Designing primers for a combinatorial library – DNA sequencing is an effective way to identify an unknown strand of DNA. Applications that require identification of strands from a combinatorial DNA library (such as those described here) would therefore benefit if the word sequences were constructed in a manner that facilitated sequencing. One requirement of a DNA sequencing reaction is that the template DNA has a short region that can serve as a unique site for hybridization of the sequencing primer. In this respect, combinatorial DNA libraries are ideal since any individual word sequence could (in principal) serve as a primer binding site. However, there are four complicating factors. First, not all word sequences have melting temperatures high enough to function well in the sequencing reaction. Second, tandem word sequences can be quite short and DNA sequencing is commonly ineffective at identifying the first ~50 bases (those within the compression zone). In that case, some or even all of the sequence would be unobtainable. Third, a sequencing reaction would need as many primers as there are word variants (in the first tandem word position) to account for all possible sequences. Fourth, depending on how the tandem word sequences were processed and with what efficiency, there may be too few sequences to serve effectively as template. These challenges can all be overcome through use of common primers that are appended to both ends of every word sequence. If the total sequence length including primers is > 100 nt and there is a sufficient number of template molecules, the tandem word sequence can be obtained by bi-directional sequencing (sequencing from both ends) provided there is enough overlap in the sequence outside the compression zone for alignment. If the total sequence length is shorter than 100 nt or if the number of template sequences is insufficient for the sequencing reaction to occur then the tandem word sequence can first be amplified by PCR. The PCR amplicon can be sequenced directly if it is longer than 100 nt or cloned and sequenced if it is shorter than 100 nt.

Primers for combinatorial word sets are constructed from the group of words that remain after the final word set is created. The process of designing these primer sequences, which is analogous to the process used for creating the combinatorial library, is given below:

1. Creating the primer at the 5'-end of the tandem word sequence

    a. Choose a word from the remaining members of the ranked list (650-64) that can be concatenated to any of the 'A' words without forming a junction that will engage in a junction mishybridization with any of the 64 set words or their complements. The complement of this word must also not mishybridize with stability above the cut-off to any of the word-set junctions.

    b. Add nucleotides to the 5'-end of the candidate primer to achieve a total length of 18 nt. Added nucleotides are chosen randomly from the group A, T and C with the added criteria that: the 5'-most nucleotide is always a C; the first nucleotide added adjacent to the candidate primer is not C; and, that at most two consecutive Cs are present in the added nucleotide sequence.

c. Analyze the binding energies between the complement of the newly created primer sequence and all of the word:word junctions in the set using the PairFold software.

d. If the binding energy between the primer complement and any word:word junction is stronger than the cut-off would allow then go to step b.) above and repeat the process with an alternate set of appended nucleotides.

e. If no acceptable primer is created then go to step 1.a.) above and repeat the process with a new word chosen from the ranked list.

f. Analyze the binding energies between the newly created primer sequence and all of the word set complements using the PairFold software.

g. If the binding energy between the primer and any word complement is stronger than the cut-off would allow then go to step 1.b.) above and repeat the entire process with an alternate set of appended nucleotides.

h. If no acceptable primer is created then go to step a.) above and repeat the entire process with a new word chosen from the ranked list.

i. If no acceptable primer sequence can be created with any of the words from the ranked list then the mismatch binding energy cut-off must be relaxed to the point when an acceptable primer sequence can be created.

2. Creating the primer binding site for the 3'-end of the tandem word sequence

a. Choose a word from the remaining members of the ranked list that can be concatenated to any of the 'D' words without forming a junction that will engage in a junction mishybridization with any of the 64 set words or their complements. The selected word and its complement must also not mishybridize with stability above the cut-off to any of the word-set junctions (including the junctions between the 5'-primer and the A_i words).

b. Add nucleotides to the 3'-end of the candidate primer to achieve a total length of 18 nt. Added nucleotides are chosen randomly from the group A, T and C with the added criteria that: the 3'-most nucleotide is always a C; the first nucleotide added adjacent to the candidate primer is not C; and, that at most two consecutive Cs are present in the added nucleotide sequence.

c. Analyze the binding energies between the complement of the newly created primer sequence and all of the word:word junctions in the set (including the junctions associated with the 5'-primer) using the PairFold software.

d. If the binding energy between the primer complement and any junction is stronger than the cut-off would allow then go to step 2.b.) above and repeat the process with an alternate set of appended nucleotides.

e. If no acceptable primer is created then go to step 2.a.) above and repeat the process with a new word chosen from the ranked list.

f. Analyze the binding energies between the newly created primer sequence and all of the word set complements (including the complement of the 5'-primer) using the PairFold software.

g. If the binding energy between the 3'-primer and any complement is stronger than the cut-off would allow then go to step 2.b.) above and repeat the entire process with an alternate set of appended nucleotides.

h. If no acceptable primer is created then go to step 2.a.) above and repeat the entire process with a new word chosen from the ranked list.

If no acceptable primer sequence can be created with any of the words from the ranked list then the mismatch binding energy cut-off must be relaxed to the point where an acceptable primer sequence can be created.