

The Sequence Manipulation Suite is now hosted by Bioinformatics.Org, an organization that promotes the development of Open Source software for the biological sciences. It can be accessed at <http://bioinformatics.org/sms> as well as from numerous mirror sites, most of which can be found using Web search engines.

Several minor enhancements have been made to the Sequence Manipulation Suite, including the addition of new features and options to many of the programs. For example, Filter DNA/Protein can now remove specific types of non-DNA and non-protein characters, or it can replace them with alternative characters of your choice. You could use this added versatility, for example, to quickly replace all the IUPAC degenerate characters in a DNA sequence with “N” to make the sequence suitable for programs that do not recognize two-fold or three-fold degenerate sites. GenBank Feature Extractor now has an option that allows sequence feature segments to be converted to uppercase, making it possible to view features in the context of the entire DNA sequence described in the GenBank record. Furthermore, coding sequences that are converted to uppercase can be pasted into other programs in Sequence Manipulation Suite that generate protein translations (Primer Show, Rest and Trans Map, Show Translation, and Translate). These all have a new translation option (called “uppercase”), which causes the uppercase segments to be treated as one continuous reading frame. Intervening lowercase segments are ignored during the translation, allowing genomic sequences to be viewed with their biologically relevant protein product. Multiple Align Show now includes more options for controlling the appearance of sequence alignments. The precise font type, size, and color of text that is used in the output can be adjusted. Also, different highlight colors

can be assigned to different groups of residues. This feature can be used to reveal protein regions that have conserved biochemical properties. Finally, a new numbering option allows alignments made using partial sequences to contain the correct numbering relative to the complete sequences.

One limitation of the Sequence Manipulation Suite is that the programs, which are written in JavaScript, execute slowly in certain Web browsers. To address this problem, and to provide features difficult to implement in JavaScript, some of the more popular utilities have been rewritten in Perl. These have been combined into a new program called Sequence Extractor, which can be accessed through its Web interface at <http://bioinformatics.org/seqext>. The primary purpose of Sequence Extractor is to generate useful reference figures for molecular biologists working with protein and DNA sequences. It accepts sequence files in raw, FASTA, GenBank, or EMBL format. It parses the sequence record and then generates an html Web page that displays the DNA sequence with protein translations, restriction sites, and primer annealing sites (primer sequences can be entered with the sequence record). What sets Sequence Extractor apart from other restriction map and feature viewing programs is that JavaScript code is embedded throughout the resulting figure to make it much more dynamic than static text output. For example, pointing to any restriction site label in the figure displays the precise position of the site and the recognition sequence of the enzyme in the Web browser's status bar. Similarly, pointing to primer sites or protein translations displays additional information not shown in the figure text. Clicking on two restriction site or two appropriate primer site labels causes the sequence between them to appear in a

separate window, essentially performing a virtual restriction digest and band isolation or PCR reaction. When primer sites are clicked, the sequence of the primers replaces the underlying “template” sequence, allowing engineered restriction sites to be incorporated into the product. These features greatly assist in the isolation of sequence fragments of interest, and in the construction of DNA constructs *in silico*. Finally, the protein translations and DNA sequences corresponding to coding regions can be obtained in a separate window using links that Sequence Extractor provides at the top of the figure. In summary, Sequence Extractor can be used in place of many of the utilities in the Sequence Manipulation Suite, and provides enhanced output not available from other existing programs.