

Automated primer design software  
version 5.6.0

John K. Everett  
Thomas B. Acton    Gaetano T. Montelione

[www-nmr.cabm.rutgers.edu/bioinformatics/Primer\\_Primer/](http://www-nmr.cabm.rutgers.edu/bioinformatics/Primer_Primer/)



Contents:

- 3 Promotional flyer
- 4 Brief introduction to DNA & PCR cloning
- 16 Setting up your vectors.xml library file
- 20 Optional vector.xml entries
- 20 Clontech's In-Fusion cloning setup
- 22 Software walk through
- 30 Software options
- 37 Entering your targets
- 40 Domain parsing
- 42 Target parsing short hand
- 43 Creating mutagenic primer sets
- 44 Selecting your vectors
- 47 Viewing your results

# Primer Prim'er

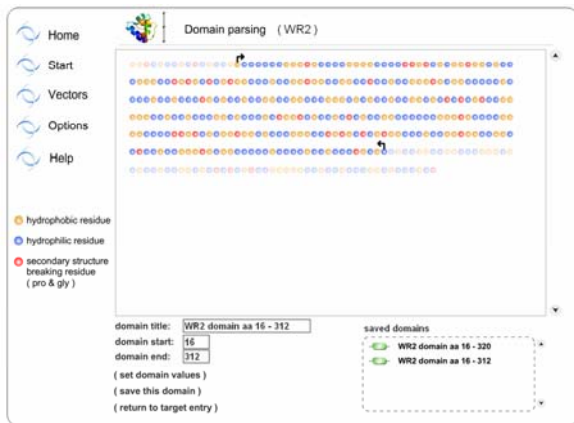
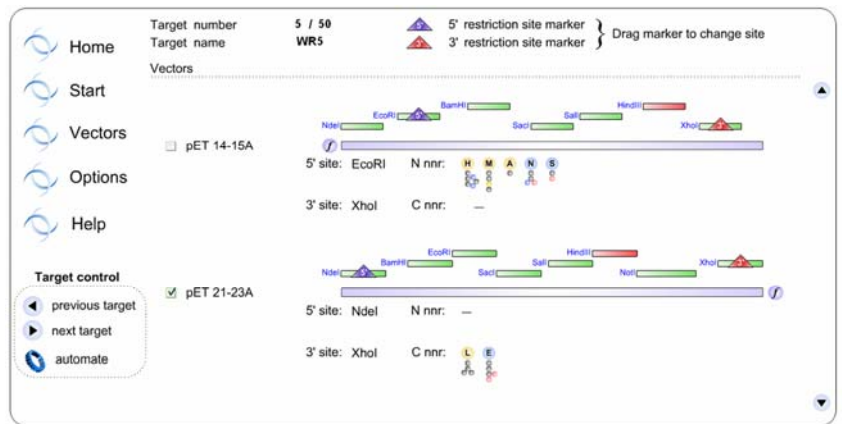
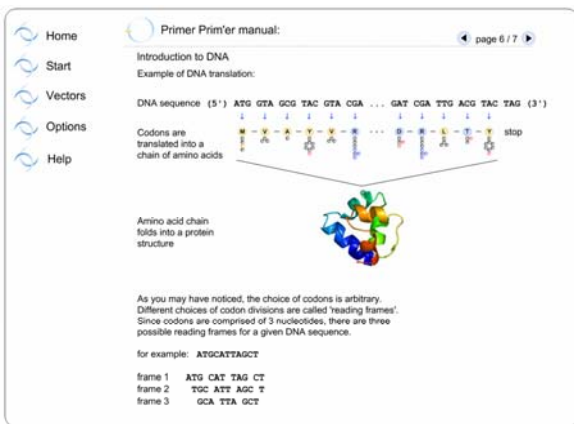
( automated PCR primer design software )

[www-nmr.cabm.rutgers.edu/bioinformatics/Primer\\_Primer/](http://www-nmr.cabm.rutgers.edu/bioinformatics/Primer_Primer/)

Primer Prim'er ( PP ) is a PCR primer design tool that completely automates the primer design process. PP generates vector specific PCR primer sets designed to amplify and insert DNA targets into your labs vectors. PP is designed to be a teaching tool as well as a powerful tool for structural genomic efforts.

PP calculates more than just the target annealing region of PCR primers. PP introduces endonuclease restriction sites into calculated primer sets. Restriction sites are embedded into target sequences when applicable and additional nucleotides are added in order to preserve frame with vector based fusions and to ensure proper endonuclease cleavage.

PP is very customizable. Aside from being able to define and employ your own vectors, a variety of settings can be tailored to your needs. PP offers tools such as a protein domain editor and virtual gels.



## Brief introduction to DNA & PCR cloning.

Nature stores all inheritable information in molecules of DNA (deoxy-ribonucleic acid).

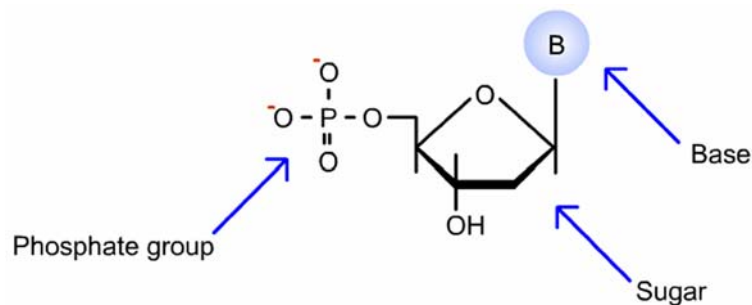
The stored information is of two types:

Protein amino acid sequences

RNA (ribonucleic acid) nucleotide sequences

DNA is a chain of linked molecules called nucleotides.

A DNA nucleotide is a deoxy-ribose sugar molecule with a phosphate group and a base group.



The phosphate group is used to link nucleotides together.

The ribose sugar serves as a scaffold to support and position the base group. The base group is used to encode genetic information.

The base group is a variable.

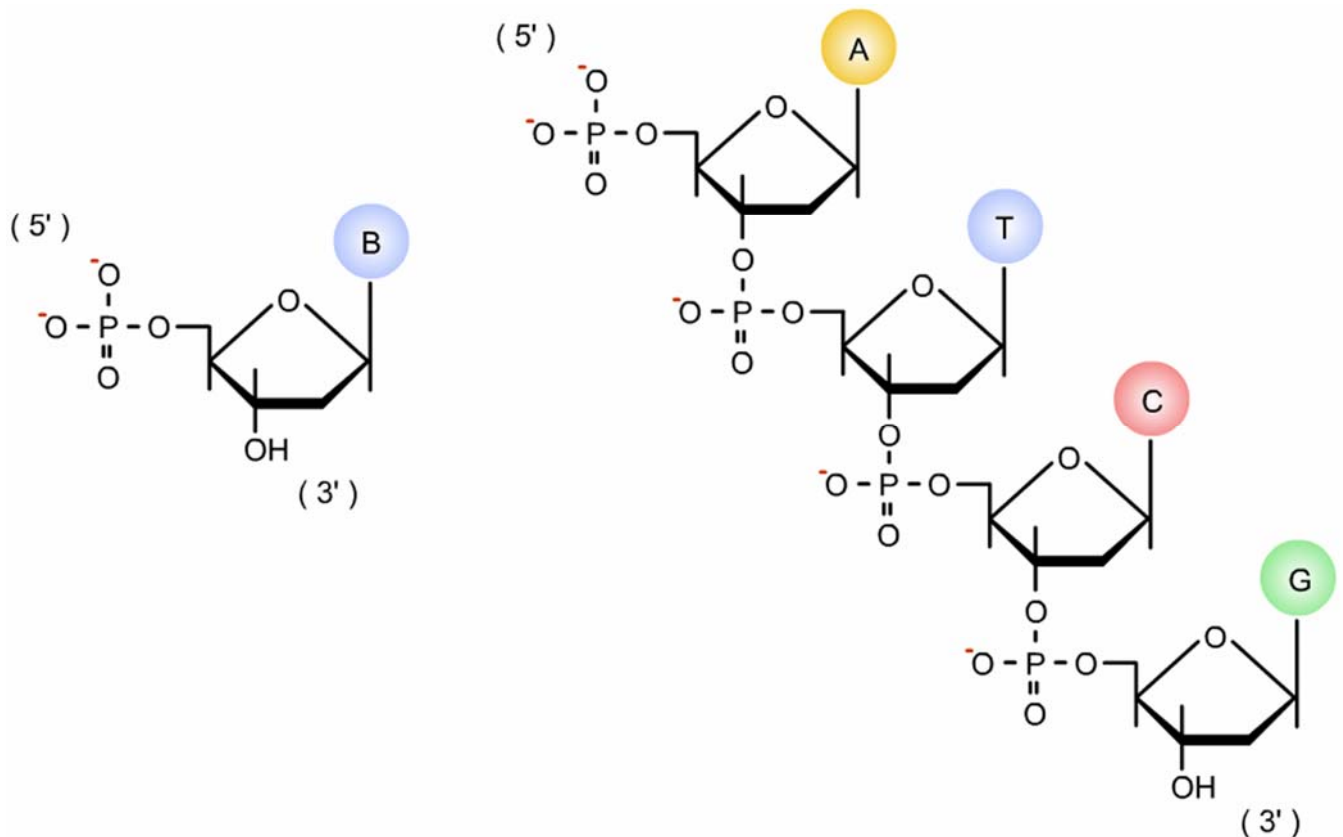
It can take one of four forms:

1. Adenine [ A ]
2. Thymine [ T ]
3. Cytosine [ C ]
4. Guanine [ G ]

Nucleotides possess an orientation. The phosphate group is called the 5' ( five prime ) end and the hydroxyl group (OH) of the ribose sugar is called the 3' ( three prime ) end. This nomenclature is derived from the positions of the carbons in the sugar.

Nucleotides form a DNA strand by forming bonds between their 3' hydroxyl groups the 5' phosphate groups of other nucleotides.

The 5' and 3' nomenclature is applied to DNA strands as well.



Nucleotide bases preferentially and discriminately bind to one another.

Adenine (A) and Thymine (T) bind to each other.



Cytosine (C) and (G) bind to each other.



Due to the structure of the bases, CG bonds are stronger than AT bonds.

Each base pairing adds to the strength of association between two DNA strands (double stranded DNA). A common description of this strength is "melting temperature".

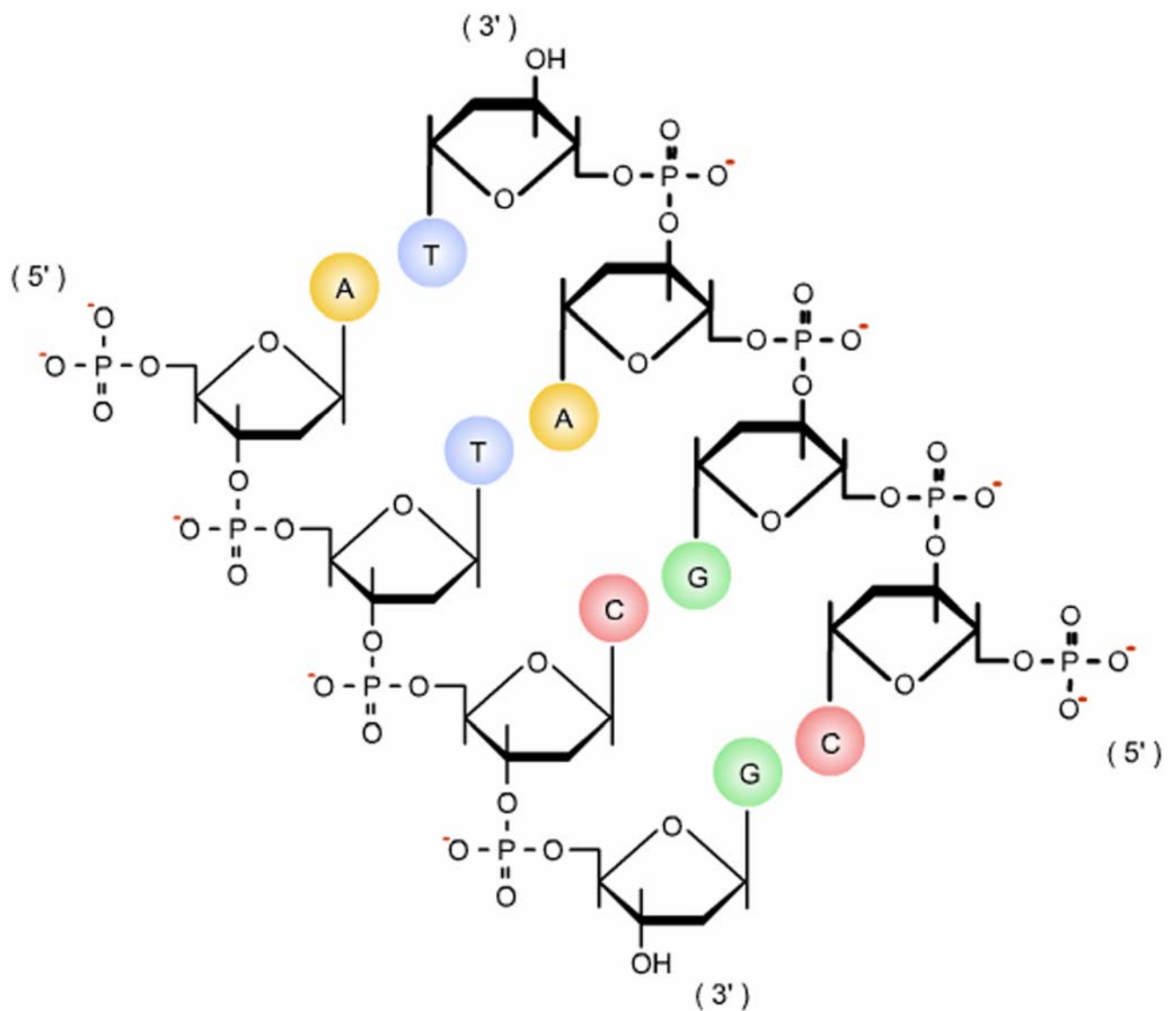
The melting temperature of double stranded DNA is the temperature at which the DNA strands will separate.

Longer double stranded segments require a greater input of energy to separate their stands than shorter segments and hence possess a higher melting temperature.

DNA does not exist as a single strand except during DNA replication. A DNA strand is coupled with a partner strand. The partner strand has the opposite orientation of the first strand.

The partner strand possesses bases that are complimentary to the first. For example, if a base in the first strand is an (A) then the adjacent base in the partner strand will be a (T).

The double stranded DNA twists into a double helix structure.



DNA encodes information within its sequence of bases ( A, T, C, G ).

Since DNA is encodes information within its sequence of bases, we need only to list the bases:

```
( 5' ) ATG GTA GCG TAC GTA CGA TCG ATC GAC GAT CGA TTG TAG ( 3' )  
( 3' ) TAC CAT CGC ATG CAT GCT AGC TAG CTG CTA GCT AAC ATC ( 5' )
```

Since the bases of each strand are complementary, we only need to list one strand to define a segment of DNA. It is customary to list the 5' to 3' strand:

```
( 5' ) ATG GTA GCG TAC GTA CGA TCG ATC GAC GAT CGA TTG TAG ( 3' )
```

The vast majority of DNA in most organisms encodes protein sequences. Protein sequences are made up of linked amino acids.

Amino acids are the building blocks of proteins.  
There are 20 standard amino acids.

Three consecutive nucleotides encode one amino acid are called a codon. Codons that do not encode an amino acid are called stop codons and mark the end of protein sequences.

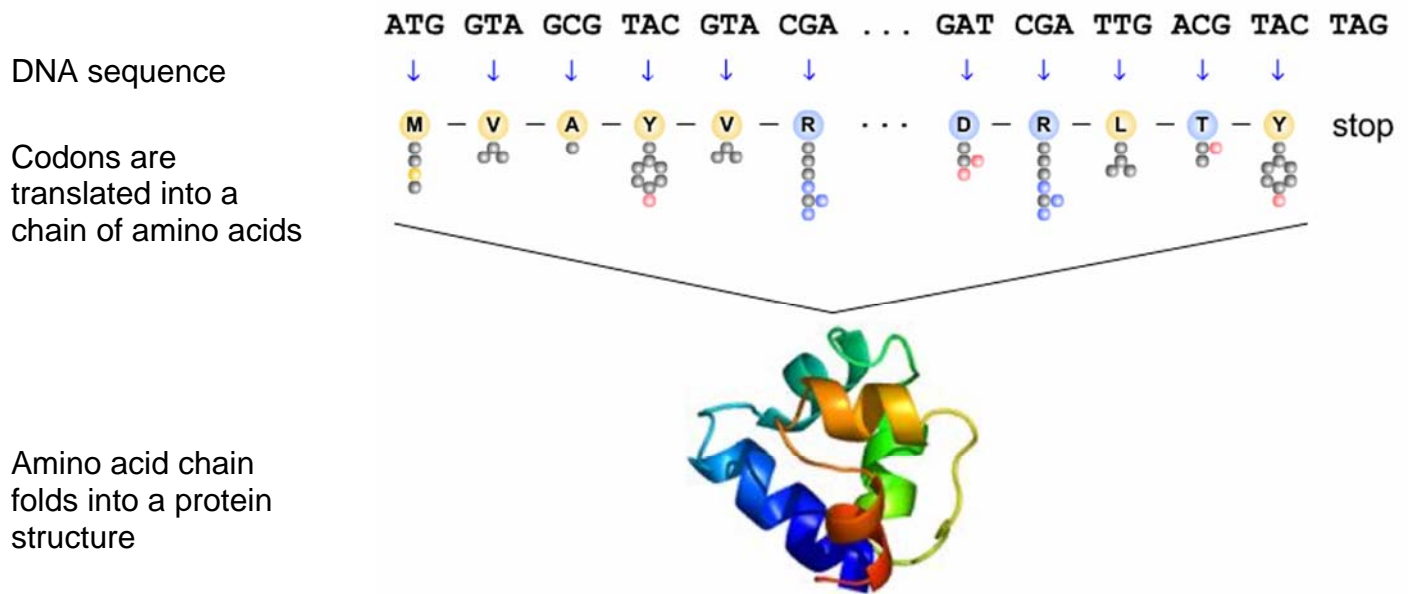
A single amino acid is often encoded by several different codons.

The process of reading DNA codons and linking together amino acids is called translation. DNA is translated in the 5' to 3' direction.

DNA is translated by a large complex of proteins and RNAs called a ribosome. The ribosome does not read the double stranded DNA directly but rather reads a single strand copy called messenger RNA.

The ribosome complex reads the DNA codons ( detailed in the messenger RNA ) and links together the encoded amino acids to create the encoded protein.

## Example of DNA translation:



The first amino acid in a protein is called the amino terminus.  
The last amino acid in a protein is called the carboxy terminus.

As you may have noticed, the choice of codons is arbitrary.  
Different choices of codon divisions are called 'reading frames'.  
Since codons are comprised of 3 nucleotides, there are three possible reading frames for a given DNA sequence.

for example:

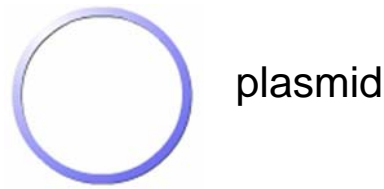
sequence: ATGCATTAGCT

frame 1 ATGCATTAGCT

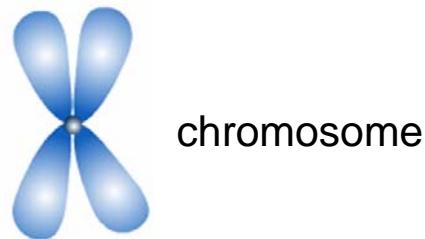
frame 2 ATGCATTAGCT

frame 3 ATGCATTAGCT

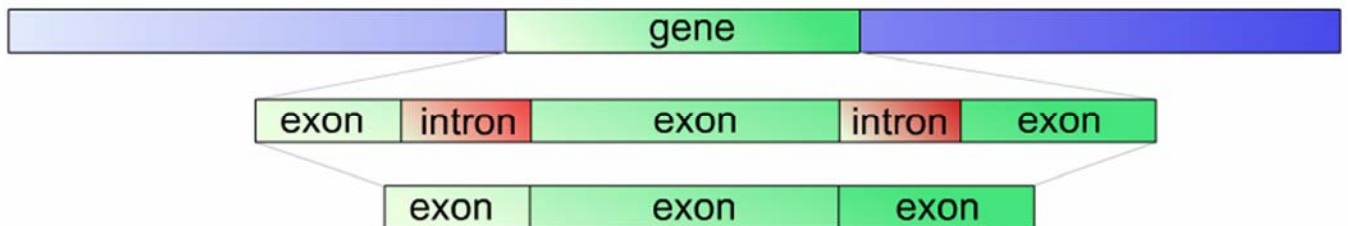
Simple organisms such as bacteria store their DNA in a circle of double stranded DNA called a plasmid.



Multi-cellular organisms organize their DNA into structures called chromosomes. Chromosomes are formed when DNA is wrapped around specialized chromosome forming proteins.



Within these structures, segments of DNA that encode proteins are called genes. One gene encodes one protein. Very often, not all of the nucleotides in a gene encode the protein. Segments of the gene that encode its protein are called exons and the segments that do not are called introns. Since introns do not encode a gene's protein they are spliced out before translation.

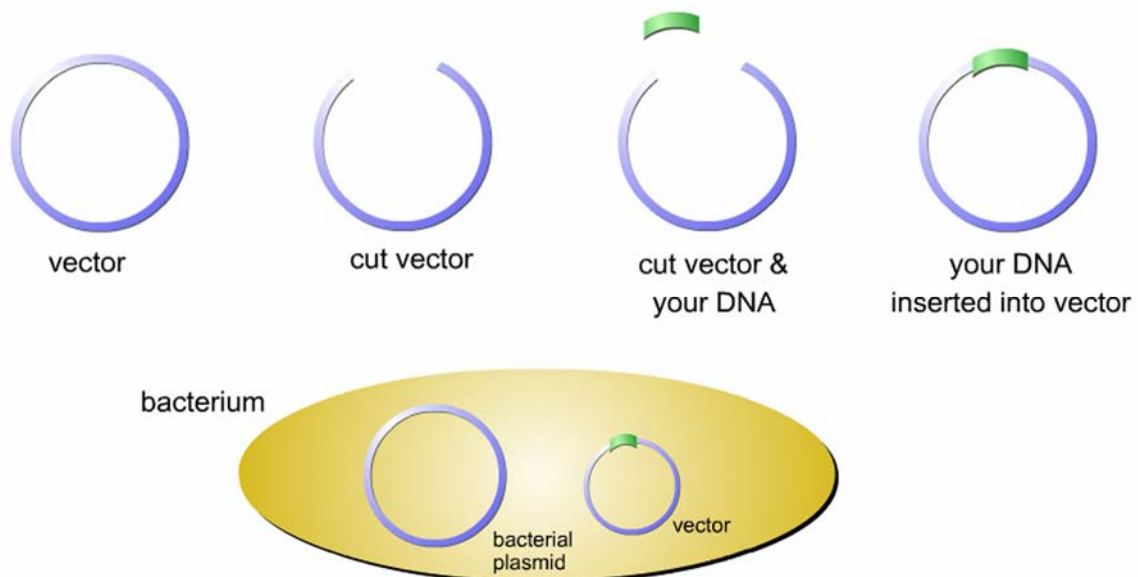


When researchers want to study a protein, they need to purify it from its native organism or produce it artificially. Since most proteins are naturally expressed in relatively low concentrations, it is often advantageous to produce them artificially. The simplest technique used to produce a protein at a high concentration is to transfer the gene that encodes the protein into a bacterium.

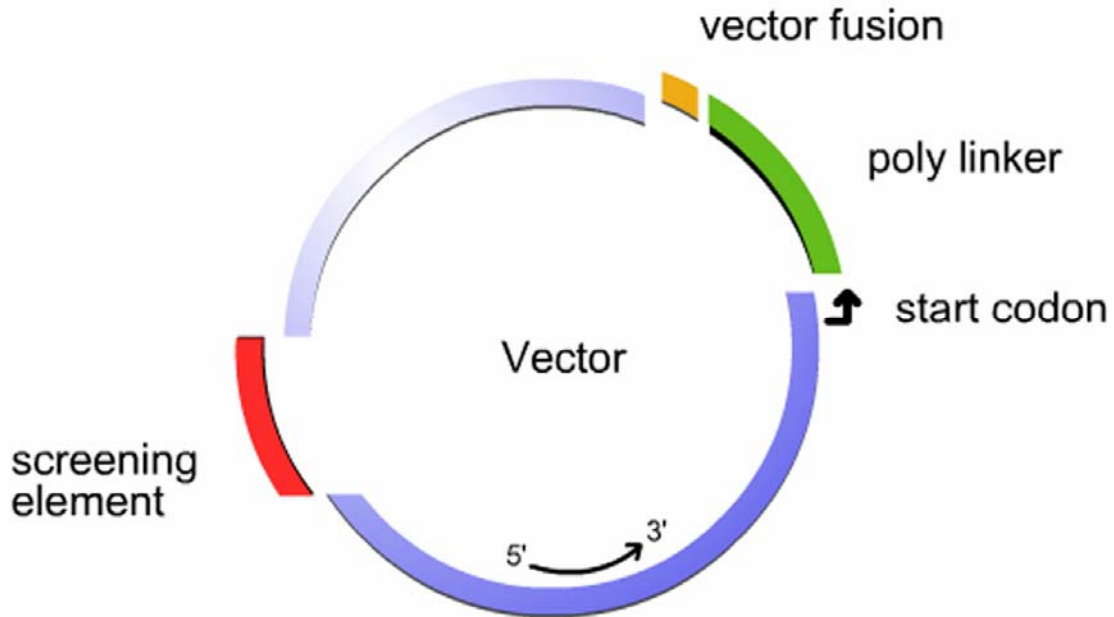
Bacteria are relatively simple organisms which store their DNA in a plasmid (a circular loop of DNA). Bacteria can support more than one plasmid. Researchers insert a gene of interest into an artificial plasmid (often called a vector) which is then inserted into a bacterium.

Vectors are cut open with specialized proteins called restriction endonucleases. Additional DNA, such as the gene encoding your protein, is inserted into a vector after which a specialized protein called ligase seals the vector.

When a gene is translated into a protein, the process is often referred to as “protein expression” as well as “translation”.



Vectors used for cloning have a basic anatomy.



- Poly linker** Region of the vector into which your DNA will be inserted. This region possesses DNA sequences that are recognized by restriction endonucleases and viral recombinases (specialized virus proteins that swap one segment of DNA for another).
- Start codon** This is the point in the vector where translation of your protein will begin. The start codon resides before the poly linker so that your inserted DNA will be translated.
- Vector fusion** Before or after the poly linker, vectors may possess additional DNA that encodes amino acids that are to be translated with your DNA. These additional amino acids often encode short amino acid segments that aid in protein purification.
- Screening element** This region of the vector encodes one or more proteins that allow researchers to identify bacterium that have taken in the vector. This region normally encodes proteins that confer anti-biotic resistance or allow the bacterium to survive in specific conditions.

The biochemical reactions that insert your DNA into a vector require several nano grams of your DNA. Several nano grams of your DNA contains several hundreds of thousands of copies of your DNA.

In order to replicate additional copies of a segment of DNA, researchers use a relatively simple yet brilliant biochemical technique called PCR. PCR is short for "Polymerase Chain Reaction".

"Polymerase" is the name of a specialized protein that replicates double stranded DNA. PCR employs a special form of polymerase commonly referred to as "Taq". Taq is a specialized polymerase that works well at high temperatures.

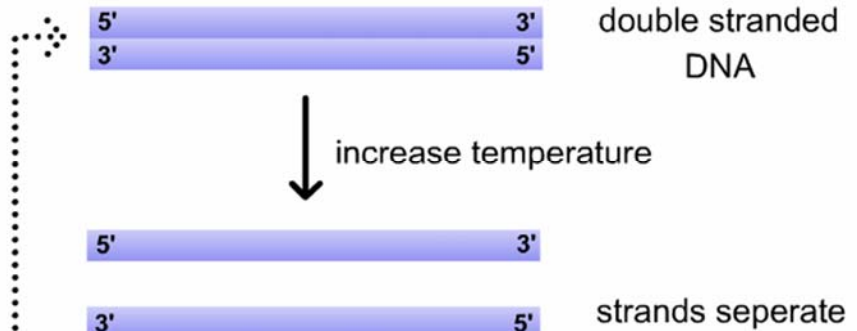
PCR requires small segments of DNA called primers that bind to the beginning and end of the DNA segment that is to be replicated.

The primer that binds to the beginning of the DNA segment is called the 5' or forward primer. The primer that binds to the end of the DNA segment is called the 3' or reverse primer.

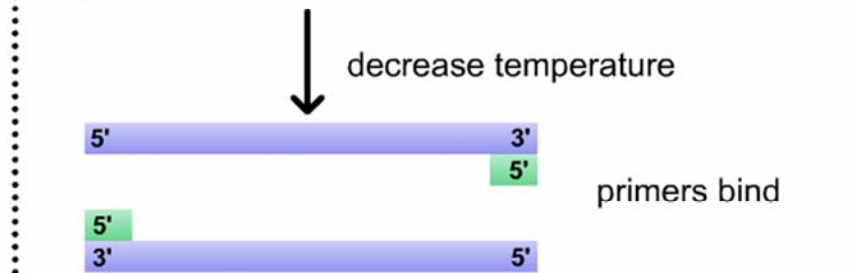


The PCR reaction is carried out in a device called a thermocycler. A thermocycler repeatedly changes the temperature of the PCR reaction.

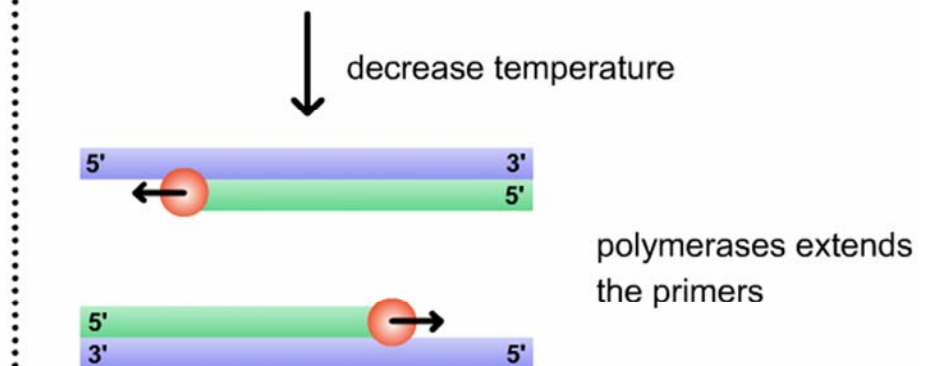
1. The thermocycler increases the temperature of the reaction causing DNA strands to separate.



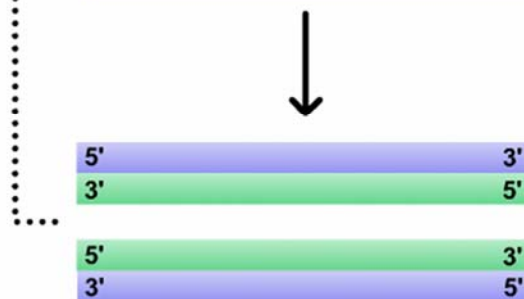
2. The thermocycler then decreases the reaction temperature and allows the PCR primers to bind to the separated strands.



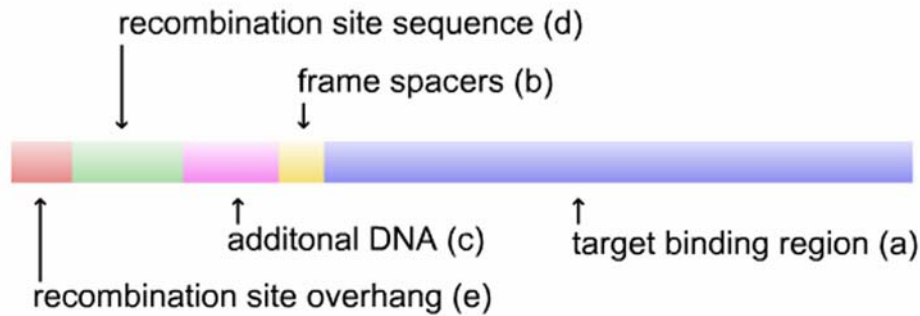
3. The thermocycler then decreases the reaction temperature again and allows the polymerases to recognize the primer - DNA strand complex. The polymerases then replicate the DNA strands.



4. The process is then repeated. The DNA found between the primers is replicated exponentially.



Primers used for cloning have a basic anatomy.



- a. This region binds to your target DNA.
- b. In certain circumstances, it is necessary to shift the reading frame so that the ribosome reads your DNA correctly.
- c. Additional DNA can be added if desired. For example, researchers can add additional DNA that encodes additional amino acids.
- d. In order to insert your DNA into a vector, the vector has to be cut with two restriction endonuclease proteins. Two cutting proteins are used rather than one in order to guarantee that your DNA is inserted in the correct orientation. Your PCR replicated DNA needs to be cut with the same restriction endonucleases as the vector. When restriction endonucleases cut DNA, they leave distinctive staggered ends. These distinctive ends self associate. For example, if one end of your DNA is cut with restriction endonuclease (X), that end will associate with the end of the vector that was also cut with endonuclease (X).
- e. Each restriction endonuclease requires a minimum number of nucleotides on each side of its recognition sequence in order to cleave properly. These additional nucleotides need to be added to primers.

Setting up your vectors.xml file.

PP is unique in that it works directly with your own vector systems. PP requires your vectors to be detailed in an XML formatted data file named "vectors.xml". This file must be in the same directory as the PP application.

XML ( eXtensible Markup Language ) is not a programming language but rather a standardized set of rules used to add structure to data using a system of tags.

XML tags always appear in pairs.  
Each pair has a starting tag and an ending tag.

Starting tags possess the name of the tag between brackets, i.e. <my\_tag>.

Ending tags look almost the same except that they possess a "/" before the tag name, i.e. </my\_tag>.

Information associated with XML tags resides between the starting and ending tags:

```
<my_data> data </my_data>
```

vector file setup:

1. All vector data must be placed between <vectors> tags.

```
<vectors>
```

```
    all of your data
```

```
</vectors>
```

2. The data that describes each of your vectors will be placed between

a pair of <vector> tags.

```
<vectors>
  <vector>
    data detailing your first vector
  </vectors>

  <vector>
    data detailing your second vector
  </vector>
</vectors>
```

3. There are four tags that describe a vector:

<name> Name of your vector.

<fusions> If your vector possesses encoded protein fusions that will be expressed with your target DNA, PP needs to know which terminus of the expressed protein these fusion will be attached to.

1. If your vector encodes no fusions omit the <fusions> tag.

2. If your vector encodes only an amino terminal fusion add:

<fusions>N</fusions>

3. If your vector encodes only a carboxy terminal fusion add:

<fusions>C</fusions>

4. If your vector encodes both an amino and carboxy terminal fusion add:

<fusions>NC</fusions>

<linker> The nucleotide sequence of the vector's poly linker is detailed between the <linker> tags.

If the vector does not encode an amino terminal fusion, the first instance of 'ATG' in the poly linker will be considered the start codon.

For amino terminal fusions, the first three nucleotides of the poly linker will be the first codon following the last codon of the amino terminal fusion DNA.

For carboxy terminal fusions, the last three nucleotides of the poly linker will be the codon preceding the first codon of the carboxy terminal fusion DNA.

<re\_site> Restriction endonuclease sites found in the vector's poly linker sequence are detailed between the <re\_site> tags.

Four tags are used to describe a restriction endonuclease site:

- <name> name of the restriction endonuclease site.
- <start> poly linker nucleotide number upon which the restriction endonuclease sequence starts.
- <stop> poly linker nucleotide number upon which the restriction endonuclease sequence stops.
- <overhang> number of nucleotides before and after the restriction endonuclease site required for proper cleavage.

i.e.

```
<re_site>
  <name>    EcoRI  </name>
  <start>   9     </start>
  <stop>    14    </stop>
  <overhang> 4    </overhang>
</re_site>
```

Example of a simple yet complete vectors.xml file:

```
<vectors>
  <vector>
    <name>pET 14,15A</name>
    <fusions>N</fusions>
    <linker>CATATGGCGAATTCTGCGGGATCCTCTGACTGGAAGC</linker>
    <re_site>
      <name>NdeI</name>
      <start>1</start>
      <stop>6</stop>
      <overhang>8</overhang>
    </re_site>
    <re_site>
      <name>EcoRI</name>
      <start>9</start>
      <stop>14</stop>
      <overhang>4</overhang>
    </re_site>
  </vector>
</vectors>
```

Optional vector.xml entries.

As new cloning technologies emerge, they are often incorporated into the software. These updates often require additional information about the vector systems you are using.

Clonetech's In-Fusion cloning technology.

Clonetech has introduced a recombinant cloning technology, named In-Fusion, that uses a proprietary enzyme that recombines 15 nucleotides (nt) of complimentary DNA. With this technology, researchers can insert PCR products into linearized vectors without the need to digest the PCR product with restriction endonuclease or ligate the PCR product / vector complex. As long as each end of the PCR product has 15 nt homologous to the ends of the linearized vector, the In-Fusion enzyme will insert and ligate the PCR product into the vector.

Since the PCR product is not digested with restriction endonucleases, the most 5' and 3' restriction sites can always be used even if they are found in target sequences. Using the most 5' and 3' site reduces the number of non-native residues between the inserted PCR product and vector encoded fusions.

In-Fusion primer sets can be calculated on the fly by simply selecting the "In-Fusion primers" option at the top of the software's results screen (p 47). Selecting this option will convert all of the previously calculated restriction endonuclease primer sets to In-Fusion primer sets. The software will automatically use the most 5' and 3' cloning sites of each vector. This behavior can be disabled in the Options section. If disabled, the software will use the restriction endonuclease sites used in the initial restriction endonuclease primer set calculations.

In order to calculate In-Fusion primer sets, additional information needs to be added to your vectors.xml file.

Add the following information to each vector record:

<pre\_linker> 15 nt (5' -> 3') before the <linker> sequence

<post\_linker> 15 nt (5' -> 3') after the <linker> sequence

In addition to the <pre\_linker> and <post\_linker> information, the number of nucleotides each restriction site adds to the required 15 nt of homology needs to be defined for each restriction endonuclease record <re\_site>. Use an <infusion\_nt\_count> tag to define the number of nucleotides each restriction endonuclease site contributes to the required 15 nt of homology. The number of nucleotides each site adds is a function of how the site is cleaved. Details on how to determine each site's contribution can be found on Clontech's web site ([www.clontech.com](http://www.clontech.com)).

An example of a vector record if the information required for creating In-Fusion primer sets.

```
<vector>
  <name>pET 14-15A</name>
  <fusions>N</fusions>
  <pre_linker>ATGGGGCCATCACCATCACCATCAC</pre_linker>
  <linker>AGCCATATGGCGAATTCTGCTTCTCTCGAGATC</linker>
  <post_linker>ATCCGGCTGC</post_linker>
  <re_site>
    <name>NdeI</name>
    <start>4</start>
    <stop>9</stop>
    <overhang>8</overhang>
    <infusion_nt_count>4</infusion_nt_count>
  </re_site>
  <re_site>
    <name>EcoRI</name>
    <start>12</start>
    <stop>17</stop>
    <overhang>3</overhang>
    <infusion_nt_count>5</infusion_nt_count>
  </re_site>
</vector>
```

Software walk through.

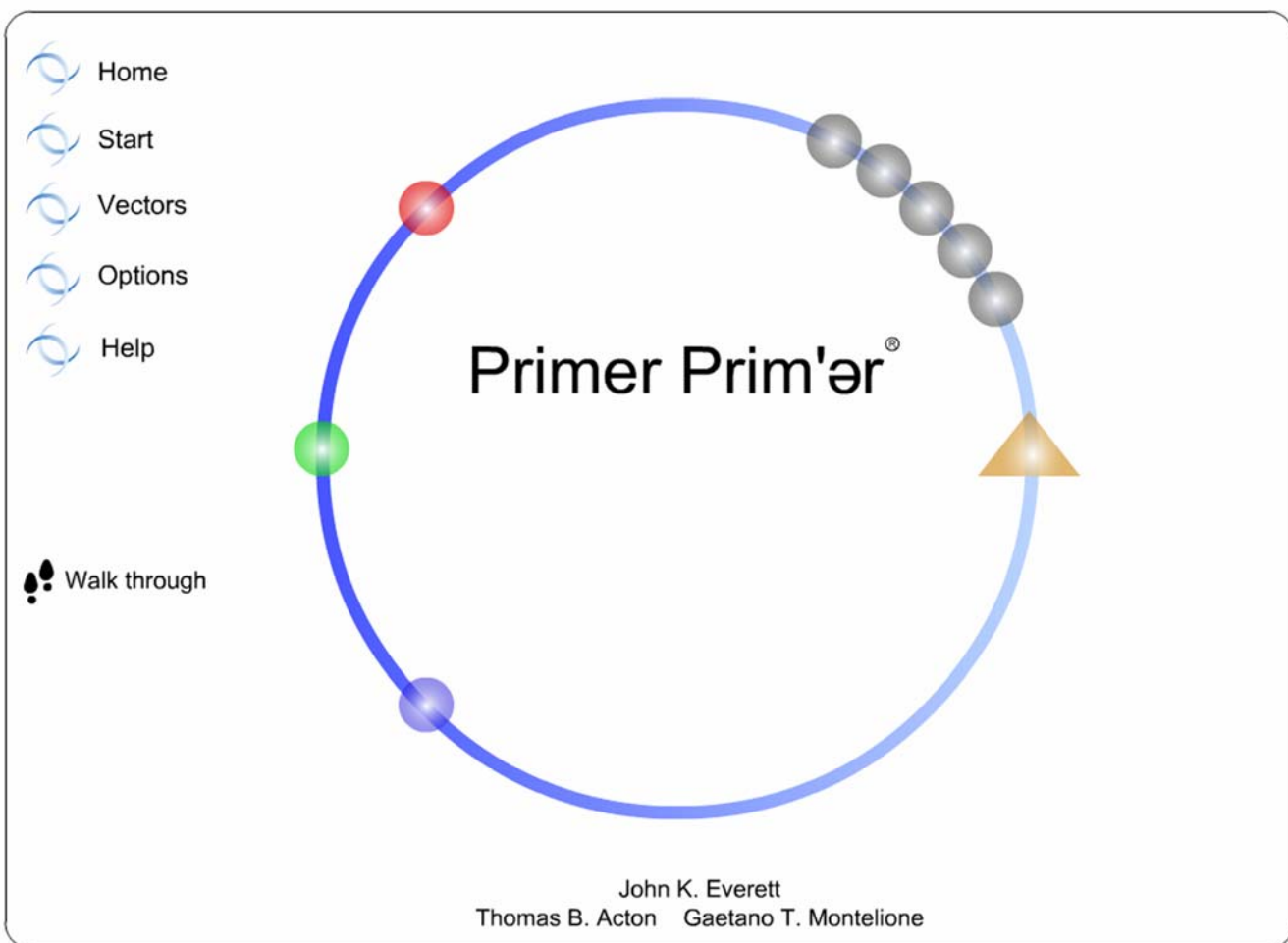
When PP loads, you are brought to the program's home screen.

This screen is just your starting point.

From here, there are four possible routes:

1. press "Start" to start the primer design process.
2. press "Vectors" to view the vectors defined in your vector library file.
3. press "Options" to change PP's options.
4. press "Help" to view the PP's version of this manual.

The "Walk through" button takes you through the software walk through section of this manual.

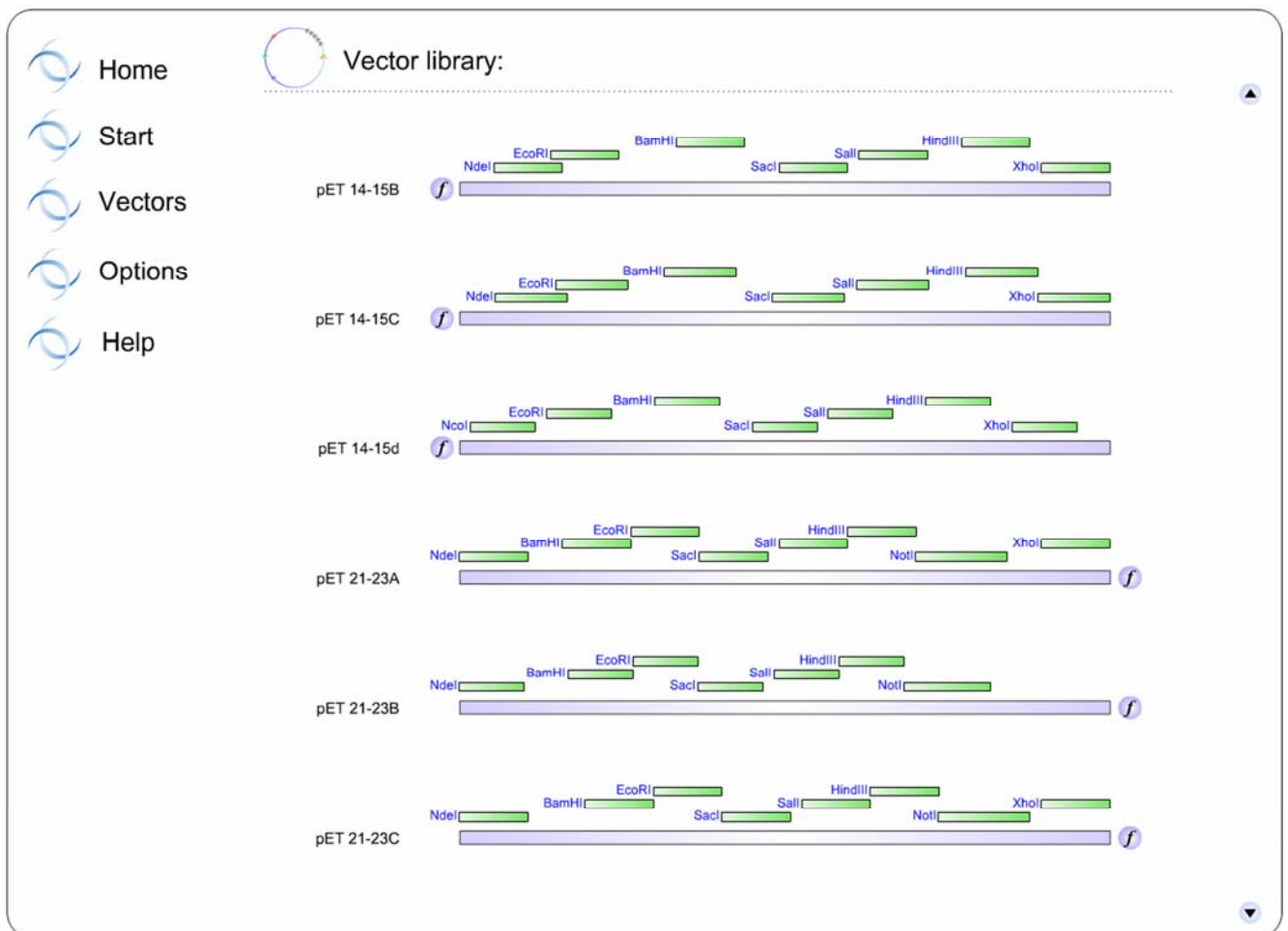


home screen > **vector library**

The vector library is a graphic representation of the vectors detailed in PP's vectors.xml file.

The poly linker of each vector is represented by a blue bar. Restriction endonuclease sites and their position in the poly linker are depicted by green bars.

Vector based fusions are represented by a  icon.



The screenshot shows a web interface for a vector library. On the left is a navigation menu with icons and text for 'Home', 'Start', 'Vectors', 'Options', and 'Help'. The main area is titled 'Vector library:' and contains a list of six vectors: pET 14-15B, pET 14-15C, pET 14-15d, pET 21-23A, pET 21-23B, and pET 21-23C. Each vector is represented by a horizontal bar with a blue segment on the left and a purple segment on the right. Above each bar, green bars indicate the positions of various restriction endonuclease sites, labeled with their names: NdeI, EcoRI, BamHI, SacI, Sall, HindIII, XhoI, and NotI. A small blue circle with a white 'f' icon is placed at the beginning of the bar for pET 14-15B, pET 14-15C, and pET 14-15d, and at the end of the bar for pET 21-23A, pET 21-23B, and pET 21-23C.

home screen > **options**

The options section allows you to customize specific PP parameters and request services.

For example, you can select which primer melting temperature calculation will be employed during the primer design process. Services such as the extraction of domain sequences and the wavering of their boundaries can be requested.

Home Start Vectors Options Help

Options:

Tm Calculation Thermodynamic Cation Conc. 50 mM  
Primer Conc. 50 nM

Codon completion gastnqdekrcmvilhpfyw

Tm range lower 60°C higher 70°C

Stop codon TAG

Automated domain parsing  enable  disable

⚠ Targets must follow a predefined format in order to be parsed automatically.  
Parsing instructions can also be defined in target titles. Consult the manual for details.

exclude the parent  apply parsing instructions to the parent

**Amino terminus**  
number of steps 1  
amino acids changed per step 2  
step direction  
step into and away from the sequence

**Carboxy terminus**  
number of steps 1  
amino acids changed per step 2  
step direction  
step into and away from the sequence

Custom ends Forward 5' 3'  
Reverse 5' 3'

home screen > **start ( target entry )**

Pressing the "start" button on the home screen takes you to

the target entry screen.

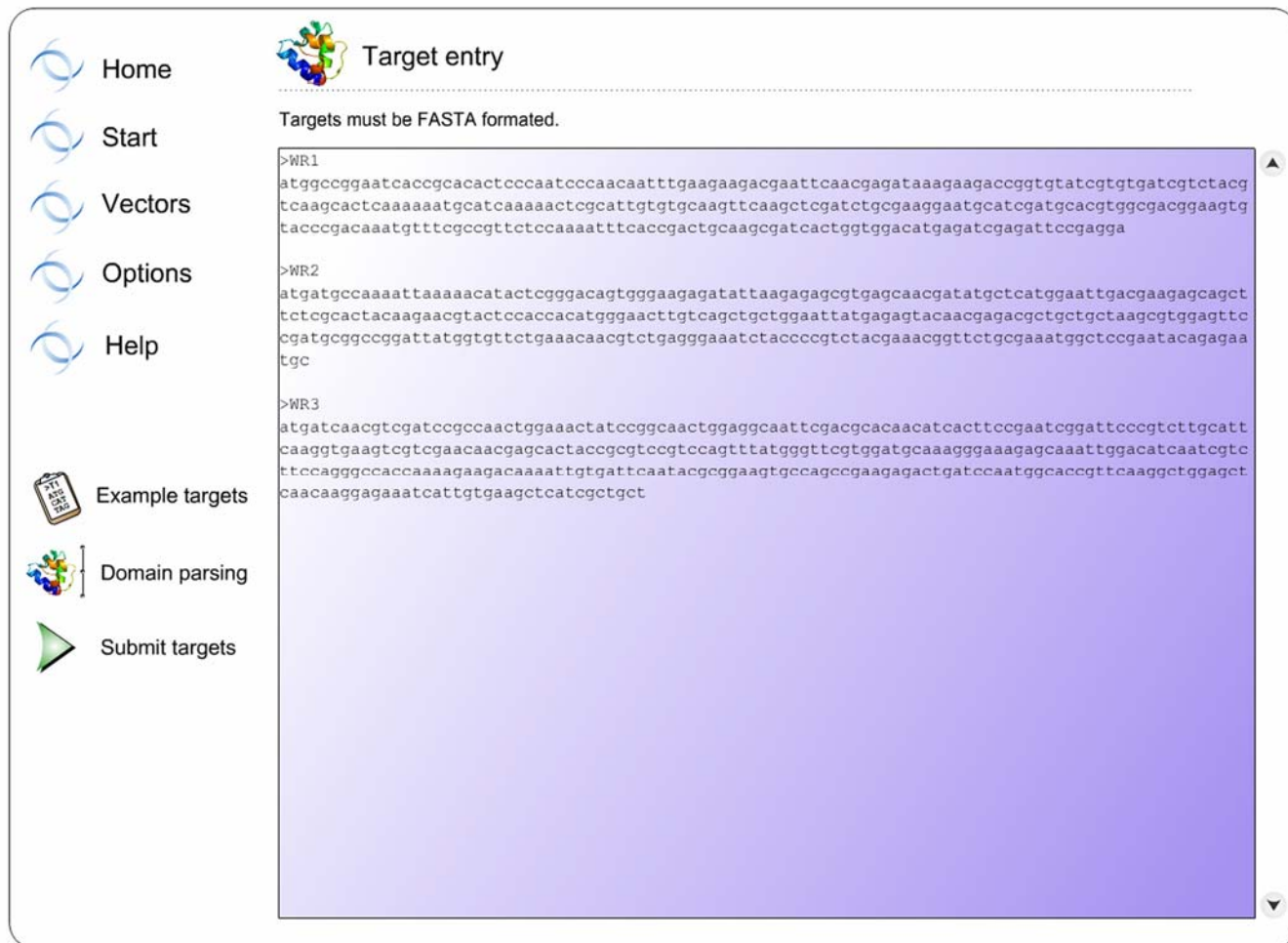
This is the first step in the primer design process.

Enter FASTA formatted targets and then press the "Submit targets" button.

For examples of properly formatted targets, press the "Example targets" button.

Press the "Domain parsing" button to start PP's domain parsing editor.

Once your targets are submitted, they are check for errors and potential problems.

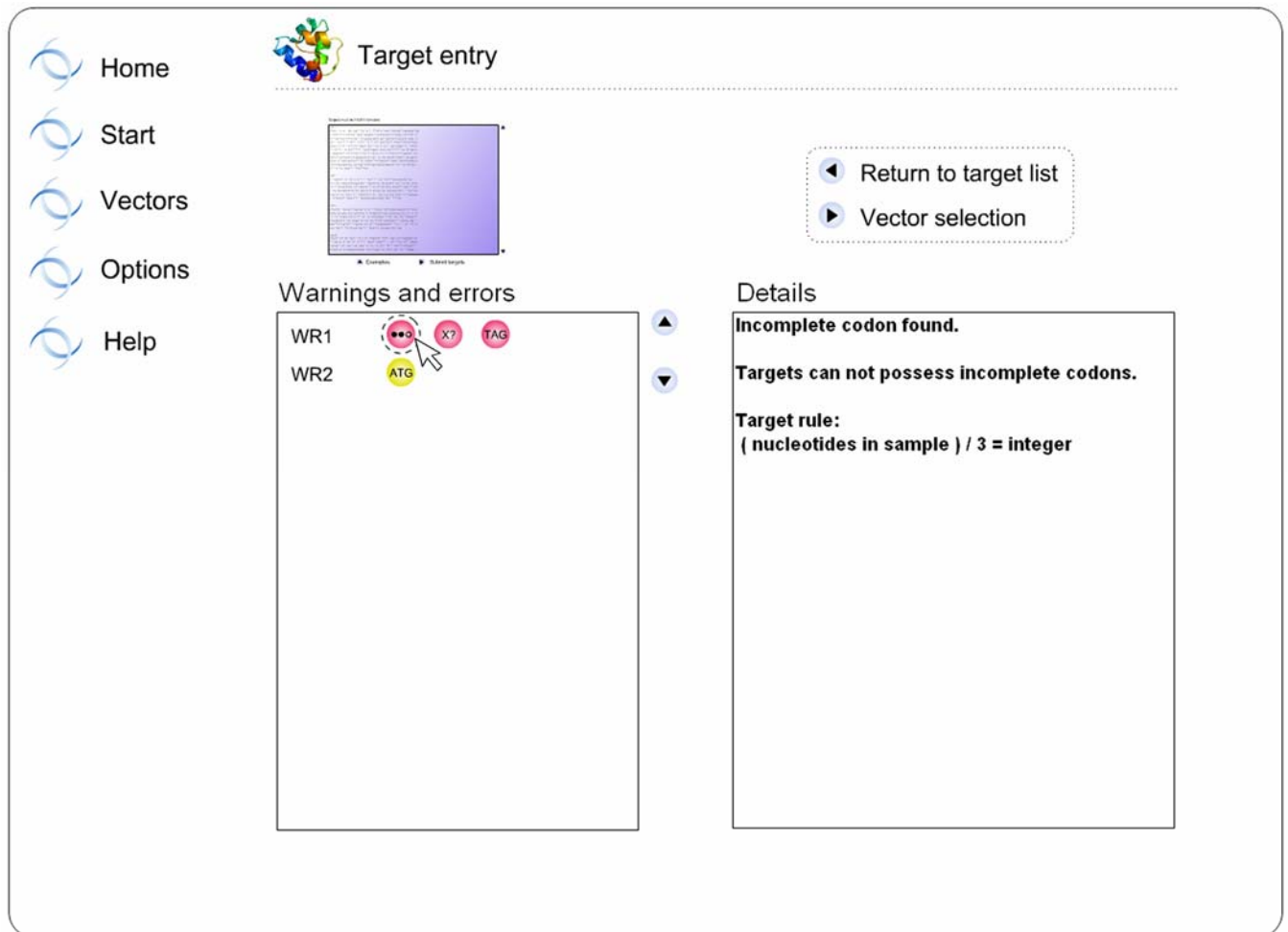


home screen > start ( target entry ) > **target verification**

Errors and warnings are reported in an iconic format in the

"Warnings and errors" window. Placing your mouse over an icon will display the details about why the error or warning is being reported. Details are shown in the "Details" window next to the warnings and errors window.

Press the "Return to target list" button to return to your target list. If you are satisfied with your targets, press the "Vector selection" button.



home screen > start ( target entry ) > **domain parsing**

PP offers a graphical domain editor. Target sequences are represented by spheres color coded based on their hydrophobicity and secondary structure impact. Several domains can be defined per target. PP designs primer sets to extract the defined domains. Domains can also be defined and their boundaries wavered using a target parsing short hand written into a target's title.

Domain parsing (WR2)

Home  
Start  
Vectors  
Options  
Help

● hydrophobic residue  
● hydrophilic residue  
● secondary structure breaking residue ( pro & gly )

domain title: WR2 domain aa 16 - 312  
domain start: 16  
domain end: 312  
( set domain values )  
( save this domain )  
( return to target entry )

saved domains  
● WR2 domain aa 16 - 320  
● WR2 domain aa 16 - 312

home screen > start ( target entry ) > target verification

## > vector selection

For each target a list of available vectors is displayed.  
The 5' and 3' cloning sites are marked with triangular markers.

You can move the markers to different sites with your mouse in order to change the cloning sites. To aid in cloning site selection, the non-native residues (nnr) that will be expressed are listed. Changing the cloning site marker dynamically updates the nnr list. Cloning sites that are present in a target are colored red and are not allowed.

In order to design a primer set for a particular vector, simply check the box next to the name of the vector.

You can move forward or backwards through your target list or automate the vector selection process by pressing the "automate" button to select the same vector(s) for all of your targets.

Target number 5 / 50  
Target name WR5

5' restriction site marker  
3' restriction site marker } Drag marker to change site

Vectors

pET 14-15A

5' site: EcoRI N nnr: H M A N S  
C nnr: —

3' site: XhoI

pET 21-23A

5' site: NdeI N nnr: —

3' site: XhoI C nnr: L E

Target control

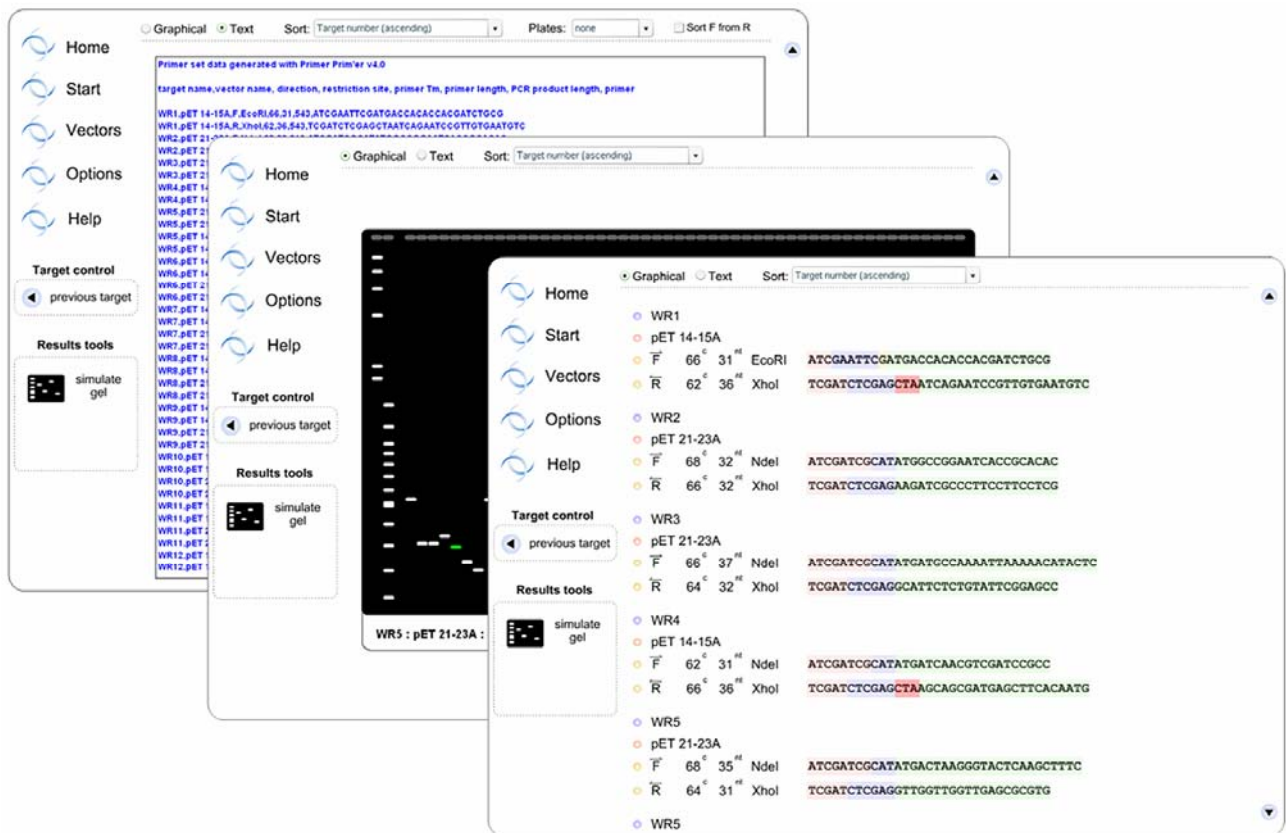
◀ previous target  
▶ next target  
⚙ automate

home screen > start ( target entry ) > target verification

> vector selection > results

Your primer sets are displayed after vectors are selected for each target. There is a graphical format as well as a text format.

Your primer sets can be sorted on several criteria such as target name or PCR product length. Visualization tools such as a simulated agarose gel are available as well.



Software options.

## Tm calculation



Tm Calculation Thermodynamic Cation Conc. 50 mM  
Primer Conc. 50 nM

PP offers four melting temperature calculations for determining the melting temperature ( $T_m$ ) of primer / target complexes.

Simple:  $T_m = 2 \cdot AT + 4 \cdot GC$

Complex:  $T_m = 81.5 + (16.6 \cdot \log[Na]) + (0.41 \cdot \%GC) - (675 / nt)$

Thermodynamic:  $T_m = (dH / dS + 1.987 \cdot \ln(1 / \text{conc})) \cdot ((16.6 \cdot \log[Na]) - 273.2)$

Small mismatch:  $T_m = 81.5 + (0.41 \cdot \%GC) - (675 / N) - \%mismatch$

The complex and thermodynamic formulas require details about the PCR reaction buffer.

\* We recommend using the simple formula since most commercial PCR kits are designed to employ this formula.

## Codon completion



Codon completion `gastnqdekr cmvilhpfyw`

In order to preserve frame with vector based fusions, it is often necessary to add one or two nucleotides to a primer sequence. These additional nucleotides lead to the expression of a non-native residue. Since any of the four standard nucleotides may be used to preserve frame, the resulting non native residue may be tailored to some extent. In the interest of improved solubility and structure determination, the addition of a more innocuous amino acid residue such as serine rather than a large hydrophobic residue such as tryptophan is often advantageous. PP accepts a list of non-native amino acids ( one letter codes ), from most preferred to least.

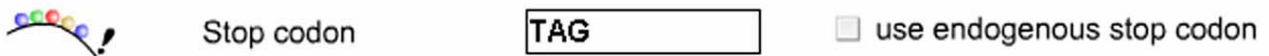
Based on this hierarchy, PP selects the added nucleotides in such a manner as to encode the most favored possible non native residue.

## Tm range



All created primer sets possess a Tm near the center of a given temperature range. The temperature range is defined here. PP will always try to create primers that end with one or more C or G nucleotides. Increasing the Tm range increases the chance that primers will end with a C or G nucleotide. We recommend a Tm range of 60C - 70C while employing the simple 2AT + 4GC Tm formula.

## Stop codon

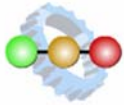


3' primers that are designed for vectors that do not possess a carboxy terminal fusion require the addition of a stop codon.


The stop codon is defined here. More than one stop codon may be used. Entering "TAGTAG" will add two stop codons to 3' primers.

PP removes 3' stop codons from your target list. If a stop codon is needed, the stop codon defined here is added. If you wish to use the stop codon(s) found in your target list, check the "use endogenous stop codon" box. Enabling this feature will tell PP to use the stop codon(s) that come with a target and to include them in the reverse primer's Tm calculation.

## Automated domain parsing



Automated domain parsing  enable  disable

 Targets must follow a predefined format in order to be parsed automatically. Parsing instructions can also be defined in target titles. Consult the manual for details.

exclude the parent  apply parsing instructions to the parent

<u>Amino terminus</u>		<u>Carboxy terminus</u>	
number of steps	<input type="text" value="1"/>	number of steps	<input type="text" value="1"/>
amino acids changed per step	<input type="text" value="2"/>	amino acids changed per step	<input type="text" value="2"/>
step direction	<input type="text" value="step into and away from the sequence"/>	step direction	<input type="text" value="step into and away from the sequence"/>

PP offers the calculation of primer sets that extract a subset of nucleotides from a target sequence. PP refers to this extraction as "target parsing". Target parsing allows for the extraction of domain sequences from sequences that encode multi-domain proteins. Target parsing can also be used to create series of truncations of a target sequence.

Parsing instructions can be written directly in a targets title.

i.e. >title parse: N,50,io,3,2

Read the "Target parsing shorthand" section of this manual for details.

Rather than working with the parsing shorthand, you may opt to use this parsing interface.

The parsing instructions defined here will only be applied to targets that are recognized as domain targets. PP will identify a target as a domain of another target ( parent target ) if it possesses the same name with one additional letter added to the end of the name.



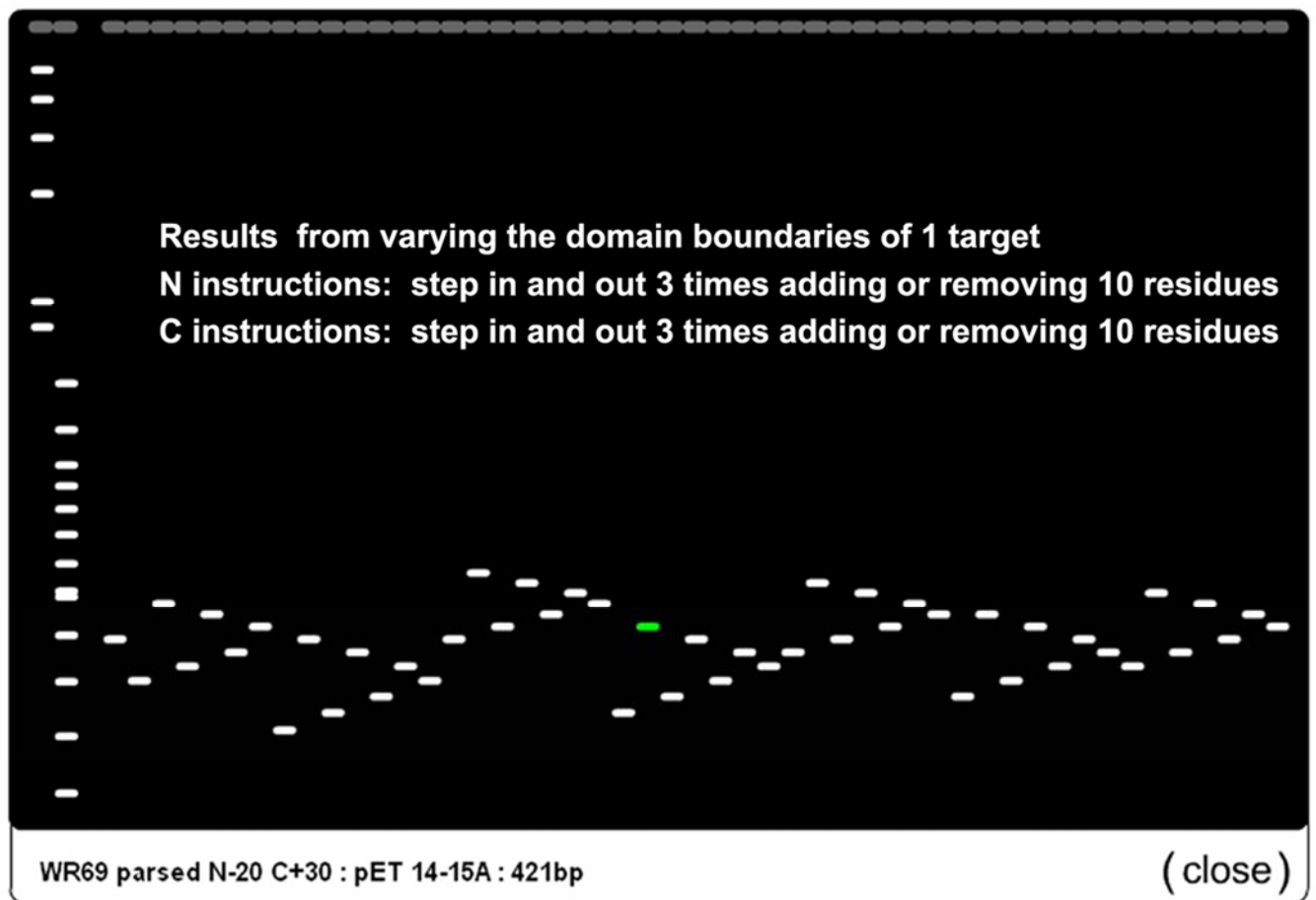
c. step direction: you can instruct PP to step into a target sequence ( truncate ), step away from the target sequence ( elongate ) or step into and away from a target sequence.

If both amino terminal and carboxy terminal instructions are given, All possible combinations of both instruction sets will be created.

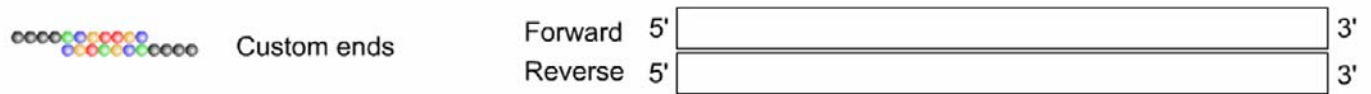
Domain targets generated by the automated domain parsing service  
Share a naming convention:

Domain name **parsed N (+/-) # residues changed C (+/-) # residues changed**

Use the "Simulate gel" tool in the results section to simulate your PCR results. This tool helps visualize your truncations.



## Custom ends



This option provides the ability to create cloning sites that are not found in the vectors.xml file. These custom sites are simply added onto the beginning or end of target sequences.

Enter a ( 5' – 3' ) nucleotide sequence into the text field labeled 'Forward' to create a custom 5' cloning site. Enter a ( 5' – 3' ) nucleotide sequence into the text field labeled 'Reverse' to create a custom 3' cloning site.

The custom sites will appear as gold colored sites for each vector during the vector selection step of the primer design process. Simply drag either the 5' or 3' cloning site marker onto the custom sites in order to employ these sites.

These sites will not be embedded into target sequences and stop codons will not be introduced into reverse primers.

Target number 2 / 3

Target name WR2

5' restriction site marker (blue triangle)

3' restriction site marker (red triangle)

Drag marker to change site

Vectors

Home

Start

Vectors

Options

Help

Target control

previous target

next target

automate

pET 14-15A

5' custom

NdeI

EcoRI

BamHI

SacI

Sall

HindIII

XhoI

3' custom

5' site: NdeI

N nnr: S H

C nnr: -

3' site: XhoI

pET 14-15B

5' custom

NdeI

EcoRI

BamHI

SacI

Sall

HindIII

XhoI

3' custom

5' site: NdeI

N nnr: S H

C nnr: -

3' site: XhoI

## Endonuclease fodder



Endonuclease fodder

5' site	5'	<input type="text" value="CCCG"/>	3'
3' site	5'	<input type="text" value="CCCG"/>	3'

The efficiency of restriction endonuclease enzymes (RE) is often dependent on the presence of additional nucleotides flanking their sequence recognition sites. Since these sites are added to the end of primer sequences, additional nucleotides need to be added to the primers to ensure proper RE cleavage. The number of flanking nucleotides is restriction endonuclease specific and is defined in the vector library file. The type of nucleotides used to flank RE cleavage sites is not terribly relevant to RE efficiency but we have found that using CG rich sequences enhance PCR performance. The option section provides fields for entering nucleotide sequence to be used for 3' and 5' RE flanking sequences.



In-Fusion cloning

Use the most 5' and 3' restriction sites sequences even if they are in the target sequence.

Clontech's In-Fusion cloning technology, described on page 20, does not require restriction endonuclease digestion of PCR products therefore the program automatically uses the most 5' and 3' restriction sites of each vector in order to reduce the expression of non-native residues. Uncheck this option to use user selected restriction sites when calculating In-Fusion primer sets.

Entering your targets.

All targets are to be entered in FASTA format. FASTA format is a simple DNA sequence format with the following guide lines:

1. DNA sequences require a one line title starting with a ">".
2. The next and following lines possess the DNA sequence associated with the title.

example of one target:

```
>my target
ATGCAGTCGATCGATCTAGCACGTCGAT
CACAGCAGTCAGTCGATTTAGCGCATCG
```

example of multiple targets:

```
>first target
ATGCACGTACGTAGCATCGATGCTAGCAT
CAGTCAGTCAGTCGATCGTACTAGCTACA
>second target
ATGCGGCGCGATCGACGATCGATCGATGT
CAGGCACGATCAGTCGATCGATGACACAC
```

If PP is being used to create protein expression constructs, the targets must possess complete codons. The number of nucleotides in each target must be a multiple of three.

Also make sure that your targets do not possess introns.

Enter as many targets as desired into the target entry field shown below.

For examples of properly formatted targets, press the "Example targets" button.

Once your targets are entered, press the "Submit targets" button.

To define domains within a given target, press the "Domain parsing" button to enter PP's domain parsing environment. This is described in detail next ( p37 ).

Target entry

Targets must be FASTA formatted.

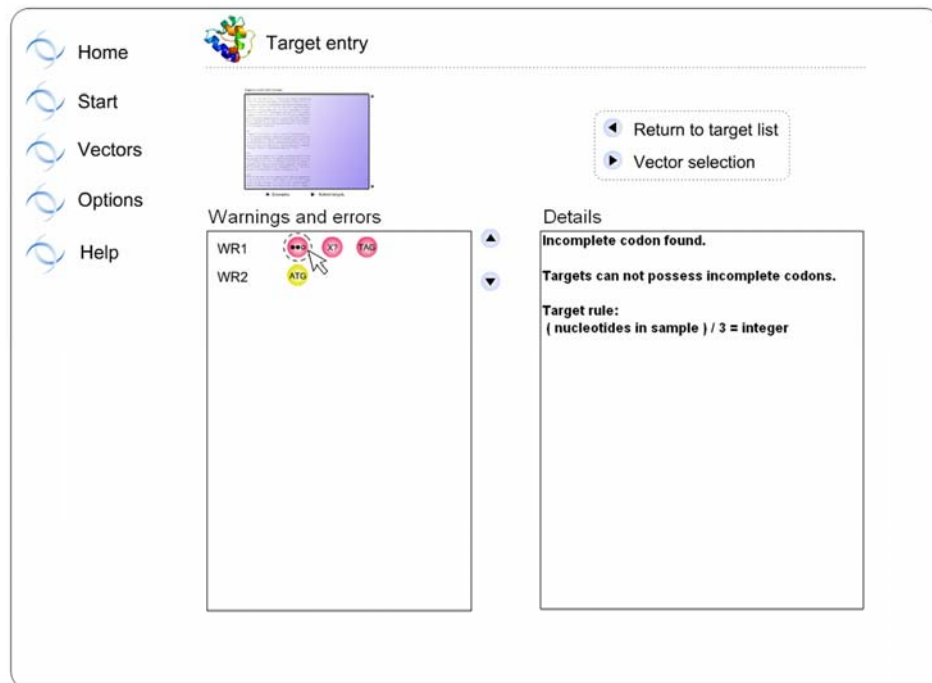
```
>WR1
atggccggaatcacccgcacactcccaatcccaacaattgaagaagacgaattcaacgagataaagaagaccgggtgatcgtgtgatcgtctacg
tcaagcactcaaaaaatgcatcaaaaactcgcatgtgtgcaagttcaagctcgatctgcgaaggaatgcatcgatgcacgtggcgacggaagtg
taccgcacaaatggttcgccgttctccaaaatttcaccgactgcaagcgatcactggtggacatgagatcgagatcccgagga

>WR2
atgatgccaaaattaaaaacatactcgggacagtgggaagagatattaagagagcgtgagcaacgatatgctcatggaattgacgaagagcagct
tctcgactacaagaacgtactccaccacatgggaacttgtcagctgctggaattatgagagtacaacgagacgctgctgctaagcgtggagttc
cgatgcggccggattatggtgttctgaaacaacgtctgagggaaatctaccccgctctacgaaacgggtctgcgaaatggctccgaatacagagaa
tgc

>WR3
atgatcaacgtcgatccgcacactggaactatccggcaactggaggcaattcgacgcacacacatcacttccgaatcggattcccgctcttgcaat
caaggtgaagtcgtcgaacaacgagcactaccgctccgtccagtttatgggttcgtggatgcaaagggaaagagcaaatggacatcaatcgctc
ttccagggccacaaaagaagacaaaattgtgattcaatacgcggaagtgccagccgaagagactgatccaatggcaccgttcaaggtggagct
caacaaggagaaaatcattgtgaaqctcatcgctgct
```

Once targets are submitted, the targets are screened for errors and potential problems. Errors and warnings are reported in an iconic format in the "Warnings and errors" windows. Placing your mouse over an icon will display details about why the error or warning is being reported. Details are shown in the "Details" window next to the warnings and errors window.

Press the "Return to target list" button to return to your target list. If you are satisfied with your targets, press the "Vector selection" button.



icons and their meanings:



Warning: This target does not start with a start codon.



Error: This target possesses an incomplete codon.  
The number of nucleotides in a target must be a multiple of 3.



Error: This target possesses a stop codon that is not the last codon.



Error: This target possesses a nucleotide character that is not an A, T, C or G.

## Domain Parsing.

PP allows you to define protein domains within any of your targets.

1. Enter one or more targets into the target entry field.
2. Press the “Domain parsing” button.
3. If more than one target was entered into the target entry field, you will be asked to select which target you would like to work with.

PP’s Domain parsing environment will be displayed.

Each amino acid in your target will be displayed as a color sphere.

Yellow sphere	hydrophobic amino acid
Blue sphere	hydrophilic amino acid
Red sphere	secondary structure breaker ( Proline, Glycine, Stop )

Home  
Start  
Vectors  
Options  
Help

● hydrophobic residue  
● hydrophilic residue  
● secondary structure breaking residue ( pro & gly )

Domain parsing (WR2)

domain title:   
domain start:   
domain end:   
( set domain values )  
( save this domain )  
( return to target entry )

saved domains

- WR2 domain aa 16 - 320
- WR2 domain aa 16 - 312

Placing your mouse over a sphere will list which residue it is.

Clicking on a sphere will produce a menu for that sphere with the following commands:

Start domain	mark this residue as the start of a domain
End domain	mark this residue as the end of a domain
Remove marker	remove a marker form this residue if present
Cancel	close this menu

You can use these commands to mark the beginning and end of a domain.

Text fields are provided for typing in the beginning and ending residues of a domain. If the text fields are used rather than the menu system, press the “set domain values” button to update the spheres.

Give your domain a title if PP has not already generated one and press the “save this domain” button to send the currently defined domain to the saved domain list ( graphic listing of all saved domains for the current target ).

Click on a saved domain icon to display the saved domain menu. This menu possesses the following commands:

Delete	delete this saved domain
Display	update the spheres to reflect this saved domain
Cancel	close this menu

When you have defined all of the domains for your target, press the “return to target entry” button. This will return you to the target entry screen. The DNA entailed in your saved domains will be appended to your original target list.

Target parsing short hand.

Target parsing instructions can be added directly to target titles.  
Parsing instructions follow the key phrase "parse:".

```
>WR1 parse: N,50,io,5,3 C,150,io,2,3  
atgtagctagctagctagctagctagctagctagctgata  
cagtcgatcgatcgatgctagctagctagctagctagcaa  
cgatcgtattatcgatcgtagctagctagctcgatgcatac
```

```
>WR1A parse: N,WR1,io,5,3  
cagtcgatcgatcgatgctagctagctagctagctagcaa  
cgatcgtattatcgatcgtagctagctagctcgatgcatac
```

Parsing instructions are terminus specific.

Parsing instructions for either terminus must posses 5 parameters.

- (a) terminus,
- (b) title of parent sequence or domain boundary residue #,
- (c) direction ( i | o | io ),
- (d) # of manipulations,
- (e) # of residues changed per manipulation

a. Either N ( amino terminus ) or C ( carboxy terminus )

b. Either the title of the parent sequence ( this target is a subset of a larger target ) or the domain boundary ( amino acid number ).

c. Direction of the manipulation:

i into the domain sequence

o outward from the domain sequence

io into and outward from the domain sequence

d. The number of manipulations ( steps ) from the domain boundary

e. The number of residues to change per manipulation ( step size )

Creating mutagenic primer sets.

Creating small mutations in vector DNA is can be accomplished quickly by using stratagene's® quick change technology or other recombinant technologies.

These strategies employ a forward and reverse primer both of which possess the desired mutation. The mutations are located in the center of the primers and bulges out during annealing. As long as the primers are long enough, this bulging does not effect PCR efficiency.

Primer Prim'er can easily create these mutagenic primers. Add the key word "mutate:" followed by the desired mutation into target titles. Mutations are defined with this format:

**native residue abbreviation    residue number    mutant residue abbreviation**

For example:

For target WR1, mutate phenylalanine 50 to an alanine

```
>WR1    mutate: F50A  
atgcgatcgatcacagtcgatcgatgcatgcatcgtacgatcga  
tcatcgatcacacggcgtatatagctagctagctagctagctaa  
ccggcgtagcatcgatcgatcgatcgatcgatgcatcgatttac
```


Targets that possess the "mutate:" key word in their titles do not appear in the vector selection step since mutagenic primer sets are vector independent.

The small mismatch Tm formula is used to calculate the Tm of mutagenic primers regardless of which Tm formula is selected in the options section.

## Selecting your vectors.

For each target, the list of available vectors is displayed.

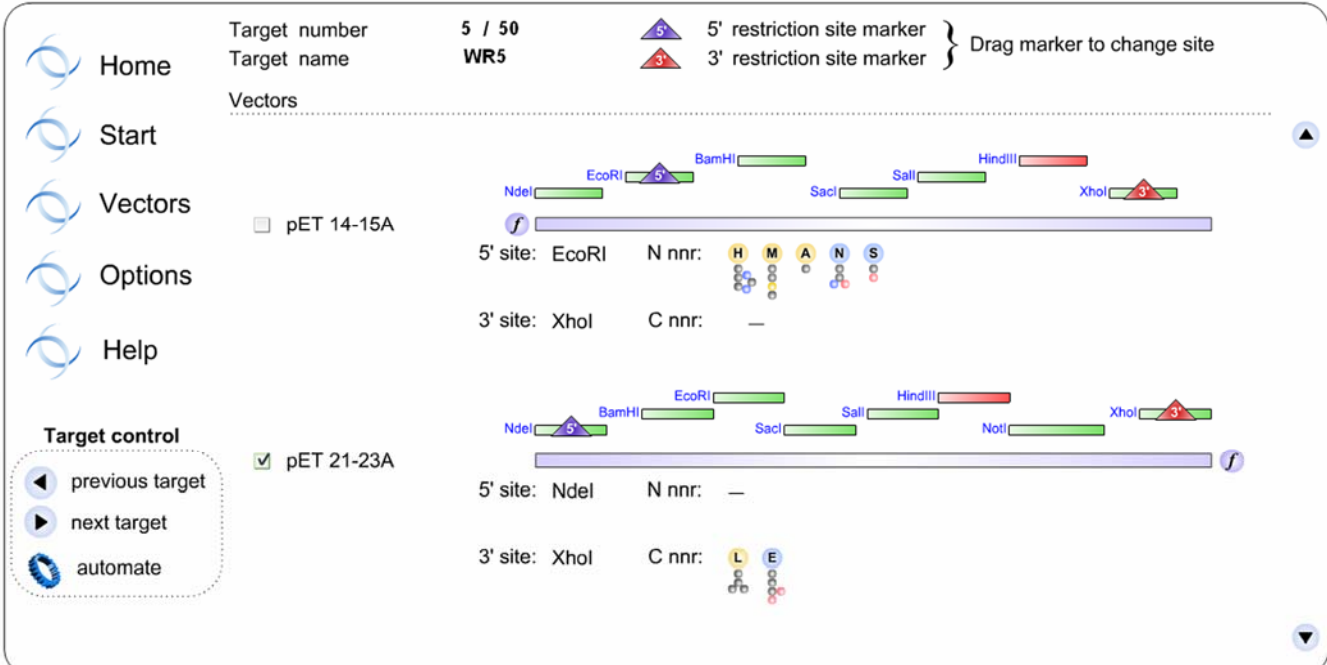
The poly linker of each vector is displayed as a blue bar and the available restriction endonuclease sites are depicted as bars above the poly linker. The sites are colored green if they are not present in the current target and are colored red if they are.

If a vector possesses a fusion, it is depicted with a  icon.

The 5' and 3' restriction endonuclease sites are marked with triangular markers:



You can move the markers to different restriction endonuclease sites with your mouse in order to change the sites.



Target number 5 / 50  
Target name WR5

5' restriction site marker  
3' restriction site marker } Drag marker to change site

Vectors

pET 14-15A

pET 21-23A

5' site: EcoRI N nnr: H M A N S  
3' site: XhoI C nnr: -

5' site: NdeI N nnr: -  
3' site: XhoI C nnr: L E

To aid in restriction endonuclease site selection, the non-native residues (nnr) that will be expressed are listed.

The nnr for each terminus is listed:

N nnr ( Amino terminal non-native residues )  
 C nnr ( Carboxy terminal non-native residues )

Changing the restriction endonuclease site markers dynamically updates the nnr lists.

The labels of the non-native residues are color coded:

yellow: predominantly hydrophobic residue  
 blue: predominantly hydrophilic residue

a -- is displayed if there are no nnr for a particular termini.

Simply check the box next to the name of a vector in order to design a primer set for it.

Target number 5 / 50  
 Target name WR5

5' restriction site marker  
 3' restriction site marker } Drag marker to change site

Vectors

pET 14-15A

5' site: EcoRI N nnr: H M A N S  
 3' site: XhoI C nnr: -





pET 21-23A

5' site: NdeI N nnr: -  
 3' site: XhoI C nnr: L E

Target control

◀ previous target  
 ▶ next target  
 ⚙️ automate

To move from one target to the next ( or previous ), use the "Target control" panel.

-  previous target      Move to the previous target
-  next target      Move to the next target
-  automate      \*Automate the vector selection process
-  display primers      Display your primer sets

\* When pressed, PP selects the currently selected vectors for each additional target while choosing the most possible 3' and 5' restriction sites.



The primer sets can be sorted on different criteria. Pressing the “sort results” button located atop of the results screen displays the available sorting options.

Grouping options

- do not group targets
- group targets according to NESG organism identifiers  
( organisms are identified by the first or first two letters of target titles )

Sorting options

- Sort by target number
- Sort titles A - Z
- Sort titles Z - A
- Sort PCR result length smallest to largest
- Sort PCR result length largest to smallest
- Sort PCR result length smallest to largest using staggering

( sort ) ( exit )

Calculated primer sets can be grouped. The software can currently group results into NESG (North East Structural Genomics consortium) organism groups which are defined in target titles. In order to include a target in a particular group, start the target title with an organism identifier found in appendix A.

If targets are grouped, then the sorting function is applied within each group.

The results can also be viewed in a text format by changing the results view to "Text".



Primer set data generated with Primer Prim'er 5.4.1

target name,vector name, direction, restriction site, primer Tm, primer length, PCR product length, re sites in target, primer

WR1,pET 14-15A,F,NdeI,66,32,544, NcoI,CCCGCCCGCATATGACCACACCACGATCTGCG  
WR1,pET 14-15A,R,XhoI,62,36,544, NcoI,GCCCGCTCGAGCTAATCAGAATCCGTTGTGAATGTC  
WR2,pET 14-15A,F,NdeI,68,32,313, EcoRI,CCCGCCCGCATATGGCCGGAATCACCGCACAC  
WR2,pET 14-15A,R,XhoI,66,35,313, EcoRI,GCCCGCTCGAGCTAAAGATCGCCCTTCCTTCCTCG  
WR3,pET 14-15A,F,NdeI,66,37,313,,CCCGCCCGCATATGATGCCAAAATAAAAACATACTC  
WR3,pET 14-15A,R,XhoI,64,35,313,,GCCCGCTCGAGCTAGCATTCTCTGTATTGGAGGCC  
WR4,pET 14-15A,F,NdeI,62,31,346, SacI,CCCGCCCGCATATGATCAACGTCGATCCGCC  
WR4,pET 14-15A,R,XhoI,66,36,346, SacI,GCCCGCTCGAGCTAAGCAGCGATGAGCTTCACAATG  
WR5,pET 14-15A,F,NdeI,68,35,298, HindIII,CCCGCCCGCATATGACTAAGGTAAGCTCAAGCTTTC  
WR5,pET 14-15A,R,XhoI,64,34,298, HindIII,GCCCGCTCGAGCTAGTTGGTTGGTTGAGCGCGTG  
WR6,pET 14-15A,F,NdeI,64,33,238, BamHI,CCCGCCCGCATATGTTTTCTCTCCACGACAAGG  
WR6,pET 14-15A,R,XhoI,68,38,238, BamHI,GCCCGCTCGAGCTAGGAAGTACTTTGTGTGACTCT  
WR7,pET 14-15A,F,NdeI,66,33,205,,CCCGCCCGCATATGCTCTACACAATTGCTGCCG  
WR7,pET 14-15A,R,XhoI,62,35,205,,GCCCGCTCGAGCTAGTAAGAGCCCAATCAGCATC  
WR8,pET 14-15A,F,NdeI,64,33,541, HindIII,CCCGCCCGCATATGTCTGTCCAATTGTCTGATG  
WR8,pET 14-15A,R,XhoI,66,34,541, HindIII,GCCCGCTCGAGCTAGGCAACTTCAGCGGCAGCTC  
WR9,pET 14-15A,F,NdeI,66,33,536, XhoI BamHI,CCCGCCCGCATATGCCAGACAACCAAGGATC  
WR9,pET 14-15A,R,HindIII,68,37,536, XhoI BamHI,CCGAAGCTTCTAGTGAAACTCAAAAACCTCATTCTTC  
WR10,pET 14-15A,F,NdeI,64,33,440, XhoI EcoRI,CCCGCCCGCATATGAACTCCATCGTCAATGGAC  
WR10,pET 14-15A,R,HindIII,62,34,440, XhoI EcoRI,CCGAAGCTTCTATTGGAATTCCTTCACTCTCG  
WR11,pET 14-15A,F,NdeI,64,32,298, SacI,CCCGCCCGCATATGGCAACCTCAACATCCACC  
WR11,pET 14-15A,R,XhoI,64,35,298, SacI,GCCCGCTCGAGCTAATCTTGGTCTTCTCTCTCC  
WR12,pET 14-15A,F,NdeI,66,34,217,,CCCGCCCGCATATGATACTAGGAGATCCAAGACG  
WR12,pET 14-15A,R,XhoI,66,38,217,,GCCCGCTCGAGCTAAGAACTTGTTCAGCAGCAATTC  
WR13,pET 14-15A,F,NdeI,60,33,334,,CCCGCCCGCATATGCAACAAAATATCCGAATCC  
WR13,pET 14-15A,R,XhoI,66,37,334,,GCCCGCTCGAGCTAGTAGTGGTTGTGTGCGCAAATTC  
WR14,pET 14-15A,F,NdeI,66,34,344, XhoI SacI,CCCGCCCGCATATGAGCCACACTCAACAACAAAC  
WR14,pET 14-15A,R,HindIII,66,34,344, XhoI SacI,CCGAAGCTTCTAATTAGAGCTCTCCGGCTCTTCG  
WR15,pET 14-15A,F,NdeI,64,31,499, NcoI SacI,CCCGCCCGCATATGGGGGAGGGCCAAATACC  
WR15,pET 14-15A,R,XhoI,64,35,499, NcoI SacI,GCCCGCTCGAGCTAATGATGTCATGGTGTCCACC  
WR16,pET 14-15A,F,NdeI,64,34,436,,CCCGCCCGCATATGGAGTACTTTACATCCATC  
WR16,pET 14-15A,R,XhoI,66,36,436,,GCCCGCTCGAGCTAAGCTGAAATACGGGCAGACTTC  
WR17,pET 14-15A,F,EcoRI,64,31,501, HindIII SacI NdeI,CCCGAATTCGATGCTTCCCACTTCGAGACC  
WR17,pET 14-15A,R,XhoI,64,40,501, HindIII SacI NdeI,GCCCGCTCGAGCTACTGTAGAATTTGAATTTAACAATTA  
WR18,pET 14-15A,F,NdeI,64,34,518, XhoI,CCCGCCCGCATATGTCTCACATTTTTGATGCGTCT

The text formatted results are comma delimited according to the key listed on top of the text window. Text formatted results can be divided into plates by selecting the desired plate division from the plates menu.

Text formatted results can also be sorted on primer direction. Checking the "Sort F from R" option will list the forward (5') primers separately from the reverse (3') primers.

PP can simulate how your PCR results will look on an agarose gel. Press the "simulate gel" button in the "Results tools" section to view a simulated gel.

The NewEngland Biolabs<sup>®</sup> lamda digest and 100bp DNA ladders are simulated in the first two lanes and your PCR results are simulated in the following lanes.

Place your mouse over any of the simulated bands to view information about the band.

Band information format:

Target name : vector name : PCR product length

The screenshot shows a web-based interface for simulating a gel. On the left is a navigation menu with links for Home, Start, Vectors, Options, and Help. Below the menu are sections for 'Target control' (with a 'previous target' button) and 'Results tools' (with a 'simulate gel' button). The main area displays a simulated gel image with multiple lanes. A single band in the third lane from the left is highlighted in green. At the top of the interface, there are radio buttons for 'Graphical' (selected) and 'Text', and a 'Sort:' dropdown menu set to 'Target number (ascending)'. At the bottom of the gel image, the text 'WR222 parsed N+30 C-20 : pET 14-15A : 421bp' is displayed, along with a '(close)' button.

## Primer creation notes:

1. A stop codon will automatically be added to reverse primers unless they are designed for vectors with a carboxy terminal fusion.
2. If possible, PP will embed cloning sites into target sequences in order to design the smallest possible primers. If a vector possesses a fusion, cloning sites will only be embedded if the fusion reading frame is preserved.
3. When designing primers PP will first look at all possible primers that fall within your  $T_m$  range which is defined in the options section. If none of the possible primers end with a C or G, the primer closest to center to the  $T_m$  range is chosen. If one or more of the primers within the  $T_m$  range does end with a C or G, the primer closest to the center of the  $T_m$  range is chosen.

## Appendix A.

NESG organism abbreviations used in target titles.

A. sp (As)	E. carotovora (Ew)
A. pernix (X)	E. coli (E)
A. tumefaciens (At)	F. nucleatum (N)
A. aeolicus (Q)	G. gallus (Gg)
A. thaliana (A)	G. sulfurreducens (Gs)
A. fulgidus (G)	H. ducreyi (Hd)
B. cereus (Bc)	H. influenzae (I)
B. halodurans (Bh)	H. sp (Hs)
B. subtilis (S)	H. pylori (P)
B. thuringiensis (Bu)	H. sapiens (H)
B. fragilis (Bf)	H. cytomegalovirus (C)
B. thetaiotaomicron (Bt)	L. plantarum (Lp)
B. henselae (Bn)	L. lactis (K)
B. longum (Bl)	L. pneumophila (Lg)
B. bronchiseptica (Bo)	L. monocytogenes (Lm)
B. parapertussis (Bp)	M. thermoautotrophicum (T)
B. pertussis (Be)	M. jannaschii (Mj)
B. burgdorferi (Bb)	M. maripaludis (Mr)
B. taurus (Ba)	M. mazei (Ma)
B. melitensis (L)	M. herpesvirus (Mh)
C. elegans (W)	M. musculus (Mm)
C. jejuni (B)	M. bovis (Mb)
C. crescentus (Cc)	M. genitalium (Mg)
C. tepidum (Ct)	M. pneumoniae (Mp)
C. violaceum (Cv)	N. meningitidis (M)
C. acetobutylicum (Ca)	N. europaea (Ne)
C. perfringens (Cp)	Other (O)
C. diphtheriae (Cd)	P. gingivalis (Pg)
D. radiodurans (Dr)	P. aeruginosa (Pa)
D. vulgaris (Dv)	P. putida (Pp)
D. melanogaster (F)	P. syringae (Ps)
E. cuniculi (Eu)	P. furiosus (Pf)
E. faecalis (Ef)	P. horikoshi (J)

R. solanacearum (Rs)  
R. norvegicus (Rn)  
R. virus (Rv)  
R. palustris (Rp)  
S. cerevisiae (Y)  
S. cholerae (Sc)  
S. typhimurium (St)  
S. pombe (Sb)  
S. oneidensis (So)  
S. flexneri (Sf)  
S. aureus (Z)  
S. epidermidis (Se)  
S. agalactiae (Sa)  
S. mutans (Sm)  
S. pneumoniae (Sp)  
S. pyogenes (D)  
S. avermitilis (Sv)  
S. coelicolor (R)  
S. solfataricus (Ss)  
T. acidophilum (Ta)  
T. volcanium (Tv)  
T. maritima (V)  
T. thermophilus (U)  
U. urealyticum (Uu)  
V. cholerae (Vc)  
V. parahaemolyticus (Vp)  
X. axonopodis (Xa)  
X. campestris (Xc)  
X. fastidiosa (Xf)