User's Guide

Sequence Analysis Software

DNAMAN

Lynnon Corporation

Lynnon Corporation
383 ave St-Louis
Pointe-Claire, Québec
Canada H9R 2A3

Tel. (514) 447-2514
Fax (514) 447-2516

E-mail: **mail@lynnon.com**

Technical Support: **support@lynnon.com**

Web Site: **http://www.lynnon.com**

TABLE OF CONTENTS

Chapter I Introduction

DNAMAN is a sequence analysis software package for computers with Microsoft Windows or the Apple OSX systems. This package provides effective and convenient tools for molecular biologists to deal with frequently used analyses in research.

User's guide for DNAMAN

This document introduces the functions of DNAMAN and provides brief instructions on how to use them.

Features of DNAMAN

- DNA and protein sequence editing
- DNA sequence conversion
- Multiple sequence alignment, profile editing and analysis
- Phylogenetic tree analysis
- Dot-matrix comparison of DNA or protein sequences
- DNA sequence assembly and editing
- RNA/DNA secondary structure analysis
- Direct accessing BLAST Web Service of NCBI or local server
- Enhanced motif search in sequence and database
- Small interfering RNA search
- Restriction analysis
- Drawing sequence maps with publication-quality
- Restriction pattern prediction
- Electronic cloning
- Reconstructing restriction maps from restriction fragments

- Silent mutation analysis to create/destroy restriction sites
- Directed mismatch to create/destroy restriction sites
- Translation and codon usage analysis
- Protein hydrophobic/hydrophilic profile analysis
- Protein characterization: sequence composition and prediction of isoelectric point.
- Protein secondary structure prediction
- Reverse translation
- Design of PCR and sequencing primers
- Characterization of thermodynamic properties of DNA or primer sequences
- Mispriming analysis
- Management of oligo, DNA and protein databases
- Generation of random sequences
- Internet access with integrated Web browser (MS Windows only)
- Multi-processing to unleash computer power

Chapter II Installing & Getting Started with DNAMAN

This chapter shows how to install DNAMAN on your computer. It also introduces the conventions of the DNAMAN software, and helps you to get started after the installation.

II.1 System requirements

Before installation, make sure your computer has the minimum hardware and software requirements.

- CPU

  Intel X86 and other compatible microprocessors (AMD…)
- **Operation system**
  Microsoft: Windows 7/8
  Apple: OSX 10.6 or higher
- **Hard drive**
  A hard disk with 100 MB of available
- **Memory (RAM)**
  1GB (4 GB recommended)
- **Others**
  Monitor with high resolution: minimum 1024x768
  Network connection and Internet browser for online access

II.2 Setup

To install DNAMAN on PC:

The Setup program can be downloaded or shipped on media such as CD/DVD/Memory stick.

1. Locate the ***Setup Program*** on hard drive or media.
2. Double click the ***Setup*** program.
3. Follow instructions of the setup program. The setup program copies necessary files into the hard drive of your computer, and builds the DNAMAN program group on the ***Start*** menu.
4. After the setup, the DNAMAN program group is shown on the screen.

To install DNAMAN on OSX:
The DNAMAN for MacOSX Installer program installs DNAMAN on your computer.

1. Locate the ***DNAMAN Installer Program*** on hard disk or media.
2. Click the **DNAMAN for Installer** file.
3. Follow the instructions of the installer program. The installation program copies files into the hard drive of your computer.
4. After copying all files to the hard drive, click the **OK** button to finish the installation.

II.3 Understanding DNAMAN

Sequence editing

DNAMAN provides a text editor to edit sequence files. This editor has the same functions as WordPad in your Windows system, or TextEdit on your MacOSX system. You may add or remove any text or sequence in files. You may also use ***Copy*** and ***Paste*** functions for editing the files. Some sequence analysis results are also shown as text file, therefore, these results can be edited using the editor. When you open a sequence file in DNAMAN/GenBank/FASTA, DNAMAN will load sequence into channel for analysis. ***If sequence format is not recognized, analysis functions may not immediately available. You must load the sequence into a sequence channel.***

Sequence analysis

DNAMAN provides 20 sequence channels to keep active sequences in memory. Active sequence must be loaded into a sequence channel. You may switch from one sequence to another during different analyses without going back to disk files. The sequence may automatically loaded into a channel when a DNAMAN, FASTA or Genbank file is opened on the desktop. You can simultaneously analyze and compare all sequences in the channels.

Analysis functions using multiple sequences, such as multiple alignment and sequence assembly, do not require sequences loaded into channels.

Sequence alignment, comparison and assembly

You may directly use disk files for multiple alignment and sequence assembly. It is not necessary to load these sequences to channels. To perform these tasks, choose the corresponding commands on the menu bar, select sequence files or channels and fill out appropriate parameters. DNAMAN remembers the sequence files and parameters used in these analyses.

II.4 Quick start with DNAMAN

If you use DNAMAN for the first time, you may follow the instructions described below.

To perform sequence analysis:

1. Choose the ***File*** | ***Open …*** menu.
2. Select the **Example1.Seq** file in the DNAMAN folder from ***Open File Dialogue Box***. Click ***Open*** button to open the file. The sequence is automatically loaded to ***Sequence channel 1***.
3. Choose the ***Restriction*** | ***Restriction Analysis*** menu to perform restriction analysis on **Example1**.
4. After restriction analysis, choose the ***Protein*** | ***Translation*** menu to translate **Example1** to amino acid sequence.

To draw a restriction map without sequence:

Choose the ***Restriction*** | ***Draw Map*** menu. A plasmid map is drawn in a restriction map window. Double-click on the window to display the map properties and make modification.

Tips: All objects in a restriction map can be moved around by drag-and-drop. Restriction sites and Text objects can be aligned when they are selected.

II.5 Getting help

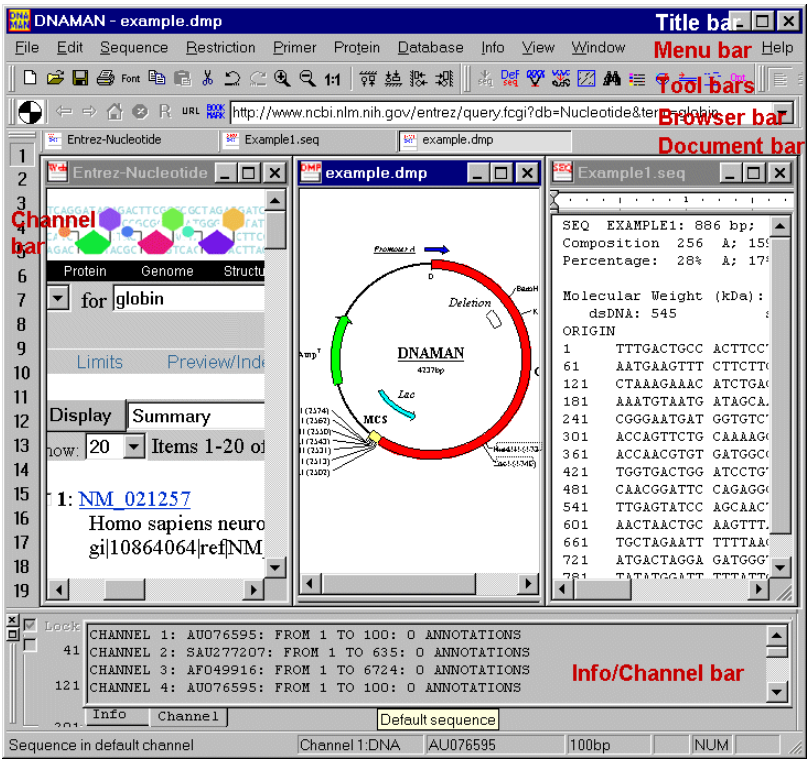
The **DNAMAN Help** provides on-line access to the detailed information of DNAMAN functions and the instructions on how to use them. Choose the ***Help | Help Topics*** command to open the Help window.

Chapter III DNAMAN Basics

This chapter provides the basic conceptions behind the DNAMAN software. It is important to understand these basics prior to performing sequence analysis.

III.1 DNAMAN Sequence Analysis Workspace

DNAMAN sequence analysis Studio is the integrated environment that all analysis tools run. The Studio provides user interfaces of all analysis tools. Users may access any analysis tools and view the current running processes through these interfaces. The studio is shown in the following figure.



III.2 Sequence channels

DNAMAN uses **Sequence Channels** to keep active sequences in the computer memory. The **Sequence Channels** increasing

analysis efficiency by decreasing disk accessing time. Once a sequence is loaded into a **Sequence Channel**, you can perform many functional analyses, such as restriction analysis, translation analysis, sequence search and comparison and PCR primer design. There are twenty channels in DNAMAN, any of them can be chosen as **Default Sequence Channel**. Default sequence is the one you are currently working on. Click a number on the Sequence channel bar will activate the corresponding channel as default.

Both DNA and protein sequences can be loaded into channels. DNAMAN checks the composition of a sequence to determine whether it is a DNA or protein fragment. DNA sequences are automatically loaded into channels. Protein sequences will be confirmed by users while DNA sequences will not. You may change the sequence type using the **Analysis Definition** function if sequence type is not correctly defined. The sequence in the default channel can be displayed in **Info/Sequence** bar.

III.3 Menus

DNAMAN has eleven **Main Menus** that are displayed on the top of the DNAMAN Sequence Analysis Studio. Each main menu has several menus, some of which have submenus. A command to carry out a function is described as a ***Main Menu* | *Menu* | *Submenu***.

Example:

a) Choose the ***File* |*Open*** command to open a DNAMAN file.

b) Choose the ***Sequence* | *Load Sequence* | *From Database*** command to load a sequence record from database into the default sequence channel.

III.4 Tool bars

There are several toolbars associated with the menu commands: **File/Edit** bar, **Sequence Analysis** bar, **Format** bar (PC only), **Browser** bar (PC only), **Info/Channel** bar and **Document Selection** bar (PC only). These toolbars are designed for most frequently used menu commands and allow users to access them

easily. You may show or hide any of these toolbars using the commands in the ***View*** menu.

1. ***File/Edit bar***
   This tool bar contains basic functions for file handling and content editing. You may create a new text file, open any type of DNAMAN documents, save and print any documents by accessing tools on this bar. You may also change the font of the document and access copy/paste/cut/undo/redo functions of the edit menu. You may zoom in/out a graphic document, or align objects of the graphic document.

   Clicking the activated button on the tool bar will process the corresponding command. Grayed out buttons indicate the commands are not available.

2. ***Sequence analysis bar***
   This tool bar contains ten frequently used analysis functions: Load Sequence from Selection, Analysis Definition, Restriction Analysis, Translation, Dot-Matrix Comparison, Multiple Alignment, Sequence Assembly, Sequence Search, PCR Primer Design and Drawing Restriction Map. The Option button is also on this tool bar. The Option command allows you to access the properties of a document or a drawing.

   Clicking the activated button on the tool bar will process the corresponding command. Grayed out buttons indicate the commands are not available.

3. ***Format bar (PC only)***
   This bar is used only for text documents. There are three commands for changing current font (Bold, Italic and Underline) and four for paragraph (Left, Center, Right-Justified and Insert of Bullets).

4. ***Browser bar (PC only)***
   This bar is used to browse Internet/Intranet information. It has basic functions of Microsoft Internet Browser: ***Back, Forward, Go Home, Stop*** and ***Refresh***. You may start a

new browser page by clicking the URL button. The animation button on this bar indicates some tasks may be in processing.

5. ***Log-cmd/Channel bar***
   This bar contains ***Log-cmd*** and ***Channel*** panels. In Windows version of DNAMAN, there are two additional panels: ***File*** and ***Database***. The ***Log*** panel is used to show the progress of current process such as Multiple Sequence Alignment and Sequence Assembly. The ***Channel*** panel shows the information or sequence of the current sequence channel. The ***Command*** panel allows users to enter DNAMAN related commands. The ***File (PC only)***panel can be used to navigate DNAMAN files and ***Database(PC only)*** to browse sequence databases.

6. ***Document Selection bar (PC only)***
   This bar contains all opened documents in the studio. You may switch the active document from one to another by clicking corresponding button in this bar.

III.5 Document types

There are eight document types in DNAMAN.

1) Text (Sequence) Document

MacOSX version: Sequence document uses general text format (***Text Document***). Text document is displayed in a **Text window**. DNAMAN provides a special **Text** editor for sequence editing. Sequences are save as plain text file.

PC version: Text document is displayed in a **Text window**. DNAMAN provides the **Rich Text** editor for sequence editing. Sequences should be save as plain text file and the ".seq" extension should be added to the file name. If you have inserted graphics, or modified text fonts in your document, the rich text format must be used in order to keep these objects. The ".rtf" extension name should be added for saving files with rich text format.

2) Restriction Map Document

Restriction map is saved in DMP format. DNAMAN map file is in binary format and stores restriction map objects. DMP file can be only opened and edited with DNAMAN.

PC version only: DNA map is an OLE object and DNAMAN is the OLE server. You may incorporate DMP files into any other OLE client applications (e.g. Microsoft or Corel Office). You may also copy a DMP file as graphic drawing object to any other programs that accept drawings. DMP format provides specific editing functions for drawing restriction map. You can use LBDraw to assemble many restriction maps of DNAMAN and make sophistic diagrams.

3) General Graphic Document

DNAMAN presents some analysis results, such as restriction pattern and dot-matrix comparison, in a graphic window.

MacOSX version: The graphic presentations are saved in image file format and can be copied and pasted between applications. DNAMAN saves graphic presentations in image format, however, does not provide functions to edit the saved image files. You can only edit these graphic presentations before saving them in image files. You may load an image file into DNAMAN and copy it to other drawing programs and perform modification.

PC version: The graphic presentations are saved in Enhanced metafile format. EMF is a clipboard format used to copy and paste graphics between Windows applications. DNAMAN saves graphic presentations in EMF format, however, does not provide functions to edit the saved EMF files. You can only edit these graphic presentations before saving them in EMF files. You may load an EMF file into DNAMAN and copy it to other drawing programs and perform modification. Some Windows graphic editing applications such as MS PowerPoint can be used to edit EMF files.

4) Multiple Alignment Document

DNAMAN saves multiple sequence document in MSD format. MSD file is in text format, however, includes special keywords for the Multiple Alignment Editor of DNAMAN (MASED). You can load MSD files in MASED and perform multiple sequence analysis. You may also open MSD files in text editor if necessary.

5) Sequence Assembly Document

Sequence assembly document is saved in SAF format. SAF file is also in text format and includes special keywords for the Sequence Assembly Editor of DNAMAN. Assembly editor is used to edit and visualize contigs. You may also open SAF files in text editor if necessary.

6) Trace File Document

DNAMAN can be used to view and analyze chromatogram files produced by automated DNA sequencing machines such as ABI series of Applied BioSystem. DNAMAN can open ABI and SCF files, however, cannot save any modification to these files.

7) Dot-Matrix Document

DNAMAN stores dot-matrix analysis results in graphic format.

8) Phylogenetic Tree Document

DNAMAN saves phylogenetic tree file in text format with extension PTR. A Phylogenetic tree file can be opened or edited as text.

Chapter IV Sequence and Text Files

This chapter introduces the DNAMAN sequence files. It also illustrates how to use, handle and edit your sequence files.

DNAMAN sequence files are in text format. You may edit a sequence file with the DNAMAN text editor or any text editors. In order to visualize or compute the length of any part of sequence, a fixed font must be used in the editor. DNAMAN uses **Fixed Font** as the default font the editor. You may change to other fonts using **Info|Settings** menu in Windows version, or **Preferences** menu in OSX version. When you open a sequence file with other word processor, choose a fixed font to correctly align the sequence.

IV.1 Characters in DNA sequence

DNAMAN recognizes the characters "A" (Adenine), "C" (Cytosine), "G" (Guanine), "T" (Thymine) and IUPAC code as components of a DNA sequence. These letters are case insensitive and both uppercase and lowercase letters are accepted.

IV.2 Characters in protein sequence

DNAMAN recognizes all the alphabetic letters as components of a protein sequence, except the letters "B", "J", "O" and "Z". Letter "X" stands for any amino acid. These letters are case insensitive and both uppercase and lowercase letters are acceptable.

IV.3 ORIGIN format

The "**ORIGIN**" keyword is used to separate the sequence content and comments in a sequence file. This keyword has to be ***placed in a separated line and at the beginning of the line***. Any text before the keyword is considered as comments and will not be involved in sequence analyses. If a sequence file does not contain the keyword, you may simply type "ORIGIN" at the beginning of the sequence content.

For example:

1. Nucleotide sequence file

```
ORIGIN
1  ATGACAAAAC ACTCATGTAT TACGGGAATG
31 ATGGTGTCTA TGGATCGTTC AATTGCATCT
```

2. Amino acid sequence file

```
ORIGIN
1  MTKHSCITGM MVSMDRSIAS CMIMHMLNQF
31 SCACESGIEY PATCASASIN V*
```

In absence of the ORIGIN keyword, a text file may still be used for analysis. The content of the file can be loaded into a sequence channel or used for other analyses (multiple sequence alignment, sequence assembly) upon user's confirmation.

IV.4 Annotations

You may define annotations for sequence analysis (**Defined Annotation**), or place annotations in sequence for visualization (**In-sequence Annotation**).

1. Defined Annotation

Defined Annotations are used in sequence analysis, such as translation of nucleotide sequence to protein sequence, and drawing sequence maps. Defined Annotations are placed before the keyword “ORIGIN”. The format is similar to GenBank files. Under the keyword “FEATURES”, you may define annotations in a DNAMAN sequence file. For example:

```
FEATURES
     mRNA            join(196..486,765..972,1422..1516)
                     /name="a-globin"
     CDS             join(358..486,765..972,1422..1516)
                     /gene="a-globin"
     terminator      1835..1837
     polyA_signal    1900..1905
     polyA_signal    1994..1999
     mRNA            join(6008..6230,6318..6522)
                     /name="b-globin"
     CDS             join(5825..5916,6008..6230,6318..6446)
```

```
                     /gene="b-globin"
ORIGIN
...
```

DNAMAN does not consider all entries under FEATURES as annotation. ALL ANNOTATIONS MUST BE DEFINED IN A SYSTEM FILE (e.g. Annotat.dat). For example:

```
DNAMAN Sequence Annotation list
////
total number=18
0 Intron
1 Exon
2 RBS
3 CDS
4 Sig_peptide
5 Mat_peptide
6 promoter
7 enhancer
8 polyA_signal
9 terminator
10 5'UTR
11 3'UTR
12 misc_binding
13 protein_binding
14 Repeat_region
15 repeat_unit
16 rep_origin
17 primer_binding
```

This file is saved in the DNAMAN system folder. You may edit it to add records. Removing records is not recommended, since databases may have used these records and deleting may result in misleading information of annotations.

2. In-sequence Annotation

If you like to include some information after the ORIGIN keyword in a text file, you can add it as In-sequence annotations. An In-sequence annotation will not be recognized as a part of the sequence in analysis. In-sequence annotations should be placed between parentheses '( )'.

Choose the ***File | Open*** command to display a sequence file on the DNAMAN Studio. After typing the annotations in the sequence, choose the ***File | Save*** command to save this file. Annotations can be more than one line, however, each line has to start with '(' and end with ')'.

Annotations in a sequence do not appear in the windows of sequence analysis results, such as sequence composition, conversions, alignment or translation.

Reformatting sequence using the ***Edit |Format | Sequence*** command eliminates all annotations in a sequence.

IV.5 Editing

The sequence editor has all the basic functions of a text editor, such as ***Font selection, Copy, Cut, Paste, Find ...*** You can format a text document with different fonts, paragraph styles.

Windows version only: You can also insert graphics into the editor from other Windows applications. In fact, the text editor provides the same functions as the Microsoft application **WordPad**. Consult Microsoft documentation for more information about **WordPad**.

The standard text editing functions are:

Undo:	reverses the last text edit command.
Redo:	reverses the last Undo command.
Select All:	selects the whole content in a text or a graphic window
Cut:	cuts the selected text and store it in Clipboard
Copy:	copies the selected text into Clipboard
Paste:	pastes the content from Clipboard
Delete:	deletes the selected text
Clear all:	deletes the whole text in a text window
Find:	finds a text
Replace:	replaces a text by another
Next:	finds or replaces the next occurrence.
Font:	changes the font, style and size of the letters in a sequence.
Format:	changes the formats of paragraph, tab, word wrap and bullet style.

The sequence specific editing functions are:

Uppercase: changes the selected text to uppercase letter.
Lowercase: changes the selected text to lowercase letter.
Sequence Format: groups the selected sequence to make it easy to read, count and compare.

IV.6 Formatting sequence

By default, DNAMAN displays sequences in six columns and ten letters per column, or in one column with sixty bases (or amino acids). In some cases, you may want to arrange sequences differently. To format sequence:

1. Select the sequence you want to change.
2. Choose the ***Edit | Format | Sequence*** command to open the **Sequence Format** dialogue box.
3. Type a number in the **Column number** box to define the number of columns per line.
4. Type a number in the **Characters per column** box to define the number of bases or amino acids per column.
5. Uncheck the ***Number Label*** box if you do not like to label sequence position.
6. Click **OK** to format the sequence and then save this document.

The ***Edit | Font*** command allows you to change the Font, Style and Size of the letters of your sequences. You can format sequence in different styles by using the functions of ***Edit | Format | Sequence*** and ***Edit | Font***.

IV.7 Sequence input from keyboard

Choose the ***Edit | Enter Sequence*** command to open the Enter Sequence dialogue box. Type letters in this box. The cursor position is indicated in the **Position** box.

To facilitate entering the sequence, the numbers 1, 2, 3, 4 and 0 can be used, which correspond to the letters A, C, G, T and N, respectively.

After the sequence has been entered into the box, click the **OK** button to display the new sequence formatted with ORIGIN (see the section **IV.3**) in a text window.

You may also choose an empty sequence channel and type letters directly in the **Sequence Content** bar.

Proofreading *(Windows version only)*

To check whether the correct nucleotides have been entered from the keyboard, you can use the Proofreading function. By clicking the **Speak** button in the **Enter Sequence** dialogue box or the **Analysis Definition** dialogue box, you can open a **Speak Sequence** box. To use this function, a sound driver must be installed on your system.

Place the cursor at any position and press the **Start** button to start the reading. The default reading speed is 120 bases per minute and it can be changed. Click the **Pause** button to stop the speaking whenever you want.

IV.8 Exporting data

DNAMAN allows you to **copy** any text or graphics into the clipboard so that you can **paste** it to your document. Using this exporting method, you can easily integrate any sequence analysis result (text or graphics) into your publication and make an impressive presentation.

IV.9 Printing

Choose the ***File | Print*** command to print out entire text content in a text window.

Chapter V General Graphic Document

This chapter introduces graphic documents used in DNAMAN.

DNAMAN presents some analysis results using graphic presentations. These results can be saved as general graphic document in EMF format. The following analysis results or information windows are in this category:

- Graphic panel of motif search
- Graphic panel of sequence assembly
- Dot matrix comparison
- Homology trees and phylogenetic trees
- Restriction pattern
- Map reconstruction
- Translation overview
- Profiles of protein hydrophobicity, hydrophilicity and secondary structures
- Profiles of DNA thermodynamic properties
- Primer self-complementarity
- Information windows of Genetic Code, Methylases, Amino acids and Nucleotides.

V.1 Editing

DNAMAN provides various editing tools for specific graphic presentations. Depending on the category of sequence analysis, you may consult corresponding chapters for the handling of the presentation objects.

Adding Text Object is a common editing tool for all general graphic documents. To add a text object in the graph, choose the ***Edit | Add Text*** command. The **Text** dialogue box is opened. You may write any text contents that will be added in the graphic window. You may change the font, style, size and color of the text, and display it horizontally or vertically.

You may modify the newly added text objects by double-clicking them.

V.2 Copying

MacOSX version: DNAMAN allows you to copy entire or partial graphic content into the Clipboard and then export it to other applications.

1. Choose the ***Edit*** **|** ***Select All*** command to select the whole graph.
2. Or place the cursor in a graphic window. Press the left mouse button and then drag it to select the graphic content you want to copy.
3. Choose the ***Edit*** **|** ***Copy*** command to copy the selection to the Clipboard.

Windows version: DNAMAN allows you to copy an entire or partial graphic content into the Clipboard and then export it to other applications for Windows. There are two formats: EMF and BMP.

Copy in EMF format:

1. Choose the ***Edit*** **|** ***Select All*** command to select the whole graph.
2. Choose the ***Edit*** **|** ***Copy*** command to copy the selection to the Clipboard. The graphic content in the Clipboard is in EMF format.

Copy in Bitmap (BMP) format:

1. Place the cursor in a graphic window. Press the left mouse button and then drag it to select the graphic content you want to copy.
2. Choose the ***Edit*** **|** ***Copy*** command to copy the selection to the Clipboard. The graphic content in the Clipboard is in BMP format.

Note: EMF format shows better graphic quality than BMP format.

V.3 Saving graphic files

You may save a graphic file in EMF format by choosing the ***File*** **|** ***Save as*** command. This file can be opened in DNAMAN. However, the existing content cannot be changed (you may still add texts in the graph). DNAMAN does not provide tools to

modify the saved files. You may use other graphic applications such as PowerPoint to edit the graphic files.

V.4 Printing

Any graphic content can be printed using the ***File | Print*** command. Choose the command to print out the whole graphic content in a graphic window.
You may reduce or enlarge the graphics in size by clicking the Zoom toolbars  .

V.5 Profile plotting

Profile plotting is displayed in graphic windows. This section explains how to handle the plot.

DNAMAN may illustrate the properties of DNA or protein molecules in a graphic window by plotting the property values against the positions. The results of three functions are shown in profile plotting window: 1) DNA thermodynamic properties, 2) Protein hydrophobic/hydrophilic profiles, and 3) Protein secondary structure profiles.

The plot is presented in a rectangle with the property values as ordinate and positions as abscissa. When you place the cursor (cross shape) on the plot, the current property value and position are displayed on the left corner of the screen (Status bar). In addition to the tools described in the above sections for handling a graphic window, the following tools allow you to perform more detailed analyses on the profiles.

1) Graphic positions
You may move the plot positions within the graphic window. Place the cursor on the up-left corner of the plot. When the cursor switches to Cross, press and hold the left mouse button and drag the plot.

2) Graphic sizes
You may change the size and shape of the plot. Place the cursor on the low-right corner of the plot. When the cursor switches to

⇔, press and hold the left mouse button and drag the plot to desired size and shape.

3) End positions

There are four buttons on the top of the graphic window. The third and fourth buttons control the end positions of the plot. To change the left end, you may press the third button then enter a Start position in the text box. You may also press the thumb beside the button to increase or decrease the number. Use the same method to changes the right position of the plot by operating the fourth button.

4) Export data

You may export the plot data to a text file by pressing the **Export** button. DNAMAN lists all plotting data in a table. You may save the table in a file and import it to other graphic/table programs such as Microsoft Excel.

5) Options

You may modify the properties of the plot by pressing the **Options** button.

In the **Options** dialogue box, DNAMAN lists all sets of data in the plot. You may check or remove any of the sets for display in the plot.

The **Window Size** parameter is used to average the property values. Increasing the **Window Size** may smooth the profile. If the size is 1, the profile represents the exact property value of each position. If the size is greater than 1, DNAMAN will repetitively average the property values for each position in the plot.

You may change the scale of ordinate by defining the values of **Scale from** and **to** (Y-Scale).

Chapter VI Sequence Handling and Conversion

This chapter explains how to handle DNA and protein sequences in DNAMAN.

VI.1 Loading sequence into channels

Active sequence can be selected once and analyzed using available tools. The active sequence must be loaded into a sequence channel. DNAMAN provides 20 sequence channels to keep active sequences in memory. You may switch from one sequence to another during different analyses without going back to disk files.

A sequence channel contains not only the sequence, but also related information: 1) Sequence name, 2) sequence type (DNA or Protein), 3) Linear or circular DNA, 4) Analysis region and 5) Annotations.

There are several methods to load sequences into sequence channels. The easiest approach is to open a DNAMAN or GenBank sequence file in the Workspace and the sequence is loaded into the default channel automatically. This option can be turned on/off using the ***Info | Settings*** menu and check the ***Load sequence in channel when opening file*** option in the ***File*** page of the ***Settings*** dialogue box.

Other methods of loading sequence into channel are shown in the six sub-menus of ***Sequence | Load Sequence*** menu.

VI.1.1 From selection

If a sequence file is opened in a text window, the ***From Selection*** command is activated by selecting a sequence content. Choose the activated ***Sequence | Load Sequence | From Selection*** command to load the selection into the default sequence channel. You may load any part of text into a sequence channel. If annotations present in the selection, they will be automatically excluded from the sequence content. The "ORIGIN" keyword is ignored and has no effect.

When should I use manual loading?

1) My sequence is only a part of the document
2) I will analyze only a part of entire sequence
3) The sequence is already opened in a text window
4) The sequence is very long (> 1000 kb) and other method may take too long time.

Windows version only: This function applies also to the Browser of DNAMAN. You may load a sequence directly from a browser window.

VI.1.2 From sequence or GCG file

DNAMAN sequence files (with ORIGIN format), or GCG files can be retrieved directly from disk.. Choosing the ***Sequence | Load Sequence | From Sequence File*** or ***GCG File*** command will open an **Open File** dialogue box.

Choose the sequence file and press the **Open** button to load the sequence. If the file name is not in sequence type (*.seq), you may select all files (*.*) in the **File of type** section.

VI.1.3 From GenBank file

A GenBank file may contain more than one sequence. The sequences all start with the ORIGIN, and end with //. Annotations may also be described in the file. Any sequence in the file can be directly loaded into the DNAMAN memory using ***Sequence | Load Sequence | From GenBank File***. When you click a sequence name in the sequence list box, the sequence information appears in the **Information** box. You may choose different individual sequence, and click the **Load** button to load it into the default sequence channel.

VI.1.4 From database

The ***Sequence | Load Sequence | From Database*** command is activated by selecting a database as the default database. Use this command to open a **DNAMAN Database** dialogue box. When you click a record name, the information of the record appears in the **Information** box. Click the **Load Seq** button to load this sequence into the default sequence channel.

VI.1.5 Loading multiple sequences

You may simultaneously load many sequences into sequence channels using the ***Sequence | Load Sequence | Multiple*** command. These sequences can be in one or many files (e.g. multiple alignment files and Genbank files may contain many sequences). The sequences will be loaded into the default channel and the next available channels.

Large sequences, such as genomic sequences, may be stored in a few files. You may combine all the sequences from different files into one by using the ***Sequence | Load Sequence | MultiSeq into One Channel*** command.

VI.2 Analysis definition

You may change the sequence properties using the **Analysis Definition** function.

Choosing the ***Sequence | Current Sequence | Analysis Definition*** command opens the **Analysis Definition** dialogue box.

You can define the following properties of a sequence:

1) **Name**: the name of a sequence is a string less than 32 letters. Any letter can be used in the name, but not SPACE (" ").
2) **Type**: DNAMAN provides an option for users to change the type of sequence. Use this option if DNAMAN has assumed a wrong type to your sequence. If you choose a different sequence type than its original one, DNAMAN will ask you to confirm change and then close the dialogue box.
3) **DNA sequence type**: you may define the sequence as circular or linear DNA.
4) **Analysis region**: you may define a region, instead of the entire sequence, for analysis. The region definition is especially useful if you like to perform restriction analysis or translation on a part of DNA sequence. To check

whether the region has been defined correctly, click the **Show Ends** button to show the first 10 and the last 10 letters of the defined region. Clicking the **All** button defines the entire sequence for analysis.

5) **Initial position**: If the sequence is DNA, clicking the **Rotation** button allows you to define any position of a circular DNA as the initial point. This function is useful especially after using the DNA Cloning function. Rotation of cloned target sequences can help you to better organize the order of elements, e.g. Vector and Insert.
6) **Proofreading (PC version only)**: you may proofread a DNA sequence by clicking the **Speck Sequence** button.
7) **Annotations**: you may add or remove annotations of the sequence by clicking corresponding buttons.

VI.3 Sequence conversion

DNAMAN can display the current sequence and its related sequences such as reverse complement sequence and double stranded sequence. When a single-stranded sequence is shown, DNAMAN also indicates its nucleotide composition and the predicted molecular weights.

Choosing the ***Sequence | Display*** command invokes a dialogue box. You may check six options for DNA sequences: 1) Sequence and Composition; 2) reverse sequence; 3) complementary sequence; 4) reverse complementary sequence; 5) double stranded sequence and 6) RNA sequence. If the default sequence is ***Protein***, only the first option is available.

If annotations have been defined in a sequence, you have options to include or exclude any of these annotations. The required sequence information will be displayed in a text window. With the features of **Excluding** and **Display Only** sequence annotations, you may exclude some sequences (e.g. Introns) from the original sequence, or get exclusively the interesting regions (e.g. Exons). The annotations can also be displayed as **lower case** while the other part of sequence shown as **upper case**.

VI.4 Draw sequence map

You may draw a map of the current sequence using the ***Sequence | Draw Sequence Map*** command. If elements have been defined within the sequence, DNAMAN will incorporate them in the map. You may also add or remove the elements by editing the map. Sequence maps are in DMP format. See the section **XI.4** about how to modify a DMP file.

VI.5 Plot DNA properties

You may plot the thermodynamic properties of a DNA molecule using the ***Sequence | Plot DNA Properties*** menu. The chart is drawn in a graphic window. See the section **V.5** about handling of the plot window.

The chart contains three curves: GC Content (GC%), Melting Temperature (Tm), and Entropy (dH).

The Tm, or melting temperature, characterizes the stability of the DNA hybrid formed between an oligonucleotide and its complementary strand. The Tm is calculated using the nearest-neighbor thermodynamic values method (Breslauer et al, 1986, ***Proc. Natl. Acad. Sci.*** 83:3746).

Tm=dH/(A+dS-Rln(Cd))-273.15+16.6*(log10(Cs))

Where:

dH is the sum of the nearest-neighbor enthalpy changes for hybrid formation,

dS is the sum of the nearest-neighbor entropy changes for hybrid formation,

R is the molar gas constant (1.987 cal K-1 mol-1),

A is a constant for helix initiation,

Cd is the DNA concentration, and

Cs is the salt concentration.

VI.6 Trace files

DNAMAN reads in ABI and SCF sequence trace files and graphically displays the results. Use the ***File | Open Special | ABI/SCF Trace*** menu and choose a trace file to open. The data for each nucleotide is plotted and the assigned nucleotide (G, A, T, C or N) in the trace file is overlayed on the graphs

The trace plot graph properties are controlled by the left panel. A horizontal slide controls the density of sequence peaks. You may condense or enlarge the graph by sliding to left or right. The left vertical slide controls the height of trace peaks and the right vertical slide can be used to amplify trace signals.

Any of the four signals can be turned on and off by check/uncheck the corresponding letter. You may also change the color for each nucleotide by double-clicking the letters.

For comparison of trace graph, you may load up to 16 files in the same panel by pressing the "***Add File***" button. All trace files are managed in a file list box. You may display or hide files in the list by check/uncheck their left buttons.

VI.7 Secondary Structure Prediction

DNAMAN predicts secondary structures of DNA/RNA molecules using the algorithm developed by Zuker and colleagues (Zuker and Stiegler, 1981, ***Nucleic Acids Res*** 9: 133). Energy parameters have been used from recent development (Mathews et al, 1999, ***J. Mol. Biol.*** 288: 911, Walter et al, 1994, ***Proc. Natl. Acad. Sci.*** 91: 9218, SantaLucia, 1998, ***Proc. Natl. Acad. Sci. USA*** 95: 1460, Allawi & SantaLucia 1997, ***Biochemistry*** 36: 10581, Allawi & SantaLucia 1998, ***Biochemistry*** 37: 2170). The algorithm uses dynamic programming to find out minimum energy secondary structures of a DNA/RNA molecule. Free energy is calculated using the Nearest-Neighbor method.

Under the ***Sequence | Secondary Structure*** menu item you may choose the ***Current Sequence*** or ***All Available Sequences*** for folding. These commands open a dialogue box. You may modify the sequence if the ***Current Sequence*** is used. DNAMAN may fold DNA and RNA molecules and you must decide which type

to use. After folding, DNA shows the structure in a graphic diagram. If you like to see the structure in text format, choose the **Show structure text in graph** option. There are other parameters used in the folding methods such as energy parameters, maximum sequence length, folding temperature ... You may find them using the Info | Settings menu. However, it is not recommended to modify them except for experts.

Press the **OK** button to perform analysis. After finishing analysis, DNAMAN displays the secondary structure in a graphic window. The minimum free energy of the structure is also shown in the graphic window.

The function of secondary structure prediction is threaded, therefore, you may use other tools of DNAMAN during the analysis.

You may evaluate the free energy of a given secondary structure. Choosing the ***Sequence | Secondary Structure | Structure Energy*** menu opens a dialogue box. Enter a sequence in the top edit box and its structure in the bottom edit box. Press the **Energy** button and the free energy is calculated using available information.

VI.8 Random sequences

In some cases, you may need to work on random sequences. DNAMAN generates random sequences from three approaches:

1) New random sequence

   Use this function to produce a new random sequence. You can define the composition in each element.

2) Current sequence

   Use this function to randomize the current sequence. The generated sequence has the same composition as the current sequence, but the order of all elements is randomized.

3) Fragmentation

Use this function to make fragments of the current sequence. You may produce a number of random fragments from the current sequence. The random sequences are stored in the default folder. You may find out the path of the default folder using the ***Info*** **|** ***Settings*** menu.

Chapter VII Sequence Assembly

This chapter explains how to assemble DNA fragments and edit contigs derived from assembly.

VII.1 Assembly of DNA fragments

Choosing the ***Sequence | Sequence Assembly*** command shows the **Sequence Assembly** dialogue box. In order to achieve the best results, adequate parameters should be chosen for processing sequence assembly. The number of sequences for assembly is currently limited by 30,000.

VII.1.1 Sequence sources

DNA fragment data can be retrieved from sequence files, sequence channels and DNA databases. Press the **Add File** button to retrieve sequences from disk files, the **Folder** button from a folder (all files in the folder), the **Channel** button from channels, or the **Database** button from the default database. You may remove a sequence from the list by selecting it and pressing the **Remove** button, or remove all sequences in the list by pressing the **Clear** button.

The sequence files for assembly may be in DNAMAN format with the **ORIGIN** keyword. You may also use any other common sequence formats. Trace files (ABI and SCF) derived from automated sequencing are also accepted in sequence assembly. DNAMAN also accepts files in FASTA, GenBank, SAF and MSD formats.

DNAMAN may remove ambiguous flanking regions of a DNA fragment. You may check the **Remove flanking regions when >** option, and set a ***Number*** of ambiguous bases in a considering ***Window***. For example: 2 ambiguous bases in 10. In this case, DNAMAN calculates ambiguity of every 10 bases. Flanking regions with >2 ambiguous bases will be removed.

There is an option to **Remove vector sequence** from all source sequences. In this case, you must load the vector sequence into the default sequence channel. If there are many vectors for the source sequences, it is recommended to combine all vectors in

one sequence file and load it into the default sequence channel. When this option turned on, DNAMAN will compare all input sequences against the vector. Flanking regions containing the vector sequence will be removed. The shorter the vector sequence is, the faster the comparison will be. Therefore, you can accelerate the process by removing unnecessary sequence in the vector.

When you work with a sequencing project, it is suggested to make a database for the project. You can save all the sequences as well as vectors of the project in the database. The database approach makes your sequencing project well organized and facilitates sequence assembly.

VII.1.2 Methods of sequence assembly

DNAMAN assembles sequences in two steps. The first step is alignment to produce scores for assembly. The second step is to assemble fragments using the score matrix. To generate alignment score matrix, you may choose **Quick Alignment** or **End Comparison** method. The **Quick Alignment** method is recommended in most cases of sequence assembly.

The **End Comparison** method compares progressively the end of each fragment. No insertion is introduced to the sequence ends. Overlapping is determined according to the qualification criteria. The **End Comparison** method should be used only for a small number of fragments while the overlaps are short and quality of overlapping is high.

With the **Quick alignment** method, alignments are performed using the quick alignment algorithm (Wilbur & Lipman, 1983, ***Proc. Natl. Acad. Sci.*** USA, 80:726) with all sequences in both plus and minus strands. The sequences will be assembled according to qualification criteria. Inserts will be added in overlapping region if necessary.

There are two methods for **Final Assembly**: **Quick** and **Optimal**. **Quick** method uses Wilbur & Lipman algorithm and will complete assembly in short time. **Optimal** method uses

Smith & Waterman and provides better assembly results. **Optimal** method is recommended in most cases.
There are two approaches to assemble sequences: **de novo** and **Re-sequencing**. **Re-sequencing** will use the longest sequence in the list as template and assemble all other sequences to the template. **de novo** approach does not use any template and will try to align all sequences according to score matrix. If template available, it is recommended to use **Re-sequence** approach since it provides best results and uses less computation resources.

Parameters for sequence assembly are:

- *Minimum Overlap*: Minimum number of nucleotides in overlaps. The default value is 80 bases. Reduce this number if overlaps are shorter.
- *Identity* >=: Minimum percentage of identity in overlap region. The default level is 90%. Reduce this number if ambiguity level seems high.
- **Quick Alignment** method parameters: **k-tuple, gap open penalty, gap extension penalty** and **window size** and diagonals per 500bp. The default values are 4, 6, 0, 4 and 3, respectively. These settings are optimal for general sequence assembly.
- **End Comparison** method: ***Maximum Overlap***. This parameter defines maximum length that the overlap region of any pair of sequences.

VII.1.3 Assembly analysis

Click the **Assemble** button to start searching for overlapping sequences. DNAMAN performs firstly pairwise comparison of all input sequences and keeps updating the results. The number of overlaps found during searching is indicated.

The process of sequence assembly is threaded, therefore, you may perform other tasks during sequence assembly.

There are two options for processing sequence assembly.

1) **Minimize window while running**. The Assembly dialogue box will be minimized while processing sequence assembly. You may check this option if you like to perform other tasks during sequence assembly.
2) **Reorder sequences**. The sequence order will be changed from the original input order. All sequences in each contig will be grouped together.

   DNAMAN will keep trace file information in the assembly results. It will show the positions in the trace file when you click sequence bases in the Sequence Assembly Editor.

You may stop assembly any time during the process by pressing the **Stop** button.

If all sequences cannot be assembled to one contig, they may be divided into different groups. The number of contigs is indicated during assembly process. Isolated sequences will be ignored in assembly result. Click the **Show Result** button to display assembly results in a Sequence Assembly Editor.

VII.2 Sequence assembly editor

The Sequence Assembly Editor facilitates the visualization and editing of assembled sequences. There are three windows in a sequence assembly editor: **Name list window, Sequence window** and **Graphic window.**

VII.2.1 Name list window

All names of assembled sequences are shown in this window. The corresponding sequence is right beside the name of the Sequence window. You may change the display order of these sequences by drag-and-drop the names to move them up or down. You may also change the name of any sequences by double clicking it.

VII.2.2 Sequence window

All sequences are listed in this window. The consensus sequence is place on top of the window. You cannot directly edit the consensus sequence, however you may modify it by editing each overlapping sequence. Any modification of source sequence may result in the change of the consensus sequence.

The edit functions for assembly sequences are the following:

1) Add or remove bases
2) Add or remove gaps
3) Select a block of one sequence and delete it.
4) Change selected sequences to uppercase letters
5) Change selected sequences to lowercase letters
6) Find a sequence in the project.

VII.2.3 Graphic window

The graphic window contains the diagram of sequence assembly. There are three elements in the graph.

1) A straight line represents the consensus sequence. The length and position of the consensus sequence is indicated at the two ends of the straight line.

   You can move the whole diagram within the graphic window by placing the cursor at the **beginning** of the consensus sequence. When the cursor switches to  , press the left mouse button and then drag and drop the diagram to the location you want. You can also change the length of the straight line (consensus sequence). Place the cursor at the **end** of the consensus sequence. When the cursor switches to east-west arrow shape, press the left mouse button and then drag and drop the end to anywhere

2) If a rectangle appears on the straight line in the diagram, it indicates an interruption in the consensus sequence. The rectangle is used to divide different sequence groups. You may make the interruption larger or smaller by moving it to left or right. In this case, the sequences in Sequence window will move according to your adjustment.

3) Each source sequence is shown as a double arrow line. Arrows indicate the directions of the sequences. By pointing the cursor to an arrow line, you can move it upwards or downwards. The names of all input sequences can be displayed on the arrow or at the left side of the panel.

See the section **V.1** for adding text contents in the **Sequence Assembly** diagram.

VII.2.4 Options of sequence assembly editor

Click the **Options** button in the Graphic window to select options for the editor.

1) **Show sequence row in color.** Each sequence will be show with color background to align with name window and facilitate sequence identification.
2) **Show double stranded sequence.** The consensus sequence and all input sequences are shown in double-stranded form.
3) **Show sequence name in graph**. You may show the name on the left side of the graphic panel or on the sequence arrow.
4) **Show translation**. DNAMAN may translate the consensus sequence in three (single strand) or six (double stranded) reading frames.
5) **Show ambiguity in consensus**. If this option is checked, DNAMAN shows IUPAC code in ambiguous positions of consensus sequence. Otherwise, the most present base is placed in such positions.

VII.2.5 Sequence export

Click the **Export** button in the Graphic window to export sequences in a text window. You have the option to export all sequences or the consensus sequence only for short assemblies. Long assemblies will be exported consensus sequences only.

Chapter VIII Sequence Search

DNAMAN provides many methods of searching for sequences. You may search for a set of DNA or protein sequences, and display the results in a graphic window. You may also search for direct repeats, mirror repeats, inverted repeats or stem-loop structures. DNAMAN also searches for amino acid sequences and possible open reading frames of a given DNA sequence.

VIII.1 BLAST Search

To search for homologous sequence, you may query a BLAST Web Server, on the Internet such as http://www.ncbi.nlm.nih.gov/blast/ or your local BLAST server. The server address can be defined with the ***Info | Settings*** command in the ***Misc*** tab. The query sequence will be compared with the DNA or protein sequence databases in NCBI using Basic Local Alignment Search Tool (Blast). You may find more information about the BLAST server from the NCBI web site. You may also refer to the publication of Altschul, S.F. *et al*., 1990, *J. Mol. Biol.* 215:403.

DNAMAN formats the default sequence to query the BLAST Server. There are five formats:

1) **Blastn** compares a nucleotide query sequence against a nucleotide sequence database.
2) **Blastx** compares the six-frame conceptual translation products of a nucleotide query sequence against a protein sequence database.
3) **Tblastx** compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.
4) **Blastp** compares an amino acid query sequence against a protein sequence database.
5) **Tblastn** compares a protein query sequence with a nucleotide sequence database dynamically translated in all six reading frames.

DNAMAN invokes Web browser to access the BLAST server after the menu command chosen. Default parameters are usually used in the search. You may modify the search parameters according to your needs. The server provides guidance on how to choose the parameters and you may access the help by clicking the parameter name.

You may use the ***Sequence | Current Channel | Analysis Definition*** command to define any region of the default sequence as a query sequence. You may also modify the content of a Blast document by deleting or adding any nucleotide or amino acid sequence.

VIII.2 Searching for nucleotide sequences

DNAMAN searches for nucleotide sequences from both strands of the current DNA sequence and presents the searching results in a graphical window or a text window.

Choose the ***Search | Sequence*** command to open the Search window, in which you can search for nucleotide sequences and defined consensus sequences as well.

VIII.2.1 Query formats

Clicking the **Query** button opens the **Enter Sequence** dialogue box. Type the nucleotide sequence to search. There are several searching formats (the letters are case insensitive):

type	AGGCNNNGATG
search	AGGCNNNGATG *(N = A, C, G, or T.)*
type	AGGC(N3-10)GATG
search	AGGC and GATG with 3-10 bases between them.
type	AGGC(X)GATG
search	AGGC and GATG with any bases between them.
type	AGGC[**ACGT/TCGA**]GATG
search	AGGC**ACGT**GATG or AGGC**TCGA**GATG

type	TGATGA(N3-12)CTGA(N5-9)GTCAGT(N2-6)TGCA(N8-66)GATCC(N2-8)GCATT <mismatch=2 1 1 1>
search	All six elements with defined base number between and with mismatch less than the required values.

In the last format, there is no limitation of element number. You may also use **IUPAC** code in any of the formats. Choose the ***Info* | *Nucleotides*** command to display the IUPAC code table (see the section **XVI.6**).

For example:

type	AGGWCGAT *(W = A or T.)*
search	AGGACGAT or AGGTCGAT

Type "GATT", for example, in the **Search** dialogue box and then click the **OK** button to display the list of the found sequences.

Motif search can be done with unlimited elements in query sequence, e.g. TGATGA(N3-12)CTGA(N5-9)GTCAGT(N2-6)TGCA(N8-66)GATCC(N2-8)GCATT <mismatch=2 1 1 1>. The motif search allows mismatch defined for each element.

DNAMAN allows you to search for more than one sequence and list all sites. Click the **Query** button again to open the Search dialogue box. Type another sequence, e.g. AATAAA, in the dialogue box and click the **OK** button. DNAMAN will add the found sequences in the list of the searching results as well as in the searching diagram. The number following the found sequence indicates the group of the query sequence.

Clicking the **Export** button displays the searching results in a text window. You can print the text by using the printing command.

There are two panels in the search window: **Sequence List** and **Graphic Presentation**.

VIII.2.2 Sequence list window

Sequence List window contains four columns.

1) **Position**: the positions of the subsequences. There is a check box beside the position. You may check and uncheck the box. The subsequence position will be indicated in the **Graphic Presentation** if it is checked.
2) **Strand**: the + or - strand where the subsequence is located.
3) **Sequence**: the subsequence and its flanking regions. The flanking regions are usually 5 bases upstream and downstream of the subsequence.
4) **Group**: the group the subsequence belongs to and the initial query sequence. You may search for many queries on the target sequence. Each query results in many subsequences that belong to the same group.

You may change the size of each column by resizing the corresponding header.

VIII.2.3 Graphic presentation

DNAMAN also displays the searching results in a graphic presentation. There are two lines in the presentation; the upper line (**Sequence Line**) represents the target sequence and the lower line (**Zoom Line**) shows the zoom of the upper line.

The graphic presentation is a general graphic document. You may handle it with tools described in the chapter V (Adding text objects, copying graphs…).

The positions of all found subsequences are shown on the **Sequence Line**. Different subsequence groups are indicated with different colors. If a subsequence is checked in the **Sequence List** window, its position is marked with an arrow.

How to zoom: Place the cursor on one of the two long-vertical lines, the cursor switches to an up-arrow ⇧. By moving the arrow you can focus on a smaller region. While you sliding the up-arrow along on the **Sequence Line**, the position is indicated on the left corner of the screen.

How to display sequence: Place the cursor on the **Sequence Line** or **Zoom Line**. Press and hold mouse left button while moving to select a region. Release the mouse button and a dialogue box appears to show the selected sequence.

How to move graph: Place the cursor at the **left** end of the **Sequence Line**. When the cursor switches to  , press the left mouse button and then drag and drop the diagram to wherever you want. You may change the relative position of the two lines by moving the **Zoom Line**. Place the cursor at the left end of the **Zoom Line**, press the left mouse button and then drag and drop the **Zoom Line** to an appropriate position.

How to resize graph: You may change the length of the **Sequence Line** and **Zoom Line**. Place the cursor at the right end of the **Sequence Line** or **Zoom Line**. When the cursor switches to ⇔, press the left mouse button and then drag and drop the line to an appropriate position.

You may remove the checked subsequences by clicking the **Rem. Check** button, or clear the list by using the **Rem. All** button.

VIII.2.4 Sequence searching options

Clicking the **Options** button opens the **Sequence Searching Options** dialogue box. You may modify the following parameters:

1. **Searching region**. You may search entire or a part of the current sequence. Define start and end positions of the searching region. *Note that the Zoom region should be within the defined searching region*.
2. **DNA both strands**. You may search for a nucleotide sequence on both strands of the current sequence by

checking the option. Uncheck this option if you want search only on the sense sequence. This option does not apply to protein sequence.

3. **Show scale**. You may display the scale on the searching diagram by checking the **Show scale** option. The size of scale may be modified.

4. **Show zoom**. You may display or hide the **Zoom Line** on the searching diagram by checking/unchecking the **Show zoom** option. The zoom positions can be changed to desired values.

5. **Color mark**. Mark the searching results in different colors in the diagram by checking the **Color Mark** option.

6. **Line size**. You may change the line thickness of the **Sequence Line** and **Zoom Line**. The size of each mark can be modified in the **Mark Height** box.

7. **Group**. You may **Check/Uncheck** one or more groups in the subsequence list. You may also **Remove** one or more groups from the list. Enter the numbers of the groups in appropriate boxes. Numbers should be separated by a space if you enter more than one number.

VIII.2.5 Searching for consensus sequences

You may search for consensus sequences, such as promoters and regulatory factor binding sites, in the graphic search window.

Clicking the **Consensus** button opens the **Consensus Sequence** dialogue box. You may select individual consensus by clicking the sequence name. The related information will be displayed in the **Sequence Information** box. You may also search for the entire list of the consensus sequences by clicking the **Search All** button.

The consensus sequence information is stored in a data file. You can edit it by clicking the **Edit** button. There is a simple format for editing the data file. The information for each consensus sequence is separated by the "//" and consists of three lines, the

name, the consensus sequence and the related information. After editing, choose the ***File*** **|** ***Save*** command to save the changes.

IUPAC code can be used in a consensus sequence.

VIII.3 Searching for a set of sequences

You may search for a set of sequences, such as a list of consensus sites. The results are shown in text format.

Choose the **Search | Nucleotide Sequence Set** or **Protein Sequence Set** command. A dialogue box appears where you may choose the file containing the search list.

Click the **File** button to browse and select the list file. The list file may be in DNAMAN consensus format or a user-defined format. If **DNAMAN consensus sequence list** is chosen, you may ignore the parameters thereafter. A sample of DNAMAN consensus file may be found in the CONSENS folder of the DNAMAN program. If **User-defined sequence list** is chosen, you should enter the information about the list file format. DNAMAN extracts listed sequences according to provided information.

1. **Name, Sequence and Reference in one line**. Check this option if each line defines a sequence in the file.
2. **Symbol for the beginning of data**. A list file may contain some necessary information prior to the sequence list. The symbol, e.g. "ORIGIN", is used to separate the information and sequence list.
3. **Symbol or position for the beginning of sequence**. Enter the start position of a sequence if **Name, Sequence and Reference in one line**. Otherwise, enter the keywords.
4. **Symbol or position for the end of sequence**. Enter the end position of a sequence if **Name, Sequence and Reference in one line**. Otherwise, enter the keywords or nothing.

Example 1:
List file content:

```
CONSENSUS ID   SEQUENCE              COMMMENTS
ASD23          GGGCGTAACCCATTTTC  ! reference book
…
…
…
```

You should

1. Check **Name, Sequence and Reference in one line**
2. Enter "CONSENSUS ID" in **Symbol for the beginning of data**.
3. Enter "15" in **Symbol or position for the beginning of sequence**.
4. Enter "!" in **Symbol or position for the end of sequence**.

Example 2:
List file content:

```
ORIGIN
NAME ID:
  ASD23
DNA Sequence:
  GGGCGTAACCCATTTTC
Reference:
  reference book here
…
…
…
```

You should

1. Uncheck **Name, Sequence and Reference in one line**
2. Enter "NAME ID" in **Symbol for the beginning of data**.
3. Enter "DNA Sequence:" in **Symbol or position for the beginning of sequence**.
4. Enter nothing in **Symbol or position for the end of sequence**.

The searching results are shown in a text window. If a sequence is found, DNAMAN displays

1. The consensus name and sequence.
2. All found sites in the plus strand of DNA sequence with positions, actual sequences and flanking regions +/-5 bases.
3. All found sequences in the minus strand.

DNAMAN does not show any information on the sequences if they are not found.

VIII.4 Searching for repeat sequences

DNAMAN searches for direct repeat and mirror repeat sequences. Choose the **Search | Direct Repeats** or **Mirror Repeats** or **Inverted Repeats** menu. Type the minimum length (bases) of a repeat sequence. The result shows the repeat sequences found in the current sequence. If a sequence repeats more than twice, the repeat will not be directly indicated in the results. You can find it by checking the position and the sequence of each repeat. For example, if the sequence AAGCTGCGTG appears at positions 50 and 205, and AGCTGCGTG is at 506, the results will be:

Length(bp)	Sites	Direct Repeat Sequence
10	50, 205	AAGCTGCGTG
9	51, 506	AGCTGCGTG
9	206, 506	AGCTGCGTG

If the number of repeat sequences is more than 200, DNAMAN will ask you to increase the minimum length of repeat unit, or define a smaller analysis region using the ***Sequence | Sequence Channel | Analysis Range Definition*** command (see the section **VI. 2**) for searching.

VIII.5 Searching for amino acid sequences

DNAMAN searches for an amino acid sequence (in one letter code) and its variations from the **six reading frames** of the current sequence.

Mismatch can be any position in the amino acid sequence.

For example**:**

1. Type VXPSC and set the mismatch parameter at 0 to search for VXPSC, in which "X" corresponds to an undefined amino acid at the position 2.

2. Type VSPSC and set the mismatch parameter at 1, in this case, the mismatch amino acid can be at any position.

Protein consensus sequences

Click the **Consensus** button in the **Search** dialogue box to open the consensus sequence box. Choose one consensus sequence and it will be loaded to the Search dialogue box. You may edit the consensus list. The editing methods for an amino acid sequence are the same as for DNA consensus sequence. For the file format, consult the SEARCHNT.DAT file in the CONSENS folder of the DNAMAN program.

VIII.6 Searching for open reading frames

You may search for open reading frames (ORFs) from six reading frames of the current sequence. By default, DNAMAN uses the universal genetic code table for the searching. In some cases, you may want to use **CTG** instead of **ATG** or both as the start codons. DNAMAN allows you to make these changes by choosing the ***Protein | Genetic Code Table*** command and then selecting an appropriate genetic code table (see the section **XII.1**). DMANAN will search the ORFs according to the start and stop codons in the selected genetic code table.

DNAMAN shows only the two largest ORFs from each reading frame.

Possible Open Reading Frame in example (1-886)

Strand	RF	AA Num	Position	Sequence
Plus	3	121	219-584	ATG...TAA
Plus	2	119	173-532	ATG...TGA
Plus	1	73	1-222	......TGA
Plus	2	40	2-124	......TAA
Plus	3	27	732-815	ATG...TAA
Plus	1	16	721-771	ATG...TGA
Minus	1	109	379-708	ATG...TGA
Minus	2	35	440-547	ATG...TAG
Minus	2	30	653-745	ATG...TAA
Minus	3	15	3-50	......TGA
Minus	3	14	843-887	ATG......

```
Minus    1   8      265-291     ATG...TAA
```

If you want to overview all the ORFs of the current sequence, you can use the ***Protein | Overview*** command to display the six reading frames in a graphic presentation (see the section **XII.2**).

VIII.7 siRNA Sequence Search

Small interfering RNA (siRNA) or short interfering RNA or silencing RNA, is a class of double-stranded RNA molecules, less than 25 bps in length. its most notable role is in the RNA interference (RNAi) pathway to affect the expression of specific genes with complementary nucleotide sequence.

You may search/design siRNA from a DNA template by choosing the **Search | siRNA** menu.

Default parameters for selection of siRNA:

- Target sequence length: 19
- Start with A: No
- End with TT: No
- R at position 3: No
- Y at position 21: No
- No polyN(>3) seq: Yes
- PolIII expression vectors: No
- CDS region defined: No
- delatG of pentamer at 5'AS must >: -9.8 kcal/mol
- delatG difference of pentamer 5'S - 5'AS must <: -0.1 kcal/mol

BLAST search of the target sequences must be performed to avoid undesired similar sequences.

Chapter IX Sequence Comparison

With DNAMAN, you can compare two or more DNA or protein sequences. ***Dot matrix comparison*** provides a graphic view of the similarity of two sequences. Sequence alignment allows you to find homology regions in two or more sequences.

IX.1 Dot matrix comparison

The ***Dot Matrix Comparison*** function compares two DNA sequences or two protein sequences in a **dot matrix plot.** The abscissa represents one sequence and the ordinate represents the other sequence. The algorithms used in Dot Matrix comparison are described by *Y. Huang* and *L. Zhang* in ***Bioinformatics*** Volume **20**: 460-466 (2004).

IX.1.1 Parameters of dot matrix comparison

Choosing the ***Sequence | Dot Matrix Comparison*** command opens a Dot Matrix comparison window. Press the **Options** button to open a dialogue box.

You may change the following parameters in the **Dot Matrix** dialogue box:

1. **Sequence type**. Choose the **DNA** option for DNA sequence comparison or **Protein** for protein sequence. Available sequence channels are listed in the combo box for your selection. To compare DNA sequence to protein sequence, choose the **DNA/Protein** option.
2. **Sequence 1 section**. You may select a channel sequence by pressing the **Channel** button. Channel sequences are available only if they match the sequence type. You may also choose a sequence from disk files by clicking the **File** button. In this case, an **Open File** dialogue box shows up for selection. Sequence from default database can also be selected by clicking the **Database** button (if database available). You may use a part of sequence for comparison by defining start and end positions of the sequence.
3. **Sequence 2 section**. Use the same selection method of **Sequence 1** to define **Sequence 2**. If the **DNA/Protein** type

has been elected in step 1, you may enter DNA sequence for **Sequence 1** and protein sequence for **Sequence 2**.

4. **Comparison**. There are four methods of comparison: **Fast, AA Coding, TFT** and **Sensitive**. **Window size**, **Mismatch** and **Both Strands** are also involved in comparison. See the **Comparison method** section for more information.
5. **Show sequences**. DNAMAN will display the sequences if you select a region in the dot matrix plot. If the homology between the two sequences is high enough, DNAMAN may show the alignment. In order to show the alignment, you should check the **Align when identity>** option and enter the minimum identity level for alignment. Select a method for alignment between **Optimal** and **Fast**. The method should be chosen according to the **Max Length** value for alignment. Long sequence may require **Fast** method while short sequence may work with **Optimal** alignment.
6. **Annotations**. DNAMAN may show annotations on the plot. This option facilitates the identification of target sequence regions. All available annotations are listed in a list box. You may select all annotations for display. You may also ignore some of the annotations by un-selecting them in the list box. Un-check the **Annotations** button to ignore all of them.
7. **Plot box**. The dot matrix plot consists of a rectangle **frame** and the **dots** in the plot. The frame box is defined by the position(X and Y) and size(Width and Height). You may change the thickness of the lines used for the **frame** and **dots**. **Tick marks** can be defined for each axis. **Gridlines** option facilitates the location of dots on the plot. **Zoom box** option shows the region being zoomed. **Plotting during analysis** option allows visualization of analysis progress, however, may slow down the overall comparison. **Plot from TopLeft** option will show start points (0,0) at top left corner. Without this option checked, DNAMAN plots from bottom left corner.

Click the **OK** button to validate the parameters and performing comparison.

IX.1.2 Comparison methods

For dot matrix comparison methods, DNAMAN uses the algorithms described in ***Rapid and Sensitive Dot-matrix Methods for Genome Analysis***, *Bioinformatics* (2004) **20**:460-6. The methods initially locate similarity regions between two sequences using a fast word search algorithm, followed with an explicit comparison on these regions. Since the initial screening removes most of random matches, the computing time is substantially reduced. The methods produce high quality dot-matrix plots with low background noise. Space requirements are linear, so the algorithms can be used for comparison of genome size sequences. Computing speed may be affected by highly repetitive sequence structures of eukaryote genomes.

There are four methods of comparison:

1. **Fast**. Use this method to compare long DNA sequences (e.g. >1 megabases).
2. **AA Coding**. This method compares DNA sequences using their corresponding amino acid codes. The algorithm is similar to that of Fast method. This method is more sensitive than the Fast method, however, is slightly slower.
3. **TFT** (Two for Three). This method is similar to AA Coding. Instead of using amino acid code, it searches and compares the first two bases out triplets. This method is slower but slightly more sensitive than the AA Coding method.
4. **Sensitive**. This method compares the two sequences progressively from the beginning to end. This is the most sensitive but slowest method. It is good for comparison of protein sequences or short DNA sequences.

The sensitivity of comparison is also determined by the parameters of **Window size** and **Mismatch**. DNAMAN suggests the Window size according to the sequence length and Mismatch values. If the dot plot appears too noisy or insensitive, you may modify these values.

DNAMAN draws a dot where the two sequences are matched with a length equal or more than the **Window size** and the mismatch size less than the defined number. The allowed

mismatch size increases proportionally to the length of match size. The dot may be shown as a line if homologue region is much larger than the **Window size**. A long line in the matrix indicates a long homologous region.

IX.1.3 Zooming dot matrix

To zoom in a specific region in plot, use mouse left button to select the area. There are 16 levels of zoom to enlarge the region of interest. To zoom out, double-click mouse left button. To zoom in again the pre-defined region, double-click mouse right button (click while holding the ***control*** key on MacOSX).

DNAMAN may display the sequences of a selected region on screen using mouse **right** button(holding the ***control*** key on MacOSX). The sequences can be automatically aligned if the homology between the two sequences is high enough. This homology value is defined in the **Show Sequences** section. The **Align when identity >** option must be checked to enable the function. To show sequence alignment of sequence 1 with the minus strand of sequence2, the **Seq2+Strand** option must be un-checked.

IX.1.4 More about dot matrix

Dot plot may show the homology of both DNA strands if the option is checked in the dialogue box. In the plot window, the two check box, **Seq2+Strand** and **Seq2-Strand,** indicates the which strand (or both) is shown. Black dots show the matches between sequence 1 and the + strand of sequence 2, and red dots show matches between the sequence 1 and the - strand of sequence 2.

You can change the size and shape of the plot frame. Place the cursor on the **low-right** corner of the plot. When the cursor switches to ⇔, press the left mouse button and drag it to change the size and shape of the plot.

You can also move the plot within the graphic window. Place the cursor on the **up-left** corner of the plot. When the cursor switches to  , press the left mouse button and drag the plot.

When the cursor is a cross on the dot matrix plot, the coordinates of Sequence 1 and Sequence 2 are displayed on the left corner of the screen.

IX.2 Two sequence alignment

When you load two or more DNA or protein sequences into sequence channels, the ***Sequence | Alignment | Two Sequence Alignment*** menu is activated. You may accurately compare two sequences with this function. Another advantage of using this function is that you may quickly align any region of the two DNA or protein sequences by defining analysis regions (see the section **VI. 2**).

A dialogue box appears by selecting the ***Sequence | Alignment | Two Sequence Alignment*** command. You should specify the following parameters.

1. **Sequences** and **Type**. Choose the **DNA** option for DNA sequence. All channels containing DNA sequences are available in the two list boxes. Choose any of them to compare against another. You have the option of **Best alignment of dsDNA**. With this option checked, DNAMAN compares the first sequence to both strands of the second sequence, and chooses high score strand for final alignment. If you choose the **Protein** option, all channels containing protein sequences will be available in the two list boxes. With this option, you may select a scoring matrix for optimal alignment. You have also an option to indicate similar amino acids in the alignment results.
2. **Alignment Method**. You may choose a Quick alignment or one of dynamic alignment methods. The algorithm of the Quick alignment method is derived from the publication of Wibur and Lipman, 1983, ***Proc. Natl. Acad. Sci. USA*** 80:726. The Smith and Waterman (local) alignment method is derived from the publications of Smith and Waterman,

1981 ***Advances in Applied Mathematics*** 2: 482**,**. Feng and Doolittle, 1987, ***J. Mol. Evol.*** 25:351 and Thompson, et al., 1994, ***Nucleic Acids Res.*** 22:4673. Myers&Miller (global) and Needleman&Wunsch (global) use also dynamic alignment methods.

Alignment parameters for all mthods:

1) **Gap open penalty** is a negative score for opening each gap.

2) **Gap extension penalty** is a negative score for extending each residue in an existing gap.

Quick alignment parameters:

1) ***K-tuple*** indicates the minimum number of identical bases in an exactly matching fragment. **The K-tuple value is from 2 to 7 for DNA and from 1 to 3 for protein**. The default level is 4 for DNA and 2 for protein. Increasing the K-tuple may decrease the sensitivity for alignment and slightly speed up the alignment.

2) ***Window*** defines the size of window alone side with a significant diagonal. Default value is 4.

2) ***Diagonals*** defines the number of significant diagonals used for alignment. Default value is 4.

3. **Show Identity with**. You may choose one of the three options to show identical base/amino acid in the final alignment results: ***** , **|** , or **:**.

Click the **OK** button to perform alignment and display the alignment results in a text window. DNAMAN includes the percentage of homology in the result window.

The identity (%) of the two sequences is calculated using the following formula:

$$\textbf{Identity (\%)} = \frac{\text{Number of identical bases or AAs between two sequences}}{\text{Length of aligned sequence - Length of all gaps}} \times 100$$

DNAMAN shows also the gaps introduced in the alignment.

$$\textbf{Gap (\%)} = \frac{\text{Number of gaps of both sequences}}{\text{Length of aligned sequence}} \times 100$$

Chapter X Multiple Alignment and Phylogenetic Analysis

With DNAMAN, you may analyze the homology relationship among the family members using multiple sequence alignment and phylogenetic tree analysis.

X.1 Multiple alignment

DNAMAN provides fast and optimal alignment methods for aligning a large number of DNA and protein sequences. The multiple alignment function is compatible with many other sequence analysis programs. A multiple sequence editor (MASED) is used to handle alignment results. The editor can produce homology and phylogenetic trees in graphic windows, and export alignment results in different sequence formats.

X.1.1 Performing multiple alignment

1. Choose the ***Sequence | Multiple Sequence Alignment*** command to open the **Multiple Alignment** dialogue box. There are four steps to setup multiple alignment sequences and parameters.
   2. In the first step, choose sequences for alignment. There are four methods for sequence input:
      a) From file: select sequence files from a dialog box. You may select more than one files from the box.
      b) From folder: select all sequence files from a folder. If there are a large number of sequences for alignment, you may move all of them in one folder and choose it for alignment.
      c) From channel: select one all more channels for alignment.
      d) From database: select one or more sequences from default database. The database must be set to default using Database Manager prior to selection.

      You may input sequences using one or more methods. Remove a sequence from the list by selecting it and pressing the **Remove** button, or the **Clear** button remove all sequences in the list. DNAMAN checks files when

they are added or removed from the list, and indicates the number of files and sequences.

Choose the type of sequence: ***DNA*** or ***Protein***. During alignment DNAMAN calculates the composition of each sequence to confirm your selection. If a sequence does not appear to be the selected type, you will see a warning message. You may align DNA sequences according to translated amino acid sequences. In this case, select the **Protein** type even though the input sequences are DNA, and choose one of the three reading frames for translation in the second page.

3. Press **Next** button to go to the second step, select one of the four alignment methods: 1) ***Full alignment***; 2) ***Profile alignment***; 3***) New sequences on profile alignment*** and 4) ***Fast alignment***. See next sections for the description of these methods.
4. Choose the output sequence order as **Aligned** or **Input**.
5. If you have chosen the **Protein** alignment type and selected DNA sequences for alignment, you may align DNA sequences according to translated amino acid sequences. In this case, choose one of the three reading frames for translation.
6. For DNA sequence alignment, you may check the **Try both strands** option. DNAMAN will compare both strands and use the one with higher homology to the first sequence for alignment.
7. Check the **Run in background** option. This option allows you to work on other functions while performing multiple alignment. Multiple alignment involves in heavy calculation and may take long times.
8. Check the **Show progress** option. With this option, DNAMAN will show the progress of alignment in the ***Info/Channel*** panel at the bottom of the Studio.
9. Press the **Next** button to go to the **Pairwise Alignment** page.
10. Choose **Quick Alignment** method.
11. Press the **Default Parameters** button to load the default parameters for alignment. You may change any of these parameters according to your needs.

12. If you have chosen the **Fast Alignment** method in the first page, press the **Finish** button to start alignment. Otherwise, press the **Next** button to go to the **Multiple Alignment** page.
13. Press the **Default Parameters** button to load the default parameters for alignment. You may change any of these parameters according to your needs.
14. Press the **Finish** button to start alignment.

X.1.2 Input formats for multiple alignment

DNAMAN is able to recognize eight common sequence formats that are differentiated by the keywords or initials in the beginning of each file. Sequence files in these formats can be used for multiple alignment without conversion. If your sequence files do not contain the corresponding keywords or initials, you may add these text to the files and then save them.

DNAMAN accepts following sequence formats:

DNAMAN format (keyword: ORIGIN)
- Single or multiple sequences

GenBank format (keyword: LOCUS and ORIGIN)
- Single or multiple sequences

EMBL/Swiss Prot formats (keyword: ID)

GCG/MSF format (Initial: PileUp)

CLUSTAL format (Initial: CLUSTAL)

FASTA format (Initial: >)

NBRF/PIR format (Initial: >)

GDE format (Initial: # for DNA, % for protein)

X.1.3 Fast alignment method

If the **Fast Alignment** option is chosen in the second page of **Multiple Alignment** dialogue box, DNAMAN performs the multiple alignment with a fast alignment method. With this method, DNAMAN aligns each pair of sequences, constructs a homology tree from the results of pairwise alignment and finally build up alignment based on the homology tree with the previous established alignment. Using these fast alignment methods, you

can quickly align a large number of DNA or protein sequences. If these sequences have low degrees of divergence, the fast alignment delivers relatively accurate results.

You may choose the **Quick Alignment** or **Dynamic Alignment** method for the pairwise alignment in the second page of the dialogue box. The **Quick Alignment** performs pairwise alignment with all sequences using the method developed by Wibur and Lipman, 1983, ***Proc. Natl. Acad. Sci. USA*** 80:726**.** The **Optimal Alignment** performs pairwise alignment with all sequences using a dynamic method). The **Dynamic Alignment** method aligns sequences more accurately, but may be much slower than **Quick Alignment** with long sequences (>1000).

There are four parameters in the **Quick Alignment** method:

1. **Gap penalty** is a negative score for each gap insertion. This score is the fixed penalty that is not related to the size of a gap.

2. **K-tuple** defines the minimum number of identical residues as an exactly matching fragment. Increasing the K-tuple value decreases the sensitivity for alignment but speeds up the alignment.

   - For DNA alignment, the K-tuple size is 1 to 6.

   - For protein alignment, the K-tuple size is 1 to 3.

3. **No. of Top diagonals** are used to define the number of diagonals with the most K-tuple matches. Increasing the **Top Diagonals** value may increase a little sensitivity for alignment but slightly reduces alignment speed.

4. **Window size** is the number of diagonals around each of the top diagonals. Increasing the Window size may increase the sensitivity for alignment but slightly reduces the alignment speed.

There are three parameters in the **Dynamic Alignment** method:

1. **Gap open penalty** is a negative score for opening each gap.

2. **Gap extension penalty** is a negative score for extending each residue in an existing gap.

   - For DNA alignment, the default penalty is 5.

   - For protein alignment, the default penalty is 0.1.

3. **Weight matrix**. You may assign a transition (A->G, C->T) weight for DNA alignment. For protein alignment, you must choose one of the weight matrices for similarity calculation.

X.1.4 Optimal Alignment

If you choose the **Full Alignment**, **Profile Alignment**, or **New Sequences on Profile** option in the first page of Multiple Alignment dialogue box, DNAMAN performs the multiple alignment with an optimal alignment method. DNAMAN performs firstly the pairwise alignment with all necessary sequences, constructs a homology tree from the results of pairwise alignment and finally build up alignment based on the homology tree with an optimal group alignment.

There are three types of optimal alignment:

1. **Full Alignment**

   If the sequence files consist of one or more multiple alignment profiles, DNAMAN will disregard the original alignments existed in the profiles and realign them completely.

2. **Profile Alignment**

   Two multiple alignment profiles should be inputted for alignment. DNAMAN aligns the two profiles without disturbing the original alignment existed in each profile.

3. **New Sequences on Profile**

   One multiple alignment profile and one or more sequences can be inputted. DNAMAN aligns the profile without

changing its original alignment with other sequences. The profile file has to be the first one in the sequence list box.

* Multiple alignment profile is a set of aligned sequences. DNAMAN allows you to export multiple alignment sequences into a text window and save the data as a multiple alignment profile.

In all the three methods, DNAMAN constructs a homology tree using the same approach as the **Fast Alignment**. You may choose the **Quick Alignment** or **Dynamic Alignment** method for similarity calculation. After the tree construction, dynamic programming is finally used to optimize group alignment (Feng and Doolittle, 1987, ***J. Mol. Evol.*** 25:351; Thompson, et al., 1994, ***Nucleic Acids Res***. 22:4673). This method generates better alignments but could be very slow with long sequences.

There are several parameters in the final multiple alignment:

1. **Gap open penalty** is a negative score for opening each gap.

2. **Gap extension penalty** is a negative score for extending each residue in an existing gap.

   - For DNA alignment, the default penalty is 5.

   - For protein alignment, the default penalty is 0.1.

3. **Delay divergent** sets up an identity level (%) that allows DNAMAN to align the more divergent sequences later.

4. **Protein weight matrix** determines the similarity scores of non-identical amino acids.

- BLOSUM (blocks substitution matrix). From Henikoff and Henikoff, 1992, ***Proc. Natl. Acad. Sci. USA*** 89:10915. This is the default protein weight matrix.

- PAM (percent accepted mutation matrix). From Dayhoff et al., 1978, in "***Atlas of Protein Sequence and Structure***", Vol 5, Suppl.3, pp 345, National Biomedical Research Foundation, Silver Spring, Maryland, USA.

- Gonnet. These matrices were derived from almost the same procedure as PAM but are much more up to date and are based on a larger data set. They appear to be more sensitive than PAM.

- Identity Matrix. This matrix assigns equal weight to each amino acid.

5. **Protein gap parameters** increase or reduce the gap penalty scores depending on the location of a gap.

   - Residue specific penalties **increases** the cost of opening a gap that is adjacent to the amino acids of A, C, E, F, I, L, M, Q, V and W, but **reduces** the cost of D, G, K, N, P, R, S and T (Pascarella and Argos, 1992, ***J. Mol. Biol.*** 224:461).

   - Hydrophilic penalties **reduces** gap penalties in any run of 5 hydrophilic residues (G, P, S, N, D, Q, E and R).

   - Gap separation **increases** the gap penalty where the gaps are too close to each other.

   - End gap separation **penalizes** the end gaps like the internal gaps. Opening end gaps are usually without any penalty.

X.2 Multiple Alignment Sequence Editor of DNAMAN (MASED)

The Multiple Alignment Sequence Editor of DNAMAN (MASED) is an efficient tool for multiple sequence editing. You may use MASED to visualize multiple alignments and analyze sequence homology. The results of the Multiple Alignment function are loaded directly into MASED. You may also use MASED to analyze the multiple alignments produced by other programs such as GCG. MASED recognizes all multiple sequence formats listed in the pages 7-12. When you use the ***File | Open special | Multiple alignment*** command to open these files, DNAMAN loads them into MASED.

X.2.1 Properties of MASED

Press the **Options** button in MASED to open property pages. There are three pages to define the properties of the current alignment: ***Preferences, Shading Homology*** and ***Elements***.

1) ***Preferences*** page

In the ***Preferences*** page you may change the following properties of MASED:

Sequences to display

By default, all sequences in MASED are shown. However, you can select some of them to display. Any sequence deselected in the display box will be hidden in the MASED window. You may press the **All** button to select all sequences in the list.

Sequence type

You can specify the sequence as **DNA**, **Protein**, or as detected by DNAMAN. The **Auto detect** option will assume the sequences as DNA or protein according to their compositions. This method makes correct assumptions in 98% cases. If your sequences have extreme compositions, you should specify them as DNA or protein sequences.

If you specify the sequences as DNA, you may display translated amino acid sequences in MASED. By choosing one the three reading frames, you may display the amino acid sequences in the editor.

Letters per line

When you print out, copy or export alignment as DNAMAN text file, the text length of each line has to be defined. You should choose a line length according to the paper size and text font. The default length is 40.

Text font

You may select any font for MASED. However, it is highly recommended to use fixed size fonts. Only fixed size fonts make the sequences correctly aligned.

Fill end gaps

Check this option if you like to have all sequence shown with the same length. DNAMAN fills shorter sequences with dots. If this option is not checked, DNAMAN will stop at the last letter of a sequence and leave empty space thereafter.

Show consensus sequence

MASED will display consensus sequence of the alignment if you check the **Show Consensus Sequence** option. Consensus residues can be defined as **Identical** bases or amino acids at each position, or as **Homology** residues at each position. If you choose **Homology** residues, the cutoff level for homology residues must be defined in the **Shading Homology** page.

Show graphic overview

If you have selected this option, a graphic presentation is shown on the top of the sequence position labels. The graphic pattern represents the homology level along the sequence profile. Strong signal means high homology level. A red rectangle box on the graphics indicates the sequence region currently displayed on the MASED window. By clicking on the graphic overview, you can move the displayed window to any region of the sequence profile. The overall percentage of the sequence identity is shown on top of the graphic presentation. This percentage is calculated according to the homology level of each position of all the sequence.

Synchronize gaps

This option defines how to edit a sequence. If this option is checked when you want to insert or remove gaps in the middle of a sequence, DNAMAN will make changes at downstream to make next sequence block unchanged. Therefore, insertion or removal of gaps will have little compact for downstream sequence. If this option is not checked, insertion or removal of gaps will modify entire downstream sequence.

2) ***Shading Homology*** page

In the **Shading Homology** page you may modify the following parameters:

Highlight homology level

DNAMAN may shade a position to show the homology. There are five options available**: Non, 100%, >=75%, >=50%** and **>=33%**. If you choose the **>=50%** option, all positions with 50% or higher homology will be highlighted. For example, if there are 10 sequences in the multiple alignment and 5 or more sequences show the letter "A" in position 2, position 2 will be highlighted.

Shading type

You may highlight homology regions with **Colors**, or **Blocks**, or both. If the **Colors** option is checked, you may define any color for different levels of homology. If the Blocks option is checked, you may define the thickness of lines to highlight the blocks.

Shading similar amino acids

This option applies only to protein sequences. If this option is checked, DNAMAN will shade similar amino acids. The similar amino acid is defined in the "similaa.dat" file in the DNAMAN system folder

3) ***Elements*** page

You may define elements to display regions on the multiple alignment.:

Add a new element

Press the ***New*** button to display the ***Element Definition*** dialogue box. There are two types of elements: directional and non-directional. An **Arrow** element is directional. You may direct the arrow from left to right by entering the numbers in ascending in the Start (bp) and the End (bp) columns (e.g. From 200 to 100). You may reverse the arrow from right to left by entering the number in descending (e.g. From 100 to 100). **Bar** element is non-directional and you can enter the numbers in either way. **Eight colors (patterns)** are available in the Color column of the dialogue box. You

may define the Element name in the corresponding box. All Elements are shown in the list box.

Modify an Element

Double-clicking the element name in the list box displays the **Element Definition** dialogue box. You may modify any the name, type and size of the element.

Remove an Element

Click the element and press the **Remove** button.

X.2.2 Functions of MASED

There are many functions of MASED for sequence analysis and editing:

- Shading homology regions
- Changing list order
- Moving a sequence fragment between gaps
- Truncating aligned sequence fragments
- Adding and deleting gap insertions
- Exporting multiple sequences in different formats
- Copying
- Producing trees
- Performing restriction analysis of DNA sequences
- Predicting secondary structures of protein sequences
- Comparison hydrophobic/hydrophilic properties of protein sequences

Shading homology regions

In a multiple alignment profile, homology regions are easier to visualize when they are highlighted. In MASED, these regions can be highlighted according to their homology level. Depending on your preference, either line blocks or colors can be used to shade the homology regions.

Press the **Options** button in MASED to open the property pages. See the previous section for the information of **Shading homology**.

If you choose the ***File | Print Preview*** command, DNAMAN displays the aligned multiple sequences with the shading colors in a printing view window.

Changing list order of multiple alignment sequences

All sequence names are listed in the left panel of MASED. To change the output sequence order:

1. Place the cursor on the name of the sequence you want to move.
2. Drag and drop the name to the desired place.
3. Repeat step 1 and 2 to change the positions of other sequences.

Moving a sequence block between gaps

You may want to move a sequence fragment that is misplaced between gaps.

1. Select the sequence you want to move. For example, selecting CTGAA:

```
CAC..CTGAA............TTACTGCCAAGGATATCCTCGAC
```

2. Upon selecting, the cursor switches to an arrow. Move the arrow to a location where you want to insert the fragment. Now, click the left mouse button to place the fragment there.

```
CAC..............CTGAATTACTGCCAAGGATATCCTCGAC
```

You may move the fragment only between the gaps that do not affect sequence position.

Truncating aligned sequence fragments

If you are interested in only a small region of the multiple alignment, you may delete flanking regions of the alignment. In other words, you may delete both ends of the alignment.

1. Place the cursor in the beginning or at the end of the first sequence of the alignment.
2. Use the **Shift+Arrow** keys or the mouse to select the fragments.

You can truncate a part of one sequence by selecting the sequence from its beginning or the end and using the **Delete** or the **Backspace** Key to delete the selected sequence.

You cannot delete any nucleotide or amino acid located in the middle of the alignment.

Adding and deleting gap insertions

A dot represents a gap insertion in the multiple alignment. DNAMAN allows adding or deleting one or more dots in the alignment. When you add or delete gaps, the downstream alignment may or may not be altered by this change. See the **Synchronize gaps** option in the **Properties of MASED**.

```
     1         10         20         30         40         50
Seq1 TTTGACTGCCACTTCCTCGAAGAAGGTTTACTGCCAAGGACATTCTGGAC //
Seq2 CAC........CTGAA...........TTACTGCCAAGGATATCCTCGAC //
```

Adding 5 dots between CT and GAA of Seq2:

```
     1         10         20         30         40         50
Seq1 TTTGACTGCCACTTCCTCGAAGAAGGTTTACTGCCAAGGACATTCTGGAC //
Seq2 CAC........CT.....GAA......TTACTGCCAAGGATATCCTCGAC //
```

Exporting multiple alignment to text window

When you click the **Output** button in MASED, a popup menu appears. Choose the ***Sequence File*** menu to export the alignment into a text window using one of the seven formats. The sequence in the text file is not marked with shading colors to indicate the identical residues.

If there are only two sequences in MASED, choose the **Two Sequence Alignment** command. If the sequences are DNA you can display the identical bases in *, | or : . If the sequences are protein, DNAMAN will indicate identical residues as well as similar amino acids.

Example:

```
Alignment of 1(upper line) and 2(lower line)
Identity=14%  Similar residues=56%

-----------------

1     RFSEVSSVFTTRLHTLNGLFSLCALTRFIRENRISHLMIHTGKIAALSILLKKLTGVRLI
      :: :.|.|:.  ::.:. | .|.   |::  |  ..: .:.:.: .:   :| . .:|
```

```
1     HYIGFSKVYKRLFQKWTRLDPLPYSQRIL..NIRDKVTTQEDSVIVIHNSMKLYRQIRER

61    FVKHNVVANKTDFYHRLIQKNTDRFICVSRLVYDVQTADNPFKEKYRIIHNGIDTDRFP.
       .. .:| :. : :.. :..|..::|..|.:: .. ..  |  .   |:.||: .: :.
59    NPNAKLVMHMHNAFEPELPDNDAKIIVPSQFLKAFYEERLPAAAVS.IVPNGFCAETYKR
```

Copying in MASED

In many cases such as preparing posters or slides, you may want to generate a document in which some residues in multiple sequence alignment are identified. DNAMAN allows you to copy a part of multiple sequence alignment from MASED to text window or to other Windows applications.

1. Press the Options button to open MASED options dialogue box. In the Preferences page, define line length in the Letters per line box. In the **Shading homology** page choose a shading method. Close the dialogue box by clicking the **OK** button.
2. Place the cursor at the position you want to start the copy and select a sequence region.
3. Choose the ***Edit | Copy*** to activate the **Copy As** menu. Click the **Graphics** menu to copy the selection as graphics. If you click the **Text** menu only the text will be copied.
4. Paste the alignment to a text window or other Windows applications.

Phylogenetic analysis

See the section **X.3.**

Performing restriction analysis of DNA sequences

This function can be used to compare the restriction properties of aligned DNA sequences. You can choose any enzymes for restriction analysis. The results are shown in a text window. See the section **XI.3.1** about ***Restriction analysis with multiple DNA sequences***

Predicting secondary structures of protein sequences

DNAMAN uses the DSC method to predict secondary structures of the aligned protein sequences. See the section **XII.9** for the description of the function and results.

Comparison of hydrophobic/hydrophilic properties of protein sequences

You can plot the hydrophobic/hydrophilic profiles of the aligned protein sequences. All these sequences are shown in the same plot and indicated in different colors. See the section **XII.8** about hydrophobic/hydrophilic plots.

X.2.3 Saving multiple alignment profile in MASED

You may save the content of MASED, or export it to a text window. If you choose the ***File | Save*** menu, MASED will save the alignment in DNAMAN2 format. The saved multiple alignment profile can be retrieved and edited later.

X.2.4 Output formats of multiple alignment in MASED

Multiple alignment in MASED can be exported to a text window in any of the seven formats:

DNAMAN 1 format(default format)
DNAMAN 2 format
GCG/MSF format
CLUSTAL format
FASTA format
NBRF/PIR format
GDE format

There are some special options of exporting two-sequence alignment (see ***Exporting alignment to text window*** in the section **X.2.2**).

X.3 Phylogenetic Analysis

MASED can produce trees from a multiple alignment. DNAMAN calculates the homology matrix and establishes related distances between all pairs of sequences. When you press the **Output** button, a popup menu appears. Under the menu

item ***Tree***, you may choose a distance matrix of the alignment, a phylogenetic tree or homology tree.

Distance matrix

This matrix shows related distances between all pairs of sequences in the alignment. The low value presents a low divergence (high homology) between two sequences.

There are five methods of calculating distances.

1) *Observed Divergence.* This method uses directly unmatched residues divided by compared length between two sequences. There is no correction applied to distances.
2) *Kimura.* This method uses *Observed Divergence*, then applies Kimura 2-parameter correction (Kimura, 1980; ***J. Mol. Evol.*** 16:111-120).
3) *Jukes & Cantor (DNA only).* This method uses *Observed Divergence*, then applies correction of Jukes and Cantor (Jukes and Cantor, 1969; ***Evolution of Protein Molecules***, Academic Press, Kimura and Ohta 1972; ***J. Mol. Evol***. 2:87-90).
4) *Poisson Correction (Protein Only).* This method uses *Observed Divergence*, then applies correction of Poisson distribution of amino acid substitutions. This distance is for estimating the number of amino acid substitutions per site under the assumption that the number of amino acid substitutions at each site follows the Poisson distribution.
5) *Maximum Likelihood.* This option uses a maximum likelihood method that searches for an estimate of the phylogenetic distance between two sequences with mutation rates estimated from the actual sequences (Hasegawa et al, 1985, ***J. Mol. Evol.***, 22. 160, Tamura & Nei, 1993, ***Mol. Biol. Evol.*** 10:512). For **DNA** model, there are three parameters: Alpha/Beta ratio (default=4), AlphaY/R ratio (default=1) and Beta12 ratio (default=1). You may choose the Poisson distribution model, however, a proportional model applies if the Data Frequency option is checked. There are four models for **Protein** sequence analysis: Jones, Taylor & Thornton (Jones et al, 1992, ***CABIOS*** 8:275), Dayhoff (Dayhoff et

al 1978,in ***Atlas of Protein Sequences and Structure*** , 5: 345), Mitochondrial (Adachi, J. and Hasegawa, M., 1996, ***Mol. Biol. Evol***. 13: 200) and Poisson. A proportional model is used in stead of Poisson if the Data Frequency option is checked.

Homology tree

This tree is setup with the distance matrix using the UPGMA method (Sneath and Sokal, 1973, ***Numerical Taxomomy***, San Francisco, USA). The matrix can be built up only with *Observed Divergence*.

Phylogenetic tree

This tree is setup with the distance matrix using the Neighbor-Joining method (Saitou and Nei, 1987, ***Mol. Biol. Ecol***. **4**:406-425). You may bootstrap a phylogenetic tree to show statistically to be typical of the variation (Felsenstein, 1985, ***Evolution*** 39: 783). Bootstrapping involves creating a new data set by sampling randomly with replacement, so that the resulting data set has the same size as the original, but some characters have been left out and others are duplicated. The method assumes that the characters evolve independently.

Bootstrapping is a threaded process in DNAMAN. Therefore, you may perform other analysis functions during bootstrapping.

X.3.1 Plotting of homology tree

Homology tree is drawn in a graphic window. It shows related homologies between two sequences or groups.

The sequence names are listed at the left of the graph. Sequences with high homology are grouped together. A rule of homology level is placed on top of the graph. You may adjust the position and size of the trees in the graph.

- **Move tree**

  Place the cursor on the rule. By pressing and holding the left button of the mouse, you may move the tree on the screen.

- **Change shape and size**

  1. Place the cursor on the most right vertical line from the root of the tree, or on the last horizontal branch (bottom) of the tree.
  2. When the cursor switches to an arrow, press and hold the left button of the mouse, you may change the shape and size of the tree.

There is an option to display or hide homology number on the branches. By clicking the **%On/Off** button, you may switch the option.

X.3.2 Plotting of phylogenetic tree

Phylogenetic tree shows related homologies between any two sequences in a multiple alignment.

A phylogenetic tree is opened in a window with two panels: an Option panel with parameter options and text box containing the description of the tree, and a Graphic panel showing the graphic view of the tree.

Tree Description

The tree is described using variations of "Newick 8:45" format adopted by the NEXUS format and by PHYLIP. The description starts at root and follows the visit of all nodes. If the node is a leaf, the leaf is recorded and then return to its immediate ancestor. If the branch leading to the leaf has a length, the length is recorded immediately after the leaf with a colon, e.g. "Mouse:0.025". There are three indicators if the node is not a leaf. 1) A left parenthesis "(" indicates the first visit and the next node is the leftmost. 2) A comma "," separates the descendants of the same ancestor. 3) A right parenthesis ")" shows all descendants of the same ancestors have been visited. A node label (e.g. a bootstrap value) may be indicated immediately after the ")" (e.g. ")99"). If the branch leading to the node has a length, the length is recorded immediately after the leaf with a colon, e.g. "):0.025". A semi colon ";" indicates the termination of description.

Tree Graph

Phylogenetic tree is drawn in a graphic window. You may choose options in Option/Description Panel. There are three types of three: **Standard, Circular and Radial.** You may also choose display in four options to show**: Rooted tree, Branch Length, Seq Weight** and **Bootstrapping**.

1. **Standard Tree**. This is default tree drawn from the tree description. The sequence names are listed at the right of the graph. Sequences with high homology are grouped together. A scale of distance is placed on top of the graph. You may adjust the position and size of the trees in the graph.

   - **Move tree**. Place the cursor on the scale. By pressing and holding the left button of the mouse, you may move the tree on the screen.

   - **Change shape and size**.

     Place the cursor on the most left vertical line from the root of the tree, or on the last horizontal branch (bottom) of the tree. When the cursor switches to an arrow, press and hold the left button of the mouse to change the shape and size of the tree.

2. **Circular Tree**. Tree is drawn in circular form.
3. **Radial Tree**. DNAMAN draws the tree in a radial diagram according to the tree description. You may move the tree as described above, but may not change the shape or size using the method described above.
4. **Rooted Tree**. If this option is chosen, DNAMAN redraws the tree with a root. The root option reduces the overall weight differences among all sequences. You may move the tree or change the shape and size as described above.
5. **Branch Length**. DNAMAN draws the tree and labels the lengths on all branches if this option is checked. This option applies to Unrooted tree and Rooted tree, but not Radial tree.
6. **Sequence Weight**. DNAMAN draws the tree and labels the sequence weights of all leaves if this option is checked.
7. **Bootstrapping**. DNAMAN draws the tree and labels the bootstrapping values if this option is checked. This option

applies to Unrooted tree and Rooted tree, but not Radial tree. DNAMAN can display Bootstrapping values or Branch length, but not both options.

You may change the display options of phylogenetic tree by using the **Edit | Options** command, or double-click the tree window. This command opens an **Options** dialogue box. You may change all options described above, move or resize the graph and change text fonts used in the graph.

Chapter XI Restriction Analysis

DNAMAN provides many functions to help users performing restriction analysis. You may analyze restriction sites and draw a restriction map on a given sequence with a list of enzymes. You may also predict a restriction pattern on agarose gel when one or more DNA fragments are digested with restriction enzymes. You may use the electronic cloning function to generate new DNA constructs. The Silent Mutation and Directed Mismatch functions may help you to create or remove desired restriction sites on a target sequence.

XI.1 Restriction site analysis

The restriction site analysis function includes many features of enzyme cutting information on a DNA sequence. Choose the ***Restriction | Restriction Analysis*** command to perform restriction analysis on the current DNA sequence. A **Restriction Analysis** dialogue box appears.

There are many options in the first page of **Restriction Analysis**:

1) **Results: Show summary**
   Check this option to display the summary of restriction analysis in a text window. The summary includes the total number of enzyme cutting sites, a list of cutting enzymes with recognition sequences and their cutting positions, a list of cutting sites ordered by position and a list of non-cutting enzymes.

2) **Results: Show sites on sequence**
   Check this option to display the sequence and cutting sites in a text window. The first letter of enzymes is located exactly on top of each cleavage base.

   If this option is checked, you may check the **With double-stranded sequence** option to display double-stranded sequence with restriction sites.

3) **Results: Draw restriction map**

Check this option to display the restriction map in a Map window. You may modify the map with numerous tools as described in the Restriction Map section.

If this option is checked, you may check the **With enzyme position** to add enzyme positions with enzyme names on the restriction map. You may check also the **With double-stranded sequence** option to display double-stranded sequence in the map window. Check the **Including annotations** option to display annotations associated with the sequence in the map.

4) **Results: Draw restriction pattern**
Check this option to display the restriction pattern of the target sequence(s) digested with selected enzymes. The restriction pattern is shown in a graphic window. This option must be checked if you like to use the electronic cloning function.

5) **Results: Ignore enzymes with more than**
Check this option and enter a number in the behind box. If a restriction enzyme gives sites more than the number, DNAMAN will not show the enzyme.

6) **Results: Ignore enzymes with less than**
Check this option and enter a number in the behind box. If a restriction enzyme gives sites less than the number, DNAMAN will not show the enzyme.

7) **Target DNA: Circular**
Check this option if you want to correctly draw a restriction map or find restriction pattern of a plasmid sequence.

8) **Target DNA: All sequences in channels**
Check this option if you want to analyze and compare the restriction properties of all sequences in the channels. It is only activated when more than one channels contain DNA sequences.

9) **Target DNA: dam methylation**
Check this option if the DNA is dam methylated.

10) **Target DNA: dcm methylation**
Check this option if the DNA is dcm methylated.

For more information on the dam and dcm methylases and the affected restriction enzymes, choose the ***Info* | *Methylase*** command.

Press the **Next** button to go to the **Enzyme Selection** page. You may select:

- a restriction data file,
- a restriction enzyme list according to enzyme properties,
- restriction enzymes for analysis.

1) **Selecting an enzyme data file**

All available enzyme data files are listed in the **Enzyme File** box. Enzyme data files are stored in the DNAMAN system folder. DNAMAN provides two enzyme files, restrict.enz (180 enzymes) and dnamanre.enz (2524 enzymes). Select one enzyme data file in the **Enzyme File** box. You may save your own enzyme file from this dialogue box. After selecting an enzyme list for restriction analysis, save the list by pressing the **Save List** button.

2) **Selecting an enzyme list**

You may shorten the selection list from a enzyme data file by restricting the cutting properties of enzymes.

DNAMAN provides two options **Cutter** and **End** that allow you to quickly select desired enzymes:

Cutter	***End***
- All	- All
- ≥ 5	- Blunt
- ≥ 6	- 5' Overhang
	- 3' Overhang

Note that only A, C, G and T are considered for cutting length. For example, AccI "GT(A/C)(T/G) AC" is not in the list of ">=6 cutter".

DNAMAN lists available enzymes according to these two options from the enzyme data file.

3) **Selecting enzymes for restriction analysis**

In the dialogue box, the available enzymes in the enzyme data file are listed in left box. Select enzymes from the left to right enzyme box by **double-clicking** the enzyme names. You may select all the enzymes listed in the left enzyme box by clicking the **Select All>>** button.

To remove an enzyme from the selected enzyme box, double-click on the enzyme name. Clicking the **Clear<<** button removes all the enzymes from the right box.

You may select any number of enzymes for restriction analysis. However, if you have checked the **Draw restriction pattern** and **All DNA in sequence channels** options in the **Restriction Analysis** page, there is a limit of enzymes you can choose. If there are two DNA sequences in sequence channels, you cannot select more than two enzymes. If there are more than two sequences available in channels, you can choose only one enzyme for analysis.

XI.2 Custom restriction enzyme data file

DNAMAN provides two restriction enzyme data files: "**restrict.enz"** and "**dnamanre.enz**" which include 180 and 2524 restriction enzymes respectively. DNAMAN also allows users to create new enzyme files according to their needs.

Creating a custom restriction enzyme data file

1. Choose the ***Restriction | Restriction Analysis*** to display a dialogue box and go to the **Enzyme Selection** page.
2. Select enzymes by double-clicking the enzyme names in the left box. For example, select the restriction enzymes listed in the puc18's multiple cloning sites.
3. Click the **Save list** button to open the Open File dialogue box.

4. Type a filename, e.g., puc18mcs.
5. Click **OK**. DNAMAN will add "enz" as the extension name. The new enzyme file name will be **puc18mcs.enz**.

Editing restriction data file

You can edit the default restriction enzyme file.

1. Choose the ***Restriction | Restriction Analysis*** to display a dialogue box and go to the **Enzyme Selection** page.
2. Select the name of the restriction enzyme data file you want to edit from the **Enzyme File** box.
3. Click the **Cancel** button to exit. The selected data file will be the default data file.
4. Choose ***Info | Restriction Enzymes*** to open an **Enzyme Information** dialogue box.
5. Click the **Edit** button to open the enzyme data file. Follow the steps in page 13-2 to edit this file.

XI.3 Results of restriction analysis

According to the options you have checked in the **Restriction analysis** page, DNAMAN shows the analysis results in different windows.

XI.3.1 Summary in text window

Restriction analysis with single DNA sequence

DNAMAN shows the number of enzymes that has been used for screening and the number of cutting sites found in the sequence. The analysis results are displayed in three parts:

1. The cutting sites listed in alphabetical order of restriction enzymes. Enzyme recognition sites and cutting number and position are also shown.
2. The cutting sites listed in position order.

3. The names of non-cutting enzymes are listed.

If you have selected **only one** or **two** enzymes, the sizes of the restriction fragments of the single and/or double digestion will be displayed.

For example, PstI and SacI are selected in restriction analysis. DNAMAN adds the following information in the result table:

Fragments generated by digestion (bp):

For the **linear DNA**:
Cut by PstI: 424, 610, 792, 975, 2297
Cut by SacI: 320, 1783, 2995
Double digestion: 94, 198, 226, 610, 698, 975, 2297

For the **circle DNA**:
Cut by PstI: 424, 610, 792, 3272
Cut by SacI: 320, 4778
Double digestion: 94, 198, 226, 610, 698, 3272

If you have checked the **Show sites on sequence** option, DNAMAN displays the sequence and cutting sites in the **Summary** window. The first letter of enzymes is located exactly on top of each cleavage base.

Restriction analysis with multiple DNA sequences
If you perform restriction analysis with multiple sequences, DNAMAN shows a table with all cutting enzymes on each sequence. For example, the result of cutting **pUC18** and **Insert** with following enzymes:

```
Restriction analysis on 2 sequences
Methylation: dam-No  dcm-No

BalI        TGG/CCA
Insert      1446

DraI        TTT/AAA
pUC18       1094  1786  1805
Insert      3232  3251  3943
```

```
HindIII     A/AGCTT
pUC18       281
Insert      29

MstI        TGC/GCA
pUC18       426   1449
Insert      262   1358  1456  3588

NaeI        GCC/GGC
Insert      403   771   931   1285

NarI        GG/CGCC
pUC18       446
Insert      414   435   549   1206
```

XI.3.2 Restriction map

DNAMAN draws a restriction map with all restriction enzyme sites available on the sequence. The map will be circular if you have checked the **Circular** option in the **Target DNA** section. The restriction site positions will be labeled on the names of restriction sites if you have checked the **With enzyme position** option in the **Results** section. The resulting map provides a starting point for drawing a complete DNA map.

XI.3.3 Restriction pattern in graphic window

See the section **XI.5** for **Restriction pattern**.

XI.4 Drawing restriction map

DNAMAN provides convenient tools to produce publication-quality restriction maps. These tools can be used to draw linear or circular restriction maps. You may also use this drawing tool for other projects, such as PCR strategy diagrams, gene structural maps, etc.

XI.4.1 Starting drawing map

There are two accesses to draw a restriction map:

1. Use the ***Restriction | Draw Map*** command if DNA sequence is not available. A plasmid map will be drawn immediately in a **Restriction Map** window. Since sequence information is not available, there is no sequence associated with the map. You may double-click in the window to modify the properties of the map.

2. If DNA sequence is available, you can draw a restriction map according to the data from restriction analysis. Choose the ***Restriction | Restriction Analysis*** command and select ***Restriction Map*** option to start drawing the restriction map. There is a DNA sequence associated with the map. Any modification of DNA fragment on the map may result in the change of sequence.

XI.4.2 Map properties

By double-clicking the map window, a dialogue box will appear for modification of the map properties.

The first page contains general information about the restriction map. You may define the name and the size of the map. You may also change the type of the map to **linear** or **circular**.

There are two parameters to define the location and the size of the map;

1) **Map frame.**
   The size of a plasmid circle, or the length of a linear fragment, is defined by map **Diameter**. The value is in logical point units that are based on your screen resolution. The **Size** (thickness) of the frame is also in logical point units.

2) **Map center position**.
   The center is defined as X position, from left to right of the screen, and Y position, from top to bottom. The unit is in logical point as well.

There are three options to change the appearance of the map:

1) **Show scale of linear map**. This option applies only to linear map. If checked, DNAMAN will draw a scale to show the length unit of the linear Map.

2) **Show end positions**. This option applies to both linear and circular map. In a linear map, the start and end positions are labeled if this option is checked. In a circular map, the top position will show up when this option is checked.

3) **Show enzyme position in new site(s).** If this option is checked, DNAMAN will display the restriction sites together with their positions when they are added to the map.

This page contains also reference information about the map. You should enter the **Author** of the map and other relative information such as description.

The second page contains the list of elements in the map. From this page, you may add new element(s), modify or remove existing elements. The **Name** button allows you to link an element to an available text object in the map.

The third page contains the list of text objects in the map. From this page, you may add new text object(s), modify or remove existing text objects.

The fourth page contains the list of sites in the map. From this page, you can add new restriction site(s), modify or remove existing sites.

If the map contains DNA sequence, the fifth page shows the information about **Sequence View**. In this page:
1) You may choose different fonts to display the sequence, enzymes and position labels.
2) If the **Show double stranded sequence** option is checked, the sequence in the map will be double stranded.
3) When the **Scroll sequence to match cursor** option is checked, the sequence region will be scrolled and shown when you move the cursor on the map.

4) If the **Scroll on clicking object** option is checked, DNAMAN scrolls the sequence to display the region of the object when you click it on the map.
5) During sequence modification, the restriction map may be updated if the sequence length change is more than a defined value by **Recalculate site locations when length change>=**. Set the value to zero if you like to update immediately after any modification.
6) You may add new restriction sites to the may by screening enzyme cutting sites again. When you press the More restriction sites option, DNAMAN allows you to select restriction enzymes, and looks for new sites on the sequence.

XI.4.3 Methods of map drawing

DNAMAN provides the following methods for drawing maps:

1. Adding a site
2. Adding an element
3. Adding a text object
4. Adding a Box object
5. Deleting a Box object
6. Deleting or editing a site
7. Deleting or editing an element
8. Deleting or editing a Text object
9. Moving a site name, an element or a Text object
10. Text alignment
11. Changing text font
12. Inserting or removing a DNA fragment
13. Copying or cutting a DNA fragment
14. Pasting a DNA fragment
15. Adjusting the scale of linear map
16. Resizing restriction map
17. Moving restriction map
18. Rotating and opening circular restriction map
19. Editing sequence

1. Adding a site

1) Place the cursor to the desired position on the map (the coordinate is displayed on the left corner of the screen). If

the enzyme site is within an element, you should first move the element out of the map frame (inward or outward in a circular map, up or down in a linear map).

2) When the cursor is a cross, click to display a popup menu.
3) Select the ***Add Site*** command to open the **New Site** dialogue box.
4) Type the site name. You may also define the position number in the position box.

2.Adding an element

Click on the map frame to display a popup menu.

Clicking the **Add Element** menu opens the **Element** dialogue box. You should fill in the five parameters (type, start position, end position, size and color) to create the element.

Arrow element is directional. Direct the arrow to clockwise or from left to right by entering the numbers in ascending in the Start (bp) and the End (bp) columns (e.g., Amp element from 3000 to 3500). You may reverse the arrow for anti-clockwise or from right to left by entering the number in descending (e.g., Lac element from 2800 to 2300).

Bar element is not directional and you can enter the numbers in either way.
To draw an element crossing the position +1 of a circular map, the end position should be the real position plus the size of the plasmid. For example, the Promoter A is from +4200 to +50, therefore, type 4200 in the box of Start (bp) column and 4350 in the box of the End (bp) column that is sum of 50 plus the size of the plasmid 4300.

Eight colors (patterns) are listed in the Color column of the dialogue box:

3. Adding a Text object

Choose the ***Edit*** **|** ***Add*** **|** ***Text*** command to open the **Text** dialogue box. Place the cursor to where the text will be and

click mouse left button. Type the text in the edit box. You may change the font, style, size and color of the text. Choose the **Vertical** option if you like the text to be drawn vertically. After clicking the **OK** button, the text object will be added in the map.

3. Adding a Box object

Choose the ***Edit*** **|** ***Add*** **|** ***Box*** command. Place the cursor to where the Box object will be and click mouse left button. A box appears on the map. You may move it when the cursor is placed at the top-left corner, change its size when the cursor at the bottom right corner

5. Removing or modifying a Box object

A Box object can be one of these shapes: box, rounded box, circle, line, and arrow lines. The shape can be changed when double-clicked. In the box list box, double-click the object to change its property. To remove a Box object, select it using mouse and choose the ***Edit*** **|** ***Delete*** command.

6. Removing or editing a site

1) Double-click on the site name.
2) Delete the name in the edit box for removing the site, or modify the name for editing.

   If you want remove more than one site, use mouse to select all of them, then choose the **Edit | Delete** menu to remove all.

7. Removing or editing an element

1) Double-click on the element.
2) Select a NULL type to delete the element, or modify the parameters for editing.

8. Removing or editing a Text object

1) Double-click on the Text object.
2) Delete the name in the edit box for removing the Text object, or modify the text or font for editing.

9. Moving a site name, an element or a Text object

Place the cursor on the object, then drag and drop it to an appropriate position.

10. Text alignment

You may align Text objects and site names in four directions and arrange them horizontally or vertically with a defined interval.

1) Press the left mouse button to select the Text objects and site names. There are six commands activated in the **Edit** menu.
2) Select the **Edit | Object Align Left** command to align the objects to the left.
   Select the **Edit | Object Align Right** command to align the objects to the right.
   Select the **Edit | Object Align Top** command to align the objects along the top.
   Select the **Edit | Object Align Bottom** command to align the objects along the bottom.
3)-Select the **Edit | Object Align As Column** command to open the **Column** box where type a number to set up the vertical alignment interval.
   Select the **Edit | Object Align As Row** command to open a dialogue box where type a number to set up the horizontally alignment interval.

11. Changing text fonts

Double-clicking a text object to open the **Text** dialogue box. Clicking the **Font** button, then select an appropriate font, size style and color and click **OK**.

You can also select a group of text objects, then change the font. There is an option to display the text horizontally or vertically.

12. Inserting or removing a fragment

1) Place the cursor in the position where you want to insert or delete a fragment.

2) Click the cross shape cursor to display a popup menu.

3) Select the ***Insert Fragment*** command to add a fragment in the map. If you want to delete a fragment choose the ***Remove Fragment*** command.

4) In the **Add Fragment** dialogue box, you can change the inserting position by typing another number.

   In the **Remove Fragment** dialogue box, type in the start and the end positions of the fragment you want to be removed.

5) Click **OK** to continue.

13. Copying and cutting a fragment

1) Place the cursor in the approximate position where you want to copy or cut a fragment.

2) Click the cross shape cursor to display a menu.

3) Select the ***Copy Fragment*** command to copy a fragment to the Clipboard. If you want cut a fragment and place it into Clipboard, choose the ***Cut Fragment*** command.

4) In the **Fragment Positions** dialogue box, type in the start and the end positions of the fragment you want to copy or cut.

5) Click **OK**. The defined fragment will be copied or cut into the Clipboard.

14. Pasting a DNA fragment

Be sure a fragment has been copied or cut into the Clipboard.

1) Place the cursor in the approximate position where you want to insert or delete a fragment.

2) Click the cross shape cursor to display a menu.

3) Select the ***Paste Fragment*** or ***Paste in Reverse Direction*** command.

4) In the **Paste Fragment** dialogue box, you can change the position by typing another number.

5) Click **OK**.

   If you paste a DNA fragment with sequence to a map without sequence, the sequence in the fragment will be

ignored. If you paste a DNA fragment without sequence to a map containing sequence, DNAMAN will ask if you want to insert "N"s in the map sequence, or remove existing map sequence.

15. Adjusting scale of a linear map

Double-clicking on the text of the scale to open a dialogue box. You may change the scale by typing an appropriate number and clicking **OK**.

16. Resizing a map

You may increase and decrease **the length of a linear map** by placing the cursor to the **right** end of the map, then pressing the left mouse button. When the cursor switches to a small square, drag it to left or right to make adjustments to the map.

The **size of a circular map** can also be enlarged or reduced by placing the cursor to the circle of the map (not on an element). When the cursor is a cross, press the left mouse button and then make adjustments by moving the cross cursor inward or outward.

17. Moving a map

You can move the restriction map within the graphic window. Place the cursor in the **center** of a circular restriction map, or the left end of a linear map, then press the left mouse button, and drag and drop the map to the desired place.

18. Rotating and opening circular restriction map

DNAMAN allows you to rotate the restriction map. There are two methods.

Method 1:

Check the **Show end position** option when setting map properties, the top position appears on the map. Double-click this number to open the **Text** dialogue box and type a new number to rotate the restriction map.

Method 2:

1) Check the **Show end position** option when setting map properties, the top position appears on the map. When you move the cursor under or above this number, the cursor switches to the ⇑ shape.
2) Press and keep the mouse button down to rotate the restriction map.

When you draw a cloning diagram, you may want to illustrate the insertion on a restriction map. DNAMAN allows you to open the circular restriction map.

1) Place the cursor above the center of a circular restriction map and the cursor then switches to the ⇔ shape. If you want to open the circular map at a position rather than at +1, you should first rotate the desired position to the "top" position.
2) Press and keep the mouse button down to open the restriction map. If you want to open it wider or narrower, place the cursor to an end and then press the left mouse button.

Note that the action of resizing, rotating or opening a restriction map results in the recalculation of the site positions.

19. **Editing Sequence**

If the sequence is available in the map, you may modify it with editing functions, e.g. insertion and deletion of bases, sequences. ***Copy*** and ***Paste*** may also be used. The map will be automatically updated when you make any modification of the sequence.

If you paste a DNA fragment containing sequence to a map without sequence, the sequence in the pasted fragment will be ignored. If you paste a fragment without sequence to a map containing sequence, DNAMAN will ask you if the letter N should be added to the map or all sequence in the map should be removed.

XI.5 Management of map files

DNAMAN provides a useful function to manage the map files. It is recommended to store all map files in a designated folder. The default folder is "seqmap" in the DNAMAN folder. All map files in this folder will be displayed when you choose the ***Restriction | Restriction Map Files*** command.

In the **Map File List** dialogue box, you may find the information in any map file by clicking the file in the list box. DNAMAN shows the map name, map length, and the type of the map (Linear or Circular). DNAMAN shows also the author and description of the map. The number of restriction sites, elements and text objects are also displayed.

You may highlight a map file and click the **Open** button to open the file in DNAMAN. You may also list the information of all map files by clicking the **Show All** button. The content of the list may be printed out using the **File | Print** command.

XI.6 Restriction pattern

DNAMAN predicts the pattern of restriction fragments that are separated by using agarose gel electrophoresis. DNAMAN assumes that electrophoresis is under ideal conditions, and all the fragments migrate in a logarithmic function of the inverse of their sizes

.

Choose the ***Restriction | Restriction Analysis*** menu to start restriction analysis on the one or multiple DNA sequences. Check the **Results: Draw restriction pattern** option.

After selecting restriction enzymes and pressing the **Next** button, the analysis results are shown in a graphic window. Depending on the number of DNA sequences and the number of enzymes you have selected, the restriction pattern may be different.

1) One DNA sequence cut with more than two enzymes:
   The pattern shows all restriction patterns of the DNA cut by each enzyme.

2) One DNA sequence cut with two enzymes
   The pattern shows all restriction patterns of single digestion of the DNA with each enzyme and double digestion with the two enzymes.

3) Two DNA sequences cut with two enzymes
   The pattern shows all restriction patterns of single digestion of the two DNA sequences with each enzyme and double digestion with the two enzymes.

4) More than two DNA sequences cut with one enzyme
   The pattern shows all restriction patterns of each DNA digested with the enzyme.

The fragment size range in the pattern diagram is from 1 bp to 50 kb. To change the resolution of the pattern, you may set different sizes at the bottom of gel. The default size of the bottom size is 100 bp. The **Bottom** control allows you to modify this value.

You can get detailed information about the fragments on the gel. Place the cursor on a fragment and click the left mouse button to open a menu box, the size of the fragment and the cutting enzymes are indicated in a pop-up box.

You may change the size and shape of the gel. Place the cursor on the **low-right** corner of the gel. When the cursor switches to ⇔, press the left mouse button and drag it to change the size and shape of the gel.

You can also move the gel within the graphic window. Place the cursor on the **up-left** corner of the gel. When the cursor switches to  , press the left mouse button and drag the gel.

XI.7 DNA cloning

In a cloning process, there are two major steps: generating an **Insert** and a **Vector** fragments and ligating them to create a

new clone. DNAMAN mimics this process to produce a new cloned sequence. This function can help you design a cloning strategy prior to laboratory works.

Generating Insert and Vector fragments

Insert and **Vector** fragments can be obtained from the same or different source sequences.

1. Perform restriction analysis with the source DNA, check the **Results: Draw restriction pattern** option. The analysis results are shown in a restriction pattern graph.
2. Move the cursor to the desired fragment, then click the left mouse button to display a popup menu. If the enzyme cutting sites are available, this menu allows you to load the fragment as **Vector** or **Insert**.
3. Select the fragment as the **Vector** or the **Insert** that will be stored in the memory.

Repeat the process to load the other source of DNA for cloning if necessary. The ***Restriction*** | ***Cloning*** menu becomes activated when the **Vector** is loaded into memory.

Ligating Insert and Vector

Choose the ***Restriction*** | ***Cloning*** command to perform ligation. A **Cloning** dialogue box appears to display the information of the Insert and Vector. Check the Vector and Insert information to make sure that no error occurs. Otherwise, you must back to restriction analysis and load the correct fragment into the memory.

There are some conventions for DNA fragments in the **Cloning** table:

1) The direction of a fragment is defined from the 5' end to the 3' end. The ligation is performed between the 5' end of the insert and the 3' end of the vector, and between the 3' end of the insert and the 5' end of the vector.

2) For each fragment, the blunt end is **Blunt**. The symbols '**+**' and '**-**' are related to the **upper strand**.

5' end +: 5' protruding termini; 5' end -: 5' recessed termini
3' end +: 3' protruding termini; 3' end -: 3' recessed termini

Example:

```
5'-AGCTGCGCGCGCTTTATTATAGCTGTAGCTA-3'
3'-    CGCGCGCGAAATAATATCGACATCG  -5'
```

The ends are: 5' end +AGCT
3' end +TA

```
5'-AGCTGCGCGCGCTTTATTATAGCTGTA    -3'
3'-    CGCGCGCGAAATAATATCGACATCG  -5'
```

The ends are: 5' end +AGCT
3' end - CG

3) Fragment end may be modified by clicking the button behind the position box. There are three options: **Keep**, **Fill in** and **Delete**.

4) **Insert** sequence may be reversed by checking the **Reverse insert orientation** option.

5) If you use linkers in the cloning construction, the linker sequences are related to the upper strand. The linker **Vector-Insert** is between the 3' end of the Vector and the 5' end of the Insert, and the linker **Insert-Vector** is between the 3' end of the Insert and the 5' end of the Vector.

6) You may perform Vector self-ligation by checking the **Vector recircularization** option.

If the ends of the insert and vector are not compatible, a message box will appear to indicate the incompatibility.

XI.8 Reconstruction of restriction map

When you have a DNA fragment without sequence information, you may want to find out its restriction map before any further work. The process might involve in digesting the DNA with single and double enzymes, estimating the sizes of all the restriction fragments using

electrophoresis, and deducing the restriction map according to the fragment sizes. Reconstruction of restriction maps may be complicate. The difficult increases when more fragments are produced by enzyme digestion. DNAMAN use a dynamic process to find all possible combinations of restriction fragments that match the digestion patterns.

In the experiment, you should use two restriction enzymes to digest the target DNA. The size of all the fragments in single and double digestion must be estimated from electrophoresis as accurate as possible. Choose the ***Restriction | Map Reconstruction*** command to open a dialogue box. Fill out the experimental data into to the table. The sums of the total fragment size of Enzyme A, Enzyme B and Enzyme A+B must be equal.

The size of the DNA fragments should be estimated accurately. You can have an error bar for the estimation. The default error is set to 10%. You may decrease it if you are confident on the estimation.

The deduced restriction map is shown in a graphic window. If more than two maps are found from the provided data, DNAMAN draws only the first two maps. The number of the possible maps is indicated at the bottom of the graphic window. In this case, the more accurate maps may be found by decreasing the error rate.

If you have more than two enzymes for digestion, perform separately the analyses of each enzyme combination. Use only two enzymes for each analysis, then construct the final map from the combination of individual maps.

XI.9 Silent mutation

The function of **Silent Mutation** allows you to change restriction properties of the current DNA sequence without changing the coding amino acid sequence. This function searches for a potential region that can be altered to create or destroy restriction enzyme sites.

1. Choose ***Restriction*** **|** ***Silent Mutation*** to open the **Silent Mutation** dialogue box. This command is activated when the current channel contains a DNA sequence.
2. Select a restriction enzyme file. "restrict.enz" is the default enzyme file in DNAMAN.
3. Select the one of the **Cutter** options.
4. Define the region you want to be analyzed in the **Seq** box.
5. Click the **Seq** button to confirm the both ends of the selected region and the coding amino acid sequence in Reading frame 1.
   * If you are interested the Reading frame **2**, increase the one base of the start point in the **Seq** box.
   * If you are interested the Reading frame **3**, increase the two bases of the start point in the **Seq** box.
6. Click the **OK** button to perform analysis and display the results in a text window.

Analysis results:
Silent mutation analysis is performed on the following three sequences to find different restriction sites.

WT: Wild type sequence

RT1: Degenerated sequence **1** from the amino acid sequence translated from WT sequence in reading frame 1.

RT2: Degenerated sequence **2** from the amino acid sequence translated from WT sequence in reading frame 1.

RT1 and RT2 are degenerated sequences according to the universal genetic code. Most amino acids can be reverse-translated to one triplet using IUPAC code, but two triplets must be used for Leu, Arg and Ser. Thus, RT1 represents one reverse translated sequence and RT2 represents another one.

Analysis results show all possible restriction sites in the three sequences. By comparing the difference of restriction sites

among WT, RT1 and RT2, you can find suitable positions to generate a silent mutation.

Example:

```
W.T. TTTGACTGCC ACTTCCTCGA TGAAGGTTTT 30
RT1  TTYGAYTGYC AYTT YCTNGA YGARGGNTTY 30
RT2  TTYGAYTGYC AYTTYTTRGA YGARGGNTTY 30

TCTAGA XbaI
W.T.:
RT1 : 15
RT2 :
```

The DNA sequence in RT1 at the position 15 is **YCTNGA**. According the IUPAC code, the YCTNGA represents the ambiguous sequences of

TCTAGA,	CCTAGA
TCTCGA,	CCTCGA
TCTGGA,	CCTGGA
TCTTGA,	CCTTGA.

You can create a mutant by changing your original sequence "CCTCGA" at the position 15 to "TCTAGA". It will produce a restriction site of XbaI at that location without altering the coding amino acid sequence.

XI.10 Directed mismatch

Directed mismatch allows you to create or remove restriction sites on the current DNA sequence or its variants by incorporating single or double mismatches at a site near the mutation. Using this function you can create or destroy restriction sites in order to distinguish the wild type allele from a common mutant allele.

1. Choose the ***Restriction*** | ***Directed Mismatch*** command to open the **Directed Mismatch** dialogue box.
2. Select a restriction **enzyme file** and the **Cutter** option. "restrict.enz" is the default enzyme file in DNAMAN.
3. Type a number in the **Mutation at** box to select the mutant position.

4. Type A, C, G or T code in the **for** box to define the mutation.

6. Click the **Site** button to confirm the sequence and mutation.

7. Select one of the **Accept Mutation** options:
   Non option searches for restriction sites without any mismatch.
   <= 1" option searches for restriction sites that can accept a single mismatch.
   <= 2 option searches for restriction sites that can accept double mismatches.

Example:

```
1       TTTGACTGCCACTTCCTCGATGAAGGTTTTACTGCCAAGG
        ------------------C--------------------
```

Mutation at position 20

Maximum Mismatch = 2

W.T. indicates the wild type sequence
'*' indicates mutation position
'^' indicates mismatch

AflII

```
Mutant:  CTGAAG
Enzyme:  CTTAAG
          * ^
```

A single mismatch "T" near the mutant allele "C" creates the enzyme site of AflII in the mutant sequence, but not in the wild type sequence.

```
Eco57I
Mutant:    CTGAAG
Enzyme:    CTGAAG
           *
```

The enzyme site of Eco57I exists in the mutant without incorporating any mismatch.

```
XbaI
W.T. :  CCTCGA
Enzyme: TCTAGA
```

^ ^ *

Double mismatches “T” and “A” near the wild type allele “A” creates the enzyme site of XbaI in the wild type sequence, but not in the mutant.

Chapter XII Protein Sequence and Translation Analysis

DNAMAN provides various functions for DNA translation and protein sequence analyses. Seven different genetic code tables are available for translation analysis. You may also add user defined genetic codes to the list. Translation overview may help you to visualize graphically the locations of open reading frames. You may calculate amino acid composition of a protein sequence, estimate its isoelectric point, plot its hydrophobic/hydrophilic profile and predict its secondary structures.

XII.1 Genetic code table

You may choose one of genetic code tables for translation analysis. Universal code is the most used in translation analysis for all biological species. However, in some case you may need to work with a special genetic code for specific works, such as translating yeast mitochondrial DNA to protein.

DNAMAN provides seven genetic code tables for your choice. In addition, you may create your own genetic code for your work. You may select any of these codes as the default genetic code. The default genetic code will be used in all functional analyses concerning sequence translation.

XII.1.1 Selecting default genetic code table

Choose the ***Protein*** | ***Select Genetic Code Table*** command to open the **Genetic Code Table** selection box. All available codes are listed in the box. Click the code name to select it as default code.

XII.1.2 Adding and editing genetic code table

To add a new genetic code table or edit an existing one, click the **Edit** button in the **Genetic Code Table** selection box. The table file is in text format. You may edit it according to your preference. However, there are some rules to make DNAMAN correctly use this file. To add a new code table:

1. Start with "//" in the first line.

2. Type the code name in the second line.
3. Enter the table. Use one letter codes for amino acids, lowercase letter for start codon, * for stop codon. You can define up to five start codons and five stop codons in a genetic code table. The code order should follow the standard universal genetic code table from top to bottom, from left to right.

You may visualize the genetic table by choosing the ***Info*** **|** ***Genetic Code Table*** command. DNAMAN displays the universal genetic code table and some variations.

A. Universal code
B. Vertebrate mitochondrial code
C. Yeast mitochondrial code
D. Insect mitochondrial code
E. Protozoa mitochondrial code
F. Plant mitochondrial code
G. Ciliate protozoa nuclear code

XII.2 Translation overview

To facilitate the searching for ORFs on a large DNA sequence, DNAMAN provides a graphic presentation of translation overview on the six reading frames of the current DNA sequence. With the overview, you may easily locate the translation start and stop positions. However, the overview cannot give correct prediction when introns are present.

In the graph, the top straight line represents DNA sequence. This line is equally divided into ten parts as scale. Six diagrams are used to represent the six reading frames of the sequence and each line represents one of them. On each diagram, the Initial codons are shown as short vertical lines above each sequence line and the Stop codons shown under the sequence line. The top three diagrams show the reading frames on **Plus** (sense) strands and the bottom three diagrams show the **Minus** (anti-sense) strands. The names at left indicate the Reading Frames.

Since translation initial and stop codons vary with different genetic code tables, the diagrams will change when you select another genetic code table.

You may change the location and length of the diagrams. Place the cursor on the **right** end of the first line. When the cursor switches to ⇔, press the left mouse button and drag it to change the length of the diagrams. To move the diagrams, place the cursor on the **left** end of the first straight line (scale). When the cursor switches to  , press the left mouse button and drag the diagrams to another location.

When the cursor is a cross on the overview graphics, the position coordinate is displayed on the left corner of the screen.

There are two methods to display sequence coding information from the Overview:

1) Double click an ORF bar on the graph. DNAMAN will show the amino acid sequence of the ORF with or without DNA sequence.
2) Select a region of any reading frame with mouse. DNAMAN will display the DNA sequence with amino acid sequence in the selected reading frame and region.

When the **Show Position Preference** option is checked, DNAMAN shows codon bias curve of all reading frames. In each reading frame, the bottom is 0, top 100% and middle line 50%.

You may modify the Overview graph by pressing the **Options** button. A dialogue box appears to show the options.

1) **Show Star/Stop codon positions**. If this option is checked, DNAMAN shows all Start and Stop codons on the diagrams with vertical lines. You may hide these lines if the option is unchecked.
2) **Show possible ORFs**. If this option is checked, DNAMAN draws Arrow bars on all possible open reading frames. Arrow indicates the direction of translation. If the Use Start codon is checked, an ORF always starts with a Start codon.

If it is unchecked, an ORF may start with any codon except Stop. A possible ORF is defined by **Minimum length**. Any continuous reading frame larger the minimum length would be considered as a possible ORF. You may change the Arrow bar size in the **Width** box.

3) **Export only protein sequence on double-click**. By double clicking on a possible ORF, DNAMAN pops up a Text window to show the translation of the ORF. If this option is checked, DNAMAN shows only amino acid sequence. Otherwise, both DNA and protein sequences are shown in the translation window.
4) **Show Position Preference**. Check this option to display position bias of the sequence. The position can be any one of the three code of amino acid. The Preference code can be any one or more of the A/C/G/T bases. The Window size is used to smooth the curve. When Show Detail is checked, the Window size is set to 1. In this case, every codon bias is displayed.
5) **View**. There are six boxes to define the sequence start and end positions, and the position and size of the graph in the window. You may modify them according to your preference.

XII.3 Translating DNA sequence to protein

Choose the ***Protein*** | ***Translation*** command to open the ***Translation Parameters*** dialogue box. You may choose necessary options to display translation results.

1) Amino Acid symbol

You may select **One letter** or **Three letter** code for amino acids.

2) Reading Frame

You may choose Reading **Frame 1**, **Frame 2** and **Frame 3.**. You may also choose the **All** option to translate the DNA sequence in all three reading frames.

3) Show DNA

No DNA sequence appears when the **No** option is chosen. Double DNA sequence shows up when the **Double Strand**

option is selected. With this option, **All** six reading frames are translated to amino acids.

4) Annotations
If there are annotations in the sequence, you have the options to deal with them.

a) **Excluding** some or all annotations. Choose the **Exclude** option and select one or more annotations. DNAMAN will translate all sequence except the regions defined in selected annotations.
b) **Using only** some or all annotations. Choose the **Use only** option and select one or more annotations. DNAMAN will translate only the sequence regions defined in selected annotations.
c) **Including** all annotations. Do not select any of the annotations. DNAMAN will translate all sequence regardless of annotations.

5) Result Format: **AA per line**
You may define the number of amino acids of each line in the output result.

6) Resulting AA sequences to channel No option
Check this option if you like to perform further analysis of the translated amino acid sequences.

7) Space between 3-letter code
When 3-letter code and only one of the three reading frame are checked, this option will group protein and/or DNA sequence in three.

If IUPAC code resent in the default DNA sequence and the coding amino acid is undefined, "X" may appear in that position.

DNAMAN can translate a part of the default DNA sequence to amino acid sequence. You may define the analysis region, such as an open reading frame, for translation. Choose the ***Sequence | Current Channel | Analysis Definition*** command to specify the region you want to analyze. Then choose the translation command to display the results.

```
1     CTAAAGAAACATCTGAGGTGGTTAAAAGCTCTCCCTCGTGTCACCCCCTTTTATGCAGTCAAATGTAATGATAGCAAAGCCAAGACGAAG
91    ACGAGAGAATGACAAAAACACTCATGTATTACGGGAATGATGGGTGTCTATGGATCGTTCAATTGCATCTTGTATGATCATGCACATGTTAA
1           M  T  K  H  S  C  I  T  G  M  M  V  S  M  D  R  S  I  A  S  C  M  I  M  H  M  L  N
181   ACCAGTTCTGCAAAAGCGGCCTAAACCAGATGACGGCTGCTACTCCTGCAGCATATGGGGCAACGGATTCCAGAGGCCAACAATACATTA
29      Q  F  C  K  S  G  L  N  Q  M  T  A  A  T  P  A  A  Y  G  A  T  D  S  R  G  Q  Q  Y  I  M
271   TGCGATGCTCCTGTGCCTGTGAAAGTGGAATTGAGTATCCAGCAACTTGTGCTTCAGCTAGTATTAATGTATAAATAATAAATACTTAGA
59       R  C  S  C  A  C  E  S  G  I  E  Y  P  A  T  C  A  S  A  S  I  N  V  *
361   AACTAACTGCAAGTTTAGTCATTGAATTTAGGGCATTTGGGGGACCATTAACTTAATTCTTGCTAGAATTTTTTAAGTGTTTTTTTAAAA
451   GTTTAGGTTTGGCATAAACACGACAAAAAA
```

The above example shows a DNA sequence and the translated amino acid sequence. There are three methods for further analysis of the protein sequences. In the **Translation Parameters** dialogue box:

1) Choose the **Resulting AA sequences to channel No** option. Amino acid sequences will be loaded directly to channels for analysis.
2) Select one of the three reading frames (not **All**), and not check the **Display with DNA sequence** option. The resulting text file will be formatted and can be directly used for analysis. Save this file and load it into a sequence channel as protein sequence.
3) Alternatively, you may select the sequence and load the selection using the ***Sequence | Load Sequence | From Selection*** menu.

XII.4 Reverse translation

Using the reverse translation function allows you to find the DNA sequences based on amino acid sequence. This function is commonly used to degenerate oligonucleotide sequences from a known peptide sequence.

Choose the ***Protein | Reverse Translation*** command to open a **Reverse Translation** dialogue box. There is a Channel List box and a Sequence Text box. DNAMAN lists all available protein sequences in the **Sequence Channel** list box. You may select any of the available sequences and show it in the Sequence Text box. You may also modify these sequences or enter a new sequence in the Sequence Text box.

Four options for how to perform reverse translation:

1) Show all possible codons. This is default option. With this option, DNAMAN displays the reverse translated

DNA sequence with ambiguous nucleotides at variant position in a text window. Check IUPAC code for information about the ambiguous nucleotides. You can easily edit the sequence and select the preferred nucleotides at ambiguous positions.

The following three options are available only if you have checked the **Codon preference file** option. DNAMAN will choose precisely nucleotides for each position of the reverse translated DNA sequence, and not incorporate any ambiguous nucleotides. See the next section (**Codon Usage Analysis)** for information about the codon usage file.

2) Use the most frequently used codon(s). DNAMAN will check the codon usage table and use only the most frequently used codons of each amino acid for reverse translation.
3) Use alternatively two most frequently used codons. DNAMAN will check the codon usage table and use alternatively the two most frequently used codons of each amino acid for reverse translation.
4) Use alternatively all codons. DNAMAN will check the codon usage table and use alternatively all codons used by each amino acid for reverse translation.

XII.5 Codon usage analysis

You may analyze the codon usage of a DNA sequence in a specific reading frame. You may also save and open a codon usage table.

1) Display codon usage

Choose the ***Protein*** **|** ***Codon Usage*** **|** ***Reading Frame 1*** or ***Reading Frame 2*** or ***Reading Frame 3*** command to display the codon usage table of the current sequence. The codon usage table is shown in a graphic window.

As reference of the table, DNAMAN shows the sequence name, translation region and the reading frame number. The total codon number is also indicated. In the table, the first codons

are listed at left, the second on top and the third at right. The codon usage is shown as **Amino acid/Number/ Frequency(in 1000)/ Preference(in 100)**.

Amino acid: amino acid translated from the codon
Number: codon number presented in the sequence
Frequency(in 1000): codon frequency in the sequence
Preference(in 100): percentage preference for the code

Example: the GUG codon in a sequence of 357 amino acids shows as **V 10/28/53**.
Interpretation:

V: GUG codes for Valine (V)
10: there are 10 GUG codons in the sequence
28: GUG represents 2.8% in the sequence (10/357).
53: GUG represents 53% of codons coding for Valine (V).

DNAMAN analyzes only the **Analysis Region** of the current sequence. Use the ***Sequence* | *Current Channel* | *Analysis Definition*** command to define the open reading frame of the current DNA sequence.

Codon usage table is drawn in a graphic window, you may modify the graphic content. See the section **V.1** for adding text objects to the codon usage table.

2) Save/Append Codon usage file

You may save a codon usage table to a file or add the current codon usage to an existing table. Choose the ***Protein* | *Save/Append Codon Usage File*** command to open a **File Open** dialogue box. You may enter a new file name to save the codon usage table to a file, or select an existing file to append the codon usage to the table. The codon usage file can be used for reverse translation (Section **XII.4 -Reverse Translation**).

3) Open Codon usage file

Choose the ***Protein* | *Open Codon Usage File*** command to open a **File Open** dialogue box. Select an existing codon usage file and click the **Open** button. DNAMAN shows the codon usage table in a graphic window, as described in the **Display Codon Usage** section.

XII.6 Amino acid composition and protein properties

You may use DNAMAN to find out protein properties from its amino acid sequence. Choose the ***Protein*** **|** ***Sequence Composition & Properties*** command to display the properties of the current protein sequence. The results are shown in a text window. DNAMAN calculates and shows the following properties of the current sequence:

1) Sequence length and deduced molecular weight (MW).
2) Theoretic isoelectric point (pI) based on the sequence information.
3) Theoretic UV absorption of the protein at five wavelengths: 276, 278, 279, 280, 282 nm. The UV absorption is expressed in extinction coefficients by molar concentration (M) and weight concentration (g/L).
4) Amino acid composition. The composition of each amino acid in a protein sequence can be expressed in weight composition (grams per 100 grams of amino acids) or in molar composition (moles per 100 moles of amino acids).

XII.7 Prediction of protein charge and pI

Since proteins contain charged amino acids, such as Aspartate and Lysine, they exhibit certain charges in different pH buffers. Isoelectric point (pI) is the pH at which the protein charge becomes to zero. The number of charged residues principally determines the pI and charge of a protein. DNAMAN computes the theoretic pI and charges of the current protein sequence in a user-friendly interface.

1. Choose the ***Protein*** **|** ***Charge & pH*** command to display a dialogue box.
2. Select the number of disulfide bonds of the protein. You should provide this information to increase the accuracy of prediction.
3. The predicted pI is shown in the dialogue box. The pI may vary with disulfide bond number in the protein.
4. If you enter a pH value, then press **Enter** from the keyboard or press the **Charge ?** button, DNAMAN will

calculate the charge of the current protein sequence at the given pH.

5. If you enter a charge value, then press the **pH ?** button DNAMAN will compute the buffer pH value for the protein to exhibit the charge.

XII.8 Hydrophobic and hydrophilic profiles

DNAMAN provides graphic presentations of hydrophobic and hydrophilic profiles of the current protein sequence. The plot is based on the method developed by Hopp and Woods (1981, ***Proc. Natl. Acad. Sci.*** 78: 3824). The hydrophobic properties of a protein may help you to locate hydrophobic clusters and hydrophilic properties can be used to find potential antigen positions in the protein sequence. For more information about plotting methodology, please refer to Hopp and Woods (1981) and Kyte and Doolittle (1982, ***J. Mol. Biol.*** 157: 105).

Choose the ***Protein | Hydrophobic Profile | Current Sequence*** or ***Protein | Hydrophobic Profile | Current Sequence*** command to display the profile of the current protein sequence in a graphic window. You may also choose ***the All Available Sequences*** menu if you like to compare the profiles of all available protein sequences in channels.

The plot shows the hydrophobicity or hydrophilicity of an amino acid against its position. Averaging the values of a group of neighbor amino acids can smooth the profile. The average group length is defined by **Window Size** in the **Options** dialogue box. The default value is 6. If the **Window Size** is 1, the profile represents the hydrophobicity (or hydrophilicity) value of each amino acid of the protein sequence. With a value greater than 1, DNAMAN will repetitively average the hydrophobicity (or hydrophilicity) values along the peptide chain in the plot.

Y axis scale can be set in the **Options** dialogue box. See the section **V.5** for more information on how to handle **Profile Plot**.
* A **Stop** is indicated as a **Circle**.

* An unknown amino acid is indicated as a **Square**.

XII.9 Prediction of protein secondary structures

DNAMAN predicts the secondary structure of the current protein sequence using the **DSC** (Discrimination of protein Structure Classes) method developed by King and Sternberg (***Protein Science***, 1996, **5**: 2298-2310). This method can provide relative accurate and reliable protein secondary structure prediction.

1. Choose the ***Protein | Secondary Structure*** command to start prediction for the current protein sequence.
2. A dialogue box appears. There are two essential parameters.
   - **Filter Level**. This parameter is used to filter the prediction according to prediction rules. The default level is 1.
   - **Remove singlet** option. This option is checked by default. In this case, if a residue has different structure from its neighbors, it will be changed according to its neighbors' structure.
3. If you are predicting the secondary structure of multiple alignment protein sequences, there are three parameters to set.
   - **Remove poor aligned regions** option. This option is checked by default. In this case, DNAMAN will ignore the poor aligned regions for prediction. Poor region is defined as the two following parameter.
   - **Length** of poor region. The default is 40 residues.
   - **Homology <(%).** The default is 20%. If a region with a defined Length has less than this percentage, DNAMAN will considerate this region as poor aligned region.
4. Click **OK** to display the results in a text window and a graphic window.

The prediction of protein secondary structures is shown as a table in a text window. You may save the table and plot it with other Graphic/Table programs, such as Microsoft Excel. For the amino acid at a given position, DNAMAN displays the probability of forming alpha-helix, beta-sheet, or random coil.

The most possible structure will be assigned to this residue. An overall prediction accuracy will be calculated.

DNAMAN also presents the prediction in a graphic window. The structure probabilities are plotted against the positions. See the section **V.5** for more information about the handling of **Profile Plot**.

XII.10 Prediction of protein transmembrane segments

DNAMAN predicts the transmembrane segments of the current protein sequence using the method described by Persson and Argos (Prediction of transmembrane segments in proteins utilising multiple sequence alignments ***J. Mol. Biol.*** 1994, 237: 182). The method calculates the propensity values of the middle (Pm) and the end (Pe) of sequence segments. DNAMAN plots the average propensity values of each position along the sequence and determines the transmembrane regions. Eight-residue segments are considered as potential cores of transmembrane segments and elongated if their Pm values are above a threshold. Pe values are considered as stop signals.

If a multiple sequence alignment is used, the contribution from each sequence is weighted according to its dissimilarity relative to the other aligned sequences. The results are plotted on a graph and also shown in a text window.

XII.11 Prediction of protein signal peptide

DNAMAN predicts the signal peptide of the current protein sequence using the method described by von Heijne (A new method for predicting signal sequence cleavage sites. ***Nucleic Acids Res.***: 1986, 14:4683). The method uses weight matrices to locate signals in proteins sequences and determines signal peptides based on the scores. DNAMAN reports all possible sites when their scores are higher than the threshold. The weight matrices are different between prokaryote and eukaryote and users must check the option when a sequence is from Prokaryote organisms.

Chapter XIII Oligo Database

Many laboratories synthesize and store a large number of oligoes for the use of DNA sequencing, blotting, PCR, *etc*. Without proper management, it may be difficult to handle the information of these oligoes. DNAMAN provides an oligo database manager that can help you effectively organize and use these oligoes.

XIII.1 Managing oligo databases

Choosing the ***Primer | Oligo Database | Manager*** command opens a **Database Manager** dialogue box where all the oligo databases are listed.

Clicking any database name displays the following information of the selected database on the top of the dialogue box:

- File name,
- Number of records,
- Type of database,
- Date of last update

You may create a new oligo database, delete an existing database or change the password of the selected database by using corresponding buttons in the dialogue box.

Creating a new oligo database

You may create a new oligo database by giving a file name and password. Clicking the **New** button opens a **New Database** dialogue box. You should type in 8 or fewer characters as a filename. DNAMAN will automatically add the oligo database extension: **odb**.

Password can be used to protect a database. DNAMAN will ask for the password when anyone wants to modify (add or remove records, delete database) the database. However, anyone can freely retrieve information of records in the database without password.

The database is not protected if you do not enter any letter in the password box. In this case, all users can modify the database.

Changing password

DNAMAN allows you to change the password of a database.

1. Choosing ***Primer | Oligo Database | Manager*** opens the dialogue box.
2. Clicking the database selects it as the default database.
3. Clicking the **Password** button opens a **Password** dialogue box.
4. DNAMAN will ask you for the old password, the new password and confirmation of the new password. The new password will be effective only if the old password is confirmed.
5. Click **OK** to validate your change.

The changing password function allows you switch your database between protection and non-protection states.

Show Records

Pressing the Show Records button opens the database records in a list window. The list window shows all information about the records and facilitates record search or identification. The list can be printed out using Printing function of DNAMAN.

Deleting database

To erase an oligo database, you should first select the database as the default database by following the step 1 and 2 described in the **Changing password** section. Click the **Delete** button to delete the database.

For password protected database, DNAMAN will ask you for the password prior to deleting.

XIII.2 Editing records

The basic functions of editing include adding or deleting oligo records and modifying the record information. Choosing the ***Primer | Oligo Database | Edit Record***. command brings up the **Oligo Database** dialogue box.

First, you should select the database you want to edit by clicking the database filename listed in the **Database** box. The information of the number of records and the date of the last update is displayed on the top of the dialogue box.

DNAMAN lists the information of a record to seven fields: **Name, Source, Memo, Length, GC content, Melting temperature** and **Sequence**. If you have a large number of records in a database, you may use the first four fields as the sorting keys to selectively display the records in the **Record List** box.

You may query a list of record by setting the parameter for oligo length. There are three options:

Enter	Click the **Show** button
23	listing the oligoes with a length of 23,
>23	listing the oligoes with a length greater than 23,
<23	listing the oligoes with a length smaller than 23.

Adding oligo record

Clicking the **Add** button in the **Oligo Database** dialogue box adds a "**NEW**" record in the oligo list. All the seven fields are empty in the new record. You should fill out the information of the **Name**, **Source** and **Memo** boxes. The maximum length of each box is 69 characters. The maximum length of an oligonucleotide is 200 bases.

After entering the oligo sequence, click on the **Save** button. The name of the new record appears in the oligo list. According to the oligo sequence, DNAMAN calculates the length, GC content and melting temperature and displays the information in the Length, GC% and Tm boxes. The thermodynamic Tm is calculated using the following formula:

$Tm=dH/(dS-Rln(Cd))-273.15+16.6*(\log_{10}(Cs))$

Where **dH** is the enthalpy, **dS** is the entropy, **R** is 1.987 cal K^{-1} mol^{-1} , **Cd** is the DNA concentration (50 nM is used), and **Cs** is the salt concentration (50 mM).

Deleting oligo record

In the oligo list, click on the record name you want to delete. Press the **Delete** button to remove the oligo record from this database. DNAMAN will ask you for the password, if the database is protected.

Modifying record information

You may edit the information of an oligo record in the database. The fields of "Name", "Source", "Memo", and "Sequence" are editable, but those of "Length", "Tm", and "GC%" are not. After editing, you should click on the **Save** button to save the changed information immediately. If you go to another record of the database without saving, the changed information will be lost. If the sequence is modified, DNAMAN will recalculate and display the length, GC% and Tm of the oligo.

XIII.3 Search against current DNA

Choosing the ***Primer*** | ***Oligo Database*** | ***Search Against Current DNA*** command brings up a dialogue box for selection of an oligo database. This function searches the current default DNA sequence using all records of the selected database.

XIII.4 Search against DNA database

Choosing the ***Primer*** | ***Oligo Database*** | ***Search Against DNA Database*** command brings up a dialogue box for selection of an oligo database. This function searches all records of the default DNA database using all records of the selected database. It may take long time for search large oligo and/or DNA databases.

XIII.5 Importing records from text file

Importing oligo sequences from a text file is a convenient tool to add a large number of oligo records to a database.

You may use a text file produced by the DNAMAN program with the function of exporting sequences (see the next section). You may also choose the ***File | New*** command to open a new file and manually enter the information of oligo records. For a new text document, you must follow the format rules: 1) Each record is separated by "//", 2) Leave a space after the name of each field 3) Other information about the oligo sequence. For example:

//
NAME ExampleA
SOURCE Example 1 seq.
MEMO PCR primer; from 4 to 27; sense; 0.5 mg stored in box1
LENGTH 23
GC% 56.5
TM 57.
SEQ GACTGCCACTTCCTCGATGAAGG
//

The information of each field must be in single line. It is necessary to provide the name and the sequence in a record. The information related to other fields can be ignored. For example:

//
NAME ExampleA
SOURCE Example 1 seq.
SEQ GACTGCCACTTCCTCGATGAAGG
//
NAME ExampleB
SEQ CCACTTTCACAGGCACAGGAGCA
//

Save this file by choosing the ***File | Save as*** command. You may also use any other word processor to create an oligo record text file.

To import the file to database, choose the ***Primer | Oligo Database | Import from Text File*** command to open the **File to ODB** dialogue box.

In the dialogue box, select an oligo database to add records. Then click the **File** button to open the **File Open** dialogue box where select the file containing the oligo records. The password will be asked if the database is protected. Click the **OK** button in the dialogue box to process. DNAMAN displays a message box indicating the number of records imported from the text file.

XIII.6 Exporting records to text file

Use this function to overview or print out the records of a database. You may also save this file and import it to another oligo database.

Choosing the ***Primer | Oligo Database | Export to Text File*** command opens a **Database** dialogue box for record selection.

Select the database you want to export. You may export all records of the database by clicking the **Select All** button. All the selected records will be exported into a text window. You may save or print out the text document.

Example:

```
//
NAME ExampleA
SOURCE Example 1 seq.
MEMO PCR primer; from 4 to 27; sense; 0.5 mg stored in box1
LENGTH 23
GC% 56.5
TM 57.0
SEQ GACTGCCACTTCCTCGATGAAGG
```

Chapter XIV Primer Analysis

DNAMAN provides functions to characterize oligonucleotides, design PCR/Sequencing primers and analyze possible mispriming sites of a primer on the target sequence.

XIV.1 Loading primer into memory

Primer analysis starts with loading an oligonucleotide sequence into memory. You may use the oligo sequence in memory for many functional analyses.

You may enter sequence directly from keyboard or retrieve an oligo sequence from database. For keyboard input, choose the ***Primer | Load Primer | From Input*** command to open the **Input Primer** dialogue box, then type a nucleotide sequence. For retrieve from database, choose the ***Primer | Load Primer | From Oligo Database*** command to open the **Oligo Database** box. Select one of the available databases and double click the record you want to load it into memory.

XIV.2 Self-complementarity

By choosing the ***Primer | Self-complementarity*** command, you can searches for the most possible self-complementarity conformation of an oligonucleotide.

The structure stability is filtrated using the method of minimum free energy and nearest-neighbor developed by Zuker and colleagues (Zuker and Stiegler, 1981, ***Nucleic Acids Res*** 9: 133). Energy parameters have been used from recent development (SantaLucia, 1998, ***Proc. Natl. Acad. Sci. USA*** 95: 1460, Allawi & SantaLucia 1997, ***Biochemistry*** 36: 10581, Allawi & SantaLucia 1998, ***Biochemistry*** 37: 2170). The algorithm uses dynamic programming to find out minimum energy secondary structures of a DNA molecule. A most stable structure will be displayed in a graphic window.

You may modify the graphic content as described in Chapter V.

XIV.3 Melting temperature

Choosing the ***Primer*** **|** ***Melting Temperature*** command opens the **Melting Temperature** dialogue box. The oligo sequence in memory is displayed in the **Oligo sequence** box. DNAMAN calculates the length, GC% and the molecular weight of the primer according to the sequence and displays the results in the dialogue box.

The melting temperature of an oligonucleotide is calculated with three methods:

1) **Thermodynamic Tm**:

   The Tm is calculated using the nearest-neighbor thermodynamic values method (Breslauer et al, 1986, ***Proc. Natl. Acad. Sci.*** 83:3746). This method is less accurate for longer oligos. The formula for the Tm is

   $$Tm=dH/(A+dS-Rln(Cd))-273.15+16.6*(\log_{10}(Cs))$$

   Where **dH** is the enthalpy, **dS** is the entropy, **R** is 1.987 cal K^{-1} mol^{-1}, **A** is a constant for helix initiation, **Cd** is the DNA concentration, and **Cs** is the salt concentration.

2) **Hybridization Tm:**

   This method is generally used for DNA or RNA hybridization, especially in presence of high salt and formamide. It is more accurate with longer oligos. The Tm is calculated using the following formula:

   for DNA:DNA hybridization:

   $$Tm=81.5+16.6*(\log_{10}[Na^+])+0.41*[\%(G+C)]-0.63*(\%Formamide)-500/L-1.5(\%Mismatch)$$

   for DNA:RNA hybridization:

   $$Tm=79.8+18.5*(\log_{10}[Na^+])+0.58*[\%(G+C)]+11.8*[\%(G+C)]^2-0.5*(\%Formamide)-820/L-1.5(\%Mismatch);$$

   for RNA:RNA hybridization:

   $$Tm=79.8+18.5*(\log_{10}[Na^+])+0.58*[\%(G+C)]+11.8*[\%(G+C)]^2-0.35*(\%Formamide)-820/L-1.5(\%Mismatch);$$

   * where L is the length (bases) of the oligonucleotide.

3) **GC+AT Tm:** the estimated Tm is the sum of the contribution of each base: 2°C for A and T and 4°C for G and C.

There are several parameters involved in Tm calculation.

DNA concentration is used only in thermodynamic Tm. You may increase oligo DNA concentration in order to increase Tm.

Salt concentration affects thermodynamic Tm and hybridization Tm. In the **[Na+][mM]** box, you may type a suitable salt concentration of your experiments. The value must be an integer greater than 0.

Hybridization type is required for Hybridization Tm. In the **DNA/RNA** box, enter the hybridization type:

D:D for DNA and DNA hybridization
D:R for DNA and RNA hybridization
R:R for RNA and RNA hybridization.

Formamide concentration affects only the hybridization Tm. You may enter an integer in the "**Formamide (%)**" box.

Mismatch number also affects only the hybridization Tm. You may type in the "**Mismatch**" box a mismatch number of base pairs between the oligo and DNA.

With all parameters entered, click the **Show Tm** button to display the melting temperatures. You may change the sequence in the **Oligo sequence** box and then click the **Show Tm** button to recalculate the length, GC%, MW and Tm of the oligo.

Click the **Report** button to display the analysis results in a text window. You may save it as a text document.

Example:

```
Oligo:
5'-GACTGCCACTTCCTCGATGAAGG-3'
```

```
SEQ  Primer 1: 23 bp;
Composition: 5   A; 7   C; 6   G; 5   T; 0  OTHER
Percentage:  21% A; 30% C; 26% G; 21% T; 0% OTHER

Molecular Weight (kDa): ssDNA: 7.06

Hybridization type:  DNA:DNA
Salt concentration: 50 mM
Formamide: 0 percent
Mismatch: 0 bp
Thermo Tm = 60.6 °C %GC Tm = 57.0 °C  GC+AT Tm = 72.0
°C
```

XIV.4 Complementarity with DNA

DNAMAN searches for the complementary sequences between the oligonucleotide and both strands of the current sequence. The two longest complementary sequences will be shown in a text window, if the paired bases in continuous are equal to or greater than 10 bp.

```
Primer-EXAMPLE1(1-886) complementarity.

First complementarity in continuous: 21 bp
5'-TGGGGACCAACGTGTGATGGC-3' Primer
   ||||||||||||||||||||
3'-ACCCCTGGTTGCACACTACCG-5' (376) Strand -

No second possible complementarity
```

XIV.5 Two primer complementarity

You may examine the complementarity of two oligo sequences with DNAMAN. To perform this function, the second primer should be loaded into the DNAMAN memory by choosing the ***Primer | Two Primer Complementarity*** command.

DNAMAN searches for complementary sequences between two primers. The results indicate the continuous and discontinuous complementary sequences between the two primers.

```
Two primers complementarity.
Max complementarity in continuous: 4 bp, free
energy=-3.40 Kcal/mol
5'-TGGGGACCAACGTGTGATGGC-3'
```

```
                    ||||
     3'-CTTAATTCAATTACCAGGGGGT-5'

Max complementarity in discontinuous: 7 bp
5'-TGGGGACCAACGTGTGATGGC-3'
       ||   | ||     | |
    3'-CTTAATTCAATTACCAGGGGGT-5'
```

XIV.6 PCR primer design

You may use DNAMAN to select PCR primers that satisfy your requirements to amplify template DNA. Many parameters are designed to filtrate PCR primers, such as the range of PCR product size, primer lengths, Tm, GC compositions and salt concentration. Other parameters are used for rejection of "low quality" primers.

Choose the ***Primer | Design PCR Primers*** command to open the **Design PCR Primer** dialogue box. There are three steps to setup parameters to design primers.

Step 1: Primer filtration
Primers are selected according to the physical properties.

- Range of PCR product size. The default range is from 400 to 600 bp.
- Sequence region for the sense primer
- Sequence region for the antisense primer
- Range of primer length. The default range is from 18 to 21bases.
- Range of primer Tm. The default range is from 62 to 65 °C.
- Range of GC content. The default range is from 40% to 60 %.
- DNA primer concentration. This parameter is used to calculate Tm. The default value is 50 nM.
- Salt concentration. This parameter is used to calculate Tm. The default value is 50 mM.

- Check the **Shortest primers only** option if too many primers are found after selection.

Parameters to disqualify primers:

- 3’ end dimer. A single primer may form dimers if the nucleotides at the 3' end are self-complementary. A high quality primer should have least length of the 3' end dimer structure. The default length is 3 bases.
- PolyN. PolyN may hybridize to non-specific regions of genomic DNA. High quality primer should have minimum consecutive identical nucleotides. The default number is 3 base.
- Hairpin structure. Hairpin structures reduce the potential hybridization of primer to target DNA. This parameter defines the maximum nucleotides to form a hairpin structure. The default number is 3 base pairs.
- 3’ unique bases. The longer the 3' unique bases are, the lower the mispriming possibility will be. This parameter defines the minimum length of nucleotides at the 3' end complementary with DNA. The default length is 6 bases.

If the PCR product will be used as hybridization probe of Southern blot, check the **Product for hybridization** option. There are four parameters to determine the thermodynamic properties of the PCR product.

- Tm range. This parameter defines the hybridization Tm of the PCR product in Southern blot. The default values are from 70 to 90°C.
- Range of GC content. The default range is from 40% to 70 %.
- Salt concentration. This parameter is used to calculate Tm. The default value is 200 mM.
- PolyN. PolyN in a probe may result in high background in Southern blot. High quality probe should have minimum consecutive identical nucleotides. The default number is 8 base.

Click the **Next** button to start the analysis. DNAMAN will select primers according to the parameters.

Step 2: Refinement and pair selection.
In the second page, all primers meeting the above criteria are displayed in two list boxes**: Sense primers** and **Antisense primers**. If the primers are used for DNA sequencing, it is not necessary to perform the next steps. You may press the **Export List** button to display the primer list in a text window.

To improve the quality of PCR primers, you may disqualify some primer pairs with the following criteria:

- Primer-Primer interaction. High number of complementary bases may result in hybridization of the two primers with them-self. The primer-primer complementary bases should be as few as possible. The default number is 7.
- Primer Tm difference. High difference makes the efficiency of hybridization difficult to control. Amplification is better performed when the difference is lower. The default difference is 2 degrees.
- Mispriming analysis on target sequence. Mispriming may result in amplification of unexpected PCR products. Check this option will increase the primer quality. You can set up a cut-off score for the analysis. Primers with high score of mispriming will be disqualified. See the next section for information on mispriming analysis
- Restriction analysis. Restriction analysis may eliminate some undesirable primers. You have options to keep primers with restriction sites or keep primers without any restriction site.

Click the **Next** button to enter the final step.

Step 3: Final.

All qualified primers are listed in one box. You can set the order of the list according to primer positions, product sizes,

sum of paired Tms, Tm differences or mispriming scores. Clicking any pair of primers shows its characteristics. Press the **Finish** button to report the list of primers in a text window.

XIV.7 Mispriming analysis

Mispriming analysis is designed to find all possible annealing sites of a primer on the default DNA sequence. This function is useful in primer selection for PCR and DNA sequencing. A weight matrix is used to differentiate the importance of primer positions. Since the matches of a primer to target DNA at 3' end is more important than 5' end for PCR amplification, more weight is given to 3' end. In order to increase the specificity of PCR primers, you should always check if there are secondary annealing sites between selected primers and target DNA.

Choose the ***Primer | Load Primer*** command to load an oligo sequence from **Input** or from a **database**. Choose ***the Primer | Mispriming Analysis*** command to set parameters for mispriming analysis. The following parameters are used to find out mispriming sites:

- **Score Matrix** section. 1) Score of Perfect Match. The default is 2. 2) Score of Mismatch. The default is -2. 3) Score of G-T pair. The default is 1.
- **Position Weight Matrix** section. 1) Length from 3' end. The length of position weight matrix beginning from 3' end. The default is 9. 2) Weight from 1 to. Set a maximum weight for the 3' end, the default is 10. In the default case, the first base at 3' end weighs 10. The second base weighs 9… The ninth base weighs 2. All other base weighs 1.
- **Gap penalty.** You may set gap open and extension penalties. The default value of both penalties is -2.
- **Scan for both strands of oligo**. Check this option if you want to perform mispriming analysis for both strands.

- **Cut-off** score can be determined automatically if you check this option. You may also set your own score as well. The default cut-off score is 90.

Press the **OK** button to perform the analysis.

The results are shown in a text window. A list of possible binding sites of the primer is displayed in the order of score. High score sites are shown first and low score last. No site or only the perfect match site will be reported, if no secondary binding site is found. It is highly recommended not to use primers with secondary binding sites for PCR and sequencing works.

Chapter XV DNA and Protein Databases

Database functions of DNAMAN provide users with tools to organize DNA and protein sequences for different subjects. Database is derived from sqlite3 and can handle a large number of records.

XV.1 Database management

Choosing the ***Database*** | ***Manager*** or ***Edit Record*** command to open the **Database** dialogue box. There are many functions in database management: creating a new database, connecting to a database, executing SQL query, and listing all records.

Create a new database

Click the **New** Database button to open a **New Database** dialogue box. You may enter the name of new database.

Connect to a database

Click Connect Database to open the list of databases available in DNAMAN. Select one as the default database. All database related functions in DNAMAN are performed with the default database.

Print records of a database

DNAMAN displays all records of the default database. The list shows all information about the records and facilitates record search or identification. The list can be printed out by right clicking and choosing ***Print record list*** menu. Double-clicking a record on the list displays the sequence record for editing.

Copy records of a database

You may copy record list to clipboard by right clicking and choosing ***Copy record list*** menu. The list can be pasted to other programs such as Excel.

XV.2 Adding record

Choosing the ***Database*** **|** ***Edit Record*** command opens the DNAMAN **Database** dialogue box. There are two ways of adding sequences to the database. Clicking the **Add File** button to add files and the **Add Entry** button for manual enter of sequence.

Databases accept a few sequence formats: GenBank, FASTA, and DNAMAN ("ORIGIN" format). The initial "**ORIGIN**" is required for retrieving a DNA or protein sequence file as DNAMAN format. For the sequence without the initial, you can simply add the "ORIGIN" in the beginning of the sequence, and then save it (see the section **IV.3**). Example:

```
SEQ  EXAMPLE1: 886 bp;
ORIGIN
1    TTTGACTGCC ACTTCCTCGA TGAAGGTTTT ACTGCCAAGG ACATTCTGGA
51   CCAGAAAATT AATGAAGTTT CTTCTTCTGA TGATAAGGAT GCCTTCTATG
101  TGGCAGACCT GGGAGACATT CTAAAGAAAC ATCTGAGGTG GTTAAAAGCT
```

The information relating to this new record is shown in seven fields:

- Sequence Name
- Definition
- Keywords
- Source
- Reference
- Memo
- CDS (coding region)

GenBank files have the ORIGIN initial. If one file contains more than one DNA or protein sequence; all these sequences will be added automatically into the default DNA or protein database as separate records.

XV.3 Deleting record

Choose the ***Database*** **|** ***Edit Record*** command to open a dialogue box. Select the record name you want to delete from the **Sequence List** box, then click the **Delete Entry** button to erase the record and its related information.

XV.4 Editing record information

The information relating to a specific record can be edited. In the **Sequence Database** box, all records are listed in alphabetical or record order. You may change order by clicking list header.

If a large number of sequences are stored in a database, it may be difficult to find out. In this case, you may use the filter function to decrease the listing number. There are eight filter keys: Name, Definition, Keywords, Source, Ref, Annotation, Length and Misc. In any one of these boxes, type either a partial or the full content, then click the **Filter** button to start filtering.

After editing a record, click the **Save** button to save the modification.

XV.5 Scanning sequence similarity

DNAMAN scans all the sequence records in the default database to search for homology sequences of the current sequence. If the default database contains DNA sequences, the default sequence has to be DNA sequence. If the default database contains protein sequences, the default sequence has to be protein sequence. DNAMAN will use fast alignment method to scan the database for sequence similarity. You have the choice of using fast alignment or optimal alignment for final output of the result.

Choose the ***Database | Scan for Sequence Similarity*** command to open a **Pairwise Alignment** dialogue box. The scanning is performed with pairwise alignment. After scanning, qualified records will be aligned to the query sequence. The **Cutoff** score is used to qualify a record as homology sequence. There are two scores of cutoff: one is used during database scanning and the other is used in final alignment. If a record fails any of the two qualifications, it will not be selected.

You can the parameters for alignment; **K-tuple** and **Gap penalty** for fast alignment (see the section **IX.2**), **Gap open** and

Gap extension for optimal alignment (see the section **IX.2**). If you scan for protein sequences, you can select different weight matrices, and have the option to show similar amino acids in the alignment.

The retrieved sequence information is shown in a text window. The homology of two sequences is calculated using the formula described in the section **IX.2**.

XV.6 Searching in database with single sequence

You can search for a nucleotide sequence in a DNA database, and an amino acid sequence in a DNA or protein database.

Nucleotide Sequence

Choose the ***Database | Search for Nucleotide Sequence*** command to open a dialogue box. Searching for a DNA sequence is performed on both strands of all sequence records in the DNA default database. The formats for searching are the same as searching in the current sequence (see the section **VIII.1**).

The results are shown in a text window.

```
Search in Database:
  ACCGGAT

Seq:EXAMPLE3 + strand
No.    Pos     Sequence
#1     457     ccttgACCGGATtgtgg

Seq:PUC18    + strand
No.    Pos     Sequence
#1     2207    tagttACCGGATaaggc

Seq:PUC18    - strand
No.    Pos     Sequence
#1     328     cgcttACCGGATacctg

Seq:PBR322    + strand
No.    Pos     Sequence
#1     2679    cgcttACCGGATacctg

Seq:PBR322    - strand
```

```
No.     Pos      Sequence
#1      1531     tagttACCGGATaaggc
```

Amino acid sequence

Choose the ***Database | Search for Amino Acid Sequence*** command to open a dialogue box. You may search for an amino acid sequence from all protein sequence records in the default protein database. You may also search for an amino acid sequence from six reading frames of all sequence records in the default DNA database.

The formats for searching are the same as the searching in the current sequence (see the section **VIII.1**).

The result is shown in a text window.

Examples:

1) search in DNA database for amino acid sequence

```
Search in Database for AA sequence:
  PLMST
Maximum mismatch: 1

Name:PUC18
Strand plus
Reading frame 3:
255     CCTCTAGAGTCGACC
         P  L  E  S  T
1077    CCAATGATGAGCACT
         P  M  M  S  T

Name:PBR322
Strand minus
Reading frame 2:
401     CCAATGATGAGCACT
         P  M  M  S  T
```

2) search in protein database for amino acid sequence

```
Search in Database for AA sequence:
  PLMS
Maximum mismatch: 1

Name:record8
1634    ldkyfPLDSevtii
```

XV.7 Searching in database with multiple sequences

You may search a database for a set of sequences. Use the ***Database | Search for Nucleotide Sequence Set*** menu for DNA sequence and the ***Database | Search for Protein Sequence*** Set for protein sequences. The formats of query sequence set are described in the section **VIII.2 Searching for a set of sequences.**

Chapter XVI Data Information and Settings

Some of data information used in DNAMAN, such as restriction enzyme databases and amino acid properties, can be accessed and modified by users. When you become an expert of DNAMAN, you can make some modifications to take advantage of the software.

XVI.1Restriction enzymes

DNAMAN allows users to create and edit custom enzyme data files (see the section **XI.2**).

Displaying restriction enzyme information

1. Choose ***Info* | *Restriction Enzymes*** to open the **Restriction Enzymes** dialogue box.
2. Select an enzyme file in the **File** box.
3. In the **Enzyme List** box, clicking an enzyme name displays the related enzyme information:

   - Recognizing sequence and cut position
   - Available from which company
   - Isoschizomers
   - Compatible ends
   - Miscellanies

Editing restriction enzyme information

1. Choose the ***Info* | *Restriction Enzymes*** command to display **Enzyme Information** box.
2. Select an enzyme file in the **File** box.
3. Select the enzyme name you want to edit from **Enzyme List** box.
4. Click the **Edit** button and the **Edit Record** dialogue box appears.
5. Edit related information.
   * There are many supplier names listed in the **Commercial source** box:

6. If you want to add a commercial source that is in the list, click that name.
7. If you want to add a new commercial source that is not in the list, click the **Add Source** button. In the **New Commercial Source** box, type the name you want to provide.
8. Click the **Save** button.

Adding new restriction enzyme

1. Choose the ***Info*** **|** ***Restriction Enzymes*** command to display **Enzyme Information** box.
2. Select an enzyme file in the **File** box.
3. Click the **Add** button to open the **Add Record** dialogue box.
4. Fill in the boxes of the **Name**, **Recognition sequence**, **Cut position**, **Isoschizomer**, **Compatible enzymes** and **Miscellanies**.
5. Select the names of the available commercial source. Click the **Add Source** button to add a new name (see the section **XVI.1**).
6. Click the **Save** button.

Removing a restriction enzyme from the list

1. Choose the ***Info*** **|** ***Restriction Enzymes*** command to display **Enzyme Information** box.
2. Select an enzyme file in the **File** box.
3. Select the enzyme name you want to **remove** from **Enzyme List** box.
3. Click the **Remove** button to delete that enzyme from the list.

XVI.2Enzyme file conversion

You may convert an enzyme file from GCG format to DNAMAN format. This is especially useful when you like to import a REBase file of New England Biolabs to DNAMAN. You may download the REBase file from the Internet using GCG format and use this function to import to DNAMAN.

1. Choose the ***Info*** **|** ***Convert Enzyme File*** command to display the **GCG RE file to DNAMAN** box.
2. Press the **File** button to select a GCG restriction enzyme file in the **File** box.
3. Enter a file name for DNAMAN restriction enzyme file. Give a name only. Do not enter any extension name. DNAMAN will automatically add a required extension name to it.
4. Click the **OK** button.

XVI.3 Methylases

There are two site-specific DNA methylases, *DAM* and *DCM*, in many strains of *E.coli*. Reference: McClelland. and Nelson. 1988. Gene 74:291. Choosing the ***Info*** **|** ***Methylase*** command displays the following table where the restriction enzyme sites are indicated in uppercase letters. The enzymes involved are used to screen methylase information in restriction analysis (see the section **XI.1**)

XVI.4Genetic code table

There are seven genetic code tables used in DNAMAN. Choose the ***Info*** **|** ***Genetic Code*** command to display the universal genetic codes and their variations. Do not use this command to modify the genetic code information. You must access from the ***Protein*** **|** ***Select Genetic Code*** command for the modification of the code information.

XVI.5Amino acid property table

Amino acid property information is used to compute protein pI and charges, and also used to plot hydrophobic/hydrophilic profile. Choose the ***Info*** **|** ***Amino Acid Properties*** command to display the information.

XVI.6Nucleotide table

Choose the ***Info*** **|** ***Nucleotides*** command to display the information about nucleotides.

XVI.7 Setting DNAMAN system configurations

Choose the ***Info*** **|** ***Settings*** command to set up DNAMAN system configurations.

There are four pages to set the configuration parameters of Directory, DNA analysis, Protein analysis and others.

Folder Settings

The folder settings are important in DNAMAN. The program may not functional properly if folders are not correctly set. There are seven folders required in DNAMAN:

1. DNAMAN **system** folder. All DNAMAN system files are stored in this folder. DNAMAN cannot work without this directory.
2. **Database** folder. This is the folder where you can find all the databases.
3. **Multiple alignment** folder. You may use this folder to store your multiple alignment data.
4. **Consensus data** folder. DNAMAN keeps consensus data files in this folder.
5. **Sequence maps** folder. This is the folder that DNAMAN uses for managing sequence maps.
6. **Temporary** folder. This is the folder that DNAMAN creates and stores temporary files.
7. **Default** folder. DNAMAN starts always in this folder.

All these directories must be correctly set up. You may change them by pressing the corresponding button and choosing an appropriate folder.

File Settings

DNAMAN uses some pre-set files for DNA sequence analysis. There are six file names in this page.

1. **Nucleotide property** file. DNAMAN uses this file to display nucleotide properties.

2. **Enzyme** file. This is the current restriction data file used in DNAMAN for restriction analysis.

3. **Annotation** file. This file stores all sequence annotations recognized by DNAMAN. It is highly recommended not to change this file. You may add or remove annotations in the list by pressing the **Edit Annotation List** button. Always avoid removing records. If a record has been used in a sequence database, deleting the annotation record results in false information in the database.

4. **Methylase** file. DNAMAN uses also this file for restriction analysis, when dam and/or dcm methylations are involved.

5. **Search list** file. This file contains your consensus sequence information for sequence search.

6. **BLAST template** file. DNAMAN uses this file as template to generate BLAST document.

You may modify these file names. However, you must be sure that the files exist, and contain correct information for sequence analyses.

Protein Analysis Settings

DNAMAN uses also some pre-set files for protein sequence analysis. There are five parameters in this page.

1. **Genetic code table** file. This file contains all genetic tables. DNAMAN uses this file to find different genetic tables for translation analysis.

2. **Current genetic code** table. This is the current genetic table used in DNAMAN for translation analysis.

3. **Amino acid property** file. DNAMAN uses this file to display amino acid properties.

4. **Similar amino acid** file. This file contains amino acid similarity matrix. DNAMAN uses this matrix to define the similarity between amino acids for alignment of two protein sequences.

   You may generate a new file for amino acid similarity by clicking the **Generate Similarity Matrix** button.

You may modify these file names. However, you must be sure that the files exist and contain correct information for sequence analyses.

Miscellaneous settings

This page contains a few important settings of DNAMAN.

1. **Save channel information before exit**. You may save existing channel information before the termination of current session of DNAMAN. When this option turned on, DNAMAN will keep all channel information for the next session of DNAMAN.
2. **Large child window**. If this option is checked, DNAMAN opens a child window with a lager size, instead of the regular size.
3. **Fast restriction analysis**. Check this option to accelerate restriction analysis when using a large enzyme database file. This option avoids multiple reading of the enzyme file.
4. **Print - Show page number**. Check this option if you like to display the page number when you print out a document.
5. **Print - Show file name path**. Check this option if you like to display the file information when you print out a document.
6. **Color list.** Click each color to change it. You may press the Default Color List button to restore the default used in DNAMAN.
7. **Internet.** BLAST server address can be defined in this section. The server can be at NCBI or your internal one. For PC version only, you may use the integrated Internet browser of DNAMAN by checking the option in this section. If you uncheck the Use internal browser option, DNAMAN will launch your default browser when you want to browser the Internet.

Advanced settings

This page contains advanced settings that are not recommended to modify.

Chapter XVII Questions and Answers

This chapter answers some frequently asked questions about the DNAMAN program.

1. I have installed the DNAMAN program on my computer, but I do not know how to work with it.

A. Read the section **II.3** and **II.4** of this document to understand DNAMAN and get started with the program.

2. I opened a sequence file, but I could not analyze it. All analysis menus are grayed out.

A. You should load the sequence into a channel in order to analyze it. **Chapter III** of this document describes the basics of DNAMAN.

3. What is the largest sequence can DNAMAN handle?

A. In theory, DNAMAN can handle sequences up to 2 giga-bases ($2x10^9$). The limitation is the computer resources. It takes a lot of memory and computing power to handle such a long sequence. We have tested the program with sequences of several mega-bases. The functions of multiple alignment and sequence assembly have different limitations in sequence input.

4. What are the limitations of multiple alignment?

A. You may use up to 32000 sequences with a maximum length of 64000 bases. Again, the limitation is the computer power. It may be really slow with long sequences. We have tested the program with 1000 sequences of approximately 300 amino acids.

5. What are the limitations of sequence assembly?

A. You may use up to 32000 DNA fragments with a maximum length of one giga-bases. Again, the limitation is the computer power. It may be really slow with long sequences. We have tested the program with 10000 sequences of approximately 300 bases.

6. May I use DNAMAN to prepare my posters?

A. Yes, you may. You may prepare DNA maps, cloning strategies and other graphics with DNAMAN. These maps can be enlarged and printed out in several pages.

7. May I use other programs within DNAMAN?

A. Yes, you may. Use the ***File | Run*** menu to launch a program. You have the option to display the results in the DNAMAN studio.

8. What is the DNAMAN browser for? May I use my favorite browser, instead of the integrated one?

A. The DNAMAN browser is designed to browser the databases on the Internet, or your Intranet. You have an option to use your favorite browser by choosing the ***Info | Settings*** menu.

9. What is a multiple threaded function? Which functions of DNAMAN are multiple threaded?

A. A multiple threaded function runs in its own thread. It will not block a program to run other modules. If a function is not multiple threaded, users must wait for the end of the function in order to start a new function. With a multiple threaded function, users may start other functions immediately. Multiple threaded functions utilize better multiple-processor computers. The multiple alignment of DNAMAN is multiple threaded. You may launch up to 16 processes of multiple alignment running in the same time. The sequence assembly function is also multiple threaded. Dot-matrix comparison, PCR primer selection and producing trees from multiple alignment fall also into this category.

Chapter XVIII References

Adachi, J. and Hasegawa, M. (1996). Tempo and mode of synonymous substitutions in mitochondrial DNA of primates. . ***Mol. Biol. Evol***. 13: 200.

Allawi, HT & SantaLucia, JJr (1997). Thermodynamics and NMR of Internal G·T Mismatches in DNA. ***Biochemistry*** 36: 10581.

Allawi, HT & SantaLucia, JJr (1998). Nearest Neighbor Thermodynamic Parameters for internal G·A mismtaches in DNA. ***Biochemistry*** 37, 2170.

Allawi, HT & SantaLucia, JJr (1998). Thermodynamics of internal C·T mismatches in DNA. ***Nucleic Acids Res.*** 26: 2694.

Altschul, S. F., Gish, W., Miller, W., Myers, E.W., and Lipman, D. J. (1990). Basic local alignment search tool. ***J. Mol. Biol.*** 215: 403.

Bairoch, A. and Apweiler, R. (1996). The SWISS-PROT Protein Sequence Data Bank and Its New Supplement. TREMBL. ***Nucleic Acids Res.*** 24***:*** 21.

Bairoch, A., Bucher, P., and Hofmann, K. (1996). The PROSITE Database, Its Status in 1995. ***Nucleic Acids Res.*** 24***:*** 189.

Benson, D.A., Boguski, M.S., Lipman, D.J., and Ostell, J. (1996). GenBank. ***Nucleic Acids Res.*** 24***:*** 1.

Bolton, E.T. and McCarthy, B.J. (1962) A General Method for the Isolation of RNA Complementary to DNA. ***Proc. Natl. Acad. Sci.*** 48:1390.

Breslauer KJ, Frank R, Blocker H, and Marky LA. (1986) Predicting DNA duplex stability from the base sequence. ***Proc. Natl. Acad. Sci.*** 83:3746.

Casey, J. and Davidson, N. (1977) Rates of Formation and Thermal Stabilities of RNA:DNA and DNA:DNA Duplexes at High Concentrations of Formamide. ***Nucleic Acids Res.*** 4:1539.

Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. (1978). A model of evolutionary change in proteins. In ***Atlas of Protein Sequences and Structure*** , (M.O. Dayhoff, ed.), Vol 5, Suppl. 3, pp. 345-352, National Biomedical Research Foundation, Washington, D.C., USA.

Eisenberg, D., Sweet, R.M., and Terwilliger, T.C. (1984). The Hydrophobic Moment Detects Periodicity in Protein Hydrophobicity. ***Proc. Natl. Acad. Sci.*** 81*:* 140.

Feng, D.F. and Doolittle, R.F (1987). Progressive Sequence Alignment as a Prerequisite to Correct Phylogenetic Trees. ***J. Mol. Evol.*** 25*:* 351.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. ***Evolution*** 39: 783

Freier S.M., Kierzek R., Jaeger J.A., Sugimoto N., Caruthers M.H., Nielson T., and Turner D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. ***Proc. Natl. Acad. Sci.*** 83:9373.

Garnier, J., Osguthorpe, D.J., and Robson, B. (1978). Analysis of the Accuracy and Implications of Simple Methods for Predicting the Secondary Structure of Globular Proteins. ***J. Mol. Biol.*** 120: 97.

Gill, S.C. and von Hippel P. H. (1989). Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. ***Analytical Biochemistry.*** 182: 319.

Grantham, R., Gautier, C., Guoy, M., Jacobzone, M., and Mercier R. (1981). Codon Catalog Usage Is a Genome Strategy Modulated for Gene Expressivity. ***Nucleic Acids Res.*** 9: r43.

Gribskov, M. and Burgess, R.R. (1986) Sigma factors from E. coli, B. subtilis, phage SP01, and phage T4 are homologous proteins. ***Nucleic Acids Res.*** 14: 6745.

Gribskov, M., Devereux, J., and Burgess, R.R. (1984). The Codon Preference Plot: Graphic Analysis of Protein Coding Sequences and Prediction of Gene Expression. ***Nucleic Acids Res.*** 12: 539.

Hasegawa M., Kishino H., and Yano T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. ***J. Mol. Evol.***, 22. 160

Henikoff, S. and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. ***Proc. Natl. Acad. Sci.*** 89: 10915.

Higgins, D.G. and Sharp, P.M. (1988). CLUSTAL: A Package for Performing Multiple Sequence Alignment on a Microcomputer. ***Gene*** 73: 237.

Higgins, D.G. and Sharp, P.M (1989). Fast and Sensitive Multiple Sequence Alignments on a Microcomputer. ***Computer Applications in the Biosciences*** 5: 151.

Hopp, T.P. and Woods K.R. (1981) Prediction of Protein Antigenic Determinants from Amino Acid Sequences. ***Proc. Natl. Acad. Sci.*** 78:3824.

Ikuta, S., Takagi, K., Wallace, R.B. and Itakura, K. (1987). Dissociation Kinetics of 19 Base Paired Oligonucleotide-DNA Duplexes Containing Different Single Mismatched Base Pairs. ***Nucleic Acids Res.*** 15:797.

Jones., D.T., Taylor., W.R. and Thornton., J.M. (1992). The rapid generation of mutation data matrices from protein sequences. ***CABIOS*** 8:275

King, R.D. and Sternberg, M.J.E. (1996) Identification and Application of the Concepts Important for Accurate and reliable

Protein Secondary Structure Prediction . ***Protein Science*** 5:2298

Kyte, J. and Doolittle, R.F. (1982). A Simple Method for Displaying the Hydropathic Character of a Protein. ***J. Mol. Biol.*** 157: 105.

Lipman, D.J. and Pearson, W.R. (1985). Rapid and Sensitive Protein Similarity Searches. ***Science*** 227: 1435.

Maizel, J.V. and Lenk, R.P. (1981). Enhanced Graphic Matrix Analysis of Nucleic Acid and Protein Sequences. ***Proc. Natl. Acad. Sci.*** 78: 7665.

Maniatis, T., Fritsch, E.F. and Sambrook. (1988) Molecular Cloning: a Laboratory Manual. Clod Spring Harbor Laboratory. Cold Spring Harbor, N.Y. Page: 9.51 and 11.46.

Mathews, D.H., Sabina, J., Zucker, M., and Turner, H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Provides Robust Prediction of RNA Secondary Structure. ***J. Mol. Biol.*** 288: 911

McClelland, M. and Nelson, M. (1988) The Effect of Site-Specific DNA Methylation on Restriction Endonucleases and DNA Modification Methyltransferases - a View. ***Gene*** 74:291.

Needleman, S.B. and Wunsch, C.D. (1970). A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. ***J. Mol. Biol.*** 48: 443.

Pearson, W.R. and Lipman, D.J. (1988). Improved Tools for Biological Sequence Analysis. ***Proc. Natl. Acad. Sci.*** 85: 2444.

Pearson, W.R. (1990). Rapid and Sensitive Sequence Comparison with FASTP and FASTA. In ***Methods in Enzymology*** , (R.F. Doolittle, ed.), 183, 63-98, Academic Press, San Diego, California, USA.

Pearson, W.R. (1995). Comparison of Methods for Searching Protein Sequence Databases. ***Protein Science*** 4: 1145.

Persson, B. and Argos, P. (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments ***J. Mol. Biol.*** 237: 182.

Roberts, R.J. and Macelis, D. (1996). REBASE - Restriction Enzymes and Methylases. ***Nucleic Acids Res.*** 24: 223.

Rodriguez-Tome, P., Stoehr, P.J., Cameron, G.N., and Flores, T.P. (1996). The European Bioinformatics Institute (EBI) Databases. ***Nucleic Acids Res.*** 24: 6.

Rychlik, W. and Rhoads, R.E. (1990). Optimization of the Annealing Temperature for DNA Amplification in virtro. ***Nucleic Acids Res.*** 18: 6409.

Saitou, N. and Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. ***Mol. Biol. Evol***. 4: 406.

SantaLucia, JJr (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. ***Proc. Natl. Acad. Sci. USA*** 95: 1460.

Schwartz, R.M. and Dayhoff, M.O. (1979). Matrices for Detecting Distant Relationships. In ***Atlas of Protein Sequences and Structure*** , (M.O. Dayhoff, ed.), 5, Suppl. 3, 353-358, National Biomedical Research Foundation, Washington, D.C., USA.

Smith, T.F. and Waterman, M.S. (1981). Comparison of Bio-Sequences. ***Advances in Applied Mathematics*** 2: 482.

Smith, T.F., Waterman, M.S., and Sadler, J.R. (1983). Statistical Characterization of Nucleic Acid Sequence Functional Domains. ***Nucleic Acids Res.*** 11: 2205.

Studier, J.A. and Keppler, K.J. (1988). A Note on the Neighbor-joining Algorithm of Saitou and Nei. ***Mol. Biol. Evol.*** 5: 729.

Sugimoto N, Nakano S, Yoneyama M, and Honda K. (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. ***Nucleic Acids Res.*** 24:4501

Tajima, F. and Nei, M. (1984). Estimation of Evolutionary Distance between Nucleotide Sequences. ***Mol. Biol. Evol.*** 1: 269.

Tamura, K. and M. Nei. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. ***Mol. Biol. Evol.*** 10:512

Thompson, J.D., Higgins, D.G., and Gibson T.J. (1994). CLUSTALW: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting Position-Specific Gab Penalties and Weight Matrix Choice. ***Nucleic Acids Res.*** 22:4673.

von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. ***Nucleic Acids Res.***: 14:4683

Walter, A., Turner, D., Kim, J., Lyttle, M., Müller, P., Mathews, D., and Zuker, M. (1994). Coaxial stacking of helices enhances binding of oligoribonucleotides. ***Proc. Natl. Acad. Sci.*** 91: 9218

Wilbur, W.J. and Lipman, D.J. (1983). Rapid Similarity Searches of Nucleic Acid and Protein Data Banks. ***Proc. Natl. Acad. Sci.*** 80: 726.

Wilbur, W.J. and Lipman, D.J. (1984). The Context Dependent Comparison of Biological Sequences. SIAM ***J. Appl. Math.*** 44: 557.

Zuker, M. and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamic and auxiliary information, ***Nucleic Acids Res*** 9: 133

INDEX

T

U

V

W

Z

User's Guide

LBDraw

for Microsoft Windows

(This software is not available for Mac OSX)

Lynnon Corporation

Lynnon Corporation
383 ave St-Louis
Pointe-Claire, Québec
Canada H9R 2A3

Tel. (514) 447-2514
Fax (514) 447-2516

E-mail: **mail@lynnon.com**

Technical Support: **support@lynnon.com**

Web Site: **http://www.lynnon.com**

Trademarks

CONTENTS

Chapter I Introduction

Welcome to LBDraw.

LBDraw is a simple and powerful drawing program. Working with DNAMAN and LBDraw, you can easily create diagrams for the presentations of molecular biology works, which are usually difficult to build with other drawing programs.

I.1 What LBDraw does

LBDraw is a Windows drawing program. It is designed to draw and assemble drawing objects. If you are familiar with DNAMAN, you have probably drawn some restriction maps. LBDraw can help you to assemble all of these maps, and make comprehensive presentations for your works.

I.2 Understanding Object Linking and Embedding

Object Linking and Embedding, or OLE, is the main propose of LBDraw. OLE has been introduced in Windows since the 3.1 version. It allows users to exchange information in a file that contains elements created in different applications. For a detailed description of OLE, refer to the MS-Windows documentation.

The major advantage of OLE is that you can combine the capacities of two or more applications in one document. The standard *Copy* and *Paste* commands in the Edit menu (without OLE capacities) allow you to exchange elements from one application to another, but they could not be edited after pasting into the application. If you want to make any modification, you have to delete them and then insert revised elements in their place. With an OLE-capable application, you can transfer and share information in a dynamic way. By embedding or linking objects, you can create a document that contains information generated by different applications, and

you edit the information by a direct call to those applications from inside the document.

There are two categories of OLE applications:

1. OLE server: applications whose objects can be embedded or linked into other documents.

2. OLE client: applications that can accept embedded or linked objects.

Some applications, such as LBDraw, may act as both a server and a client. Others act only as one or the other. DNAMAN is an OLE server application.

How LBDraw works as an OLE client:

You can embed and link the objects from other applications into LBDraw document. To edit these objects, you can open these applications from within LBDraw, make your changes, and save the changes. For example, if you have embedded a DNAMAN restriction map (DMP file) into a LBDraw document, and decide to modify a part of the restriction map, you open DNAMAN from within LBDraw and edit the map, and then save the file. The saved file contains all modifications of the restriction map.

I.3 Using Online Help

LBDraw comes with a Online help file which provides instructions of how to perform each of the program's operations. You can access the Online help file in a number of ways. The most convenient one is by pressing F1.

Chapter II Using LBDraw

II.1 Creating a new LBDraw document

When LBDraw starts, a new document is automatically created. You can work on the new document. With the multiple document interface (MDI) of LBDraw, more documents can be opened in the same time. Use the ***File | New*** menu to create another new document.

II.1.1 Setting up drawing environment

From the ***View*** menu or the File/Edit/Object toolbar, you can select parameter to set up your drawing environment.

Zoom

LBDraw shows the actual page size on screen. You can adjust the view size of the page using the **zoom** command on the File/Edit/Object toolbar. The default level of zoom is 50%.

Ruler Lines

LBDraw provides on-screen grid lines as rulers of your drawing window for determine the size and position of objects. When you move mouse on the drawing window, the cursor position is indicated on the left side of the **Status bar**. When you select the ***View | Ruler Lines*** command, the grid lines will appear on the screen. You can remove the grid lines by select another time ***Ruler Lines*** from the ***View*** menu. The position of the mouse cursor is always displayed even though the grid lines disappear.

Paper Color

Paper color lets you match the screen color to the paper you plan to print out. When you choose the ***View |Paper Color*** menu, a color selection dialogue box appears. Choose a color from either the Spot or create a custom color palette. When you click on **OK** button, the screen will display the selected color.

Object Outlines

LBDraw provides an option to outline objects from other applications. Using this option, you can distinguish objects of other applications from native LBDraw. Enable this option by choosing the ***View | Object Outlines*** menu. Repeat the action to disable it.

Tool Bars and Status Bar

By default, LBDraw displays both **Tool Bar** and **Status Bar** on the screen. You can hide them to get more visible area of the screen, but you usually need them for drawing.

Toolbars

Toolbars include buttons for some of the most common commands in LBDraw. There are two tool bars in LBDraw: one shows the commands of ***File***, ***Edit*** and ***Object***, and the other shows the drawing tools. A check mark appears next to the menu items when the toolbars are displayed.

The File/Edit/Object toolbar is usually displayed across the top of the application window, below the menu bar. The drawing tools toolbar is placed at the left side of the application window. The toolbars provide quick mouse access to many tools used in LBDraw,

To hide or display the toolbars, choose them from the View menu.

Click	To
	Open a new document.
	Open an existing document. LBDraw displays the Open dialogue box, in which you can locate and open the desired file.
	Save the active document or template with its current name. If you have not named the document, LBDraw displays the Save As dialogue box.
	Print the active document.
	Remove selected data from the document and stores it on the clipboard.
	Copy the selection to the clipboard.
	Insert the contents of the clipboard at the insertion point.
	Insert a DMP object of DNAMAN.
	Choose Selection tool.
	Choose Text tool.
	Choose Straight Line tool.
	Choose Arrow Line tool.
	Choose Multiple Lines tool.
	Choose Arrow Lines tool.

 Choose Rectangle tool.

 Choose Rounded Rectangle tool.

 Choose Ellipse tool.

 Choose Polygon tool.

 Choose Curve tool.

 Choose DNA tool.

 Choose Needle tool.

 Choose Virus tool.

 Choose Phage tool.

 Choose Flask tool.

 Choose Tube tool.

 Align the selected objects to top.

 Align the selected objects to bottom.

 Align the selected objects to left.

 Align the selected objects to right.

Status Bar

The **Status Bar** describes the action to be executed by the selected menu item or depressed toolbar button, and keyboard

latch state. A check mark appears next to the menu item when the Status Bar is displayed.

The status bar is displayed at the bottom of the LBDraw window. To display or hide the status bar, use the ***Status Bar*** command in the ***View*** menu.

The left area of the status bar describes actions of menu items as you use the arrow keys to navigate through menus. This area similarly shows messages that describe the actions of toolbar buttons as you depress them, before releasing them. If after viewing the description of the toolbar button command you wish not to execute the command, then release the mouse button while the pointer is off the toolbar button.

The right areas of the status bar indicate which of the following keys are latched down:

Indicator Description

CAP The Caps Lock key is latched down.

NUM The Num Lock key is latched down.

SCRL The Scroll Lock key is latched down.

II.2 Using drawing objects

II.2.1 Choosing Drawing Objects

Click on the Tool Bar to select one of the object tools:

II.2.2 Adding Text

1. Click on  to select Text tool.

2. Click the left button of the mouse on the position which you want to draw a text object. A **Text Properties** dialogue box appears.

3. Type the text you want to enter.

4. Press on the **Font** button. Choose Font dialogue box appears.

5. Choose a font, font size, text color and effects for the text, then click OK to back.

6. Press OK button when you finish. The text will appear on the drawing.

7. Click on  to stop using Text Tool.

8. If you want to change the orientation of the text, place the mouse cursor at the top right position of the text. Press and hold down the left button of the mouse. Drag and spin the text to find the desired orientation, then release the mouse left button.

II.2.3 Drawing Straight Line

1. Click on  to select Straight Line tool.

2. Place the cursor at the point which you want the line to start and click once.

3. Move the cursor the point which you want the line to stop and click once.

4. Repeat step 2 and 3 if you want draw another straight line.

5. Click on  to stop using Straight Line Tool.

6. Adjust the line position by moving the its end points.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical straight line.*

II.2.4 Drawing Arrow Line

1. Click on  to select Arrow Lines tool.
2. Place the cursor at the point which you want the line to start and click once.
3. Move the cursor the point which you want the line to stop and click once.
4. Repeat step 2 and 3 if you want draw another arrow line.
5. Click on  to stop using Arrow Line Tool.
6. Adjust the line position by moving the its end points.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical arrow line.*

II.2.5 Drawing Multiple Lines

1. Click on  to select Multiple Line tool.
2. Place the cursor at the point which you want the line to start.
3. Click the left button of the mouse to make the first point of the multiple lines. Make sure that the mouse button is released immediately
4. Move the cursor to a desired position and click once the left button of the mouse to make a straight line.
5. Repeat step 4 for more straight lines.
6. To add the line, double click on the left button of the mouse.

7. Click on  to stop using Multiple Lines Tool.

8. Adjust the line positions by moving the each point.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical straight line.*

II.2.6 Drawing Arrow Lines

1. Click on  to select Arrow Lines tool.

2. Place the cursor at the point which you want the line to start.

3. Click the left button of the mouse to make the first point of the multiple lines. Make sure that the mouse button is released immediately

4. Move the cursor to a desired position and click once the left button of the mouse to make a straight line.

5. Repeat step 4 for more straight lines.

6. To add the line, double click on the left button of the mouse.

7. Click on  to stop using Arrow Lines Tool.

8. Adjust the line positions by moving the each point.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical straight line.*

II.2.7 Drawing Rectangles

1. Click on  to select Rectangle tool.

2. Place the cursor at the point which you want to have a corner. Press and hold down the mouse button.

3. Move the cursor to draw a rectangle. Continue moving until you have a desired rectangle.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another rectangle.

6. Click on  to stop using Rectangle Tool.

7. Adjust the rectangle size and position by moving the its corners or sides.

** Press and hold the Shift key while drawing, you can make a square.*

II.2.8 Drawing Rounded Rectangles

1. Click on  to select Rounded Rectangle tool.

2. Place the cursor at the point which you want to have a corner. Press and hold down the mouse button.

3. Move the cursor to draw a rectangle. Continue moving until you have a desired rectangle.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another rounded rectangle.

6. Click on  to stop using Rounded Rectangle Tool.

7. Adjust the rounded degree by moving the inner point of its top-right corner.

8. Adjust the rectangle size and position by moving the its corners or sides.

** Press and hold the Shift key while drawing, you can make a rounded square.*

II.2.9 Drawing Ellipses

1. Click on  to select Ellipse tool.

2. Draw an ellipse in exactly the same way as you draw a rectangle. The ellipse fits inside the rectangle which you define.

3. Click on  to stop using Ellipse Tool.

4. Adjust the size and shape of the ellipse by moving the corners of the rectangle you have drawn.

** Press and hold the Shift key while drawing, you can make a circle.*

II.2.10 Drawing Polygons

1. Click on  to select Polygon tool.

2. Place the cursor at the point which you want to start drawing.

3. Click the left button of the mouse to make the first point of the polygon. Make sure that the mouse button is released immediately.

4. Move the cursor to a desired position and click once the left button of the mouse to make another point.

5. Repeat step 4 for more points.

6. To add the last point, double click on the left button of the mouse.

7. Click on  to stop using Polygon Tool.

8. Adjust the position of each point by dragging it.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical straight line.*

II.2.11 Drawing Curves

1. Click on  to select Curve tool.

2. Place the cursor at the point which you want to start drawing.

3. Click the left button of the mouse to make the first point of the curve. Make sure that the mouse button is released immediately.

4. Move the cursor to a desired position and click once the left button of the mouse to make a straight line. There are four points on this line to make it as curve.

5. Repeat step 4 for more lines.

6. To add the last line, double click on the left button of the mouse.

7. Click on  to stop using Curve Tool.

8. Adjust the curve shape by dragging the corresponding points.

II.2.12 Drawing DNA molecules

1. Click on  to select DNA tool.

2. Place the cursor at the point which you want to start. Press and hold down the mouse button.

3. Move the cursor to draw a rectangle shape with DNA image. Continue moving until you have a desired size.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another DNA molecule.

6. Click on  to stop using DNA Tool.

7. Adjust the DNA size and position by moving the its corners or sides.

II.2.13 Drawing Needles

1. Click on  to select Needle tool.

2. Place the cursor at the point which you want to start. Press and hold down the mouse button.

3. Move the cursor to draw a needle. Continue moving until you have a desired needle.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another rounded rectangle.

6. Click on  to stop using Needle Tool.

7. Adjust the handle orientation by moving the inner point of its top-right corner.

   If the needle is filled with liquid, you can adjust the liquid level by moving the inner point of its top-right corner.

8. Adjust the needle size and position by moving the its corners or sides.

II.2.14 Drawing Viruses

1. Click on  to select Virus tool.

2. Place the cursor at the point which you want to start. Press and hold down the mouse button.

3. Move the cursor to draw a rectangle shape with a virus image. Continue moving until you have a desired size.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another virus image.

6. Click on  to stop using Virus Tool.

7. Adjust the virus size and position by moving the its corners or sides.

II.2.15 Drawing Phages

1. Click on  to select Phage tool.

2. Place the cursor at the point which you want to start. Press and hold down the mouse button.

3. Move the cursor to draw a rectangle shape with a phage image. Continue moving until you have a desired size.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another phage image.

6. Click on  to stop using Phage Tool.

7. Adjust the phage size and position by moving the its corners or sides.

II.2.16 Drawing Flasks

1. Click on  to select Flask tool.

2. Place the cursor at the point which you want to start. Press and hold down the mouse button.

3. Move the cursor to draw a flask. Continue moving until you have a desired flask.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another flask.

6. Click on  to stop using Flask Tool.

7. Adjust the flask orientation by moving the inner point of its top-right corner.

   If the flask is filled with liquid, you can adjust the liquid level by moving the inner point of its top-right corner.

8. Adjust the flask size and position by moving the its corners or sides.

II.2.17 Drawing Tubes

1. Click on  to select Tube tool.

2. Place the cursor at the point which you want to start. Press and hold down the mouse button.

3. Move the cursor to draw a tube. Continue moving until you have a desired tube.

4. Release the mouse button.

5. Repeat step 2 to 4 if you want draw another tube.

6. Click on  to stop using Tube Tool.

7. Adjust the tube orientation by moving the inner point of its top-right corner.

   If the tube is filled with liquid, you can adjust the liquid level by moving the inner point of its top-right corner.

8. Adjust the tube size and position by moving the its corners or sides.

II.3 Moving drawing objects

II.3.1 Moving to front and back

1. Select the object which you want to move. You can select only one object for moving.

2. Choose the ***Object*** **|** ***Move to Front*** or ***Object*** **|** ***Move to Back*** menu to move the object to the front or the back of the drawing.

II.3.2 Moving forward and backward

1. Select the object which you want to move. You can select only one object for moving.

2. Choose the ***Object*** **|** ***Move Forward*** or ***Object*** **|** ***Move Back*** menu to move the object one step forward or backward.

II.4 Aligning objects

1. Select the objects which you want to move. You have to select more than one object for aligning.

2. Choose the ***Object*** **|** ***Align Left*** () **or** ***Object*** **|** ***Align Right*** () **or** ***Object*** **|** ***Align Top*** () **or** ***Object*** **|** ***Align Bottom*** () menu to align all the selected objects towards left, right, top or bottom, respectively.

II.5 Saving the LBDraw file

If you have already saved your LBDraw file and want to save additional changes you have made since, select ***File*** **|** ***Save*** menu. The file will be updates with the changes.

If you have not yet saved your file, a Save File dialogue box will appear. Fill in a desired file name for the document and click on OK button to save the current file.

II.6 Exiting LBDraw

Choose the ***File | Exit*** menu to end the current LBDraw session. If you have made any change since last saved file, a message box will appear to ask you if you want to save the modification. If you answer Yes, the current file will be overwritten. If you like to save the current file under another name, you have to choose Cancel, then use the ***File | Save*** as menu. Choose No will give up all changes you have made since.

Chapter III LBDraw Menus

This chapter will discuss the menus of LBDraw

III.1 File menu commands

The File menu offers the following commands:

New	Creates a new document.
Open	Opens an existing document.
Close	Closes an opened document.
Save	Saves an opened document using the same file name.
Save	Saves an opened document to a specified file name.
Print	Prints a document.
Print Preview	Displays the document on the screen as it would appear printed.
Print Setup	Selects a printer and printer connection.
Exit	Exits LBDraw.

III.2 Edit menu commands

The Edit menu offers the following commands:

Undo	Reverse previous editing operation.
Redo	Reverse previous undo operation.
Cut	Deletes selected object data from the document and moves it to the clipboard.
Copy	Copies selected object data from the document to the clipboard.
Copy Image	Copies entire drawing objects from the document to the clipboard.
Paste	Pastes data from the clipboard into the document.
Paste Link	Pastes from the clipboard a link to data in another application.
Delete	Deletes selected object data from the document
Select All	Selects all object data of the document
Properties	Displays the properties of selected object
Insert DNAMAN Object	Inserts and embeds a DMP object of DNAMAN into the document.
Insert New Object	Inserts and embeds an object, such as a chart or an equation in a document.
Links	List and edit links to embedded documents.

III.3 Draw menu commands

The Draw menu offers the following commands:

Select	Chooses object selection tool.
Text	Chooses text insertion tool.
Straight Line	Chooses straight line drawing tool.
Arrow Line	Chooses arrow line drawing tool.
Rectangle	Chooses rectangle drawing tool.
Rounded Rectangle	Chooses rounded rectangle drawing tool.
Ellipse	Chooses ellipse drawing tool.
Polygon	Chooses polygon drawing tool.

III.4 Object menu commands

The Object menu offers the following commands:

Move to Front	Moves the selected object to the front of drawing.
Move to Back	Moves the selected object to the back of drawing.
Move Forward	Moves the selected object to the next position in z-order of drawing.
Move Back	Moves the selected object to the previous position in z-order of drawing.
Align Left	Aligns all selected objects to left.
Align Right	Aligns all selected objects to right.
Align Top	Aligns all selected objects to top.
Align Bottom	Aligns all selected objects to bottom.

III.5 View menu commands

The View menu offers the following commands:

Ruler Lines	Shows or hides ruler lines on screen.
Object Outlines	Toggle OLE object indication.
Toolbar-File/Edit/Object	Shows or hides the File/Edit/Object toolbar.
Toolbar- Drawing	Shows or hides the drawing toolbar.
Status Bar	Shows or hides the status bar.

III.6 Options menu commands

The Options menu offers the following commands:

Background Color	Changes the background color.
Auto Page Number	Updates the page number automatically
Reverse Text Angle in Printing	Reverses the text angle when printing out the document

III.7 Window menu commands

The Window menu offers the following commands, which enable you to arrange multiple views of multiple documents in the application window:

New Window	Creates a new window that views the same document.
Cascade	Arranges windows in an overlapped fashion.

Tile Horizontally	Arranges windows in non-overlapped horizontal tiles.
Tile Vertically	Arranges windows in non-overlapped vertical tiles.
Arrange Icons	Arranges icons of closed windows.
Split	Split the active window into panes.
Window 1, 2, ...	Goes to specified window.

III.8 Help menu commands

The Help menu offers the following commands, which provide you assistance with this application:

Help Topics	Offers you an index to topics on which you can get help.
About	Displays the version number of this application.

Chapter IV How to ...

This chapter will discuss briefly on how to use LBDraw functions.

IV.1 Create a drawing

Use the ***File | New*** menu to create a new drawing document.

When you start LBDraw, a new document is created automatically.

You should set the environment of the drawing

How to set environment of a drawing

When a new drawing document is created, some parameters should be set up.

1. Summary information about the document
2. Zoom size
3. Ruler
4. Background color
5. Object outlines
6. Tool bar
7. Status bar

How to use summary information of a drawing document

The Summary Information can help you to know what kind of the document is. LBDraw recommend you to fill the information for all your drawing documents.

Use the ***File*** **|** ***Summary Info*** menu to open document properties dialogue box.

The first page of the properties should be filled by yourself.

The fields for the information are:

1. Author: Type your name.

2. Keywords: Enter some important words concerning the drawing.

3. Comments: If you have something more to tell, do not forget.

4. Title: Use the document name here

5. Subject: It is always useful to explain the Title.

6. Template: If the drawing is derived from another document, it will help you to remember.

The second page is the statistics of the document.

LBDraw will keep updating the following information of the document:

1. Created: When the document was created.

2. Last saved by: If someone else modifies the drawing, ask her/him to leave her/his name.

3. Last saved: The last time when the document was saved.

4. Revision number: LBDraw records the revision number when modification occurs.

5. Total editing time: LBDraw remembers the spent time on the document.

6. Last printed: LBDraw tells you when the document was printed the last time.

How to use the ruler

LBDraw provides on-screen grid lines as rulers of your drawing window for determine the size and position of objects. When you move mouse on the drawing window, the cursor position is indicated on the left side of the Status bar.

When you select ***Ruler Lines*** from the ***View*** menu, the grid lines will appear on the screen. You can remove the grid lines by select another time **Ruler Lines** from the ***View*** menu. The position of the mouse cursor is always displayed even though the grid lines disappear.

How to select background color

When you start LBDraw, the background color is white. You can change the background from white to other colors by choosing the ***Background Color*** command from the ***View*** menu. You can change it at anytime while drawing.

To select a different background color

1. choose the ***View*** | ***Background Color*** menu to open **Choose Color dialogue box**.

2. Select the color.

3. Click on **OK** button.

How to use Object Outlines

LBDraw provides an option to outline an OLE object. Using this option, you can distinguish objects between OLE drawing and native LBDraw.

Enable this option by choosing the ***View | Object Outlines*** menu. Disable it by choosing another time the same menu.

IV.2 Manage LBDraw files

LBDraw is a multiple document interface(MDI) application. You can work with many drawing document in the same time.

Please refer to following chapters for:

How to

Open and Save a LBDraw File

Embed a Drawing

Edit an Embedded Drawing

Link a Drawing

Edit a Linked Drawing

How to open and save LBDraw file

To open a LBDraw file

1. From the File menu, choose Open. The Open dialogue box appears.
2. If the file you want to use is on a different drive, select the drive you want from the Drives box.
3. In the Directories box, choose the directory you want to open.
4. From the list of files, select the file you want to open.

Some applications provide a check box to specify that the file be read-only; that is, changes cannot be made to the file. If you want the file to be read-only, select this check box.

5. Double-click the filename, or choose the OK button.

To save an existing LBDraw file

From the File menu, choose Save.

To save a new file or an existing file under a new name

1. From the File menu, choose Save As. The Save As dialogue box appears.

2. If you want to save the file on a different drive, select the drive you want from the Drives box.

3. In the Directories box, choose the directory in which you want to save the file.

4. In the File Name box, type a name for the file.

If you don't specify a filename extension, LBDraw will add one. The default extension (.MMP) appears in the Save File As Type box. To use a different extension, select the extension from the Save File As Type list, or type an extension.

5. Choose the OK button.

How to embed a drawing

There are two ways to embed a drawing. You can start from LBDraw, or you can start from the document where you want the drawing to appear. Check the documentation of the other applications you are using for details on how to embed a drawing starting from within them.

To embed a drawing starting from LBDraw

1. Open LBDraw.
2. Create a drawing, or open a document that contains the drawing you want to embed.
3. If you want to save the drawing, choose Save from the File menu.
4. Select the drawing.
5. From the Edit menu, choose Copy.

   A copy of the drawing is placed onto the Clipboard.
6. Open the application into which you want to embed the drawing.
7. Place the insertion point in the drawing area at the location you want the drawing to appear.
8. From the application's ***Edit*** menu, choose Paste.

How to edit an embedded drawing

You can edit an embedded drawing in LBDraw using its application (must be an OLE server).

To edit an embedded drawing

1. Double-click the embedded drawing in the destination document.

   LBDraw opens, displaying the drawing.
2. Edit the drawing as needed.
3. Click outside of the application area.

   The embedded drawing in the document will be updated to reflect your changes.

How to link a drawing

You have to save a drawing before you can link it to another document. When you make changes to the drawing, the changes appear in every document that contains a link to the drawing.

To link a drawing starting from LBDraw.

1. Open LBDraw.
2. Create and save a drawing, or open the LBDraw drawing you want to link.
3. Select the drawing.
4. From the Edit menu, choose Copy.

   A copy of the drawing is placed onto the Clipboard.
5. Open the application into which you want to link the drawing.
6. Place the insertion point in the drawing area at the location you want the drawing to appear.
7. From the Edit menu, choose Paste Link.
8. Save the application with the drawing you intend to link.

You now have the original LBDraw document containing the drawing, plus a link to that document in the application in which you are working.

How to edit a linked drawing

You can edit a linked drawing in LBDraw using its application (must be an OLE server).

Any changes you make to a linked drawing appear in each document that contains the drawing.

To edit a linked drawing

1. In the destination document, double-click the linked drawing.

   LBDraw opens, displaying the drawing.

2. Edit the drawing as necessary.

   The linked drawing is updated as you make your changes. If the Update option in the Links dialogue box from the Edit menu is set to manual, you must update the link by selecting Update Now.

3. From the File menu in LBDraw, choose Save and then Exit.

IV.3 Work with editing tools

To start edit any object, you should select it first. LBDraw provides standard selection and editing tools for graphic editing.

1. Select/Deselect Object(s)

2. Copy, Cut, and Paste Object(s)

3. Link and Embed Object(s)

How to select/deselect object(s)

There are different kinds of selections you can have.

To select one object:

1. Choose the ***Draw | Select*** menu, or click on the ***Selection*** toolbar.

2. Click the left button of the mouse on the desired object.

To select all objects of the drawing:

Choose the ***Edit | Select All*** menu allows you to select all objects of the drawing.

To select a group of objects:

1. Choose the ***Draw | Select*** menu, or click on the ***Selection*** toolbar.

2. Draw a rectangle with your mouse. All objects overlapped with the rectangle will be selected.

To add another object to the selected group:

1. Choose the ***Draw | Select*** menu, or click on the ***Selection*** toolbar.

2. Press and hold down the Shift key.

3. Click the left button of the mouse on the desired object.

You can deselect the selected object(s) by clicking the left mouse button on the surface outside the selected object(s).

You can also deselect any one of the selected objects:

1. Choose the ***Draw | Select*** menu, or click on the ***Selection*** toolbar.

2. Press and hold down the Shift key.

3. Double click the left button of the mouse on the desired object.

How to copy, cut, and paste

You can use the Copy, Cut, and Paste commands on the Edit menu to transfer object(s) to and from the Clipboard. You can also copy and paste an entire drawing of LBDraw document to other applications.

To copy or move object(s), and then paste them into another LBDraw document

1. Select desired object(s) using selection tools.

2. To copy the object(s), choose Copy from the Edit menu or press CTRL+C. Or, to move the object(s), choose Cut from the Edit menu or press CTRL+X.

3. Open the LBDraw document into which you want to paste the object(s).

4. From the Edit menu, choose Paste or press CTRL+V.

The cutout appears on the drawing. You can now move the object(s) where you want them to appear.

5. Move the cursor inside the object(s).

6. Drag the object(s) to where you want them.

7. Click outside the object(s) to paste it, or select any tool.

To copy object(s) to another location in the same file

1. Select desired object(s) using selection tools.

2. Move the cursor inside of the object(s).

3. Press and hold down CTRL, and then click on the object(s).

4. Release CTRL.

5. Use mouse to drag the object(s) away from the original location. When the object(s) are where you want them, release the mouse button.

6. Click outside the object(s) to paste it in their new position, or select any tool.

You may not be able to copy any object from a part of drawing document to other applications, but you can copy entire drawing to other applications.

1. Choose the ***Edit*** **|** ***Copy Graph*** menu to copy entire drawing to the Clipboard.

2. Paste into the other application

How to link and embed

The method you use to insert a LBDraw drawing into another file determines how you make changes to it. You can link and embed information only if the application you are transferring the drawing to supports object linking and embedding.

Method	**Description**
Copy	This method copies a drawing to a document in a different application. Once the drawing is in the new document, you can change it only by deleting it and copying a new drawing in its place. Use the Copy command on the Edit menu to copy the drawing onto the Clipboard.
Embed	You can edit embedded drawings from within the other application's document. Use the Copy command on the Edit menu to copy the drawing onto the Clipboard.
Link	Linking dynamically connects two files so that information in one file automatically updates when information in the other file changes. To link an object, use the Copy command on the Edit menu to copy the drawing onto the Clipboard.

How to insert DNAMAN object

You can copy a DMP object from DNAMAN Restriction Map, then paste it into a LBDraw document. You can also create an DMP object within LBDraw.

To create a DNAMAN DMP object in LBDraw:

1. Choose the ***Edit*** **|** ***Insert DNAMAN*** menu, or click on the ***DNAMAN*** toolbar.

   A DNAMAN working area appears on the window.

2. Double click on the DNAMAN working area. A dialogue box appears to ask you the essential information to draw a DNA map.

3. Fill in the information, then you can start to edit the DMP object.

IV.4 Work with text tool

You can draw a text on the drawing window by using Text tool.

This section tells you how to:

1. Enter and Edit Text.

2. Change Text Orientation.

How to enter and edit text

To enter text:

1. Choose the ***Draw*** **|** ***Text*** menu, or click on the ***Text*** toolbar.

2. Click the left button of the mouse on the place where you want to draw a text object. A **Text Properties** dialogue box appears.

3. Type the text you want to enter.

4. Press on the **Font** button. Choose Font dialogue box appears.

5. Choose a font, font size, text color and effects for the text, then click OK to back.

6. Press OK button when you finish. The text will appear on the drawing.

To edit the text:

1. Choose the ***Draw*** **|** ***Select*** menu, or click on the **Selection** toolbar.

2. Double click the left button of the mouse. The **Text Properties** dialogue box appears. You can edit the text and change its properties.

How to change text orientation

LBDraw provides two methods to change the text orientation.

If you know the exact orientation angle of the text, in degrees:

1. Choose the ***Draw*** **|** ***Select*** menu, or click on the ***Selection*** toolbar.

2. Double click the left button of the mouse. The **Text Properties** dialogue box appears.

3. In the Orientation section, enter the angle of the text in degree, relative to the bottom of the page.

4. Press OK button when you finish typing.

If you do not know the exact orientation angle of the text:

1. Choose the ***Draw*** **|** ***Select*** menu, or click on the ***Selection*** toolbar.

2. Place the mouse cursor at the top right position of the text. The cursor will change shape to a North-East arrow.

3. Press and hold the left button of the mouse. Drag and spin the text to find the desired orientation.

4. Release the mouse left button.

IV.5 Work with Straight Line and Arrow Line tools

LBDraw provides Straight Line and Arrow Line tools. The two tools are essentially the same. The Arrow Line will draw an arrow at the end of the straight line.

To draw a Straight Line:

1. Select the Line or Arrow Line tool.

2. Move the cursor into the drawing area.

3. Press the mouse button to anchor one end of the line, and then drag the cursor. A flexible line stretches from the anchor point to the mouse position.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical straight line.*

4. When you are satisfied with the line, release the mouse button.

5. Double click the mouse button to edit the properties of the line. A Line Properties dialogue box appears. You can change the size and color of the line.

IV.6 Work with Multiple Lines and Arrow Lines tools

LBDraw provides Multiple Lines and Arrow Lines tools. The two tools are essentially the same. The Arrow Lines will draw an arrow at the end of the multiple lines.

To draw multiple lines and arrow lines:

1. Select the Multiple Lines or Arrow Lines tool.

2. Move the cursor into the drawing area.

3. Click the left button of the mouse to start drawing the object. Make sure that the mouse button is released.

4. Move the cursor to a desired position and click once the left button of the mouse to make a straight line.

5. Repeat step 4 for more straight lines.

6. To add the last line, double click on the left button of the mouse.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical straight line.*

7. You can adjust the position of any point by dragging and dropping it.

8. Double click the lines to edit its properties. You can change the size and color of each line.

IV.7 Work with Rectangle and Rounded Rectangle tools

LBDraw provides tools to draw Rectangles, Rounded Rectangles.

Rectangle is a box with four straight corners and Rounded Rectangle is a box with four rounded corners.

To draw a rectangle:

1. Select the Rectangle or the Rounded Rectangle tool.

2. Move the cursor into the drawing area.

3. Press the mouse button to anchor one corner of a flexible box. Drag the cursor. The flexible box stretches from the anchor point to the mouse position.

** Press and hold the Shift key while drawing, you can make a square.*

4. When you are satisfied with the box size, release the mouse button.

5. Change the rectangle size by moving any side line.

You can change the shape of a rounded rectangle by moving the top right corner of the box.

6. Double click the mouse button to edit the properties of the box. A Rectangle Properties dialogue box appears. You can change the size and color of each line. You can also fill in the rectangle with desired color.

IV.8 Work with Ellipse tool

LBDraw provides tools to draw circles and ellipses.

To draw a circle or ellipse

1. Select the Ellipse tool.

2. Move the cursor into the drawing area.

3. Press the mouse button to anchor one point of a flexible ellipse. Drag the cursor. The flexible ellipse stretches from the anchor point to the mouse position.

** Press and hold the Shift key while drawing, you can make a circle.*

4. When you are satisfied with the circle or ellipse, release the mouse button.

5. Double click the mouse button to edit the properties of the box. A Ellipse Properties dialogue box appears. You can change the size and color of the line. You can also fill in the ellipse with desired color.

IV.9 Work with Polygon tool

LBDraw provides tools to draw polygons.

To draw polygons

1. Select the Polygon tool.

2. Move the cursor into the drawing area.

3. Click the left button of the mouse to start drawing the polygon. Make sure that the mouse button is released.

4. Move the cursor to a desired position and click once the left button of the mouse to make one point.

5. Repeat step 4 for more points.

6. To add the last point, double click on the left button of the mouse.

** Press and hold the Shift key while drawing, you can make a horizontal or vertical straight line.*

7. You can adjust the position of any point by dragging and dropping it.

8. Double click the polygon to edit its properties. You can change the size and color of each line. You can also fill in the ellipse with desired color.

IV.10 Work with Curve tool

LBDraw provides tools to draw curves.

To draw curves

1. Select the Curve tool.

2. Move the cursor into the drawing area.

3. Click the left button of the mouse to start drawing the curve. Make sure that the mouse button is released.

4. Move the cursor to a desired position and click once the left button of the mouse to make a straight line. There are four points on this line to make it as curve.

5. Repeat step 4 for more lines.

6. To add the last line, double click on the left button of the mouse.

7. You can adjust the curve shape by dragging and dropping corresponding points.

8. Double click the curve to edit its properties. You can change the size and color of the curve.

IV.11 Work with DNA tool

LBDraw provides tools to draw DNA molecules.

To draw a DNA molecule:

1. Select the DNA tool.

2. Move the cursor into the drawing area.

3. Press the mouse button to anchor one point of a flexible DNA molecule. Drag the cursor. The flexible DNA molecule stretches from the anchor point to the mouse position.

4. When you are satisfied with the DNA molecule, release the mouse button.

5. Double click the mouse button to edit the properties of DNA molecule. A Properties dialogue box appears. You can change the size and color of the molecule.

IV.12Work with Needle, Flask and Tube tools

LBDraw provides tools to draw needles, flask and tube. These tools are essentially the same. These objects are different in shape.

To draw an object:

1. Select the Needle, Flask or Tube tool.

2. Move the cursor into the drawing area.

3. Press the mouse button to anchor one corner of a flexible object shape. Drag the cursor. The flexible object stretches from the anchor point to the mouse position.

4. When you are satisfied with the object, release the mouse button.

5. Change the object size by moving any side line.

6. Change the orientation of the object by moving the top right corner of the box.

7. If the object is filled with liquid (color), adjust the liquid level by moving the top right corner of the box.

8. Double click the mouse button to edit the properties of the box. A Properties dialogue box appears. You can change the size and color of drawing lines. You can also fill in the object with desired color.

IV.13 Work with Virus and Phage tools

LBDraw provides tools to draw virus and phage. These tools are essentially the same. These objects are different in shape.

To draw an object:

1. Select the Virus or Phage tool.

2. Move the cursor into the drawing area.

3. Press the mouse button to anchor one corner of a flexible object shape. Drag the cursor. The flexible object stretches from the anchor point to the mouse position.

4. When you are satisfied with the object, release the mouse button.

5. Change the object size by moving any side line.

6. Double click the mouse button to edit the properties of the box. A Properties dialogue box appears. You can change the size and color of drawing lines. You can also fill in the object with desired color.