

Introducción al diseño “*in silico*” de primers.

Introducción

Esta guía es una aproximación, más o menos detallada, al diseño de primers por medios informáticos. En la primera parte revisaremos algunos conceptos generales pertinentes a este tema. En la segunda parte encontrará un taller que le permitirá poner en práctica la mayoría de los conceptos revisados.

Para una revisión del procedimiento de PCR consulte los siguientes enlaces:

<http://www.dnalc.org/ddnalc/resources/pcr.html>

http://ls126.molsci.org/PCR_primers.html

PARTE I

Hibridación, anillamiento y temperatura de fusión.

Con el término hibridación hacemos referencia a la asociación específica y complementaria de dos cadenas sencillas de ácidos nucleicos. Esta asociación responde a un proceso físico (al que nos referimos como “anillamiento”) en el cual las dos cadenas sencillas forman enlaces de hidrógeno entre sus bases complementarias (A-T y G-C), y de esta manera se lleva a cabo la formación de una molécula de doble cadena, helicoidal y complementaria.

Resulta aparente que para llevar a cabo un proceso de anillamiento es necesario entonces contar con DNA de cadena sencilla, estado que regularmente debemos inducir en el laboratorio (¡a menos que estemos trabajando con virus de RNA!) mediante el calentamiento de la cadena doble de DNA por encima de su temperatura de fusión (melting temperature o T_m) e inmediatamente enfriando la reacción alrededor de los 0°C. De esta manera se asegura que las hebras recién separadas no vuelvan a anillarse.

La temperatura de fusión de una molécula dada depende principalmente de su contenido de G+C, de sí se trata de DNA o RNA, híbridos RNA:RNA (que tienen las temperaturas más altas), de las condiciones físico químicas de la solución en la que se trabaja, así como de la longitud de las secuencias implicadas en el proceso (a menor longitud, menor temperatura de fusión).

La longitud de la secuencias es fundamental para su asociación

La asociación de cadenas de ácidos nucleicos mayores de 16pb es un proceso extremadamente específico. Si tenemos en cuenta las posibilidades combinatorias de diversas “secuencias sonda” dada una cadena de DNA:

<i>Sonda</i>	<i>Probabilidad</i>
A, C , G o T/U	1/4
Dinucleótido (ej. AG)	1/16
Tetranucleótido (ej. AGCT)	1/256

De esta manera la probabilidad de encontrar una sonda de **16pb** es de $1/4^{294.967.296}$.

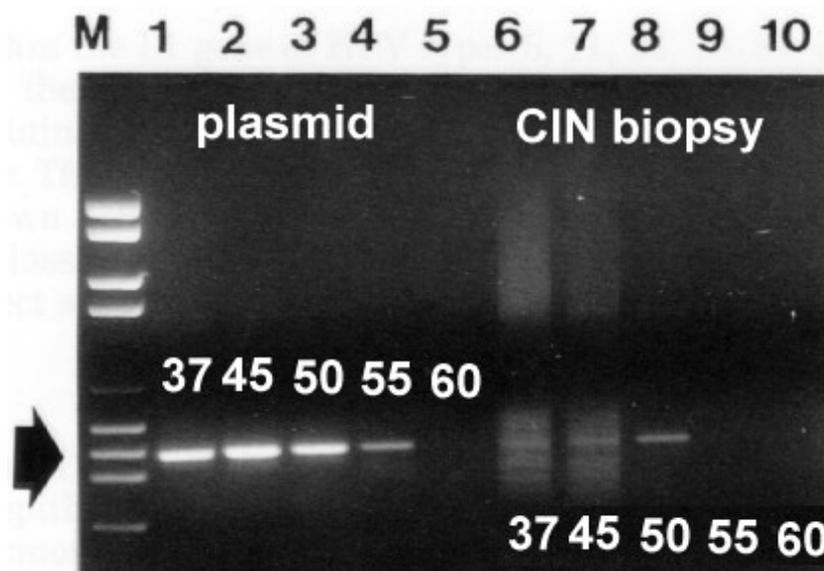
Consideraciones generales para el diseño de primers

Temperatura de anillamiento (T_a)

Como se mencionó anteriormente la temperatura de fusión de una cadena doble de ácidos nucleicos está relacionada con su longitud y contenido de G+C¹. De esta manera la temperatura de anillamiento que se escoge para una PCR **depende directamente de la longitud y composición de los primers**. Es común utilizar temperaturas de anillamiento 5°C por debajo de la temperatura de fusión más baja del par de primers utilizados (Innis and Gelfand, 1990).

Es importante tener en cuenta que una T_a demasiado baja puede llevar a eventos de inespecificidad de los primers utilizados, de tal manera que lleguen a ser anillados en secuencias diferentes a la que queremos amplificar (Rychlik *et al.*, 1990). Una T_a **demasiado alta**, por su parte, puede llevar a que el rendimiento de la amplificación sea muy bajo ya que se reduce la probabilidad de anillamiento (figura 1). Es muy importante tener consideración también que el par de primers escogidos deben **no tener T_a s demasiado diferentes**, ya que es muy probable obtener amplificaciones asimétricas (amplificaciones de una sola hebra).

Fig. 1. Efecto de T_a en la especificidad y rendimiento de amplificación de *papillomavirus humano tipo 16 (HPV 16)*.



Se amplificaron plantillas de DNA de plasmido y biopsia a diferentes T_a s. Notar que el DNA de HPV amplifica específicamente a 50°C. Imagen tomada de: Williamson and Rybicki, 1991: J Med Virol 33: 165-171

Longitud del primer

Un primer debe ser lo suficientemente complejo de tal manera que la probabilidad de anillamiento con una secuencia diferente a la de nuestro interés sea muy baja. Como se anotó anteriormente una secuencia de 16b de longitud tiene una probabilidad de ser encontrada una vez cada 4^{16} bases. Por esta razón es muy recomendable que un primer tenga, como mínimo, 17 bases de longitud.

¿Cuál cree usted que sea el resultado de utilizar primers de longitudes demasiado largas?

1 Una manera fácil de realizar el cálculo de T_m es: $T_m = 4(G + C) + 2(A + T)^\circ C$.

Temperatura y tiempo de elongación

Por lo general se encuentra entre los **70-72°C** durante **0.5-3** minutos.

Si la polimerasa tiene una actividad específica a 37°C entonces ¿de qué manera los primers, pueden ser elongados a una temperatura de anillaje tan alta?

Número de ciclos

El número de ciclos de amplificación necesarios para obtener una banda visible depende en gran medida de la concentración inicial de nuestro DNA a amplificar (Innis and Gelfand, 1990). Se recomiendan de 40-45 ciclos para amplificar 50 moléculas y entre **25-30** para amplificar **3⁵** moléculas, a la misma concentración (plateau effect).

Consejos generales para el diseño de primers:

1. Los primers deben tener una longitud de 17-28b
2. El contenido de G+C debe estar alrededor del 50-60%.
3. Los Tms entre 55-80°C son deseables .
4. Se deben evitar secuencias de 3 o más Cs o Gs en las regions terminals 3'.
5. Las terminaciones 3' no deben ser complementarias, ya que esto puede llevar a la formación de dímeros.
6. Debe evitarse los primers que formen secuencias auto-complementarias (habilidad para formar estructuras 2^o).

PARTE II

Primer3 es una aplicación que se encuentra para su libre uso en diferentes servidores web alrededor del mundo. En esta oportunidad utilizaremos la implementación de éste programa ofrecida por el **Whitehead Institute**.

Acceda a la implementación de Primer3 en el White Head Institute, en la siguiente dirección:
http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

Después de unos segundos se encontrará con una página web similar a la imagen a continuación:

Primer3 [disclaimer](#) [code](#)
[cautions](#) [FAQ](#)

pick primers from a DNA sequence (older interface). **NEW** New interface allows check for mispriming in template. **NEW**

Paste source sequence below (5'->3'), string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINES, etc.) or use a [Mispriming Library \(repeat library\)](#): NONE

Pick left primer or use left primer below. Pick hybridization probe (internal oligo) or use oligo below. Pick right primer or use right primer below (5'->3' on opposite strand).

Sequence Id: A string to identify your output.

Targets: E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the [source sequence](#) with [and]: e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.

Excluded Regions: E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the [source sequence](#) with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.

Product Size Ranges 150-250 100-300 301-400 401-500 501-600 601-700 701-850 851-1000

[Click here to specify the min, opt, and max product sizes only if you absolutely must. Using them is too slow \(and too computationally intensive for our server\).](#)

Number To Return: **Max 3' Stability:**

Max Mispriming: **Pair Max Mispriming:**

Al mover la ventana del navegador hacia abajo encontrará que primer3 cuenta con diversas secciones que le permiten controlar una amplia variedad de parámetros relacionados con el diseño de primers.

Por favor no se sienta intimidado-a por la cantidad de parámetros modificables por primer3 (alrededor de 81), la mayoría de estos son rara vez modificados, aún por los “gurúes” del diseño de primers quienes, incluso, pueden no estar seguros de la funcionalidad de muchos de ellos ;-)

Por ser este nuestro primer acercamiento al diseño de primers mediante esta herramienta utilizaremos un set de condiciones mínimo y la siguiente secuencia de ejemplo:

```
GTTCGGGAATTCGCGCGCGTTCGACCCACGCGTCCGTCCGCCTCCACCTACCAACCACCATGG
CCACGATCGCTATCATCTACTACTCCACCTACGGCCACATCGCCCAACTGGCCGAGGCCG
CCAAGAAGGGTGTGAATCCGTCGACGGCGTGACCGCCGAGATCTACCAAGTGCAGGAGA
CTCTCTCCGACGAGATTCTGGGCAAGATGCACGCCGCCCCAAGAAGGACCACCCCATCG
CCACTCTTGACACCCTGAAGGAGGCCGATGGCATCCTGTTTCGGTTTTCCCACTCGTTTTCG
GCTCGCTGCCAGCGCAGGTTAAGGCGTCTTTGACTCGGCCGGTGGCCTGTGGGCTGCTG
GGGCTCTGGTAGGCAAGCCTGCCGGTATCTTCTTCAGCACGGGCACTCAGGGCGGCGGTC
AGGAGACAACGGCCTTCACTGCCTTGACGTTCCCTAGCTCACCAGGGTCTCACCTTCGTGC
CACTGGGCTACCGTGCCCTGAGCTTTTCAACATGGACGAGCTTCACGGTGGATCCCCGT
GGGGAGCGGGAACCTTGCCGGCGGCGATGGCTCTCGCCAGCCATCGAAGCTGGAGCTGA
CGGTCGCTACCACCCAGGGTAAGTCGTTTCGCTAAAGTGGCCAAGAAGCTGGAGGCTTAAG
CACTGCGATTATTTGTGTGTGTGTGTCTTGGCTCAAGCTAAATTCTAATGCAATAAAA
TTTAGTTTTACACGG
```

Pegue esta secuencia en el campo de texto en la parte superior de la página de primer3, como se muestra a continuación:

Primer3

pick primers from a DNA sequence (older interface). **NEW** New interface allows check for mispriming in template. **NEW**

Paste source sequence below (5'→3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA undesirable sequence (vector, ALUs, LINEs, etc.) or use a [Mispriming Library \(repeat library\)](#): NONE

```
GGGCTCTGGTAGGCAAGCCTGCCGGTATCTTCTTCAGCACGGGCACTCAGGGCGGCGGTC
AGGAGACAACGGCCTTCACTGCCTTGACGTTCCCTAGCTCACCAGGGTCTCACCTTCGTGC
CACTGGGCTACCGTGCCCTGAGCTTTTCAACATGGACGAGCTTCACGGTGGATCCCCGT
GGGGAGCGGGAACCTTGCCGGCGGCGATGGCTCTCGCCAGCCATCGAAGCTGGAGCTGA
CGGTCGCTACCACCCAGGGTAAGTCGTTTCGCTAAAGTGGCCAAGAAGCTGGAGGCTTAAG
CACTGCGATTATTTGTGTGTGTGTGTCTTGGCTCAAGCTAAATTCTAATGCAATAAAA
TTTAGTTTTACACGG
```

<input checked="" type="checkbox"/> Pick left primer or use left primer below.	<input type="checkbox"/> Pick hybridization probe (internal oligo) or use oligo below.	<input checked="" type="checkbox"/> Pick right primer or use right primer below (5'→3' on opposite strand).
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="button" value="Pick Primers"/>	<input type="button" value="Reset Form"/>	

Por defecto primer3 tiene habilitadas las opciones de encontrar tanto un primer corriente arriba (pick left primer or use left primer below) como corriente a bajo de nuestra secuencia. Deshabilitando cualquiera de ellas evitaremos que el programa busque el primer en cuestión. Por otra parte, si tenemos habilitadas estas opciones e ingresamos un primer en la casilla de texto justo debajo de dichas opciones, primer3 buscará el primer recién ingresado. Por esta vez dejaremos estas opciones tal y como se encuentran.

Presione el botón “Pick Primers” y espere unos segundos.

Debe encontrarse ahora en la página de resultados de primer3:

Primer3 Output

No mispriming library specified

Using 1-based sequence positions

OLIGO	<u>start</u>	<u>len</u>	<u>tm</u>	<u>gc%</u>	<u>any</u>	<u>3'</u>	<u>seq</u>
LEFT PRIMER	510	20	59.84	50.00	4.00	0.00	TCAACATGGACGAGCTTCAC
RIGHT PRIMER	699	20	59.90	50.00	2.00	0.00	TGAGCCAAGACACACACACA

SEQUENCE SIZE: 737
INCLUDED REGION SIZE: 737

PRODUCT SIZE: 190, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00

```
1 GTTCCGGAATTCGCGCGCGTTCGACCCACGCGTCCGCTCCGCTCCACCTACCAACCACCAT
61 GGCCACGATCGCTATCATCTACTACTCCACCTACGGCCACATCGCCCAACTGGCCGAGGC
121 CGCCAAGAAGGGTGTGCAATCCGTCGACGGCGTGACCGCCGAGATCTACCAAGTGCAGGA
```

Por defecto, primer3 muestra 4 opciones de primers además de la opción principal (en la sección “Additional Oligos”, en la parte inferior de los resultados).

La mayoría de datos arrojados en la página de resultados son bastante explícitos, tal vez los más confusos sean las columnas denotadas como “**any**” y “**3'**”. Estas columnas muestran una medida de la tendencia del primer a formar estructuras secundarias y dímeros, respectivamente.

Regrese a la página inicial de primer3 con el botón de regreso de su navegador e ingrese la siguiente secuencia:

GTTCCGGAATTCGCGCGCG

en el cuadro de texto justo debajo de la opción de escogencia del primer izquierdo, como se muestra a continuación:

Paste source sequence below (5'-undesirable sequence (vector, AI

```
GTTCCGGAATTCGCGCGCGTTCGACCC
CCACGATCGCTATCATCTACTACTCCA
CCAAGAAGGGTGTGCAATCCGTCGACC
CTCTCTCCGACGAGATTCTGGGCAAGA
CCACTCTTGACACCCTGAAAGGAGGCC
GCTCGCTGCCAGCGCAGTTAAAGGCGI
GGGCTCTGGTAGGCAAGCCTGCCGGA
```

Pick left primer or use left primer below.

GTTCCGGAATTCGCGCGCG

Pick Primers

Reset Form

Presione nuevamente el botón “Pick Primers”.

Qué diferencias encuentra en los resultados arrojados esta vez por primer3 en cuanto a la tendencia de formación de estructuras secundarias y dímeros? Discuta las razones por las cuales dichos valores han cambiado.

La interfase con la cual hemos trabajado hasta ahora es una, relativamente nueva, implementación del programa, que acepta diversos rangos de tamaño para el producto que queremos amplificar (“Products size ranges”). La antigua implementación permitía definir específicamente un tamaño mínimo, uno máximo y uno óptimo. Aún es posible realizar dichas especificaciones, pero se aconseja usarlas únicamente si es estrictamente necesario, porque la utilización de éstas supone una gran carga para el servidor.

Localice y siga el enlace titulado:

Click here to specify the min, opt, and max product sizes only if you absolutely must. Using them is too slow (and too computationally intensive for our server).

Será llevado a una página con la siguiente advertencia:

Primer3

Warning: slow and expensive - Use only if you really need very precise control over the interaction of product size with other parameters. Otherwise:

Please use the new version.

Si es estrictamente necesario realice sus análisis en este formulario de lo contrario abandónelo.

La manera de operar con primer3 es básicamente la misma para cualquiera de las otras 80 opciones disponibles. Como podrá notar el programa resulta bastante sencillo de utilizar, la clave para su utilización se encuentra en el conocimiento que tengamos acerca de los diferentes parámetros que van a guiar nuestra reacción de PCR, tales como: T_m , T_a , etc.

Varié las diferentes opciones disponibles en primer3 y familiarícese con ellas, y aplique las diferentes “reglas” o consejos generales que se encuentran al comienzo de esta guía para el diseño de primers.

Lamentablemente no es posible asegurar con un 100% de certeza que un primer diseñado mediante una herramienta de software va a ser completamente efectivo, pero sin duda le ayudará a aproximarse a la solución más óptima de una manera muy rápida.

Pero ... ¿es esto todo lo concerniente al diseño de primers?

No!, por supuesto que no, aún es posible realizar algunos otros análisis.

Por ejemplo mediante mFold² podemos visualizar el tipo de estructura secundaria susceptible de ser formada por nuestros primers.

Este tipo de visualización puede resultarnos de ayuda a la hora de corregir ciertos parámetros.

Visite la siguiente dirección:

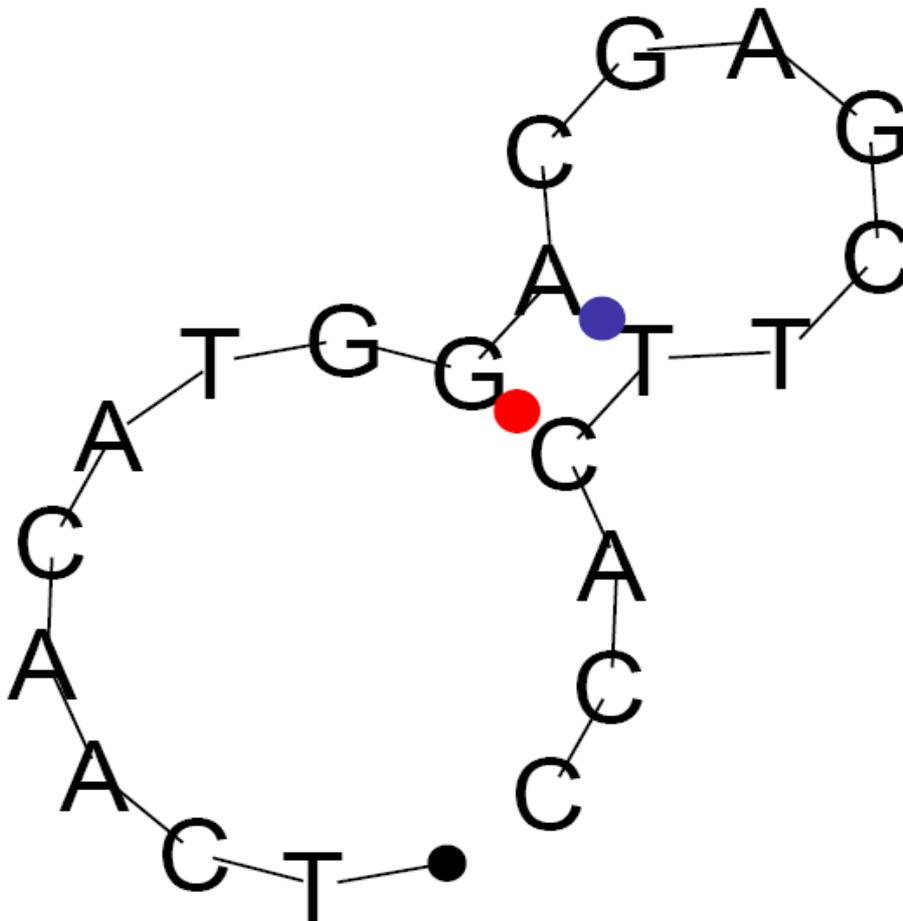
<http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html>

O alternativamente: http://mfold.burnet.edu.au/dna_form

En cualquiera de los dos casos se encontrará con un formulario donde podrá ingresar la secuencia del primer a analizar. Para este ejemplo use el primer izquierdo diseñado en el paso anterior:
TCAACATGGACGAGCTTCAC

Deberá ingresar una cuenta de correo electrónico válida, donde los resultados le serán enviados.

La siguiente imagen muestra el resultado del análisis de estructura secundaria de nuestro primer mediante mFold:



2 Este es un programa que permite realizar predicciones de estructura secundaria de DNA o de RNA. Si requiere mayor información acerca de los métodos que subyacen al programa visite:
<http://helix.nih.gov/docs/online/mfold/node1.html>

Por otra parte es muy recomendable que realice un análisis BLAST, con los primers recién diseñados, de tal manera que se asegure que dichos primers no sean complementarios a ninguna de las “posibles secuencias” involucradas en su reacción de PCR (por ejemplo, vectores de clonación).

La metodología para realizar dicho análisis es la misma que se sigue para cualquier análisis BLAST, para este efecto puede revisar la guía titulada BLAST y recuperación de secuencias.

Guía elaborada por Andrés M. Pinzón V., del **Centro de Bioinformática** del Instituto de Biotecnología en la Universidad Nacional de Colombia y del **Laboratorio de Micología y Fitopatología** de la Universidad de los Andes, y está distribuida bajo licencia:



Bogotá Colombia - Julio de 2006. Revisado en Enero de 2007

Cualquier sugerencia o inquietud dirigirla a:

ampinzonv@unal.edu.co ó andrespinzon@gmail.com