Introduction

SimFit was developed out of a project to explore the use of positive rational functions in non–Michaelian steady state enzyme kinetics and also the development of a theory to define positive and negative cooperativity as a continuous function of the degree of saturation in a binding isotherm using the zeros of the Hessian of the binding polynomial.

Numerical analysis software

Stand–alone programs were written to generate data then add pseudo–random error in order to create artificial data sets to study the usefulness and power of statistical tests for parameter estimation and model discrimination with biochemical models. Optimal design for data spacing using model discrimination techniques was also investigated. These programs relied on the NAG library for subroutines to perform simulation, advancing the solutions of systems of nonlinear differential equations and constrained nonlinear optimisation.

However, in order to be Open Source and so completely free of any commercial constraints and make the programs useful to other research groups, it became necessary to provide a method for non–programmers to write their own models and it was also decided to create an Open Source replacement for some 220 NAG subroutines with the same arguments as the NAG library so that users with NAG licences could use either the NAG routines or the SimFit replacements.

Graphical software

In order to provide visual display of critical points and asymptotes in such functions a graphics package was written in the PostScript language in order to generate device–independent EPS hardcopy for direct inclusion into LaTeX documents. An advanced interface was provided to allow extensive editing of graphs permitting the export and import of templates and metafiles to allow for retrospective resumption of editing.

The Encapsulated PostScript files created are written in such a way that it is easy to edit titles, legends, plotting symbols, line types and colours retrospectively using a text editor like Notepad and also to create collages. Subsequently the graphics package was extended to transform the ASCII text PostScript code inline to display and export all Windows graphics types and eventually SVG graphics which can also be edited retrospectively using a text editor.

Curve fitting

The package contains two advanced curve fitting programs, one for arbitrary models and one for systems of nonlinear differential equations. Both offer a library of pre–compiled models but also provide for users to supply their own models written using standard mathematical notation. In practise it became clear that many users require programs dedicated to one type of model or a family of models and prefer programs that automatically calculate parameter starting estimates from the data set provided.

This is done by scaling the variables to coordinates ranging from 0 to 1 then fitting the data for small values in order to estimate slopes at the origin and fitting large values to estimate asymptotic slopes and values followed by re–calculating the parameters and estimates in user–coordinates. Programs were developed in this way to fit polynomials, exponentials, positive rational functions, ligand–binding curves, flow cytometry data and growth curves. Statistics to estimate goodness of fit, parameter confidence limits, and model discrimination are now provided as well as graphs of best fit curves.

Statistics

As the project had already made available a large number of subroutines for standard statistics it was decided to expand these into a fully–featured statistics package containing all the usual options for statistical analysis but also containing numerous additional items such as nonparametric methods, survival analysis, generalised linear modeling, power as a function of sample size, calibration analysis, data smoothing, analysis of binomial and trinomial proportions, partial least squares, false detection rates, and meta analysis.

Code to perform multivariate analysis was developed to perform correlation analysis, cluster analysis, principal components, including calculation and plotting of dendrograms, 2D and 3D biplots and K-means clusters (plotting principal components if required).

Test files

To make the interface more user–friendly a library of more than 800 data sets was created so that every time a procedure is selected an appropriate test file is supplied so that users can see the format required and execute that procedure with correctly formatted data before submitting their own data for analysis. In addition an Excel macro was created so that users can export data from spreadsheets into data files in SɪMFɪT format. Data transfer using copy and paste through the clipboard was also provided.

Results

In order to preserve the results from analysis an interface was developed so that as each program executes a text file is created containing all tables of results. Up to 100 such results files are archived but these are organised into last–in first–deleted order and an interface was created to extract tables for inclusion into Word or LATEX documents.

Documentation

By this stage there was a collection of about forty programs each with a dedicated help section available from the SimFit front page and a set of warning messages to alert users of inconsistencies in their data sets when errors were encountered. A reference manual was also created which describes every procedure available and gives the mathematical formulas for all the statistics that are used for hypothesis testing.

Finally a set of tutorials with worked examples was written to explain exactly how to use a selected SimFit program. A document containing all these tutorials collected together with an index was created.

Publications

1. Sigmoid curves, non-linear double reciprocal plots and allosterism.
   Bardsley,W.G. & Childs,R.E. (1975) *Biochem. J.* **149**, 313-328

2. An analysis of non-linear Eadie-Hofstee-Scatchard representations of ligand binding and initial-rate data for allosteric and other complex enzyme mechanisms.
   Childs,R.E. & Bardsley,W.G. (1976) *J. theor. Biol.* **63**, 1-18

3. The 3:3 function in enzyme kinetics. Possible shapes of v/S and $(1/v)/(1/S)$ plots for third degree steady-state rate equations.
   Bardsley,W.G. (1977) *J. theor. Biol.* **65**, 281-316

4. Factorability of the Hessian of the binding polynomial. The central issue concerning statistical ratios between binding constants, Hill plot slope and positive and negative co-operativity.
   Bardsley,W.G. & Waight,R.D. (1978) *J. theor. Biol.* **72**, 321-372

5. Concerning the thermodynamic definition and graphical manifestations of positive and negative co-operativity.
   Bardsley,W.G. & Wyman,J. (1978) *J. theor. Biol.* **72**, 373-376

6. An analogue of the binding polynomial for the case of ligands binding to an aggregating macromolecule.
   Woolfson,R. & Bardsley,W.G. (1980) *J. Mol. biol.* **136**, 451-454

7. Relationships between the magnitude of Hill plot slopes, apparent binding constants and factorability of binding polynomials and their Hessians.
   Bardsley,W.G., Woolfson,R. & Wood,R.M.W. (1980) *J. theor. Biol.* **85**, 247-284

8. A new approach to the measurement of sigmoid curves with enzyme kinetic and ligand binding data.
   Bardsley,W.G. & Wright,A.J. (1983) *J. Mol. biol.* **165**, 163-182

9. Use of the F test for determining the degree of enzyme-kinetic and ligand-binding data. A Monte Carlo simulation study.
   Burguillo,F.J., Wright,A.J. & Bardsley,W.G. (1983) *Biochem. J.* **211**, 23-34

10. Inhibition of enzymes by excess substrate. A theoretical and Monte Carlo study of turning points in $v(S)$ graphs.
Bardsley,W.G., Solano-Munoz,F., Wright,A.J. & McGinlay,P.B. (1983) *J. Mol. Biol.* **169**, 597-617

11. Critical points and sigmoidicity of positive rational functions.
Wood,R.M.W. & Bardsley,W.G. (1985) *Amer. Math. Month.* **92(1)**, 37-48

12. The F test for model discrimination with exponential functions.
Bardsley,W.G., McGinlay,P.B. & Wright,A.J. (1986) *Biometrika* **73**, 501-508

13. Optimal design for model discrimination using the F test with non-linear biochemical models. Criteria for choosing the number and spacing of experimental points.
Bardsley,W.G., McGinlay,P.B & Roig,M.G. (1989) *J. theor. Biol.* **139**, 85-102

14. A statistical theory for the interpretation of altered flow cytometry profiles in terms of the binding of ligands to cell surface receptors and changes in gene expression.
Bardsley,W.G. & Kyprianou,E.K. (1995) *J. Math. Biol.* **34**, 271–296

15. A computer program to study the optimal design for model discrimination with life science models.
Bardsley,W.G. & Melikhova,E.M. *Computers and Chemistry* (1996) **20** 145–157

16. Using ASCII text files in post-fix notation (reverse Polish) to define mathematical models and systems of differential equations for simulation and nonlinear regression.
Bardsley,W.G. & Prasad,N *Computers and Chemistry* (1997) **21**, 71–82

17. Contributions of individual molecular species to the Hill coefficient for ligand binding by an oligomeric protein.
Edelstein,S.J & Bardsley,W.G. *J. Mol. Biol* (1997) **267**, 10–16