and Protocols in the series Methods in Molecular Biology. Humana Press, Totowa, NJ, 2000, pp. 365-386.

[2] Abd-Elsalam, K.A. Bioinformatic tools and guideline for PCR primer design. African Journal of Biotechnology, 2(5): 91-95, 2003.

[3] Cabre, A., Girona, J., Zalba, G., Moreno, M.U., Diez, J., Masana, L. Generation of eight adjacent mutations in a single step using a site-directed mutagenesis kit. Clin Chem Lab Med, 42(4): 384-386, 2004

[4] Stratagene QuikChange® Site-Directed Mutagenesis Kit Manual. http://www.stratagene.com/products/displayProduct.aspx?pid=131

[5] Nielsen, D.A., Novoradovsky, A., Goldman, D. SSCP primer design based on single-strand DNA structure predicted by a DNA folding program. Nucleic Acid Research, 1995, 23: 2287-2291.

[6] SantaLucia, J. Jr., Allawi, H.T., Seneviratne, P.A. Improved nearest-neighbor parameters for predicting dna duplex stability. Biochemistry, 35: 3555-3562, 1996.

[7] Zuker, M. *Mfold* web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res., 31 (13): 3406-3615, 2003.

[8] Zuker, M. Predicting Nucleic Acid Hybridization and Melting Profiles. Genome Informatics, 14: 266–268, 2003.

[9] http://sgc.ucsd.edu/~anovo/SDM/SDM_master.php

[10] Sommer, S.S., Groszbach, A.R., Bottema, C.D. PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. Biotechniques, 12(1): 82-87, 1992.

[11] Hornshoj, H., Stengaard, H., Panitz, F., Bendixen, C. SEPON, a Selection and Evaluation Pipeline for OligoNucleotides based on ESTs with a non-target Tm algorithm for reducing cross-hybridization in microarray gene expression experiments. Bioinformatics, 20(3): 428-429, 2004.