General Instruction for use of the modular assembly algorithm:

Enter your DNA sequences of interest in the fasta file format. This can be done as a batch file where the fasta files for different genomic regions are concatenated. An example input file is linked at the top of the page.

Enter the minimum score for the pair of ZFPs that you would like to have for the target sequences. The scores for each pair of ZFPs range from 0 to 18 with 3 points possible per finger. If there is a 5' G in the triplet recognized by a finger that is worth a single point. The other 2 points per finger are received if and only if 2 conditions are met: 1) a 3' G is present in the triplet, 2) the following base is either G or T (A or C on the complementary strand). This allows the use of an RSD motif at the -1, 1 and 2 positions of the finger. See figure below:

The contribution of each module to DNA-recognition can be approximated by the number of guanines at the 5' and 3' positions within its DNA triplet
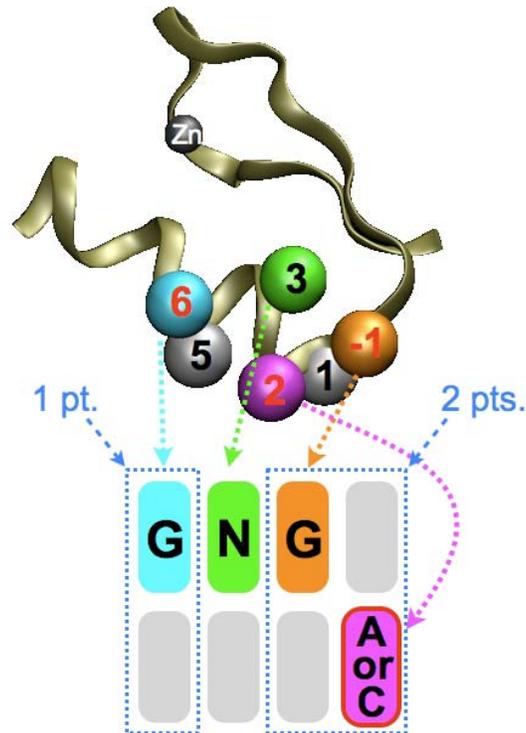
Choose the type of output file that you desire and hit submit. This will provide a list of hits sorted in descending order by score. The site position indicates the start position of the ZFN site in the sequence. The ZFN site has the 2 ZFP recognition elements in caps and the intervening and flanking bases in lowercase. The distance between the ZFN site is indicated by the gap. The 5p (left site) and 3p ZFP sites indicate the ZFP recognition sequences, where the 5p site is the reverse complement of the ZFN site sequence. The triplet that is recognized by each of the 6 fingers is indicated. Next the presence of a restriction enzyme (RE) site is indicated. Finally the option to output

information on the zinc finger modules that need to be assembled is present. This provides information on the clones that need to be assembled (link to document at the bottom of the page). The information on each particular clone is found in the linked xls file at the bottom of the page. The protocol for this process is also found at a link at the bottom. Finally more information on the RE sites present in the spacer region is indicated where sequences in brackets indicates that multiple bases are allowed at a specific position (e.g. [ACGT] = N).

Please contact scot wolfe (scot.wolfe@umassmed.edu) or Nathan Lawson (nathan.lawson@umassmed.edu) if you have any questions or concerns.