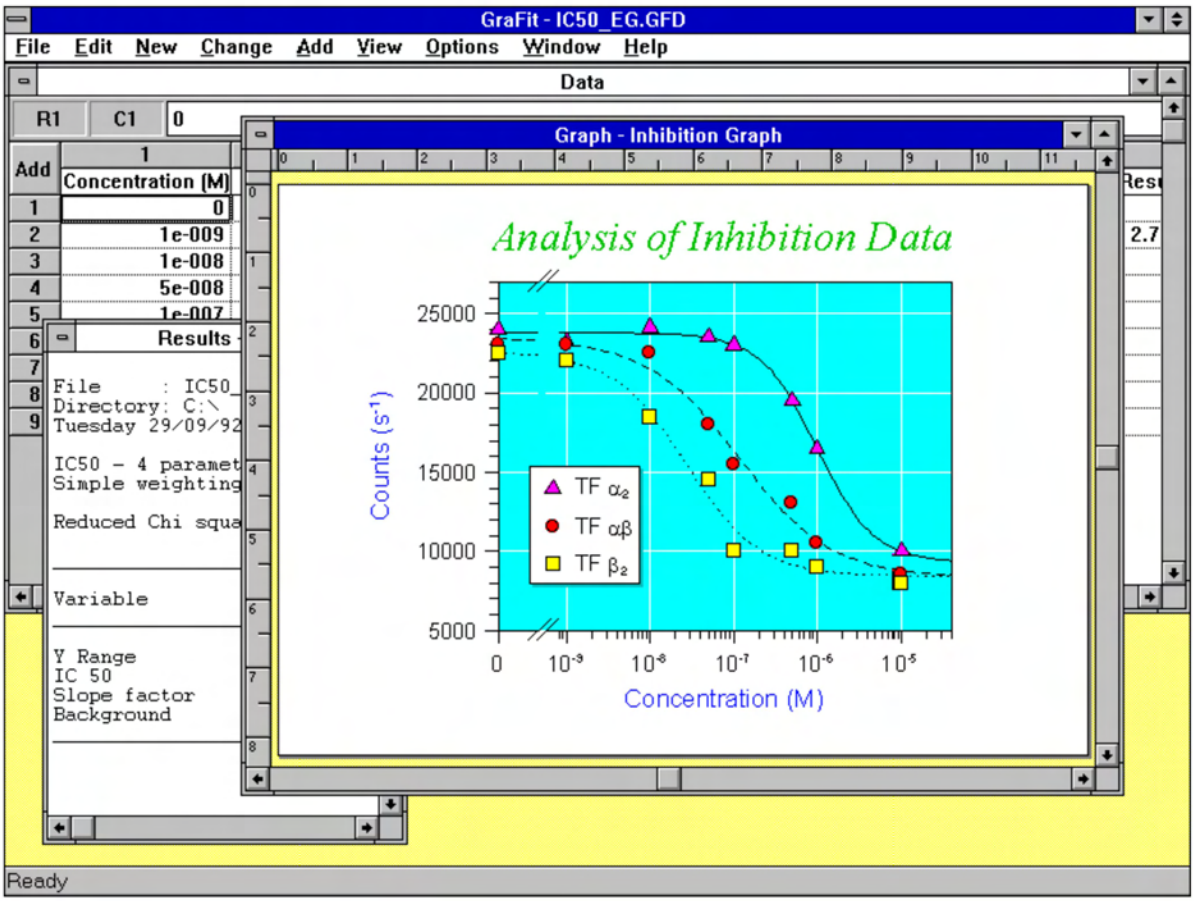
GraFit User's Guide



Erithacus Software

Erithacus Software

GraFit
User's Guide

Version 3.0

Program and Manual Designed and Written by

Robin J. Leatherbarrow

Biological Chemistry
Department of Chemistry
Imperial College of Science, Technology & Medicine
London SW7 2AY
U.K.

VERSION 3

Citing GraFit in Publications

When citing GraFit, simply treat this software publication as you would a book. For example:

Leatherbarrow, R.J. (1992) *GraFit* Version 3.0, Erithacus Software Ltd., Staines, U.K.

Preface to Version 1

Computers play an increasing role in laboratory work. It is rare that a present day laboratory is without at least one computer. Indeed, their proliferation is such that they often outnumber personnel! Over the last few years computer hardware has become freely available, with increasingly powerful performance coupled to ever reducing price. However, scientific computer software, with some notable exceptions, has lagged behind the capabilities of the computer. With GraFit I have tried to produce a software package that will be useful to scientists involved in data handling. The program uses the capabilities of the computer to the full, and hopefully satisfies some of the needs that professional scientists have for professionally written software tools.

The examples in this manual reflect, to a large degree, my own biases with regard to experimental equations for analysis. Therefore if there seems an inordinate number of equations relating to enzyme systems the reader is asked to indulge my excesses. The structure of the program makes it suitable to analyze any equation of the form

$$y = f(x)$$

and so it is applicable to a vast number of experimental situations.

May, 1989 ROBIN J. LEATHERBARROW
London

Acknowledgements

In the preparation of this program I have benefited from encouragement and advice from many sources. My thanks to all my friends and colleagues who have given help and advice, including all the members, past and present, of the Biological Chemistry section, Dept. of Chemistry, Imperial College of Science, Technology and Medicine, London U.K., especially Alan Russell, Tim Wells, Alan Fersht and Sanjoy Ray.

I am particularly indebted to my wife, Marcella, who proof-read the manual and put up with the ups and downs associated with program development.

Version 2

My thanks to the many people who have given advice for improving the program. In particular, I am grateful to Bjarke Ebert, Bernard Mangold, Sanjoy Ray, Martin Reddington, Walter Ward and Tim Wells for many constructive suggestions. Once again, this program and manual would not have been possible without the help and encouragement of Marcella.

Version 3

Many of the improvements in this version are due to the excellent suggestions that have come from various GraFit users. My thanks to all who contributed, particularly to Pietro Belfiore, John Gebler, Douglas Kell, Steve Matthews and David Tew for many valuable recommendations, and to Martin Reddington, Walter Ward, Tim Wells and Marcella, for their continued advice.

Contents

1 Getting Started

Welcome

GraFit combines presentation-quality scientific graphs with comprehensive data fitting in the Microsoft® Windows™ environment. Together, this results in a powerful package that is easy to use, produces stunning output, and can also fulfil the needs for quantitative data analysis. Data can be fitted using various types of theoretical models, and your own equations can easily be added to those supplied with the program. Comprehensive graph plotting facilities allow you to display your data on-screen and edit the graphs interactively before sending the output to your printer or plotter.

What You Need

Any Windows-compatible computer (80486 or better is recommended) will run the program. The program will make use of a numeric coprocessor if one is fitted into your computer. A mouse is required, and at least 4 MB of memory is needed together with 5 MB of hard disk space.

To run GraFit you need Microsoft® Windows™ version 3.1 or greater.

The GraFit Package

The GraFit package comprises the following:

- GraFit program disks, 5¼" or 3½".
- The GraFit user manual (this manual).
- User registration card.

Please complete your registration card and return it to Erithacus Software as soon as possible. This will ensure that you are informed of the latest updates to the program.

About this Guide

This manual describes how you can use GraFit to examine your data, and to plot graphs. The basic operations of the program are described in Chapters 2–5, an overview of creating graphs is found in Chapter 6, specific aspects of configuring graphs in Chapters 7–20, and data fitting in Chapters 21–32. We recommend that you read the manual fully, and try out some of the examples described.

If you are not familiar with Microsoft® Windows™ you should consult your Windows documentation before using GraFit.

Notational Conventions

The following notational conventions are used throughout this manual:

Convention	Meaning
SMALL CAPITALS	Small capitals indicate keys to be pressed. For instance, A indicates "press the 'A' key". If more than one key needs to be pressed this is written as (for instance) SHIFT END which would mean "hold down the 'Shift' key and then press the 'End' key".
`Monospace`	Monospace type indicates text that should be entered on the computer keyboard. For example, to obtain a directory of drive A: when in DOS you would enter `DIR A:` RETURN
italics	Italicized text indicates a placeholder that should be replaced by the required text. For example, to change directory when in DOS the command is `CD` *directoryname* The italicized text should be substituted here by an appropriate directory name, such as `CD GRAFIT`
Bold	Bold text is used for menu item commands, e.g. the **Edit** menu. Commands followed by an ellipsis (...) result in a dialog box.

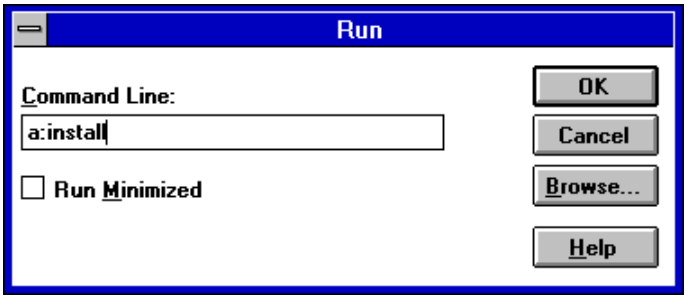
Installation

The GraFit program disk #1 has a file called README.TXT that contains any last-minute information about the program. To view this text enter
`TYPE README.TXT`
at the DOS prompt. Alternatively, you can read this file into the Windows Notepad application.

Before installing GraFit, it is essential to have Microsoft® Windows™ version 3.0 or later installed on your computer. GraFit installation is carried out from within Windows.

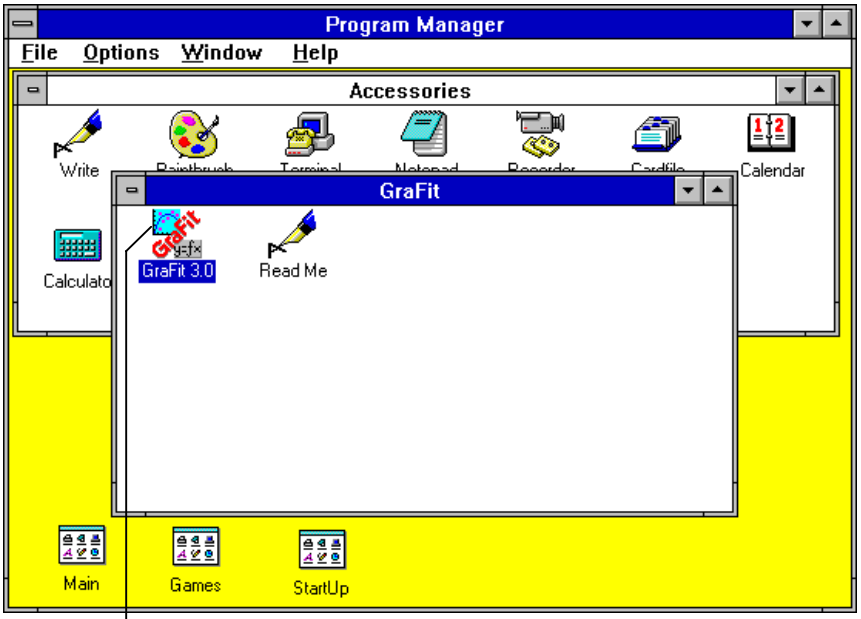
To install GraFit

- If Windows is not yet running, start it by typing
  WIN RETURN
- Insert the GraFit program disk #1 into drive A.
- Choose the **File Run...** command from the Program Manager menu (Windows 3.1), or the **Run…** command from the **Start** button (Windows 95) to produce the Run dialog.



- Enter
  `A:INSTALL`
  then select **OK**.
- Follow the instructions.

The installation program will copy the files to your hard disk, and build a new Program Manager group from which you can run GraFit by double-clicking on the GraFit icon.

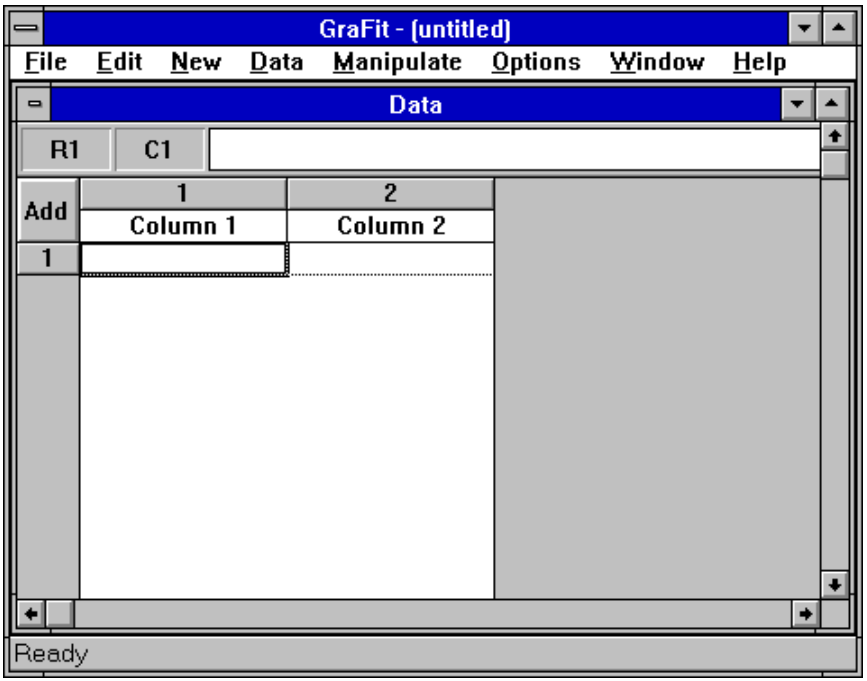


Double-click on the GraFit icon to run the program

2 GraFit Basics

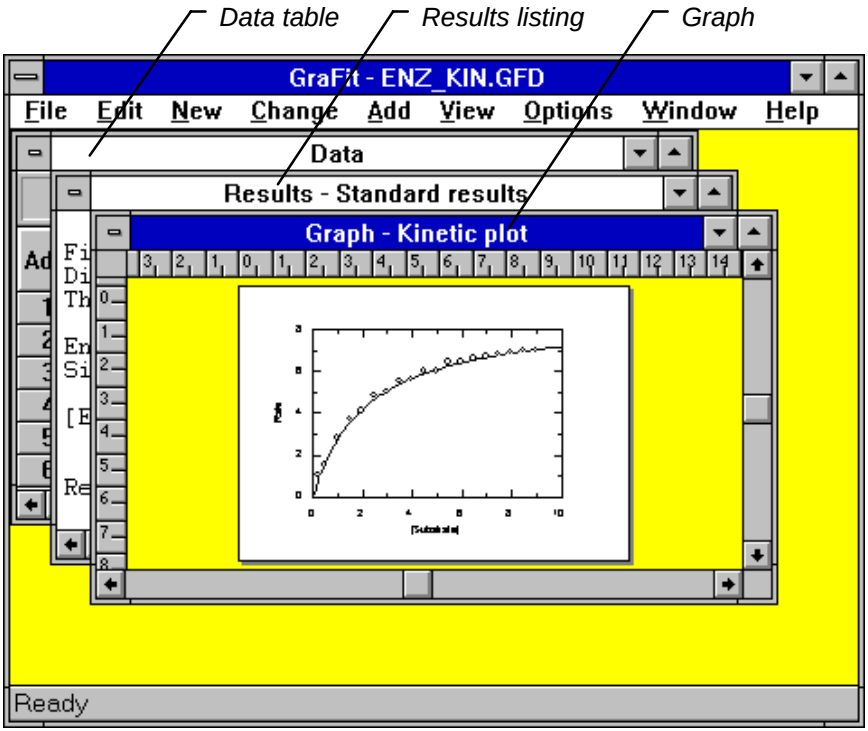
Running GraFit

To run GraFit, double-click on the GraFit icon that the installation program will add to the Program Manager. When you first start GraFit, a Data window appears. This holds the data table that is used to enter raw data values for graph plotting or data-fitting.



Use of Multiple Windows

When running GraFit, you will create graphs and fit data to generate results listings. Graphs and results are held in separate sub-windows created within the main GraFit window.



GraFit always has one, and only one, Data window, but may have any number of Results and Graph windows. If more that one data table is needed, it is possible to run another copy of GraFit, using the ability of Windows to run several programs simultaneously.

Selecting a Window

To select one of these windows, click on it using the mouse. The selected sub-window is drawn on top of the other windows. Alternatively, choose the window from the **Window** menu.

Menus

Different menus are present on the main GraFit window, depending upon whether the Data, a Results or a Graph window is selected. Commands to manipulate and fit data are only present when the Data window is selected; commands to modify graphs are present only when a Graph window is selected.

Data window menu

Results window menu

Graph window menu

Common to all are **File** (open and save files), **Edit** (copy to and paste from Clipboard), **New** (create new graph or results windows), **Window** (select window) and **Help**.

Getting Help

Extensive help is available when using GraFit. To get help, select the **Help** menu, and choose the **Help Index** command. Alternatively, press the F1 key.

Context Sensitive Help

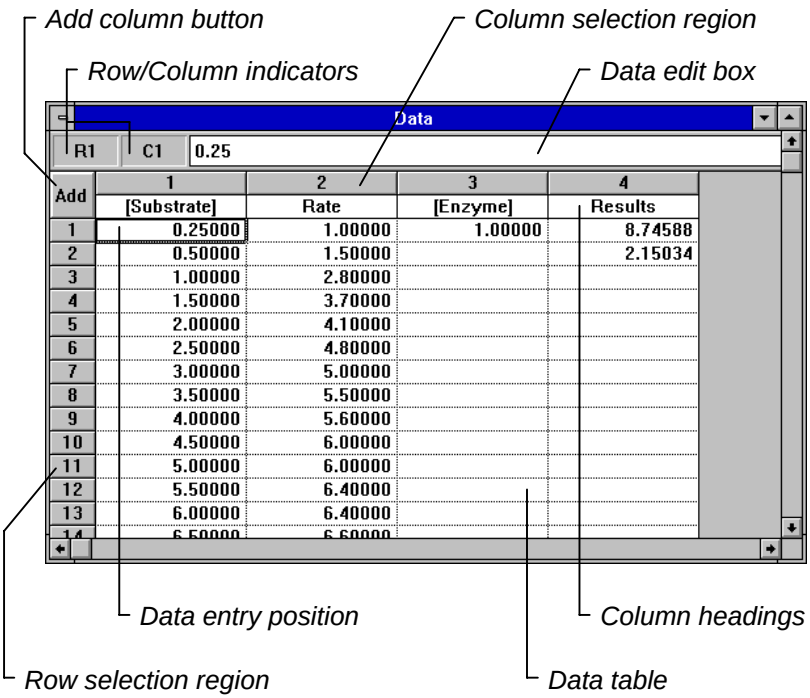
The help provided by pressing F1 is context sensitive. This means that if you have a dialog box on screen, and you are not sure what some of the selections mean, pressing the F1 key will bring up help on that specific dialog. Context sensitive help is available for dialog boxes and for error messages.

3 Entering Data

The Data Window

A GraFit data file always contains a single Data window. When first created, it has two empty data columns, labeled “Column 1” and “Column 2”, although more columns can be added as required.

The parts of a Data window are shown in the following illustration.



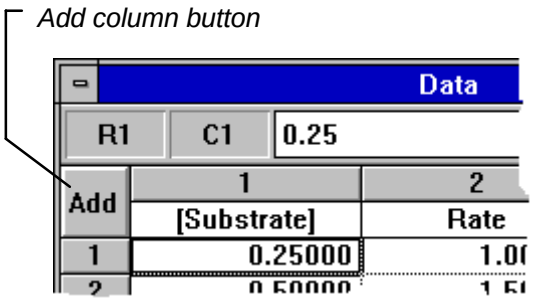
Position row 1 column 1 in the above figure is marked by a dark outline; this is the current data entry position. Data entries are displayed as they are entered, and may be edited, in the data edit box. When the entry is complete, pressing RETURN, TAB or a cursor key, or clicking the mouse on a different position in the data table, will accept the data, and move the data entry position.

The following table describes the elements in a Data window. Certain sub-regions are distinguished by the mouse cursor changing shape when it is over them.

Element	Description	Cursor
Add column button	Click on the **Add** button to add an additional data column to the table.	
Column selection region	Click and drag the mouse on the column numbers to select entire columns of data.	
Row/Column indicators	These show the current data entry position.	
Data edit box	As data are entered, the text is shown in the data edit box. This behaves in the same way as any other edit box within Windows, and allows the data values to be edited.	
Data entry position	This indicates the position where newly entered data is placed. Move this position using the cursor keys, or by clicking the mouse on a different part of the data table.	
Row selection region	Click and drag the mouse on the row numbers to select entire rows of data within the data table.	
Column headings	These label the various data columns. Edit them by double-clicking on the heading.	
Data table	The data values themselves are listed in the data table. Click the mouse on a value to move the data entry position; drag the mouse to select larger regions of the data table.	

Adding Extra Data Columns

The starting data table contains only two columns. To add extra columns press the **Add** button at the top left of the data table.



Alternatively, use the **Data Add Column...** command.

Naming Data Columns

The default column heading names are not very informative, and it is useful to change the name to reflect the contents. To change the name of a particular column, double-click on the heading. Alternatively, select the **Data Rename Column...** command.

Data Types

Each column of data can hold up to 8196 data values (although only around 1500 can be displayed). These columns hold the data for all operations in GraFit, including plotting graphs and curve fitting, and are also used as a place to store fitted results. The number of columns is limited only by available memory, and so it is possible to have the results from a series of related experiments together in memory. Any combination of columns may be used to generate graphs or for data fitting.

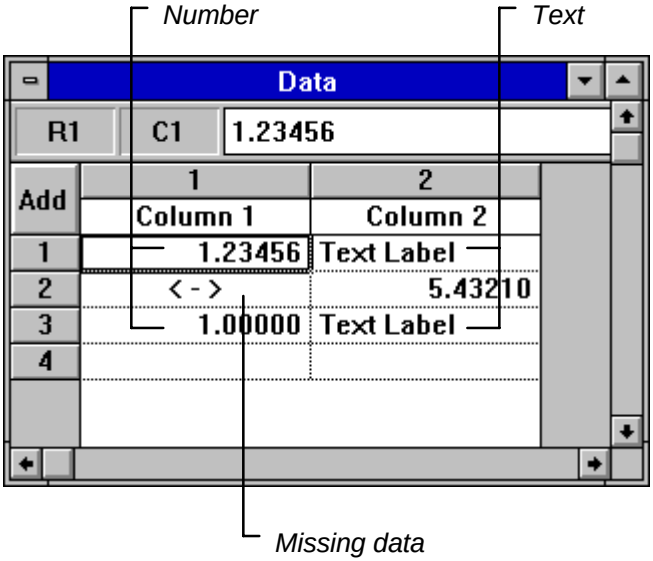
Within a column, each individual cell holds a data value which may be one of three different types.

Numbers Numbers are the most common data form that will be entered into GraFit. They are used to plot *x*/*y* scatter plots, and for data fitting by regression analysis. Very small or large numbers can be entered in exponential format, for example `1e7` is equal to 10000000, and `1e-23` represents 10^{-23}.

Text Text entries are used for the labels on the category axes of charts.

Missing Data Experimental data often contain certain values that are not available (perhaps because that particular experimental sample was the one that fell on the floor!). In order to maintain a neat data array, GraFit allows you to insert “missing” values by leaving the data entry blank. All the graphing and data fitting elements of the program will recognize such points, and omit them from the analysis.

The representation of these different data types is illustrated in the following diagram.



Moving the Data Entry Position

When you have entered your data value you will need to move to the next data entry position. This is done by pressing TAB or RETURN after entering the value. The following keys control movement of the data entry position within the data table.

To	Press
Move the entry point to the right	TAB or RIGHT
Move the entry point to the left	SHIFT TAB or LEFT
Move the entry point up	UP
Move the entry point down	DOWN
Move to the next entry point to the right, or if already in the extreme right hand column move to column 1 of the next row	RETURN
Move to the first column of a row	CTRL LEFT
Move to the last column of a row	CTRL RIGHT
Move to the first row of a column	CTRL UP
Move to the last row of a column	CTRL DOWN
Move to row 1 column 1	CTRL HOME

It is often convenient to press RETURN between entering data points.

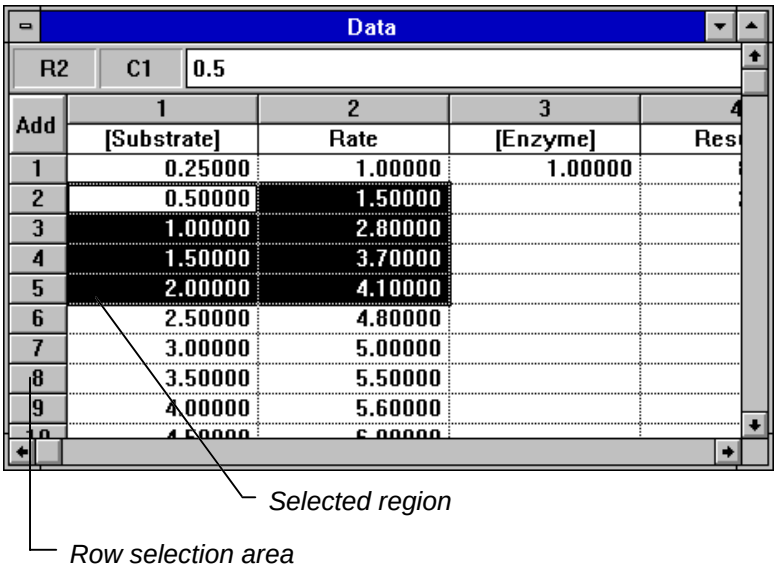
Moving the data entry position using the mouse

To move the data entry position using the mouse, move the mouse pointer to the position you require, and click the left mouse button.

Selecting Regions of the Data Table

Selecting Part of the Data Table

To select a portion of the data table, drag the mouse over the required region. Alternatively, hold down the SHIFT key, and use the cursor keys to extend the selection. The selected region is displayed inverted, as shown below.



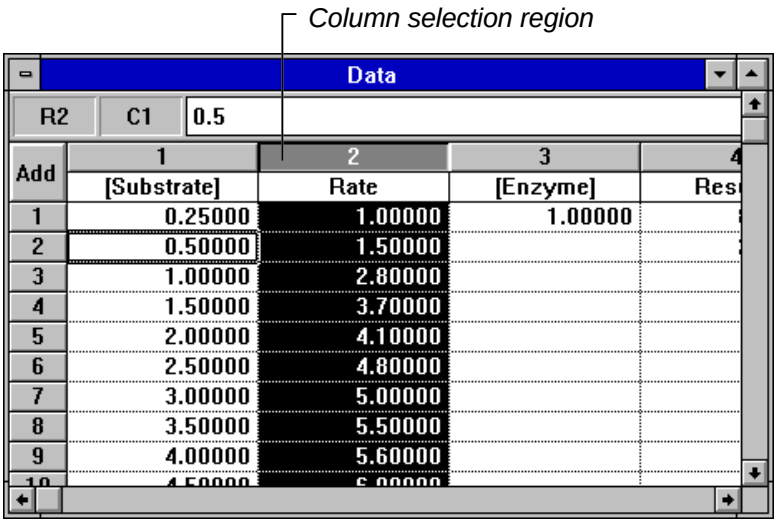
Clicking the mouse within the row selection area results in the entire row being selected.

To make a more extensive selection:

1. Click on the start position.
2. While holding down the SHIFT key, use the scroll bars to move the window as necessary.
3. Still holding down the SHIFT key, and click on the end position for the selection.

Selecting Entire Columns

Whole columns of data are selected by clicking on the column selection region. The selected column is displayed inverted.



Drag the mouse along the column selection region to extend the selection.

Selecting columns using the keyboard

Columns can be selected from the keyboard by pressing F6 until the column selection region becomes highlighted. Move the selection using the cursor keys, and extend it by holding down the SHIFT key while using the cursor keys.

Copying and Pasting Data

Selected regions of the data table are copied to the Clipboard using the **Edit Copy** command, and pasted using **Edit Paste**. Data may also be pasted to and from other applications in the same way.

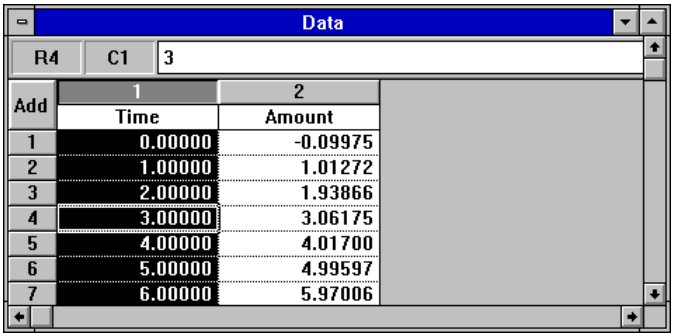
Copying and Pasting Regions of the Data Table

1. Select the region of data that you wish to copy.
2. Choose the **Edit** menu, and select **Copy**.
3. Move the data entry position to where the data is to be copied.
4. Choose **Edit Paste**. The data will be inserted at the data entry position.

Copying and Pasting Columns

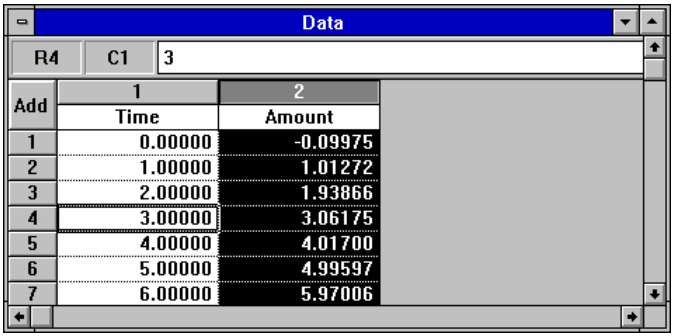
Entire columns of data can also be copied and pasted. Pasting into the data table when a column is selected results in the pasted values being placed in a new column, immediately to the right of the selected column.

1. Select the column(s) of data that you wish to copy.

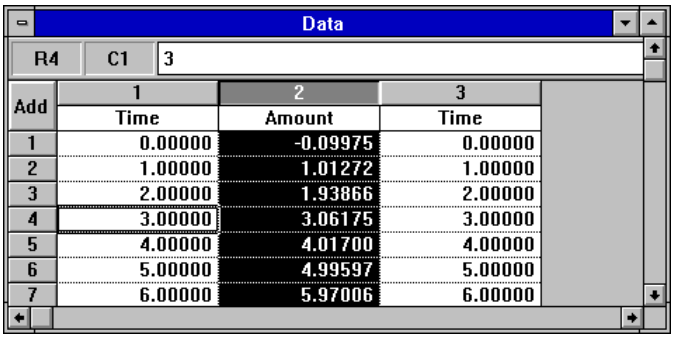


2. Choose the **Edit** menu, and select **Copy**.

3. Select the column after which the data are to be inserted.



4. Choose **Edit Paste**.



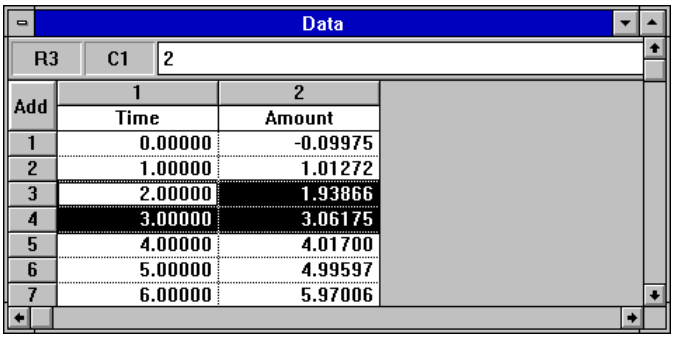
The data are pasted into the table.

Deleting Data

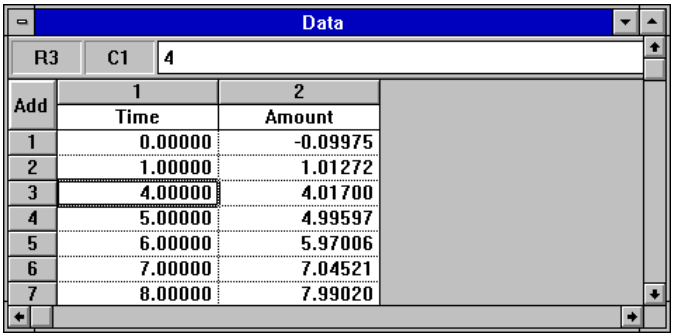
The **Edit Clear** command, or the DEL key, can be used to delete the currently selected data or columns.

Deleting a Region of Data

1. Select the data value(s) to be deleted.



2. Choose **Edit Clear**, or press DEL.



The remaining data values in the columns are moved up to fill the gap left by the deletion.

Deleting Columns of Data

1. Select the column(s) to be deleted.
2. Choose **Edit Clear**, or press DEL.

Inserting Values into the Table

Values are inserted into the data table using the **Edit Insert Gap** command. This command inserts a blank value at the data entry position, and moves the rest of the data down the column. If this command is chosen when a region of the data table is selected, the gap will be the same size as the selection.

Displayed Data Format

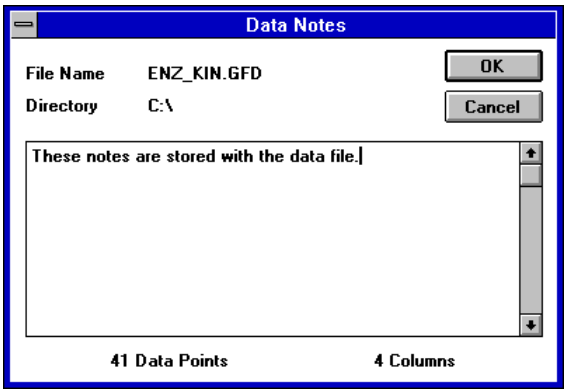
The numbers in the table are initially displayed in a fixed decimal format. If this is not appropriate for your data (for example if you are entering very small numbers) you can change the format on a column by column basis by double-clicking on the data table, or by using the **Options Data Format...** command (see Chapter 28 for a full description of display data formats).

Adding Notes to the Data File

It is possible to store notes with a data file. Such notes are useful for adding comments regarding experimental conditions etc.

To view or edit the notes:

- Select the **Data** menu and choose the **Edit Notes...** command.



- Edit the notes as appropriate.

4 Working with Data Files

You can read and save GraFit data from the **File** menu. A GraFit data file (.GFD file) holds the data from the data table, plus the contents of any Graph or Results windows that have been created. Files can be deleted using the Windows File Manager application.

Opening a Data File

You can open a new or an existing data file from the **File** menu of the main GraFit menu.

Creating a New File

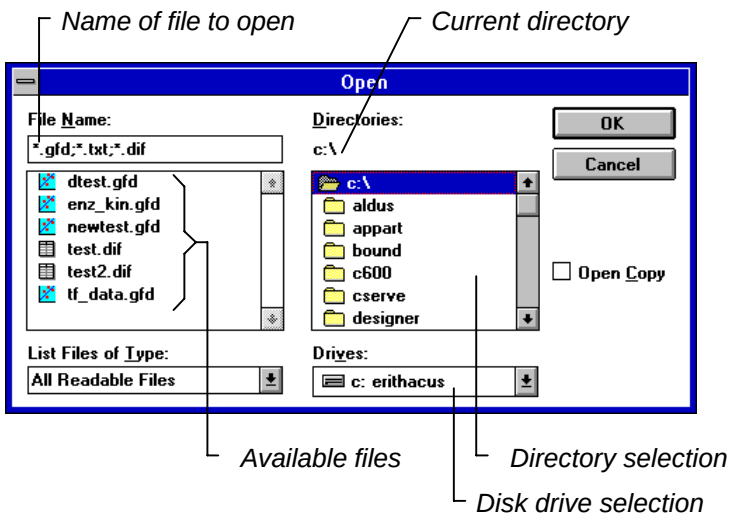
To create a new, empty data file:

- Select the **File** menu and choose **New**.

Opening an Existing File

To open an existing GraFit data file, follow the following steps:

1. Select the **File** menu and choose **Open...**. The Open dialog box will be displayed.



The available files are classified according to type by the small icon drawn next to the file name.

Icon	File type
	GraFit data file (.GFD). GraFit data files hold the data table, together with any associated graphs and results listings.
	Data file that can be read by GraFit.
	Text file. Text files may be converted to data, or imported as a Results window.
	Unrecognized file type.

This classification is based on the three letter file extension.

2. Select the file that you want to open from the list of those available. If necessary, change directory using the directory selection list box, or change disk drives using the Drives drop-down list box.
3. Select **OK**, or press RETURN.

To open an existing file using the mouse, simply double-click on the filename you wish to open.

Opening a Copy of an Existing File

If the Open Copy box is checked, GraFit will open a copy of the selected file. Use this option to ensure that you do not overwrite a valuable file.

Selecting Several Data Files

It is possible to select more than one data file from the list of available files by holding down the CTRL key when making your selections. If multiple files are selected in this manner, the first file will be read as normal, and the subsequent files merged in turn as if the **File Merge...** command had been used on each.

Saving a File

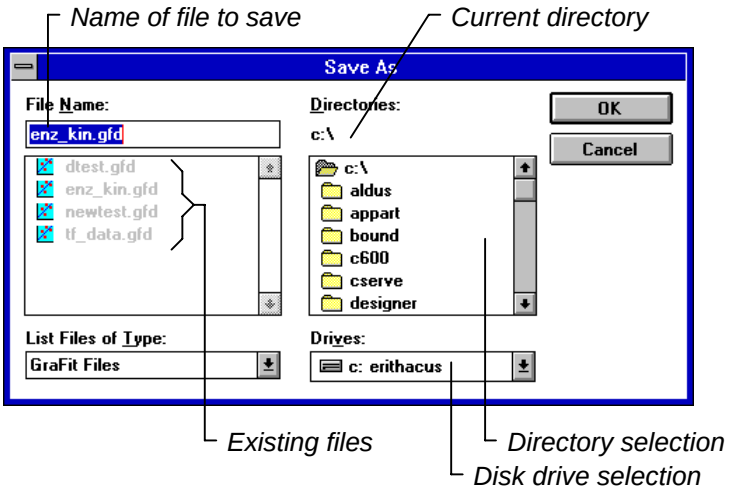
Files may be saved for further use. There are two commands for this, **Save** and **Save As...**. Both these commands are found in the **File** menu.

Saving a New File

Use the **Save As...** command to give a name to a file that you have just created. It can also be used to save the current file under a new filename, keeping the original file unchanged.

To save as a new file

1. Select the **File** menu and choose the **Save As...** command. The Save As dialog box will be displayed.



2. If you wish to change drive or directory prior to saving the file, do so using the directory selection list box, and/or the disk drive selection drop-down list box.

3. Type in a filename for the new file. If you don't give the file an extension, GraFit will append the default data file extension of .GFD. If you wish to save the file with no extension, you should enter the filename followed by a period, for example `MYDATA.`

4. Choose the **OK** button. Your file will be saved onto the disk.

The new file name will be displayed on the title bar of the main GraFit window.

Note If you give the name of a file that already exists, GraFit will ask whether you want to overwrite the existing file with the new one. Choose the **Yes** button if you wish to replace the old file, or choose **No** and select a different filename if you do not want to lose the existing file.

Saving Changes

The **Save** command saves the changes you have made to the current file. The filename remains unchanged. To save your file:

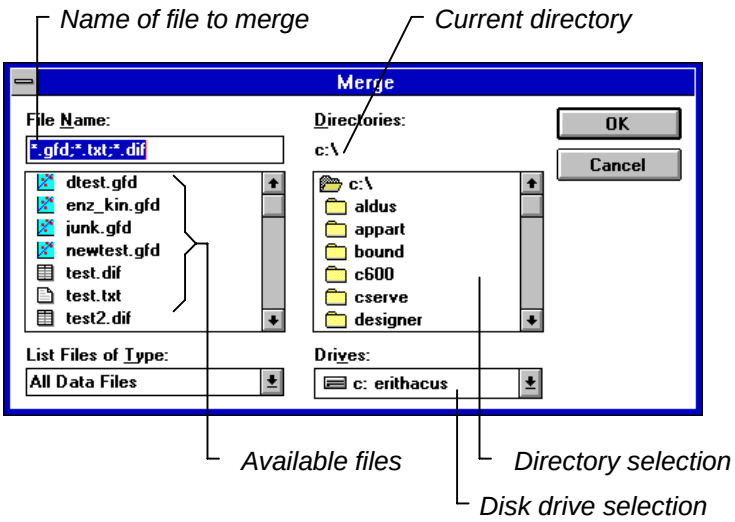
- Select the **File** menu and choose the **Save** command.

If no changes have been made to a new file, the **Save** command is inactive, and will be displayed grayed.

If you select the **Save** command for a file that does not have a filename, the command will be treated as a **Save As...** command.

Merging Data From a File

Data from a file can be merged into the current data table using the **File Merge...** command.



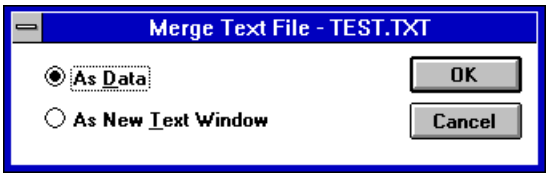
The Merge dialog box operates in a similar way to the Open dialog box. Merged data are placed in new columns at the right of the data table.

Merging Several Files

It is possible to select more than one data file from the list of available files by holding down the CTRL key when making your selections. All the selected files will be merged into the data table in turn.

Merging a Text File

When a text file is chosen from the **File Merge...** command, there are two possible ways of merging the file: as data, and as a Results window. The program determines which you require by the following dialog box.

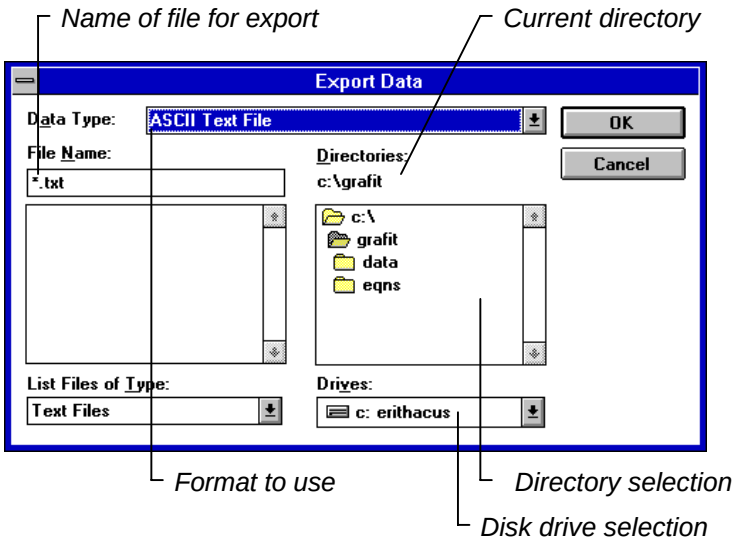


Make the appropriate choice, and press **OK**.

Exporting Data From GraFit

The data in the data table may be copied to other applications via the Clipboard, as described in Chapter 3, or can be exported to a file using various file formats. The format used in the GraFit .GFD file cannot be read by other programs, and so it is necessary to create a file in a standard format if it is to be used by another application.

To export data, select the **File Export Data...** command.



Choose the data format to use for the file, enter a file name, and press **OK**.

Available Data Formats for Export

The formats that are available depend on which import/export filters were added during installation, but may include plain ASCII, Data Interchange Format (DIF), and Comma Separated Variables.

5 Printing

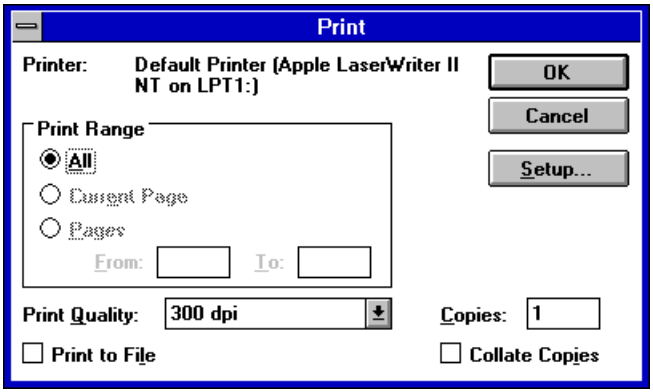
Printing

The contents of the active window are sent to the printer using the **File Print...** command. Depending on which window is on top, this command will print:

- The contents of the main data window as a table of numbers.
- Results listings.
- Graphs.

Printing the Data Table

When the data table is uppermost, the **File Print Data...** command will print the contents of the table.



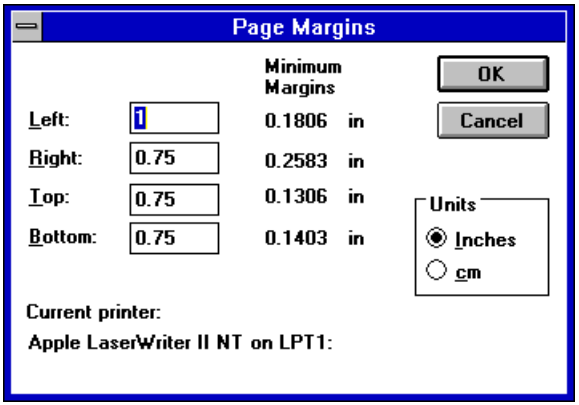
Enter the number of copies to be printed, and press **OK**. Depending upon the printer in use, it may be possible to alter the print quality.

Printing to File

If the Print to File box is checked, a disk file will be created instead of sending data to the printer.

Setting the Margins For Text Output

Textual output from GraFit is printed within margins that may be altered using the **File Text Page Setup...** command before printing. Selection of this command results in the following dialog box.



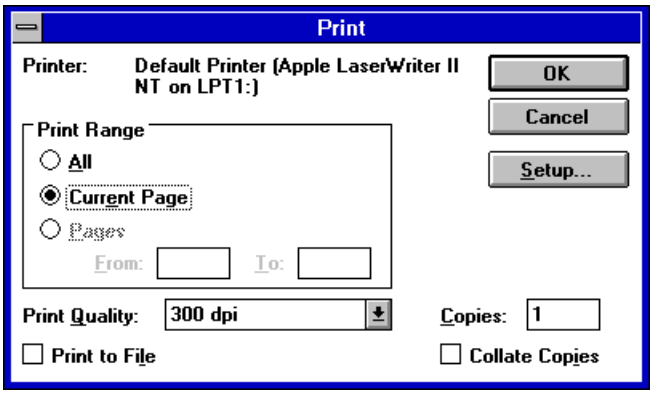
The margin settings are specified as follows:

Error! Objects cannot be created from editing field codes.

GraFit obtains the dimensions and orientation of the page from the printer driver, and so it is important that the driver has been correctly initialized. The default margin settings are: left margin, 1.00 inches; right, top and bottom margins, 0.75 inches. Enter the margins that you require (either in inches or in centimeters — choose between inches and centimeters from the Units section of the dialog box). The dialog box also shows minimum values for these margins. Most printers cannot print right up to the edge of the paper and the minimum values displayed indicate how small the margins can be. GraFit will prevent you from entering margin settings smaller than these.

Printing Results

When a Results window is active, the contents of the window are printed using the **File Print Results...** command.



The default Print Range is "Current Page"; this will print the contents of the uppermost Results window only. If the All option is selected, all Results windows will be printed in turn. The function of the **Setup...** button is described later.

Enter the number of copies to be printed, and press **OK**. Depending upon the printer in use, it may be possible to alter the print quality.

Printing to File

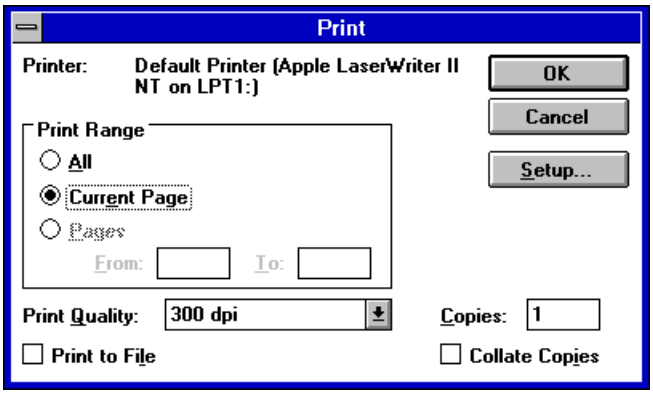
If the Print to File box is checked, a disk file will be created instead of sending data to the printer.

Setting the Margins for Output

Margin settings for the printed results are specified using the **File Text Page Setup...** command. The operation of this command is as described above for printing data.

Printing Graphs

The **File Print Graph...**command is used to print the currently selected graph window, and results in the following dialog box.



The default Print Range is “Current Page”; this will print the contents of the uppermost Graph window only. If the All option is selected, then all Graph windows will be printed in turn. The function of the **Setup...** button is described below.

Enter the number of copies to be printed, and press **OK**. Depending upon the printer in use, it may be possible to alter the print quality.

Printing to File

If the Print to File box is checked, a disk file will be created instead of sending data to the printer.

Selecting a Printer and Setting Printer Options

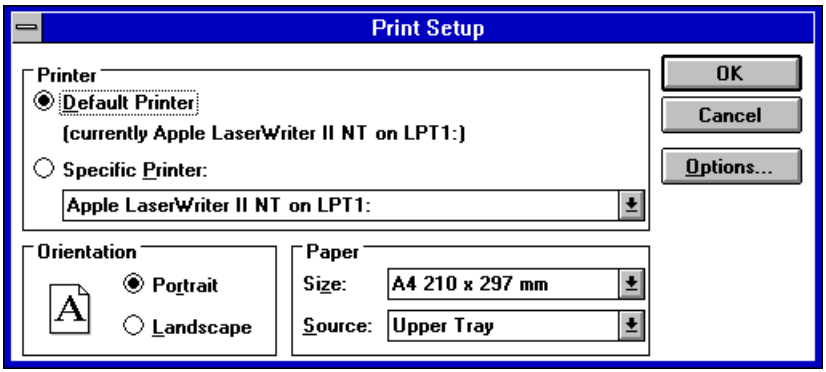
When first the program is started, GraFit will choose the Windows default printer, configured with the default options, for output. In many cases you will not need to alter these settings when running GraFit.

Printer selection and configuration operates at three levels.

1. Permanent settings. These are set using the Control Panel application (see your *Microsoft Windows User's Guide* for full details). Any changes will remain in place when Windows is next run.
2. Settings for the current GraFit session. These are made using the **File Print Setup...** command. These settings are used for the current session of GraFit, and do not affect any other Windows applications that might be running at the same time.
3. Settings for the current print job. Within GraFit, it is also possible to change the print settings for an individual printing job. This is done using the **Setup...** button that appears on the Print dialog.

Print Setup

The printer and the printer settings are set using the **File Print Setup...** command, or by choosing the **Setup...** button on the Print dialog. The following dialog box is displayed.



This dialog box allows the printer to be selected, along with the printing orientation and the paper size and source. For some older printer drivers it is not possible to select the orientation and paper settings, in which event these settings will be grayed.

The **Options...** button allows printer-specific configuration, which gives further control over the printed output.

Printer Limitations

The output from GraFit is ultimately limited by the quality of the printer used. Although the output attempts to be independent (apart from resolution) of the print device used, certain printers have intrinsic limitations or advantages, as explained below.

Output From Plotters

On screen, and when using conventional matrix printers, it is possible to erase part of an image by overdrawing in a background color. Therefore one object may be placed 'on top' of another. Such effects are not available on a plotter, as it is not meaningful to 'draw in a background color' for such devices. Plotters can only use vector-based text, and we recommend setting the default text to the Modern or Roman fonts.

Color

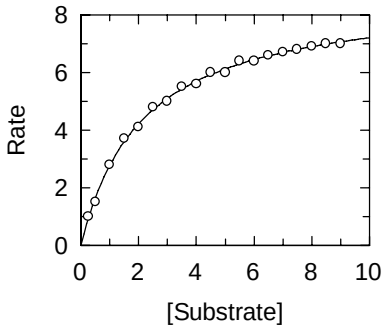
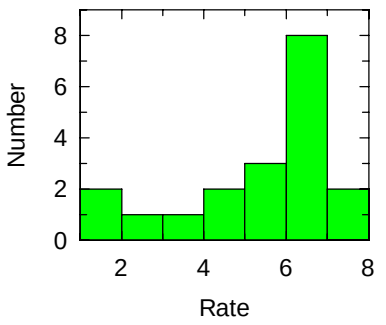
Printers that can produce colored output (e.g. HP PaintJet, HP plotters, color PostScript printers) will generate output that matches the colors of the screen image as closely as possible. On monochrome devices, colors will be replaced by shades of gray.

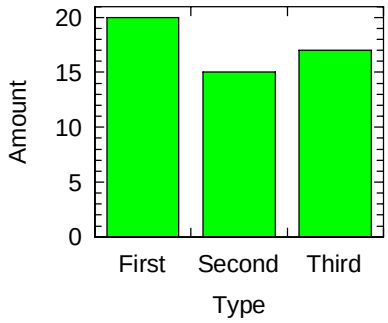
6 Graph Basics

GraFit can be used to create publication-quality graphs in various sizes, shapes and styles. This chapter describes the types of graph that can be produced, and how regions of a graph can selected, sized and edited.

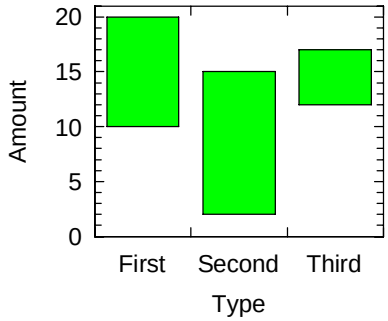
Graph Types

GraFit divides graphs into different types depending on the form of data they display. For example, column charts (which have numeric data only on their *y* axes) and scatter plots (which have numeric data on both axes) are distinct graph types that cannot be interconverted. On the other hand, scatter plots with double *y* axes, offset axes, logarithmic axes, or error bars are considered to differ only in style, and may be interconverted as required. The following graph types are available.

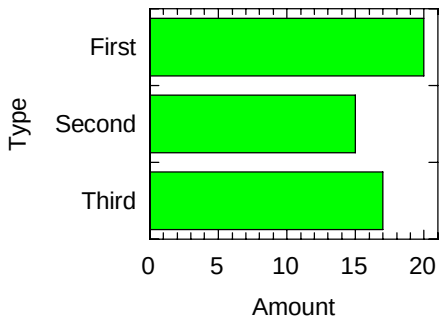
Graph or Chart Type	Data Content
	**X/Y graph** (scatter graph) has numeric data for both the *x* and *y* axes.
	**Histogram** has numeric data for both the *x* and *y* axes. Note that despite the similar appearance, this means that a histogram is fundamentally different to a column chart.



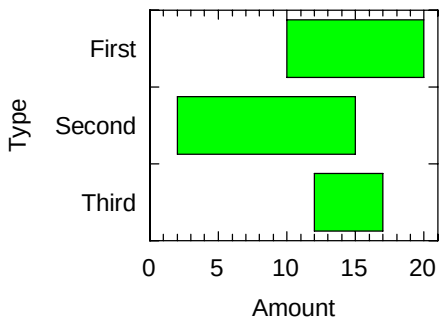
Column chart has numeric data for the *y* axis, and category names for the *x* axis.



Low/High column chart has two numeric values for the *y* axis (the upper and lower limits of the columns), and category names for the *x* axis.



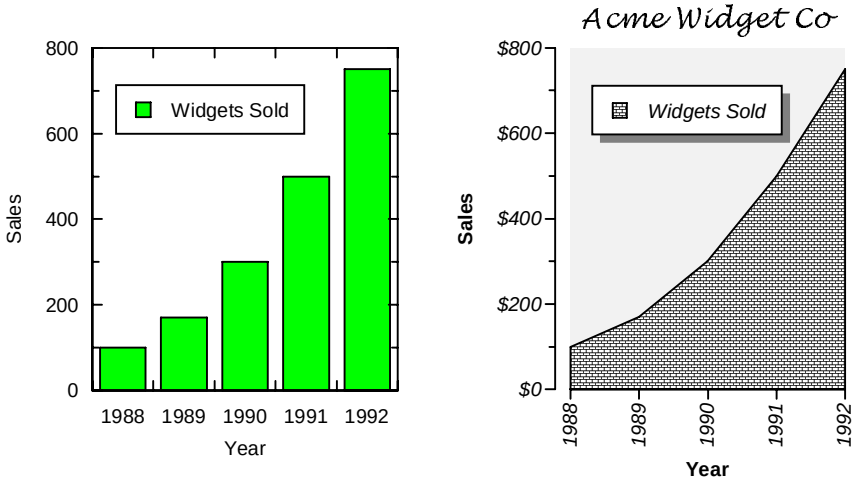
Bar chart has numeric data for the *x* axis, and category names for the *y* axis.



Low/High bar chart has two numeric values for the *x* axis (the upper and lower limits of the bars), and category names for the *y* axis.

Graph Styles

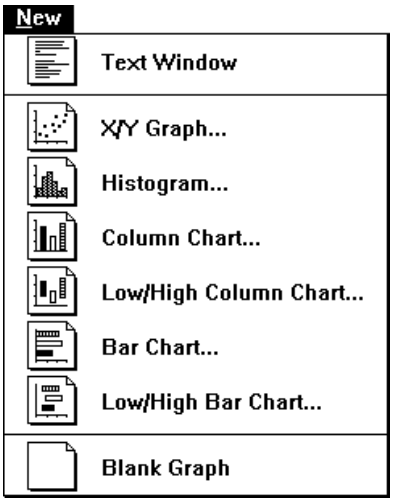
Within any particular graph type, a wide range of different styles are available. For example, the following graphs are both column charts, and differ only in the style in which they display the same data.



Customized graph styles can be selected during graph creation, or applied later. In addition, individual aspects such as text or data point styles can be edited as required. A full description of graph style selection is given in Chapter 20.

Creating a New Graph

Graphs are created using the **New** menu.

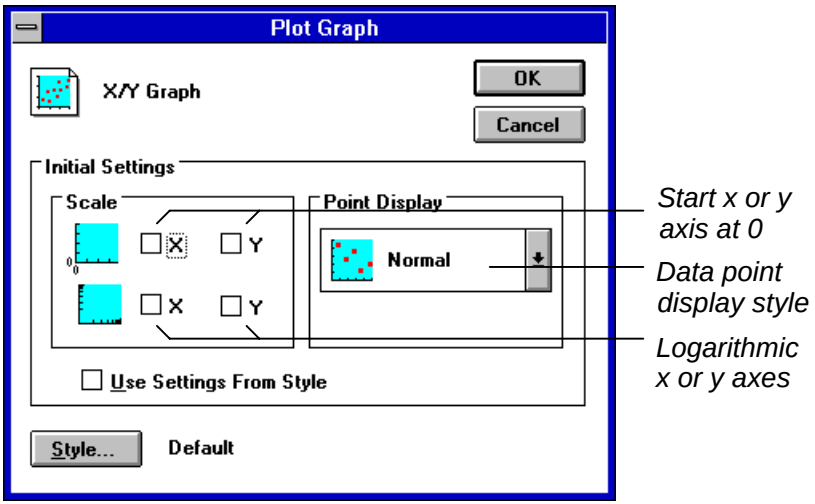


*The **New** menu is used to create a new graph window*

The data used to plot the graph must be present in the data table. Any column can be used to represent the *x* and *y* data. All the commands for creating new graph windows work in a similar manner, and in each case the program will produce a dialog box that allows the graph style to be specified, followed by a dialog box from which the data columns are selected.

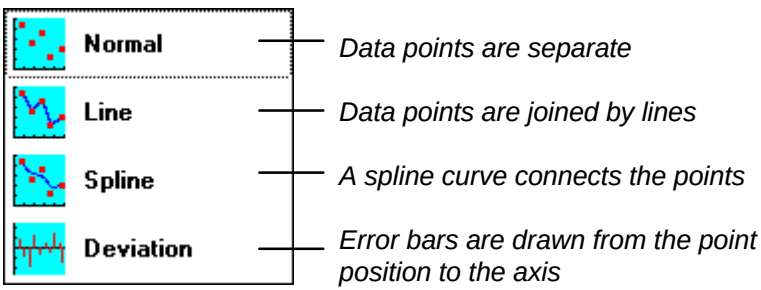
Creating a Scatter (X/Y) Graph

The **New X/Y Graph...** command will create a new scatter plot. After selecting this command, you will have the opportunity to set the various aspects of the graph style from the following dialog box.

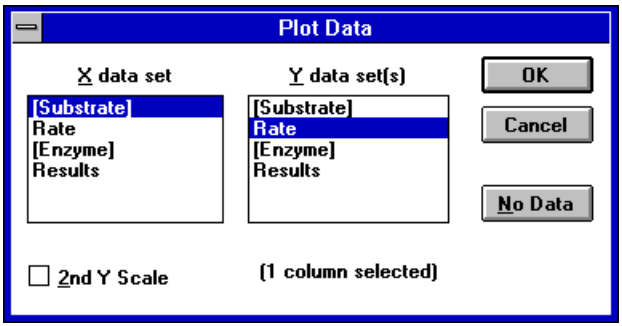


The style on which the graph is based is indicated at the bottom of the dialog box, and may be specified by pushing the **Style...** button. A full description of style selection is given in Chapter 20.

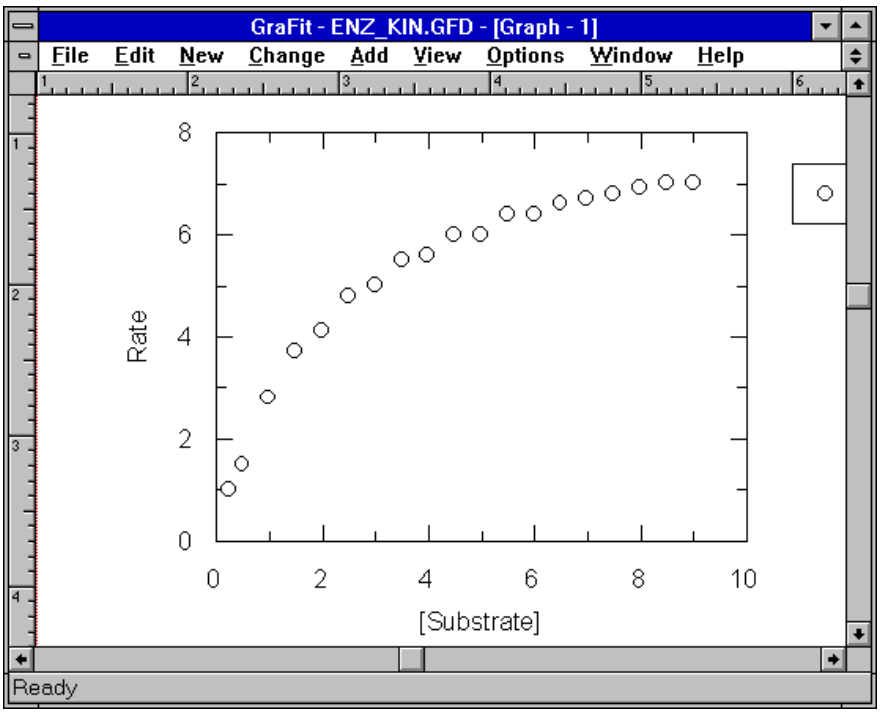
The Initial Settings region of the dialog box lets you initialize some commonly altered features of the graph, including whether or not the axes are logarithmic, or the scale should begin at *x* or *y* = 0. The Point Display section can optionally be set to display data points joined by lines or spline curves, and to display the data in the form of an error plot. This is illustrated below.



After selecting the initial graph style, it is necessary to indicate which columns hold the data points. This is done from the following dialog box.

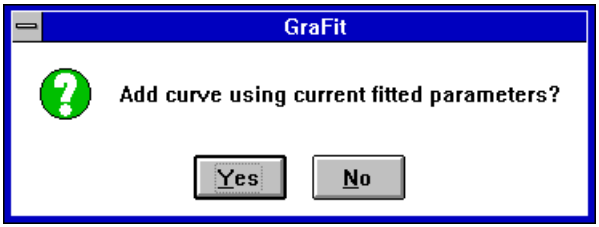


The column that holds the *x* data is selected, along with the column that holds the *y* data. More than one *y* data column can be chosen by holding down the CTRL key when making the selections. After pressing **OK** a new Graph window will be created that holds the graph, as shown below.



Adding Curves to the Graph

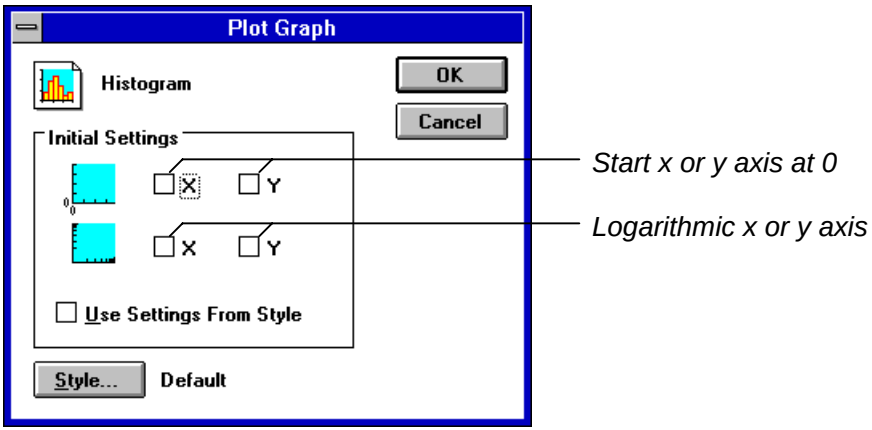
If, prior to plotting the graph, the data have just been fitted, the program will ask whether you wish to show the fitted curve on the graph.



Select **Yes** to add the curve. Theoretical curves can also be added explicitly by the **Add Curve...** command, as is described in Chapter 13.

Creating a Histogram

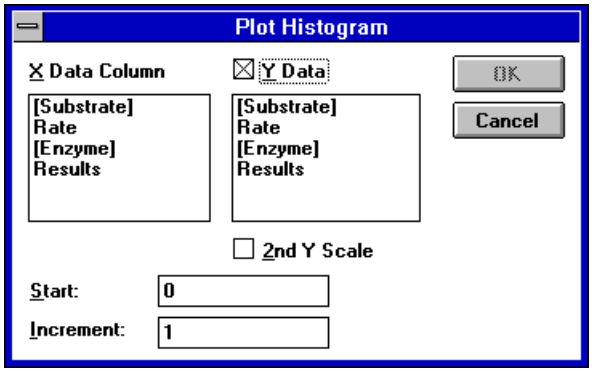
The **New Histogram...** command will create a new histogram plot. After selecting this command, you will have the opportunity to set the various aspects of the graph style from the following dialog box.



The style on which the graph is based is indicated at the bottom of the dialog box, and may be specified by pushing the **Style...** button. A full description of style selection is given in Chapter 20.

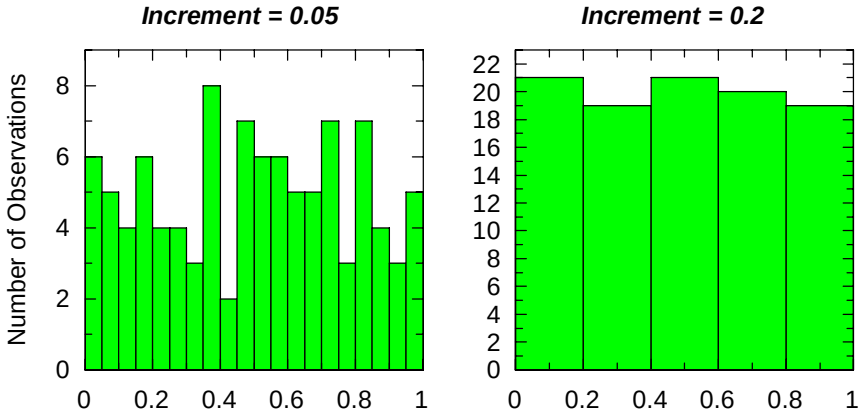
The Initial Settings region of the dialog box lets you initialize some commonly altered features of the graph, including whether or not the axes are logarithmic, or the scale should begin at x or $y = 0$.

After selecting the initial graph style, it is necessary to indicate which columns hold the data points. This is done from the following dialog box.



If the Y Data box is unchecked, it is assumed that there is one each of the *x* data values. If it is checked, the selected *y* column is taken to hold the number of entries corresponding to the values in the *x* column.

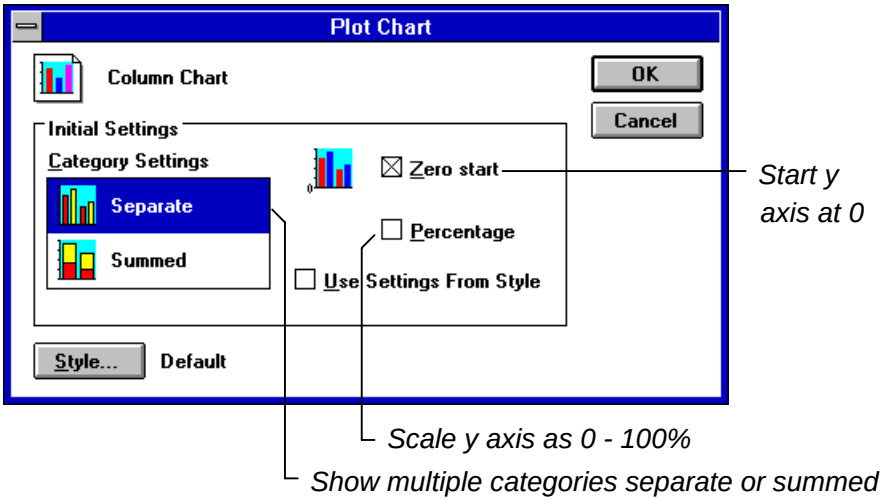
The histogram is grouped by taking values starting from the Start value, with the bars spaced as specified by the Increment value. The effect of changing the Increment value is illustrated below.



These group sizes can be altered later by double-clicking on the histogram symbol within the data legend (see Chapter 14).

Creating a Column Chart

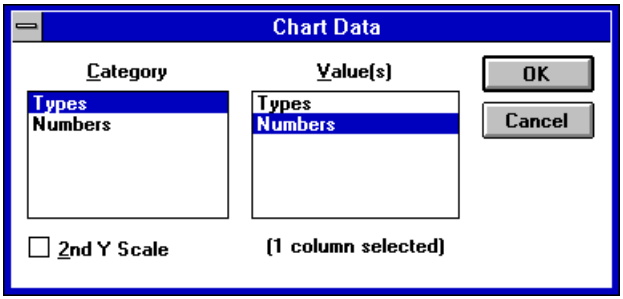
The **New Column Chart...** command will create a new column chart. After selecting this command, you will have the opportunity to set various aspects of the graph style from the following dialog box.



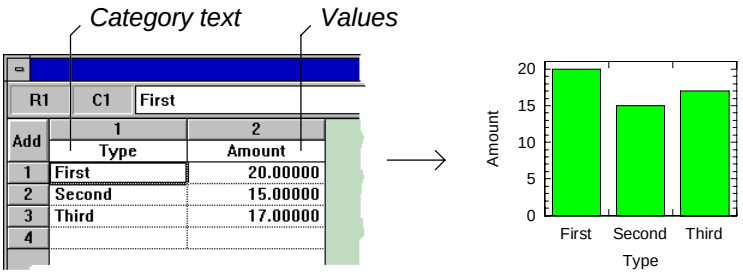
The style on which the graph is based is indicated at the bottom of the dialog box, and may be specified by pushing the **Style...** button. A full description of style selection is given in Chapter 20.

The Initial Settings region of the dialog lets you initialize some commonly altered features of the graph.

After selecting the initial graph style, it is necessary to indicate which columns hold the data. This is done from the following dialog box.



The column that holds the category text is selected, along with the column that holds the numeric values.

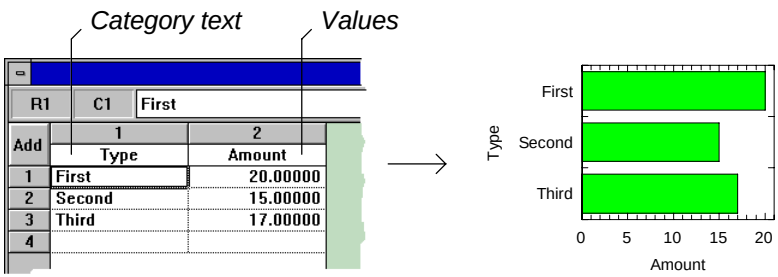


If the category column contains numbers, the format of the number in the data table is used when displaying the value on the graph. If necessary, the format of the numbers in the category column can be changed prior to plotting the graph — see Chapter 28 for details.

More than one value column may be chosen by holding down the CTRL key when making the selection. After pressing **OK** a new Graph window will be created that holds the graph.

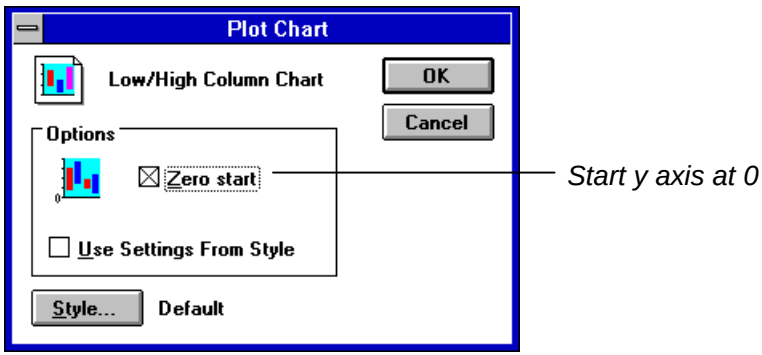
Creating a Bar Chart

Bar charts are similar to column charts except the category axis is the *y* axis of the graph. To create a bar chart, use the **New Bar Chart...** command. The process of making a bar chart is then almost identical to that described above for column charts.



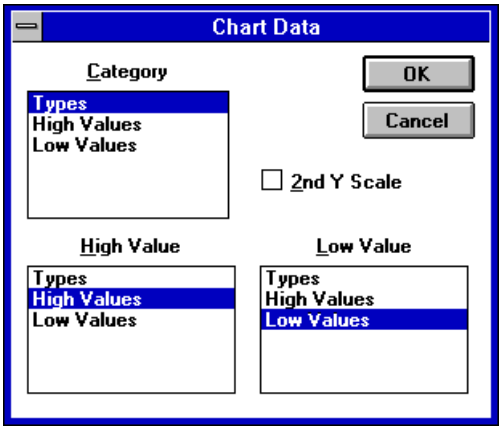
Creating a Low/High Column Chart

The **New Low/High Column Chart...** command will create a new column chart that has upper and lower limits to the columns. After selecting this command, you will have the opportunity to set various aspects of the graph style from the following dialog box.

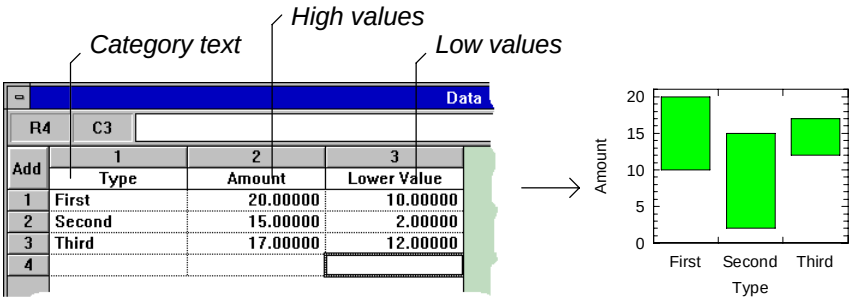


The style on which the graph is based is indicated at the bottom of the dialog box, and may be specified by pushing the **Style...** button. A full description of style selection is given in Chapter 20.

After selecting the initial graph style, it is necessary to indicate which columns hold the data. This is done from the following dialog box.



The column that holds the category text is selected, along with the columns that hold the high and low numeric values.

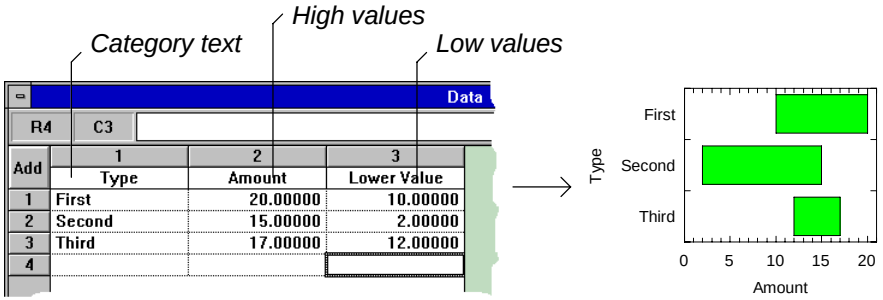


If the category column contains numbers, the format of the number in the data table is used when displaying the value on the graph. If necessary, the format of the numbers in the category column can be changed prior to plotting the graph — see Chapter 28 for details.

After pressing **OK** a new Graph window will be created that holds the graph.

Creating a Low/High Bar Chart

Low/high bar charts are similar to low/high column charts except the category axis is the *y* axis of the graph. To create a low/high bar chart, use the **New Low/High Bar Chart...** command. The process of making a low/high bar chart is then almost identical to that described above for low/high column charts.



Creating a Blank Graph

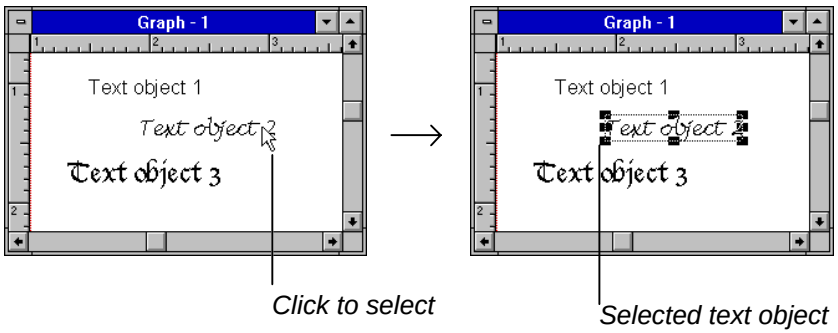
It can sometimes be useful to create a new graph window that has no graphs. Graphs, pictures and text can be added later, as required. To create a new blank graph, use the **New Blank Graph** command.

Graph Objects

GraFit graphs comprise a series of independent objects that can be selected, moved, sized and edited using the mouse. Graph objects include the graphs themselves, plus any text, lines, boxes etc. that annotate the graph. However, even a simple graph is not a single object — each graph is made from many parts (for example the legend, the scale, the axis labels), each of which can be selected, moved and edited.

Selecting an Object

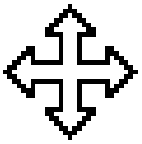
To select a graph object, move the mouse pointer over the object and click the left mouse button. The selected object will be marked by a dotted box.



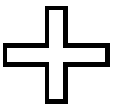
Types of Selected Object

Movable and Non-Movable Objects

When the cursor is over a selected region, the shape of the cursor changes. The shape indicates whether the selection can be moved.



Movable Region

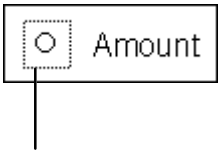


Non-movable Region

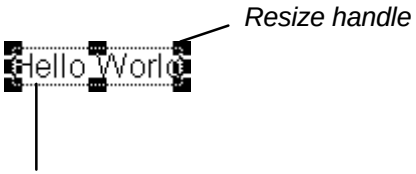
Regions that can be selected, but which cannot be moved, are usually part of a larger object. For example, the axis scale of a graph can be selected, but only the entire graph can be moved.

Resizable and Fixed Size Objects

Some graph objects have a fixed size, which cannot be altered. Others can be resized as required. Resizable objects, when selected, have square black 'handles' along the edges and corners of the selected region.



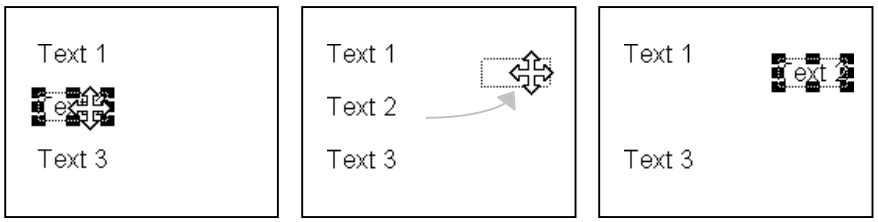
Selected data symbol
Cannot be resized



Selected text
Can be resized

Moving an Object

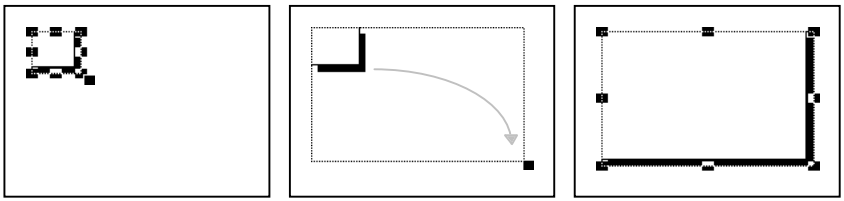
To move a movable object, hold down the left mouse button when the cursor is over the selection, and use the mouse to drag the object to the required location. Release the button when the object is in position.



1. *Move the cursor over the selection*
2. *Drag the selection to the new position*
3. *Release the mouse button*

Resizing an Object

To change the size of a resizable object, move the mouse cursor over one of the resize handles, hold down the left mouse button and drag the outline until it is the required size.



1. Move cursor over the resize handle

2. Drag to give the required size

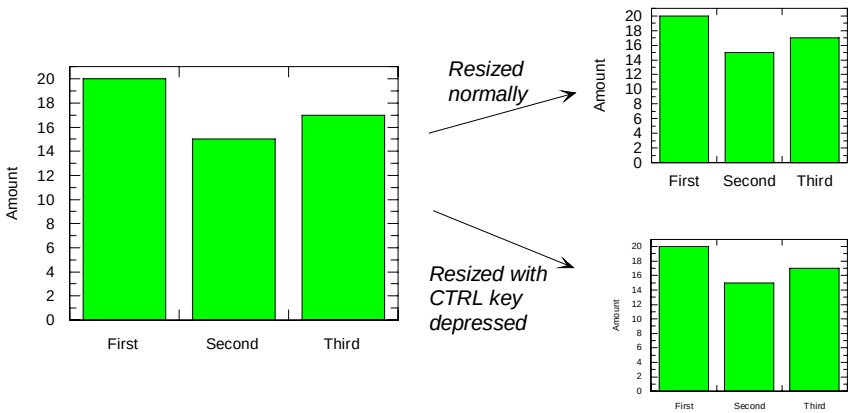
3. Release the mouse button

Maintaining the Aspect Ratio

To maintain the aspect ratio of an object (i.e. to keep the relative width and height the same after resizing), hold down the SHIFT key during resizing.

Changing the Size of Text When Resizing

By default, changing the size of the selected region does not alter the size of any associated text. However, if the CTRL key is depressed during resizing, any text associated with the selection will also be resized.



Editing the Object

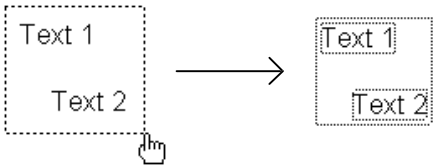
To alter the way that the selected object is displayed, double-click on the object. This will produce a dialog box specific to the selected object, which will allow various aspects of the object to be edited.

Moving Groups of Objects

A number of graph objects can be moved together as a group. It is possible to group several objects in two ways.

1. Dragging the mouse

Hold down the left mouse button over a blank region of the graph, and drag to enclose the required objects.



Drag outline to enclose objects

Grouped objects

2. Holding down the SHIFT key when selecting

If the SHIFT key is pressed when selecting an object, the object is added to the current group.

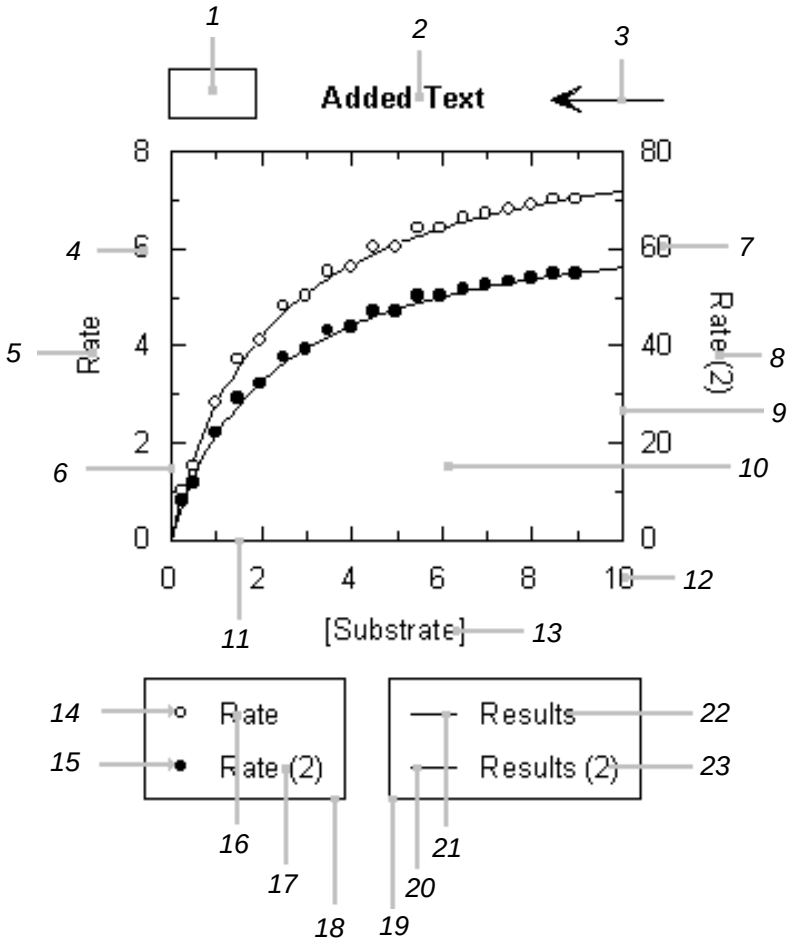
To move a group of objects, drag the outline around the group to the required location.

Aligning Objects

Objects that have been selected as a group as described above can be aligned using the **Edit Align...** command. Choose the alignment required (top, bottom, left, right etc.) from the resulting dialog box.

Selectable Graph Regions

Graph windows have many regions that can be selected and modified according to your choice and application. These are summarized on the figure below. The relevant chapter that describes how these different regions are edited is given in the table that follows.

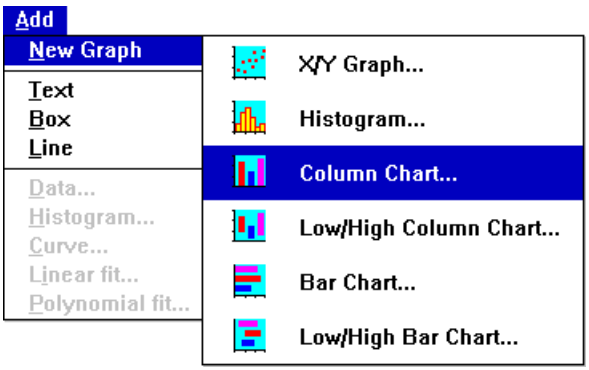


#	Region	Editable Properties	Chapter
1	Box	Outline style and color, background color, shadow style and color.	17
2	Added text	The text itself, text size, font style, color, alignment, orientation.	12
3	Line/Arrow	Line style and color. End style (arrows etc.).	18
4	*Y* axis scale numbers	Text size, font style, color, alignment, orientation. Number formatting.	11
5	*Y* axis legend	The text itself, text size, font style, color, alignment, orientation.	12
6	*Y* axis	Axis scale selection, tick mark options, line styles, axis offset, axis breaks, color.	10
7	Second *y* axis scale numbers	Applies to the second *y* axis (if present). Text size, font style, color, alignment, orientation. Number formatting.	11, 16
8	Second *y* axis legend	Applies to the second *y* axis (if present). The text itself, text size, font style, color, alignment, orientation.	12, 16
9	Second *y* axis	Applies to the second *y* axis (if present). Axis scale selection, tick mark options, line styles, axis offset, axis breaks, color.	10, 16
10	Main graph area	Background color. Charts: category spacing, display type, 3D effect.	9
11	*X* axis scale numbers	Text size, font style, color, alignment, orientation. Number formatting.	11
12	*X* axis legend	The text itself, text size, font style, color, alignment, orientation.	12
13	*X* axis	Axis scale selection, tick mark options, line styles, axis offset, axis breaks, color.	10
14, 15	Data symbol	Data symbol style, color, joining lines, data interval (histogram), display style (charts), error bar settings.	14, 15 (errors)
16, 17	Legend text	The text itself, text size, font style, color.	14
18	Legend box (data)	Outline style and color, background color, shadow style and color.	14

19	Legend box (curves)	Outline style and color, background color, shadow style and color.	14
20, 21	Curve symbol	Curve style, color, display range, parameter values.	14
22, 23	Legend text	The text itself, text size, font style, color.	14

Adding Graphs to an Existing Graph

To add a further graph to a graph window, use the **Add New Graph** command, and select the type of graph required.



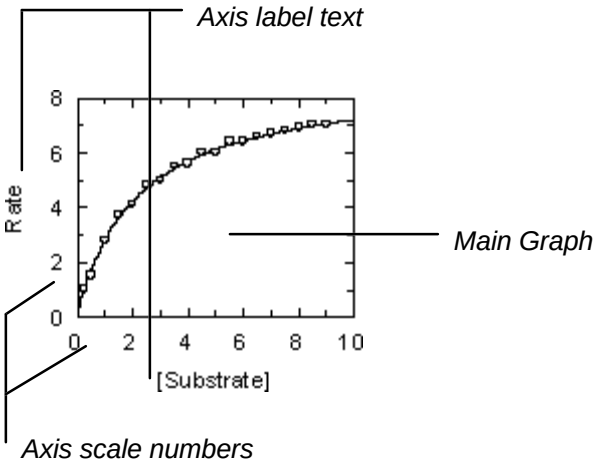
Add New Graph menu

The additional graph is specified in a similar manner to when creating a new graph window. After adding the graph, position and size it as required. Up to 64 separate graphs can be added to each graph window.

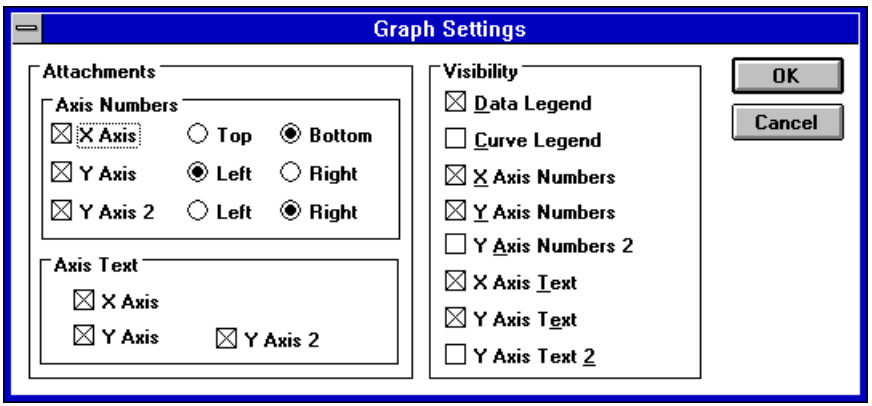
When only one graph is present, all menu commands apply to this graph. However, when several graphs are present it is often necessary to indicate which graph is to be acted upon. This is done by selecting part of the required graph prior to choosing the command.

Altering the Graph Attachment and Visibility Settings

When a graph or chart is created using the default styles, whenever the main graph is moved, the surrounding axis text and scale numbers are also moved to remain adjacent to their axes.



This is because the axis numbers and text are specified to be attached to the main graph. However, it is also possible to position these various elements independently. This is done using the **Options Graph Settings...** command, which results in the following dialog box.



Using the settings in the Attachments portion of this dialog box it is possible to select whether the different axis scale numbers and text are locked to the main graph or freely positioned. It is also possible to specify whether these labels are placed at the left, right, top or bottom of their respective axes.

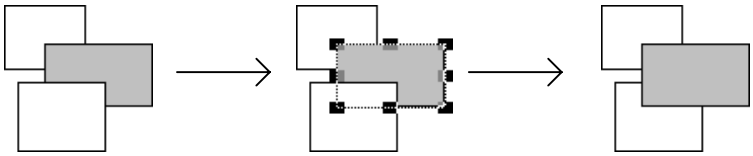
Attachment settings can also be set be selecting an attachable region, and choosing the **Edit Attach to Graph** or **Edit Detach from Graph** command.

Hiding or Showing Parts of the Graph

It is sometimes not desirable to display all the parts of a graph. The Visibility section of this dialog box controls whether various parts of the graph are displayed or not.

Ordering of Graph Objects

The various objects that comprise a graph are drawn in a series of layers. Objects that are "on top" overlay objects that are "underneath". To alter the position of an object it is necessary to select it and choose the **Edit Bring to Front** or **Edit Send to Back** commands.



Select box *Bring to Front*

When new objects are added to a graph, they are automatically added on top of all the other graph objects.

Printer Limitations

When one object is drawn on top of another on screen, the effect is to obscure the background object. For certain printers, however, it is not possible to overdraw underlying objects in this way. Plotters will draw all objects in the order they occur, but the images on top will simply overlay items underneath. Another situation where it is impossible to obscure underlying objects is when using in-built fonts with certain printers, including the HP LaserJet II.

Deleting Graph Objects

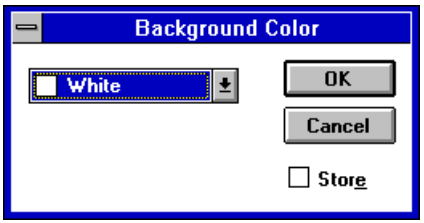
The **Edit Clear** command, or alternatively the DELETE key, can be used to delete graphs or graph objects. The region to be deleted must first be selected using the mouse. Portions of the graph that may be deleted in this way are:

- Added text
- Pictures
- Added boxes
- Added lines or arrows
- Whole graphs (when any part of the graph is selected)
- Data sets (when the appropriate data symbol is selected)
- Curves (when the appropriate curve symbol is selected)

7 Graph Page Settings

Changing the Graph Window Background Color

By default, the background color of the graph window is that specified for window text backgrounds by the Windows Control Panel application. This color may be altered to represent more closely the background color of the output device. To alter the background color, double-click on the background part of the graph window to give following dialog box.

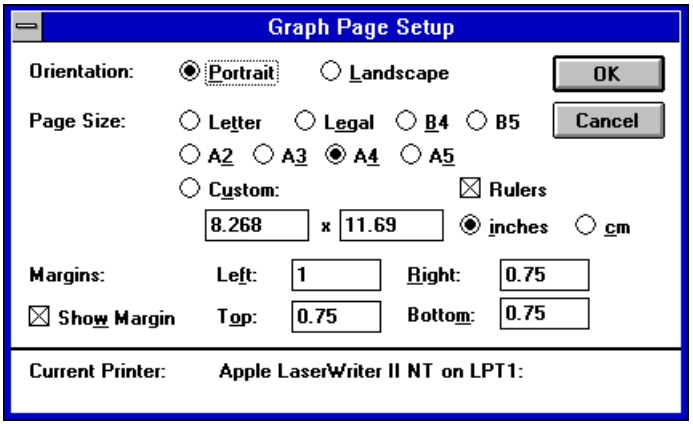


Select the required color from the drop-down list box (see Chapter 8 for a description of how colors are selected in GraFit). If the Store box is checked, the selected color will become the default when GraFit is next started.

Note The background color is for display purposes on-screen only, and changing this color will not result in the printed output background being altered. If you do wish to alter the background color when printing, this can be done by adding an opaque box of an appropriate color as a backdrop to the graph — see Chapter 17 for details of how to add boxes to a graph.

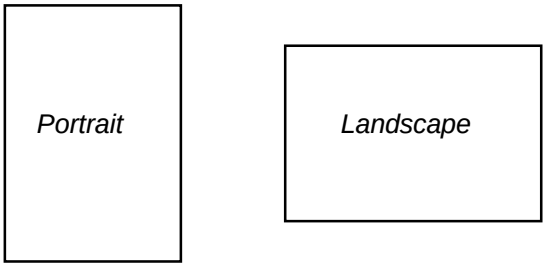
Setting the Graph Page Size

When a graph is created, the page size and orientation are set to match those of the currently selected printer (see Chapter 5 for details on selecting printers). It is possible to change these settings using the **File Graph Page Setup...** command. This results in the following dialog box.



Graph Orientation

The graph orientation is selected from Portrait or Landscape. Note that it is essential to set the printer page orientation to match that of the graph as required.



Page Size

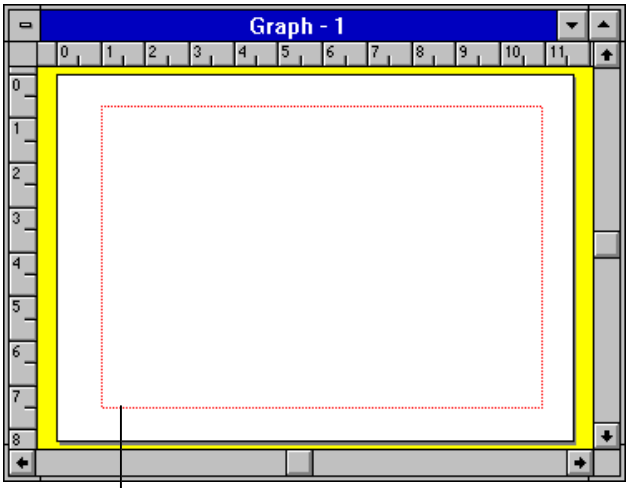
Select the page size from those listed, or enter custom sizes as required.

Scale Units

Choose from centimeters or inches. The units selected will be used to display the rulers on the graph window.

Margin Settings

The margin settings can be altered as required by entering values into the edit boxes.



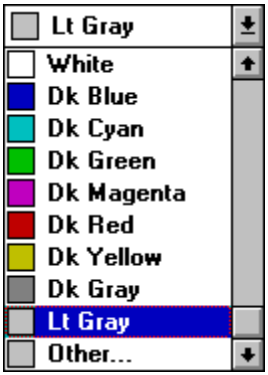
Margins are displayed on the graph solely to assist in positioning graphs so that they are not placed too close to the edge of the page. The margin lines are not printed, and can be turned off if required by unchecking the Show Margin box.

8 Use of Color in Graphs

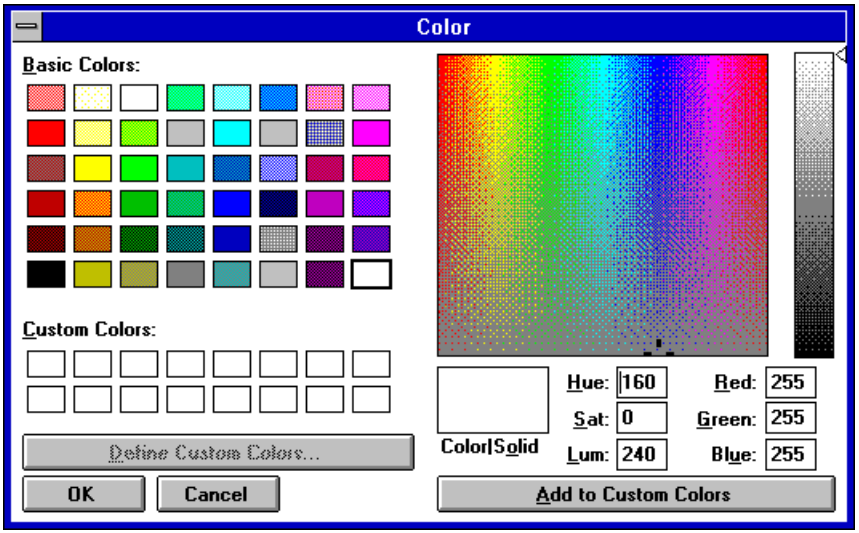
Use of Color

GraFit allows full use of color in your graphs. Colors are selected from the Windows palette of 24 million colors. The number of colors actually available depends upon the type of device (monitor, printer) you are using; the standard EGA and VGA displays provide sixteen different colors, and extended color modes are available from some manufacturers to give still higher resolutions. To ensure compatibility with all devices, GraFit allows access to the full spectrum of colors, but also provides convenient short cuts to the common primary colors.

The colors for various regions of the graphs are selected from color drop-down list boxes, which are found in many dialog boxes for editing parts of the graph.



The main available colors are listed, and can be selected as required. It is also possible to choose a custom color by selecting the **Other...** option. This produces the color selection dialog box.



Choose the exact color from the spectrum available, or enter precise Red Green Blue (RGB) or Hue Saturation Luminosity values for the color required.

Printing Colors on a Monochrome Printer

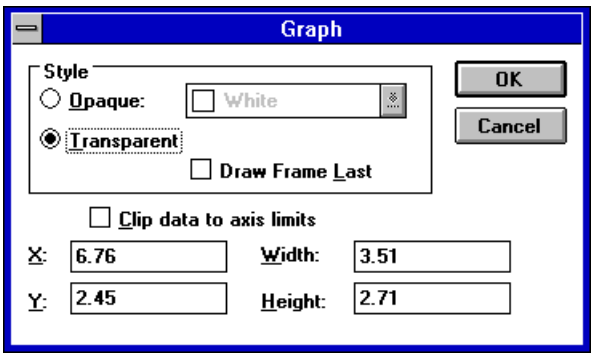
When colors are printed on a monochrome printer, the colors are translated into shades of gray.

9 The Main Graph Region

Selecting the main graph itself allows the graph to be positioned and resized. In addition, several settings relating to the graph display are configured by double-clicking on the graph.

X/Y Scatter Graphs and Histograms

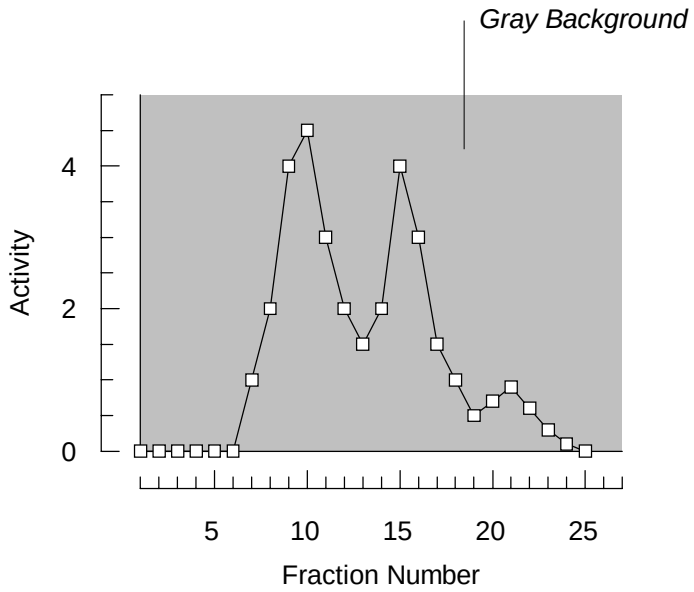
Double-clicking on one of these graph types results in the following dialog box.



Graph Style Settings

Background

Graphs can be drawn with an opaque or a transparent background; if Opaque is selected it is possible to choose the color from the color drop-down list box. In the graph below, the graph background has been set to light gray. The axes have been offset (see Chapter 10) to emphasize the effect.



Order of Drawing Data and Frame

If, as is the default for a scatter graph, the Draw Frame Last box is unchecked, the data points will be drawn after the frame and axes. Checking this box results in the frame and axes being drawn over any data, and is the default for histograms and charts.

Error! Objects cannot be created from editing field codes.

Data Clipping

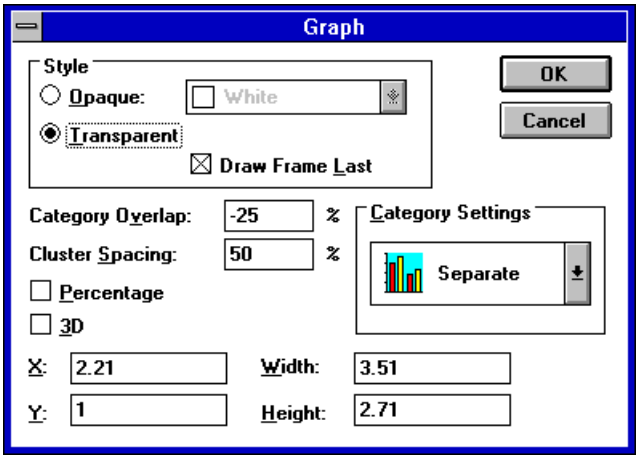
When data are plotted on a graph, it is possible to set the scale so that some of the data lie outside the selected axis limits. Unless the Clip Data to Axis Limits box is checked, GraFit will allow the display of data points that lie just outside the axis limits, but will always clip data points to the limits of the main graph region. By checking the Clip Data to Axis Limits, the display of data points is more restrictive, and only those points that lie within the specified axis limits will be displayed. By default this box is unchecked.

Graph Size and Position

The graph size and position can be specified precisely by entering values into the edit boxes. The measurements are given in the currently selected graph units, as chosen by the **File Graph Page Setup...** command (Chapter 7).

Column and Bar Charts

Double-clicking on one of these graph types results in the following dialog box.



Graph Style Settings

These are specified as described above for *x*/*y* scatter graphs.

Category Overlap and Cluster Spacing

These settings control the display of the columns or bars, as shown below for a column chart.

Error! Objects cannot be created from editing field codes.

Category Overlap

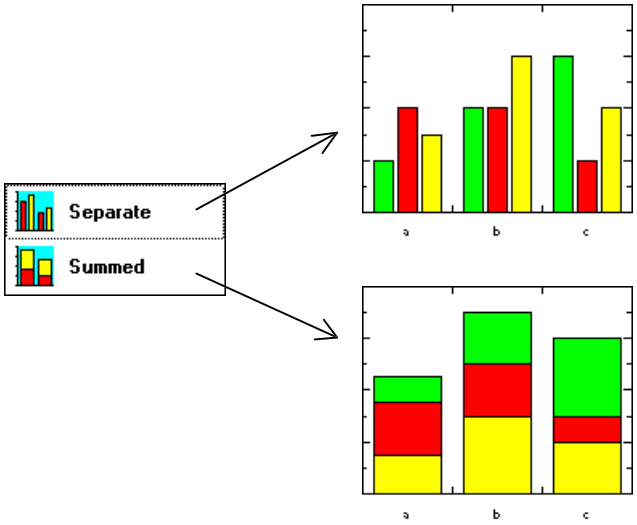
This indicates the degree of overlap for multiple values of the same category. A value of 100% gives complete overlap. Negative values result in a space between the values, for example, -25% means that the values are separated by a quarter of the width of any one column.

Cluster Spacing

This indicates the separation between the last value of the cluster and the edge of the category division

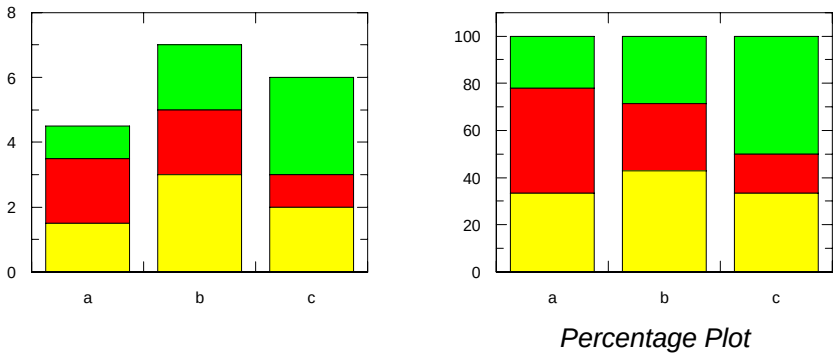
Category Settings.

Categories may be displayed as separate values, or summed. Choose the desired style from the drop-down list box.



Percentage

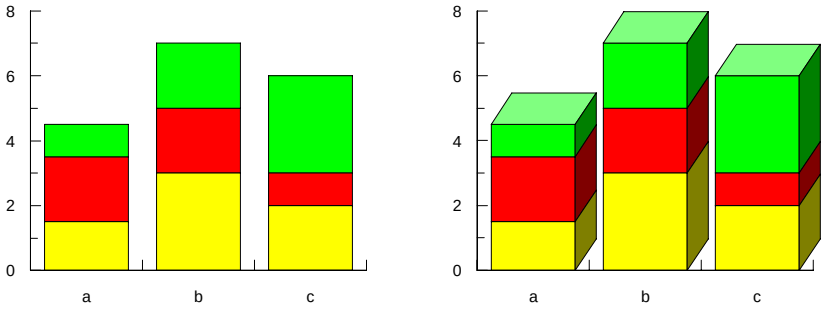
If the percentage box is checked, the values in each category are scaled so that they represent percentages of the total.



Percentage Plot

3D

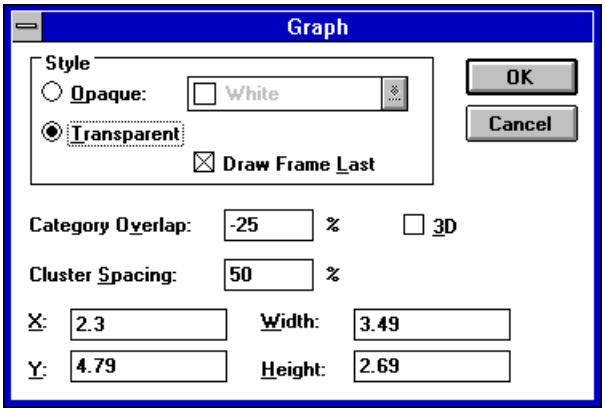
If the 3D button is checked, the graph is displayed with a 3D effect.



3D Effect

Low/High Column and Low/High Bar Charts

Double-clicking on the graph results in the following dialog box.



Graph Style Settings

These are specified as described above for *x*/*y* scatter graphs.

Category Overlap and Cluster Spacing

These are specified as described above for bar and column charts.

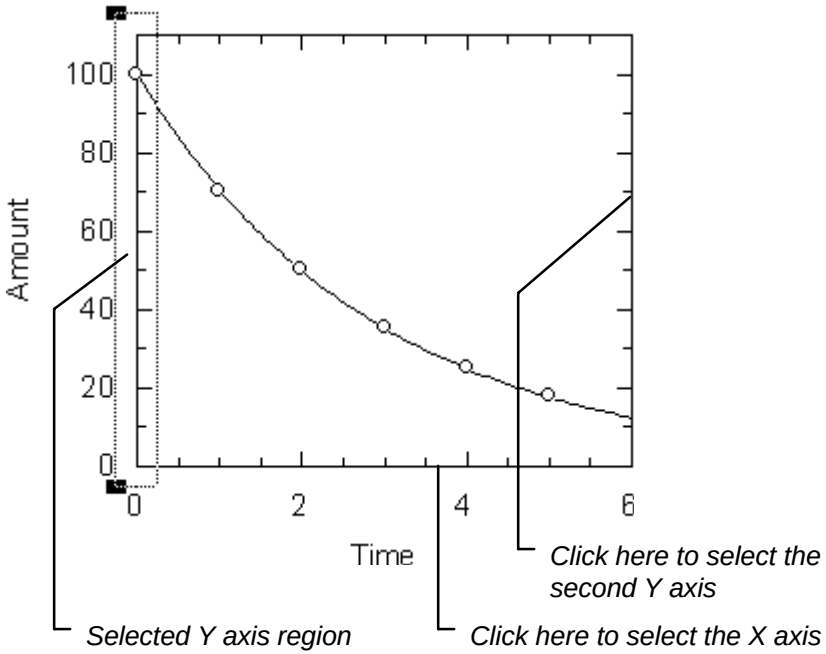
3D Effect

Checking the 3D box results in a 3D effect to the displayed graph.

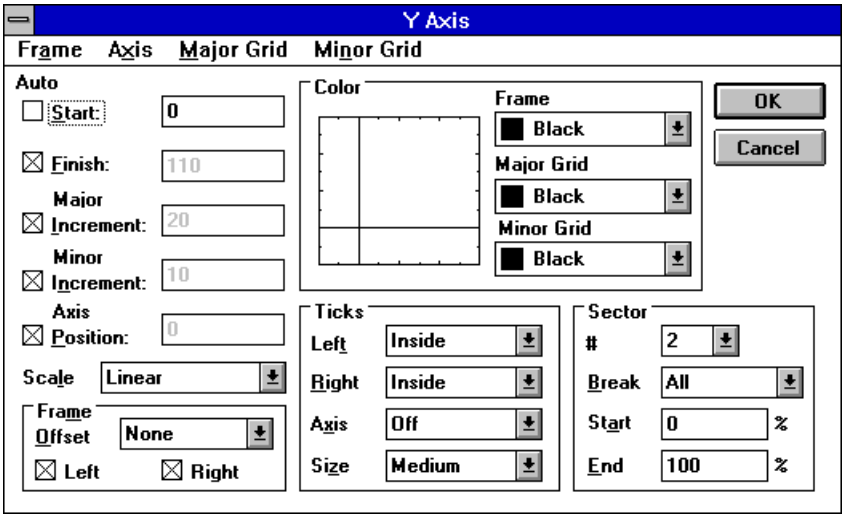
10 Graph Axis Scales and Tick Marks

Selecting the Axis Regions

To select the graph axis scale, click near the left edge of the graph (*y* axis), the right edge of the graph (second *y* axis [if present]), or the upper or lower edge (*x* axis) of the main graph. A single click will select the axis region, as shown below for the *y* axis.



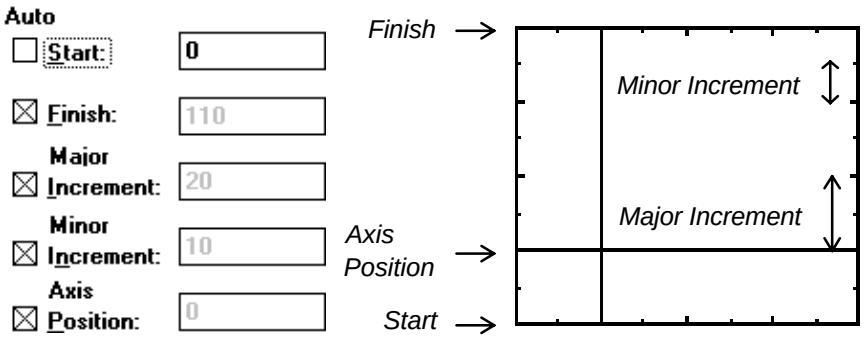
Note that selected axis regions cannot be resized (it is necessary to resize the graph as a whole) or moved (although it is possible to move the whole graph if required). Double-clicking on the selected *y* axis region produces the following dialog box.



This dialog box controls all aspects of the *y* axis scaling and style ranging from the scale of the axis to the type of tick marks used to annotate the graph. Similar dialog boxes are produced when double-clicking on the *x* axis region, or the second *y* axis region.

Altering the Scale

The scale of the axis is specified within the scale selection portion of the axis dialog box. Scaling may be performed automatically, by checking the Auto boxes, or manually by unchecking one or more of these boxes and entering appropriate values into the edit boxes.



Five different aspects of scaling can be specified.

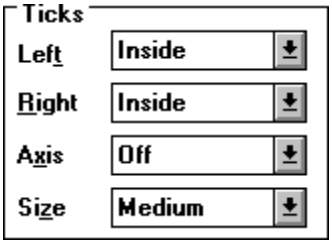
1. The start position of the axis.
2. The finish position of the axis.
3. The spacing of the major increment. The major increment specifies how frequently the axis is numbered. A large tick mark is drawn on the axis to show the number position.
4. The spacing of the minor increment. Between the major increment tick marks are placed smaller ticks spaced as defined by the minor increment size.

5. The position where the axis line is drawn. This is normally drawn at 0 or at the start position, but may be altered as required.

Any number of these parameters may be set automatically by checking the appropriate boxes. When the dialog box is first displayed the numbers in the edit boxes show the current values for the scaling parameters.

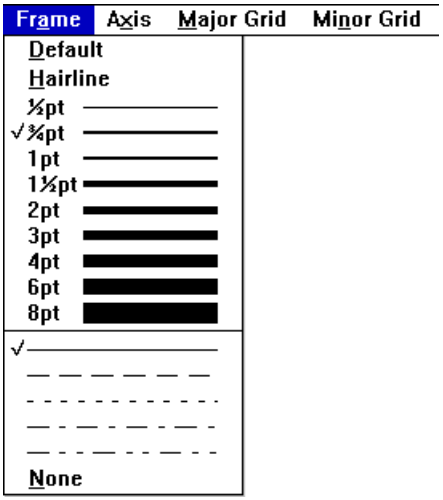
Altering the Tick Marks

Tick marks are drawn on the frame (the perimeter) of the graph, and/or on the axis line. They are used to indicate the major and minor spacing increments as defined above. Ticks can be drawn on any side of the graph, or be switched off as required. These selections are made from the Ticks section of the axis dialog box.

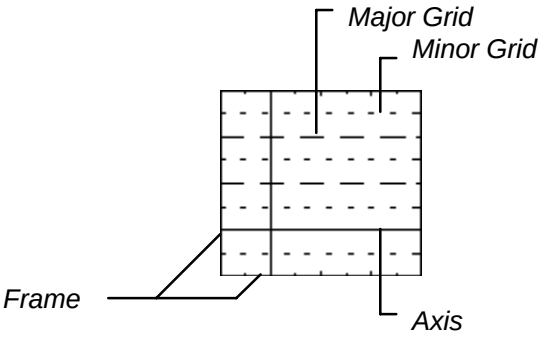


Changing the Line Style or Thickness

The graph frame, axis, tick marks and grid lines can be drawn in various line styles using pens of several different thicknesses. Selection of these styles is from the menu bar of the dialog box, which allows individual styles to be specified for the **Frame** (the box that is drawn around the graph; this style also applies to the tick marks drawn on this frame), the **Axis** line, the **Major Grid** lines and the **Minor Grid** lines. Each of these selections has a similar menu.



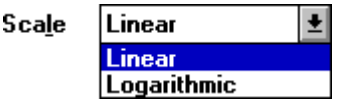
This menu allows the thickness and style of the line to be chosen. It also allows the specified line to be switched off, as is the default for both the grid lines. The menus are divided into two regions. The top part specifies the thickness of the line. The thickness is given in points (a printing measurement, there are 72 points to the inch). Two other sizes may be selected: hairline, which the thinest line available; and default, the result of which depends upon the printer in use but which is usually one pixel wide. The current setting is shown ticked. The lower part of the menu is used to specify the line style. Choose from solid or various styled lines, or select 'None' if no line is to be drawn. For the *y* axis, the line styles apply as shown below.



The settings specified are for the specific axis selected, in this case the *y* axis. To alter the *x* axis settings it is necessary to activate the dialog box by double-clicking on this axis.

Selecting Logarithmic Axes

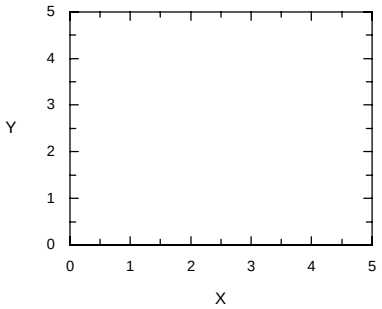
The axis may be scaled linearly (the default) or logarithmically.



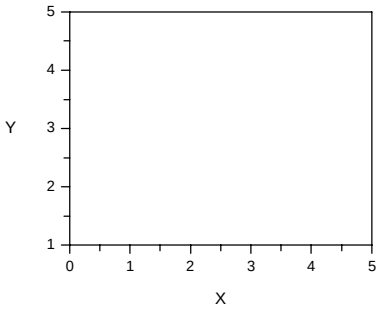
Select the Linear or Logarithmic ($\log_{10}$) option as required.

Examples of Different Axis Styles

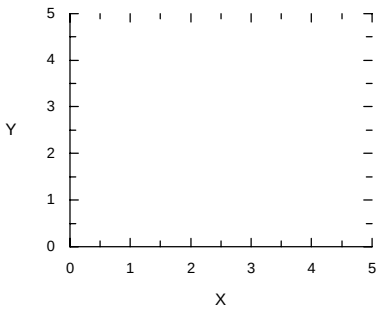
The various options available allow many different styles of graph to be drawn. Below is a selection of some of the different styles that can be produced.



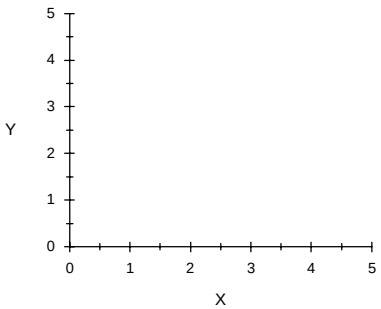
Default style



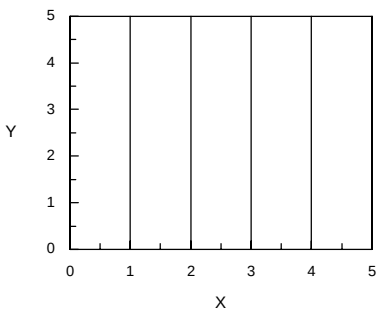
Ticks on outside of axes, frame drawn



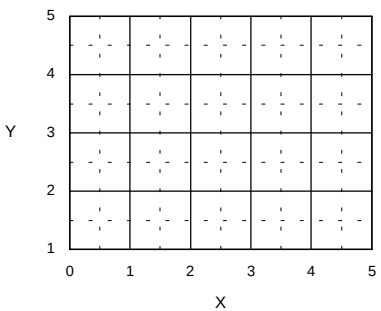
Ticks inside all edges, axes drawn, frame off



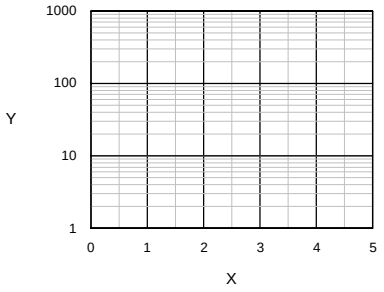
Small ticks on both sides of axes, axes drawn, frame off



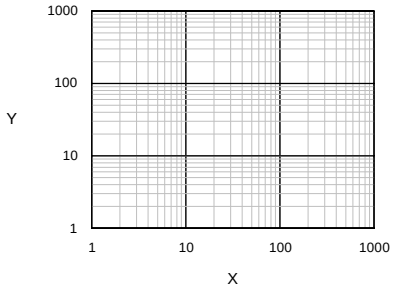
Solid lines for major grid on X axis



Solid lines for major grids, dotted lines for minor grids



Light gray solid lines for minor grid, Y axis logarithmic

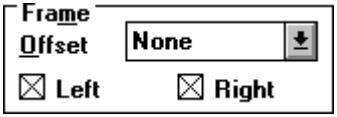


Both axes logarithmic

It is possible to save different axis styles, and apply these saved styles to a graph, using the **Options Save Style...** and **Options Apply Style...** commands. These are described in Chapter 20.

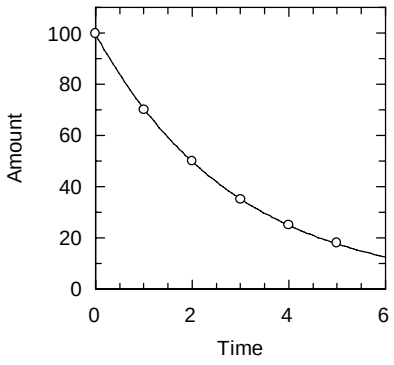
Frame Drawing and Offset

The graph frame comprises the four edges around the graph. The display of these can be switched on or off individually, and they can also be offset from the main graph body. From the *y* axis dialog it is possible to control whether the left and right edges are drawn; the *x* axis dialog controls the top and bottom edges. The relevant portion of the *y* axis dialog is shown below.

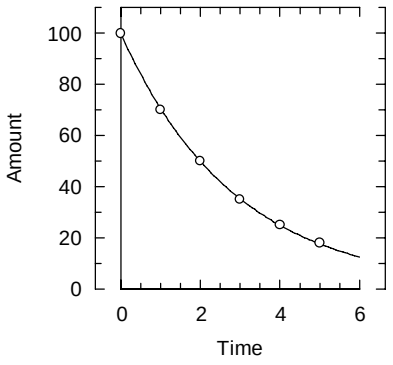


Checking or unchecking the Left and Right boxes controls whether these edges of the frame are drawn.

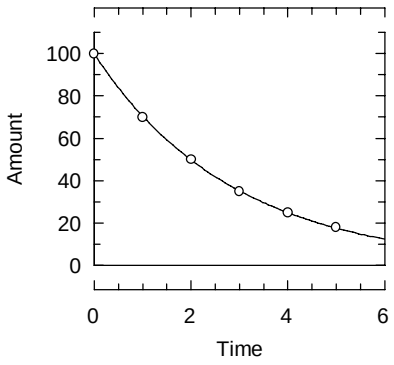
The Frame section also sets an optional offset of the frame edge from the graph body. Various amounts of offset can be incorporated, and the effect of offsetting the frame is shown below.



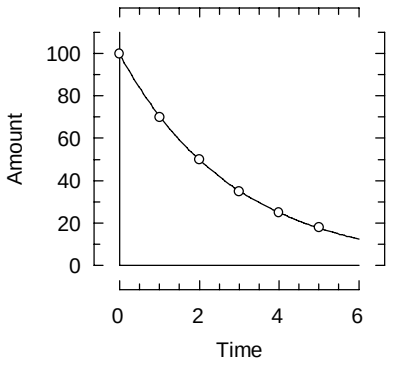
No offsets



Offset set from y axis dialog box



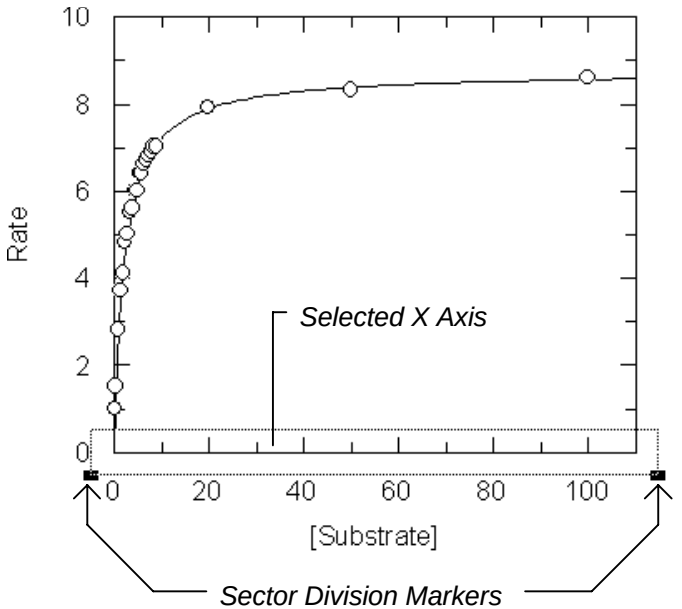
Offset set from x axis dialog box



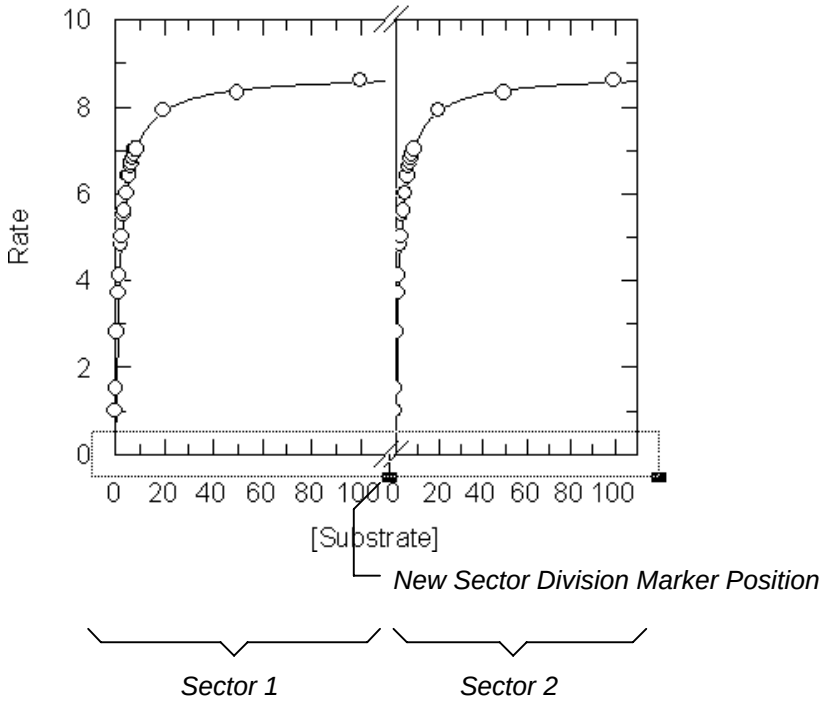
Both axes offset

Split Axis Scales

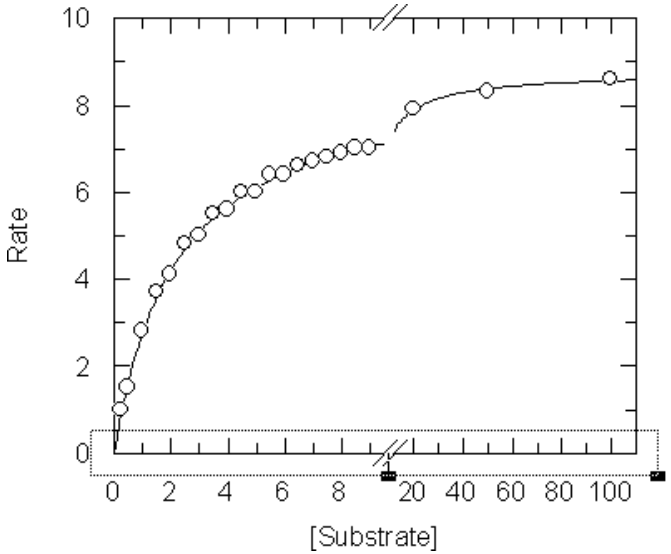
Any numeric axis (*x*, *y* or second *y*) may be split into two or three sections, each of which can be scaled independently. This is useful if the data cover a large range, and it is necessary to show the beginning and end at different magnifications. Below is shown a data set that has many data points at low *x* values, and fewer points at high *x*. The *x* axis has been selected.



At the ends of the selected axis are the sector division markers. Axis breaks are controlled by moving these markers to the place where the break is to be drawn. When they are placed at the ends of the axis, as is the default shown above, no axis breaks are present. To incorporate an axis break, drag either of the sector division markers to the location where you wish the break to occur; using two markers allows up to two breaks to be incorporated. After dragging the left marker, the screen appearance is as below.



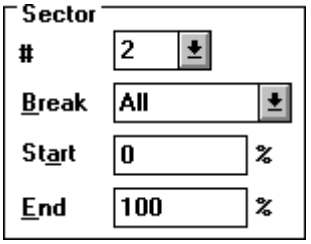
The result of this is to divide the *x* axis into 2 sectors; moving the right division marker would allow a third sector to be created. When first created, the sectors each have the same scale settings, and so the graph appears to be duplicated in the two regions. To produce a sensible plot it is necessary to set appropriate scales for each sector (the autoscaling options are not very suitable for selecting sensible scales when split axis ranges are present). This can be done by double-clicking on the selected axis within the sector required. Each sector can then be scaled separately. After setting scales for sector 1 of 0 to 9.5, and for sector 2 of 12 to 110, the effect on this graph is as shown below.



Both the *x* axis and the *y* axis or both may be divided into sectors using the sector division markers. The break position(s) can be adjusted as required by dragging the marker. To remove the axis break, drag the markers so that they are at the extreme edges of the axis selection.

Sector Options

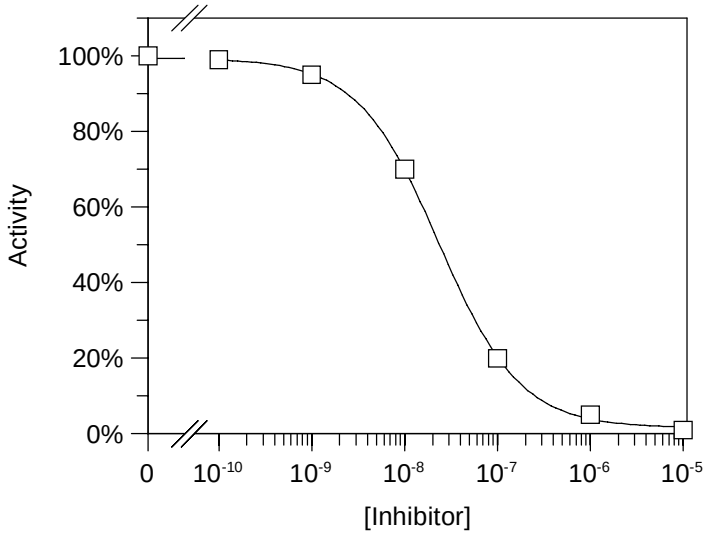
The Sector section of the axis dialog box allows certain sector drawing settings to be controlled.



The sector number (#) can be set from the top drop-down list box. The Break box controls if and where the axis breaks (the "//" marks) are displayed. The Start and End settings control the start and end positions of the plotted graph region within the designated sector. By default these are 0% to 100%, i.e. the graph occupies the whole sector. To prevent overlap between sector axis scale numbers it can help to set the sector start value to be slightly greater than 0 (say 5%) and/or the sector end value slightly less than 100 (say 95%). Changing these values can also be useful when using the second *y* scale (see Chapter 16).

Incorporating a Zero Value on a Logarithmic Plot

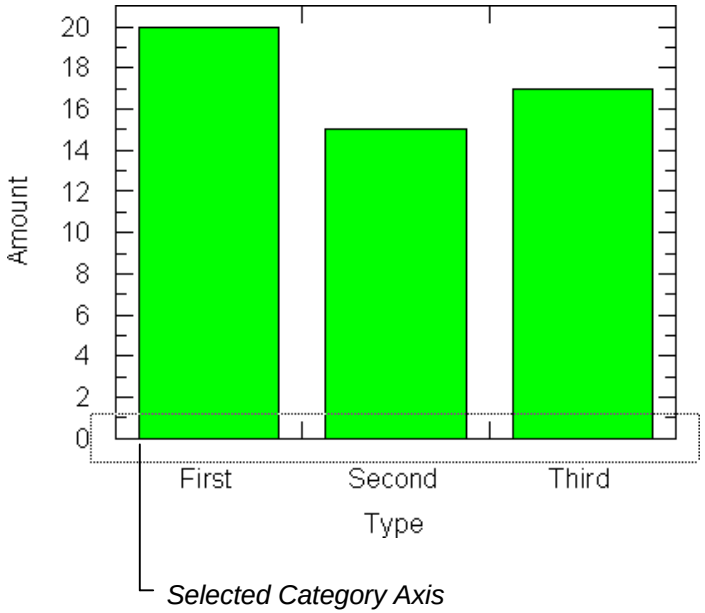
When displaying some types of data, it is sometimes useful to incorporate a zero value on a logarithmic axis. While, strictly speaking, this cannot really be done, it is common to display such a value, especially on dose-response curves. As the individual sectors of a GraFit graph can be scaled individually, this effect can be produced by using a split axis, having a small first sector that is scaled linearly (make sure that the axis Finish value and tick increments in this sector are such that only the 0 tick is drawn — in the example below the End value of the first sector was set to 1e-11), and a larger second section scaled logarithmically. After doing this it is possible to produce graphs such as the one below.



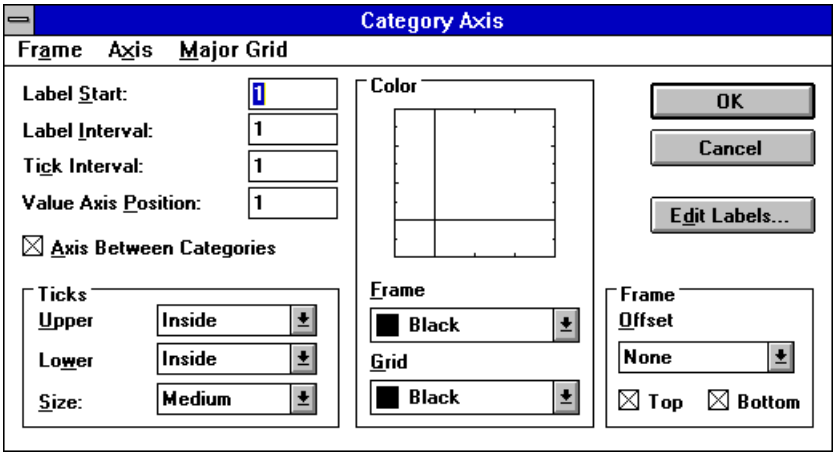
In this example, the Start value for the second *x* sector was set at 5% to improve the appearance of the graph.

Category Axis Scales

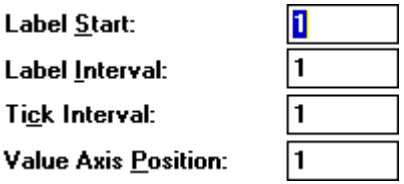
For charts, double-clicking the category axis controls the scale, tick and grid settings.



Double-clicking on the category axis results in the following dialog box.



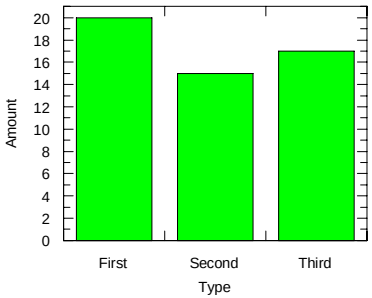
Altering the Scale



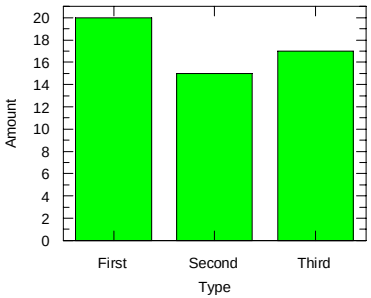
It is possible to select how often the category labels are displayed. By default all the available labels are drawn, and the Label Interval is set to 1. A setting of 3 would show every third label, etc. A Label Start of 2 combined with a Label Interval of 4 would display the 2nd, 6th, 10th... labels. The Tick Interval setting controls how frequently the ticks are drawn on the axis.

The Axis Value setting controls where the axis line is drawn. When the Axis Between Categories box is checked, the axis line is drawn through the center of the selected category — if this is not desired then the axis line can be disabled by setting the Axis Line style to None from the menu bar.

The Axis Between Categories box setting determines whether the tick marks are placed centrally, or between the categories.



Axis Between Categories Checked



Axis Between Categories Unchecked

Tick Settings and Frame Offsets

Tick settings and frame offsets are controlled in the same way as for numeric axes, as was described earlier in this chapter.

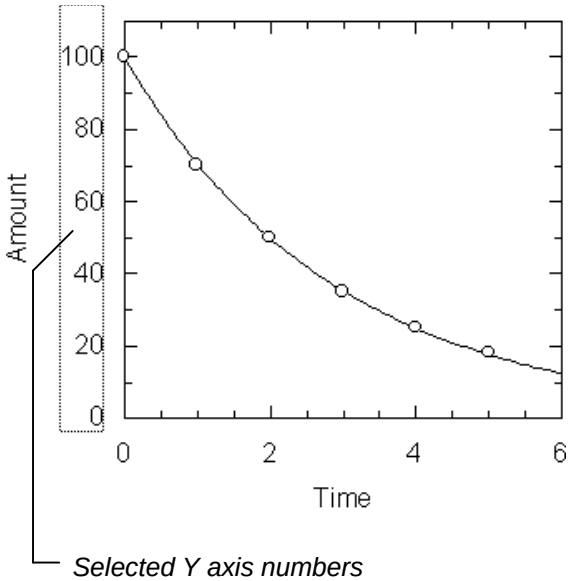
11 Scale Numbers and Category Legends

Configuring the Scale Numbers

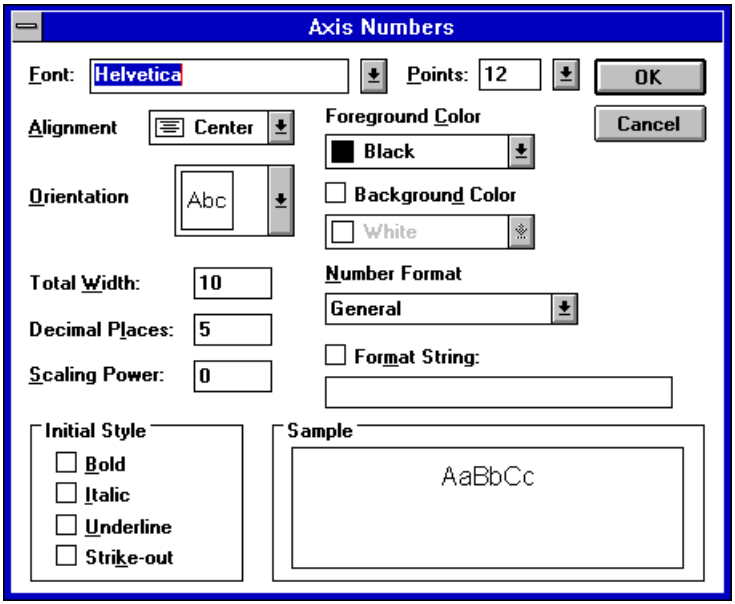
A numeric scale is present on both the *x* and *y* axes of a scatter graph, and on the value axis of a chart. It is possible to alter the position, font, color and formatting of these numbers. By default the axis number labels are anchored to the main graph, although they can be moved independently by use of the **Options Graph Settings...** command. However, their size is fixed by the size of the main graph and they cannot be sized independently, although altering the main graph size will result in the number labels being resized.

Y Axis Scale Numbers

Clicking on the *y* axis numbers will select the numeric scale, as shown below.



Double-clicking on *y* axis number region produces the following dialog box.



From this dialog box it is possible to configure the displayed text style and the manner in which the numbers are formatted.

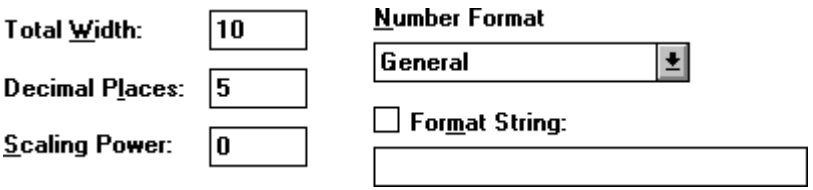
X Axis Scale Numbers and 2nd Y Axis Scale Numbers

The *x* axis numbers (and second *y* axis numbers if present) are selected and edited in a similar manner to that described above for the *y* axis scale numbers.

Altering Numbering Text Style

The style of text used to write the numbers is selected from the above dialog box as described in Chapter 12.

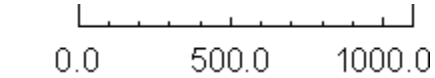
Number Display Format



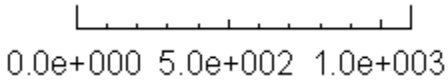
Number Format

The format used to display the axis scale can be selected from fixed decimal, exponential, general and power formats. Number formats are described in Chapter 28, and are controlled by the settings of the total field width and number of decimal places. The power format is used to display large or small numbers as a power of ten, for example "1 x 10^{-23}".

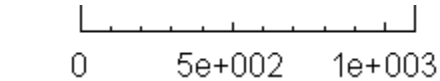
The following *x* axis scales show the effect of changing the number format type. The decimal places setting for each was 1.



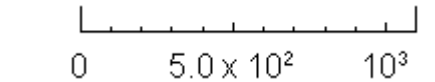
Fixed



Exponential



General



Power

Scaling Power

Besides the numeric format used, it is possible to enter a number for the 'Scaling Power'. The scaling power is used when the numbers plotted are very small or very large, and you do not wish to include the exponent in the scale. For example, if the axis was drawn between 40,000 and 80,000, and you wanted to label the axis 4, 6, 8 rather than 40,000, 60,000, 80,000 this could be done by specifying a scaling power of –4. The scaling power is *P*, where

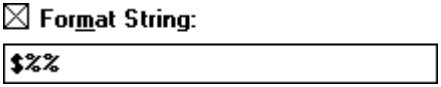
$$\text{Number Shown} = \text{Actual Number} \times 10^{P}$$

In this case, the number shown would be $40{,}000 \times 10^{-4} = 4$. Alternatively, you may have entered values in (for example) μs, and wish to incorporate the missing factor of 10^{-6} in your plot. In this case the scaling factor is –6.

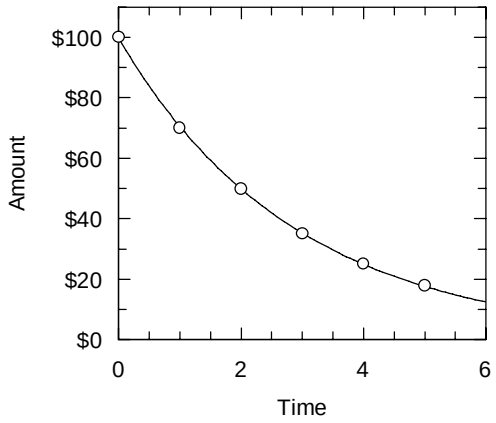
The scaling power must be a whole number, and may be positive or negative. A scaling power of 0 will result in the actual number being displayed ($10^0 = 1$).

Format String

When the Format String box is unchecked, the numbers are displayed normally. When this box is checked, the text entered into the Format String edit box is displayed in place of each number, with the first occurrence of the text string "%%" substituted by the appropriate number. For example, the following entry for the *y* axis numbers:

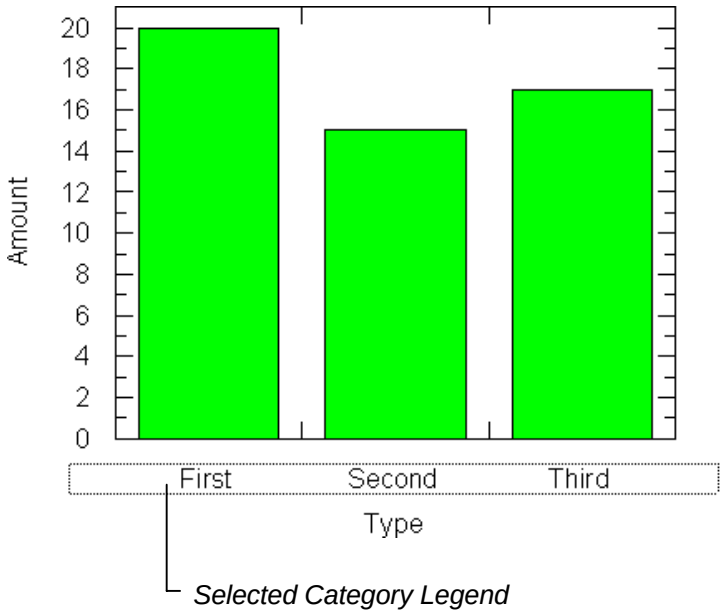


would result in the axis numbers being displayed as below.

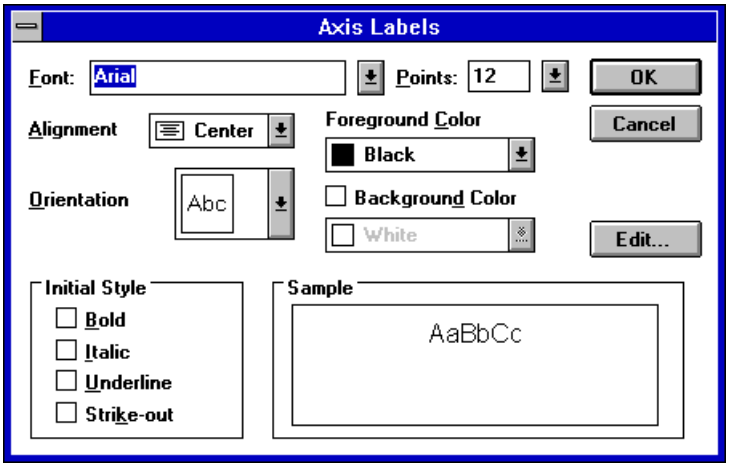


Configuring Category Legends

Charts have one numeric axis, and one category axis. The various categories have text labels that are displayed adjacent to the axis.



The display format, and label text, of these category legends can be altered by double-clicking on the selected legend. The resulting dialog box is shown below.

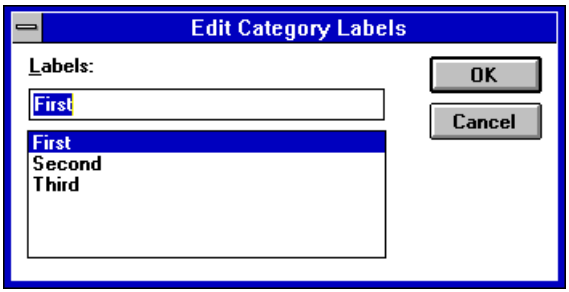


Altering Category Legend Text Style

The style of text used to write the category names is selected from the above dialog box as described in Chapter 12.

Editing the Category Legend Text

To edit this text, press the **Edit...** button to give the following dialog box.

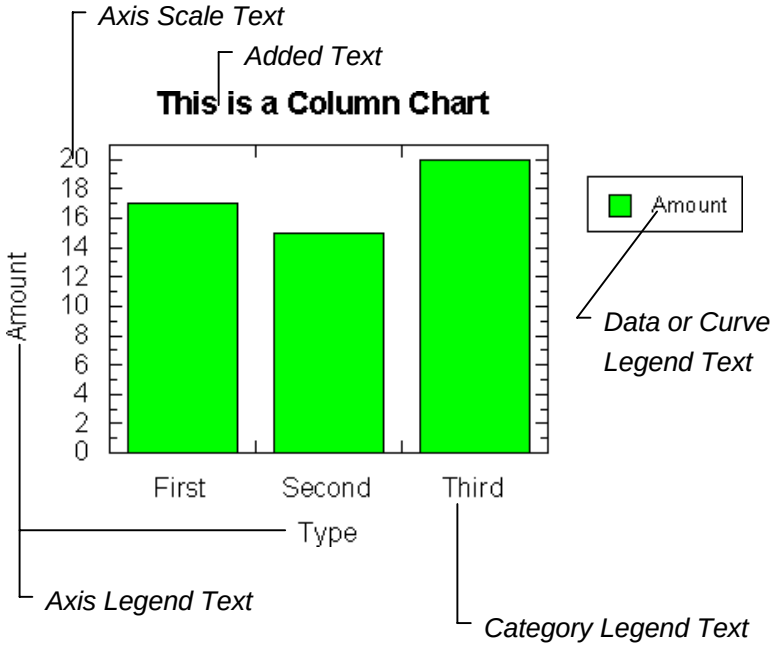


Click on the label to be edited, and enter the new label text. This text may include formatting commands for superscripts etc., as described in Chapter 12.

12 Graph Text

Text in Graphs

Textual information is present in various parts of the graph. The display of each type of text can be modified, and the settings of the font style, size alignment and orientation can be edited. The following regions of the graph contain text that can be selected.



In each case, double-clicking on the text produces a dialog box that allows the text style to be configured as well as making region-specific settings.

Specific Text Region Settings

Axis Legend Text

Double-clicking on the axis legend text allows the axis label to be edited as required.

By default, the axis legend text is locked in position, but may be positioned independently of the main graph if required (use the **Options Graph Settings...** command to change this).

Axis Scale Text

Double-clicking on the axis scale text allows the number format to be specified. See Chapter 10 for further details.

By default, the axis scale text is locked in position, but may be positioned independently of the main graph if required (use the **Options Graph Settings...** command to change this).

Category Legend Text

Double-clicking on the category legend text allows the legends to be edited. See Chapter 11 for further details.

By default, the category legend text is locked in position, but may be positioned independently of the main graph if required (use the **Options Graph Settings...** command to change this).

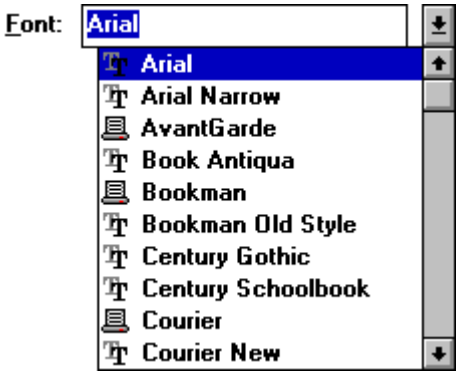
Data or Curve Legend Text

Double-clicking on the data or curve legend text allows the legends to be edited. See Chapter 14 for further details.

Text Style Settings

The style of any displayed text is altered by double-clicking on the text. This produces a dialog box, from which the style may be selected. All dialog boxes that control text display allow the style to be set in a similar manner, as described below.

Font Name



The font to be used can be chosen from the list or typed into the edit box. A small icon by the left of the font name shows the type of the font.

Icon	Type
	Device-specific font. These fonts are specific to the printing device in use. GraFit will use the most closely matching Windows font for the screen display, but it is possible that there will be some differences between the screen representation, and what is printed.
	TrueType font. TrueType fonts, which are only available in Windows version 3.1 or later, are fully scalable to any size. The fonts displayed on screen will closely match the printed output.
(none)	Windows vector font. Vector fonts are useful for plotters, but do not produce high quality output on matrix or laser printers.

All the currently installed Windows fonts are listed. To add further fonts to your system, consult your *Microsoft Windows User's Guide*.

Font Size

The size of the font used is specified in points. A point is a standard printing measurement; there are 72 points to the inch. Font sizes can be entered to a resolution of 0.1 points, although on some devices only certain discrete point sizes are available. For best results it is advisable to use TrueType fonts, which can be scaled to any size, but which are only available in Windows version 3.1 or higher.

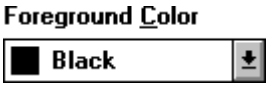
Font Styles

The font chosen can be drawn with four attributes selected, as shown above. Each is independently selected, so it is possible to have bold italic underlined text etc. The font style shown on screen is chosen to best represent the *printed* output. For example, if italic style is chosen for a printer that does not support italic text then the font will not appear italic on screen.

The style selected is designated the Initial Style because it is possible to enter formatting information into a piece of text in order to modify the style for portions of the text. See p89 for further information.

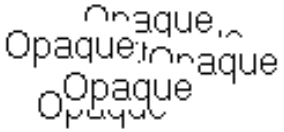
Text Color

The color of the text is set from the Foreground Color drop-down list box.



Text Background

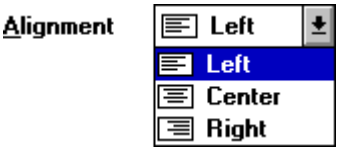
Text can have an opaque background or a transparent background depending on whether the Background Color box is checked or not. The difference between text with opaque and transparent backgrounds is shown below.



The color of the opaque background is set from the drop-down list box.

Text Alignment

The text drawn within GraFit can be aligned to the left or right, or may be centered.

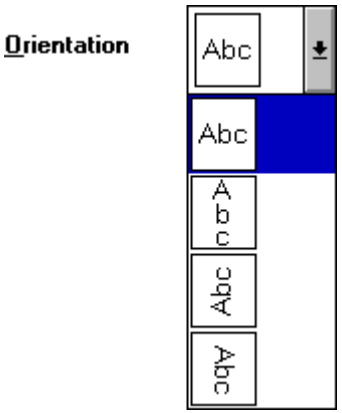


The effect of text alignment is shown below.

Left	Center	Right
1000	1000	1000
100	100	100
10	10	10
1	1	1

Text Orientation

The orientation of the text is set from the Orientation box.



Choose the orientation required from those illustrated.

Problems With Rotated Text

Windows versions before 3.1 did not have full capability to display rotated fonts, although certain printer drivers, notably the PostScript driver, allowed such fonts to be printed. With the advent of TrueType fonts incorporated into Windows 3.1, rotated text is fully supported. In order to print and display rotated fonts correctly you must either

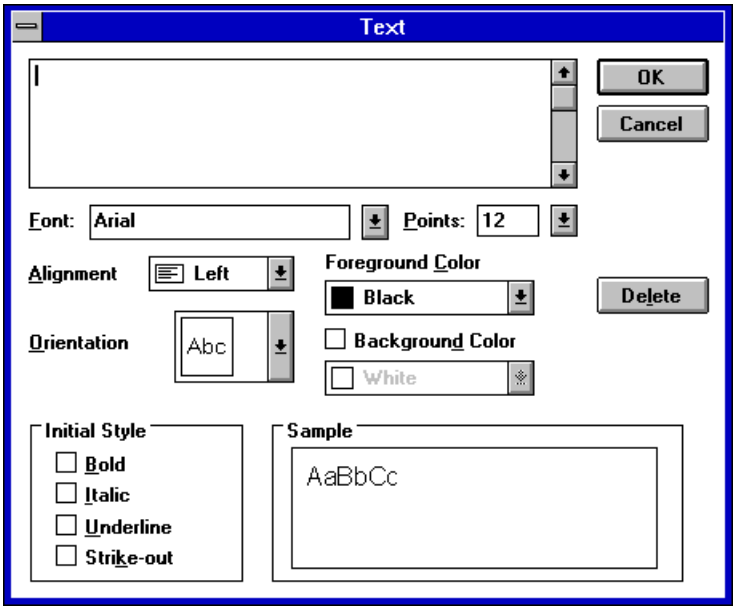
- Use a printer that supports rotated fonts. Such printers include PostScript printers and the HP LaserJet III.
- Use Windows version 3.1 and TrueType fonts.

Adding Text to a Graph

Text labels are added to the graph using the **Add Text** command. After selecting this command, the cursor will change shape to a cross, and you should click the mouse at the position where the text is to be placed.

Cross-shaped Cursor

The following dialog box is produced.



Enter the text in the edit box and press **OK** to generate the text label. The text style is specified from the remainder of the dialog box, as described above.

Deleting Text

Only text that was added using the **Add Text** command can be deleted. Other text is attached to a graph, and can only be deleted by removing the entire graph. To delete a piece of selected text double-click on the text to produce the Text dialog box, and select the **Delete** button. Alternatively, use the **Edit Delete** command to delete selected text.

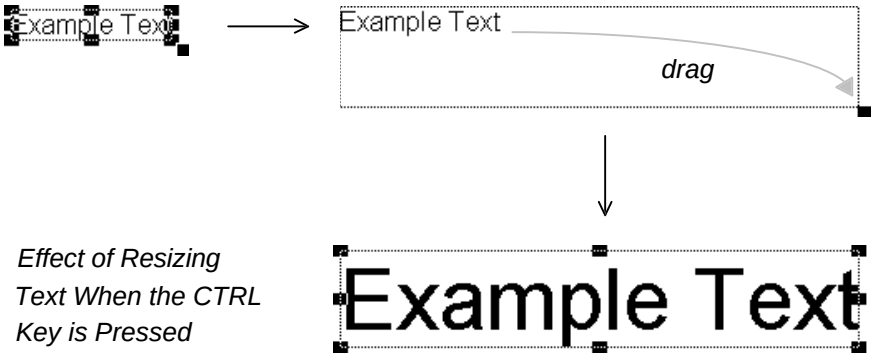
Text Wrap

The text placed by **Add Text** is drawn within a rectangular region. The text is wrapped to fit within the left and right limits of this region, and the height of the region is set to allow all the text to be displayed. When the text is selected, the shape of the region can be altered by dragging the resize handles. The effect of changing the size of this region is to cause the text to be reformatted. This is illustrated below.

Error! Objects cannot be created from editing field codes.

Rescaling Text

If the CTRL key is held down while resizing text, the effect is to rescale the text without altering the text wrapping.

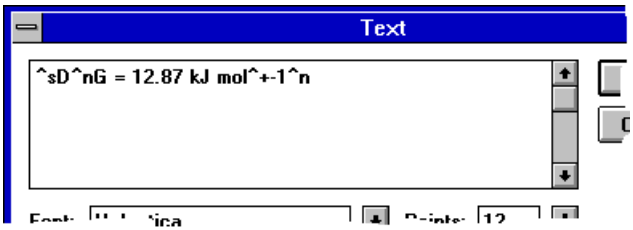


Effect of Resizing Text When the CTRL Key is Pressed

Extended Text Formatting Commands

Any text that is used in a graph may include sections of bold, italic, underlined, superscripted, or subscripted text, or any combination. It may also include characters from the Symbol font (this includes a full set of Greek characters). A full list of characters that are available in the standard and Symbol fonts is given in Appendix 2.

To specify that a section of text needs to be printed in this manner, it is necessary to enter simple formatting commands into the text. For example, the following entry



results in this output.

$\Delta G = 12.87\ kJ\ mol^{-1}$

Formatting Commands

Formatting commands can be included in any text that is on a graph, including axis labels, category labels on charts, and legend text. The formatting commands are single letters that are prefixed by the '**^**' character; to include the '**^**' character within a text string it is necessary to write '**^^**'. A list of these commands is given below:

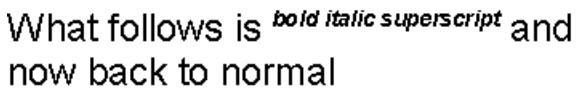
Command	Action
+	Text that follows will be superscripted
-	Text that follows will be subscripted
n	Reverts to the normal, starting text format
b	Bold

i	Italic
u	Underline
s	Text that follows will use the Symbol font. The symbol font includes Greek characters and mathematical symbols. The available characters are shown in Appendix 2.

It is allowed to intermix the various attributes, for example, bold italic superscripted text could be specified as:

```
What follows is ^b^i^+bold italic superscript^n and now back to normal
```

which would result in:



13 Adding Data and Curves To A Graph

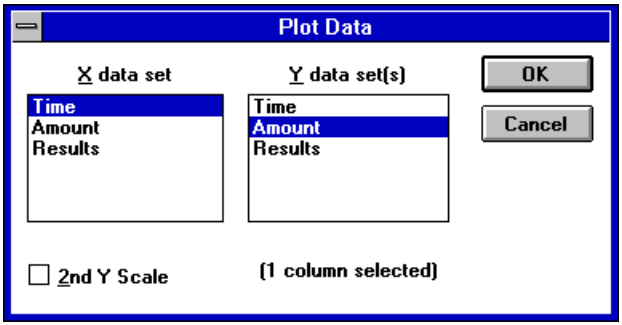
When graphs are first created they will usually display one or more data sets (see Chapter 6). Additional data sets and theoretical curves can be added to the graph using the commands described in this chapter. For information about editing the display characteristics of the data or curve, see Chapter 14.

Adding Data to the Graph

To add additional data to the graph, select the **Add Data...** command.

Adding *X*/*Y* Data to a Scatter Graph

The following dialog box is produced.



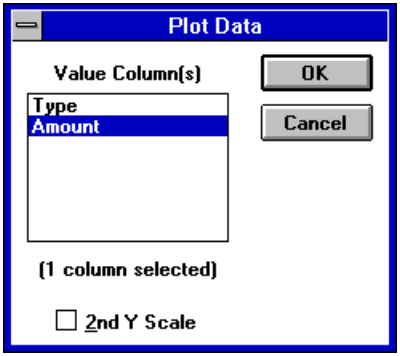
Several *y* data sets may be plotted by selecting more than one column from the Y data set(s) list box; to make multiple selections it is necessary to hold down the CTRL key when choosing items from the list box. However, each of these *y* data sets must have the same *x* data set — if they have different *x* data then you will need to repeat the **Add Data...** command for each.

2nd Y Scale

If the 2nd Y Scale box is checked, the data are added to the second *y* axis scale. See Chapter 16 for details of using more than one *y* axis scale.

Adding Data to a Column or Bar Chart

The following dialog box is produced.



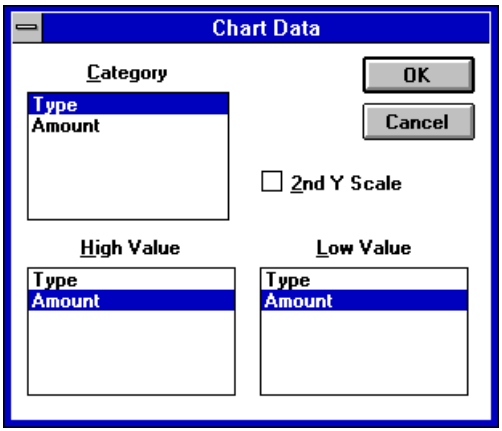
Several value data sets may be plotted by selecting more than one column from the Value Columns(s) list box; to make multiple selections it is necessary to hold down the CTRL key when choosing items from the list box.

2nd Y Scale

If the 2nd Y Scale box is checked, the data are added to the second *y* axis scale. See Chapter 16 for details of using more than one *y* axis scale.

Adding Data to a Low/High Column or Bar Chart

The following dialog box is produced.



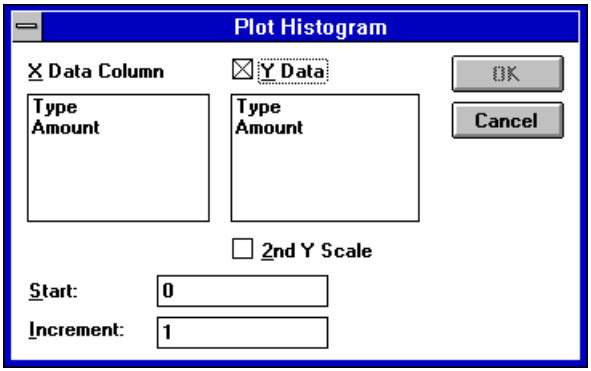
Select the columns that hold the category axis, the high data value and the low data value.

2nd Y Scale

If the 2nd Y Scale box is checked, the data are added to the second *y* axis scale. See Chapter 16 for details of using more than one *y* axis scale.

Adding Histograms

Histograms are added using the **Add Histogram...** command. The following dialog box is produced.



If the Y Data box is unchecked, it is assumed that there is one each of the *x* data values. If it is checked, the selected *y* column is taken to hold the number of entries corresponding to the values in the *x* column.

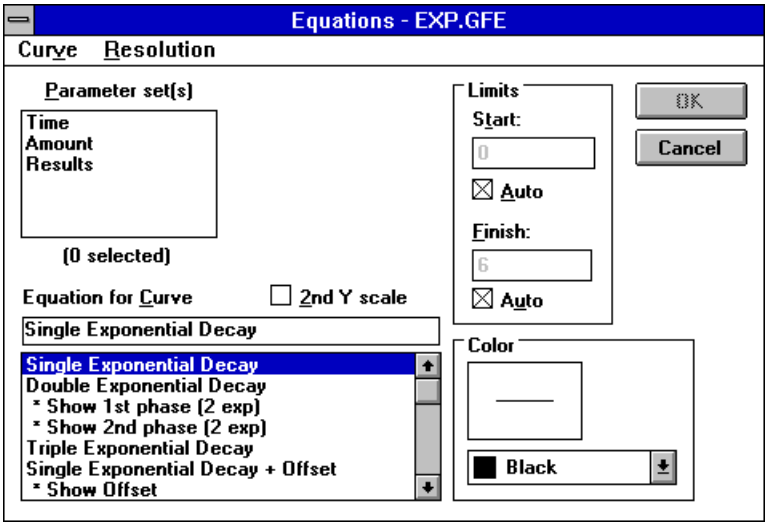
The histogram is grouped by taking values starting from the **Start** value, with the bars spaced as specified by the **Increment** value (see Chapter 6).

2nd Y Scale

If the 2nd Y Scale box is checked, the data are added to the second *y* axis scale. See Chapter 16 for details of using more than one *y* axis scale.

Adding Curves to the Graph

Data fitting results in a set of parameter values that describe the data (see Chapters 21 and 22). To display this theoretical curve on the graph, chose the **Add Curve...** command. The following dialog box is produced.



Select the column that contains the parameters and the equation for the curve, and then press **OK**. The equation list includes all those that are in the current equation definition file (to use a different definition file, use the **Options Definition file...** command before using **Add Curve...**). Several curves can be specified by selecting more than one item from the Parameter set(s) list box (hold down the CTRL key to choose more than one data column).

If the equation that is selected uses constants, an additional list box is present from which the column for the constants is selected. Several curves can be specified by selecting more than one item from the Constant(s) list box (hold down the CTRL key to choose more than one data column).

Curve Line Style and Resolution

The **Curve** menu is used to select the line type and thickness. The **Resolution** menu selects from low, medium and high resolution for the curve. High resolution produces the most accurate representation at the cost of longer calculation times and greater memory requirements. For most purposes medium resolution is optimum; this is the default setting.

Altering the Limits

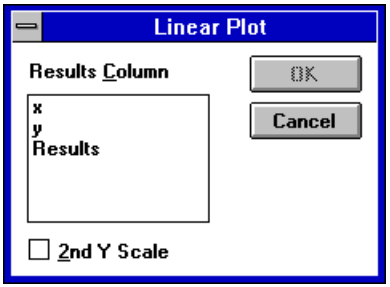
The Limits section specifies the *x* range over which the curve is drawn. If the Auto boxes are selected, the start and finish positions are set by the extent of the *x* axis. Auto selection is the default, but is not suitable for all curves. For example, certain curves are undefined at $x = 0$ and so may require defined limits.

2nd Y Scale

If the 2nd Y Scale box is checked, the data are added to the second *y* axis scale. See Chapter 16 for details of using more than one *y* axis scale.

Plotting Linear and Polynomial Fits

The results from linear regression using the **Data Linear fit...** and **Data Polynomial fit...** commands can be plotted using the **Add Linear fit...** and **Add Polynomial fit...** options (see Chapter 21 for a description of linear and polynomial regression). These commands operate in a similar manner; the **Add Linear Fit...** command produces the following dialog box.



Select the column that holds the results, and press **OK**.

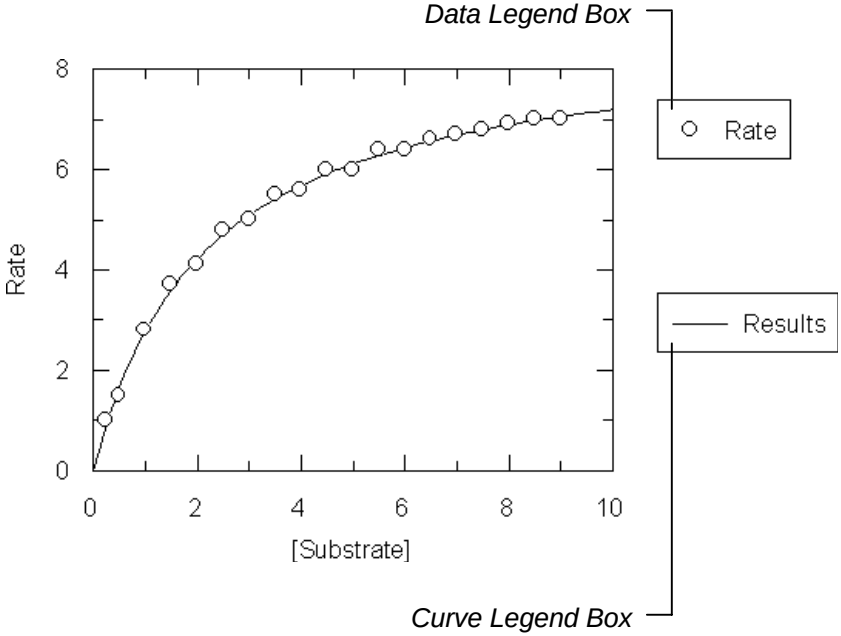
2nd Y Scale

If the 2nd Y Scale box is checked, the data are added to the second *y* axis scale. See Chapter 16 for details of using more than one *y* axis scale.

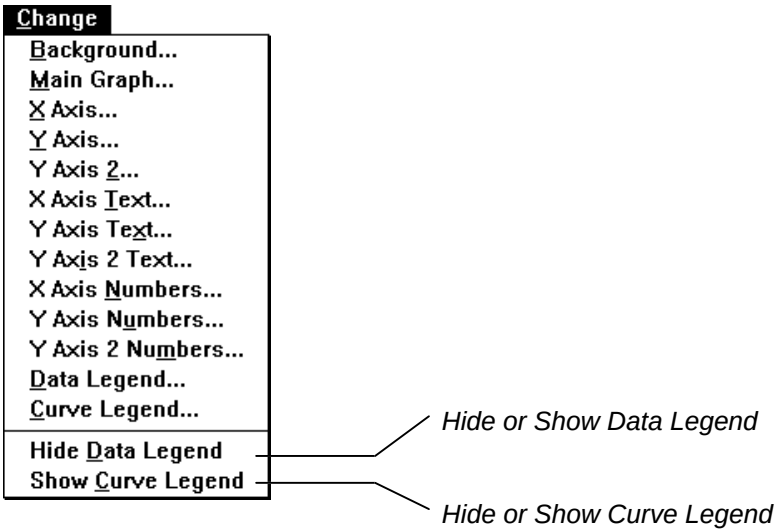
14 Editing Data and Curves

Editing Data and Curve Styles

All graph types may hold any number of data sets; *x*/*y* scatter plots may also show theoretical curves. The identity of each data set or curve is shown in the data or curve legend box. These legend boxes also provide a means of editing the style of the displayed data or curve.



Legend boxes may be displayed or hidden from the **Change** menu



If more than one graph is present, it is necessary to select part of the required graph before choosing the **Change** menu command. It is useful to display the legend in order to edit the data or curve styles (see below), although the **Change Data Legend...** and **Change Curve Legend...** commands also allow these styles to be edited.

The Data Legend Box

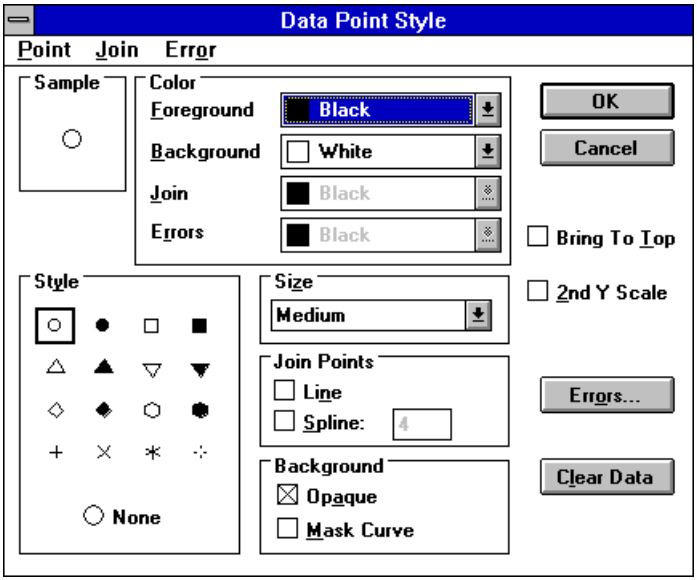
The data legend box allows various aspects of the displayed data points to be edited. When selecting the data legend box itself it is important to click near the edge of the box, as this box contains additional subregions that may also be selected. The different subregions are illustrated below.

Error! Objects cannot be created from editing field codes.

The selected legend box can be moved to any position on the screen, however it cannot be resized as the size is fixed by that of its contents. Selected subregions of the legend box cannot be moved or sized.

Data Point Symbol

The style of the points used to plot the data in an *x*/*y* scatter graph is selected by clicking on the data point symbol within the data legend box. Double-clicking on the selected data point symbol results in the following dialog box.

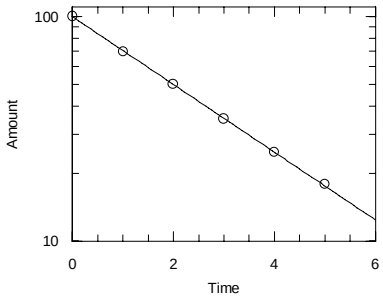


Data Point Style and Size

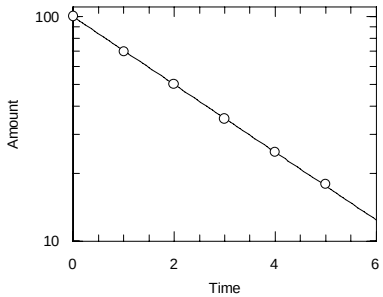
Choose one of the sixteen available data point styles, or 'None' if no point is to be drawn. Each of these styles can be drawn in various sizes, as controlled by the Size setting. The size of the point is relative to the overall graph width, and so the sizes of the points change proportionately when the graph size is altered. The line style and thickness used to draw the data points are controlled by the **Point** menu.

Background Settings

Data points may be drawn transparently over the curves, or they can overlie the curve by selecting an Opaque background. The difference is illustrated below.



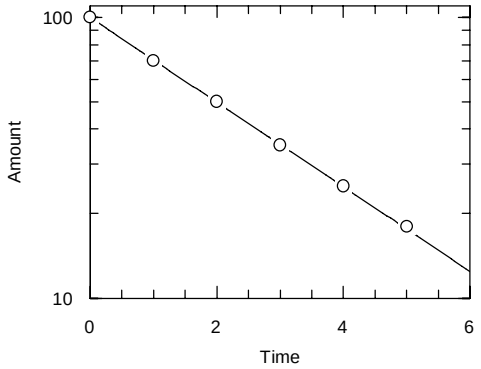
Transparent Data Points



Opaque Data Points

The colors used for the foreground and background parts of the data points are individually selectable from the Color part of the dialog box.

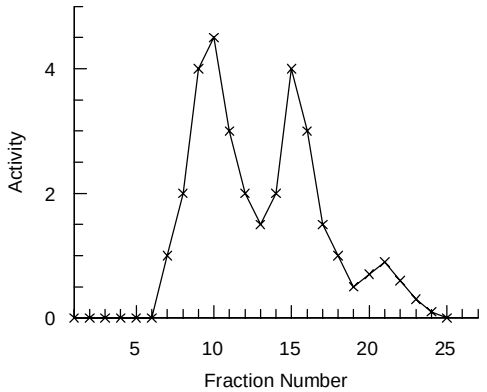
Choosing the Mask Curve box results in the following appearance for the plot.



Mask curve selected

Join Points Settings

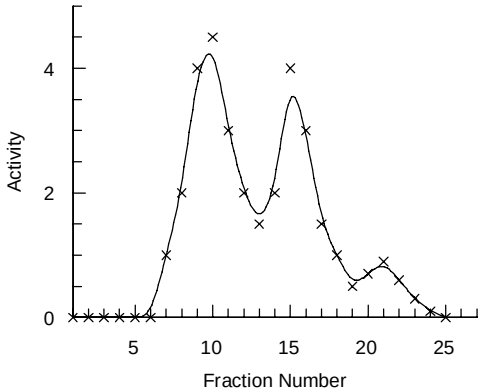
If the Line box is checked, successive data points are joined by lines. This allows graphs to be plotted without theoretical curves, but still indicating the trends in the data.



Points joined by a line

For correct representation of the data on the graph it is necessary that the points are stored in the data table in the order in which they will be joined. Use the **Manipulate Sort...** command on the values in the data table if necessary (see Chapter 25).

An alternative way of joining the points is using a spline curve, obtained by checking the Spline box. This option produces a smooth curve through the data points.



Points joined by a spline curve

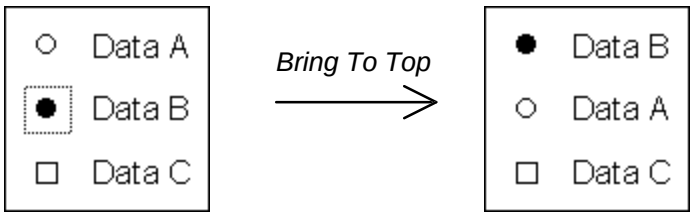
The style of the line that joins the points is set from the **Join** menu option; the color is set from the Color portion of the dialog box.

Spline Curve Order

The degree of smoothing applied to the spline curve can be set by entering a value next to the Spline box. This value, the order of the spline curve, controls how many points are involved in determining the local curvature. Valid entries range from 3 to 6, with higher values giving more averaging.

Order of Data Items in the Legend Box

When more than one data set is present it is possible to alter the order of the information shown in the data legend box. This is accomplished by checking the Bring to Top box of the dialog box. If this box is checked the currently selected data description will be repositioned as the first item in the legend box.



2nd Y Scale Selection

The data set can be displayed within the scale settings for the default *y* axis, or for the 2nd *y* axis. When the 2nd Y Scale box is checked the data are associated with the 2nd *y* axis. For further details of using more than one *y* scale, see Chapter 16.

Clearing the Data Set

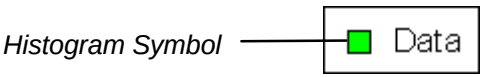
To clear the selected data set, press the **Clear Data** button. You will be asked to confirm that you wish to remove these data points from the graph; if confirmed these data will be deleted. Alternatively, chose **Edit Delete** when the data symbol is selected.

Error Bars

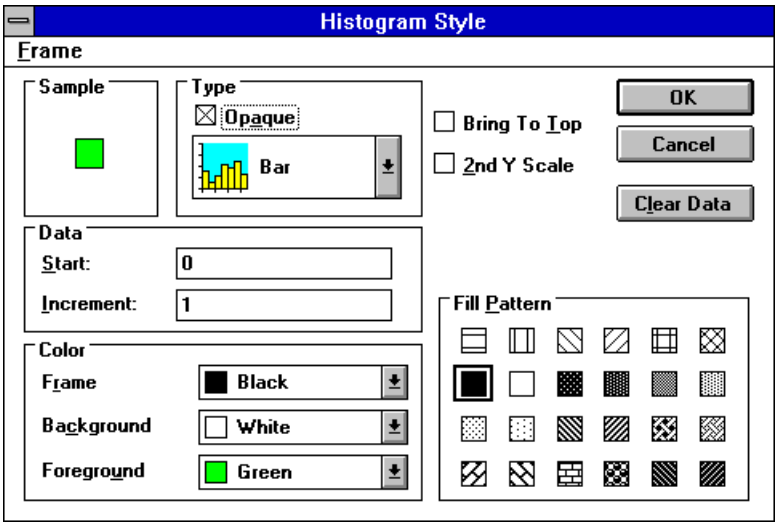
The **Errors...** button controls the drawing of any error bars on the chart. For a description of error bars, see Chapter 15.

Histogram Symbol

When a histogram is present, a square box is present in the data legend box that shows the fill style for the histogram.



Double-clicking on the histogram symbol results in the following dialog box.



Line Style

The style and thickness of the pen used to draw the histogram is chosen from the **Frame** menu.

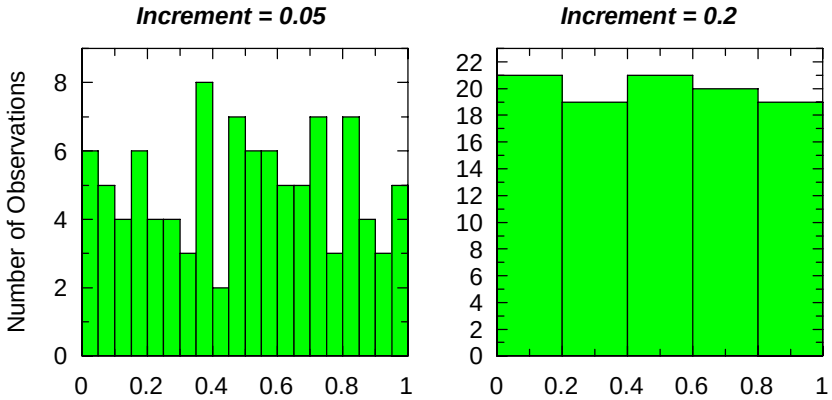
Type Section

The bars used to draw the histogram can be drawn opaque or transparent when used with the first eight fill patterns. Check the Opaque box to select opaque bars. The style of the histogram is selected from the drop-down list box.

Choose from a bar or outline representation as required.

Data Section

The data section specifies the start and increment values for the data ranges that will generate the histogram. A setting of Start = 0, Increment = 1 will result in the data being grouped in the following ranges: greater than or equal to 0 to less than 1; greater than or equal to 1 to 2; greater than or equal to 2 to 3; ... etc. The effect of changes to the increment value is illustrated below.



Fill Pattern

Two types of fill pattern are available, hatched and bitmap. The pattern is chosen by clicking on the appropriate part of the dialog box; the current selection is outlined by a dark box.

Error! Objects cannot be created from editing field codes.

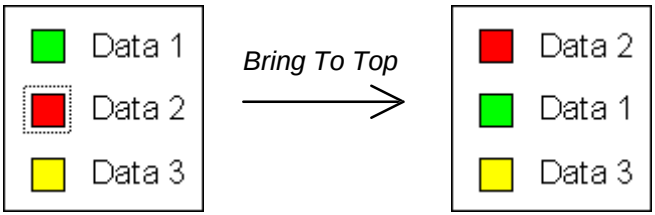
The foreground color is applied to the black portions of the pattern, and background color to the white parts. For the hatched patterns, the background is only drawn if the Opaque box is checked.

Grayscale Printing

To obtain a gray color it is possible to select one of the bitmap patterns, or to choose the Solid Foreground pattern and set the foreground color to gray from the Color selection area. Most printers will give more pleasing results when using the latter method.

Order of Items in the Legend Box

When more than one data set is present it is possible to alter the order of the information shown in the data legend box. This is accomplished by checking the Bring to Top box of the dialog box. If this box is checked the currently selected data description will be repositioned as the first item in the legend box.



2nd Y Scale Selection

The data set can be displayed within the scale settings for the default *y* axis, or for the 2nd *y* axis. When the 2nd Y Scale box is checked the data are associated with the 2nd *y* axis. For further details of using more than one *y* scale, see Chapter 16.

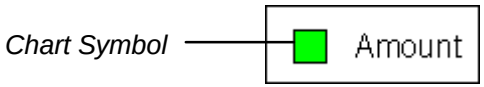
Clearing the Data Set

To clear the selected data set, press the **Clear Data** button. You will be asked to confirm that you wish to remove these data points from the graph; if confirmed these data will be deleted. Alternatively, chose **Edit Delete** when the data symbol is selected.

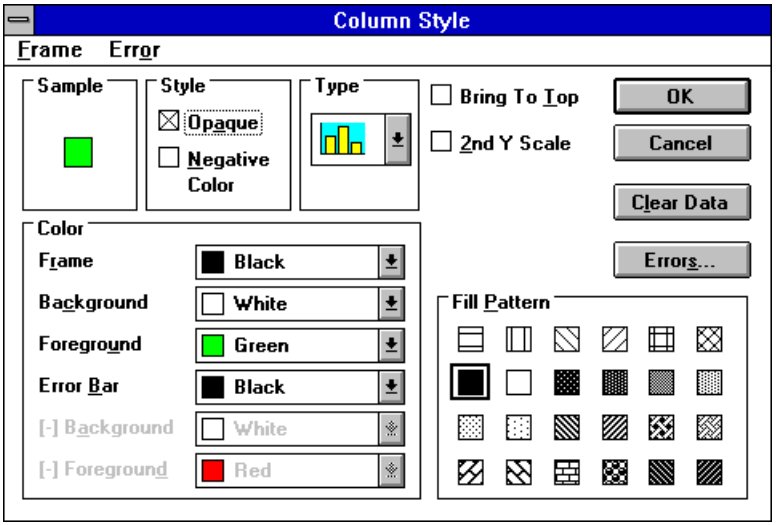
Bar and Column Chart Symbol

Editing the styles of bar and column charts is performed in a similar manner. The following section illustrates the settings for a column chart symbol, but editing settings for bar charts is done in an analogous fashion.

When a chart is present, a square box is present in the data legend box that shows the fill style.



Double-clicking on the chart symbol results in the following dialog box.

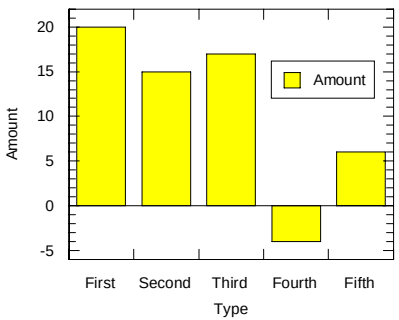


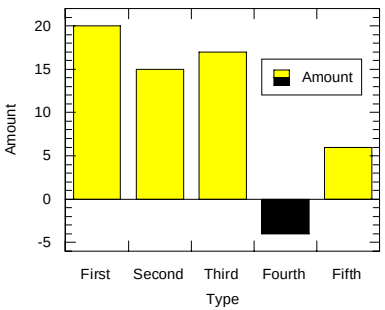
Line Style

The style and thickness of the pen used to draw the chart is chosen from the **Frame** menu. If error bars are present, the pen style used to draw them is chosen from the **Error** menu.

Style Section

The bars used to draw the chart can be drawn opaque or transparent when used with the first eight fill patterns. Check the Opaque box to select opaque bars. If the Negative Color box is checked, any negative values in the chart are drawn in a different color from the positive values. The colors for the different parts of the chart bars are controlled from the Color section.

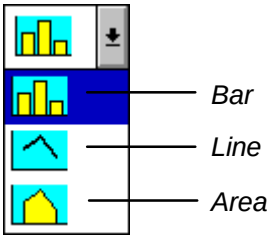




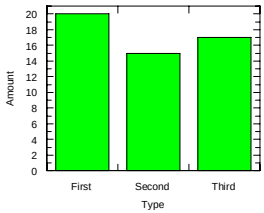
Negative Color box checked

Type Section

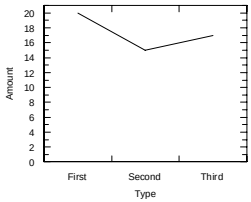
The drop-down list box sets the overall style used to represent the data.



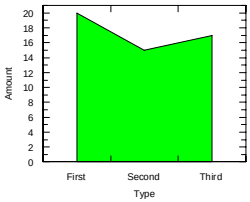
Choose the representation required. The effect of these different styles is shown below.



Bar

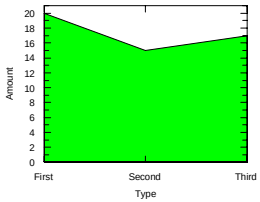


Line



Area

The Area representation differs slightly when the chart category axis style is set so that the Axis Between Categories box is unchecked (see Chapter 10). In this case the above chart would be drawn as below

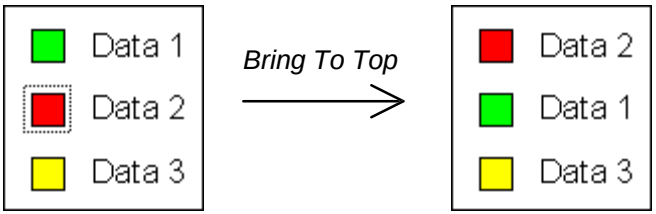


Fill Pattern

The fill patterns that are available are the same as the ones described earlier for histograms.

Order of Items in the Legend Box

When more than one data set is present it is possible to alter the order of the information shown in the data legend box. This is accomplished by checking the Bring to Top box of the dialog box. If this box is checked the currently selected data description will be repositioned as the first item in the legend box.



2nd Y Scale Selection

The data set can be displayed within the scale settings for the default *y* axis, or for the 2nd *y* axis. When the 2nd Y Scale box is checked the data are associated with the 2nd *y* axis. For further details of using more than one *y* scale, see Chapter 16.

Clearing the Data Set

To clear the selected data set, press the **Clear Data** button. You will be asked to confirm that you wish to remove these data points from the graph; if confirmed these data will be deleted. Alternatively, chose **Edit Delete** when the data symbol is selected.

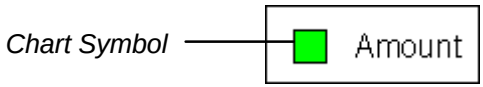
Error Bars

The **Errors...** button controls the drawing of any error bars on the chart. For a description of error bars, see Chapter 15.

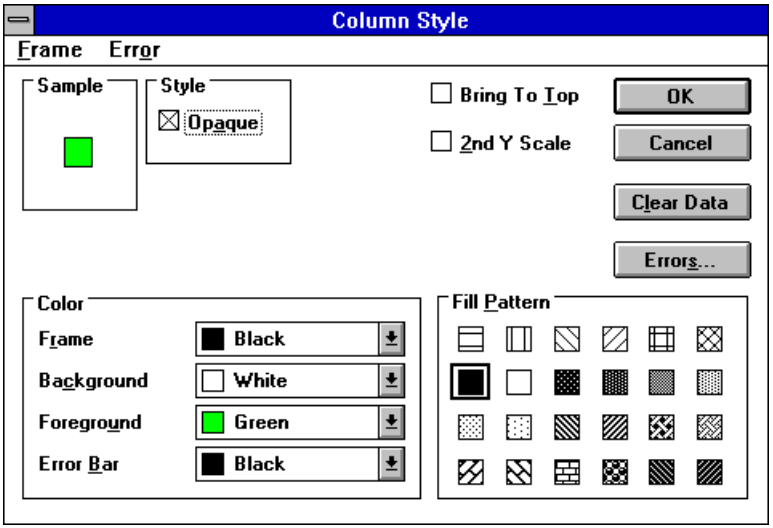
Low/High Bar and Column Chart Symbol

Editing the styles of low/high bar and low/high column charts is performed in a similar manner. The following section illustrates the settings for a low/high column chart symbol, but editing settings for low/high bar charts is done in an analogous fashion.

When a chart is present, a square box is present in the data legend box that shows the fill style.



Double-clicking on the chart symbol results in the following dialog box.



Line Style

The style and thickness of the pen used to draw the chart is chosen from the **Frame** menu. If error bars are present, the pen style used to draw them is chosen from the **Error** menu.

Style Section

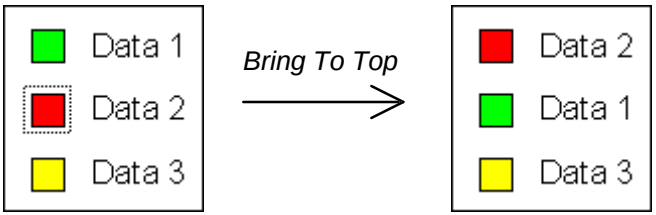
The bars used to draw the chart can be drawn opaque or transparent when used with the first eight fill patterns. Check the Opaque box to select opaque bars.

Fill Pattern

The fill patterns that are available are the same as the ones described earlier for histograms.

Order of Items in the Legend Box

When more than one data set is present it is possible to alter the order of the information shown in the data legend box. This is accomplished by checking the Bring to Top box of the dialog box. If this box is checked the currently selected data description will be repositioned as the first item in the legend box.



2nd Y Scale Selection

The data set can be displayed within the scale settings for the default *y* axis, or for the 2nd *y* axis. When the 2nd Y Scale box is checked the data are associated with the 2nd *y* axis. For further details of using more than one *y* scale, see Chapter 16.

Clearing the Data Set

To clear the selected data set, press the **Clear Data** button. You will be asked to confirm that you wish to remove these data points from the graph; if confirmed these data will be deleted. Alternatively, chose **Edit Delete** when the data symbol is selected.

Error Bars

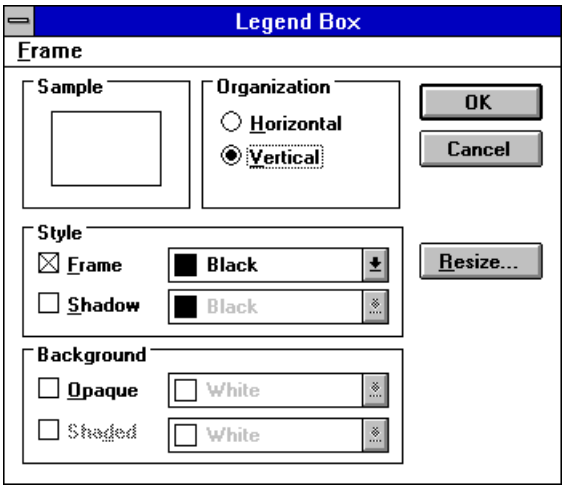
The **Errors...** button controls the drawing of any error bars on the chart. For a description of error bars, see Chapter 15.

Legend Boxes

The style of both the data and curve legend boxes can edited in the same way.

Box Style

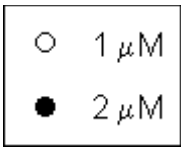
Double-clicking on the selected legend box produces the following dialog box.



The style of the legend box is altered in the same way as that of an added box, and is described in Chapter 17.

Organization

The Organization section describes how multiple sets of data points are displayed within the legend box. Choose from Horizontal or Vertical.



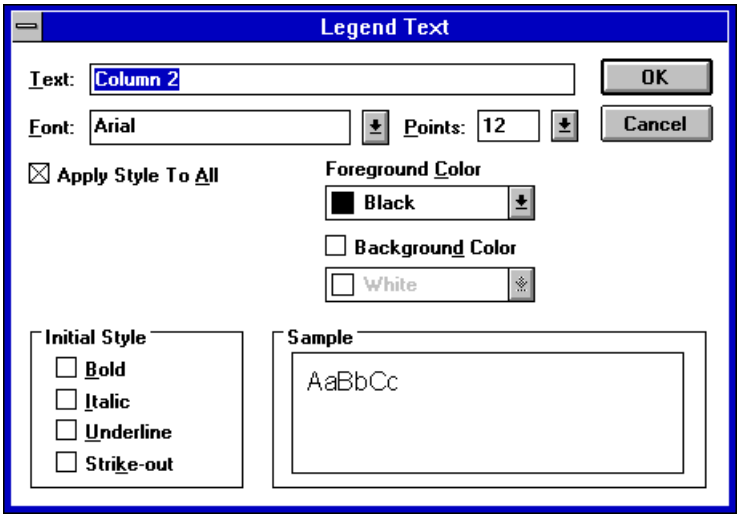
Vertical

Horizontal

Legend Text

The data legend box shows the style of the data point together with a text identifier. This text is initially taken from the *y* column name used when plotting the data. The curve legend box shows the line style of the curve, with an identifier text that is initially taken from the column that held the parameter values. In each case it is possible to edit the identifier text, or change the display style.

To select the identifier text, click on this text (if the legend box is selected it will be necessary to deselect this first by clicking on the graph background area). This identifier text is fixed in position relative to the surrounding legend box, and so cannot be moved or resized (although the whole legend box may be moved). Double-clicking on the identifier text produces the following dialog box.



This dialog box allows the text style to be specified, as described in Chapter 12. The text used for the label can be edited as required, and may include extended formatting codes so that, for example, superscripted or Greek letters can be produced (see Chapter 12).

Apply Style To All

If the Apply Style To All box is checked, any style settings will be applied to all text labels within the legend box; if unchecked the style settings will apply only to the selected label.

The Curve Legend Box

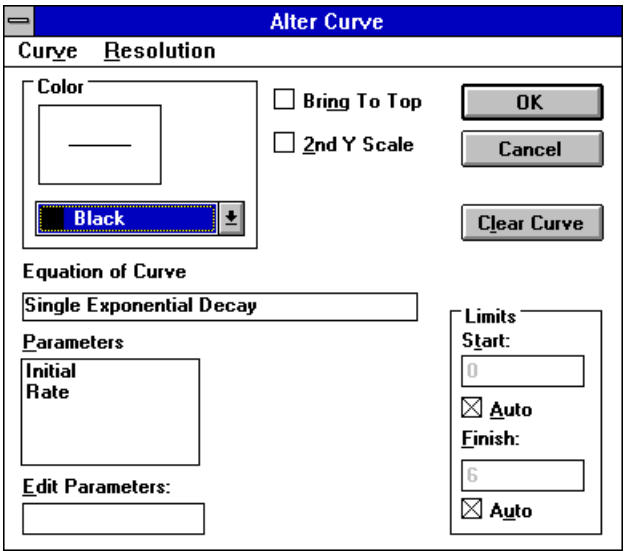
The curve legend box contains information relating to any theoretical curves drawn on the graph. Only *x*/*y* scatter graphs may incorporate theoretical curves. This legend box can be selected by clicking on the box, near to the edge. The curve legend box contains two further subregions, and so to be sure of selecting the box itself it is necessary to click close to the edge of the box. The different subregions are illustrated below.

Error! Objects cannot be created from editing field codes.

The style of the legend box can be edited by double-clicking on the box. The style settings are altered in the same way as with the data legend box, as described above.

Curve Style

Double-click on the selected curve symbol to edit the style of the line used to display the theoretical curve, or to change the parameter values that describe the curve. This produces the following dialog box.



This dialog box allows you to specify various aspects of the curve's presentation.

Curve Line Style and Resolution

The **Curve** menu is used to select the line type and thickness. The **Resolution** menu selects from low, medium and high resolution for the curve. High resolution produces the most accurate representation at the cost of longer calculation times and greater memory requirements. For most purposes medium resolution is optimum; this is the default setting.

Altering the Limits

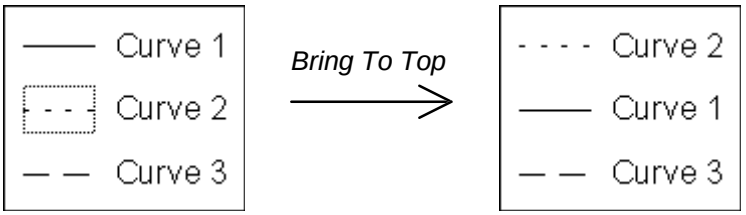
The Limits section specifies the *x* range over which the curve is drawn. If the Auto boxes are checked, the start and finish positions are set by the extent of the *x* axis. Auto selection is the default, but is not suitable for all curves. For example, certain curves are undefined at $x = 0$ and so may require defined limits.

Editing Parameter and Constants Values

The Parameters section allows alteration of the values used to define the equation. Select the parameter that you wish to change from the Parameters list box; the value of this parameter will appear in the Edit Parameters box, and may be edited as required. If the equation selected uses constants, a Constants section will also be present. This allows the values of the constants to be edited.

Order of Curves in the Legend Box

When more than one curve is present it is possible to alter the order of the information shown in the curve legend box. This is accomplished by checking the Bring to Top box of the dialog box. If this box is checked the currently selected curve description will be repositioned as the first item in the legend box.



Clearing Curves

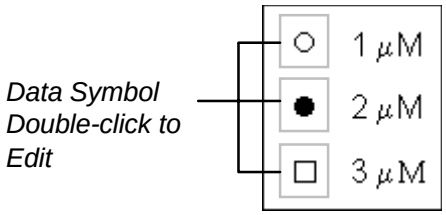
To clear the curve, select the **Clear Curve** button. You will be asked to confirm that you wish to delete the curve; selecting **Yes** will result in the selected curve being removed from the graph. Alternatively, the curve may be cleared using the **Edit Delete** command when the curve symbol is selected.

15 Error Bars

Adding Error Bars to a Graph

Error bars can be displayed on the data plotted in a scatter graph on the *x*, *y* or both axes. Data in charts can display error bars on the values. In each case, the error limits that are associated with the data are specific to the data set.

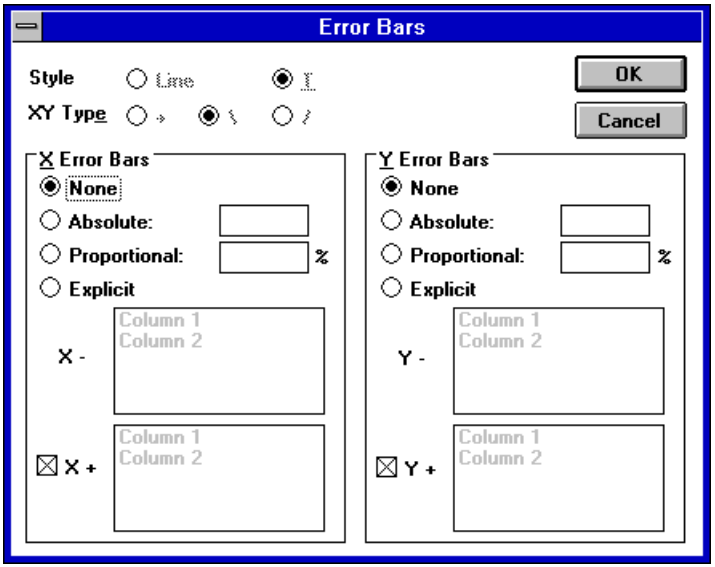
The display of error bars in a GraFit graph is a two step process. First, the data must be plotted. Second, an error bar is associated with the data set. Error bars are treated as a style that is applied to a data set. Like any other element of the data point display, they are added or edited by double-clicking on the data or chart symbol within the data legend, as described in Chapter 14.



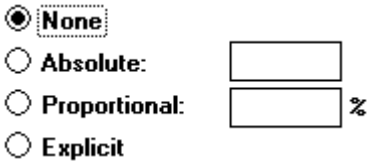
In this way, the error bar display of each data set can be individually controlled. Data on scatter graphs and charts have different characteristics that require different dialog boxes to control the various data style settings (see Chapter 14). Each allows the style and thickness of the error bar to be set from an **Error** menu item, and the color of the error bar to be selected. In addition, each has a button labeled **Errors...** which controls the error bar settings. The dialog boxes that result from selecting this button are described below.

Error Bars on *X*/*Y* Data

Data points can have error bars on the *x*, *y* or both axes. The **Errors...** button on the data point dialog box results in the following dialog box.



Error bars are specified separately for the *x* and *y* axes. Four selections are available.



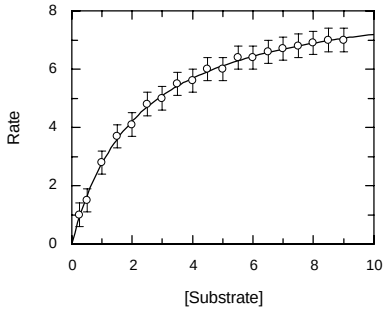
1. None. No error bars are drawn; this is the default setting.
2. Absolute. The size of the error bar is the same for each data point of the data set, this size is entered into the edit box next to the Absolute selection.
3. Proportional. The size of the error bar depends upon the magnitude of the data point, the size is expressed as a percentage of the data point value and is entered into the edit box next to the Proportional selection.
4. Explicit. The size of the error for each data point in the data set is specified individually. The error values must be stored in a data column, which is selected from the list box. The negative and positive lobes of the error bar must be specified separately; if you require these errors to be the same simply select the same column for each list box. It is also possible to draw just one side of the error bar; if the 'X +' or 'Y +' check boxes are unchecked the positive errors are omitted. This is useful to show differences from theoretical curves (see later).

Shape of Error Bar

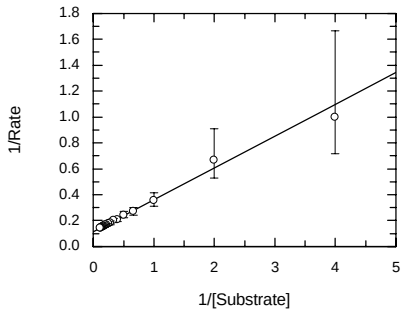
The shape of the error bar may be chosen between a simple line and a T-shaped bar. When both *x* and *y* error bars are drawn the double error bars may be selected from a cross shape, or from a diagonal error line.

Examples of Error plots

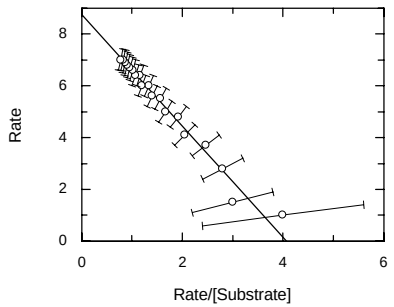
Below is shown a series of example graphs illustrating some effects that can be obtained by including error bars within a graph. Note that logarithmic scaling, when selected, is also applied to the error bars.



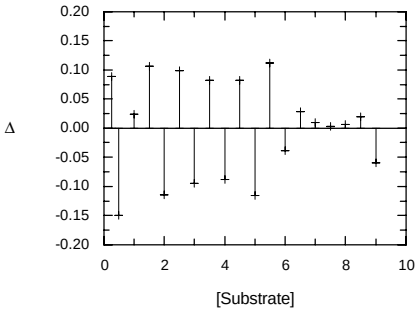
Enzyme kinetics plot. The error is present only in the Rate (y axis) measurement. Errors are simple, i.e. all data points have the same error.



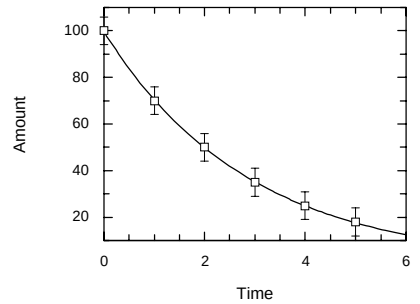
Lineweaver Burk plot of the data above. This plot takes reciprocals of both axes to obtain a linear plot. Note the large distortion of the error bars resulting from this plot; the distortion is greater above the line than below it.



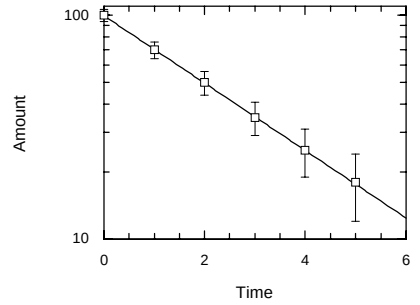
Eadie Hofstee plot of the same enzyme kinetics data. This plot rearranges the data by plotting Y versus Y/X to get a straight line. Note that in this case the error bars are present on both axes in the transformed plot.



Error plot from the above enzyme kinetics data. For this plot the errors were calculated using the **Manipulate Difference...** *command, and used as explicit 'Y-' error values; 'Y+' errors were not drawn.*



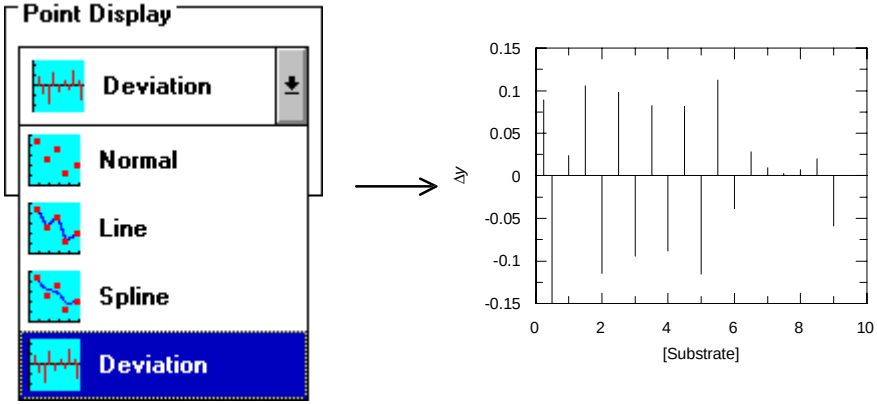
Exponential decay using a linear y axis scale. Equal error bars are drawn.



The same data and error bar selections as above, but using a semi-logarithmic plot. Note that the error bars are automatically adjusted to use logarithmic scaling.

Creating a Quick Residuals Plot

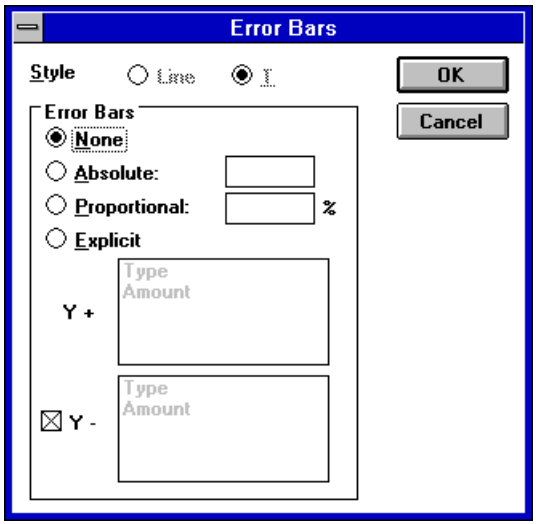
A plot of $y_{\text{experimental}} - y_{\text{calculated}}$ versus x gives a useful indication of whether the data fitting procedure is satisfactory (see Chapter 29). GraFit provides a means of creating such residuals plots in a single step. From the **New X/Y Graph...** command dialog box, select the Point Display style of "Deviation." Then choose the column holding the Δy values for the y axis of the graph, and the x data values for the x axis. This will produce the following style of residuals plot.



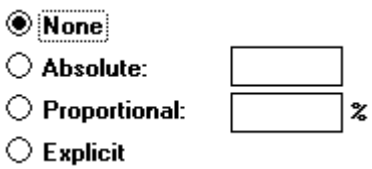
To create the Δ*y* values, either generate a Differences Column during curve fitting (Chapter 22), or use the **Manipulate Difference...** command (Chapter 25).

Error Bars on Bar or Column Charts

The **Errors...** button from the chart symbol dialog box results in the following dialog box.

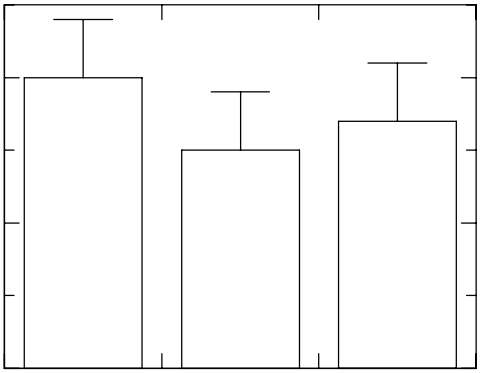


Four error bar selections are available.



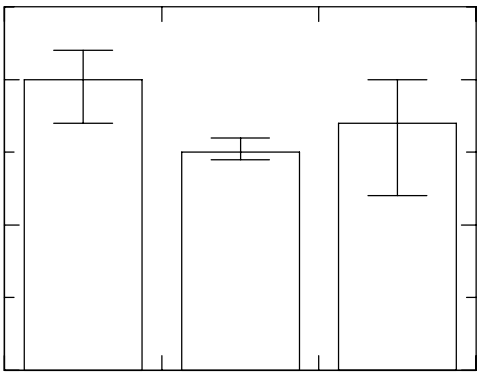
1. None. No error bars are drawn; this is the default setting.

2. Absolute. The size of the error bar is the same for each data value of the data set, this size is entered into the edit box next to the Absolute selection. Error bars are only displayed above the data value, i.e. as shown below.



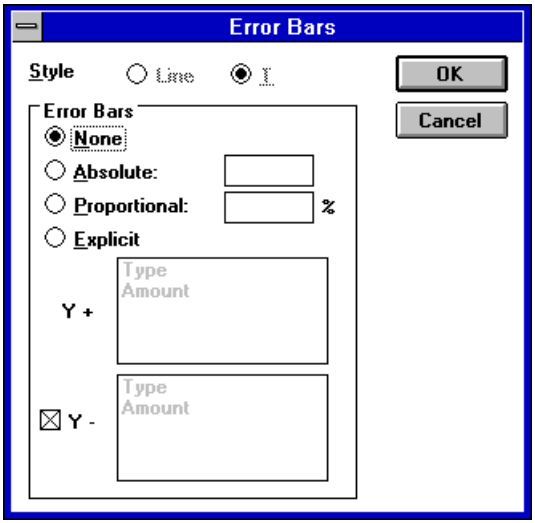
3. Proportional. The size of the error bar depends upon the magnitude of the data value, the size is expressed as a percentage of the data value and is entered into the edit box next to the Proportional selection. Error bars are only displayed above the data value

4. Explicit. The size of the error for each data value in the data set is specified individually. The error values must be present in a data column, and the appropriate column is selected from the list box marked Y+. If the Y- box is checked, a lower lobe of the error bar is drawn, with the values drawn from the data column in the adjacent list box. The effect of checking the Y- box is shown below.

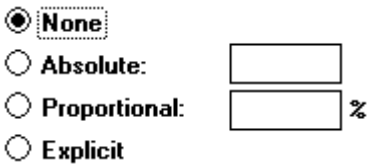


Error Bars on Low/High Bar and Column Charts

The **Errors...** button from the chart symbol dialog box results in the following dialog box.

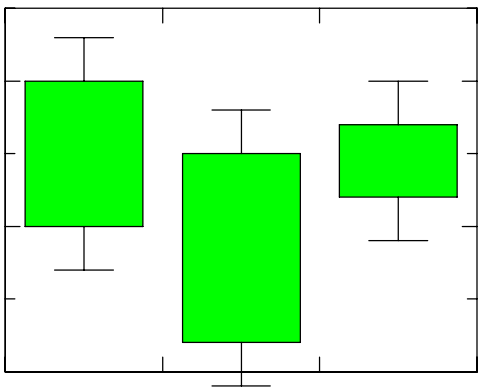


Four error bar selections are available.



1. None. No error bars are drawn; this is the default setting.

2. Absolute. The size of the error bar is the same for each data value of the data set, this size is entered into the edit box next to the Absolute selection. Error bars are only displayed over the upper data value and beneath the lower value, i.e. as shown below.



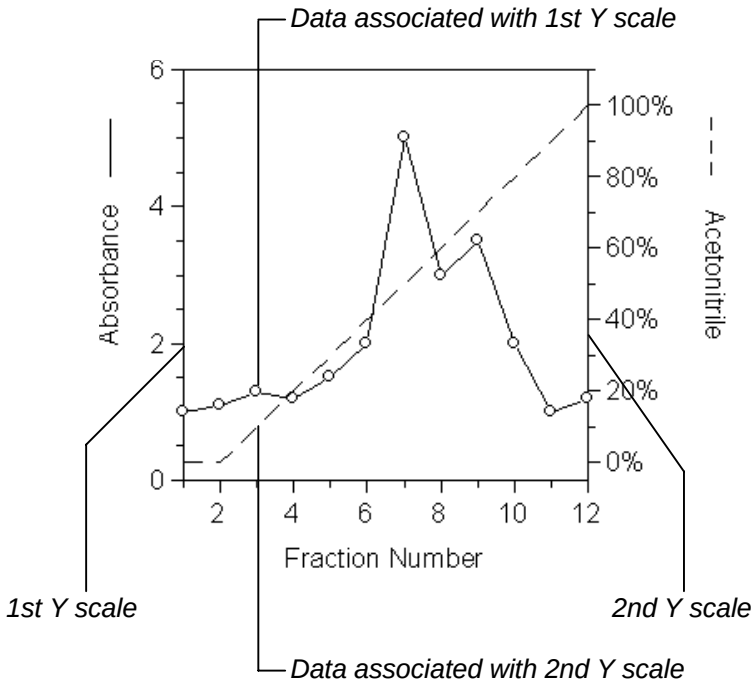
3. Proportional. The size of the error bar depends upon the magnitude of the data value, the size is expressed as a percentage of the data value and is entered into the edit box next to the Proportional selection.

4. Explicit. The size of the error for each data value in the data set is specified individually. The error values must be present in a data column, and the appropriate column to show the errors in the upper range is selected from the list box marked Y+. If the Y- box is checked, errors are shown for the lower data value, with the error values drawn from the data column in the adjacent list box.

16 Using More Than 1 Y Axis Scale

Using More Than 1 Y Axis Scale

X/Y Graphs and Column Charts may be created using up to 2 independent *y* axis scales. Any data or curves that are drawn on these graphs may be associated with the 1st (by default) or 2nd *y* scale.



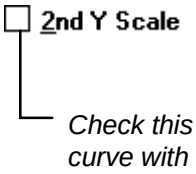
The second *y* scale, in common with the default *y* scale, has an associated scale number region and axis legend text. These regions are selected, edited and moved in exactly the same way as those of the default *y* scale (see Chapter 6).

Tick Settings

When using a second *y* scale, it is often necessary to edit the axis tick settings to provide a suitable display (the default settings can result tick marks from the first and second *y* axes overlapping each other). To change the tick mark settings, double-click on the appropriate axis, as described below.

Scale Association of Data and Curves

Whenever data or curves are added to a graph, it is possible to select whether they are associated with the first or second *y* axis. To associate them with the second *y* axis, check the 2nd Y Scale box that is present on all relevant dialog boxes.



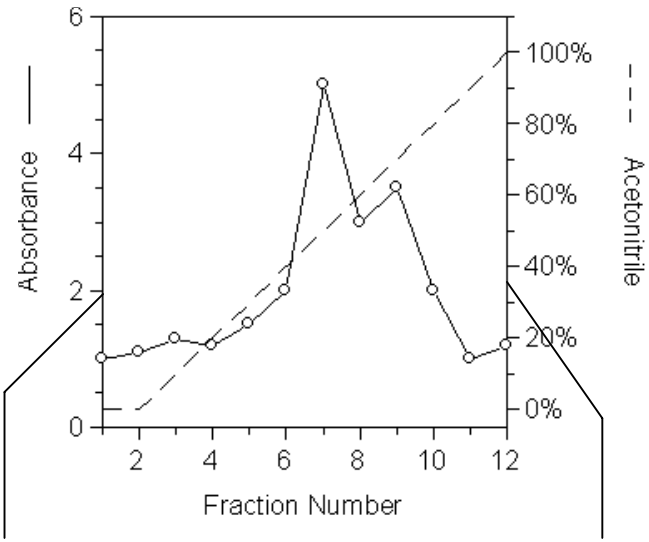
To change the association of existing data or curves, double-click on the appropriate symbol within the data or curve legend, and check or uncheck the 2nd Y Axis box.

Visibility of the 2nd Y Scale

The second *y* scale markings are only visible when one or more data sets or curves have been associated with the second *y* scale. Also, when no data or curves are so associated, as is the default, the second *y* axis legend and the second *y* axis scale numbers are hidden. The visibility of these regions and their attachment to the main graph region, as is the case for the default *y* axis legend and scale numbers, is controlled by the **Options Graph Settings...** command (see Chapter 6).

Setting 2nd Y Scaling and Tick Settings

The scale settings of the default and second *y* axis are set by double-clicking on the left or right edges of the graph respectively.



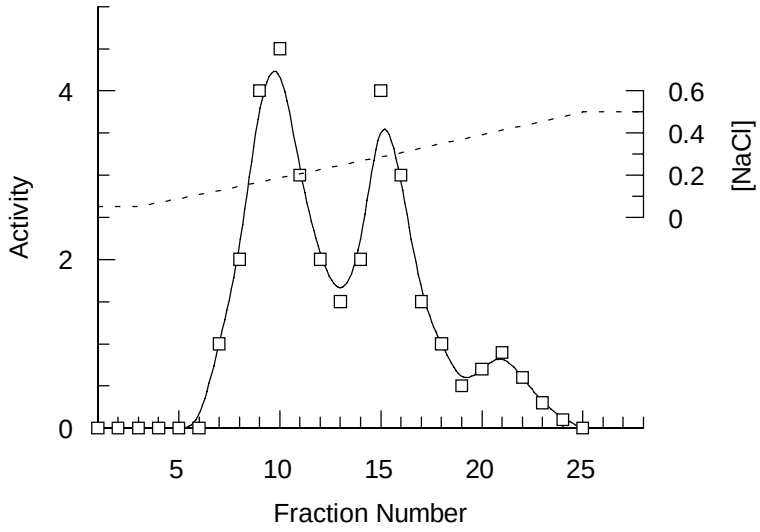
Double-click on this axis to set the first Y scale

Double-click on this axis to set the second Y scale

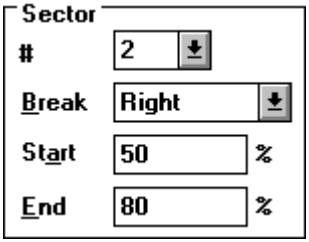
Scale settings for the second *y* axis are edited in the same manner as those of the default *y* axis, as described in Chapter 10.

Displaying the 2nd Y Scale Over a Limited Region

It is sometimes convenient to limit the display of a second *y* scale to a limited portion of the graph, such as in the plot below.



This can be achieved by double-clicking on the second *y* axis, and setting the sector Start and End values in the dialog box — in the above example these are set to be 50 and 80% respectively.



17 Graph Boxes

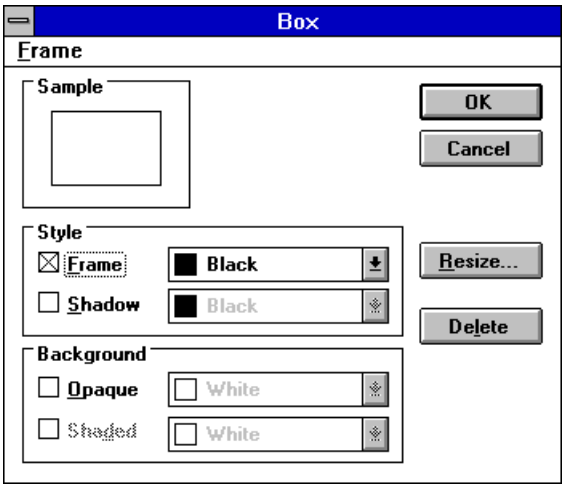
Graph Boxes

Boxes can be added to a graph window using the **Add Box...** command. Selection of this command results in the mouse cursor changing to a cross shape.

Cross-shaped Cursor

Move the cursor to the position where you require the top left corner of the box, and click the mouse button. Resize the resulting box as required.

Double-clicking on a selected box produces the following dialog box.

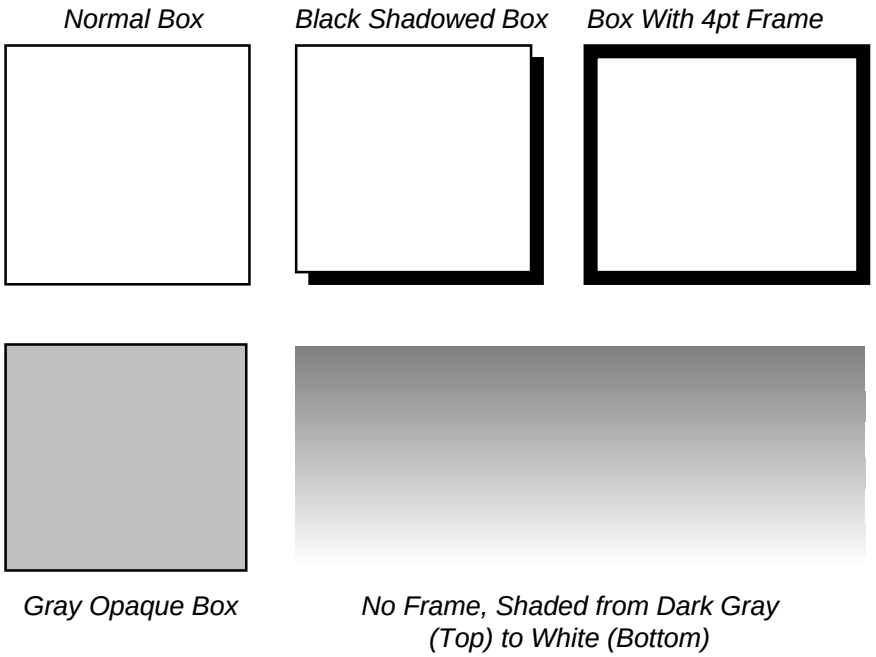


Box Styles

Boxes can be drawn in various styles. The **Frame** menu is used to select the style and width of the frame surround. The outer frame is drawn when the Framed style is checked; a shadow to the box is drawn when Shadowed is checked.

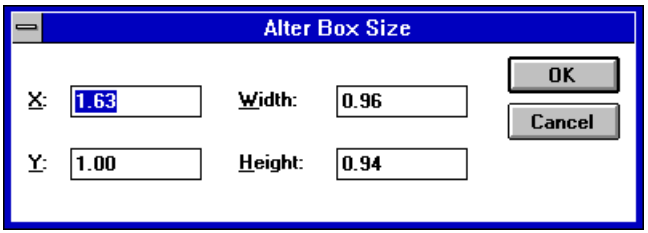
The background of the box is opaque or transparent depending upon whether the Opaque box is checked; opaque boxes will overdraw other parts of the graph. If the background is opaque a further check box labeled ‘Shaded’ becomes active. When the Shaded box is checked, the box background is shaded from the Opaque color (top) to the Shaded color (bottom).

A selection of different box styles is shown below.



Setting the Box Size and Position

The **Resize...** button can be used to set the position and size of the box precisely. The command results in the following dialog box.



The scale to use (inches or cm) is that of the currently selected graph (this can be set using the **File Graph Page Setup...** command [Chapter 6]).

18 Graph Lines and Arrows

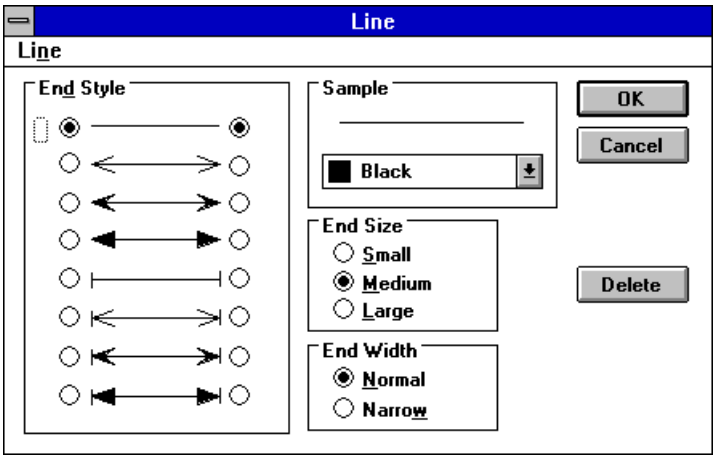
Adding and Editing Lines and Arrows

Lines and arrows are added to a graph window using the **Add Line** command. Selection of this command results in the mouse cursor changing to a cross shape.

Cross-shaped Cursor

Move the cursor to the position where you require the top left corner of the line, and click the mouse button. Position the ends of the resulting line as required after clicking on the line to select it.

Lines can have variously styled ends, giving simple lines or single and double-headed arrows. Double-clicking on the selected line produces the following dialog box.



Line Styles

Lines can be drawn in various styles. The **Line** menu is used to select the style and width of the line. Various different end styles may be selected for each end of the line. The size of the arrow head is selected from the End Size section, and it is possible to choose between Normal and Narrow arrow heads from the End Width section.

Deleting Lines

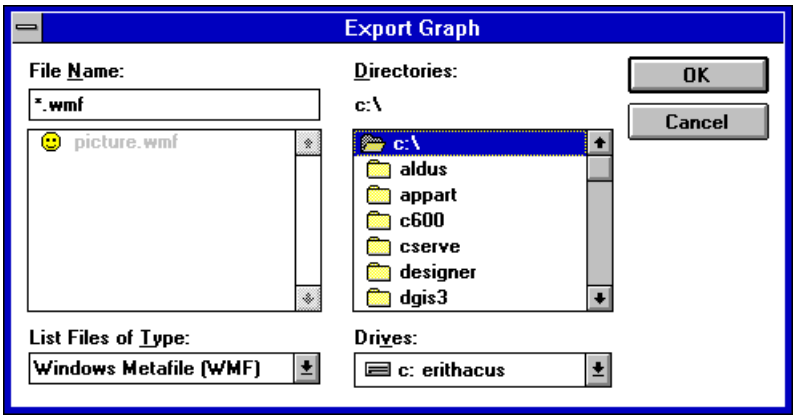
The line can be deleted using the **Delete** button on this dialog box. Alternatively, select the line using the mouse and choose the **Edit Clear** command, or press the DELETE key.

19 Exporting Graphs, Importing Pictures

Graphs may be exported to other applications via the Clipboard, or by creating a Windows Metafile (WMF file).

Exporting Graphs as WMF Files

The contents of a graph window can be saved to disk as a Windows Metafile (WMF file) using the **File Export Graph...** command. The WMF files that are created may then be read back by any Windows application that supports this format. Selection of this command results in the following dialog box.



Enter the name for the WMF file, and press OK to save the contents of the graph window in this format.

Copying and Pasting Graphs

When a graph window is active, the **Edit** menu allow graphs to be copied to or pasted from the Clipboard. If part of the graph is selected, **Edit Copy Selection** will copy a picture of the current selected region. If no part of the graph is selected, **Edit Copy All** can be chosen to copy the entire graph to the Clipboard. To paste the graph into another Windows application, activate this application and select **Edit Paste**. The graph is copied into the Clipboard in the Windows Metafile (picture) format. Applications that can accept these images include Windows Write, Microsoft Word for Windows and Aldus PageMaker.

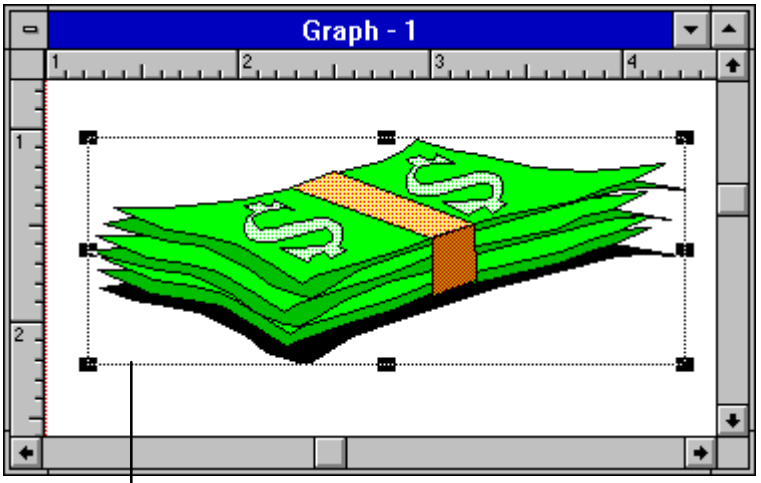
Problems With Rotated Text in Windows 3.0

Windows 3.0 has limitations in the display of rotated text, and if any rotated text is present in a graph pasted from GraFit, the text will not display correctly on-screen. However, the image

will still print correctly providing the printer supports the printing of rotated text. Windows version 3.1 or later has enhanced support for rotated text, and displays the text correctly.

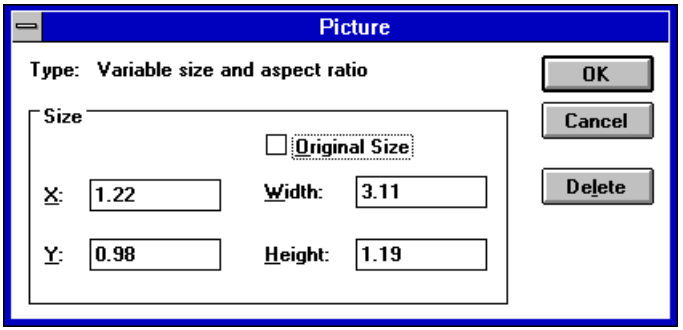
Pasting Pictures into a Graph

A picture created in another application may be included into a GraFit graph by pasting it from the Clipboard. GraFit will accept any image in the Windows picture (metafile) format.



Selected Picture

Double-clicking on a pasted picture results in the following dialog box.



The position, width and height of the image may be set, and the picture can be restored to its original size using the Default Size check box. It can be deleted by pressing the **Delete** button on this dialog box, or alternatively by using **Edit Clear** when the picture is selected.

Maintaining the Aspect Ratio of Pictures

When resizing pasted pictures it is possible to maintain the aspect ratio by holding down the SHIFT key while changing the size. This is often important when resizing pictures, in order to avoid distortion.

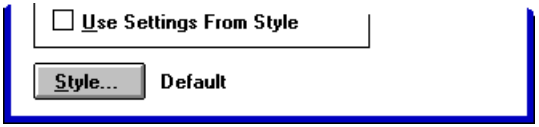
20 Graph Styles

Graph styles provide a quick way to alter the appearance of a graph or a graph object. After modifying the graph to give the required size, shape, point style, tick style, grid layout etc. the style is saved using the **Options Save Style...** command. This style, or one of the predefined styles that come with the program, can subsequently be applied to a new graph by selecting the saved style during creation, or by using the **Options Apply Style...** command to a selected graph in the Graph window.

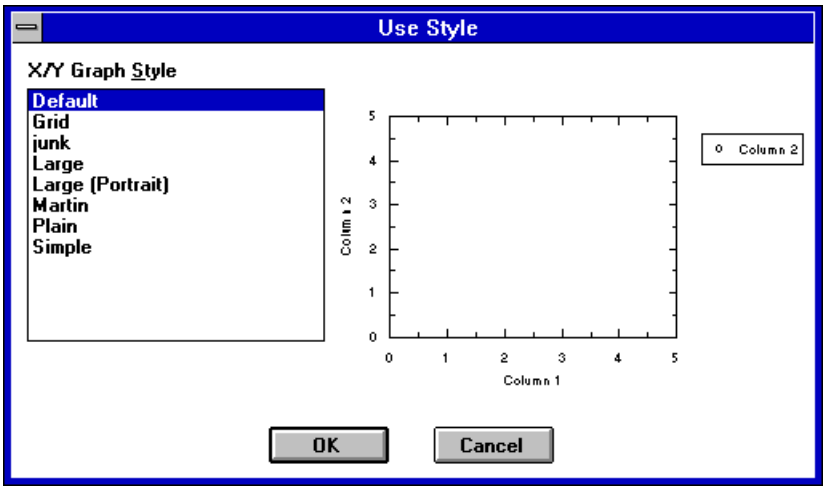
Styles can be individually applied to each of the principal graph types, to boxes, to text and to lines.

Choosing Graph Styles When Creating Graphs

When creating a graph or chart using one of the **New** menu commands, it is possible to select the style that will be used for the graph.



If the Use Settings From Style box is unchecked, default scale settings will be used; if checked the new graph will use the same scale settings as are defined in the stored style. By default the "Default" style is used. To use another style, press the **Style...** button to select the style on which the new graph is based.

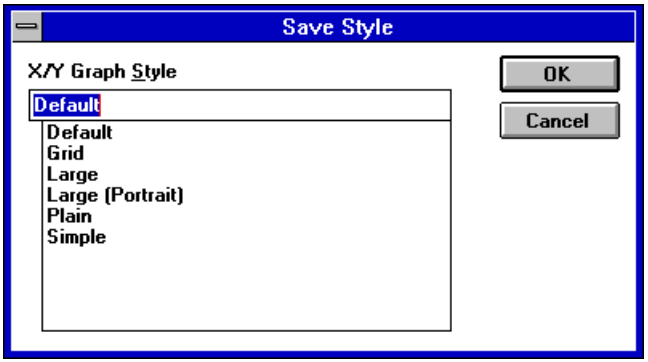


The picture on the right of the dialog box shows the graph style selected in the list box. Choose the style that is required, and press **OK**.

Saving Graph Styles

To save the style of a graph, or of a graph object (text, lines and boxes added with the **Add Text**, **Add Line**, or **Add Box** command):

1. Click on the required object or graph to select it
2. Choose the **Options Save Style...** command to give the following dialog box.

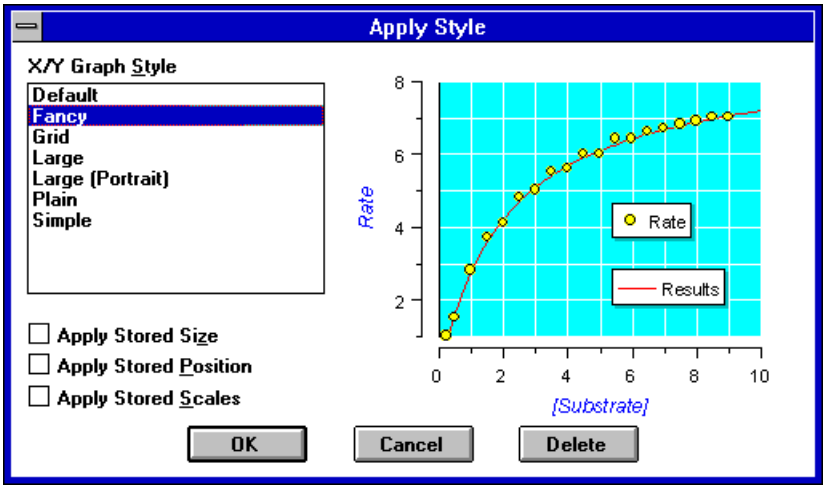


Enter a name for the style, and press **OK**. The name "Default" has special significance, and styles saved as Default will become the default style that is used by the program when it is next run. By saving styles as Default you can customize GraFit to produce graphs in your preferred manner.

Applying Graph Styles

To change the style of a graph or of a graph object (text, lines and boxes added with the **Add Text**, **Add Line**, or **Add Box** command) to one of the stored styles:

1. Click on the required object or graph to select it
2. Choose the **Options Apply Style...** command to produce the following dialog box.



The picture on the right of the dialog box shows the style selected in the list box.

Size/Position Options

The normal effect of applying a new style to an object is that the appearance of the object is altered, but not the size and position. However, if the Apply Stored Size box is checked, the size setting of the saved style is applied to the selected object. If the Apply Stored Position box is checked, the position of the selected object is set to be the same as that of the saved style.

Graph Scale Options

The normal effect of applying a new style to a graph object is that the axis scale settings are not affected. If the Apply Stored Scales box is checked, the graph scale will also be changed to that of the stored style.

Deleting a Style

To delete a style, select the style from the list and press the **Delete** button.

21 Data Fitting Basics

This chapter describes some of the basic principles behind data fitting. A more complete description of the theory is given in Chapter 29, and additional information may be found in several reviews.[1–6] The practical aspects of using GraFit to fit data are described in Chapter 22.

When Is Data Fitting Appropriate?

Data fitting is the mathematical analysis of a data set in order to analyze trends in the data values. This will usually involve linear or non-linear regression analysis of these data values, in order to define a set of parameter values that best characterize the relationship between the data points and an underlying theoretical model. While many types of data can be usefully presented in graphical form, not all such data are suitable for data fitting.

Types of Data

In general, data can be divided into three groups.

1. No Systematic Trend

There are many examples of such data, and two are displayed below.

[1]Cleland, W.W. (1967) *Adv. Enzymol.* **29,** 1-32.

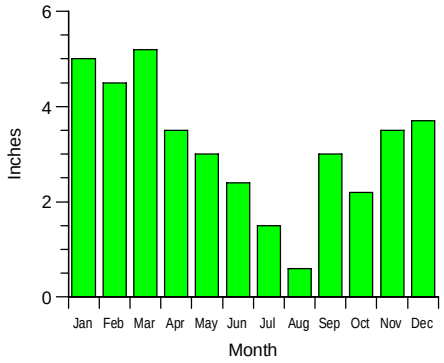
[2]Bevington, P.R. (1969) *Data Reduction and Error Analysis fo the Physical Sciences* McGraw-Hill.

[3]Duggleby, R.G. (1981) *Analytical Biochemistry* **110,** 9-18.

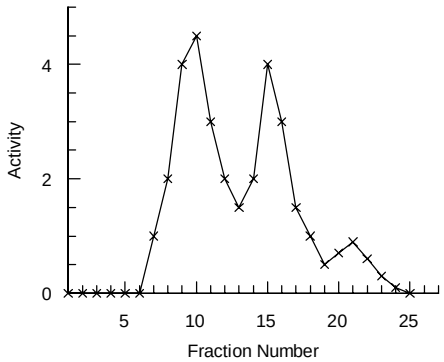
[4]Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1986) *Numerical Recipes. The Art of Scientific Computing* Cambridge University Press

[5]Motulsky, H.J. & Ransnas, L.A. (1987) *FASEB J.* **1,** 365-374.

[6]Leatherbarrow, R.J. (1990) *TIBS* **15,** 455-458.



A rainfall chart



An elution profile from a purification column

Such data are typically presented as bar or column charts, or as scatter plots with the data points joined by lines (or possibly by a spline curve). As no systematic relationship describes the data, they are not suitable for data fitting.

2. The Trend is Described by a Known Equation

This is the situation where data fitting is most useful. In this case, based on knowledge of the experiment, and the laws that govern it, the data are described by a known equation, i.e.

$$y = f(x, p_1, p_2, p_3, ...)$$

In non-mathematical terms, the observed data, y (also known as the dependent variable), varies as a function of some experimentally controlled value x (the independent variable) and one or more parameters p_1, p_2, p_3, etc. The exact form of the equation depends on the experiment and the conditions that are used: it might be a simple linear equation,

$$y = ax + b$$

or a more complex function. In either case, the purpose of data fitting is find values for the parameters in this equation that best describe the data. For a simple linear equation, as shown above, this is done by linear regression. In general, however, a more complex equation will apply. Such data are fitted by non-linear regression. GraFit allows non-linear equations of arbitrary complexity to be used for data fitting.

3. A Systematic, But Uncharacterized Trend is Present

Occasionally, data may be collected that show a distinct trend, but no known equation describes this relationship. For example, an empirically determined calibration curve may collected. If the physical basis for the shape of the data is not well defined, the appropriate equation is unknown. In such situations it can be useful to fit the data to a general purpose equation, such as a polynomial. This will allow values to be extrapolated from the standard curve, although any parameter values that are obtained will not have physical meaning. The use of data fitting for this purpose is illustrated in Chapter 30.

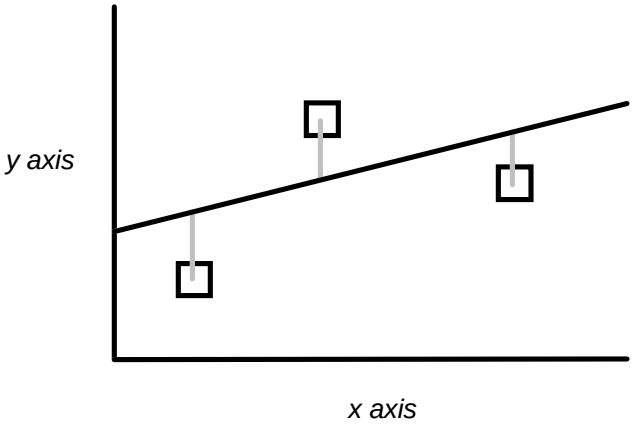
Regression Analysis

Regression analysis finds the "best fit" line or curve through a series of data. This is considered to be the one that minimizes the sum of the squares deviations of the experimental data points from the theoretical curve. GraFit uses regression to allow the analysis of experimental data.

Practical Considerations

The Errors are in the *y* Data

It is the convention to assume that the *x* data values represent the independent variable, and the *y* data values represent the dependent variable, i.e. the *x* data are what is varied in the experiment (for example time, concentration, etc.), and the *y* data are the experimentally observed values.



The vertical distance (the gray lines in the above diagram) between the data point and the curve or line is that which is minimized by normal regression analysis, i.e., it is assumed that all errors are present in the *y* data. An important aspect of experimental design is to ensure that the *x* data are as accurate as possible so that this assumption is valid.

The Data Values Have A Suitable Numerical Range

A consequence of computer calculation is that rounding errors can occur, particularly if very large or very small numbers are used. This is because computers only store decimal numbers

with a limited numerical precision. Although this should not affect most calculations, it is advisable to scale any set of very large or very small values prior to performing calculations. For example, if a series of time values were entered as 0 to 10^{-15} seconds, the rounding errors could be significant. Entering the data as 0 to 1 femtoseconds may be more appropriate.

Types of Regression Analysis

Regression acts to minimize the sum of the squares deviations of the experimental values from values calculated using some theoretical equation. However, the mechanics of the calculations are different for linear and non-linear equations.

Linear Regression

Linear regression is a technique that should be familiar to most scientists. It is the mathematical equivalent of using a ruler to draw the "best" line through a series of data points to obtain values for the slope and intercept of this line. The calculations needed to perform linear regression are relatively simple, and are even found on many scientific calculators.

Polynomial Regression

Polynomial regression is an extension of linear regression, where the equation is

$$y = a + bx + cx^2 + dx^3 + ex^4 + ...$$

Non-linear Regression

Non-linear regression allows data to be fitted in the more general case to any equation where the *y* data value can be described as a function of the *x* data and a series of parameters, i.e.

$$y = f(x, p_1, p_2, p_3, ...)$$

As a result, it is applicable to many more situations than linear regression, and is more generally useful for analyzing experimental data. The only drawback to non-linear regression is that the calculations are too complex perform using a calculator, and require a computer program such as GraFit.

Additional Considerations for Non-linear Regression

The more complex nature of the non-linear regression calculations do require some further considerations compared with linear regression.

Equation to Use

Non-linear regression can employ a vast range of different equations. It is therefore necessary to select an equation that is appropriate to the particular experimental situation before fitting the data.

Initial Estimates

The mathematical method used to perform non-linear regression acts to take starting estimates for the parameters in the equation, and optimize them during the calculations. As described in Chapter 23, GraFit provides a mechanism for providing these initial estimates automatically for some equations, but in principle it is necessary to give the calculations some rough estimates prior to data fitting.

Convergence Criteria

Unlike linear regression, non-linear regression does not provide an exact solution, but, by an iterative process, calculates successively better parameter values until a suitable tolerance is reached. The criteria for defining this tolerance are determined by the program, and can be specified when using GraFit.

Weighting Scheme

Both linear and non-linear regression analyses incorporate assumptions about the accuracy of the data points. Normally, they assume that all data points are equally accurate, and so carry equal weighting in the analysis. It is also possible to use other weighting schemes that give different weighting to one or more data values. A full description of weighting is found in Chapter 29.

Linear Regression of Transformed Data

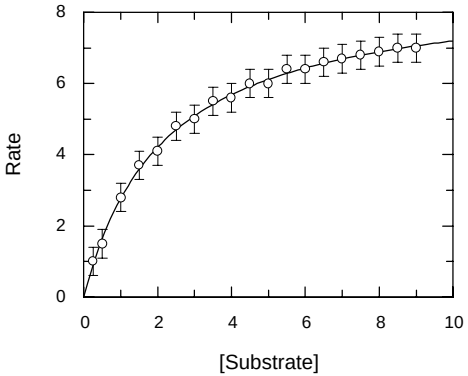
A familiar method of analyzing non-linear data graphically is to rearrange the data to a linear form, find the "best line" through the points, and so derive the parameter values. This is illustrated in the following example. Enzyme kinetic data are described by an equation of the form

$$y = \frac{Ax}{B + x}$$

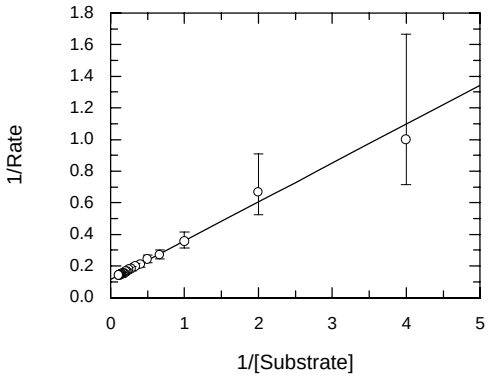
which is hyperbolic. This can be linearized by plotting $1/y$ versus $1/x$, as

$$\frac{1}{y} = \frac{1}{x} \cdot \frac{B}{A} + \frac{1}{A}$$

The ease of linear regression analysis makes it tempting to use in order to fit data that have been rearranged in this way. However, the results that are obtained are not statistically optimal. This is because the regression calculations make assumptions about the error distribution that are not usually valid after transformation. The distortion of the experimental errors by the above transformation is easily demonstrated in the following graphs.



Enzyme kinetics plot. The error is present only in the Rate (y axis) measurement. Errors are simple, i.e. all data points have the same error.



Double reciprocal (Lineweaver Burk) plot of the data above. Note the large distortion of the error bars resulting from this plot.

In this case, the error involved in the high data points of the transformed plot is much greater than in the low data points — this is simply a consequence of the reciprocal transformation. Linear regression (unless suitable compensation is made for the distorted errors) is therefore not an appropriate way to analyze the data, and for best results the original data should be fitted to the appropriate non-linear equation.

Information Obtained From Data Fitting

Data fitting provides the following information regarding the experimental data that are analyzed.

1. The values of the parameters in the equation used that best describe the data. These parameters may include (for example) rate constants, inhibition constants etc. whose values need to be determined.

2. The accuracy of these calculated parameters.

From this information it is possible to judge how well the data fit the experimental model (equation) that was used. In most cases, however, it is the numerical values of the parameters that are important, and whose values are required from the analysis.

22 Fitting Data Using GraFit

Once data are entered into the data table, either by typing in the data points or by reading in a data file, they can be fitted to one or more theoretical equations that model the results. GraFit fits data by linear regression, non-linear regression, or polynomial regression. Of these, non-linear regression is the most powerful, and the most generally applicable method of data analysis.

Preparing for Data Fitting

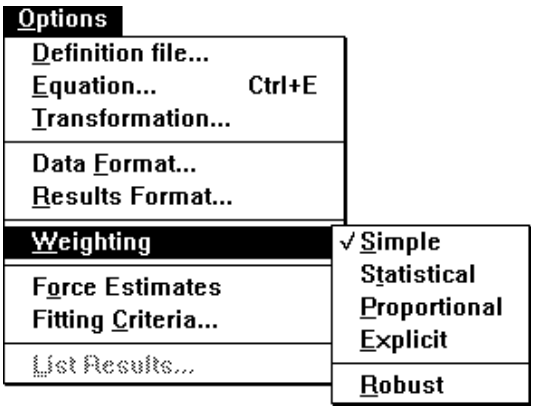
Before the data can be analyzed it is necessary to do the following.

- Enter the data values into the data table (see Chapter 3).
- Select the correct weighting.
- If using non-linear regression, select the appropriate equation.

Selecting the Correct Weighting

All types of regression analysis require information about the error distribution in your data. The reason for this is explained fully in Chapter 29, but briefly it is necessary to specify whether the inaccuracies are likely to be the same for each data point ("Simple Weighting"), are proportional to the value of the *y* data points ("Proportional Weighting") or have been individually estimated ("Explicit Weighting").

Select the weighting using the **Options Weighting** command menu. The current weighting is ticked.



Options Weighting Menu

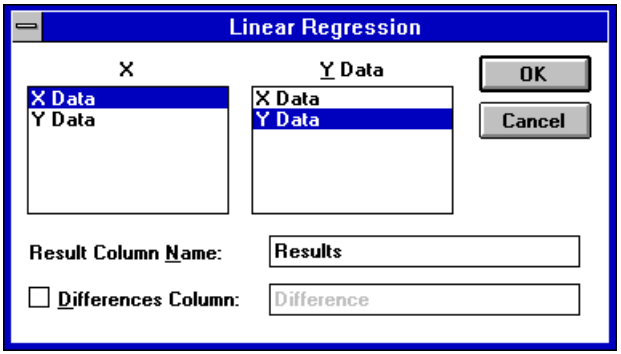
The default weighting selection of "Simple" is often the most appropriate setting.

Linear Regression

Linear regression is used to find the best line through a set of data. Although linear regression is a familiar technique, few experimental situations produce data that follow a linear trend. While it is often possible to rearrange non-linear data to produce a linear dependence, you should generally avoid doing so, and instead fit the data by non-linear regression to an equation that describes the situation directly (see Chapter 21 for an explanation).

Fitting the Data

To fit data by linear regression, select the **Data Linear fit...** command. This results in the following dialog.



It is necessary to indicate which of the data columns holds the *x* data, and which the *y* data. The parameter values that result from the analysis will be added to the data table as a new column. The name for this column is, by default "Results", but it is possible to change this by editing the name in the Result Column Name edit box.

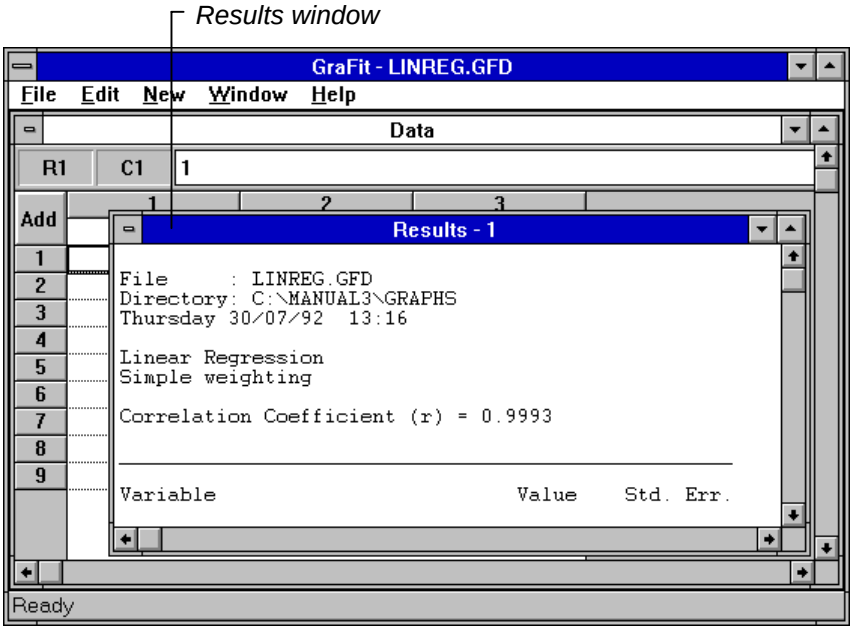
Calculated Differences

If the Differences Column box is checked, a column will be created that holds the difference between the experimental and the calculated *y* values. This can be useful to see if the data have been fitted satisfactorily, and for producing residual plots.

Results Listing

After data fitting you are given the option to create a full results listing describing the analysis.

The results listing is made to a new Results window. By default, this listing just gives the values for the fitted parameters, but if the List Calculated Values box is checked will also include a full list of the raw data plus calculated *y* data values.



Contents of the Results Listing (Linear Regression)

```
1       File     : LINREG.GFD
2       Directory: C:\MANUAL3\GRAPHS
3       Thursday 30/07/92  20:52

4       Linear Regression
5       Simple weighting

6       Correlation Coefficient (r) = 0.9993

        ____________________________________________________

7       Variable                          Value    Std. Err.
        ____________________________________________________

        Intercept                        3.1609       0.0755
        Slope                            0.9756       0.0150
        ____________________________________________________




        ____________________________________________

8                X            Y
              x data       y data     Calculated
        ____________________________________________

           1       1.0000       4.1584       4.1366
           2       2.0000       5.1291       5.1122
           3       3.0000       6.0986       6.0879
           4       4.0000       6.8696       7.0635
           5       5.0000       8.1436       8.0391
           6       6.0000       9.0842       9.0148
           7       7.0000      10.0054       9.9904
           8       8.0000      10.9216      10.9661
        ____________________________________________
```

The Results Listing (Linear Regression)

The results listing contains various sections (1–8).

1. The file name.
2. The directory where the file is located.

3. The current date and time. The format of both date and time (for example whether a 24 hour or 12 hour clock is used, and whether dates are written as Month/Day/Year or Day-Month-Year etc.) will depend upon the current format settings in the WIN.INI file. These settings are altered using the Control Panel application.
4. The type of analysis used.
5. The weighting employed.
6. The value for the correlation coefficient.
7. The main results, showing the values of the calculated parameters with the standard errors of these parameters.
8. A full listing of the experimental and calculated values for the complete data set. This section is only present if the List Calculated Values box was checked when creating the listing.

Number Format for the Results Listing

The format for the numeric values listed in the results is controlled using the **Options Results Format...** command (for a full description of numeric formatting, see Chapter 28). It is possible to relist the results after changing the formatting by selecting the **Options List Results...** command. This will create a new Results window using the revised format settings.

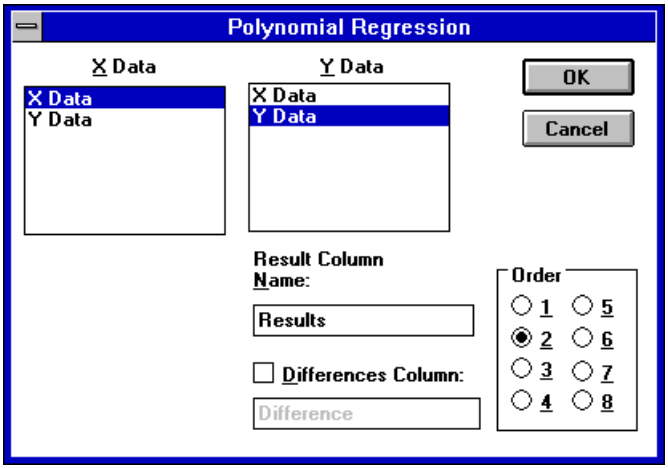
Polynomial Regression

Polynomial regression is simply an extension of linear regression, where the equation is the power series:

$$y = a + bx + cx^2 + dx^3 + \ldots + mx^n$$

The order of the polynomial is *n*, which means that a 1st order polynomial has two terms (*a* and *b*), and is equivalent to linear regression.

To fit data using polynomial regression, choose the **Data Polynomial fit...** command. The process of polynomial fitting is very similar to that described above for linear regression.



Select the columns that hold the *x* and *y* data values, and enter a name for the column that will hold the parameter results, in the same way as for linear regression. The order of the polynomial fitting is chosen from the Order section. Fitting may use any order of polynomial from 1 to 8.

Calculated Differences

If the Differences Column box is checked, a column will be created that holds the difference between the experimental and the calculated *y* values. This can be useful to see if the data have been fitted satisfactorily, and for producing residual plots.

Non-linear Regression

Non-linear regression is often the most appropriate way to analyze experimental data. Data are fitted using an equation that specifically describes how the data point values vary as a function of the independent variable (*x*) and one or more parameter values. Within GraFit it is possible to define any equation that is of the form

$$y = f(x, p_1, p_2, \ldots)$$

i.e., *y* is the dependent variable which is a function of *x* the independent variable, and p_1, p_2, ... the parameters describing the data. GraFit allows up to 30 parameter values in an equation, and permits up to 10 independent (*x*) variables. The process whereby new equations can be defined is given in Chapter 23, but for the most part, data analysis will access pre-defined equations that are stored on disk.

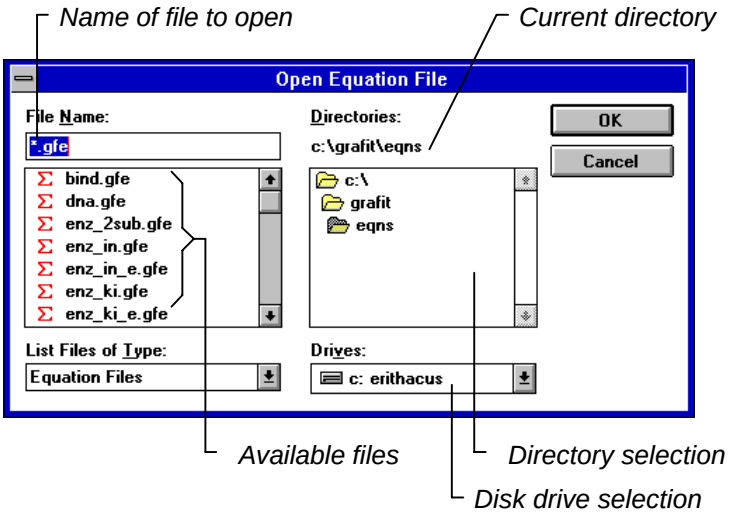
Choosing the Equation

Equations are stored in files that have the extension .GFE. Each equation file can hold any number of different equation definitions, limited only by available disk space. We suggest that you keep related equations in the same equation file, for ease of use. GraFit comes with a variety of equations that are ready for use. A full list of these is found in Chapter 32. The

process of loading an equation involves first selecting the definition file that holds it, and then choosing the equation from those available.

Loading the Equation Definition File

To do this choose the **Options** menu and select the **Definition file...** command. The Open Equation File dialog box is similar to that used for opening data files (Chapter 4).



The available files are classified according to type by the small icon drawn next to the file name.

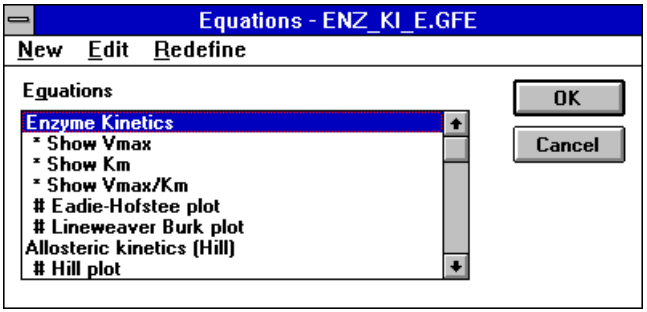
Icon	File type
Σ	GraFit equation file (.GFE). GraFit equation files hold equation and transformation definitions.
□	Unrecognized file type.

This classification is based on the three letter file extension used by the file.

Choose the equation file that is to be used, and press **OK**.

Selecting the Equation

After choosing an equation definition file you must select the specific equation to use. To allow this, a further dialog box appears containing a listing of the equations available in the current equation file.



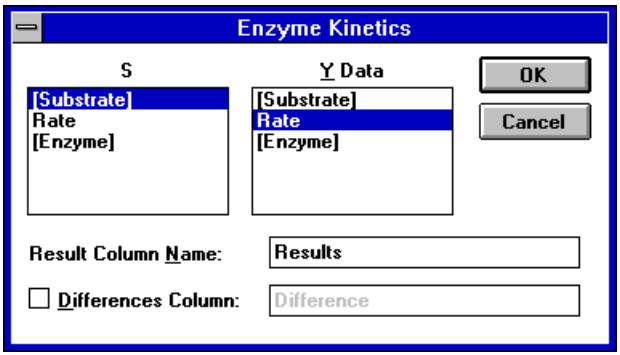
You can also obtain this dialog box by choosing the **Options** menu and selecting the **Equation...** command. Select the equation that you wish to use, and press **OK**. Alternatively, double click on the equation you require. Full details of how to view the equation definitions, and to define new equations, are given in Chapter 23.

Saving the Equation Used With the Data File

When you save your data file, GraFit will automatically record the equation and definition file that you are currently using. Reading back the data file will restore these settings.

Curve Fitting

To fit the data by non-linear regression, choose the **Data** menu, and select **Curvefit...**. The resulting dialog box, which differs slightly depending upon the equation involved, is shown below.



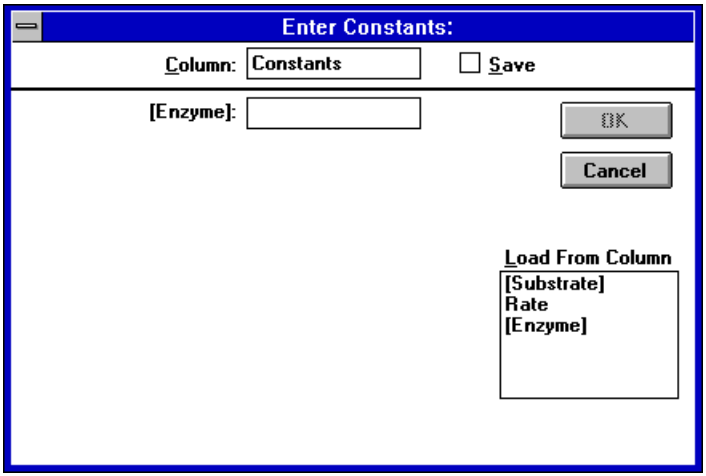
The equation used to fit the data is shown on the title of the dialog box. GraFit allows any of the columns of the data table to be used as the *x* and *y* data values. The dialog box contains two lists of the columns that are present — select one column for the *x* data, and one for the *y* data. The parameter values from curve fitting are stored in a data column, which allows their access for graph plotting. You can give a meaningful title to this column by entering a name in the Result Column Name edit box.

If the Differences Column box is checked, a column will be created that holds the difference between the experimental and the calculated *y* values. This can be useful to see if the data have been fitted satisfactorily, and for producing residual plots.

Note If explicit weighting was selected, the previous dialog box would have a third column from which to select the individual error values.

Entering Values for Constants

Some equations are defined to use "Constants". The values for these are passed to the program during data fitting, and so are useful for values that are likely to change with different data sets, for example concentration of a reagent, temperature etc. If the equation that if being used has constants, a dialog box will appear into which you must enter appropriate values before proceeding.



Saving Entered Values to the Data Table

The upper part of the dialog box allows you to store any values that are entered for further use. If the Save box is checked these values will be stored in a column; the name of the column can be entered in the Column edit box.

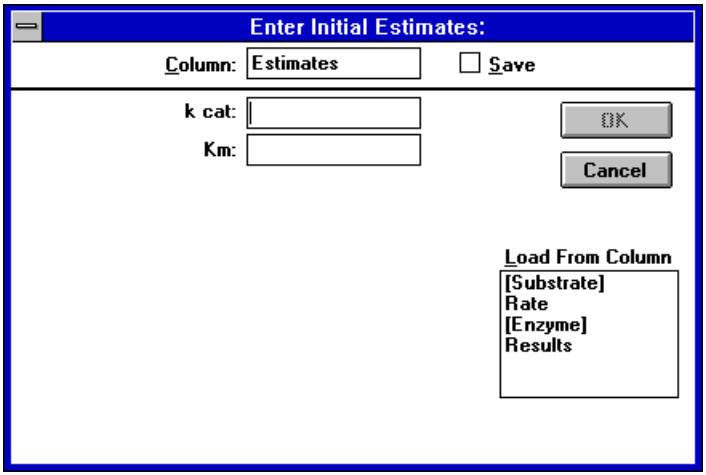
Loading Values From the Data Table

To the right of the dialog box is a list box labeled Load from Column. This contains a list of the current column names in the data table. If you click on a column name in this box, the data contained in this column will be used to fill in the values in the edit boxes.

Entering Initial Estimates

Unlike linear regression, non-linear regression requires approximate starting estimates for the parameter values. These starting values are then improved upon by iteration, until the best-fit curve is calculated. As will be seen in Chapter 23, within an equation definition it is possible to allow the program to provide these estimates automatically. However, for some equations it is difficult to do this, and so estimates will need to be given explicitly. It is also possible to force the program to allow manual entry of initial estimates by choosing the **Options Force Estimates** command.

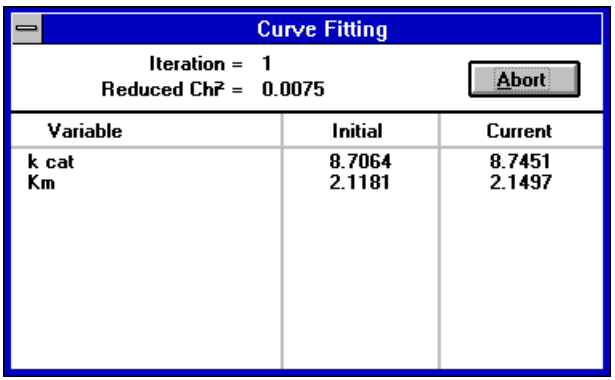
When initial estimates are required, the following dialog box will appear.



This dialog box allows values to be entered in the same way as described above for entering constants. Enter values that approximate the final result. Just how accurate these estimates need to be depends upon the equation in use, but it is often enough if they are correct to within an order of magnitude.

Progress of Data Fitting

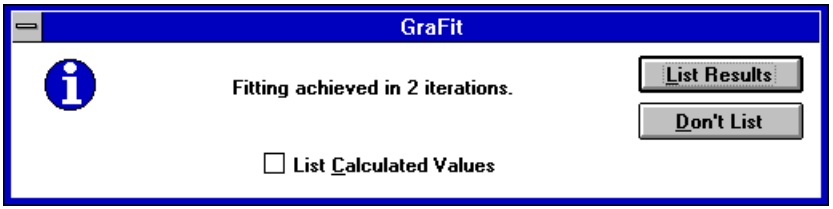
Once initial estimates have been provided, the fitting procedure begins and a dialog box appears showing the progress of the calculations. After each iteration of the data fitting cycle, the dialog box is updated to show the current parameter values.



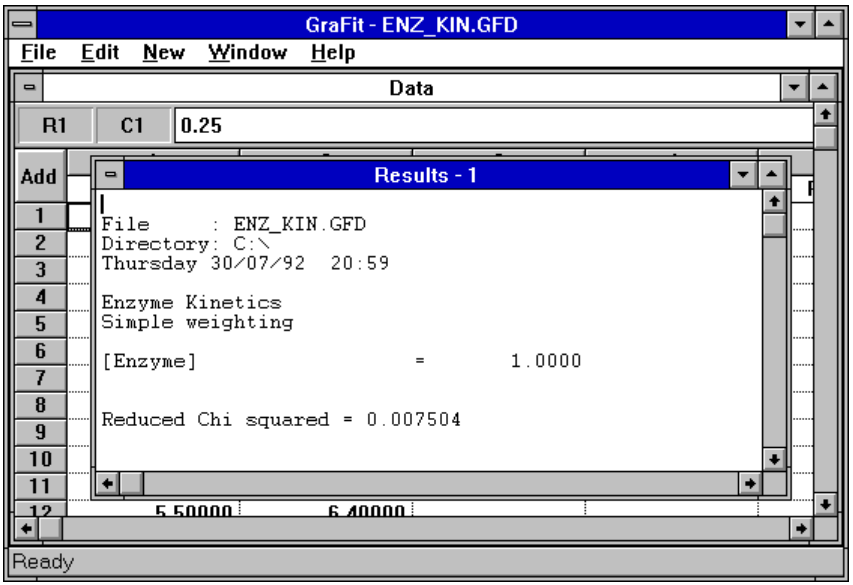
The **Abort** button may be pressed at any time to stop the calculations, if required.

Listing the Results

The calculations proceed until an acceptable fit is produced (the criteria for fitting are user definable, see p154). When fitted, the following dialog box appears.



To list the results to a new Results window, as shown below, press the **List Results** button. By default this listing just holds the fitted parameter values, but if the List Calculated Values box is checked, it will include a full list of the starting data, together with theoretical values calculated using the fitted results.



Please Note If the initial estimates for the fitting are inappropriate it is possible to "fit" the data with values that do not describe the data very well. We strongly recommended that you view a graph of the results to examine the fit obtained. It is then possible to see whether the calculated values are appropriate to describe the data. The process of creating a graph is described in Chapter 6. It can also be useful to examine the residuals to ensure that they are randomly distributed (see Chapter 29).

Contents of the Results Listing (Non-linear Regression)

```
1       File     : ENZ_KIN.GFD
2       Directory: C:\
3       Thursday 30/07/92  20:59

4       Enzyme Kinetics
5       Simple weighting

6       [Enzyme]                  =       1.0000


7       Reduced Chi squared = 0.007504

        ____________________________________________________

8       Variable                           Value    Std. Err.
        ____________________________________________________

        k cat                             8.7459       0.0889
        Km                                2.1503       0.0687
        ____________________________________________________




        ____________________________________________

9                S            Y
             [Substrate]      Rate      Calculated
        ____________________________________________

           1       0.2500       1.0000       0.9109
           2       0.5000       1.5000       1.6500
           3       1.0000       2.8000       2.7762
           4       1.5000       3.7000       3.5939
           5       2.0000       4.1000       4.2145
           6       2.5000       4.8000       4.7017
           7       3.0000       5.0000       5.0944
           8       3.5000       5.5000       5.4175
           9       4.0000       5.6000       5.6881
        ...etc.
```

The Results Listing (Non-linear Regression)

The results listing contains various sections (1–9).

1. The file name.
2. The directory where the file is located.
3. The current date and time. The format of both date and time (for example whether a 24 hour or 12 hour clock is used, and whether dates are written as Month/Day/Year or Day-Month-Year etc.) will depend upon the current format settings in the WIN.INI file. These settings are altered using the Control Panel application.
4. The type of analysis used.
5. The weighting employed.
6. The values for any constants used by the equation.
7. The value for the reduced chi^2 value, χ^2_v
8. The main results, showing the values of the calculated parameters with the standard errors of these parameters.
9. A full listing of the experimental and calculated values for the complete data set. This listing is only present if the List Calculated Values box was checked when creating the listing.

Number Format for the Results Listing

The format for the numeric values listed in the results is controlled using the **Options Results Format...** command (for a full description of numeric formatting, see Chapter 28). It is possible to relist the results after changing the formatting by selecting the **Options List Results...** command. This will create a new Results window using the revised format settings.

Speed of Calculations

The time taken for curve fitting depends upon the following.

- The type and clock speed of the central processor fitted in your computer.
- Whether you have a numeric coprocessor. A numeric coprocessor will speed up mathematical calculations by around a factor of 4. A coprocessor is integral to the 80486 processor.
- The number of data points present.
- The complexity of the equation. With more parameters the time taken to fit the data increases. In addition, the use of transcendental mathematical functions (log, sin etc.) slows the calculations down considerably, particularly if no coprocessor is present.

However, in general, the time taken for a typical calculation should be no longer than a few seconds.

Results Windows

The results from data fitting are presented as a separate Results window, which functions in a similar manner to the Windows Notepad application. Text may be added, either by typing it in or by pasting it from the Clipboard. The contents of any Results windows are saved with the data file.

Results can either be printed directly, or pasted to another application such as Windows Write, where special effects (such as alteration of fonts) can be added before printing. This facility to transfer data between applications is a feature of Microsoft Windows, and it can be put to good use in this instance if you wish to prepare high quality output suitable for publication.

Printing the Results

When the Results window is the active (uppermost) window, the **File Print Results...** command will print the contents to the currently selected printer. For more information about printing, see Chapter 5.

Pasting the Results into Another Application

To paste the contents of the Results window to another Windows application:

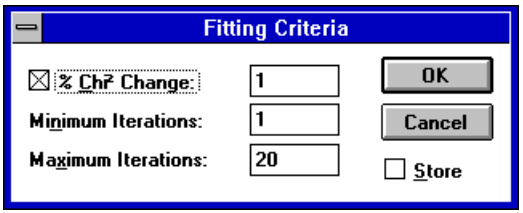
1. With the Results window selected choose the **Edit** menu and select the **Select All** command.
2. Choose the **Edit** menu and select **Copy**.
3. Activate the application that you wish to send the Results text to, and select **Paste**.

Deleting a Results Window

To remove a Results window, select the window and choose the **File Close** command.

Setting the Convergence Criteria for Non-linear Fitting

As described above, non-linear regression calculates best-fit parameters by a series of iterations that produce successively more accurate results. It is up to the program to decide at what point the values are deemed acceptable, and at which point the calculations stop. The criteria used to decide this are set using the **Options Fitting Criteria...** command. This gives the following dialog box.



The maximum number of iterations that will be used is entered in the Maximum Iterations box. The minimum number is entered in the Minimum Iterations box. If the % Chi^2 Change box is checked, the fitting procedure will also terminate when the change in the χ_v^2 value is less than the percentage value entered (providing the minimum number of iterations has been reached). If this box is unchecked, the program will continue until the maximum number of iterations is reached.

Setting the Default Fitting Criteria

If the Store box is checked, the values entered will be saved as the defaults for when GraFit is next run. If this box is not checked, the modified settings will only apply to the current session of GraFit.

23 Defining Equations

GraFit is supplied with a variety of equations for non-linear regression already defined on disk. However, there will be occasions where these predefined equations do not cover your particular experiment, and a new or modified equation is required. GraFit offers complete flexibility in the definition of equations — all the equations supplied can be edited to see how they are defined; these equations can be copied, and completely new equations added.

How Equations are Stored

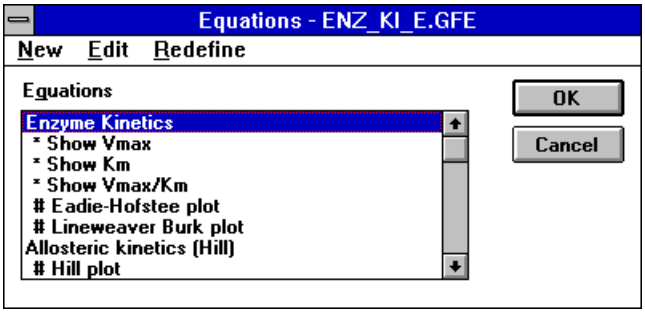
As described in Chapter 22, equation definitions are stored in files on disk, by default in files with the extension .GFE. These files, which also hold transformation definitions (see Chapter 26), can hold any number of equations. We suggest that related definitions are kept together in the same definition file for ease of access. Commonly used equations can be incorporated into several definition files if required.

Creating a New Definition File

To create a new, empty, definition file, use the **Options Definition File...** command, and enter the new file name. If this file does not already exist, a new file will be created.

Editing an Equation Definition

To edit an equation, select the **Options** menu and choose the **Equation...** command (if necessary, first select the appropriate definition file using the **Options Definition file...** command as described in Chapter 22). The resulting dialog box is shown below.



The menu bar on this dialog box allows equations to be created, copied and edited.

Edit Menu

The **Edit** menu item on this dialog box functions similarly to **Edit** on the main menu bar of Windows programs. The following commands can be selected.

Edit Menu

Cut

The **Cut** command will cut the currently selected equation definition to the Clipboard. This command is only active if a definition is selected from the equation list box.

Copy

The **Copy** command will copy the currently selected equation definition to the Clipboard. This command is only active if a definition is selected from the equation list box.

Paste

The **Paste** command will paste an equation from the Clipboard into the current definition file. The equation in the Clipboard must have been placed there by the **Cut** or **Copy** commands. If no equation is present in the Clipboard this command is inactive, and is shown grayed.

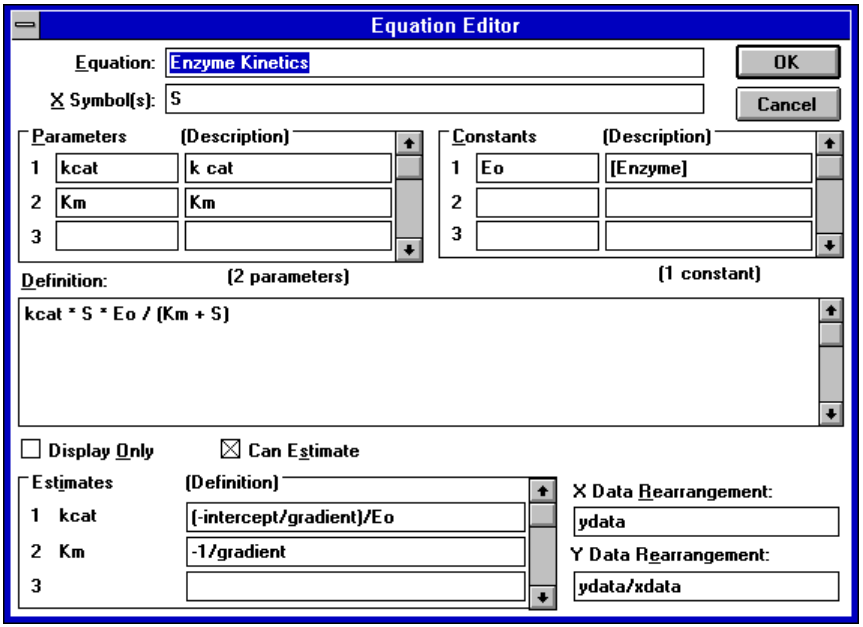
Duplicate

This command makes a second copy of the currently selected equation. It is only active if a definition is selected from the equation list box.

Hint To build a new definition file incorporating various equations from existing files it is convenient to run two or more copies of GraFit, and to cut and paste equations between them. Remember that Windows allows you to run as many programs at once as can fit into the available memory.

The Redefine and New Options

Redefine allows the currently selected equation to be viewed and altered. **New** will create a new, empty equation definition. Selection of either option results in the appearance of the Equation Editor dialog box; if **New** is selected this dialog box is empty, whereas if **Redefine** is chosen the currently selected equation is listed in the definition box, and can be edited. Selecting **Redefine** produces the Equation Editor, as shown below.



Defining An Equation

The Equation Editor holds all the information needed to define an equation used for curve fitting. Equations, as used by the GraFit program, are of the form

$$y = f(x, \text{param}_1, \text{param}_2, ..., [\text{constant}_1, \text{constant}_2, ...])$$

where *f* means "is a function of", *y* is the dependent variable, *x* is the independent variable, param_1... are the parameters in the equation that we are trying to determine and constant_1... are constants. The expression must always include at least one parameter, but need not include constants if the equation has no use for them A GraFit equation may include up to 30 parameters, 10 constants, and between 1 and 10 independent variables (*x* variables).

The equation that is shown in the Equation Editor (above) is the Michaelis-Menten equation of enzyme kinetics.

$$v = \frac{k_{cat}[S][E_0]}{K_m + [S]}$$

In this equation, the rate of reaction, *v*, varies with the concentration of substrate, [S], depending upon the catalytic rate constant, k_{cat}, and the Michaelis constant, K_m. The rate also depends upon the enzyme concentration, $[E_0]$, which is fixed for a given experiment. This equation can be broken down as follows.

Role	Symbol in this equation	Name to be used within the Equation Editor
x data	[S]	S

y data	*v*	*(not required)*
Parameters	k_{cat}, K_m	`kcat, Km`
Constants	$[E_0]$	`Eo`

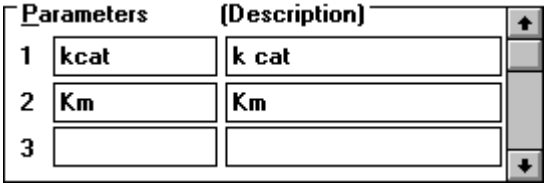
To use this equation it must be entered in a form that can be understood by the program. All equations within GraFit must be of the form *y*=... so in the equation definition it is only necessary to enter the right hand side of the expression. This must be entered in a stylized manner without the use of subscripts etc., and so for this equation we get:

```
kcat * S * Eo /(Km + S)
```

In general, the syntax used for defining equations is similar to that of BASIC. A complete description follows.

Defining Parameters

For the program to be able to use the equation definition we must also give it a list of the parameters that are present, i.e. define what is meant by k_{cat}, K_m etc. This is done in the section of the dialog box labeled Parameters.



The first three parameters are displayed, and the scroll bars can be used to bring any additional definitions into view. For each parameter there are two edit boxes that need to be filled in. Into the first is entered the name of the parameter as it is used within the equation definition. In the next it is possible to enter a fuller text description. The fuller descriptive name is used when the program is listing results, etc. However, in the equation definition itself, a symbolic name is used to represent the parameter. A symbol name must begin with a letter, contain only letters and numbers, and have no space characters in it. The full descriptive name may contain any characters and spaces.

Up to thirty different parameters may be present. It is important, however, that any parameters that are defined should occupy the first few places; for example, if three parameters are present they should be defined as parameters 1–3 rather than 2, 5 and 7.

Defining the X Symbol(s) (Independent Variables)

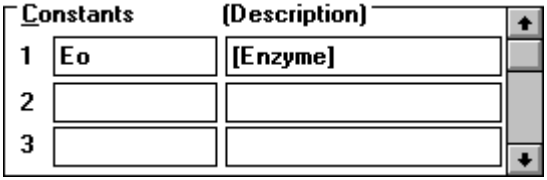
As the equation is a function of *x*, it is also necessary to let the program know the name used to represent the *x* axis. This must be defined as a symbol, so that it can be used within the equation definition. In the equation definition dialog box shown previously, this symbol is defined as `S`.

Equations With Several Independent Variables

If an equation had more than one independent variable, this would be indicated by entering several X Symbols, each separated by a space.

Defining Constants

Constants are used to allow the entry of a value into an equation definition that is fixed in any single experiment, but varies from experiment to experiment. This is the case for $[E_0]$ in the enzyme kinetics equation. When constants are included in an equation, the program will prompt for their values to be entered prior to curve-fitting the data. Up to 10 constants may be present, although many equations will include no constants.



The definition of constants is performed similarly to that of parameters. Each constant is given a full descriptive name together with a symbol name that is used in the equation definition.

Rules for Symbol Names

Symbol names are used to represent parameters, constants and independent variables within equation definitions. In addition, several predefined symbol names may be accessed. The rules for symbol names are similar to those of variable names within a BASIC program.

- The name must begin with a letter, and contain only letters and numbers. In particular, a symbol name cannot contain the space character.

  The following names are legal symbols:

  ```
  Fred
  k_on
  x23
  ```

  The following names, however, are invalid:

`2nd`	(starts with a number)
`k off`	(contains a space)
`test-2`	(contains a minus sign)
`ten:`	(contains a punctuation symbol)

- No distinction is made between upper and lower case characters within a symbol definition. Therefore the symbol names grafit, GraFit, GRAFIT and gRAfIT are all considered identical.

We recommend that you use meaningful names wherever possible when defining symbols. Using `Final` to represent the final concentration of a substance rather than `z` will make

understanding your definitions far easier for others as well as yourself. In addition, you should give a full descriptive name for your parameters and constants.

The Equation Definition

The equation definition is built up from the symbols representing parameters and constants (if any) with whatever arithmetic operators and mathematical functions are required. The syntax used is similar to that of BASIC.

Arithmetic Operators

Symbol	Operation	Example	Explanation
+	Addition	`x + y`	Add `x` and `y`
-	Subtraction	`lim-off`	Subtract `off` from `lim`
*	Multiplication	`m * x`	Multiply `m` by `x`
/	Division	`a/b52`	Divide `a` by `b52`
^	Exponentiation	`23^N`	Calculate 23^N

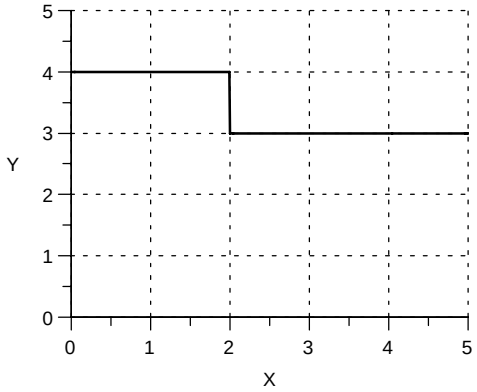
Logical Operators

Symbol	Operation	Example
>	Is greater than	`k>7`
<	Is less than	`x_value<con1`
>=	Is greater than or equal to	`test>=23.87`
<=	Is less than or equal to	`rate <= maxrate`
=	Is equal to	`i76=0`

Logical operators allow a comparison to be made. The result of the comparison is either 0 if FALSE, or -1 if TRUE. To see how these operators work, consider a graph of the form

```
(x<test)*-1 + 3
```

where `x` is the *x* axis, and `test` is a constant that has the value 2. This function is plotted below.



These comparisons can be used to introduce a discontinuity into a curve, and though they will seldom be used in equations, are extremely useful for displaying limits etc. to annotate a graph. An alternative method to obtain discontinuities is to use the `if then endif` constructs described below. Some examples of the use of such annotations may be found in Chapter 30. ***Note*** to obtain a sharp discontinuity it is necessary to create the curve with **Resolution** set to **High** (see Chapter 14).

Numbers may be entered directly into the equation definitions, and can include exponents, e.g. `27.87e-12`.

Functions

A large number of functions are provided for use in your equation definitions.

Function	Returns
`abs(x)`	Absolute value of `x`, i.e. if `x` < 0 returns `-x`, otherwise returns `x`.
`acos(x)`	Cos^{-1} of `x`.
`alog(x)`	Antilogarithm to base 10 of `x`.
`asin(x)`	Sin^{-1} of `x`.
`atan(x)`	Tan^{-1} of `x`.
`cos(x)`	Cosine of `x`.
`exp(x)`	e^x.
`frac(x)`	The fractional part of `x`. The fractional part of 13.387 is 0.387.
`int(x)`	The integer part of `x`. The integer part of 13.387 is 13.
`log(x)`	The logarithm to base 10 of `x`.
`ln(x)`	The natural logarithm of `x`.
`sin(x)`	Sine of `x`.
`sqr(x)`	The square of `x`, x^2.
`sqrt(x)`	The square root of `x`, $\sqrt{x}$
`tan(x)`	Tangent of `x`.

The expression within the brackets may be of arbitrary complexity, for example

```
exp(-k * sqrt(j - 1))
```

All the geometric functions (sin, cos etc.) use radians as their units.

The following predefined constants are also available:

Constant	Value
`_pi_`	π = 3.1415926536
`_e_`	e = 2.7182818285
`_rnd_`	A random number between 0.0 and 1.0.

Multi-line Equation Definitions

In its simplest form as described earlier, the equation definition is a single line that calculates the *y* value. For example, if the equation to be used was $y = ax + b$, then the equation definition might be written as

```
a * x + b
```

where `a` and `b` have been declared to be the symbols that represent the parameters, and `x` has been declared to be the symbol that represents the *x* data. Note that only the right hand side of the equation needs to be written, and that the multiplication between *a* and *x* needs the explicit inclusion of the multiplication operator, `*`. The equation definition may also be defined over several lines in order to simplify a complicated definition. This allows the calculation to be broken up into several steps.

Use of Temporary Variables

Temporary variables can be used to hold intermediate calculation results and simplify definitions. For example, the following definition uses three temporary variables, `a`, `b` and `c`.

```
a := 1
b := Kd + total + Capacity
c := total * Capacity
return( -(-b + sqrt( sqr(b) - 4 * a * c )) / (2*a) )
```

The remaining symbols (`Kd, total, Capacity`) are defined in the equation editor as parameters or constants. Values are assigned to the temporary variable using the assignment operator, `:=`, after which they may be used in expressions as normal. Use of temporary variables makes the above definition more easily read, and speeds up the calculation as the value for `b` need only be calculated once. The `return(`*`value`*`)` function is used to return the calculated *y* value to the program. Each line of the definition is separated from the next by pressing the RETURN key.

Note Do not confuse the assignment operator (`:=`) with the logical equality operator (`=`).

Conditional Statements

The values that are evaluated can be controlled using the `if` statement. This takes the form

```
if condition then
.....statements.....
endif
```

or

```
if condition then
.....statements.....
else
.....statements.....
endif
```

For example, it is possible to return different values to program depending upon the value of a parameter, constant or independent variable:

```
if extra=1 then
  return( a1 + b*x )
endif
if extra=2 then
  return( a2 + b*x )
endif
if extra=3 then
  return( a3+ b*x )
endif
return( a4+ b*x )
```

Comments

Comments can be added to annotate the equation definitions.

```
' Anything after an apostrophe is a comment
```

Looping Statements

The `for` statement has the following syntax.

```
for varname = startValue to endValue
........statements.......
next
```

or alternatively

```
for varname = startValue to endValue step stepValue
........statements.......
next
```

Goto Statement

Another statement that will be familiar to programmers is the `goto` statement. In GraFit equations this is implemented in the following manner:

```
:labelName
......statements.......
goto labelName
```

`labelName` can be any valid name, using the same rules as apply to the names for symbol names. The label may come before or after the `goto` statement.

Note Indiscriminate use of `goto` could result in an infinite loop.

Providing the Program With Initial Estimates

Non-linear regression works by taking initial estimates for the best-fit parameter values, and by an iterative process calculates successively better values until the best-fit values are reached. It is therefore necessary to provide the program with suitable starting estimates, which should generally be of the same order of magnitude as the actual ones. If an equation definition alone is provided, the program cannot know what initial estimates are appropriate. In this event, when the equation is used the program will ask you to enter initial estimates before fitting the data.

To relieve users the inconvenience of estimating parameter values, GraFit provides a means of generating these estimates automatically. It should be noted that the method used is not foolproof, and requires familiarity with manipulating equations. It does, however, make data fitting simpler, and can be over-ridden if the estimates provided are inappropriate. In addition, equations can be defined without provision for automatic initial estimation.

Method Used to Obtain Initial Estimates

The method used is based on performing linear regression on rearranged data. It is often possible to rearrange data in such a way that a linear plot is obtained, with the parameter values being related to the gradient and intercept. As linear regression needs no initial estimates, this can be used to estimate these parameter values providing the program knows the following.

1. What rearrangement is needed.
2. How the parameter values relate to the gradient and intercept.

This process is illustrated using the previous example, where the equation is:

$$v = \frac{k_{\mathrm{cat}}[\mathrm{S}][\mathrm{E}_0]}{K_{\mathrm{m}} + [\mathrm{S}]}$$

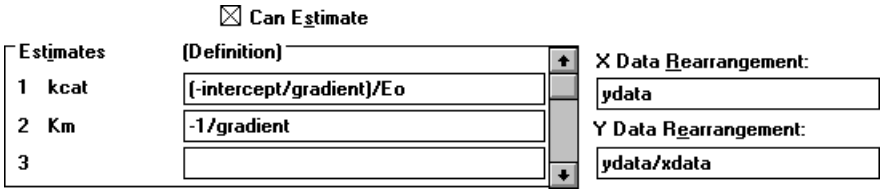
We can rearrange this to a linear form by plotting $v/[\mathrm{S}]$ against v:

$$\frac{v}{[\mathrm{S}]} = \frac{k_{\mathrm{cat}}[\mathrm{E}_0]}{K_{\mathrm{m}}} - \frac{v}{K_{\mathrm{m}}}$$

which is the so-called Eadie-Hofstee equation. Plotting v/[S] against v and allows the slope and intercept to be found, from which it is possible to calculate k_{cat} as (-intercept/gradient)/[E_0], and K_m as -1/gradient. It must be stressed that this approach does *not* provide strictly correct values for k_{cat} and K_m — this is why we wish to use non-linear regression in the first place! (If you are unclear about this consult Chapter 21.) The values provided are suitable, however, to use as initial estimates for the non-linear regression analysis. (It is instructive to compare the initial estimates provided by this method with the final values produced by the non-linear regression analysis. As they are demonstrably different, this illustrates the danger of using linear regression of a linearized function for analyzing data.)

The Initial Estimate Section

To specify that initial estimates are provided for an equation, the Can Estimate check box is used. When this box is checked the Initial Estimation section of the equation editing dialog box becomes visible, and is used to indicate how initial estimates may be obtained from the data.



If you do not want the program to calculate initial estimates, simply leave the Can Estimate check box unchecked.

Defining How to Rearrange the Data Points

For our current equation, the rearranged plot that we require is of v/[S] against v, i.e., y/x versus y. The edit boxes marked X Data Rearrangement and Y Data Rearrangement are used to define the data points for such a plot (note that the stored data points in the data table are, of course, unaffected). For this equation the definitions are:

```
ydata
```

and

```
ydata/xdata
```

respectively. The symbols "`xdata`" and "`ydata`" are predefined by the program to represent the x and y data points. These definitions allow the program to perform the rearranged plot, and any valid single line expression can be used. The edit boxes next to the parameter names are used to define how we obtain our initial estimates from linear regression on the rearranged plot. Shown in the last figure is the definition of k_{cat} as `(-intercept/gradient)/Eo`, and K_m as `-1/gradient`. Use the scroll bar to bring any further parameters into view. The symbols "`intercept`" and "`gradient`" are predefined, and represent the intercept and gradient of the rearranged linear plot.

Two further predefined symbol names may be used when defining initial estimates of your variables. The symbol `min` will give the value of the minimum *y* data point, and `max` the value of the maximum *y* data point. These values can be of benefit if you need to obtain rough estimates of limits, where the limiting value may be approximated by the maximum recorded value, etc. It is also possible to enter numeric values directly as estimates. In some situations, rough numerical values provide suitable estimates without recourse to the more complex methods described above.

Compiling the Equation

When you have finished defining the equation press the **OK** button to compile it (the **Cancel** button will abandon any changes that you have made, leaving the original definition unaltered). Compilation translates your definition into an internal form that can be executed rapidly. During compilation the definition is comprehensively checked to ensure that the correct syntax has been used. If an error is detected a message box will be displayed informing you of the mistake; after pressing the **OK** button to continue the cursor will be placed at the most likely place of error.

An Error Message Box

To obtain further information on the specific error, press the `F1` key when the error message box is present to obtain specific help information.

If compilation proceeds without error the new equation definition is stored on disk, and is selected as the current active equation for non-linear data fitting.

Equations for Display Only

When plotting graphs it can be useful to use equation definitions to allow modified curves, limits and tangents to be drawn on a graph. These equations are defined in the same way as normal equations, but need to have the Display Only box marked.

Definitions designated "Display Only" cannot be used for fitting equations. The compilation proceeds with some relaxation of error checking compared with normal equation definitions, as it is possible to define equations for Display Only that do not use all the defined parameters or constants. Display Only equations are used further in Chapters 30 and 32.

24 Automated Repetitive Data Fitting

When several data sets have been collected, it is often necessary to analyze each in turn. GraFit provides a mechanism for repetitive fitting of a series of multiple data sets by non-linear regression, and the production of and printing results and graphs from these data in an automated fashion.

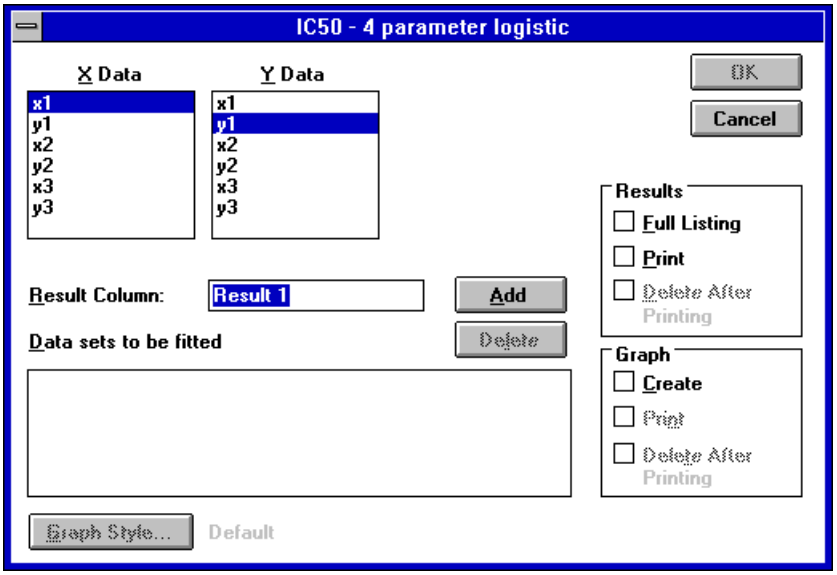
Requirements for Automated Fitting

The following requirements are necessary before fitting a series of data sets.

1. All the data must be present in the data table. If these are in separate files they can be merged using the **File Merge...** command.
2. All the data need to be fitted using the same equation. This equation must be able to generate initial estimates for the parameters (see Chapter 23).
3. The weighting used must be the same for each data set.

Procedure for Automated Fitting

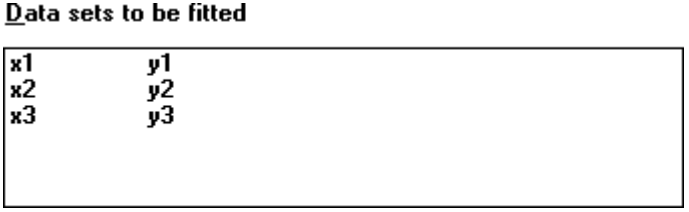
To fit a multiple series of data sets, select the **Data Multi Curvefit...** command. This results in the following dialog box.



The data sets that are to be fitted are specified from this dialog box.

Choosing the Data Sets for Fitting

The first *x* and *y* data columns are selected from the list boxes, and the name that will be used for the column created to hold the fitted parameter values is entered in the Results Column edit box. When the **Add** button is pressed this information is added to the list of data sets to be fitted. This procedure is repeated as required until the list of data sets to be fitted is complete.



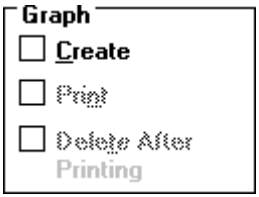
Any incorrect entry in this list may be deleted by selecting this entry and pressing the **Delete** button.

Specifying the Output

Results

Unless the Full Listing box is checked, only the fitted parameter values will be produced in the Results listings; if checked the results will include a full list of experimental and calculated values. If the Print box is checked, the results listings will be sent to the printer, and the Delete After Printing box will be activated. Checking the Delete After Printing box results in the Results listings windows being destroyed after they have been printed — this is often convenient to avoid the screen display becoming too cluttered.

Graphs



If the Create box is checked, a graph will be created to show the fitted curve. The style, size, position and scale settings of the graph are controlled by the **Style...** setting (see Chapter 20 for further information about graph styles). It can be useful to create a specific style for displaying the results to best effect, rather than relying on the default graph style. If a graph

is created, you have the option to print a copy of the graph, and to delete the graph window after printing is complete. Unless only a few data sets are fitted it is advisable to delete the graph in order to save memory and to prevent a confusing number of superfluous windows.

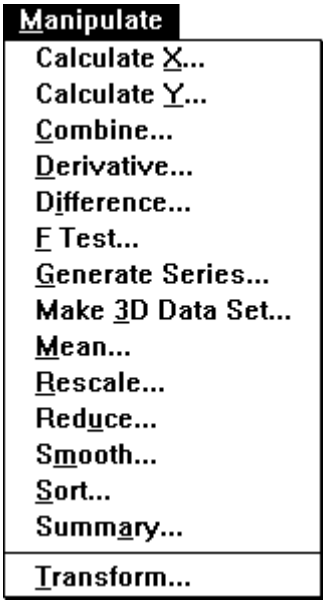
Fitting the Data

Once the data sets have been selected, they are fitted in turn. It is possible to abort the fitting at any time by pressing the **Abort** button or by canceling printing. At the end of the analysis a report is produced that shows the progress of the fitting procedures.

```
M u l t i f i t   l o g
Wednesday 23/09/92  20:53

Multiple fit number 1
Fitting achieved in 4 iterations.

Multiple fit number 2
Fitting achieved in 3 iterations.

Multiple fit number 3
Fitting achieved in 2 iterations.

Multiple fit number 4
Fitting achieved in 3 iterations.

Multiple fit number 5
Fitting achieved in 5 iterations.

Multiple fit number 6
Fitting achieved in 5 iterations.

Multiple data fitting of 6 data set(s) successfully completed.
```

25 Data Manipulation

GraFit provides several commands that allow manipulation of the data stored in the data table. These commands are accessed from the **Manipulate** menu, and are described below (except for the **Transform...** command, which is described in Chapter 26).



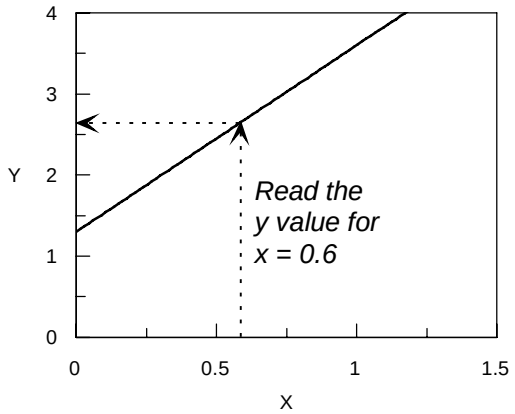
p175
p174
p177
p177
p178
p178
p181
p181
p182
p183
p185
p185
p186
p187
Chapter 26

Reading Values From the Calculated Curve

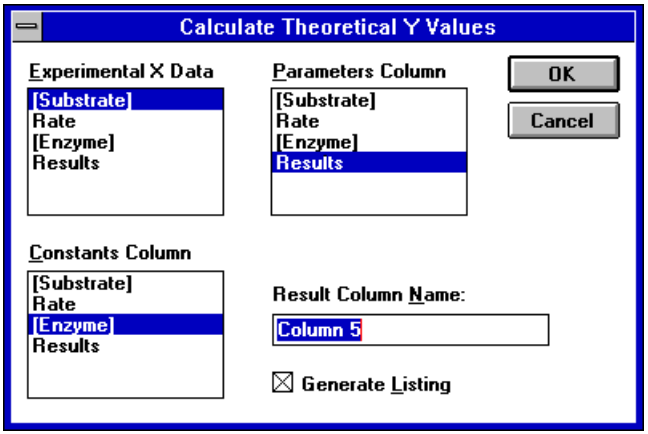
One common use for data fitting is to generate a standard curve from which the values for a series of "unknown" observations are determined. The unknown values may be either *x* or *y* data values; GraFit allows values to be read off the curve from each axis, and in addition will generate a Results listing for the analysis.

The commands used for extrapolating such values assume that the data have been fitted by non-linear regression, and that the appropriate equation has been selected using the **Options Definition File...** and/or **Options Equation...** command (see Chapter 22). However, it is also possible to read values from a linear or polynomial fit, providing appropriate equations are selected. The file POLYNOM.GFE contains linear and polynomial equations that are suitable for this purpose.

Calculating Y values



To read off the *y* axis values corresponding to one or more *x* axis values stored in a data column, select the **Calculate Y...** command. The following dialog box is produced:



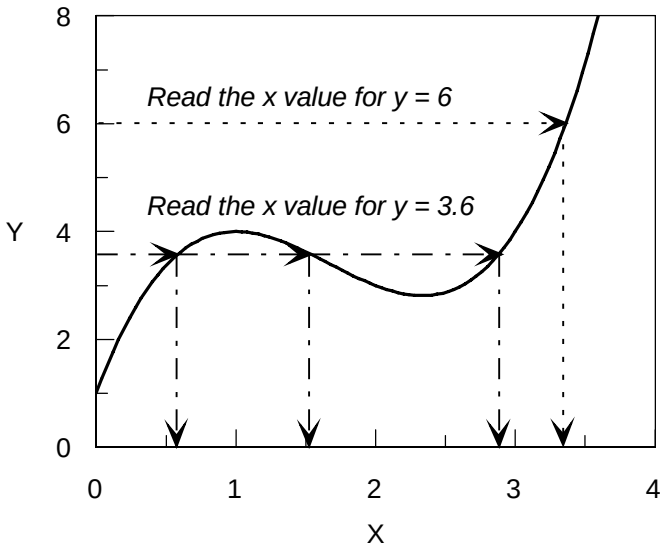
Select the column containing the *x* data values that you wish to read off the curve, the column that holds the calculated parameters describing the equation, and the column that holds the constants (if the equation in use has constants). The currently selected equation is used; use the **Options Equation...** command to alter this if necessary. The calculated values will be placed in a new column, which you may name from the dialog box.

The dialog box also contains a check box that if selected will result in the generation of a Results listing. These are placed in a new Results window, and may be printed, saved, or copied to another application as usual. An example of such a Results listing is shown below:

```
File     : ENZ_KIN.GFD
Directory: C:\
Wednesday 8/07/92  20:06

Enzyme Kinetics

Experimental data in column "Rate"
Calculated values in column "Calculated X"
Parameters from column      "Results"
Constants from column       "[Enzyme]"

Parameters:
  k cat                     =       8.7459
  Km                        =       2.1503

Constants:
  [Enzyme]                  =       1.0000


_____________________________

Experiment   Calculated
_____________________________

      1.0000       0.2776
      1.5000       0.4452
      2.8000       1.0126
      3.7000       1.5768
      4.1000       1.8977
      4.8000       2.6158
      5.0000       2.8703
      5.5000       3.6437
      5.6000       3.8278
      6.0000       4.6987
...(etc.)...
```

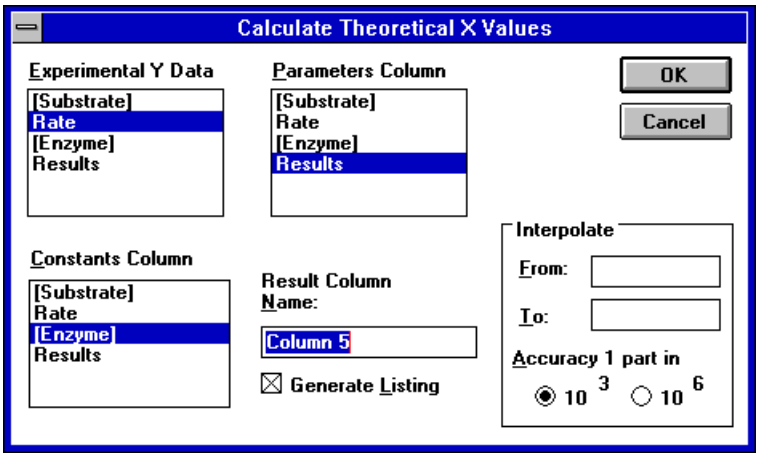
Calculating X Values

X values may also be read off a curve from a series of *y* values stored in a data column. The process is similar to that for calculating *y* values, but has an additional complication. Curves analyzed by GraFit must always be single valued in *y*, i.e. for every *x* value there is only one *y* value possible. *This is not necessarily the case for y values*. For example, consider the curve drawn below.



At $y = 6$ the value of x read from the curve is 3.4; but at $y = 3.5$ there are three x values, 0.55, 1.52 and 2.9. Clearly, whether the curve is single or multiple-valued in x depends upon the equation in use and the value of y that is used.

To overcome this complication it is necessary to specify limits to the x values between which the data are to be calculated. Selecting the **Calculate X...** command gives the dialog box below:



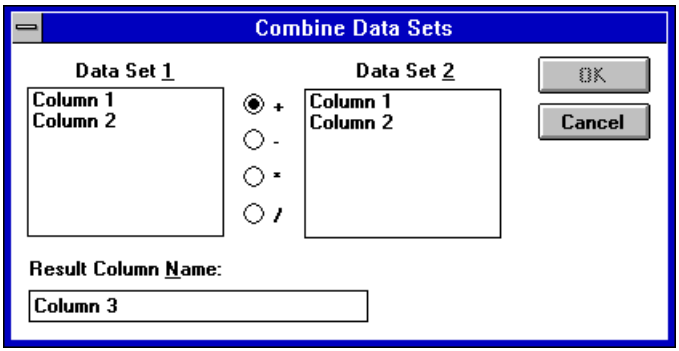
The dialog box is largely the same as that used to calculate y values. Enter into the Interpolate section the range of values (for x) between which the calculation is required. The calculation takes longer than that used for the y values, and is achieved by interpolation between the limits given. For example, to find the solution $x = 1.52$ at $y = 3.5$ in the previous graph the interpolation was performed between $x = 1$ and $x = 2$. Because interpolation is used

it is necessary to specify a required accuracy for the calculation; accuracy to 1 part in 10^6 takes twice the time of accuracy to 1 part in 10^3.

If the curve is not multi-valued in *x*, it is sufficient to use an interpolation range that covers the maximum possible *x* value range.

Combining Data Sets

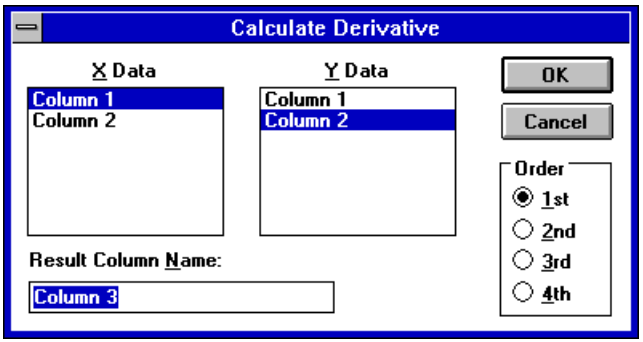
The **Combine...** command allows the data in any pair of columns to be combined using simple mathematical operators (addition, subtraction, multiplication and division), and the result stored in a new data column. The **Combine...** command gives the following dialog box.



Select one data column from each list box, the mathematical operation required (addition (+), subtraction (-), multiplication (*) or division (/)), and enter the name for the new data column. In combination with the **Rescale...** command, the **Combine...** command allows the production of difference plots, signal averaging etc.

Calculating Derivatives

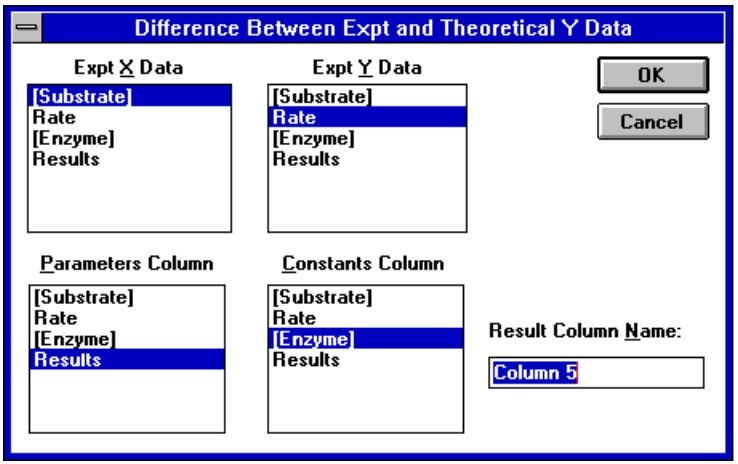
GraFit allows the calculation of the first to fourth derivatives of a column of data. The derivative calculations assume that the data points are in order (use the **Sort...** command if necessary). Unlike some derivative calculations, it is not necessary for the *y* data values to be equally spaced. To calculate derivatives, select the **Derivative...** command, which will produce the following dialog box.



Select the columns holding the *y* data values (of which the derivative is calculated) and the *x* data values, the derivative order, and enter a name for the new column that will hold the results.

Difference Between Experiment and Calculation

The difference between the experimental the calculated values is useful to indicate whether the calculation accurately reflects the variation in the data. To calculate the difference between experimental *y* values and the values described by a fitted equation, choose the **Difference...** command. The following dialog box is displayed.



The equation used is that currently selected; use the **Options Equation...** command to alter this if necessary. Select the columns that hold the *x* and *y* data values, the parameter values and the constants values (if the equation uses constants). The calculated differences are placed in a new column, which can be named from this dialog box.

Assessing the Relative Merits of Two Equations

The **Manipulate F Test...** command is used to test whether one equation fits a set of data significantly better than does a second equation. Before using this command it is necessary to fit the same data set by non-linear regression using two different equations (which must be present in the same definition file). List the results and note the reduced χ^2 (chi-squared) values from the data fitting. By comparing these values it is possible to calculate the probability that the fits are the same. A low probability value indicates that one of the two equations (that giving the lower reduced χ^2) fits the data significantly better than the other. To be significant, the probability should be lower than 0.1, and preferably lower than 0.05.

Example

The following data were obtained, and were fitted to a single exponential decay equation, and to an equation for a single exponential decay including a background offset (these equations can be found in the definition file EXP.GFE).

Single exponential decay:

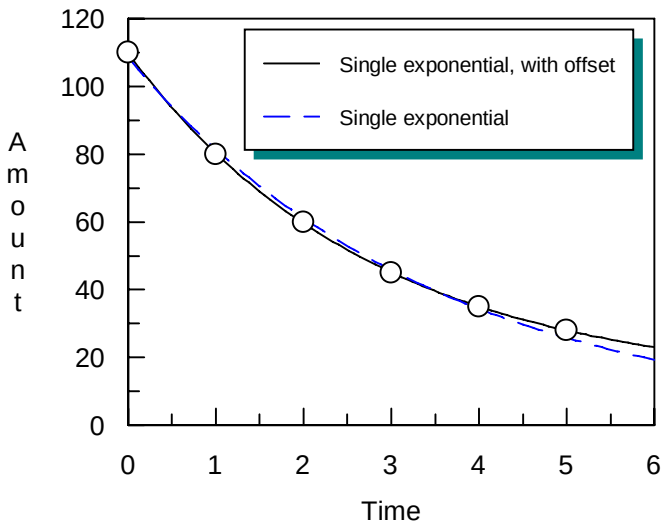
$$y = Ae^{-kt}$$

Single exponential decay including offset:

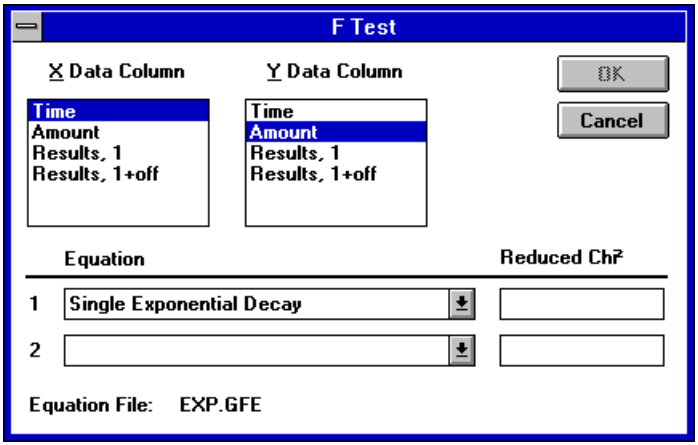
$$y = Ae^{-kt} + \text{offset}$$

Time	Amount
0	110
1	80
2	60
3	45
4	35
5	28

When fitted to a single exponential decay, the reduced χ^2 is 2.92; fitted to a single exponential decay including a background offset the reduced χ^2 is 0.09189. The second value is lower, but the difference between the two equations is not great, as is shown below.



Is this difference significant? To check this we perform an *F*-test by selecting the **Manipulate F Test...** command. This gives the following dialog box.



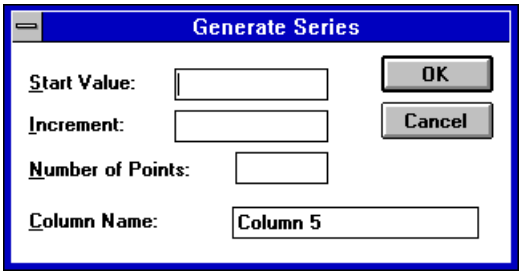
After selecting the appropriate equations, the data columns that hold the data, and entering the reduced χ^2 values, pressing **OK** will create a new Results window that holds the results of the *F*-test analysis.

```
File      : TESTEXP.GFD
Directory: C:\
Wednesday 8/07/92  18:27

Equation 1: Single Exponential Decay
Equation 2: Single Exponential Decay + Offset

Number of data points     = 6
Reduced Chi² value  (1)   = 2.82
Number of parameters (1)  = 2
Degrees of freedom   (1)  = 3
Reduced Chi² value  (2)   = 0.09189
Number of parameters (2)  = 3
Degrees of freedom   (2)  = 2
F statistic               = 30.6889

Probability               = 0.0634443
```

In this instance, the probability that the two fits are equally appropriate is 0.06, which is low. We can therefore be reasonably confident that it is more valid to fit the data using the equation that includes a background offset (providing there is a theoretical or experimental justification for using this equation).

Generating Time-Series Data

The **Generate Series...** command provides a rapid way to produce a data column containing a series of equally spaced values, for example, time-series values.



The data values generated start at the Start Value, and increase in size by the Increment. The total number of values that are produced is given by the Number of Points entry. These data values are added to the data table as a new column.

Making a 3D Data Set

The **Manipulate Make 3D Data Set...** command allows the generation of an x_1, x_2, y set of data columns from a series of x, y data column pairs.

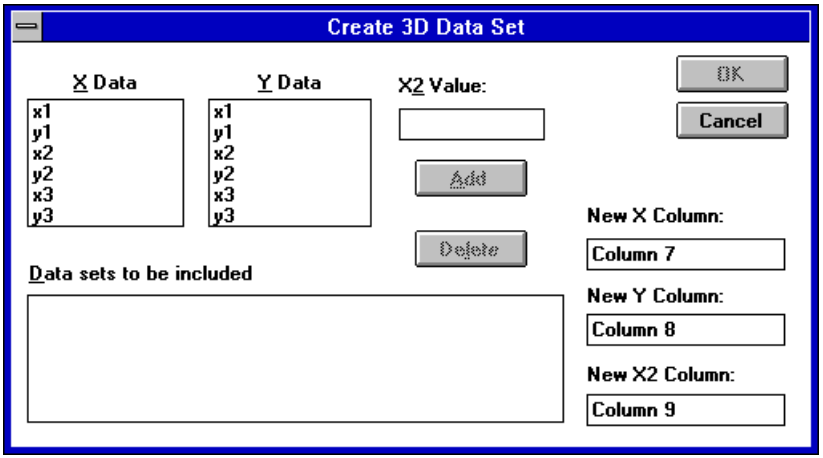
Add	1	2	3	4	5	6
	x1	y1	x2	y2	x3	y3
1	1	2	11	12	21	22
2	2	3	12	13	22	23
3	3	4	13	14	23	24
4						

Make 3D Data Set by combining the three separate x/y data pairs

Combined x *Combined y* *x2 values*

Add	1	2	3	4	5	6	7	8	9
	x1	y1	x2	y2	x3	y3	New X	New Y	X2
1	1	2	11	12	21	22	1.00000	2.00000	100.00000
2	2	3	12	13	22	23	2.00000	3.00000	100.00000
3	3	4	13	14	23	24	3.00000	4.00000	100.00000
4							11.00000	12.00000	200.00000
5							12.00000	13.00000	200.00000
6							13.00000	14.00000	200.00000
7							21.00000	22.00000	300.00000
8							22.00000	23.00000	300.00000
9							23.00000	24.00000	300.00000
10									

For data entry and for plotting results it is convenient to have such 3D data as a series of x, y pairs; however to fit such data to an equation that has two independent variables it is necessary to have the data combined into three columns: x_1, x_2, y. This command provides a convenient alternative to generating these columns using cut and paste.

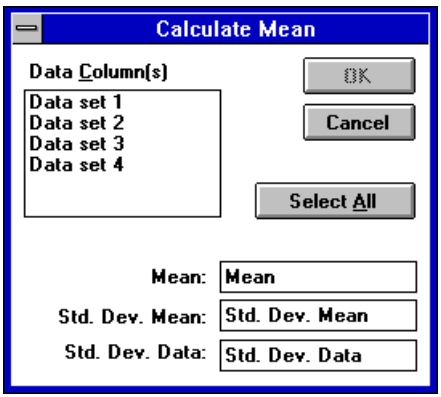


The x_1 and y data are assumed to be already present in the data table. A pair of such data columns are chosen from the list, and the x_2 value that corresponds to this is entered into the X2 Value box. The **Add** button is used to add these selections to the list of data sets to be included; repeat with as many data sets as are required. Entries can be deleted from this list by selecting an entry and using the **Delete** button. When the **OK** button is pressed, all entries in the list are merged to generate three new data columns that represent the combined 3D data set.

Calculating Mean and Standard Deviations

If several replicate data sets have been collected, it may be required to calculate the mean of the various values. In addition, the standard deviation of the data is a useful indication of the accuracy of the experiment. This value may also be used as a weighting factor during subsequent regression analysis.

The **Mean...** command calculates the mean, the standard error of the mean and the standard error of the data point for a series of columns of data. This command results in the following dialog box.



Select the columns that contain the data for which the mean values are to be calculated. The results will be placed into three new columns that will hold the calculated mean, standard deviation of the mean, and standard deviation of the data. You may enter names for these columns in the appropriate edit boxes. When you are ready press **OK** to calculate the results.

Example

Below is a sample data set containing four replicate measurements. Each of the replicates is placed in a separate column, i.e. for the first data point we have four measurements: 0.75063, 0.92515, 1.10525, 0.82366; the values for the second data point are 2.03179, 2.19798, 2.00677, 1.83295 etc.

Data set 1	Data set 2	Data set 3	Data set 4
0.75063	0.92515	1.10525	0.82366
2.03179	2.19798	2.00677	1.83295
2.84665	3.16142	2.90200	3.24426
4.15437	4.12330	3.75749	3.97285
5.04250	4.83705	4.79570	4.80954
5.98994	6.17947	5.93223	5.75233

After calculation, the data table would become:

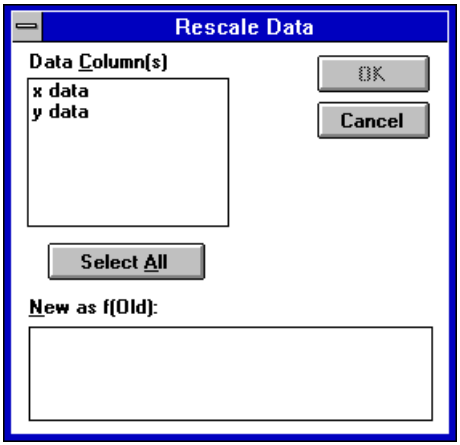
Data set 1	Data set 2	Data set 3	Data set 4	Mean	Std. Dev. Mean	Std. Dev. Data
0.75063	0.92515	1.10525	0.82366	0.90117	0.07686	0.15373
2.03179	2.19798	2.00677	1.83295	2.01737	0.07469	0.14939
2.84665	3.16142	2.90200	3.24426	3.03858	0.09699	0.19398
4.15437	4.12330	3.75749	3.97285	4.00200	0.09063	0.18126
5.04250	4.83705	4.79570	4.80954	4.87120	0.05774	0.11549
5.98994	6.17947	5.93223	5.75233	5.96349	0.08799	0.17598

The mean value for the first set of data points (0.75063, 0.92515, 1.10525, 0.82366) is 0.90117; the standard deviation of the mean is 0.07686, and the standard deviation of these data points is 0.15373.

Note The calculations use the current weighting, selected by the **Options Weighting** command. The values calculated in the example above used Simple weighting, i.e. it was assumed that the errors in the data points were of equal magnitude. It is also possible to select Statistical or Proportional weighting, which would result in slightly different calculated values. See Chapter 29 for a description of how the calculations are performed.

Rescaling Data

The **Rescale...** command allows the data values within a column to be altered. At the simplest level this command permits such actions as background subtraction or concentration correction, although it also may be used to allow more sophisticated recalculation of data. The command produces the following dialog box.



Select the data columns that you wish to rescale (the **Select All** button can be used as a short cut to select all the data columns in the list box). To select more than one column, hold down the CTRL key when making the selections. Rescaling of the data in the selected columns is performed as specified in the 'New as f(Old):' edit box. This edit box is used to enter an expression describing how the existing data is to be modified. The syntax for this expression is the same as for a single-line equation definition, as described in Chapter 23. All the functions available for use in equations may be used. One additional symbol that is defined for use in this command is `Old`. This symbol is used to access the current value in the data table.

Examples

The following examples show how rescaling may be specified and used.

Background subtraction

```
Old - 0.123
```

Subtracts 0.123 from the value in the data table.

Concentration correction

```
Old * 23.98
```

Multiplies the value in the data table by 23.98.

More complex rescaling

```
Log(Old)
```

Calculates the logarithm of the value in the data table.

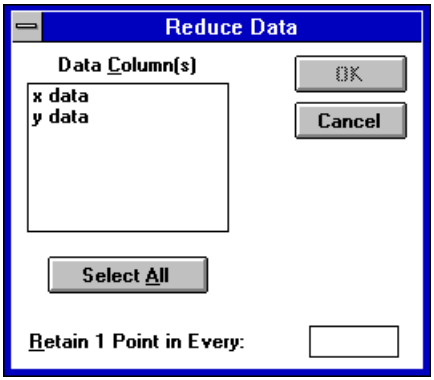
```
(old*1.98)/(old-7.876)
```

Performs more complex rescaling.

Note The rescaled values are written back into the data table, overwriting the old values. If you wish to retain the original values, duplicate the relevant columns using the Clipboard as described in Chapter 2, and rescale the copied columns.

Reducing the Number of Data Points in a Column

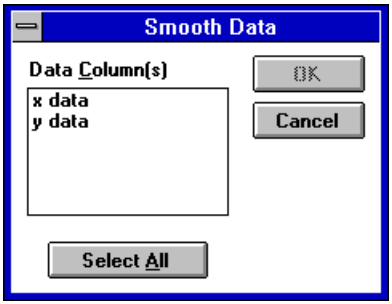
Data that are imported from other applications, or obtained from an instrumental source such as a spectrometer, often have more data values than are necessary to define the experimental curve. For example, the data obtained from a spectrometer may contain more than a thousand points; fitting so many points would be extremely time consuming, and would not necessarily give a "better" fitted value than if (say) thirty points spread over the data range were analyzed. The **Reduce...** command provides a convenient means of reducing the number of data values in a column. This command results in the following dialog box.



Select one or more columns to reduce, and enter a number in the edit box. For example, entering the number 10 in the edit box will retain one data point in every 10, i.e. if 200 data values were present in the selected column(s) data values 1, 11, 21, etc. would remain leaving 20 data points after reduction.

Smoothing Data

Smoothing of data is often desirable if the raw data values contain excessive noise. Beware, however, of the indiscriminate use of smoothing to hide poor quality experimental techniques. To smooth the data contained in one or more data columns select the **Smooth...** command, which will give the following dialog box.



Smoothing may be applied to several columns by selecting more than one column from the list box. To choose more than one column, hold down the CTRL key when making your selections.

GraFit uses a binomial smoothing technique. The smoothed data value, x_i is defined as:

$$x_i = \frac{x_{i-1} + 2x_i + x_{i+1}}{4}$$

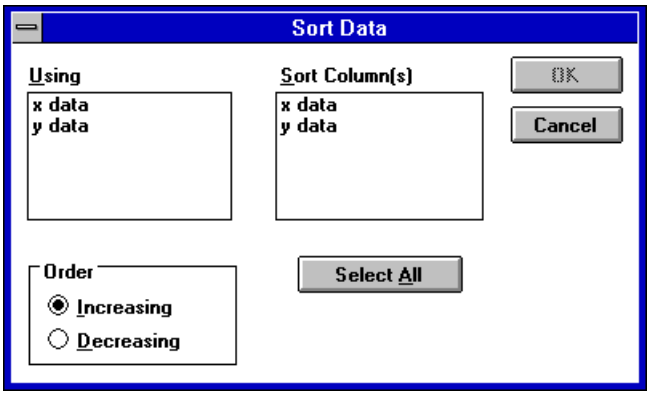
where x_{i-1} and x_{i+1} are the previous and subsequent values in the data column.

This technique requires that the data points in the column to be smoothed are first sorted; use the **Sort...** command first if necessary. If greater smoothing is required the **Smooth...** command may be repeated as often as necessary.

Note After smoothing, the results are written back into the selected data column(s), overwriting the existing data. If you wish to retain the original data, duplicate the relevant columns using the Clipboard as described in Chapter 2.

Sorting Data

The **Sort...** command sorts one or more of the data columns into ascending or descending order. This command results in the following dialog box.



Groups of related columns are sorted based on the values in one of the data columns. For example, a pair of data columns containing *x* and *y* data values may be sorted based upon the *x* data values. To do this the *x* data column would be selected from the "Using" list box, and both the *x* and *y* data columns would be selected from the "Sort Column(s)" list box. Data can be sorted into increasing or decreasing order as required.

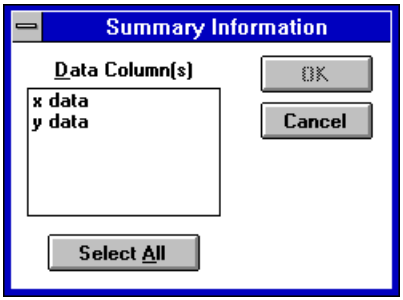
Note The sorted data values are written back into the data table, overwriting the raw data; if you wish to store the unsorted data you may copy the appropriate columns first as described in Chapter 2.

Selecting More Than One Column

More than one column may be selected for sorting by holding down the CTRL key when making the selections. As a short cut it is possible to select all the column names in the Sort Column(s) list box by pressing **Select All**.

Summary Statistics

The **Manipulate Summary...** command produces a Results window listing summary statistics for one or more columns of data.



Select the column(s) for which the statistics are required, and press **OK**. To choose more than one column, hold down the CTRL key when making your selections.

Example

Sample output from this command is shown below

```
File     : (untitled)
Thursday 9/07/92  20:13

Summary information for column "x data"
Number of data points        = 9
Minimum data value           = 1
Maximum data value           = 9
Data value range             = 8
Column total                 = 45
Mean data value              = 5
Median data value            = 5
Standard deviation (n)       = 6.6666667
Standard deviation (n-1)     = 7.5
Variance (n)                 = 2.5819889
Variance (n-1)               = 2.7386128

Summary information for column "y data"
Number of data points        = 9
Minimum data value           = 0.88753319
Maximum data value           = 8.9453902
Data value range             = 8.057857
Column total                 = 44.597205
Mean data value              = 4.9552449
Median data value            = 5.015247
Standard deviation (n)       = 6.8461091
Standard deviation (n-1)     = 7.7018728
Variance (n)                 = 2.6165071
Variance (n-1)               = 2.7752248
```

26 Transforming Data

For display purposes it is often useful to rearrange data, usually to produce a linearized plot. This process of rearranging the data we will call transformation.

To display a transformed plot requires the following steps.

1. Rearrange the data in the data table as required to give new *x* and *y* data columns that hold the transformed data. This is most easily done using the **Manipulate Transform...** command.
2. Create a graph to display these transformed data.
3. If required, add a curve or line to show how well the data have been fitted. The parameters that describe this should have been calculated from the original data, probably by non-linear regression.

Transformations

As an example of the use of transformations to linearize data we will consider the case of enzyme kinetic data. The rate, *v*, of an enzyme-catalyzed reaction varies with the concentration of substrate, [S], using the equation:

$$v = \frac{k_{cat}[E_0][S]}{K_m + [S]}$$

where k_{cat} (the catalytic rate constant) and K_m (the Michaelis constant) are the unknown parameters. The concentration of enzyme, $[E_0]$, is fixed for any given experiment, and so is defined as a constant in the equation.

Such data are often rearranged for presentation purposes to give a linear plot. Several plots are used, a common rearrangement being the Lineweaver Burk plot, where the rearranged equation is:

$$\frac{1}{v} = \frac{1}{[S]} \cdot \frac{K_m}{k_{cat}[E_0]} + \frac{1}{k_{cat}[E_0]}$$

A plot of $1/v$ versus $1/[S]$ gives a straight line. It must be stressed that such a plot will cause distortions to the errors involved (see Chapter 21), but is a useful way to visualize the data.

The data file ENZ_KIN.GFD holds a sample data file that can be transformed in this manner.

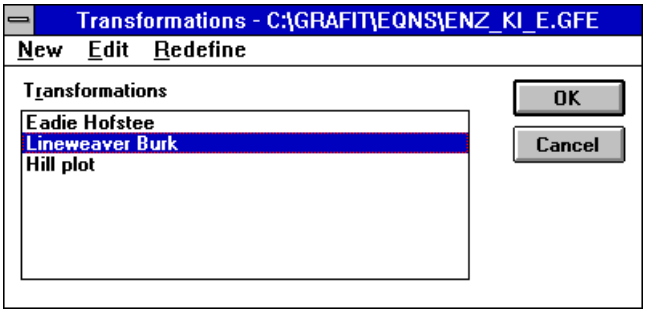
Selecting the Definition File

Transformations are stored in the same definition files as equations. To select a definition file it is necessary to use the **Options Definition File...** command, as was described in

Chapter 22. (The definition file holding an equation for fitting enzyme kinetics data is ENZ_KI_E.GFE.)

Selecting the Transformation

To choose a pre-defined transformation select the **Options Transformation...** command. The following dialog box will be displayed.

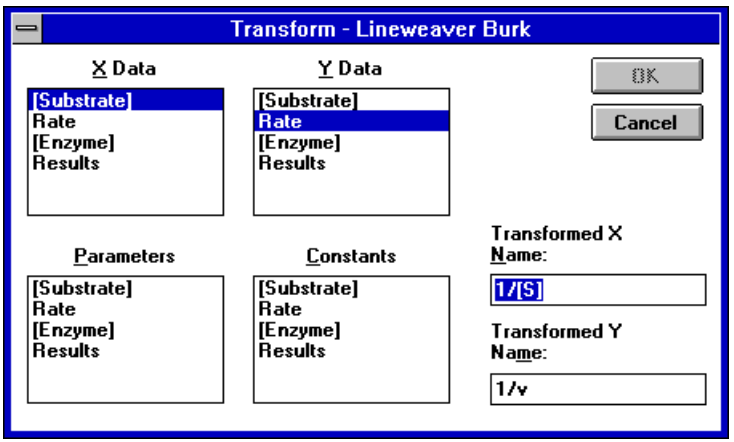


This dialog box functions similarly to that used for selecting and editing equations, which was described in Chapter 23. Select the required transformation from the list box, and press **OK**.

Transforming the Data

Data are transformed using the **Manipulate Transform...** command. The currently selected transformation is used; if no transformation has been selected, an appropriate message box will inform you. Transformation definitions can access the values of fitted parameters and constants, and so it is assumed that any data that is to be transformed will have already have been fitted by non-linear regression.

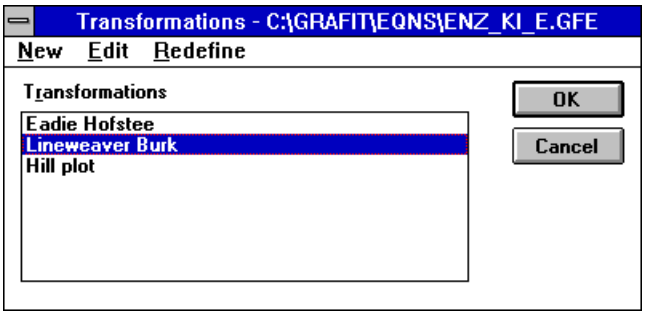
The **Manipulate Transform...** command results in the following dialog box.



It is necessary to indicate which columns hold the *x* and *y* data to be transformed, and to specify which columns hold the calculated parameter values. If any constants are associated with the equation being transformed (as is true here) it is also necessary to show which column holds the constants values. (The values of parameters and constants are vital to some transformations but irrelevant to others, this depends upon the transformation definition in use. If these values are not required it is possible to select any column to represent these data.) After supplying this information press **OK** to transform the data. The results of the transformation will be placed in two new columns that will be appended to the data set. As usual, it is possible to give meaningful names to these new columns from the dialog box.

Defining Transformations

Transformation definitions are stored, with equation definitions, in the definition file (.GFE file). As many transformations as required can be stored in this file. To add or edit transformations select the **Options Transformation...** command (only available when the data table is uppermost). The following dialog box is produced.



Commands on the menu bar of this dialog box allow the transformations to be altered or defined.

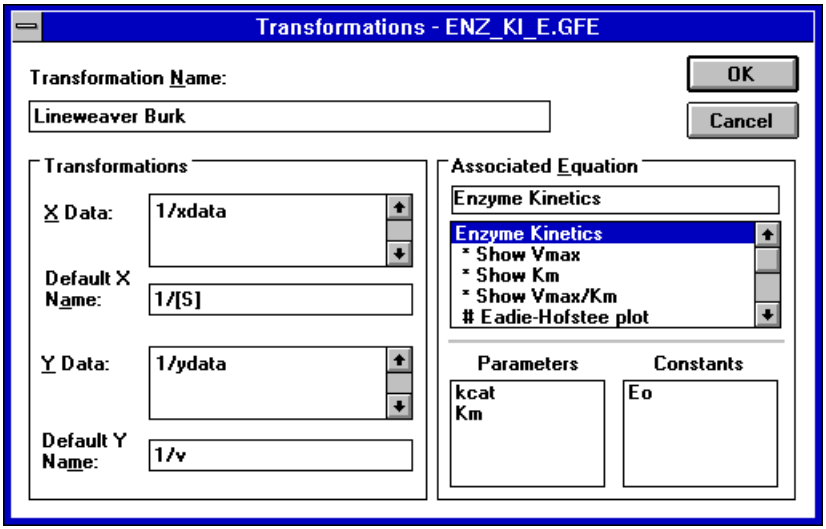
The Edit Command

The **Edit** command allows transformation definitions to be copied, cut and pasted to and from the Clipboard, and to be duplicated.

Edit Menu

The Redefine Command

Redefine allows the currently selected transformation to be edited. Choosing **Redefine** when the Lineweaver Burk transformation is selected gives the following dialog box:



The dialog box has several regions:

Associated Equation

The list shows the equations currently defined in the present definition file. All transformations are associated with an equation, and you must select the appropriate equation from this list. The symbolic names of the parameters and constants (if any) used in the selected equation are also listed. These are present to remind you of their names, as these symbols can be included in the transformation definitions.

Transformation Name

The transformation name identifies the transformation.

Transformations

Two transformations need to be entered, one each for the transformed *x* and *y* data. Use the edit boxes to enter the transformation definitions. The rules for these transformation definitions are the same as described in Chapter 23 for single-line equation definitions. The definition can be of arbitrary complexity, and can include mathematical functions, parameter names etc. The pre-defined symbols `xdata` and `ydata` represent the original *x* and *y* data values respectively. In this example we wish to take reciprocals of these data; the definitions are therefore `1/xdata` and `1/ydata`. Any algebraic expression may be entered, for example

```
xdata/(ydata + sqrt(xdata))
```

is a perfectly valid transformation definition.

When the transformation definitions have been entered, press **OK**. The definition is then checked for any mistakes in syntax, and if correct is compiled.

The New Command

The **New** command creates an empty transformation definition, and then allows you to define this new transformation.

Plotting the Transformed Data

The transformed data are treated in the same manner as any other data columns, and can be used as the *x* and *y* data for plotting a graph. Consult the section on drawing graphs (Chapter **Error! AutoText entry not defined.**) for full details of how to plot data.

Drawing the Transformed Curve

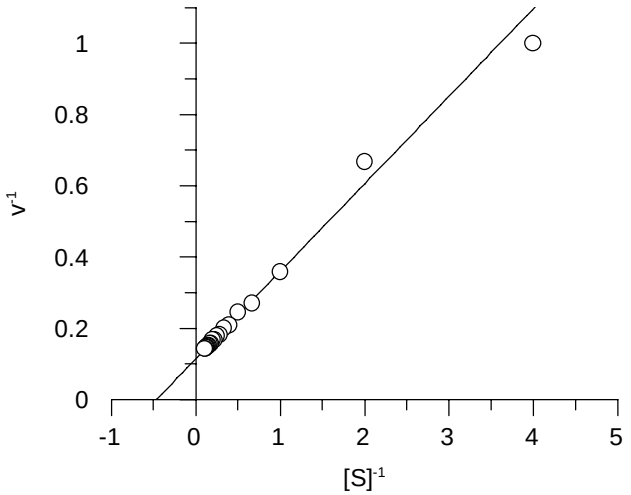
To draw the transformed curve (which in the example above gives the linear Lineweaver Burk plot) it is necessary to have an appropriate equation definition. This probably will have been defined as a "Display Only" equation to distinguish it from normal equations (see Chapter 23). For the example that we are using, the transformed equation is

$$\frac{1}{v} = \frac{1}{[\mathrm{S}]} \cdot \frac{K_\mathrm{m}}{k_\mathrm{cat}[\mathrm{E}_0]} + \frac{1}{k_\mathrm{cat}[\mathrm{E}_0]}$$

What is being plotted is $1/v$ (the *y* axis) versus $1/[\mathrm{S}]$ (the *x* axis). To show this line, we therefore need an equation that is defined as

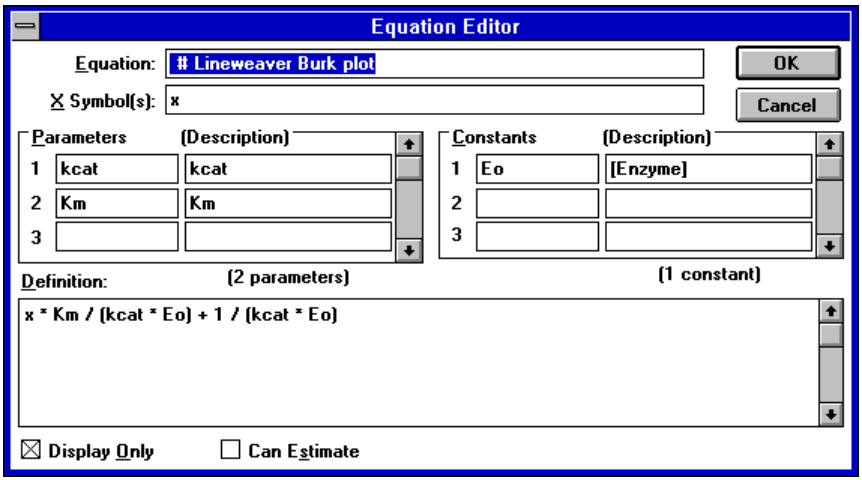
$$y' = x' \cdot \frac{K_\mathrm{m}}{k_\mathrm{cat}[\mathrm{E}_0]} + \frac{1}{k_\mathrm{cat}[\mathrm{E}_0]}$$

where *x'* and *y'* are the transformed data. To draw this line, choose the **Add Curve...** command, and choose this "equation" (in the file ENZ_KI_E.GFE, the appropriate curve to draw a Lineweaver Burk plot is named " # Lineweaver Burk plot"). The result for a Lineweaver Burk plot is shown below.



Transformed Equation Definitions

The equation definitions used to draw transformed curves are defined in the same manner as normal equations. Editing the equation used to draw the Lineweaver Burk line (choose the **Options Equation...** command, select the "# Lineweaver Burk plot" equation and press **Redefine**) shows the following dialog box.



The "equation" defined is for the transformed line, as derived earlier. The parameters and constants must be entered in the same way, and in the same order, as for the parent equation; the easiest way to do this is to use the **Edit Duplicate** command on the equation dialog box. To distinguish this from a normal equation, the Display Only box has been checked. Any transformed line or curve can be defined in an analogous manner. We suggest that you edit some of the pre-defined definitions included with GraFit for further examples.

27 Multiple Regression

What is Multiple Regression?

The most common form of non-linear equation that is analyzed by GraFit is of the form

$$y = f(x)$$

where the right hand side of this expression may contain up to 30 parameters. Such an equation has one dependent (y) and one independent variable (x). In addition, GraFit can be used to analyze equations with more than one independent variable, i.e.

$$y = f(x_1, x_2, x_3...)$$

In this case the analysis is termed multiple non-linear regression. GraFit allows a maximum of 10 independent (x) variables, along with up to 30 unknown parameters. Note, however, in all cases it is assumed that only the y axis data contain errors.

Use for Multiple Regression

Multiple regression is essential when the data to be examined are inherently multidimensional. Examples include peak-shape analysis of 2D spectra. However, there are several situations where multidimensional data are collected, but are traditionally analyzed by a performing a series of normal two-dimensional (x, y) curve fitting.

Example

The progress of an enzyme-catalyzed reaction at a fixed concentration of substrate is given by the equation:

$$A_t = A_\infty(1 - e^{-kt})$$

where A_t is the extent of reaction at time t, A_∞ the maximum extent of reaction, and k the rate constant at this concentration of substrate. The value of k from this equation is related to the kinetic parameters by the equation:

$$k = \frac{V_{max}[S]}{K_m + [S]}$$

where V_{max} and K_m are the parameters that we need to obtain from data fitting. The usual way of analyzing these data is to collect a series of progress curves (A_∞ as a function of t) at varying concentrations of substrate. For each concentration of substrate the data are fitted to the first equation, and a value for k obtained. The resulting k values are then fitted to the second equation to find V_{max} and K_m. A better solution is to recast the equation as a three-dimensional one, where the progress of the reaction varies with t and [S], which are the two independent variables. The equation is obtained by combining the previous ones:

$$A_t = A_\infty\left(1-\exp\left\{-\frac{V_{max}[S]}{K_m+[S]}\cdot t\right\}\right)$$

This allows V_{max} and K_m to be determined directly from the raw data, which has two *x* variables (*t* and [S]).

Advantages of Multiple Regression

Use of multiple regression simplifies the analysis, as the curve fitting need only be carried out once. In the above example, the value of A_∞ is fixed for the entire experiment and only found once for the multidimensional equation. However, if the data are analyzed by several applications of the first equation, the A_∞ value would be determined several times.

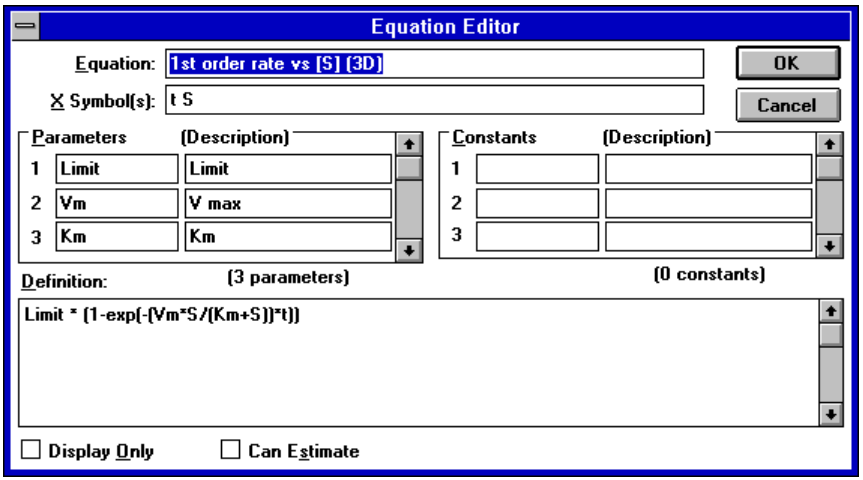
Disadvantages of Multiple Regression

The disadvantages of multiple regression are as follows.

- Conceptual. It is often difficult to visualize the problem in more than two dimensions.
- Finding initial estimates. This is sometimes difficult for the same reasons.
- Difficulty of display. The current version of GraFit does not support three-dimensional plotting; the display of four-dimensional data is even more problematic!

Defining Equations

Definition of equations having multiple independent variables is achieved in the same way as the definition of normal two-dimensional *x*/*y* equations. For example, the equation described in the example above is found in the RATE.GFE definition file named "1st order rate vs [S] (3D)". The equation editor definition for this equation is shown below.



The only difference between this definition and those that we have seen previously is that more than one X Symbol is present. In this case the symbols used to represent the two independent variables are `t` and `S`, and are entered into the X Symbol(s) edit box separated by a space. As many as ten independent variables can be entered in this way.

Fitting Data by Multiple Non-linear Regression

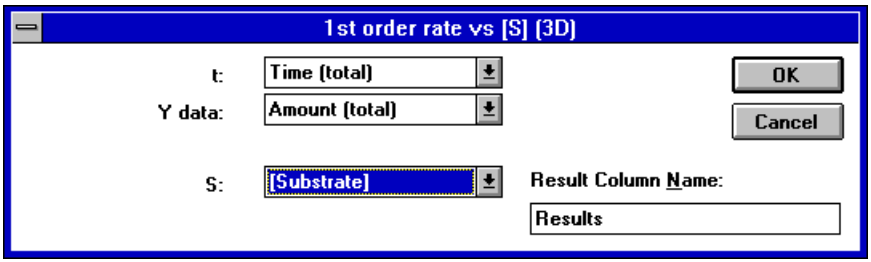
How Data Must be Arranged

For a normal data set that contains *x*/*y* data, these data must be present in the data table with the *x* data and the *y* data in separate columns. When multiple independent variables are involved, the x_1, x_2, etc. data must all be separate columns. For the example above, this means that the data table must hold at least three columns: one that contains all the time values (x_1), another with the absorbance values (y), and a third with the substrate concentrations (x_2).

In many cases, it is likely that the raw data will have been collected as a series of *x*/*y* pairs, in the example above these would be absorbance/time data for various substrate concentrations. If this is the case, before fitting it is necessary to create the required data columns by merging the various raw data sets. This could be done by cutting and pasting to amalgamate the data values, but a shortcut method is to use the **Manipulate Make 3D Data Set...** command (see Chapter 25). It is useful to retain the raw *x*/*y* data pairs for plotting the results on a graph.

Data Fitting

The file MULTRATE.GFD contains a sample data set that can be used with the above equation. After selection of the equation, and reading in the data, they are fitted using the **Data Curvefit...** command as normal. The dialog box used to select the columns for fitting is shown below.

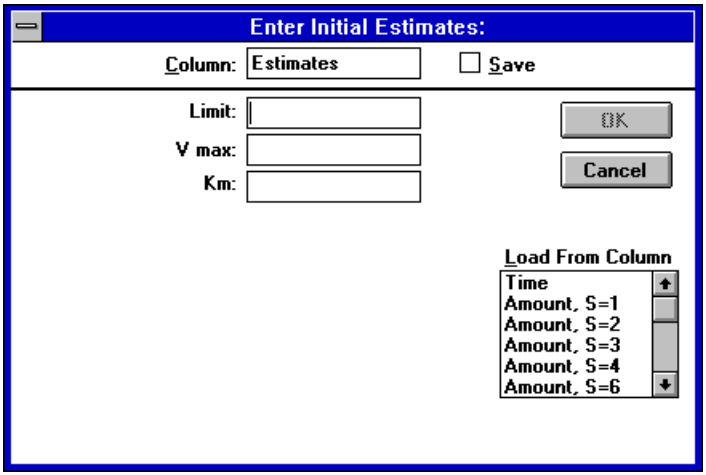


This dialog box differs from that for a normal two-dimensional equation; it is necessary to specify which column holds the x_1 data, which holds the y data and which holds the x_2 data. If explicit weighting was selected it would also be necessary to specify a column for the y standard deviations.

To fit these data, select the appropriate columns (x_1 = Time (total), y = Amount (total), x_2 = [Substrate]) and enter a name for the column to hold the results, then press **OK**.

Initial Estimates

You will next be asked to provide initial estimates for the parameter values (it is not possible for the program to give initial estimates for a multidimensional equation). Enter appropriate values into the following dialog box.



Suitable values for the example data are: A_{∞}, 0.7; V_{max}, 1.5; K_m, 15.

The Results listing generated is shown below.

```
File     : MULTRATE.GFD
Directory: C:\GRAFIT\DATA
Friday 7/08/92  20:15

1st order rate vs [S] (3D)
Simple weighting

Reduced Chi squared = 3.275e-005

_____________________________________________________

Variable                          Value    Std. Err.
_____________________________________________________

Limit                            0.5011       0.0018
V max                            1.8948       0.1175
Km                              18.9052       1.3073
_____________________________________________________




_____________________________________________________

          t            S            Y
     Time (total) [Substrate]  Amount (total)  Calculated
_____________________________________________________

   1       1.0000       1.0000       0.0480       0.0455
   2       2.0000       1.0000       0.0950       0.0869
   3       3.0000       1.0000       0.1210       0.1245
   4       4.0000       1.0000       0.1670       0.1587
   5       5.0000       1.0000       0.1990       0.1898
Etc.
```

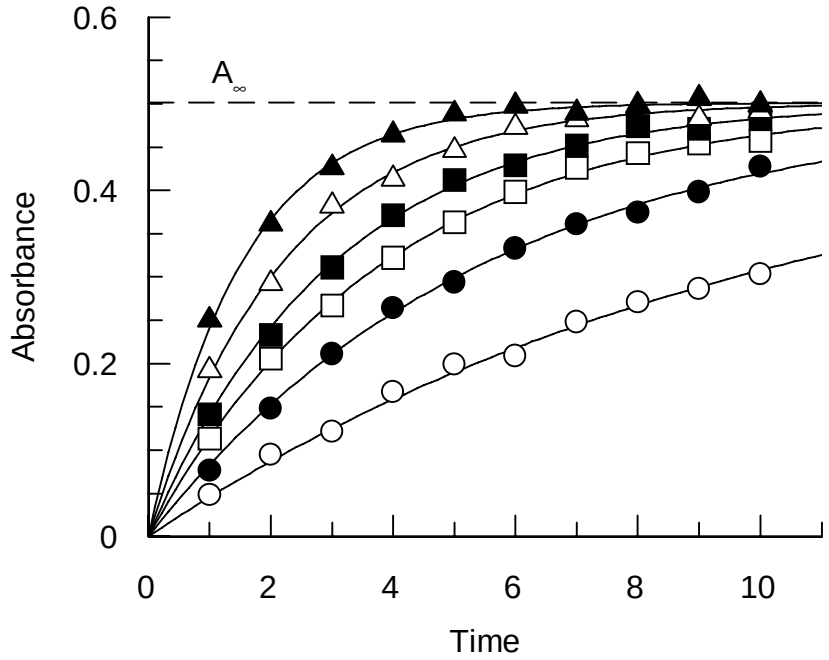
The listing is very similar to that produced for normal equations, but includes the additional *x* data.

Plotting Results from Multiple Regression

The data that form the basis for multiple regression have three or more dimensions. To represent such data would, strictly speaking, require a multidimensional plot which is not available. However, the usual representation for such data is to plot multiple curves,

representing the raw data as collected, on the same graph. The following graph shows the above fitted data; the curves were drawn using the " * Show Individual Curve (3D)" display equation, where the [S] is defined as a constant. (***Hint*** it is often easier to enter the [S] values *after* plotting the curve. Double-click on the curve symbol, then edit the values of the constants to enter the [S] values that you require.)

The A_{∞} value is indicated by using the ' * Show Limit' display equation.



Showing Curves

In the above example, it was necessary to use an equation defined for "Display Only" to show the fitted curves. This is because the equation used to fit the data had more than one independent variable, and such equations cannot be used for graph plotting. The display equation, " * Show Individual Curve (3D)", is defined in such a way that it contains only one independent variable (time). The other independent variable in the analysis, the substrate concentration, is replaced by a constant. In this way it is possible to show the fitted curve at a fixed (constant) substrate concentration, but make use of the fitted results obtained from the full multiple regression analysis.

28 Numeric Formats

The numbers in the main data table and in the Results listing can be displayed with varying precision using several types of number format.

Number Formats

There are three main ways of displaying numbers within GraFit.

Fixed Format

Fixed decimal or "F" format displays numbers using a fixed number of decimal places within a defined field width. For example, 12.700, 1000.00 and 0.001 are all fixed format number representations. Fixed format is useful to display numbers that are neither very small nor very large. The position of the decimal point is the same for a group of numbers so formatted, ensuring that columns of numbers printed in fixed format will line up, e.g.

```
  23.273
9999.999
  -0.123
```

Exponential Format

The second form of representation is exponential format, or "E" format. For example, 1.23e+023, -1.00e-017 and 9.00e+000 are all in exponential format. Exponential format is useful when handling small or large numbers. Columns of numbers printed in exponential format will also line up, e.g.

```
 1.700e+000
-8.218e+012
 1.000e-001
 1.000e-010
```

General Format

The final form of representation available within GraFit is general format, or "G" format. General format is a hybrid between the first two formats; very small or large numbers are shown in exponential format, other numbers lack the exponent although the number of decimal places is not fixed. Examples of numbers in general format are 1, 7.9, 2.3e+023. General format is useful for displaying numbers of any magnitude, but the numbers can look untidy in tabular form as they are usually of different widths.

A comparison of numbers in each format is shown below.

Fixed	Exponential	General
-0.0000	-1.0000e-017	-1e-17
22.4000	2.2400e+001	22.4
7.0000	7.0000e+000	7
-4000.0000	-4.0000e+003	-4000
52100000.0000	5.2100e+007	5.21e+007

Controlling Displayed Precision

It is possible to control the precision of the displayed number for each of these formats. To do so it is necessary to specify a field width and the number of decimal places required. These terms are defined below.

Fixed Format

The number below has a field width of 12, with 3 decimal places.

```
     999.000          Decimal Places = 3
^^^^^                 Field Width = 12
```

The final figure is rounded, if necessary, for example 6.666666 to 2 decimal spaces is 6.67. It should be noted that the number is stored internally to high accuracy (12 significant figures), only the representation shows the rounding. When used in the main GraFit data table, large numbers that would overflow their field width if displayed in fixed format will actually be printed in exponential format.

Exponential Format

The number below has a field width of 12, with 3 decimal places.

```
  9.990e+001          Decimal Places = 3
^^                    Field Width = 12
```

General Format

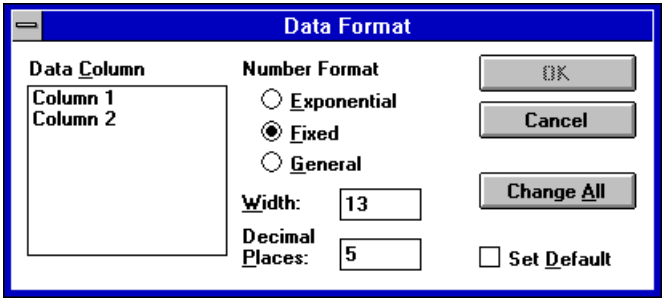
The number below has a field width of 12, with 3 "decimal places" selected.

```
         999          Decimal Places (significant figures) = 3
^^^^^^^^^             Field Width = 12
```

For numbers printed in general format the number specified as the "decimal places" controls the number of significant figures printed. Large or small numbers are printed in exponential format.

Specifying the Numeric Format of the Data Table

By default, the numbers in the data table are in fixed format using a field width of 13 and 5 decimal places. To alter the format used for the display select the **Options** command, and choose **Data Format...**. The following dialog box is then displayed.



The format for each column of data can be set individually. Choose the data column, and specify the required formats, field widths and number of decimal places. After pressing **OK** the data table will be displayed in the new format.

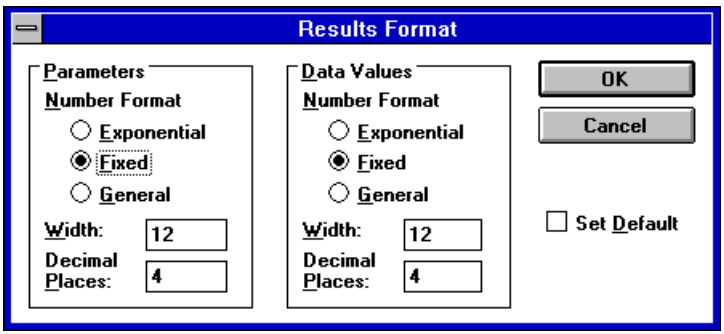
It is also possible to set the display format for an individual column by double-clicking on any of the data within that column. This will produce the above dialog box with the appropriate column selected.

Default Format Settings

When GraFit is first run, the default data format is as shown in the above dialog box. This can be altered by checking the Set Default box. If this box is checked, the format specified will be used as the default when GraFit is next run.

Specifying the Numeric Format of the Results Listing

By default all numbers in the Results listing have fixed format using a field width of 12, with 4 decimal places. To alter this format select the **Options** command, and chose **Output Format...**. The following dialog box is displayed.



The numeric format can be specified separately for the parameter listings and the *x* and *y* data value listings. Choose the required formats, field widths and number of decimal places and

press **OK**. The new numeric formats will be used when the results are next listed (using the **Data List Results...** command, or when the data are next fitted).

Default Format Settings

When GraFit is first run, the default results format is as shown in the above dialog box. This can be altered by checking the Set Default box. If this box is checked, the formats specified will be used as the defaults when GraFit is next run.

29 Theory

To make best use of GraFit, it is helpful to know some of the theory behind regression analysis. Chapter 21 provided a general overview; this section describes more of the theoretical background.

Regression Analysis

Regression analysis is a technique used to calculate the "best fit" line or curve through a data set by minimizing the deviation of the data from the curve. Mathematically, it finds values for various unknown parameters in an equation (for example, the slope and intercept of a straight line) that produce the smallest possible sum of the squares deviations of the experimental values from the calculated ones. This quantity that is minimized is denoted by the symbol χ^2 (chi-squared). The value of χ^2 includes the variance of the data point (the variance of data point i is σ_i), so that data points that have greater errors contribute less to the analysis. It is calculated as:

$$\chi^2 \equiv \sum \left(\frac{\Delta y_i}{\sigma_i} \right)^2$$

Δy_i is the difference between the experimental and calculated y data values (the subscript "i" is used to indicate the "ith" data value).

A further useful statistic is χ^2_ν, or the reduced chi-squared. This is defined as

$$\chi^2_\nu = \frac{\chi^2}{\nu}$$

where ν is the number of degrees of freedom, which is given by $\nu = N - n - 1$; N is the number of data points and n is the number of variables in the equation.

Regression analysis is the mathematical equivalent of finding the best line by eye — it has the advantage, however, of being non-subjective. It is also possible to fit the "best curve" through the data by non-linear regression. This has no manual equivalent (because humans are very poor at judging accurate curvature), but is equally simple mathematically using computational methods.

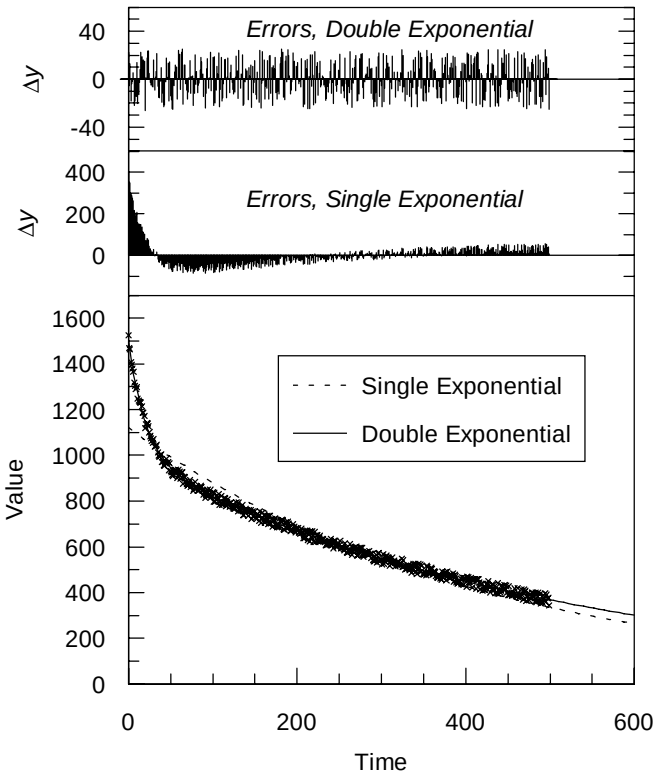
Assumptions in Regression Analysis

Regression analysis is probably the most powerful method to analyze data. However, such analysis involves several assumptions about the function chosen and the error structure of the data. These assumptions are listed below.

The Correct Equation is used

This assumption is often overlooked, but should be borne in mind when using any data analysis technique. It is important to be aware that deviations from ideality are often caused by use of an inappropriate equation. Graphical methods provide a useful means to see whether the chosen model equation is appropriate. This is because use of an inappropriate equation often results in systematic deviations from the curve, which are easy to see in a graph.

Residual plots of Δy versus x are useful to assess whether the equation is appropriate. Systematic deviations in the residuals can mean that the equation used was inappropriate to describe the data. The differences between experimental and calculated values can be generated during data fitting (see Chapter 22), or created using the **Manipulate Difference...** command (see Chapter 25). The data shown below have been fitted to a single exponential decay, and to a double exponential decay, and the residuals from each have been plotted. Clearly, the residuals from the double exponential fit are smaller and more randomly distributed, and show that this fit is superior.



Only the Y Data Contain Errors

Only the dependent variable, which by convention is always the y variable (y_i), is subject to error; the other (x_i) is known exactly. This requirement should be satisfied by suitable

experimental design — the *x* data values often comprise time or concentration values, which should be known with precision.

The Errors are Normally Distributed

The errors in the *y* data values should follow a normal distribution function, i.e. if sufficient replicates were taken they would have a Gaussian spread about the true *y* value. Two common factors that invalidate this assumption are the presence of (*i*) systematic errors, and (*ii*) outliers (data points that are wildly inaccurate, usually due to operator or machine failure). Systematic errors invalidate any data analysis, and should be removed by correct experimental design. Outliers can be eliminated by use of "robust" techniques, as will be described later in this chapter.

The Correct Weighting is Used

An explanation of what "weighting" of data in regression analysis is, why it is incorporated into all such analysis (including the subjective regression performed by drawing the "best line" using a ruler), and how the appropriate weighting can be determined are described later.

Linear Regression

The simplest case for regression analysis is when the equation is linear. For the linear equation:

$$y = a + bx$$

χ^2 is defined as:

$$\chi^2 \equiv \sum \left(\frac{\Delta y_i}{\sigma_i} \right)^2$$

$$\chi^2 = \sum \left(\frac{1}{\sigma_i^2} (y_i - a - bx_i)^2 \right)$$

where σ_i^2 is the standard deviation of the data point *i*.

The best fit to this equation will yield parameter values for *a* and *b* that produce the minimum possible value of χ^2. This is equivalent to saying that the first derivative of χ^2 with respect to *a* and *b* is equal to zero:

$$\frac{\partial}{\partial a}\chi^2 = \frac{\partial}{\partial a}\sum\left(\frac{1}{\sigma_i^2}(y_i - a - bx_i)^2\right)$$

$$= -2\sum\left(\frac{1}{\sigma_i^2}(y_i - a - bx_i)\right) = 0$$

$$\frac{\partial}{\partial b}\chi^2 = \frac{\partial}{\partial b}\sum\left(\frac{1}{\sigma_i^2}(y_i - a - bx_i)^2\right)$$

$$= -2\sum\left(\frac{x_i}{\sigma_i^2}(y_i - a - bx_i)\right) = 0$$

If we rearrange these equations it is possible to solve for *a* and *b* exactly to give values for these parameters that minimize χ^2. This is shown in the equations below.

$$a = \frac{1}{\Delta}\begin{vmatrix}\sum\frac{y_i}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2}\\ \sum\frac{x_i y_i}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2}\end{vmatrix}$$

$$b = \frac{1}{\Delta}\begin{vmatrix}\sum\frac{1}{\sigma_i^2} & \sum\frac{y_i}{\sigma_i^2}\\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i y_i}{\sigma_i^2}\end{vmatrix}$$

$$\Delta = \begin{vmatrix}\sum\frac{1}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2}\\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i y_i}{\sigma_i^2}\end{vmatrix}$$

Weighting

This analysis presupposes that we know the standard deviation, σ_i^2, for each data point. In general, we will not have such information available. However, only relative uncertainties need be known, and if we assume that all the data points have the same inherent errors, the equations simplify to give:

$$a = \frac{1}{\Delta}\begin{vmatrix}\sum y_i & \sum x_i\\ \sum x_i y_i & \sum x_i\end{vmatrix}$$

$$b = \frac{1}{\Delta}\begin{vmatrix}N & \sum y_i\\ \sum x_i & \sum x_i y_i\end{vmatrix}$$

$$\Delta = \begin{vmatrix}\sum y_i & \sum x_i\\ \sum x_i & \sum x_i y_i\end{vmatrix}$$

This assumption that the errors are constant is often made, for example in the simple linear regression calculations performed by pocket calculators. Although often true, this assumption is not always valid. In many experiments the uncertainty varies with the magnitude of the *y* value, such that the percentage error is constant. To account for the different types of error distribution we introduce a weighting factor, which is related to $1/\sigma_i^2$ and is used to multiply all elements of the above matrices.

Weighting Types

The following types of error distribution are easily accounted for, and can be selected for use in the GraFit program from the **Options Weighting** menu.

Constant error

The standard error, σ_i^2, is the same for each data point and therefore a weighting factor of 1 is used ("Simple weighting"). This type of error is often given by instrumental variations (for example, a gauge may be read to an accuracy of 0.1 units, etc.).

Statistical error

For statistical variation, where σ_i^2 varies in proportion to y_i, the weighting factor is $1/y_i$. This type of error distribution is uncommon, but is found in systems where the main errors arise from sampling, for example counting low numbers of radioactive disintegrations.

Proportional error

For many experiments the percentage error is constant. In this case, σ is proportional to *y*, leading to a weighting factor of $1/y_i^2$. A typical example of such errors might be when, for instance, a measuring instrument is accurate to 2% of the delivered volume, and the so error is proportional to the recorded value.

Individually determined error

All the above cases make assumptions about the error of each measurement. However, if sufficient replicates of each data point are available, or a complex error function is applicable, it is possible to calculate σ_i^2 for each point and use these values explicitly. This method can also be used if you know that certain values are less accurate than others (for example if they were measured in a different way) as again only relative σ_i^2 values need be given.

In practice, the use of simple weighting or proportional weighting is most often appropriate. If the errors are small the weighting function chosen has only a small effect on the calculated values. We will see later how it is possible to estimate which type of error function best describes a set of data. If any data values are close to 0, it is not appropriate to use proportional weighting, and simple weighting should be selected.

Non-linear Regression

For an equation that is not a linear function the determination of best-fit parameters becomes slightly more difficult, although the principles discussed above still apply. The difficulty arises because it is not possible to derive an exact analytical expression for minimizing χ^2. It is possible, however, to take initial estimates for the best-fit parameter values and to improve on them by a series of iterative calculations until values are obtained that approach the minimum χ^2 value.

Apart from this mathematical inconvenience, non-linear regression functions in the same way as linear regression. We still need to know the weighting that has to be applied to the data, and the calculations still aim to minimize χ^2. The advantage is that non-linear regression may be used to analyze any function that may be written $y = f(x)$, i.e. the y data value is dependent upon some arbitrary single-valued function of the x data value.

GraFit performs non-linear regression using the method of Marquart[1] using a numerical second order method to calculate partial differentials. In this way it is possible to incorporate equations of arbitrary form. For an excellent explanation of regression analysis the reader is referred to the text by Bevington[2].

Standard Errors of the Calculated Parameters

As well as giving the best-fit values for the parameters, regression analysis can provide estimates of the standard error of each parameter. GraFit calculates these errors by the matrix inversion method. These estimated standard errors are useful to show how accurate the calculated parameters are. As a rough guide, the standard error should not be greater than about 10% of the parameter value if you are to have confidence in the calculated parameters. High standard errors will be obtained if the data are of poor quality, or if the data do not allow accurate estimation of one or more of the parameters.

Robust Weighting

One of the assumptions of regression analysis is that all the data points have a normal error distribution. However, a common feature of actual experimental data is the presence of one or two observations that are well outside the error range of the other observations. The stray data points are usually due to mistakes by the experimentalist (incorrect addition of reagents, inaccurate keying in of data, etc.) and are termed outliers. If these points are included in the analysis they will disproportionately affect the results, since the square of the residuals for these points will dominate the fit. It is possible to edit the data manually to remove such points, but it is preferable to eliminate them in a less subjective manner. Robust methods

---

1 Marquart, D.W. (1963) *J. Soc. Ind. Appl. Math.*, **11**, 431-41.
The Marquardt method is sometimes also referred to as the Marquart-Levenburg method. It is perhaps best explained in reference 2.

2 Bevington, P.R. (1969) *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill Book Company.

achieve this by applying an additional weighting factor varying, for example, from 1.0 for points that have low residuals, to 0 for extreme outliers — effectively eliminating such points.

GraFit employs optional robust weighting using the algorithm of Mosteller & Tukey[3] as implemented by Duggelby[4]. An additional 'bisquare' weight, *b*, is incorporated into the analysis, and is calculated as:

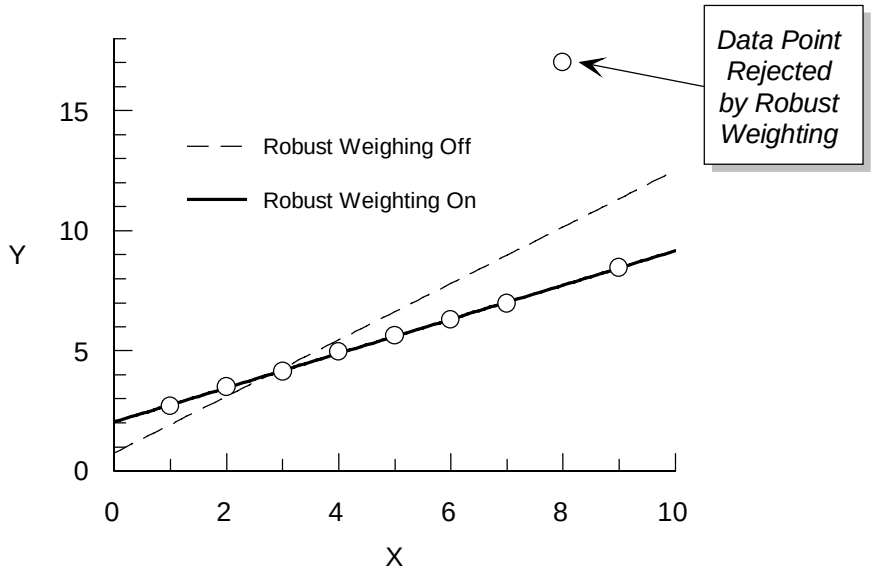
$$b_i = \begin{cases} (1-u_i^2)^2 & \text{if } |u_i| \le 1 \\ 0 & \text{if } |u_i| > 1 \end{cases}$$

where, with *z* being the residual weighted by the *a priori* weight and *N* the number of observations:

$$u_i = \frac{z_i}{6\sum |z_i|/N}$$

Including robust weighting will result in improved estimates for the parameters in the presence of outliers. This is, however, gained at the expense of more calculation, which will slow down the fitting procedure.

The effect of using robust weighting is seen in the graph below:



The data, which contain a single outlier, have been fitted to a straight line with and without robust weighting enabled. When robust weighting is omitted, the presence of this outlier

---

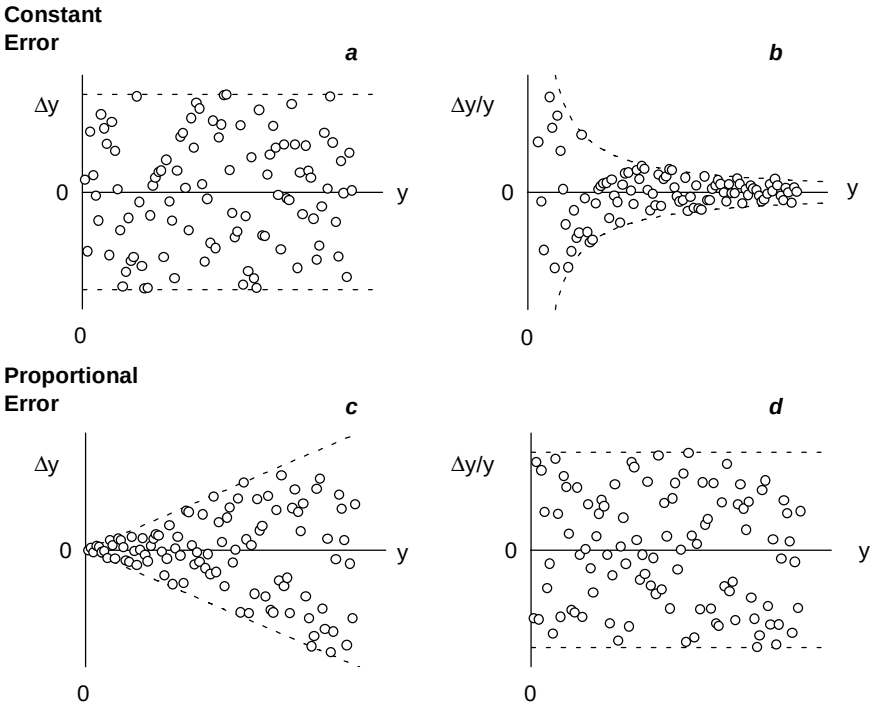
3 Mosteller, F. & Tukey, J.W. (1977) *Data Analysis and Regression.* Addison-Wesley, Reading MA, USA pp.353-65.

4 Duggleby, R.G. (1981) *Analytical Biochemistry* **110,** 9-18.

markedly affects the calculated results. With robust weighting selected, however, the rogue data point is rejected, and the calculated line now correctly follows the main trend.

Determining the Most Appropriate Weighting

For most experimental data the error is commonly either independent of the value of the observation (simple or constant errors) or is proportional to the *y* data value (proportional errors). Simple errors may be caused by instrumental uncertainties (for instance, a machine reading may be only accurate to ± some intrinsic error, independent of the magnitude of the reading). Providing sufficient data points have been sampled, it is possible to determine the correct weighting by examination of the residual error (the experimental minus the calculated value). A plot of the residuals versus the observed value (Δy versus y) is diagnostic of the type of error involved. For simple errors by definition, the residual is independent of y, as is seen in the plot (a). If the errors are of a proportional nature the residual plot will be as plot (c). A plot of the relative residual ($\Delta y/y$ versus y) results in the profiles (b) and (d) for simple and proportional errors respectively, and confirms the error distribution.



Plot of Residual Errors

Above is shown the type of residuals plots obtained when the error is constant or proportional in nature. It must be stressed that visualization of clear cut plots of this kind depends upon having many data values to define the curves. The dotted lines on the plots show the

envelope within which all points are expected to be found. Residuals can be calculated during data fitting, or by using the **Manipulate Difference...** command.

Testing Goodness of Fit — *F* Test

There are situations where it is necessary to know whether a set of data is fitted better by one particular equation or another. The *F* test is a statistical way of judging the goodness of fit by two different equations. This test compares the χ^2 values obtained from the two fits and the degrees of freedom in each, and calculates the probability that the fits are equally good. The *F* statistic is calculated from the equation below.

$$F = \frac{(\chi_1^2 - \chi_2^2)/(v_1 - v_2)}{\chi_2^2 / v_2}$$

A low probability value indicates that one of the two equations (that giving the lower χ_v^2) fits the data significantly better than the other. To be significant the probability should be lower than 0.1, and preferably lower than 0.05. To perform the *F* test, use the **Manipulate F Test...** command (see Chapter 25 for more information).

Mean and Standard Deviation

For a full description of means and standard deviations, see any text on statistics (for example, that by Bevington[2]). A brief description follows.

The mean value of a set of data points is defined as:

$$\mu \approx \bar{x} = \frac{1}{N}\sum x_i$$

where μ is the mean of the parent distribution, and $\bar{x}$ is the mean of our experimental sample. The uncertainty, or standard deviation, of the data points, σ, is estimated from the data as:

$$\sigma \approx s = \sqrt{\frac{1}{N-1}\sum (x_i - \bar{x})}$$

and the uncertainty of the mean, σ_μ, is

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} \approx \frac{s}{\sqrt{N}}$$

All the above equations assume that the uncertainties of all the data points are equal. If they are not, for reasons described earlier, we must use equations that incorporate weighting. For example, calculation of the mean becomes:

$$\mu \approx \bar{x} = \frac{\sum (x_i / \sigma_i^2)}{\sum (1/\sigma_i^2)}$$

When calculating means using the **Manipulate Mean...** command, GraFit uses the weighting selected from the **Options Weighting** menu to determine how to perform the calculations.

30 Worked Data Fitting Examples

This chapter provides some further examples of data fitting to help acquaint you with the program. The examples are presented so that they can be worked through in a step-by-step manner. The data files that are used can be found in the \GRAFIT\DATA directory, and the equation files are in \GRAFIT\EQNS. Don't worry if the examples do not seem relevant to the type of data that you need to analyze — the same principles are involved in all data fitting.

This chapter assumes that you are familiar with the basic functioning of the program, and if not you may like to consult the earlier chapters.

Using One of the Equations Provided

We will fit the data obtained from an NMR titration curve to obtain the pK_a of an ionizable group.

Background The position of a peak in the NMR spectrum (the chemical shift position) resulting from a magnetic nucleus that is part of a titratable group depends upon the equilibrium between the two ionized forms. For example, in the following equilibrium:

$$A^- + H^+ \rightleftharpoons AH$$

Nuclei close to the ionizable group, A, will have chemical shift values that depend on the ratio of A^- and AH. By definition, the acid dissociation constant, K_a is given by

$$K_a = \frac{[A^-][H^+]}{[A]}$$

It is usually more convenient to use pK_a values, where

$$pK_a = -\log_{10} K_a$$

The variation of the ratio of A^- and AH is given by the Henderson-Hasselbalch equation:

$$pH = pK_a + \log_{10} \frac{[A^-]}{[HA]}$$

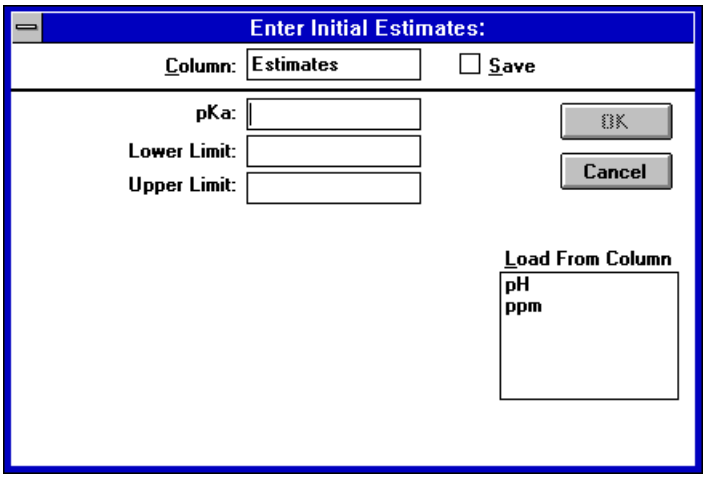
The chemical shift observed for the nucleus varies between that measured at the extreme acid range, δ_{AH}, and that at the extreme alkaline range, δ_A, depending on the ratio of $[A^-]/[HA]$.

The data obtained are chemical shift values (in units of ppm, or parts per million) at various pH readings. These values need to be fitted to a single ionization curve where the start and end chemical shift values and the pK_a are the unknown parameters.

Fitting the Data

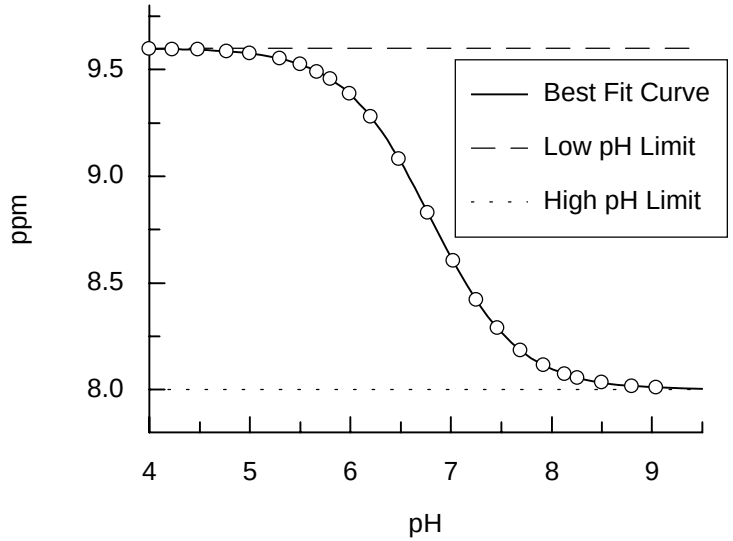
The example data are in the file PH.GFD, which can be opened using the **File Open...** command. Two columns are present: pH and Chemical Shift. The equation to be used is in the file PH.GFE. Load this file (**Options Definition file...** command) and select the equation 'pKa Determination, Single pKa'.

You are now ready to fit the data. Select the **Data Curvefit...** command. This equation is one for which initial estimates for the parameter values are difficult to provide. Therefore it is necessary to enter these initial estimates explicitly. Equations that require initial estimates will produce a dialog box similar to the one below.



The main portion of the dialog box allows you to enter initial estimates for the parameter values. As described in the Chapter 29, these estimates should be reasonably close to the true values. The program then iterates to find successively better values until a satisfactory fit is found. Obviously, to enter appropriate 'guesses' requires that you have some idea of what the parameter values should be. We describe later how it is possible to test out different values if you have no idea of the correct parameters. However, for this equation suitable parameters are: pKa, 7; Lower limit, 10; Upper limit, 8. (The Upper and Lower limits in this equation represent the chemical shifts of the fully ionized and non-ionized forms at the low pH and high pH extremes of the curve.)

When you have entered your initial estimates, press the **OK** button to start curve fitting. It is always advisable to examine the resultant fit by plotting a graph, as unsuitable initial estimates can cause sometimes the calculations to "converge" to values that describe the data poorly. The data that we have fitted here give the following curve.



It is clear in this instance that the fitted curve describes the data well.

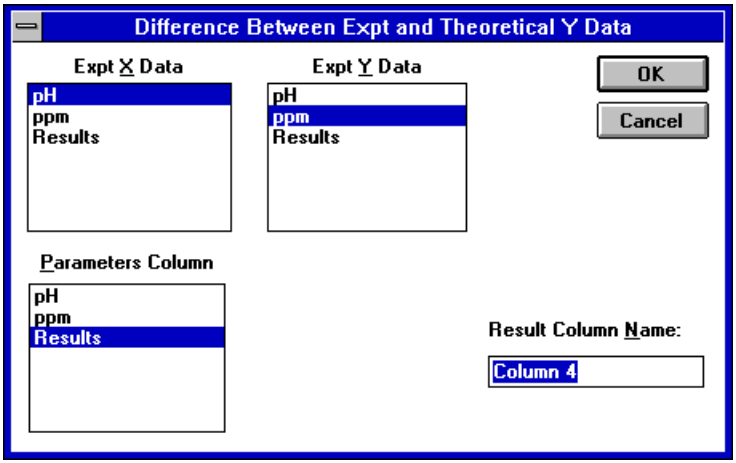
Checking the Weighting

It is often useful when performing an experiment for the first time to estimate the type of weighting that should be used. It must be stressed that for these checks to be meaningful a large number of data points should be available, and the equation describing the data must be well established.

As was described in Chapter 29, plots of residual error (Δy) versus y or $\Delta y/y$ versus y give characteristic patterns for the common simple and proportional error distributions. We will produce these plots for the data that have been fitted above.

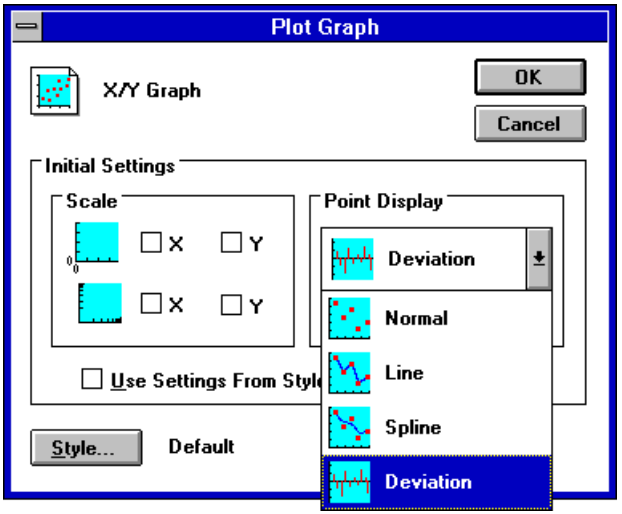
Drawing Residual Plots

We will plot the residual plot Δy versus y. To make this plot we need a column containing the Δy values. This can be generated during data fitting, or created after using the **Manipulate Difference...** command. Selecting the **Manipulate Difference...** command results in the following dialog box.

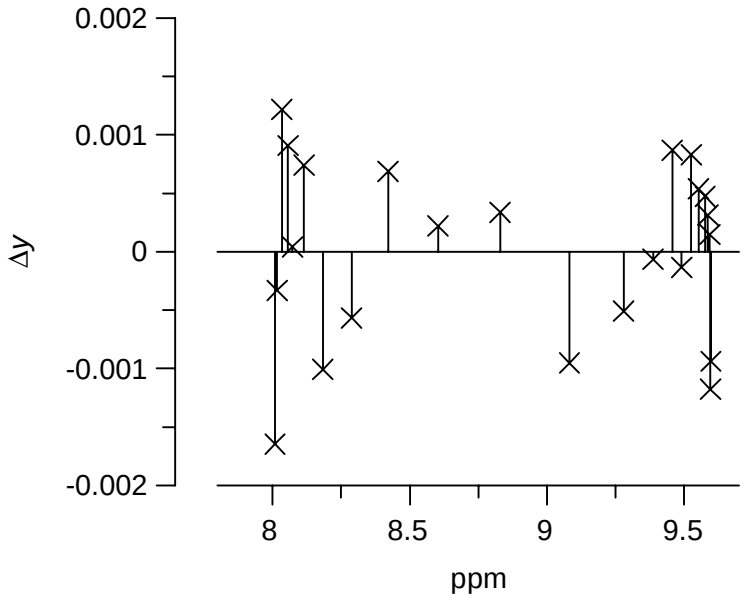


It is necessary to specify the columns that held the original *x* data (in this case the pH column), the *y* data (the ppm column), and the column holding the calculated parameter values from curve fitting. The results will be placed in a new column, which can be named using the Result Column Name edit box. The current equation is used to calculate the Δy values. Once these have been calculated we can plot a graph of Δy versus y.

The easiest way to give a residual plot is to create an X/Y scatter graph using the **New X/Y Graph...** command (Chapter 6), and to select the Deviation style from the Point Display section, as shown below.



The effect of this display style is that a line is drawn from the data point to the axis using an error bar (see Chapter 15 for a description of error bars). The graph below shows the resulting residual plot (after editing to enhance the appearance of the display).



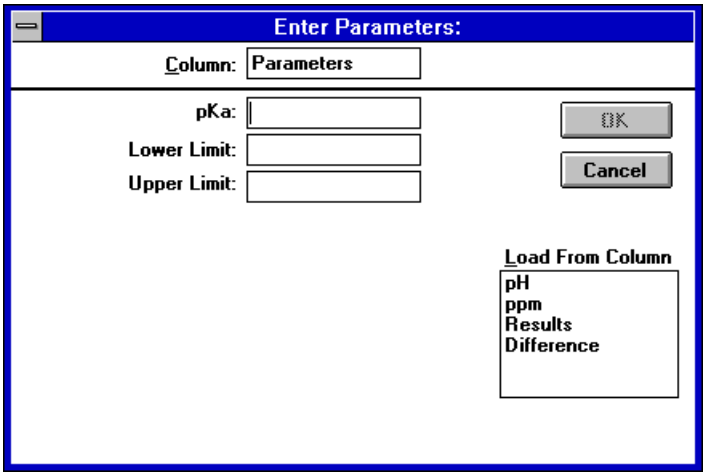
As is clear from the shape of this plot (see the discussions in Chapter 29), it was probably correct to use simple weighting with these data.

To make $\Delta y/y$ versus y plots, it is possible to use the **Manipulate Combine...** command to create a column resulting from dividing Δy by y.

Finding Initial Estimates

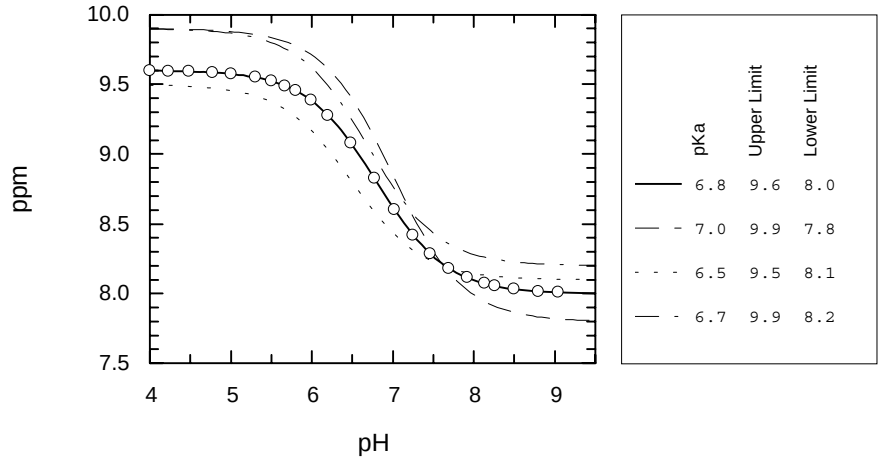
There are often times when you are uncertain about appropriate estimates for the parameter values. In these circumstances it is possible to use the graph plotting capabilities of GraFit to view the shapes of various curves that have different parameter values. From this it is possible to get some idea of what values produce reasonable curves. We will re-examine the data in PH.GFD that was fitted earlier.

To fit these data we need to know approximate values for the pK_a, the starting limit value and the end limit value. Suitable values are those giving a curve that does not deviate extensively from the data points. We will test values by overlaying curves until suitable estimates are found. First we will enter very approximate values into a new column, as the curve drawing routines expect to find their values in a column. The easiest way to create this column is to use the **Data Make Parameters Column...** command. This command results in the following dialog box.



The values that you enter will be stored in a newly created column. You can define the name of the new column using the Column edit box.

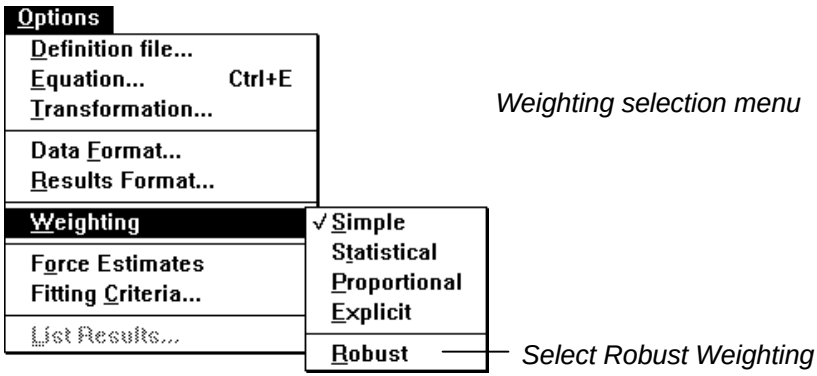
We can now examine these estimates by plotting a new graph, and using these values to define a curve. To test a new set of estimates we can do one of the following. Either enter a new set of estimates in a different column, and overplot these; or edit the values used to define the curve by double-clicking on the curve symbol and using the Edit Parameters edit box (see Chapter 14). A series of different overplots is shown on the graph below. When selecting suitable initial estimates it is necessary to use values that produce curves reasonably close to the data, but it is not usually necessary to be very precise.



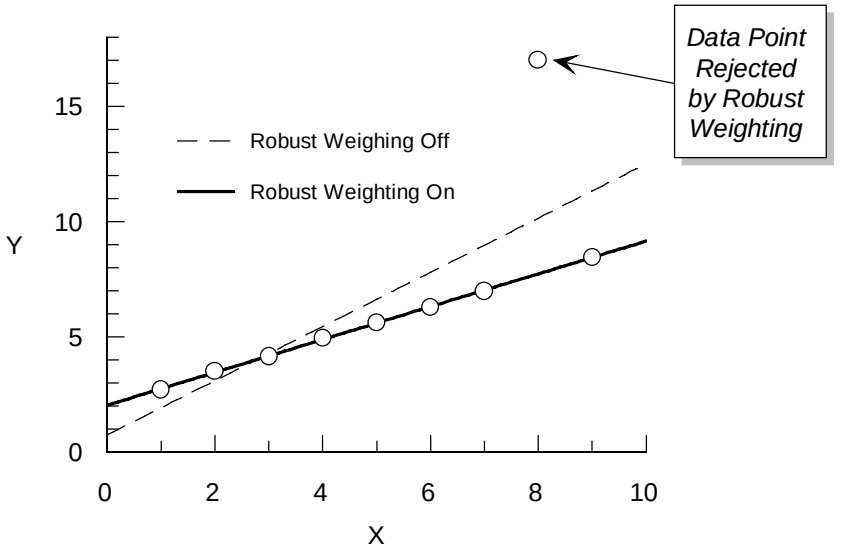
Using Robust Weighting

Robust weighting is an additional weighting that is applied to a data set. It acts to reject data points that are very much in error, and so removes the occasional rogue data point or "outlier" that may be present in a data set. A complete description of the background to robust weighting can be found in Chapter 29.

To show the detrimental effects of outliers on curve fitting, and see how robust weighting eliminates these effects, we will analyze the data in ROBUST.GRD. These data have a linear relationship, and so we need to select a linear regression equation from the equation definition file POLYNOM.GFE or LINEAR.GFE (use the **Options Definition file...** command to load the new equation definitions). ***Note*** it is not possible to use the **Data Linear fit...** command when using robust weighting. Fit the data using simple weighting with no robust weighting. Both options are chosen from the **Options Weighting** menu; when robust weighting is active a tick mark will appear beside the selection. Next fit the data with robust weighting selected.



The graph obtained by plotting the data with the best fit lines obtained with and without robust weighting is shown below.



As is apparent, without robust weighting selected the presence of the outlier biases the fit considerably. This is because the large sum of squares residual from this point has a disproportionate effect on the regression analysis. However, using robust weighting eliminates the outlier, and the regression line is drawn through the majority of the data points.

The Results listing is shown below.

```
File     : ROBUST.GFD
Directory: C:\MANUAL3\GRAPHS
Friday 7/08/92  20:58

y = a + bx1
Simple weighting

Robust weighting selected

Reduced Chi squared = 0.00223
```

Variable	Value	Std. Err.
a	2.0393	0.0351
b	0.7121	0.0067

	x1 X data	Y Y data	Calculated	
1	1.0000	2.7003	2.7514	
2	2.0000	3.5127	3.4635	
3	3.0000	4.1387	4.1755	
4	4.0000	4.9617	4.8876	
5	5.0000	5.6170	5.5996	
6	6.0000	6.2960	6.3117	
7	7.0000	6.9701	7.0238	
8	8.0000	17.0000	7.7358	!*!
9	9.0000	8.4646	8.4479	

Data values that are totally eliminated are flagged '!*!'. In this example only a single data point has been removed by the robust weighting.

Defining a New Equation (1)

There are few examples in the literature where both raw data and calculated values can be found. However, the paper by Roughton *et al.* (1955)[1] is notable because the raw data are presented with a set of calculated results. Also the equation to be fitted is somewhat complex. Because of this, these data provide a good test for a data analysis package. The Roughton paper examines the binding of oxygen to sheep hemoglobin as a function of the pressure of oxygen. Hemoglobin has four non-equivalent sites, and the binding is described by the equation:

$$\text{Saturation} = \frac{K_1 p + 2K_1K_2 p^2 + 3K_1K_2K_3 p^3 + 4K_1K_2K_3K_4 p^4}{4(1 + K_1 p + K_1K_2 p^2 + K_1K_2K_3 p^3 + K_1K_2K_3K_4 p^4)} \times 100\%$$

where p is the pressure of oxygen, and K_1 to K_4 are the equilibrium constants for the binding of four successive moles of oxygen per mole of hemoglobin.

The data are in the file HB_O2.GFD. Read this file in using the **File Open...** command. An interesting feature of these data is that some data points were collected with greater accuracy than others. To account for this it is necessary to select Explicit Weighting from the **Options Weighting** menu; the errors are to be found in the column 'Error'.

Defining the Oxygen Binding Equation

To define this equation we need to create a new definition. It is convenient to place this in a new definition file. Select the **Options Definition file...** command, and enter a new file name (for example 'OXYGEN'). When you enter a new file name in this way the program will ask you whether you really wish to create a new file; select **Yes**. Create a new equation by choosing the **New** menu command on the resulting dialog box. This will produce a blank equation edit dialog box. (See Chapter 23 for a complete description of defining equations.)

Definition of the new equation requires that we make appropriate entries in this dialog box.

---

[1] Roughton, F.J.W., Otis, A.B. & Lyster, R.L.J. (1955) *Proc. Roy. Soc. London Ser B* **144**, 29-54.

Equation Name: `Oxygen binding to hemoglobin`

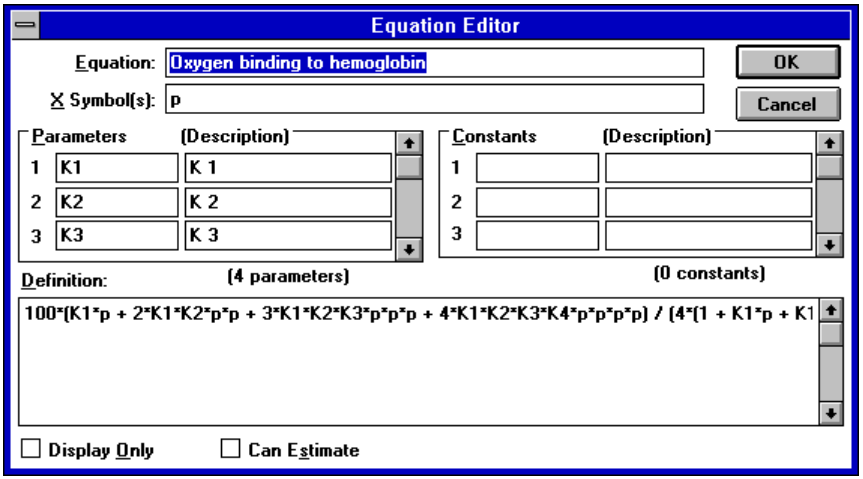
Parameter	**Description**	**Symbol Name**
K_1	`K 1`	`K1`
K_2	`K 2`	`K2`
K_3	`K 3`	`K3`
K_4	`K 4`	`K4`

X Symbol: `p`

Definition:[2]

```
100*(K1*p + 2*K1*K2*p*p + 3*K1*K2*K3*p*p*p +
4*K1*K2*K3*K4*p*p*p*p) / (4*(1 + K1*p +
K1*K2*p*p + K1*K2*K3*p*p*p +
K1*K2*K3*K4*p*p*p*p))
```

The appearance of the dialog box after entering this definition is shown below.

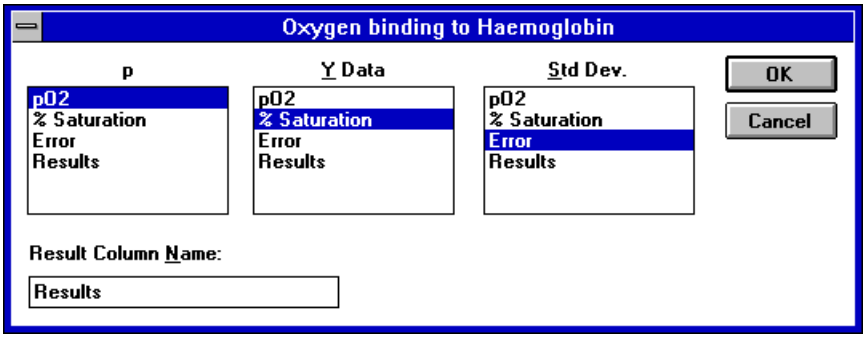


Calculate the results by fitting the data using the **Data Curvefit...** command after selecting explicit weighting. When explicit weighting is chosen it is necessary to indicate which column contains the error values. The dialog box produced is shown below.

---

[2] As this definition is for demonstration purposes only, it is defined in this way as this is the nearest form to that of the equation given. Although the definition would function in this form, it is rather inefficient to write it like this, and a better approach would be to enter the mathematically simpler:

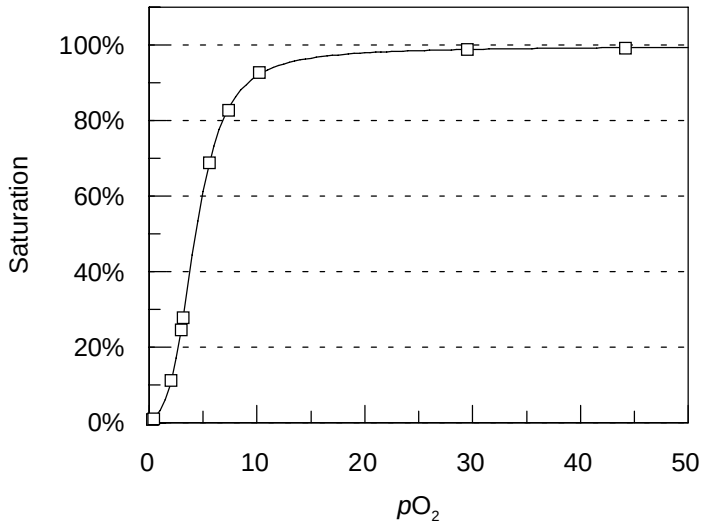
```
100 * K1*p*(1 + K2*p*(2 + K3*p*(3 + K4*p*4))) /
(4*(1 + K1*p*(1 + K2*p*(1 + K3*p*(1 + K4*p)))))
```



For comparison, the results quoted by Roughton *et al.* are given below.

	Parameter value	Standard Error
K_1	0.0961	0.0049
K_2	0.1040	0.0293
K_3	0.3352	0.1018
K_4	0.9263	0.0648

It should be borne in mind that Roughton *et al.* performed their regression analysis without the aid of personal computers or pocket calculators! The best fit curve to the data is shown below.



Defining a New Equation (2)

The hybridization of RNA and cDNA molecules is used to provide information about the genome organization and complexity of an organism. This hybridization can be described by the equation:

$$\frac{d}{D_0} = B + \sum_{i=1}^{n} P_i \left[1 - e^{-0.693(Rot / Rot_{0.5(i)})} \right]$$

where d/D_0 is the fractional hybridization at any given value of Rot (moles second L^{-1} of nucleotides of RNA), B is a background value, P_i is the proportion of the total cDNA hybridized to the *i*th component, and $Rot_{0.5(i)}$ is the Rot value at which 50% of the cDNA of the *i*th component is hybridized. In practice, the data obtained may be fitted assuming three binding components, giving a total of seven independent variables. Green *et al.*[3] have presented a sample of such data, these can be found in the file DNA.GFD.

Defining the Equation

The equation can be defined as follows:

Equation Name: `Nucleic Acid Hybridization (3 comp)`

Parameter	Description	Symbol Name
B	Background	B
P_1	Proportion Component 1	P1
$Rot_{0.5(1)}$	Rot 0.5 (1)	R1
P_2	Proportion Component 2	P2
$Rot_{0.5(2)}$	Rot 0.5 (2)	R2
P_3	Proportion Component 3	P3
$Rot_{0.5(3)}$	Rot 0.5 (3)	R3

X Symbol: `Rot`

Definition:

```
B + P1*(1-exp(-0.693*Rot/R1))
  + P2*(1-exp(-0.693*Rot/R2))
  + P3*(1-exp(-0.693*Rot/R3))
```

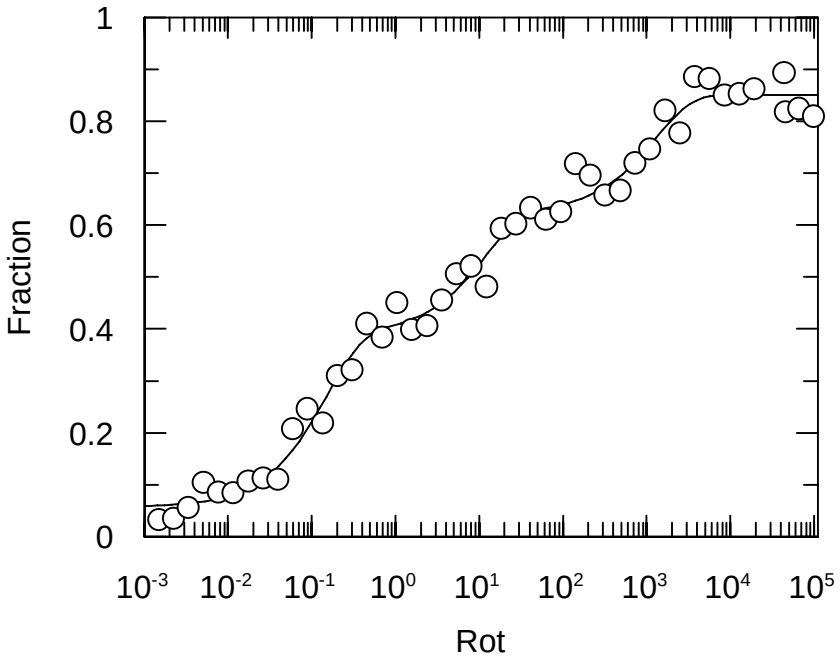
[3] Green, S., Field, J.K., Green, C.D. & Beynon, R.J. (1982) *Nucleic Acids Research* **10**, 1411-1421.

Fitting the Equation

The following initial estimates are appropriate to analyze the data.

B	0.05
P_1	0.3
$Rot_{0.5(1)}$	0.2
P_2	0.3
$Rot_{0.5(2)}$	40.0
P_3	0.3
$Rot_{0.5(3)}$	600.0

It is interesting to benchmark your computer using this equation. Green *et al.* using an Apple II computer and employing a "Patternsearch" algorithm report that these data were fitted in approximately two hours of calculation. Using a 33MHz 80486-based system, GraFit fits these data in less than 3 seconds. The calculated result is displayed below.



Multiple Regression

Multiple regression is used when the observed data values depend on more than one dependent (x) value. As an example of this, we will examine the data listed by Toney & Kirsch[4]. These data consist of observations of the rate constant of a reaction, k, as a function of the pK_a of the group involved in the reaction and the molecular volume of this group, Vol. The data are described by the equation

$$\log k = \beta.K_a + V.\mathrm{Vol} + c$$

where β, V and c are unknowns in the equation, which is a multiple linear equation having two independent variables. The data presented in the paper are in the file MULTLIN.GFD. It is possible to fit these data using the appropriate equation in the definition file LINEAR.GFE, or by defining a new specific equation. The required definitions are as follows.

Equation Name: `Multilinear kinetic equation`

Parameter	Description	Symbol Name
β	`Beta`	`beta`
V	`V`	`V`
c	`Constant`	`c`

X Symbols: `pKa Vol`

Definition:
```
beta*pKa + V*Vol + c
```

The results obtained from fitting these data are:

Variable	Value	Standard Error
β	0.39	0.05
V	-0.055	0.005
c	-0.7	0.5

Displaying Multidimensional Data

Displaying data sets that have more than one x axis can be difficult. One solution is to remove all but one dimension from a plot. For example, in the above data, a plot of ($\log k - V.\mathrm{Vol}$) versus pK_a will result in a linear graph with a slope β and intercept c. We can create the ($\log k - V.\mathrm{Vol}$) data using the data manipulation abilities of GraFit; the result for V is obtained from the curve fitting.

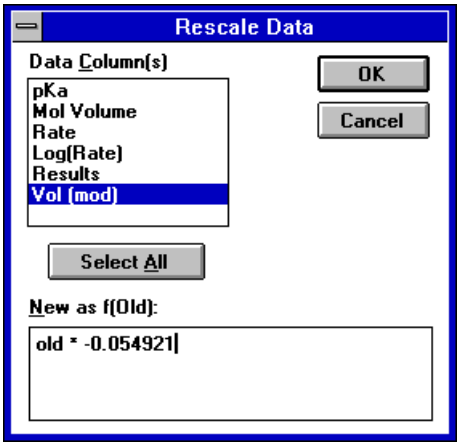
The first step is to duplicate the column containing the Vol data. To do this, click on the column number above the 'Mol Volume' label, select **Edit Copy** followed by **Edit Paste**.

[4]Toney, M.D. & Kirsch, J.F. (1989) *Science* **243**, 1485-1488.

This will duplicate the column of data (if you are unsure of how to copy entire columns of data, consult Chapter 2).

Rescaling the Data

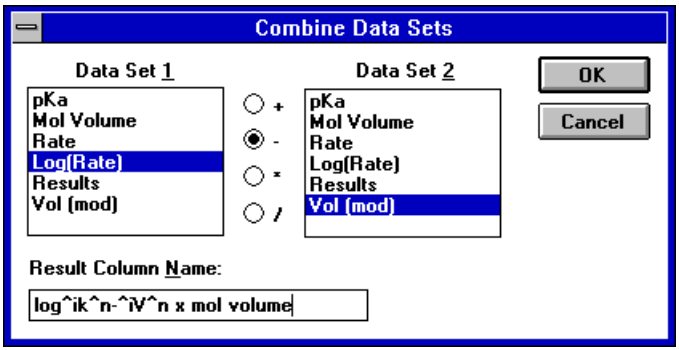
Next we need to modify the contents of this column to be *V*.Vol. At present the new column is labeled 'Mol Volume', just as the original column. It is helpful to rename this column to avoid confusion. Rename the column using the **Data Rename Column...** command, or by double-clicking on the column name — a suitable name is 'Vol (mod)'. To make the Vol (mod) column hold *V*.Vol values use the **Manipulate Rescale...** command.



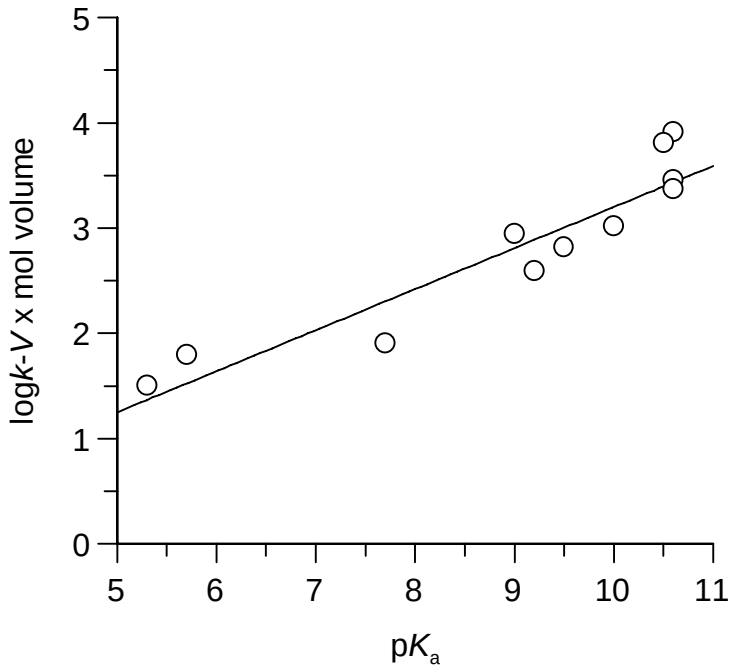
Select the Vol (mod) column, and define the new contents of the column to be `old * -0.055`. (***Hint*** it is possible to copy the value of *V* from the data table, and then paste it into the Rescale dialog box by pressing SHIFT INSERT.)

Subtracting Data Columns

We now have the value of *V*.Vol in the 'Vol (mod)' column, and log *k* in the original 'Log(Rate)' column. To create the (log*k* – *V*.Vol) data use the **Manipulate Combine...** command to subtract these data columns and place the result in a new column.



The result, plotting (log*k* – *V*.Vol) versus pK_a and using a linear equation to display the appropriate line through the data is shown below.



Reading Values from a Calibration Curve

A calibration curve records the variation of an observed value when an experimental condition is altered in a defined manner. For example, an instrumental reading may be measured for various known concentrations of a solution. The standard curve that is produced can then be used to allow, in this example, the concentrations present in an unknown solution to be determined by extrapolating values from the curve.

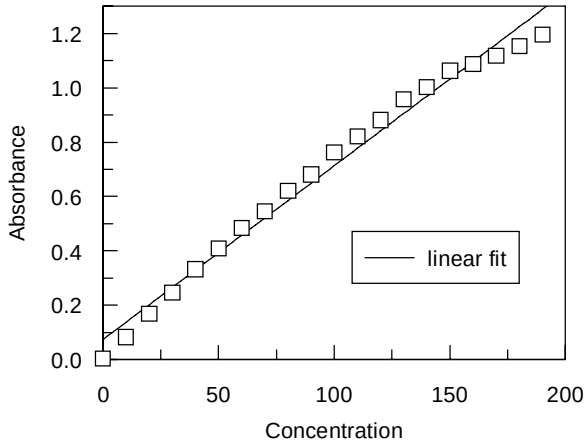
These processes can be performed using GraFit by the following procedure.

1. Fit the data to an equation that describes the observed trend.
2. Use the **Manipulate Calculate X...** or **Manipulate Calculate Y...** commands to read values from the fitted curve.

Choosing an Equation to Fit the Data

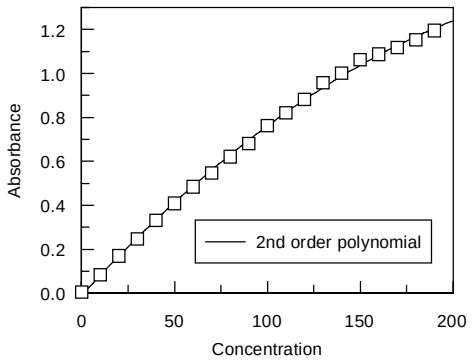
In many circumstances, the relationship between the *x* and *y* data values will be well-defined. In these cases, the data will be fitted by linear or non-linear regression in the usual manner. However, some calibration data show a clear trend of how *y* varies with *x*, but are not necessarily defined by a well-characterized equation.

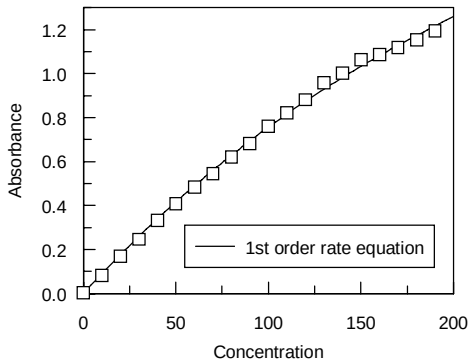
For example, the data below (present in the file STDCURVE.GFD) show the absorbance reading given by varying concentrations of a compound, and a straight line fit to the data.



In theory, the plot of absorbance versus concentration should be linear, but the data shown above describe a shallow curve. While it might be of interest to find out why this effect occurs, the only important consideration here is to allow readings to be taken from the calibration curve. Quite clearly, fitting the data to a straight line is not suitable.

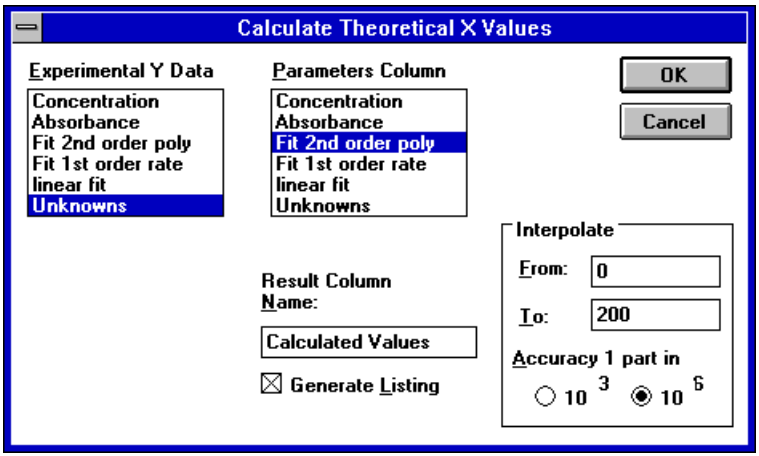
Under these circumstances, the best recourse is to fit the data to an arbitrary equation that allows a smooth curve to be drawn through the data. As there is no theoretical basis for the equation, any calculated parameter values will have no physical significance. The choice of equation will depend upon the shape of the data, and it is possible that more that one equation would prove suitable. In the case shown above it might be appropriate to use a polynomial, or any equation that can produce a gentle curve. Below is shown the results of fitting the data to a second order polynomial, and to a first order rate equation. The latter equation has no possible theoretical justification, but has a suitable mathematical form to allow a smooth curve to run through the data.





Reading Values from the Fitted Curve

Either of the fitted curves shown above will allow the concentration values that correspond to a series of measured absorbance values to be read from the curve. This is performed using the **Manipulate Calculate X...** command (described fully in Chapter 25).



This command assumes that the equation to be used has been selected using the **Options Equation...** command. The two equations that were used above can be found in the definitions files POLYNOM.GFE and RATE.GFE respectively.

The Results listing from this command, using the second order polynomial fit, is given below.

```
File     : STDCURVE.GFD
Directory: C:\GRAFIT\DATA
Sunday 27/09/92  13:15

y = a + bx + cx^2

Experimental data in column "Unknowns"
Calculated values in column "Calculated Values"
Parameters from column      "Fit 2nd order poly"

Parameters:
  a                         =    -0.008889
  b                         =     0.009148
  c                         =   -1.45e-005


_____________________________

Experiment   Calculated
_____________________________

        0.32        38.28
       0.734        95.74
        0.04         5.39
        0.92        127.2
_____________________________
```

31 Advanced Data Fitting Topics

This chapter describes some advanced data fitting techniques that can be applied using GraFit. These techniques are more complicated than those for simple data fitting, and require a thorough knowledge of how equations are defined before they can be useful.

Fixing the Value of a Parameter

There are occasions when it is desirable to fix the value of one or more of the parameters describing an equation. For example, the equation for a first order rate process is

$$y = A_{\infty}(1 - e^{-kt})$$

A_{∞} is the limiting value of y at time $t = \infty$, k is the rate constant. If A_{∞} is known from a separate experiment, then this value could be included in the equation explicitly so that only the value for k need be found by data fitting. To do this, simply define a separate equation in which the parameter to be fixed is specified as a constant. When fitting data to the new equation you will be prompted to enter the value for the constant during the analysis.

Constrained Minimization

Constrained minimization is where one or more parameter values are forced to stay within a certain range during minimization. This range may be numerical, for example forcing a parameter to have a value between 10 and 20, or relative to a second parameter, for example constraining one parameter to be always less than a second. GraFit allows complete flexibility in defining constraints within equation definitions.

Method of Constrained Minimization

GraFit can build constraints within the body of an equation definition. This gives you full control over the constraints that you use, but means that a separate equation must be defined and selected when you undertake constrained minimization. The need to constrain parameter values is likely to be extremely specific to a particular experimental situation, and so no predefined equations are provided that have constraints. The following examples give some ideas of how constraints can be built in to any equation that you use.

There are three common features that must be built into the equation in order to use constraints within GraFit.

1. When the variables satisfy the constraints, the calculated *y* value is defined by the usual equation.
2. When the constrained value(s) is outside its limit, the calculated *y* value returned by the equation must be altered, so that the value returned will result in a very poor fit to the data.

3. The method used in (2) to give the poor *y* value must give *y* values that are increasingly worse as the constrained parameter value tends away from the desired limits. Minimization will therefore act to force the values to the acceptable range.

Constraining a Parameter to be Greater than a Minimum Value

The equation for a standard rate equation is given by

$$y = A_{\infty}(1 - e^{-kt})$$

The definition for this equation, in the usual GraFit equation notation (only the right hand side of the equation need be defined), is

```
A * (1 - exp(-k * t))
```

Let us suppose that we want A_{∞} to be greater than or equal to 10. This can be done by defining the equation as

```
if A<10 then
  return (A * (1 - exp(-k * t)) + (10-A)*1000*A)
else
  return (A * (1 - exp(-k * t)))
endif
```

If A_{∞} is greater than or equal to 10 then the normal relationship is used. However, if A_{∞} is less than 10, the calculation returns a value that will give a poor fit. This deliberate miscalculation will act to force A_{∞} to be within the constraints specified. In general, to force a parameter, *v*, to be greater than a certain minimum value, *m*, you can add a term (*v*–*m*)***S* to the equation definition, where *S* is the severity by which the equation is forced to be constrained. In the above example, *S* was set to be `1000*A`. You should note that the success of the constrained minimization will depend on the value chosen for *S*, and that this is likely to vary depending on the numerical values of the data, and the equation used.

As described above, it may seem that a new equation would need to be defined for any value that we wish to use as a minimum. This need can be removed by using a constant to provide the minimum value, for example as shown below.

```
if A<m then
  return (A * (1 - exp(-k * t)) + (m-A)*1000*A)
else
  return (A * (1 - exp(-k * t)))
endif
```

where the minimum value for *k*, `m`, is defined as a constant. GraFit will allow you to enter the value for the constraint whenever you fit the data. If you wish to vary the severity of the constraint, this is also possible by defining a second constant. For example

```
if A<m then
  return (A * (1 - exp(-k * t)) + (m-A)*S*A)
else
  return (A * (1 - exp(-k * t)))
endif
```

where the constants are `m` and `S`, the severity factor.

Constraining a Parameter Value to be Within Upper and Lower Boundaries

By extension of the example above, if it is necessary to force the limiting value in the above equation, A_{∞}, to be within the range 100-200, one way of doing this would be to use the definition

```
if A<100 then
  return (A * (1 - exp(-k * t)) + (100-A)*1000*A)
else
  if A>200 then
    return (A * (1 - exp(-k * t)) + (A-200)*1000*A)
  else
    return (A * (1 - exp(-k * t)))
  endif
endif
```

As before, the exact range and severity could be defined as constants to be entered prior to fitting the equation.

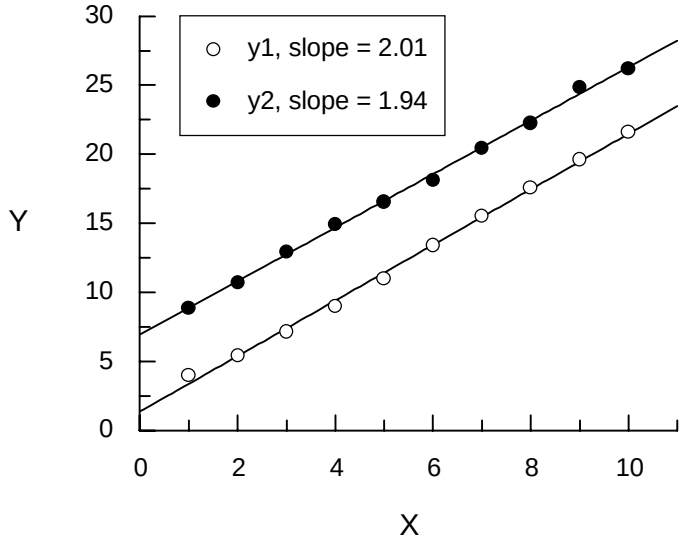
Warnings About Constrained Minimization

Although it is possible to extend the examples shown above to exert various contraints on the fitting, it must be stressed that there should be good theoretical reasons for wishing to use constraints within your data fitting. In particular, you must not use constraints to force bad data to give "correct" fitted values. In addition, the error estimates that are given after constrained minimization should be treated with extreme caution.

Fitting Several Data Sets With Some Parameters Common and Some Specific

To illustrate, imagine that you have collected two linear data sets, and that you already know that for this experiment the slope is the same for each although the intercepts are different. Using GraFit it is possible to fit these two data sets to find the best slope that is common to each, and the two intercept values that are specific to the two data sets. Indeed, it is possible to fit two or more data sets to any of the normal equations that GraFit can use and to share one or several parameters between the data sets.

The example below shows how this is done for the paired linear data. Two raw data sets are illustrated below (these data are in the example file PAIR.GFD). Fitting these two data sets individually results in similar, but not identical, values for the slope. This is shown in the figure below.



If the data sets are known to have the same slope, this can be incorporated into the data fitting as follows.

Preparing the Data for Simultaneous Fitting

To fit the two data sets simultaneously it is necessary to merge the two *x*/*y* data sets to create a merged *x* data column and a merged *y* data column, and to create a new data column that holds a value to show which data set is involved. For two data sets it is convenient if this third data column holds the value "1" or "2" respectively. For example,

x_1	y_1	x_2	y_2	***x* merge**	***y* merge**	**data set**
1	4	1	8	1	4	1
2	5	2	9	2	5	1
3	6	3	10	3	6	1
...	...	...	...	...	...	1
				1	8	2
				2	9	2
				3	10	2
				...	...	2

While it is possible to create these three new columns by cutting and pasting parts of the data table, the **Manipulate Make 3D Data Set...** command provides a short cut (see Chapter 25 for a full description of this command).

Method Used to Share Variables Between Data Sets

The column that identifies the data set is accessed as a second independent (*x*) variable within the equation definition, which is defined in such a way that that this value acts as a switch to decide which parameters apply to the data value. For the simple example described above, the independent variables are `x` and `dataset`; parameters `a1` and `a2` are the two intercepts, and `b` is the common slope. This equation is defined as

```
if dataset=1 then
  return (a1 + b*x)
endif
return (a2 + b*x)
```

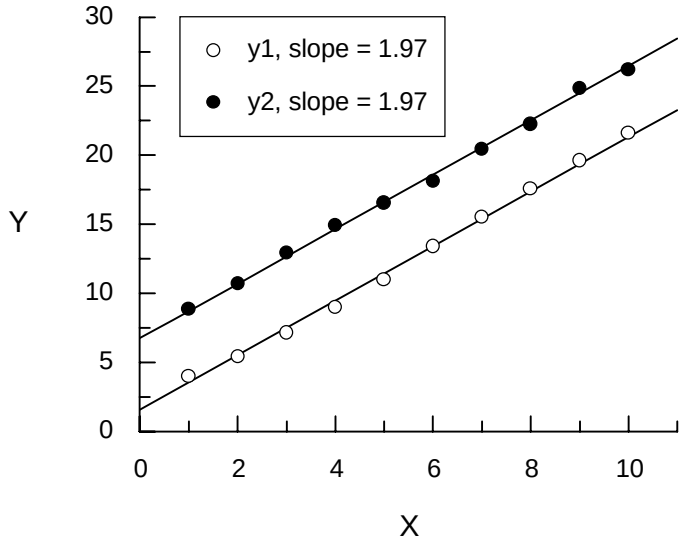
How does this work? When data from the first data set are accessed, the value of `dataset` is 1 and the equation returns

```
a1 + b*x
```

When data from the second data set are accessed, the value of `dataset` is 2 and the equation returns

```
a2 + b*x
```

Therefore the value of `a1` is determined, in effect, from only the first set of data, and `a2` from the second. However, the slope `b` is applicable to all the data. The graph below shows these data fitted to a common slope in this manner. The equations used are in the file PAIR.GFE. This file also defines two “display only” definitions that are used to plot the fitted lines using the calculated results.



It is simple to extend this approach to many data sets. The equation to find the best slope through four data sets, and a separate intercept for each is

```
if dataset=1 then
  return (a1 + b*x)
endif
if dataset=2 then
  return (a2 + b*x)
endif
if dataset=3 then
  return (a3 + b*x)
endif
' assumes that dataset must be 4
return (a4 + b*x)
```

The data would need to present in the data table as

***x* merge**	***y* merge**	**data set**
1	2	1
2	3	1
(etc.)	*(etc.)*	1
1	3	2
2	4	2
(etc.)	*(etc.)*	2
1	4	3
2	5	3
(etc.)	*(etc.)*	3
1	5	4
2	6	4
(etc)	*(etc)*	4

Example — Paired Denaturation Curves

It is important to note that this technique of sharing parameters between two or more data sets is not limited to linear equations. Any equation may be recast in this manner. For example, the equation below describes the denaturation of a protein as a function of the concentration of a denaturing solvent[1].

$$F = 1 - \frac{\exp\left(\frac{m[D] - \Delta G}{RT}\right)}{1 + \exp\left(\frac{m[D] - \Delta G}{RT}\right)}$$

where F is the fraction of the protein that is denatured, $[D]$ is the concentration of denaturant (e.g. guanidinium hydrochloride or urea), R is the gas constant, T is the temperature, ΔG is the stability of the protein, and m is a slope factor. Finding the difference in stability between two proteins is often more important than determining the absolute stability. If the proteins are highly related, it can be assumed that the value for m is the same for each, but ΔG differs.

Using the technique described above, it is possible to fit two data sets to a common m value but different ΔG values. In this case, the equation to be used can be defined as

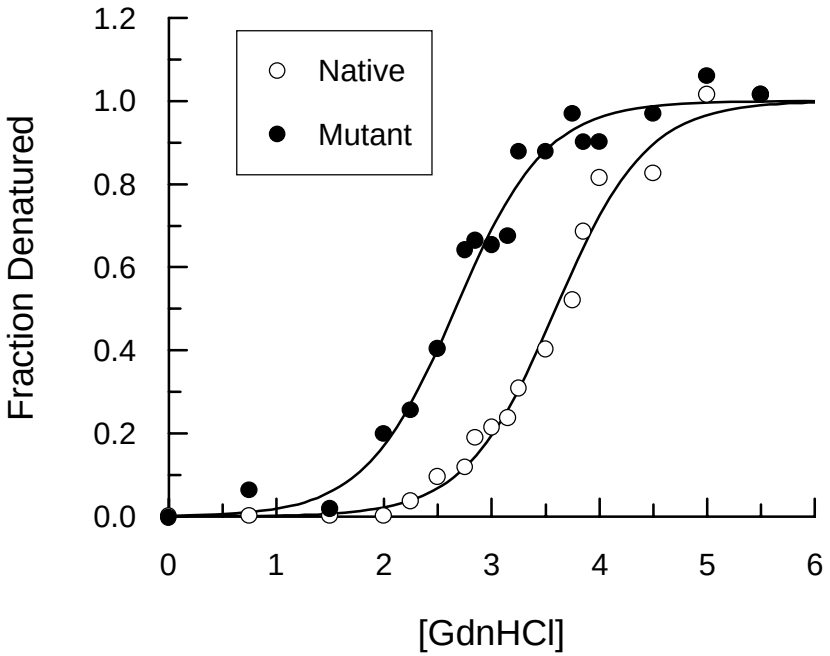
```
if dataset=1 then
  a := exp((m*D - G1)/(8.314*T))
else
  a := exp((m*D - G2)/(8.314*T))
endif
return (1 - a/(1 + a))
```

The *x* variable is the concentration of denaturant, `D`, the slope factor is `m`, the two values of ΔG are `G1` and `G2`. The temperature is `T`, and should be defined as a constant. The

[1] Jandu, S.K., Ray, S.R., Brooks, L. & Leatherbarrow, R.J. (1990) *Biochemistry* **29,** 6265-6269.

`dataset` column (x_2) must hold the value 1.0 if the *x* and *y* columns are from the first data set, and 2.0 if they are from the second. The value `a` in the above definition is a temporary value used to simplify the calculation.

The figure below shows the guanidinium hydrochloride unfolding of a native protein, and of a mutant variant of this protein. The mutation has caused the protein to become less stable, making it unfold at lower concentrations of denaturant. The solid lines indicate the fitted curves, using the above equation. The difference in stability can therefore be determined.



32 Equations Provided

GraFit comes with a large number of equations provided. These equations may be edited to see how they have been defined, and further equations can be added using the equation editor. This chapter briefly describes the equations that come with the program. It does not attempt to discuss the detailed background of each equation, as it is assumed that the user of a specific equation will be familiar with this.

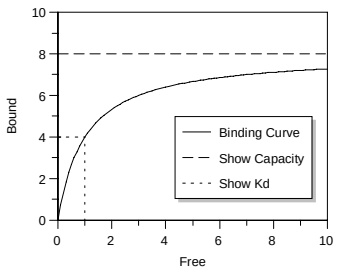
The equations that are present include ones defined for data fitting, and ones that are for graphical display purposes only. The convention used in all the equations provided is that equations for "display only" begin with ' *', for example ' * Show kcat'. Display equations used with transformed data are prefixed with ' #', for example ' # Display Linearized plot'. The equations are ordered so that display equations relating to a specific equation are listed beneath the main equation.

BIND.GFE

This definition file contains various equations describing the binding of ligands to surfaces, receptors, enzymes etc.

Equations

Ligand Binding - 1 Site



This is the simplest form of binding equation. The equation is

$$y = \frac{[\mathrm{L}].\mathrm{Cap}}{K_\mathrm{d} + [\mathrm{L}]}$$

where y is either the amount bound or is some factor proportional to it (e.g. radioactivity, absorbance etc.), or the number of moles bound per mole of receptor. The capacity for binding ligand, Cap, is either an amount bound in terms of moles of ligand, or a stoichiometric quantity in terms of moles of ligand per mole receptor depending upon the definition of y. The dissociation constant of ligand from the receptor is K_d. The concentration of free ligand, [L], is the x data axis.

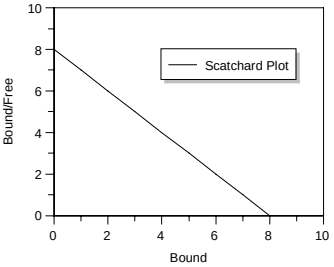
* Show Cap (1 site)

This display equation shows the capacity.

* Show Kd (1 site)

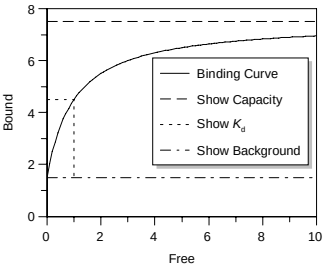
This display equation shows the dissociation constant, K_d. For best results set the curve resolution to High when using this display equation (see Chapter 14).

Display Scatchard (1 Site)



The Scatchard plot is the usual way to linearize binding data. This display equation takes the Cap and K_d values calculated by fitting to a single site binding curve, and draws the line for the Scatchard plot. Use this display curve with the Scatchard data transformation.

Ligand Binding - 1 Site + Background



Often binding experiments have an inherent background *y* value. This value can usually be determined separately by control experiments. However, such controls sometimes do not accurately reflect the real experiment, and it can be preferable to treat this background as a separate parameter to be obtained from the data fitting. The equation is simply:

$$y = \frac{[\mathrm{L}].\mathrm{Cap}}{K_d + [\mathrm{L}]} + \mathrm{background}$$

where the parameters are the same as for the single site ligand binding except for the inclusion of a background or offset value.

* Show Cap (1 site + back)

This display equation shows the capacity.

* Show Kd (1 site + back)

This display equation shows the dissociation constant, K_d. For best results set the curve resolution to High when using this display equation.

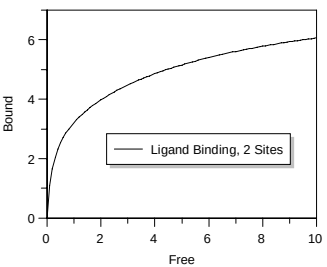
* Show Background

This display equation shows the background value.

Display Scatchard (1 Site + Back)

Use this display equation to display Scatchard plots for binding data containing a background value.

Ligand Binding - 2 Sites



Often there is more than one type of site involved in binding the ligand. This equation describes the case of two sites characterized by individual Kd and Cap values. The equation is:

$$y = \frac{[\mathrm{L}].\mathrm{Cap}_{(1)}}{K_{d(1)} + [\mathrm{L}]} + \frac{[\mathrm{L}].\mathrm{Cap}_{(2)}}{K_{d(2)} + [\mathrm{L}]}$$

* Show Cap 1 (2 sites)

This display equation shows the capacity of the first binding site.

* Show Cap 1 + Cap 2 (2 sites)

This display equation shows the total binding capacity for ligand.

* Show Curve 1 (2 sites)

This display equation shows the binding curve for the first binding site in isolation.

* Show Curve 2 (2 sites)

This display equation shows the binding curve for the second binding site in isolation.

Ligand Binding - 2 Sites + Background

This equation provides background determination for the case where two independent binding sites are present. The equation for this is:

$$y = \frac{[\mathrm{L}].\mathrm{Cap}_{(1)}}{K_{\mathrm{d}(1)} + [\mathrm{L}]} + \frac{[\mathrm{L}].\mathrm{Cap}_{(2)}}{K_{\mathrm{d}(2)} + [\mathrm{L}]} + \mathrm{background}$$

* Show Cap 1 (2 sites + back)

This display equation shows the capacity of the first binding site. This value is on top of the background.

* Show Cap 1 + Cap 2 (2 sites + back)

This display equation shows the total binding capacity for ligand, on top of the background value.

* Show Background (2 sites + back)

This display equation shows the background binding.

* Show Curve 1 (2 sites + back)

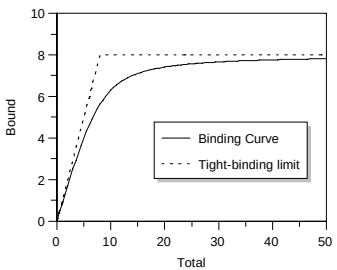
This display equation shows the binding curve for the first binding site in isolation.

* Show Curve 2 (2 sites + back)

This display equation shows the binding curve for the second binding site in isolation.

Bound vs. Total

The previous equations require the unbound ligand concentration to be known. However, although the total ligand concentration is always known (this is what you have added), the proportion bound and free is not always certain. This is particularly the case if the *y* axis is an observation that is not directly related to the bound concentration, for example fluorescence intensity or absorbance. Here the equation relates the *y* axis to the total ligand added, *t*.



$$y = -\frac{-(K_d + t + \mathrm{Cap}) + \sqrt{(K_d + t + \mathrm{Cap})^2 - 4.t.\mathrm{Cap}}}{2}$$

* Bound vs Total (tight bind limit)

This display equation shows the maximum possible amount bound (or change in absorbance etc.) for a given total amount of ligand added if the binding were infinitely tight.

Ligand binding - Cooperative

A generalized equation describing cooperative binding (or multiple binding sites of differing affinities) is:

$$y = \frac{[\mathrm{L}]^n.\mathrm{Cap}}{(K_d + [\mathrm{L}])^n}$$

where *n* is a measure of the cooperativity. For normal binding to a single site, *n* should not differ significantly from 1.0.

You can use the 'Show Cap' and 'Show Kd' display equations from 'Ligand Binding - 1 Site' to display these parameters; the 'Ligand Binding - 1 Site' equation will show the curve given by $n = 0$.

Ligand binding - Cooperative + Back

This is the same as the equation above, but includes an unknown background level in the *y* axis observations, i.e.

$$y = \frac{[\mathrm{L}]^n.\mathrm{Cap}}{(K_d + [\mathrm{L}])^n} + \mathrm{background}$$

You can use the 'Show Cap', 'Show Kd' and 'Show Background' display equations from 'Ligand Binding - 1 Site + Background' to display these parameters; the 'Ligand Binding - 1 Site + Background' equation will show the curve given by $n = 0$.

This equation is probably the best with which to fit a 'Dose response curve' (typically displayed with a logarithmic *x* axis). It is possible to read *x* axis values from the graph given a series of *y* axis "unknown" values by using the **Manipulate Calculate X...** command of the main menu.

Adair Equation - 2 sites

The Adair equation describes the binding of ligands to multiple sites on a receptor when the dissociation constants for binding of successive ligands are to be individually defined from the data. The equation for *n* sites is:

$$\mathrm{B} = \frac{\mathrm{Cap}.\{\mathrm{L}/K_{d1} + 2\mathrm{L}^2/(K_{d1}K_{d2}) + \cdots + n\mathrm{L}^n/(K_{d1}K_{d2}\cdots K_{dn})\}}{n.\{1 + \mathrm{L}/K_{d1} + \mathrm{L}^2/(K_{d1}K_{d2}) + \cdots + \mathrm{L}^n/(K_{d1}K_{d2}\cdots K_{dn})\}}$$

where B is the amount bound, L the concentration of free ligand, Cap the maximum capacity of the system to bind ligand, and K_{d1}, K_{d2} etc. the successive dissociation constants for ligand.

Adair Equation - 3 sites

Use this equation for 3 binding sites.

Adair Equation - 4 sites

Use this equation for 4 binding sites.

Transformations

Remove Background

This transformation is used for binding data that have background present; it simply subtracts the background from the *y* data.

Scatchard

The Scatchard plot is the most common means of representing binding data in a linear form. This transformation rearranges the data to allow a plot of Bound/Free vs. Bound.

Scatchard (Background present)

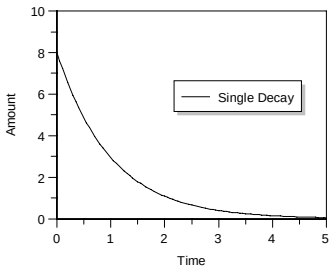
This transformation simultaneously removes the calculated background and rearranges the data to allow a Scatchard plot.

EXP.GFE

This definition file contains equations for calculating the rate constants of various exponential decay processes.

Equations

Single Exponential Decay

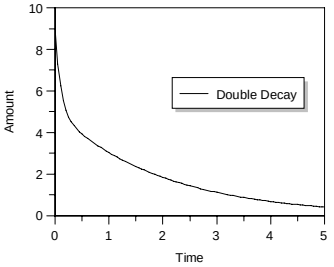


The equation for a single exponential decay is:

$$y = A_0 . e^{-kt}$$

i.e. the observed values decay with time, *t*, from an initial value, A_0, at $t = 0$, to 0 at $t = \infty$. The rate constant for the decay is *k*. (The half-life for the process, $T_{½}$, is given by $\log_e 2/k$.) To display a linearized plot of this equation, simply alter the *y* axis to a logarithmic scale.

Double Exponential Decay



If two decay processes occur simultaneously the observed values decay with time following the double exponential decay equation:

$$y = \mathrm{A}_{0(1)}.e^{-k_1 t} + \mathrm{A}_{0(2)}.e^{-k_2 t}$$

where the two processes have separate A_0 and k values.

* Show 1st phase (2 exp)

This displays the first of the two phases in isolation.

* Show 2nd phase (2 exp)

This displays the second of the two phases in isolation.

Triple Exponential Decay

For a triple exponential decay process the equation is simply

$$y = \mathrm{A}_{0(1)}.e^{-k_1 t} + \mathrm{A}_{0(2)}.e^{-k_2 t} + \mathrm{A}_{0(3)}.e^{-k_3 t}$$

Single Exponential Decay + Offset

If the observed values do not decay to 0, but to some defined offset or background value the equation for a single exponential process is

$$y = \mathrm{A}_0.e^{-kt} + \text{offset}$$

where 'offset' is the limit of the decay.

* Show Offset

Displays the offset value. Use this for all the following equations that include an offset value.

Display 1 Exp - Offset

If the data contain an offset they may be transformed to remove this offset. This display equation draws the single exponential curve through such transformed data.

Double Exponential Decay + Offset

This equation is used when a double exponential process decays to a non-zero limit value.

Display 2 Exp - Offset

If the data contain an offset they may be transformed to remove this offset. This display equation draws the double exponential curve through such transformed data.

Triple Exponential Decay + Offset

This equation is used when a triple exponential process decays to a non-zero limit value.

Display 3 Exp - Offset

If the data contain an offset they may be transformed to remove this offset. This display equation draws the triple exponential curve through such transformed data.

Transformations

Remove Offset

Data that decay to an offset value may be transformed to remove the offset. Once so transformed the plotted data may be linearized (single exponential decay only) by a semi-logarithmic plot. This transformation may be applied to any of the exponential decay equations containing an offset.

PH.GFE

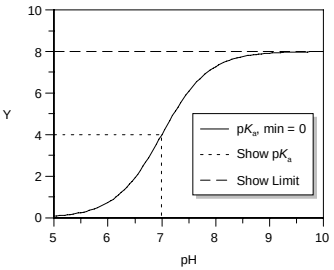
These equations are used when the observed value varies with pH. They usually involve the determination of one or more pK_a values from the experimental data.

Equations

pKa Determination, minimum = 0

This equation is used when the observed parameter varies with pH from 0 at low pH to a limiting value at high pH. Examples include various enzymic activities. The equation is:

$$y = \frac{\text{Limit}.10^{(\text{pH-p}K_a)}}{10^{(\text{pH-p}K_a)}+1}$$



* Show pKa (min=0)

This display equation is used to represent the position of the pK_a on the curve where there is a single ionization and the starting value is 0. For best results set the curve resolution to High when using this display equation (see Chapter 14).

* Show Limit (min=0)

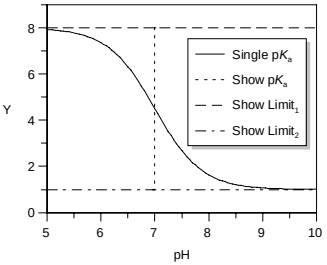
This displays the limit value of the *y* axis at high pH.

Display Linearized plot (min=0)

Displays the line for the linearized titration curve. The linearized equation is:

$$y = -\frac{K_a.y}{[\text{H}^+]} + \text{Limit}$$

pKa Determination, Single pKa



This equation is used when the observed parameter varies from one value at low pH to another at high pH, with the variation depending upon the ionization of a group. The pK_a value of this group determines the inflection point of the curve. The equation is:

$$y = \frac{\mathrm{Lim}_1 + \mathrm{Lim}_2 .10^{(\mathrm{pH}-\mathrm{p}K_a)}}{10^{(\mathrm{pH}-\mathrm{p}K_a)} + 1}$$

where Lim_1 and Lim_2 are the limits of the titration curve at low and high pH.

* Show pKa (min not 0)

This display equation is used to represent the position of the pK_a on the general ionization curve. For best results set the curve resolution to High when using this display equation (see Chapter 14).

* Show Limit 1

Displays the limiting *y* value at one extreme of pH.

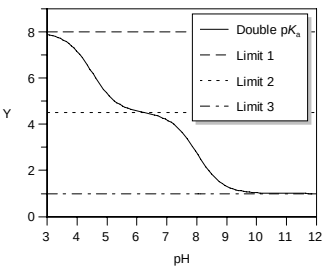
* Show Limit 2

Displays the limiting *y* value at the other extreme of pH.

Display Linearized plot (min not 0)

Displays the line for the linearized titration curve.

pKa Determination, Double pKa



This equation is used when the observed parameter varies from one value at low pH to another at high pH, with the variation depending upon the ionization of two groups. The pK_a values of these groups determine the shape of the curve.

* Show Curve 1

This shows the first of the two ionization curves in isolation.

* Show Curve 2

This shows the second of the two ionization curves in isolation.

* Show Limit 1 (double)

Displays the lower limit of the double curve.

* Show Limit 2 (double)

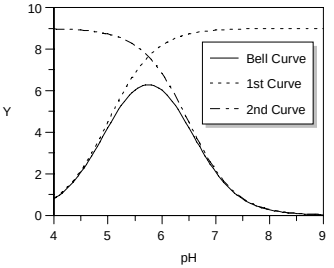
Displays the middle limit of the double curve.

* Show Limit 3 (double)

Displays the upper limit of the double curve.

Bell-shaped Double pKa curve (min = 0)

Sometimes two ionizations determine the value of the observed parameter. If the first ionization results in increasing activity and the second destroys activity this results in a bell-shaped curve.



* Display 1st Curve of Bell

This display equation is used to display the effect of the first ionization in the bell-shaped curve.

* Display 2nd Curve of Bell

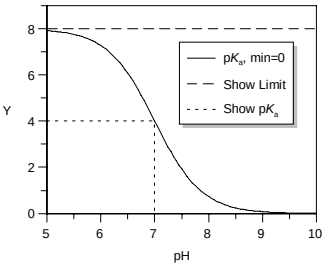
This display equation is used to display the effect of the second ionization in the bell-shaped curve.

* Display Limit of Bell

This shows the upper limit of the bell-shaped activity profile.

pKa Determination, min=0, high pH=low activity

This equation is used when the observed parameter varies with pH from a limiting value at low pH to 0 at high pH.



* Show pKa (min=0)

This display equation is used to represent the position of the pK_a on the above curve. For best results set the curve resolution to High when using this display equation (see Chapter 14).

* Show Limit (min=0)

This displays the limit value of the *y* axis at low pH.

Transformations

Linearize pKa (min=0)

This transformation linearizes a pH titration curve for display purposes.

Linearize Single pKa (min not 0)

The transformation rearranges the data for a general single titration to allow a linear plot.

POLYNOM.GFE

This definition file contains the definitions for a series of polynomial equations of the form:

$$y = a + b.x + c.x^2 + d.x^3 + ...$$

They may be used to draw a generalized curve through data where the equation is not defined. This can be useful for presentation purposes.

Equations

y = a + bx

First order polynomial (standard linear regression).

y = a + bx + cx^2

Second order polynomial.

y = a + bx + cx^2 + dx^3

Third order polynomial.

y = a + bx + cx^2 + dx^3 + ex^4

Fourth order polynomial.

y = a + bx + cx^2 + dx^3 + ex^4 + fx^5

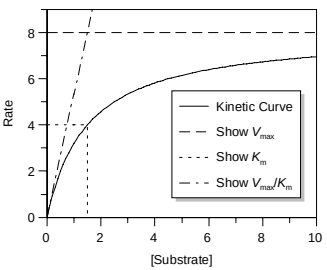
Fifth order polynomial.

ENZ_KI_E.GFE

Various equations are provided for calculating parameters from enzyme kinetic data. Two sets of these equations are provided. Definition files beginning with 'ENZ' relate to enzymes; those ending '_E', e.g. ENZ_KI_E.GFE, contain definitions in which the concentration of enzyme, $[E_0]$, is entered as a constant. From this the catalytic rate constant, k_{cat}, can be obtained. Definition files without the '_E' suffix do not require values for $[E_0]$, and instead give V_{max} values (the two are simply related, as $V_{max} = [E_0].k_{cat}$).

Equations

Enzyme Kinetics



This equation fits the basic Michaelis-Menten enzyme kinetics equation:

$$v = \frac{k_{cat}[E_0][S]}{K_m + [S]}$$

where the rate, v, varies with the concentration of substrate, [S], depending upon the catalytic constant, k_{cat}, the Michaelis

constant, K_m, and the enzyme concentration [E_0].

* Show Vmax

Displays the limiting rate, V_{max} (= k_{cat}[E_0]).

* Show Km

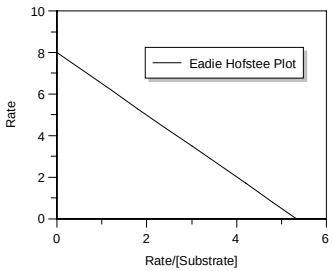
Displays the Michaelis constant, K_m. For best results set the curve resolution to High when using this display equation.

* Show Vmax/Km

Displays the linear portion at the beginning of the graph where [Substrate] ? K_m.

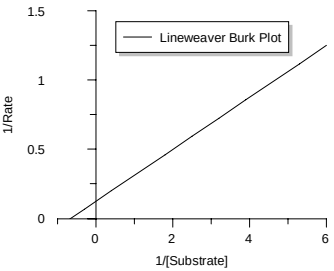
Eadie-Hofstee plot

This display equation draws the linear Eadie-Hofstee plot of enzyme kinetic data.



Lineweaver Burk plot

This display equation draws the linear Lineweaver Burk plot of enzyme kinetic data.

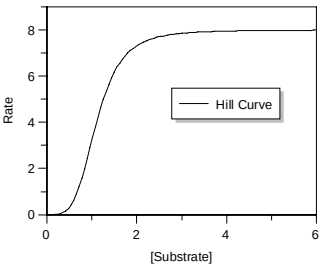


Allosteric kinetics (Hill)

The Hill equation describes enzyme kinetics for an allosteric enzyme. For equation is:

$$v = \frac{k_{cat}[E_0][S]^n}{K_m + [S]^n}$$

The Hill coefficient, *n*, equals 1 if standard kinetic behavior is followed, but differs significantly from 1 if the enzyme kinetics are cooperative.



Hill plot

This equation is used to display the linearized Hill plot where $v/(V_{max} - v)$ is plotted versus [S]. The slope is n / K_M.

Enzyme Kinetics - Contaminating S

Occasionally the reagents etc. used in an enzyme assay will contain traces of contaminating substrate. Therefore the rate will vary with added substrate, S_{add} as

$$v = \frac{k_{cat}[E_0]([S_{add}]+[S_{con}])}{K_m+[S_{add}]+[S_{con}]}$$

where $[S_{con}]$ is the concentration of contaminating substrate.

Transformations

Eadie Hofstee

Use this transformation to produce an Eadie Hofstee plot.

Lineweaver Burk

Use this transformation to give the Lineweaver Burk double reciprocal plot.

Hill plot

Produces the rearrangements required to generate the Hill plot.

ENZ_KI.GFE

The equations and transformations in this file are identical with the ones in the ENZ_KI_E.GFE file, but do not require the concentration of enzyme $[E_0]$ to be entered.

ENZ_IN_E.GFE

This file, and the accompanying ENZ_IN.GFE, contain equations describing enzyme inhibition.

Equations

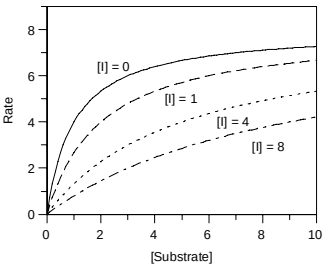
Enzyme kinetics (no inhibition)

The normal Michaelis Menten enzyme kinetics equation.

Competitive Inhibition (3D)

A multidimensional equation: rate is monitored versus [substrate] and [inhibitor]. The equation for competitive inhibition is:

$$v = \frac{[S][E_0]k_{cat}}{[S]+K_m(1+[I]/K_i)}$$



* Show Competitive inhib

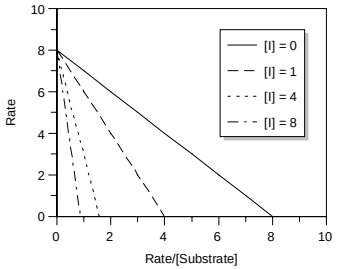
This display equation is effectively the same as that above, but the [I] is defined as a constant. This is because multidimensional equations are not allowed when plotting graphs, which are inherently two dimensional. Use this display equation to plot the curves describing the rate versus [Substrate] at various [I] values.

* Show Vmax

This display equation shows the V_{max} (= $[E_0]k_{cat}$) value.

Display Eadie (competitive 3D)

Use this to display the Eadie Hofstee plot.



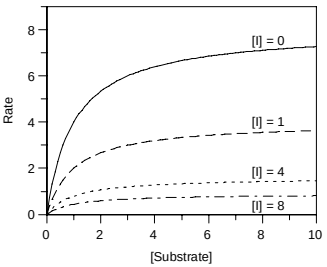
Display L-Burk (comp 3D)

Displays the Lineweaver Burk plot.

Non-competitive Inhibition (3D)

A multidimensional equation: rate is monitored versus [substrate] and [inhibitor]. If $K_{EI} = K_{ESI} = K_I$ (the usual assumption), the equation for non-competitive inhibition is:

$$v = \frac{[E_0]k_{cat}([S]/K_m)}{1+[S]/K_m+[S][I]/(K_m.K_I)}$$



* Show Non-competitive Inhibition

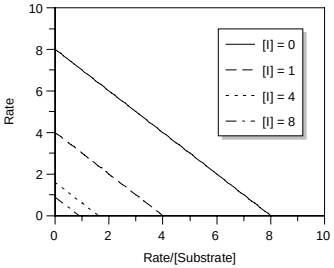
This display equation is effectively the same as the above, but the [I] is defined as a constant. This is because multidimensional equations are not allowed when plotting graphs. Use this equation to plot the curves describing the rate versus [S] at various [I] values.

* Show Vmax

Shows the Vmax value.

Display Eadie (non-competitive 3D)

Use this to display the Eadie Hofstee plot.



Display L-Burk (non-comp 3D)

Displays the Lineweaver Burk plot.

Uncompetitive Inhibition (3D)

The equation describing uncompetitive inhibition is

$$v = \frac{k_{cat}[E_0][S]/(1+[I]/K_i)}{[S]+K_m/(1+[I]/K_i)}$$

* Show Uncompetitive Inhibition

This display equation is effectively the same as the above, but the [I] is defined as a constant. This is because multidimensional equations are not allowed when plotting graphs. Use this equation to plot the curves describing the rate versus [S] at various [I] values.

* Show Vmax

Shows the V_{max} value.

Display Eadie (uncomp)

Use this to display the Eadie Hofstee plot.

Display L-Burk (uncomp)

Displays the Lineweaver Burk plot.

Full Non-comp Inhibition (3D)

If we do not assume that $K_{EI} = K_{ESI} = K_I$, the equation for non-competitive inhibition is:

$$v = \frac{[E_0]k_{cat}([S]/K_m)}{1+[S]/K_m+[I]/K_{EI}+[S][I]/(K_m.K_{EI})}$$

Transformations

Eadie Hofstee

Rearranges the data to allow an Eadie Hofstee plot.

Lineweaver Burk

Rearranges the data to allow a Lineweaver Burk plot.

ENZ_IN.GFE

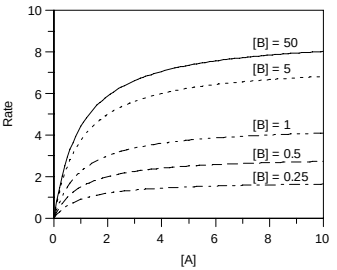
The equations and transformations in this file are identical with the ones in the ENZ_IN_E.GFE file, but do not require the concentration of enzyme [E0] to be entered.

ENZ_2SUB.GFE

The equations in this definition file are designed to analyze enzyme kinetic data involving two substrates.

Equations

2 Substrates - Ternary Complex (3D)



The equation relating the rate of reaction, *v*, to the concentration of substrates A and B and the constants K_A', K_A and K_B is:

$$v = \frac{V_{max}[A][B]}{K_A'.K_B + K_B[A] + K_A[B] + [A][B]}$$

* Show 2 Sub - Ternary Complex

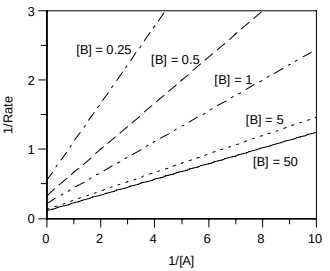
This display equation is equivalent to the equation defined above, except that [B] is defined as a constant. Use this equation to display curves at fixed [B], which is specified by the value of this constant.

Hint it is easier to alter the value of [B] after plotting the curve than to create several columns holding the [B] values. See the section on 'Changing the Curve Line Type', Chapter 14.

* Show Vmax

Shows the V_{max} value.

Display L-Burk - ternary (3D)



Displays the line for the Lineweaver Burk plot.

2 Substrates - ping pong (3D)

The equation relating the rate of reaction, *v*, to the concentration of substrates A and B and the constants K_A and K_B is:

$$v = \frac{V_{max}[A][B]}{K_B[A] + K_A[B] + [A][B]}$$

* Show 2 Sub - ping pong

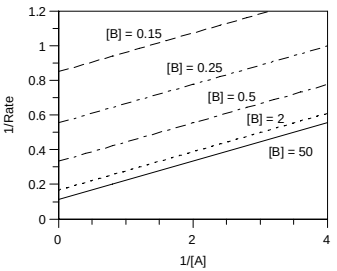
This display equation is equivalent to the equation defined above, except that [B] is defined as a constant. Use this equation to display curves at fixed [B], which is specified by the value of this constant.

* Show Vmax

Shows the V_{max} value.

Display L-B - ping pong (3D)

Displays the line for the Lineweaver Burk plot.



Transformations

Lineweaver Burk

Rearranges the data to allow the Lineweaver Burk plot.

RATE.GFE

This set of equations fits data describing a first order rate equation.

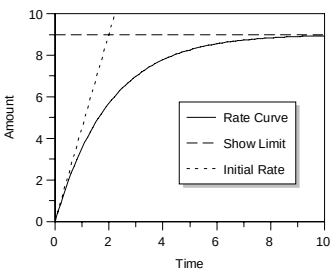
Equations

1st order rate equation

The equation for a process where the observed value increases with time is

$$A_t = A_\infty (1 - e^{-kt})$$

where A_t is the amount of A at time t, A_∞ is the maximum amount of A formed, and k is the rate constant that describes the process.



* Show limit

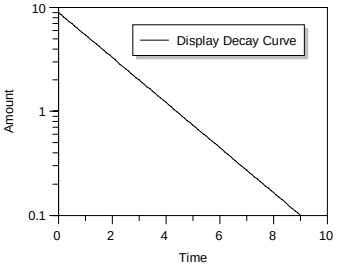
This displays the limiting value, A_∞.

* Show initial rate

The initial part of the full curve (approximately the first 10%) is linear, with the slope being $k.A_{\infty}$. This display equation shows the initial linear portion.

Display decay plot

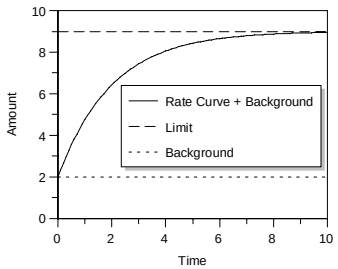
To linearize the first order rate equation, we can convert the function into an exponential decay. This displays the curve given by this rearranged equation. To linearize the curve, display the *y* axis with a logarithmic scale.



1st order rate + offset

If the observed *y* value has a background or offset at time *t*=0 (due to experimental constraints, for example a starting absorbance may not be zero) the equation becomes:

$$A_t = A_{\infty}(1 - e^{-kt}) + \text{offset}$$



* Show limit (offset)

This displays the limiting value for the equation containing an offset.

* Show Offset

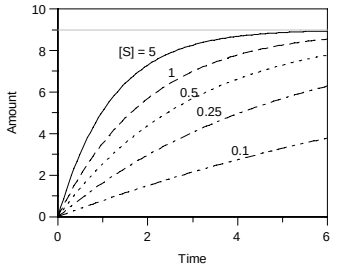
Use this display equation to show the offset value.

Display decay plot (offset)

This display equation plots the rate equation as an exponential decay. Use it with 'Decay Curve (+Offset)' transformed data.

1st order rate vs [S] (3D)

To find enzyme kinetics parameters, it is often necessary to carry out a series of first order rate reactions at different substrate concentrations. The rate constant versus [S] is then analyzed using the Michaelis Menten equation. Using this 3D equation allows these procedures to be combined, giving the 3D equation:



$$A_t = A_{\infty}\left(1 - \exp\left(-\left(\frac{V_{\max}.[S]}{K_m + [S]}\right).t\right)\right)$$

To fit this equation requires three columns: Amount, Time, [Substrate].

* Show Limit (3D)

Use this to display the limiting A_∞ value.

* Show Individual Curve (3D)

This equation is identical with the full 3D equation, but the [Substrate] has been defined as a constant. This allows this equation to be used to plot graphs, as 3D equations cannot be used for drawing curves on graphs.

1st order rate (+ back) vs [S] (3D)

This equation is the same as that above, but also includes an unknown background value, i.e. the equation is:

$$A_t = A_\infty\left(1 - \exp\left(-\left(\frac{V_{max}.[S]}{K_m + [S]}\right).t\right)\right) + \text{background}$$

* Show Limit (+ back) (3D)

Use this to display the limit value for the above equation.

* Show Individual Curve (+ back) (3D)

This equation is identical with the full 3D equation that includes a background A value, but the [Substrate] has been defined as a constant. This allows this equation to be used to plot graphs, as 3D equations cannot be used for drawing curves on graphs.

Transformations

Decay Curve

Rearranges the data to allow a plot of $A_\infty - A_t$ versus *t*. To linearize the plot rescale the *y* axis to logarithmic.

Decay Curve (+ Offset)

Rearranges the data to allow a plot of $A_\infty - A_t$ versus *t*.

LINEAR.GFE

These equations are used for normal and multiple linear regression.

Equations

y = a + bx1

Use this equation for normal linear regression.

y = a + bx1 + cx2

Multiple linear regression with two independent (*x*) axes.

y = a + bx1 + cx2 + dx3

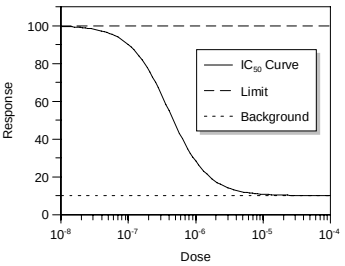
Multiple linear regression with three independent (*x*) axes.

IC50.GFE

These equations are used for analyzing radio-immuno assays or Dose-response curves.

Equations

IC50 - 4 parameter logistic



For RIA analysis or inhibition data it is usually best to fit the results using a 4 parameter logistic (see Halfman, 1981[1]). The binding or inhibition is then characterized in terms of an IC_{50} value:

$$y = \frac{a}{1+\left(\frac{x}{IC_{50}}\right)^{s}} + \text{background}$$

Here *a* is the maximum *y* range, and *s* is a slope factor. The *x* axis represents the concentration of analyte. Data fitted to this equation are usually displayed with a logarithmically scaled *x* axis. For convenience, the equation as defined attempts to provide initial estimates for the parameter values. However, as there are four unknowns this can often fail to give satisfactory results — if so turn off the automatic initial estimates (**Options Force Estimates** command) and try again, or use the "IC50 - 4 parameter logistic, no estimates" equation.

The curve shown displays inhibition with increasing dose. It is also possible to use the same equation for increasing response with dose. In this case, the *y* range is a negative value, and the "Background" is the maximum response.

* Show Limit

This displays the limiting value of the data.

* Show Background

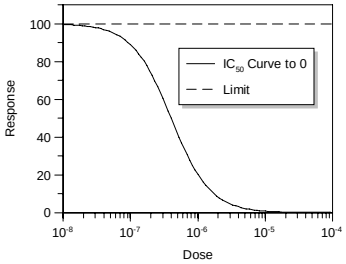
This displays the calculated background value.

---

1 Halfman, C.J. (1981) *Methods in Enymology* **74**, 481-508

IC50 - Start at 0

If the background value is known to be 0, then this equation is applicable.



* Show Limit - Start at 0

This displays the limiting value of the data.

IC50 - Start at 0, defined End

If the background value is 0, and the maximum is a defined value, then the curve is characterized by just the IC_{50} and the slope factor. In this equation the defined end value is entered as a constant.

* Show Limit - Start at 0, defined End

This displays the limiting value of the data.

IC50 - 4 parameter logistic, no estimates

As described above, the "IC50 - 4 parameter logistic" equation attempts to provide initial estimates for the parameter values. As this involves estimating four parameters, it is possible that the process will fail to produce suitable results. This equation is identical to that described above, but does not provide the initial estimates.

Appendix 1: Troubleshooting

Common Problems and Their Solutions

Program will not print

It is necessary to initialize the Windows printer driver when you first install Windows. This can be done using the Control Panel application as described in your Windows documentation.

Program will not Curve Fit

It is necessary to select an equation using the **Options Definition File...** and/or the **Options Equation...** command before fitting your data.

Program runs too slowly

Have you sufficient memory installed? We recommend a minimum of 4 Mbytes of RAM, but the performance of Windows is enhanced by increasing this to 8 Mbytes or greater.

Program prints too slowly

GraFit can produce graphs with extremely high detail. The price to be paid for this can be long printing times. If the time for printing is excessive on your system we advise reducing the resolution during printing of routine graphs, and using the highest resolution only when top quality is essential. The resolution used by the printer may be configured using the **File Print Setup...** command.

Arithmetic Errors During Data Fitting

Arithmetic errors can occur during non-linear regression if the calculations are failing to converge to a solution. This is often a result of the initial estimates being inappropriate. Try using different estimates, or, if the equation is providing these estimates automatically, use the **Options Force Estimates** command to allow you to override them.

Bugs

Although we make every effort to eliminate bugs from our code, all programs of reasonable complexity will inevitably contain various errors. If you believe that you have detected a bug, first check the following:

1. Is the effect reproducible?
2. Does the effect only occur when GraFit is running with another specific program? If so then the problem may lie with the other program.
3. Is the effect simply due to low memory? Windows can behave very oddly when memory is short, a typical effect being that menu bars fail to appear. The amount of memory free

can be tested by selecting the **Help About Program Manger...** command of the Program Manager.

If you are still convinced that you have discovered a bug, please contact Erithacus Software ***in writing*** giving the following details:

1. The version number of GraFit that you are using. This is found by selecting the **Help** menu and choosing the **About GraFit...** command.
2. The Windows version number (this is shown when Windows is first run).
3. The make and model number of your computer. Also note the amount of memory installed and whether a numeric coprocessor is present.
4. The make and model number of your printer (if the problem involves printing).
5. The printer driver installed.
6. The type and resolution of the graphics display device in use.
7. A full description of the problem. You should include sufficient detail so that we can reproduce the fault. Sample data often helps.

Appendix 2: Windows Character Set

Standard Fonts

To obtain a character from the grid, hold down the ALT key, then, using the numeric keypad, press 0 followed by the code number for the character required. For example, to produce character ™, hold down the ALT key, then press in succession the keys 0 1 5 3 on the keypad.

Code	0	1	2	3	4	5	6	7	8	9
30				!	"	#	$	%	&	'
40	(	)	*	+	,	-	.	/	0	1
50	2	3	4	5	6	7	8	9	:	;
60	<	=	>	?	@	A	B	C	D	E
70	F	G	H	I	J	K	L	M	N	O
80	P	Q	R	S	T	U	V	W	X	Y
90	Z	[	\	]	^	_	`	a	b	c
100	d	e	f	g	h	i	j	k	l	m
110	n	o	p	q	r	s	t	u	v	w
120	x	y	z	{	|	}	~	•	•	•
130	‚	*ƒ*	„	…	†	‡	ˆ	‰	Š	‹
140	Œ	•	•	•	•	‘	’	“	”	•
150	–	—	˜	™	š	›	œ	•	•	Ÿ
160		¡	¢	£	¤	¥	¦	§	¨	©
170	ª	«	¬	-	®	¯	°	±	²	³
180	´	µ	¶	·	¸	¹	º	»	¼	½
190	¾	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç
200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
220	Ü	Ý	Þ	ß	à	á	â	ã	ä	å
230	æ	ç	è	é	ê	ë	ì	í	î	ï
240	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù
250	ú	û	ü	ý	þ	ÿ				

Symbol Font

To obtain a character from the grid, hold down the ALT key, then, using the numeric keypad, press 0 followed by the code number for the character required. For example, to produce character →, hold down the ALT key, then press in succession the keys 0 1 7 4 on the keypad.

Code	0	1	2	3	4	5	6	7	8	9
30				!	∀	#	∃	%	&	϶
40	(	)	*	+	,	−	.	/	0	1
50	2	3	4	5	6	7	8	9	:	;
60	<	=	>	?	≅	Α	Β	Χ	Δ	Ε
70	Φ	Γ	Η	Ι	ϑ	Κ	Λ	Μ	Ν	Ο
80	Π	Θ	Ρ	Σ	Τ	Υ	ς	Ω	Ξ	Ψ
90	Ζ	[	∴	]	⊥	_	‾	α	β	χ
100	δ	ε	φ	γ	η	ι	φ	κ	λ	μ
110	ν	ο	π	θ	ρ	σ	τ	υ	ϖ	ω
120	ξ	ψ	ζ	{	|	}	~			
130										
140										
150										
160		ϒ	′	≤	⁄	∞	*f*	♣	♦	♥
170	♠	↔	←	↑	→	↓	°	±	″	≥
180	×	∝	∂	•	÷	≠	≡	≈	…	⏐
190	—	↵	ℵ	ℑ	ℜ	℘	⊗	⊕	∅	∩
200	∪	⊃	⊇	⊄	⊂	⊆	∈	∉	∠	∇
210	®	©	™	∏	√	·	¬	∧	∨	⇔
220	⇐	⇑	⇒	⇓	◊	〈	®	©	™	∑
230	⎛	⎜	⎝	⎡	⎢	⎣	⎧	⎨	⎩	⎪
240		〉	⌠	⌡	⎮	⎞	⎠	⎟	⎭	⎤
250	⎥	⎦	⎫	⎬	⎭					

Index

E

F

G

H

I

L

R

S

T

V

W