GeneDesign B2.0 Manual

About GeneDesign

GeneDesign is a web-based program for the design of synthetic genes. It consists of several modules that automate the tasks associated with the manipulation of synthetic sequences.

There are many ways to use GeneDesign. Probably the most common is to start with the protein sequence of an interesting gene and proceed through reverse translation to oligo design. This path is explored in the Design a Gene section of the manual.

Each GeneDesign module can be accessed and used individually, as well. Each module is covered in the GeneDesign Modules section of the manual.

Every module has help text that appears at the top of the screen and contains the basic requirements for the use of the module. The help text will also bring to light user errors in the process and help diagnose design flaws in the synthetic sequence.

The list of restriction enzymes GeneDesign currently recognizes is non-redundant; only one isoschizomer for each type of site is used. Please see Rebase for other available isoschizomers.

The GeneDesign manual can be downloaded in PDF form.

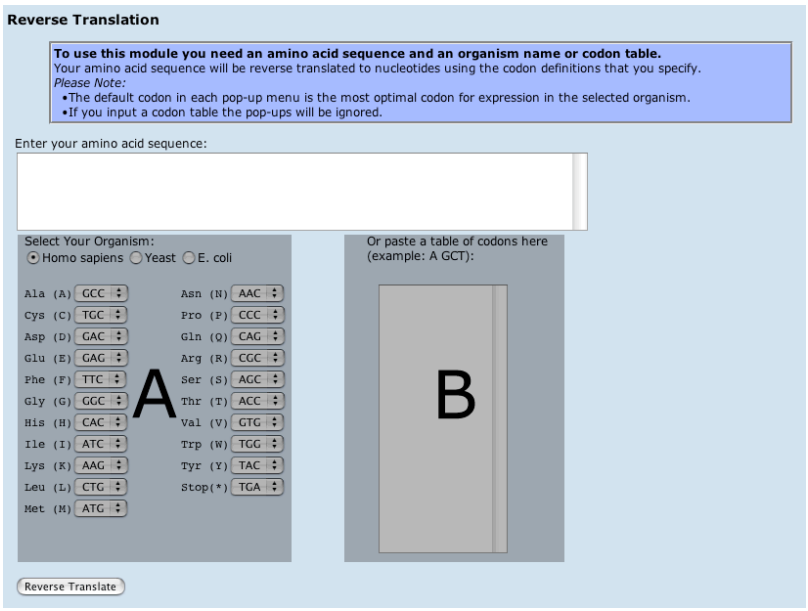
Design a Gene

The Design a Gene pathway is designed for the simple and rapid generation of oligos encoding a synthetic gene from a protein sequence. GeneDesign will perform stepwise modifications to an amino acid sequence to produce a codon-optimized nucleotide sequence that codes for the specified protein. The modified nucleotide sequence can have dozens of customizations that will allow users to proceed to complex experimental work designed to quickly elucidate the structure and function of the gene in question. The final nucleotide sequence will be comprised of segments linked by restriction sites, allowing users to swap synthetic segments with native segments or to easily introduce mutations, in order to help determine which regions are necessary for the function of the gene.

The first module in the pathway is Reverse Translation, followed by Silent Site Insertion and Oligo Design.

Design a Gene: Reverse Translation

The first module you use on the Design a Gene path is Reverse Translation. This is where you will input your protein sequence and choose the codons that will be used in the nucleotide sequence. A single codon will be assigned to each amino acid by default. GeneDesign does not allow multiple codons to be assigned to the sequence at this stage of the gene design.



First, paste your sequence in amino acids into the text area at top. To define the codon usage you can select one of the four common organisms from the radio buttons. (A) Choosing an organism will cause the pop-up boxes to automatically select the codon with the highest RSCU value in highly expressed genes in that organism (pubmed). You can then edit the selection if you wish.

You may also paste your own custom codon table into (B). The format is "AA space codon line break", so that the first line of a valid table could be "M ATG" and the second line "L CTG", for example. Any input in the codon table will supercede entries in the pop-up boxes. Your 21-codon input will be labeled the optimal codons. This version of GeneDesign uses the standard genetic code.

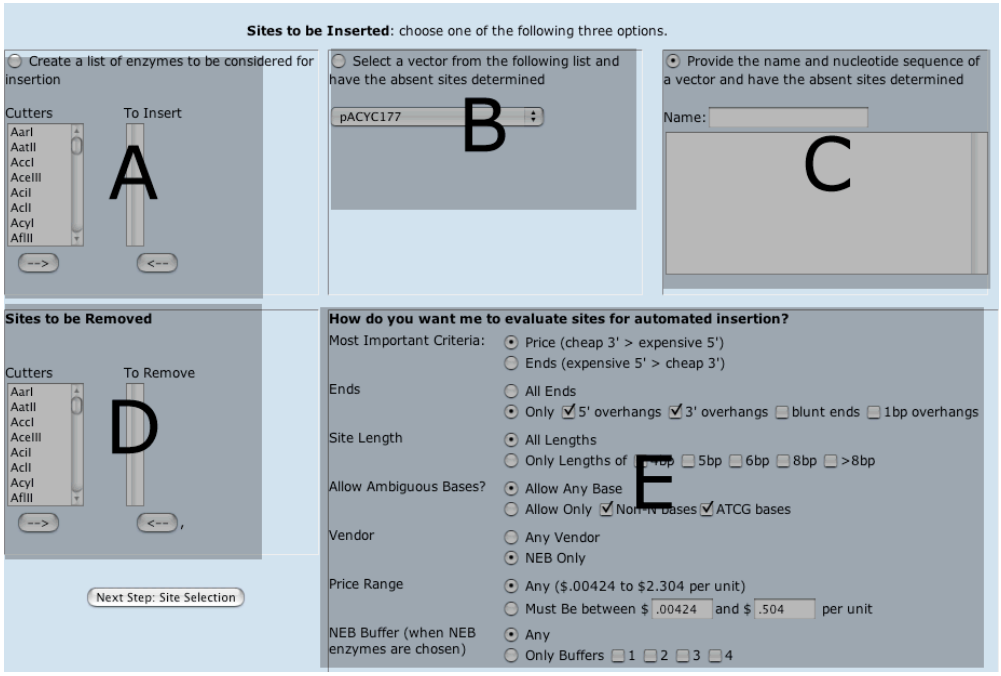
Once the codons have been defined to your satisfaction, hit "Reverse Translate". Your new nucleotide sequence will appear. There will be several buttons at the bottom that allow you to take this nucleotide

sequence to other modules. To follow the Design a Gene path, click "Silent Site Insertion".

Design a Gene: Silent Site Insertion

The second step on the Design a Gene path is the Silent Site Insertion module. This is where you will choose the vector you plan to use to carry your synthetic gene. Alternatively, you can define a list of restriction enzymes to be considered for insertion. Vector choice is an extremely important step in the process. The purpose is to instruct the program on what restriction sites you DO NOT WANT to use as special "landmark restriction sites" in your synthetic sequence, namely those already present in the vector. If you choose a very large vector with many sites in it, the program will not have any sites left to play with. Therefore, if you have a choice, use the smallest vector available to you. In practice pUC18/19 work well; derivatives lacking the multiple cloning site work even better.

Your input amino acid and first-pass (optimized) nucleotide sequences will appear at the top of the screen. Next you must choose the sites to be inserted. If you specify a vector sequence, all sites in the vector are automatically removed from consideration as landmark restriction sites by the program. There are three ways to choose eligible sites.



You may (A) simply select the sites from a list. Click on the site name in the left column and then the arrow button at the bottom of the column to move that site to the right hand column. Repeat the process on the right side to remove it. Only sites in the right column will be considered. Or you may (B) select the name of the vector you will be

using from a pull-down menu. GeneDesign has the sequences of these vectors and will determine which enzymes are absent from this vector automatically for you. Lastly, you may (C) provide a vector sequence and a name for that vector. GeneDesign will parse the sequence and determine which enzymes are absent from your vector.

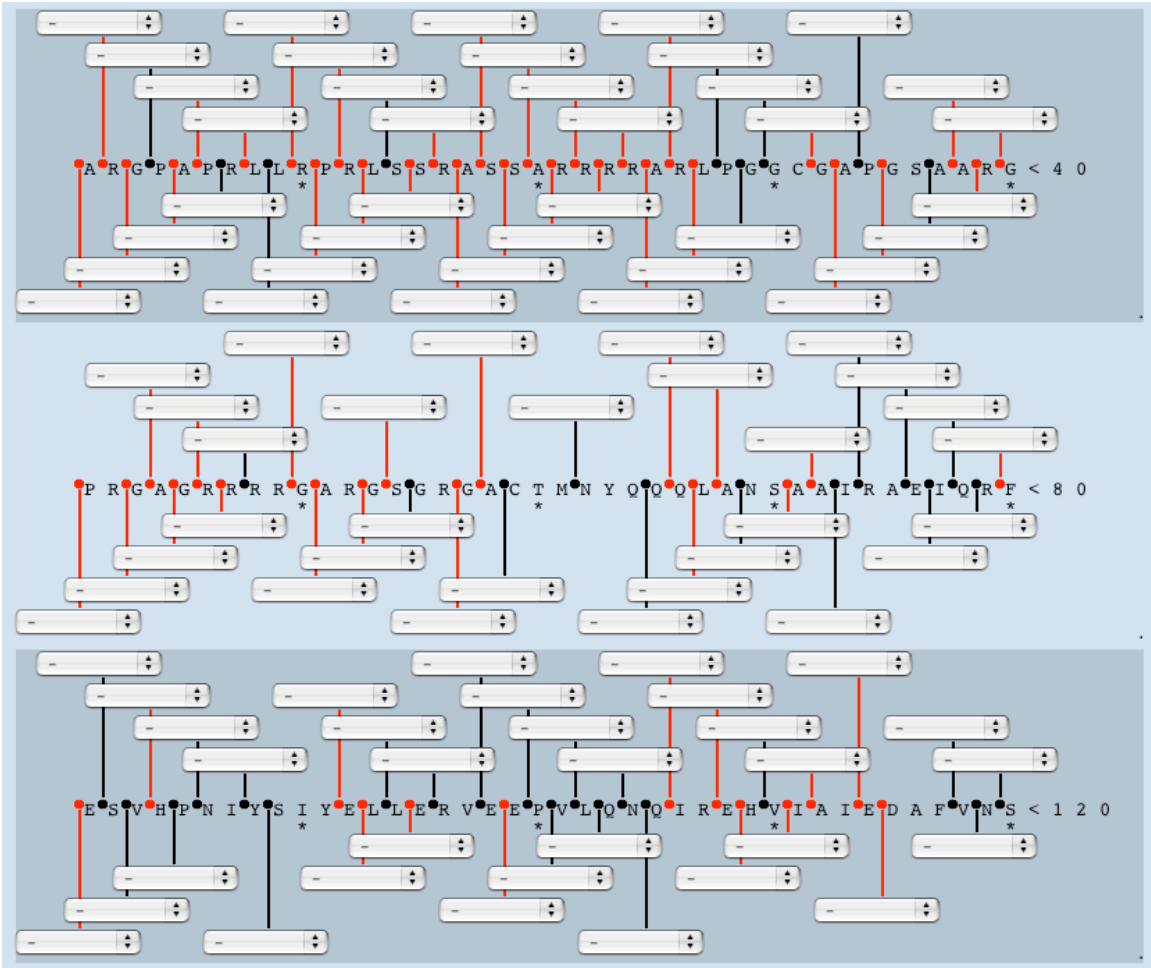
You may in addition specify a list of enzymes that must not appear in the final sequence. (D) Click on the site name in the left column and then the arrow button at the bottom of the column to move that site to the right hand column. Repeat the process on the right side to remove it. Only sites in the right column will be processed. The sequence will be searched for these sites and if they are found in your first-pass synthetic sequence, they will be removed. These sites will not be considered for landmark restriction sites.

At (E) you can define the ranking criteria for sites to be inserted. We have found the default criteria to be more than adequate for most purposes. For more information about ranking enzymes, see Enzyme Choosing.

Once you have selected the vector and/or the sites you can click "Next Step" to move on to the next step in site selection.

Design a Gene: Silent Site Insertion, part 2

The second step of Silent Site Insertion involves the actual selection of landmark restriction sites. Your protein sequence is displayed with an asterisk under every tenth amino acid. Anchored pull down menus contain all pf the possible silent, unique insertions.



Easiest method: You can make the computer select sites for you with the "Pick Sites For Me" button. Enter an amino acid interval and GeneDesign will select sites on that interval using the criteria you defined in the last screen. The program's choices will be presented to you using the same screen and you will have a chance to edit them. You can change the amino acid interval and have the program reconsider as many times as you like.

Manual method: You can make your own landmark selections. Each pull down menu is anchored by a line and a dot between a pair of amino acids; these only appear where a silent mutation is possible.

The pull down menus are populated with lists of enzymes that can be introduced at those positions. A black line indicates that none of the enzymes in the list are absent from the vector you selected (or present in the list of sites you created). A red line indicates that there are one or more enzymes in the list that are absent from the vector. Small warning: the program does not currently keep a list of your choices during manual selection, to help you keep track of which enzymes have been used already. We are working on this.

If you have the computer select the sites it will use blue lines to indicate its selections. If you like you may edit the selection of sites after the computer has made its selections, for example, if you really want a unique site at a specific position, or if you require a XhoI site somewhere in your gene. The next screen will warn you of any difficulties in landmark insertion.

Once you are satisfied with the sites that have been selected you can click the "Continue to Summary" button. The next screen will provide you with a detailed list of the sites chosen for silent insertion. The first column is the site name. The next column is the recognition sequence, followed by an asterisk if the enzyme exhibits star activity. The third column is the nucleotide position of the site. The fourth column defines the end type the enzyme will leave, 5', 3', or blunt. The fifth column is the incubation temperature of the enzyme. The sixth and seventh columns provide a general guideline to methylation sensitivity. The eighth column is vendor availability and the ninth column is the cost per unit in '04 dollars.

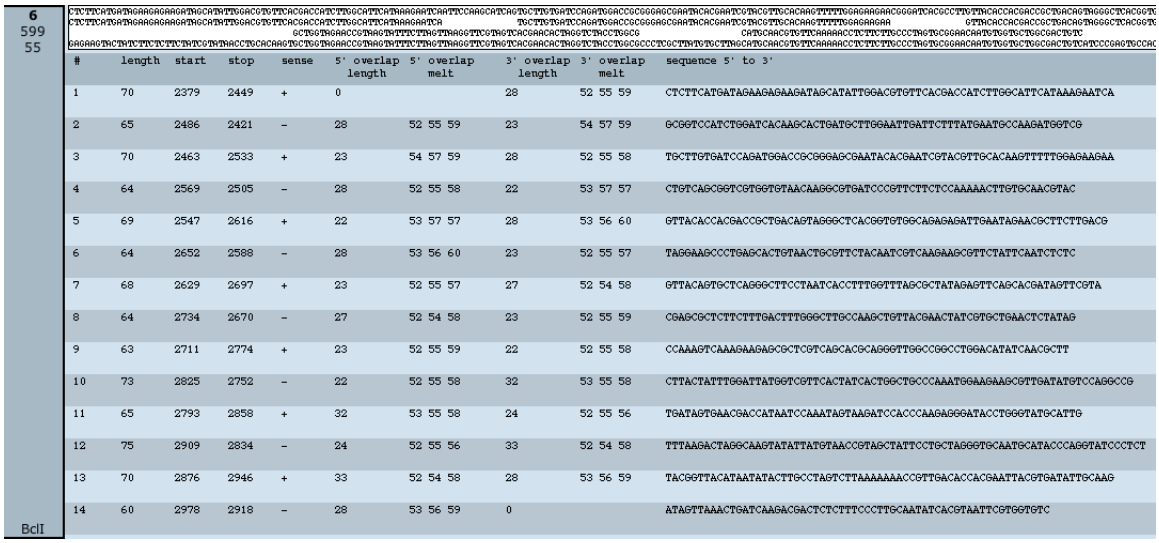
At the far left, there is a checkbox next to each enzyme name. If you used automatic site selection and are unhappy with one or more of the enzymes chosen, you can check any number of them and hit the "Reconsider" button. Automatic selection will be run again and the enzymes you select will be excluded from consideration.

Your new nucleotide sequence is at the top of this last screen. It would be wise to double check it in your favorite sequence analysis program to be sure it produces the correct amino acid sequence at this stage!

You can now take this sequence to another module. To follow the "Design a Gene" path, click "Oligo Design".

Design a Gene: Oligo Design

Now that you have a sequence, you can design the oligos. The defaults are 60bp oligos with 20bp overlaps. These values work well for us with yeast and mammalian sequences which are ~40% GC. You may want to play with these values if your sequences are outside of this range. GeneDesign will break the sequence into roughly 500bp chunks on unique restriction sites (restriction sites must be non-blunt, non-1bp overhang cutters that have cleavage sites internal to the recognition sequence) and then break those chunks into an even number of 60bp oligos. The oligos are then adjusted for even lengths and equal melting temperatures.



6 599 55 BclI	#	length	start	stop	sense	5' overlap length	5' overlap melt	3' overlap length	3' overlap melt	sequence 5' to 3'
	1	70	2379	2449	+	0		28	52 55 59	CTCTTCATGATAGAAGAGAAGATAGCATATTGGACGTGTTCACGACCATCTTGGCATTCATAAAGAATCA
	2	65	2486	2421	-	28	52 55 59	23	54 57 59	GCGGTCCATCTGGATCACAAGCACTGATGCTTGGAATTGATTCTTTATGAATGCCAAGATGGTCG
	3	70	2463	2533	+	23	54 57 59	28	52 55 58	TGCTTGTGATCCAGATGGACCGCGGGAGCGAATACACGAATCGTACGTTGCACAAGTTTTTGGAGAAGAA
	4	64	2569	2505	-	28	52 55 58	22	53 57 57	CTGTCAGCGGTCGTGGTGTAACAAGGCGTGATCCCGTTCTTCTCCAAAAACTTGTGCAACGTAC
	5	69	2547	2616	+	22	53 57 57	28	53 56 60	GTTACACCACGACCGCTGACAGTAGGGCTCACGGTGTGGCAGAGAGATTGAATAGAACGCTTCTTGACG
	6	64	2652	2588	-	28	53 56 60	23	52 55 57	TAGGAAGCCCTGAGCACTGTAACTGCGTTCTACAATCGTCAAGAAGCGTTCTATTCAATCTCTC
	7	68	2629	2697	+	23	52 55 57	27	52 54 58	GTTACAGTGCTCAGGGCTTCCTAATCACCTTTGGTTTAGCGCTATAGAGTTCAGCACGATAGTTCGTA
	8	64	2734	2670	-	27	52 54 58	23	52 55 59	CGAGCGCTCTTCTTTGACTTTGGGCTTGCCAAGCTGTTACGAACTATCGTGCTGAACTCTATAG
	9	63	2711	2774	+	23	52 55 59	22	52 55 58	CCAAAGTCAAAGAAGAGCGCTCGTCAGCACGCAGGGTTGGCCGGCCTGGACATATCAACGCTT
	10	73	2825	2752	-	22	52 55 58	32	53 55 58	CTTACTATTTGGATTATGGTCGTTCACTATCACTGGCTGCCCAAATGGAAGAAGCGTTGATATGTCCAGGCCG
	11	65	2793	2858	+	32	53 55 58	24	52 55 56	TGATAGTGAACGACCATAATCCAAATAGTAAGATCCACCCAAGAGGGATACCTGGGTATGCATTG
	12	75	2909	2834	-	24	52 55 56	33	52 54 58	TTTAAGACTAGGCAAGTATATTATGTAACCGTAGCTATTCCTGCTAGGGTGCAATGCATACCCAGGTATCCCTCT
	13	70	2876	2946	+	33	52 54 58	28	53 56 59	TACGGTTACATAATATACTTGCCTAGTCTTAAAAAAACCGTTGACACCACGAATTACGTGATATTGCAAG
	14	60	2978	2918	-	28	53 56 59	0		ATAGTTAAACTGATCAAGACGACTCTCTTTCCCTTGCAATATCACGTAATTCGTGGTGTC

Every 500bp chunk is displayed on the oligo summary page. For each chunk, the order number, length in bp, average Tm of overlaps, and 3' unique site appear in that order in the vertical bar at left. The chunk sequence and the sequences of oligos are printed for visual confirmation of appropriate alignment at the top of each chunk. Information for each oligo is contained in rows under the line up. This information includes oligo length, start and stop coordinates, and sense. It also includes data on the length and melting temperatures of the 5' and 3' overlaps. The 5' to 3' sequence of each oligo is printed at the far right.

Melting temperature is calculated in three ways. The first column uses the formula from Baldino et al. This is the same formula used by Strider. The second column uses a derivative of that formula, as used in Primer3. The third column uses the nearest neighbor (Borer et al) parameters for DNA/DNA duplexes as defined in Sugimoto et al. These

parameters are used in the formula below, derived from Rychlik et al, where ΔH is enthalpy, ΔS is entropy, R is the molar gas constant, (16.6*log[Na+]) is the salt correction, and 3.4 kcal/mol is the activation energy required to go from single to double stranded DNA.

$$Tm\ (^{\circ}C) = \frac{\Delta H - 3.4}{.001(\Delta S + R \ln(\frac{[oligo]}{4}))} - 273.15 + (16.6 * \log[Na^{+}])$$

If there are no further adjustments to be made, you can export these oligos to a delineated text file for ordering. If you select a designation string for the oligos and a delineator, you can use the "Order Form" button to kick out a list that can easily be opened in Excel or sent to the oligo producing company of your choice. The oligos will be numbered from first chunk to last, using the designation string. For example, if your designation string is "syngene", the first oligo will be called "syngene01" and the next will be "syngene02", etc.

The "Full Report" button will open in a new window a summary of all the steps you took in the Design a Gene path, from Reverse Translation to Oligo Designs.

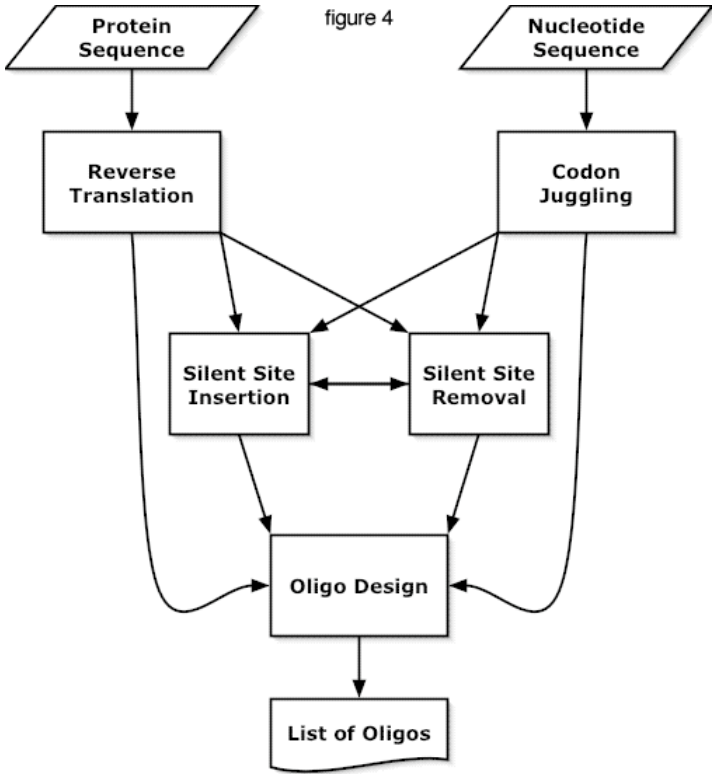
GeneDesign Modules

The GeneDesign modules have been designed for individual use and flexibility. The movement of nucleotide sequences between modules is accomplished with the single click of a button. For instance, you can go from codon juggling to oligo design, or from site insertion to site removal and back to site insertion, simply by directing the results of one module to another.

The main GeneDesign modules are Codon Juggling, Reverse Translation, Silent Site Insertion, Silent Site Removal, and Oligo Design. The chart at right illustrates the flow of information between these five modules. Not pictured is Sequence Analysis, which can be entered from any of the main modules.

There are three supplementary modules, Random DNA Generation, Enzyme Choosing and Vector Choosing. These modules do not interface with the others but provide information that is useful to the design of synthetic genes.

figure 4



GeneDesign Modules: Codon Juggling

GeneDesign's Codon Juggling module uses several algorithms to modify a coding nucleotide sequence without altering its translation. It is thus possible to generate a sequence that is optimized for expression, as different as possible from the original sequence (and still coding the same protein), or a combination of the two.

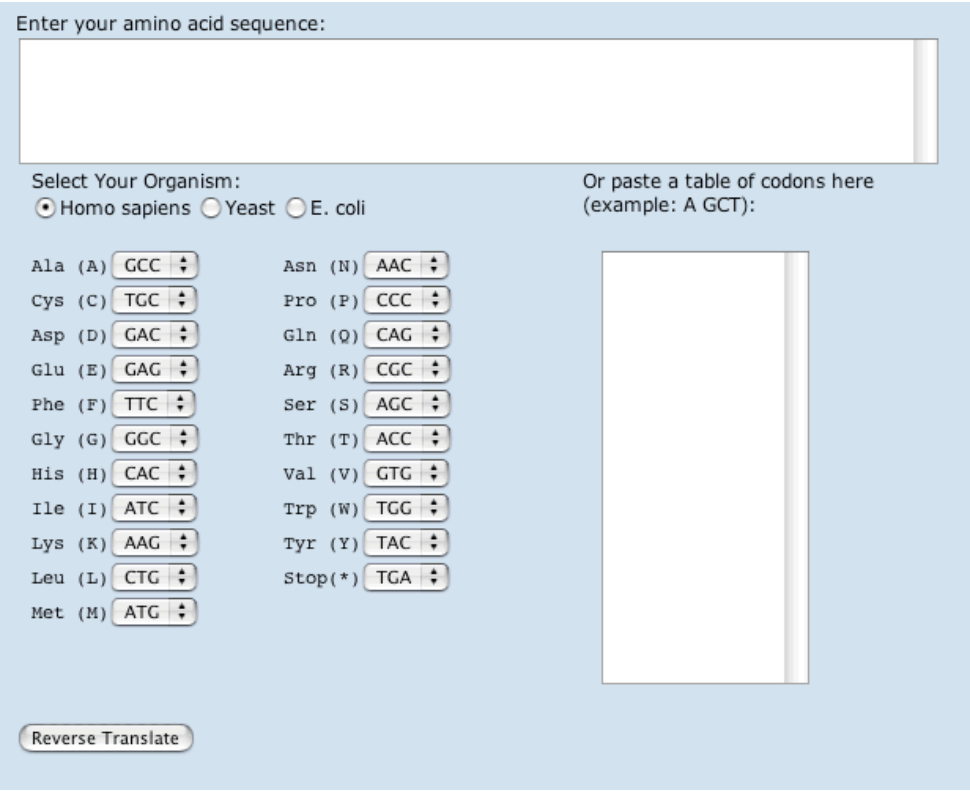
The optimized algorithm, which is only available if you are using one of the organisms recognized by GeneDesign, simply replaces every codon in the sequence with the most translationally optimal codon for that organism. If the codon is already the ideal codon it is left alone. This data comes from this paper and this one. The next most optimal algorithm uses the same data to replace the original codon with the most optimal codon that is not the original codon; that is, if the current codon is not the most optimal it becomes the most optimal, and if it is the most optimal it becomes the next most optimal.

The most different algorithm attempts to change as many bases as possible within the codon, preferably transversions. In two, three, and four codon families it switches the wobble. In six codon families it always changes the first position and the wobble; in the Serine family it changes all three positions. If two codons are equally eligible as replacements and an organism has been defined, this algorithm will take the more translationally optimal codon. if no organism has been defined, this algorithm decides ties randomly and is non-deterministic.

The random algorithm swaps the original codons out with random selections from the same family.

GeneDesign Modules: Reverse Translation

GeneDesign's Reverse Translation module takes a protein sequence to synonymous nucleotide sequence using either a user-defined codon scheme or the most optimal codons for expression in a user-selected organism.



To define the codon usage you can select one of the four common organisms from the radio buttons. Choosing an organism will cause the pop-up boxes to automatically select the codon with the highest RSCU value in highly expressed genes in that organism (pubmed). You can then edit the selection if you wish.

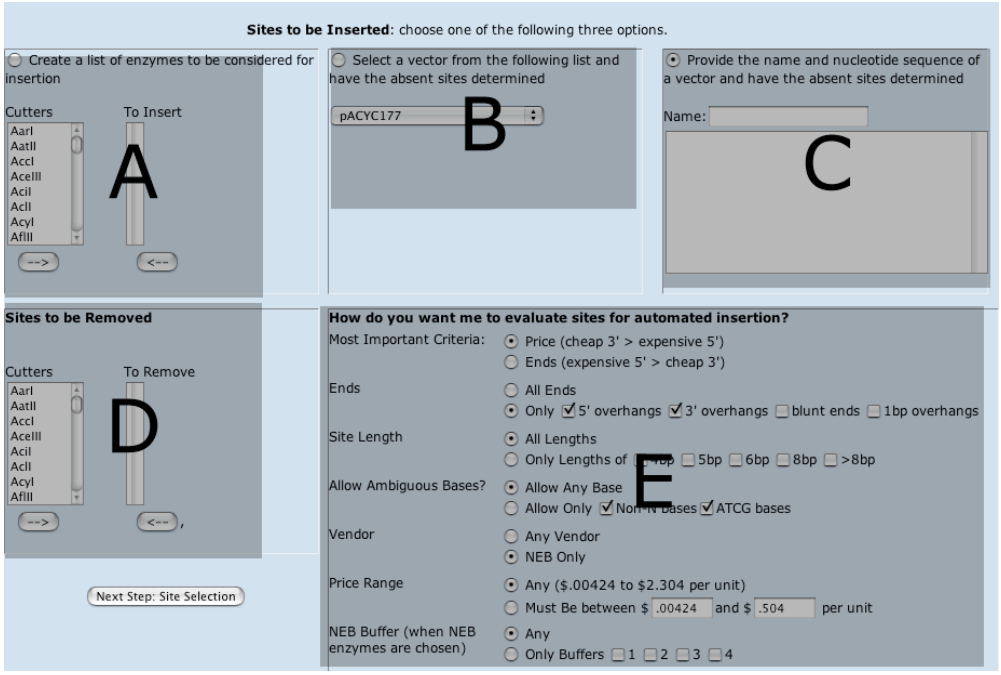
You may also paste your own custom codon table into the program. The format is "AA space codon line break", so that the first line of a valid table could be "M ATG" and the second line "L CTG", for example. Any input in the codon table will supercede all entries in the pop-up boxes, so your table must be complete. Your 21 codon input will be used as the optimal codons. This version of GeneDesign uses the standard genetic code.

Once the codons have been defined to your satisfaction, hit "Reverse Translate". Your new nucleotide sequence will appear. There will be

several buttons at the bottom that allow you to take this nucleotide sequence to other modules.

GeneDesign Modules: Silent Site Insertion

GeneDesign's Silent Site Insertion module allows you to process a nucleotide sequence for modular mutagenesis. You can choose the vector you plan to use to carry your synthetic gene, or you can define a list of restriction enzymes to be considered for insertion. Vector choice is an extremely important step in the process. The purpose is to instruct the program on what restriction sites you DO NOT WANT to use as special "landmark restriction sites" in your synthetic sequence, namely those already present in the vector. If you choose a very large vector with many sites in it, the program will not have any sites left to play with. Therefore, if you have a choice, use the smallest vector available to you. In practice pUC18/19 work well; derivatives lacking the multiple cloning site work even better.



Next you must choose the sites to be inserted. If you specify a vector sequence, all sites in the vector are automatically removed from consideration as landmark restriction sites by the program. There are three ways to choose eligible sites. You may (A) simply select the sites from a list. Click on the site name in the left column and then the arrow button at the bottom of the column to move that site to the right hand column. Repeat the process on the right side to remove it. Only sites in the right column will be considered. Or you may (B) select the name of the vector you will be using from a pull-down menu. GeneDesign has the sequences of these vectors and will determine which enzymes are absent from this vector automatically

for you. Lastly, you may (C) provide a vector sequence and a name for that vector. GeneDesign will parse the sequence and determine which enzymes are absent from your vector.

You may in addition specify a list of enzymes that must not appear in the final sequence. (D) Click on the site name in the left column and then the arrow button at the bottom of the column to move that site to the right hand column. Repeat the process on the right side to remove it. Only sites in the right column will be processed. The sequence will be searched for these sites and if they are found in your first-pass synthetic sequence, they will be removed. These sites will not be considered for landmark restriction sites.

At (E) you can define the ranking criteria for sites to be inserted. We have found the default criteria to be more than adequate for most purposes. For more information about ranking enzymes, see Enzyme Choosing.

Once you have selected the vector and/or the sites you can click "Next Step" to move on to the next step in site selection.

The second step of Silent Site Insertion involves the actual selection of landmark restriction sites. Your protein sequence is displayed with an asterisk under every tenth amino acid. Anchored pulldown menus contain all pf the possible silent, unique insertions.

Easiest method: You can have the computer select sites for you with the "Pick Sites For Me" button. Enter an amino acid interval and GeneDesign will select sites on that interval using the criteria you defined in the previous screen. The program's choices will be presented to you using the same screen and you will have a chance to edit them. You can change the amino acid interval and have the program reconsider as many times as you like.

Manual method: You can make your own landmark selections. Each pulldown menu is anchored by a line and a dot between a pair of amino acids; these only appear where a silent mutation is possible. The pulldown menus are populated with lists of enzymes that can be introduced at those positions. A black line indicates that none of the enzymes in the list are absent from the vector you selected (or present in the list of sites you created). A red line indicates that there are one or more enzymes in the list that are absent from the vector.

If you have the computer select the sites it will use blue lines to indicate its selections.

Once you are satisfied with the sites that have been selected you can click the "Continue to Summary" button. The next screen will provide you with a detailed list of the sites chosen for silent insertion. The first column is the site name. The next column is the recognition sequence, followed by an asterisk if the enzyme exhibits star activity. The third column is the nucleotide position of the site. The fourth column defines the end type the enzyme will leave, 5', 3', or blunt. The fifth column is the incubation temperature of the enzyme. The sixth and seventh columns provide a general guideline to methylation sensitivity. The eighth column is vendor availability and the ninth column is the cost per unit in '04 dollars.

At the far left, there is a checkbox next to each enzyme name. If you used automatic site selection and are unhappy with one or more of the enzymes chosen, you can check any number of them and hit the "Reconsider" button. Automatic selection will be run again and the enzymes you select will be excluded from consideration.

Your new nucleotide sequence is at the top of this last screen. It would be wise to double-check it in your favorite sequence analysis program to be sure it produces the correct amino acid sequence at this stage!

You can now take this sequence to another module.

GeneDesign Modules: Silent Site Removal

GeneDesign's Silent Site Removal module allows you to remove restriction enzyme sites without affecting the translation of the sequence.

After you input your sequence, GeneDesign will parse it for restriction sites and report back to you with a list of all of the present sites. You can check as many as you like for removal and GeneDesign will attempt to remove them all. If you defined a host organism, codon changes will be made with the optimization of the sequence in mind. If no organism was defined all changes will be random. Any errors or difficulties will be reported to you on the last page, as well as the similarity of the new sequence to the old sequence.

After site removal you can take the sequence to another module.

GeneDesign Modules: Enzyme Chooser

GeneDesign's Enzyme Chooser module allows you to search and sort the restriction enzymes used by the program by a variety of criteria.

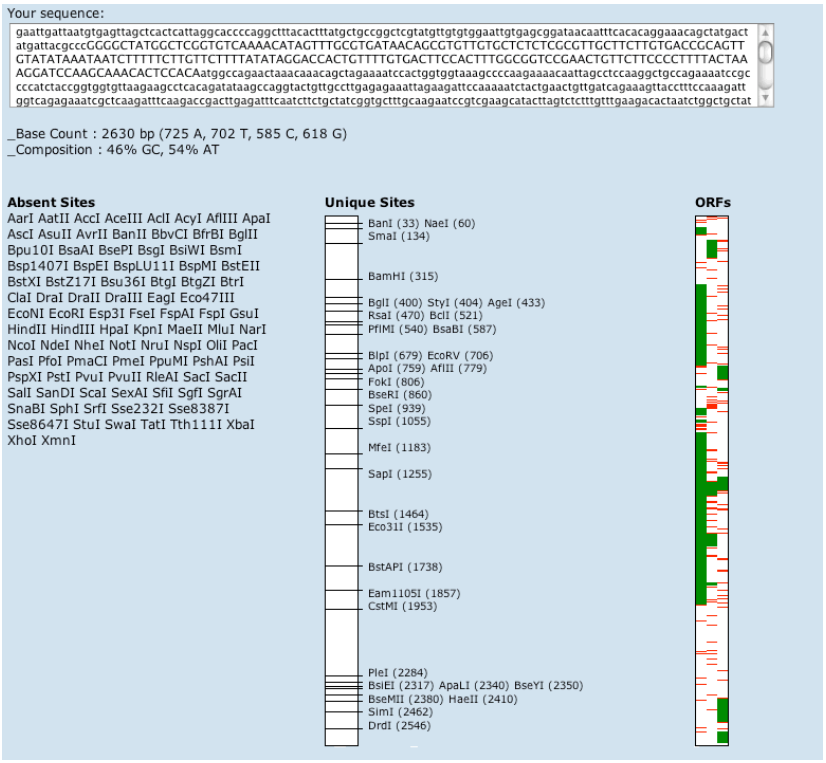
You can select enzymes by ends, site length (this does not count trailing Ns), the presence of ambiguous bases, price, the presence or absence of specific sequence in the recognition site, and whether or not the cleavage site is internal to the recognition site.

The next screen will give you a copy-able sorted list of enzymes that match your criteria, as well as an annotated list for your further edification.

GeneDesign Modules: Sequence Analysis

GeneDesign's Sequence Analysis module allows you to examine the composition and melting temperature of single sequences or groups of oligos.

Simply paste your sequence(s) into the text box, check which kind of sequence you analyzing, and click the "analyze" button. The next screen will display a base composition and a list of melting temperatures for each sequence you input.

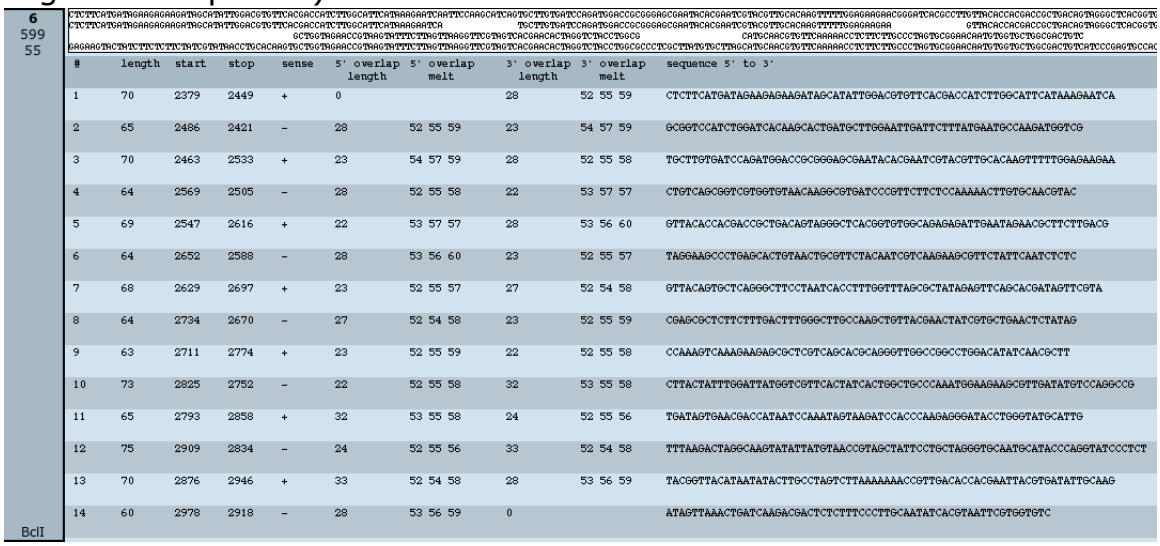


If you put in a single sequence you will also see a list of prototypes and their neoschizomers that are absent from the sequence, a map of the same that are unique, and a vertical chart of the open reading frames in the first three frames. The first column of the ORF chart is the first frame, the second column the second frame, etc. Green bars indicate an ORF; red bars indicate a stop codon. To analyze another single sequence, you do not have to go back. Simply replace the sequence in the text box with the new sequence and click "analyze again."

If you are analyzing a list of oligos, say for the purposes of verifying base content, length, or melting temperature, paste them in the the box with carriage returns between each oligo. On the next screen you will see a table. The first column is the oligo, the second the oligo number (assigned in the order the oligo were input), then the percent GC and AT. Then you will see three letter number pairs. B stands for Baldino et al, P stands for Primer3, and N stands for nearest neighbor. These are the three formulas with which the melting temperature is calculated. For more discussion of Tm, see Oligo Design (at the bottom of the section).

GeneDesign Modules: Oligo Design

GeneDesign's Oligo Design module allows you to generate a list of oligonucleotides from a nucleotide sequence. The default settings produce 60bp oligos with ~20bp overlaps. These values work well for us with yeast and mammalian sequences which are ~40% GC. You may want to play with these values if your sequences are outside of this range. GeneDesign will break the sequence into roughly 500bp chunks on unique restriction sites (restriction sites must be non-blunt, non-1bp overhang cutters that have cleavage sites internal to the recognition sequence) and then break those chunks into an even



6 599 55	#	length	start	stop	sense	5' overlap length	5' overlap melt	3' overlap length	3' overlap melt	sequence 5' to 3'
	1	70	2379	2449	+	0		28	52 55 59	CTCTTCATGATAGAAGAGAAGATAGCATATTGGACGTGTTCACGACCATCTTGGCATTCATAAAGAATCA
	2	65	2486	2421	-	28	52 55 59	23	54 57 59	GCGGTCCATCTGGATCACAAGCACTGATGCTTGGAATTGATTCTTTATGAATGCCAAGATGGTCG
	3	70	2463	2533	+	23	54 57 59	28	52 55 58	TGCTTGTGATCCAGATGGACCGCGGGAGCGAATACACGAATCGTACGTTGCACAAGTTTTTGGAGAAGAA
	4	64	2569	2505	-	28	52 55 58	22	53 57 57	CTGTCAGCGGTCGTGGTGTAACAAGGCGTGATCCCGTTCTTCTCCAAAAACTTGTGCAACGTAC
	5	69	2547	2616	+	22	53 57 57	28	53 56 60	GTTACACCACGACCGCTGACAGTAGGGCTCACGGTGTGGCAGAGAGATTGAATAGAACGCTTCTTGACG
	6	64	2652	2588	-	28	53 56 60	23	52 55 57	TAGGAAGCCCTGAGCACTGTAACTGCGTTCTACAATCGTCAAGAAGCGTTCTATTCAATCTCTC
	7	68	2629	2697	+	23	52 55 57	27	52 54 58	GTTACAGTGCTCAGGGCTTCCTAATCACCTTTGGTTTAGCGCTATAGAGTTCAGCACGATAGTTCGTA
	8	64	2734	2670	-	27	52 54 58	23	52 55 59	CGAGCGCTCTTCTTTGACTTTGGGCTTGCCAAGCTGTTACGAACTATCGTGCTGAACTCTATAG
	9	63	2711	2774	+	23	52 55 59	22	52 55 58	CCAAAGTCAAAGAAGAGCGCTCGTCAGCACGCAGGGTTGGCCGGCCTGGACATATCAACGCTT
	10	73	2825	2752	-	22	52 55 58	32	53 55 58	CTTACTATTTGGATTATGGTCGTTCACTATCACTGGCTGCCCAAATGGAAGAAGCGTTGATATGTCCAGGCCG
	11	65	2793	2858	+	32	53 55 58	24	52 55 56	TGATAGTGAACGACCATAATCCAAATAGTAAGATCCACCCAAGAGGGATACCTGGGTATGCATTG
	12	75	2909	2834	-	24	52 55 56	33	52 54 58	TTTAAGACTAGGCAAGTATATTATGTAACCGTAGCTATTCCTGCTAGGGTGCAATGCATACCCAGGTATCCCTCT
	13	70	2876	2946	+	33	52 54 58	28	53 56 59	TACGGTTACATAATATACTTGCCTAGTCTTAAAAAAACCGTTGACACCACGAATTACGTGATATTGCAAG
BclI	14	60	2978	2918	-	28	53 56 59	0		ATAGTTAAACTGATCAAGACGACTCTCTTTCCCTTGCAATATCACGTAATTCGTGGTGTC

number of 60bp oligos. The oligos are then adjusted for even lengths and equal melting temperatures.

Every 500bp chunk is displayed on the oligo summary page. For each chunk, the order number, length in bp, average Tm of overlaps, and 3' unique site appear in that order in the vertical bar at left. The chunk sequence and the sequences of oligos are printed for visual confirmation of appropriate alignment at the top of each chunk. Information for each oligo is contained in rows under the line up. This information includes oligo length, start and stop coordinates, and sense. It also includes data on the length and melting temperatures of the 5' and 3' overlaps. The 5' to 3' sequence of each oligo is printed at the far right.

$$Tm\ (^{\circ}C) = \frac{\Delta H - 3.4}{.001(\Delta S + R \ln(\frac{[oligo]}{4}))} - 273.15 + (16.6 * \log[Na^{+}])$$

Melting temperature is calculated in three ways. The first column uses the formula from Baldino et al. This is the same formula used by

Strider. The second column uses a derivative of that formula, as used in Primer3. The third column uses the nearest neighbor (Borer et al) parameters for DNA/DNA duplexes as defined in Sugimoto et al. These parameters are used in the formula below, derived from Rychlik et al, where ΔH is enthalpy, ΔS is entropy, R is the molar gas constant, (16.6*log[Na+]) is the salt correction, and 3.4 kcal/mol is the activation energy required to go from single to double stranded DNA.

Your oligos may be longer or shorter than expected because of optimizations for length and melting temperature.

If there are no further adjustments to be made, you can export these oligos to a delineated text file for ordering. If you select a designation string for the oligos and a delineator, you can use the "Order Form" button to kick out a list that can easily be opened in Excel or sent to the oligo producing company of your choice. The oligos will be numbered from first chunk to last, using the designation string. For example, if your designation string is "syngene", the first oligo will be called "syngene01" and the next will be "syngene02", etc.

GeneDesign Modules: Vector Chooser

GeneDesign's Vector Chooser module allows you to search and sort the vectors known to the program by a variety of criteria.

You can select vectors by size (kb), the number of absent sites, and the presence or absence of specific restriction enzymes.

The next screen will give you a copy-able sorted list of vectors that match your criteria, as well as an annotated list for your further edification.

This module is still under development. Any comments would be appreciated.

GeneDesign Modules: Short Sequence Removal

GeneDesign's Short Sequence Removal module allows you to alter a nucleotide sequence without changing its first frame translation by changing whole codons around the target sequence. This is exactly like the Silent Site Removal module except that the user is allowed to define the sequence to be removed.

You must provide two nucleotide sequences - a longer one as the main sequence, and a shorter one (at least 2 bp) to be removed from the first sequence. You have the option of defining an organism. If you do, the codons that are more optimal for expression in that organism will be considered first when altering sequence. If you do not select an organism, every codon replacement will be random.

GeneDesign Modules: Random DNA Sequence Generation

GeneDesign's Random DNA Sequence Generator module allows you to create a DNA sequence of order and defined composition.

You can define length (bp), AT content, and a number of sequences to generate. You can also choose to disallow stop codons in the first frame of translation (random ORFs).

The next screen will give you a copy-able list of the sequences generated.

This module is still under development. Comments would be appreciated.

GeneDesign Enzyme List

AarI (CACCTGC)
AatII (GACGTC)
AccI (GTMKAC)
AciI (CCGC)
AclI (AACGTT)
AcyI (GRCGYC)
AflII (CTTAAG)
AflIII (ACRYGT)
AgeI (ACCGGT)
AluI (AGCT)
AlwI (GGATC)
AlwNI (CAGNNNCTG)
ApaI (GGGCCC)
ApaLI (GTGCAC)
ApoI (RAATTY)
AscI (GGCGCGCC)
AsuII (TTCGAA)
AvaI (CYCGRG)
AvaII (GGWCC)
AvrII (CCTAGG)
BalI (TGGCCA)
BamHI (GGATCC)
BanI (GGYRCC)
BanII (GRGCYC)
BbeI (GGCGCC)
BbsI (GAAGAC)
BbvCI (CCTCAGC)
BbvI (GCAGC)
BccI (CCATC)
BceAI (ACGGC)
BciVI (GTATCC)
BclI (TGATCA)
BfrBI (ATGCAT)
BglI (GCCNNNNNGGC)
BglII (AGATCT)
BisI (GCNGC)
BlpI (GCTNAGC)
Bpu10I (CCTNAGC)
BpuEI (CTTGAG)
BsaAI (YACGTR)
BsaBI (GATNNNNATC)
BsaJI (CCNNGG)
BsaWI (WCCGGW)
BseMII (CTCAG)
BseRI (GAGGAG)
BseSI (GKGCMC)
BseYI (CCCAGC)
BsgI (GTGCAG)
BsiEI (CGRYCG)
BsiHKAI (GWGCWC)
BsiYI (CCNNNNNNNGG)
BsmAI (GTCTC)
BsmI (GAATGC)
Bsp120I (GGGCCC)
Bsp1407I (TGTACA)
BspEI (TCCGGA)
BspHI (TCATGA)
BspLU11I (ACATGT)
BspMI (ACCTGC)
BsrBI (CCGCTC)
BsrDI (GCAATG)
BsrI (ACTGG)
BssHII (GCGCGC)
BssSI (CACGAG)
Bst1107I (GTATAC)
BstAPI (GCANNNNNTGC)
BstEII (GGTNACC)
BstF5I (GGATG)
BstNI (CCWGG)
BstUI (CGCG)
BstXI (CCANNNNNNTGG)
Bsu36I (CCTNAGG)
BtgI (CCRYGG)
BtgZI (GCGATG)
BtrI (CACGTC)
BtsI (GCAGTG)
Cac8I (GCNNGC)
Cfr10I (RCCGGY)
CfrI (YGGCCR)
ClaI (ATCGAT)
CviAII (CATG)
DdeI (CTNAG)
DpnI (GATC)
DraI (TTTAAA)
DraII (RGGNCCY)
DraIII (CACNNNGTG)
DrdI (GACNNNNNNGTC)
EagI (CGGCCG)
Eam1105I (GACNNNNNGTC)
EciI (GGCGGA)
Eco31I (GGTCTC)
Eco47III (AGCGCT)
Eco57I (CTGAAG)
Eco57MI (CTGRAG)
EcoICRI (GAGCTC)
EcoNI (CCTNNNNNAGG)
EcoRI (GAATTC)
EcoRII (CCWGG)
EcoRV (GATATC)
EcoT22I (ATGCAT)
Esp3I (CGTCTC)
FatI (CATG)
FauI (CCCGC)
Fnu4HI (GCNGC)
FokI (GGATG)
FseI (GGCCGGCC)
FspAI (RTGCGCAY)
FspI (TGCGCA)
GsuI (CTGGAG)
HaeII (RGCGCY)
HaeIII (GGCC)
HgaI (GACGC)
HhaI (GCGC)
HinP1I (GCGC)
HindII (GTYRAC)
HindIII (AAGCTT)
HinfI (GANTC)
HpaI (GTTAAC)
HpaII (CCGG)
HphI (GGTGA)
Hpy188I (TCNGA)
Hpy188III (TCNNGA)
Hpy99I (CGWCG)
HpyCH4III (ACNGT)
KasI (GGCGCC)
KpnI (GGTACC)
Ksp632I (CTCTTC)
MaeI (CTAG)
MaeII (ACGT)
MaeIII (GTNAC)
MboI (GATC)
MboII (GAAGA)
MfeI (CAATTG)
MluI (ACGCGT)
MlyI (GAGTC)
MmeI (TCCRAC)
MnlI (CCTC)
MseI (TTAA)
MslI (CAYNNNNRTG)
MspA1I (CMGCKG)
MwoI (GCNNNNNNNGC)
NaeI (GCCGGC)
NarI (GGCGCC)
NciI (CCSGG)
NcoI (CCATGG)
NdeI (CATATG)
NheI (GCTAGC)
NlaIII (CATG)
NlaIV (GGNNCC)
NotI (GCGGCCGC)
NruI (TCGCGA)
NspI (RCATGY)
OliI (CACNNNNGTG)
PacI (TTAATTAA)
PflMI (CCANNNNNTGG)
PfoI (TCCNGGA)
PleI (GAGTC)
PmaCI (CACGTG)
PmeI (GTTTAAAC)
PpuMI (RGGWCCY)
PshAI (GACNNNNGTC)
PsiI (TTATAA)
PspXI (VCTCGAGB)
PstI (CTGCAG)
PvuI (CGATCG)
PvuII (CAGCTG)
RsaI (GTAC)
RsrII (CGGWCCG)
SacI (GAGCTC)
SacII (CCGCGG)
SalI (GTCGAC)
SanDI (GGGWCCC)
SapI (GCTCTTC)
Sau96I (GGNCC)
ScaI (AGTACT)
ScrFI (CCNGG)
SduI (GDGCHC)
SexAI (ACCWGGT)
SfaNI (GCATC)
SfcI (CTRYAG)
SfiI (GGCCNNNNNGGCC)
SgfI (GCGATCGC)
SgrAI (CRCCGGYG)
SmaI (CCCGGG)
SmlI (CTYRAG)
SnaBI (TACGTA)
SpeI (ACTAGT)
SphI (GCATGC)
SrfI (GCCCGGGC)
Sse8387I (CCTGCAGG)
SspI (AATATT)
StuI (AGGCCT)
StyI (CCWWGG)
SwaI (ATTTAAAT)
TaiI (ACGT)
TaqI (TCGA)
TatI (WGTACW)
TauI (GCSGC)
TfiI (GAWTC)
TseI (GCWGC)
Tsp45I (GTSAC)
TspEI (AATT)
TspRI (CASTGNN)
Tth111I (GACNNNGTC)
VspI (ATTAAT)
XbaI (TCTAGA)
XcmI (CCANNNNNNNNNTGG)
XhoI (CTCGAG)
XhoII (RGATCY)
XmaI (CCCGGG)
XmnI (GAANNNNTTC)
ZraI (GACGTC)

GeneDesign Version History

Changes from 1.9 to 2.0

- Revamped Enzyme List and Enzyme Selection Criteria
    - New list composed of prototypes (and their neoschizomers) that are commercially available, single cutting, Type IIs with known cleavage sites.
    - No longer use buffer, vendor, and NEB availability as criteria.
    - Added incubation temperature, methylation sensitivity, star activity and exhaustive vendor list.
    - Now able to exclude enzymes with cleavage external to recognition site.
- Added C. elegans to list of known organisms
- Oligo Design now automatically excludes blunts, 1bp overhangs, and enzymes that cut outside of recognition site.
- Codon Juggling detailed
    - Now warns if you are not using a simple coding sequence
    - Will allow you to not define an organism; hides optimization algorithms and randomizes most different
    - Define next most optimal and random as highly experimental algorithms
- Silent Site Insertion detailed
    - Allow reconsideration of automatic site selection
    - Fixed error messages about simple coding sequences
- Miscellaneous bug fixes that you probably never noticed
- Manual updated accordingly and PDF version added.
- Now with version history.