Preface

In the past decade microcomputers have revolutionized biomedical research. Almost every new scientific instrument is "computer controlled." Almost every researcher has a personal computer easily available or is readily linked to a mainframe. The improvements in computer software development and advances in methods of analysis have paralleled the computer hardware improvements.

It is clear that new ways of evaluating experimental data have enhanced the type of conclusions that may be drawn and have changed the way in which experiments are conceived and done. The biochemical community must be aware of new developments in data analysis and computer usage. The primary aim of this volume is to inform biomedical researchers of the modern data analysis methods that have developed concomitantly with computer hardware.

The process of collecting experimental data, analyzing the data, and then publishing the data and results is not a one-way street. All methods of data analysis make assumptions about the nature of the data. Specifically, they make assumptions about the types and magnitudes of the experimental uncertainties contained within the data. A biomedical researcher should carefully design the experimental data collection procedures such that they are compatible with the desired method of data analysis.

A common procedure used in the past for the analysis of nonlinear systems was to rearrange the equation describing the process into a linear form and then to use linear least-squares to determine slopes and intercepts related to the parameters of interest. Typical examples include the Lineweaver–Burk plot for analysis of Michaelis–Menton kinetic data and the Scatchard plot for analysis of equilibrium binding data.

Consider a Scatchard plot as an example. The objective of this, and many other types of plots, is to transform a set of experimental data into a straight line form; in this case, a plot of the amount of bound ligand divided by the free ligand (Y axis) as a function of bound ligand (X axis). For a ligand binding problem with a single class of noninteracting binding sites this transformation will provide a straight line. The next step is to "fit" a least-squares straight line to the transformed data points. The slope of this line is related to the ligand binding affinity and the X-axis intercept is the binding capacity. However, this approach makes an invalid assumption about the nature of the uncertainties contained in the experimental data. Fitting a least-squares straight line to the transformed

data assumes that the experimental uncertainties follow a random distribution and are parallel to the Y axis. However, in a Scatchard plot the uncertainties are nearly parallel to the Y axis at low fractional saturations and nearly parallel to the X axis at high fractional saturations. Consequently, the use of a least-squares method is not valid for the analysis of Scatchard plots. Note that this does not preclude the use of a Scatchard plot to help a researcher visualize an experiment if the actual data analysis is performed by another method.

So how can the data be analyzed? The best approach is to fit the original data, without any transformations, by nonlinear least-squares. For a more complete discussion of Scatchard plots refer to Klotz,[1,2] Munson and Rodbard,[3,4] and Johnson and Frasier.[5]

So why was the Scatchard plot developed? The Scatchard plot was developed in the 1940s before the availability of digital computers. Some nonlinear least-squares techniques were available at the time, i.e., the Gauss–Newton Method.[6] However, nonlinear least-squares techniques require too many operations to be performed without a computer in a reasonable length of time. At the time the Scatchard plot was the only "show in town." *Now that high-speed computers are available there is no reason to attempt to analyze transformed data.*

Almost every type of transformation "plot" to analyze experimental data was developed because high-speed digital computers were not available to perform the correct calculation. Incidently, this includes a semilog plot for the analysis of exponential decays.[5] These plots fail to meet the statistical requirements of linear least-squares methods. This failure is due to the required transformation of the data. The reason that these plots are still used for the analysis of data is primarily due to a lack of information about the available methods of data analysis. One purpose of this volume is to provide this information to biomedical researchers.

"On the other side of the coin," many biomedical researchers have learned to revere computers as oracles. They assume that if a computer analyzed their data then the results must be correct. *Computers are not oracles!* The results of any computer analysis are no better than the computer programs used for the analysis. Data analysis computer programs are created by people who make assumptions about the nature of

[1] I. M. Klotz, *Science* **217,** 1247 (1982).
[2] I. M. Klotz, *Science* **220,** 981 (1983).
[3] P. J. Munson and D. Rodbard, *Anal. Biochem.* **107,** 220 (1980).
[4] P. J. Munson and D. Rodbard, *Science* **220,** 979 (1983).
[5] M. L. Johnson and S. G. Frasier, this series, Vol. 117, p. 301.
[6] M. L. Johnson and L. M. Faunt, "Parameter estimation by least-squares methods," this volume [1].

the experimental data. They subsequently make assumptions about the best method of analysis based on their assumptions about the experimental data. These assumptions may not be acceptable for your experimental data. They also make compromises to save computer time and space in the memory of the computer. Computer programmers can also make mistakes. Thus, *computer programs sometimes include unwarranted assumptions and can make mistakes!*

Consequently, biomedical researchers cannot simply insert data into a computer and accept the results as gospel. Researchers must be aware of the assumptions used by their data analysis programs. They must be certain that they are using methods that are appropriate for their particular type of data. They need to validate their computer programs with real and synthetic data to ascertain that the computer programs are producing the results they expect. They should always question the results of a computer analysis, i.e., do they have physical meaning? The purpose of this volume is to help biomedical researchers meet these needs.

The chapters in this book are written for biomedical researchers by biomedical researchers. The volume is divided into three basic categories. First, basic methods such as nonlinear least-squares and maximum likelihood are described. Second, specific examples of the use of some of these methods are presented. The volume ends with introductory discussions about methods that are currently being developed, such as neural networks and fractals.

We are grateful to Nathan O. Kaplan who recognized the importance of data analysis in enzymology and conceived the idea for this volume.

LUDWIG BRAND
MICHAEL L. JOHNSON