Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

When accurate initial rate data are obtained over an extended range of substrate concentration then deviations from Michaelis-Menten kinetics such as substrate inhibition, substrate activation, sigmoidicity, and other types of cooperativity are often encountered. If artifacts such as aggregation, failing to correct for pH or ionic strength, etc. are not responsible, then the appropriate model would be positive rational function. The most important thing is to determine the degree of such a rational function, because this can be used to select a possible kinetic scheme and rate equation.

From the main SimFit menu select [A/Z], then open program **rffit** and study the default test file `rffit.tf6` which has the following data set.

S	$v(S)$	$se(v)$
0.01000	0.02100	0.00157
0.01624	0.03409	0.00256
0.02637	0.05533	0.00415
0.04281	0.08975	0.00673
0.06952	0.14532	0.01090
0.11288	0.23421	0.01757
0.18330	0.37301	0.02798
0.29764	0.57659	0.04324
0.48329	0.83187	0.06239
0.78476	1.04968	0.07873
1.27427	1.09824	0.08237
2.06914	0.98829	0.07412
3.35982	0.87359	0.06552
5.45559	0.86246	0.06468
8.85867	0.96350	0.07226
14.38450	1.08641	0.08148
23.35721	1.07355	0.08052
37.92690	0.87789	0.06584
61.58482	0.62001	0.04650
100.00000	0.40461	0.03035

The columns contain data in the following format.

1. **Column 1**: the non–negative substrate concentration S which must be in non-decreasing order.

2. **Column 2**: the non–negative initial rate $v(S)$ presumed to be dependent on substrate in column 1.

3. **Column 3**: the positive sample standard deviation of the replicate rate measurements.
   This column can be omitted or set to 1 if unweighted regression is required.

SimFit program **rffit** fits positive rational functions of the following form

$$v(S) = \frac{\alpha_0 + \alpha_1 S + \alpha_2 S^2 + \cdots + \alpha_n S^n}{\beta_0 + \beta_1 S + \beta_2 S^2 + \cdots + \beta_n S^n}$$

which will be referred to as a *n:n* function. This has $2n+1$ independent nonnegative parameters $\alpha_i \geq 0$, $\beta_i \geq 0$ so we define $\beta_0 = 1$. In addition, it is is usually the case that $\alpha_0 = 0$ so that $v(0) = 0$ and, if dead-end enzyme-substrate complexes are assumed, it will also be convenient to set $\alpha_n = 0$ to model the case where $v(\infty) = 0$.

Some observations on curve fitting n:n functions

Program **rffit** scales the data provided into internal coordinates and, by default, sets all parameters to one or to user-supplied values. For very difficult problems with widely spaced parameters, the program can estimate approximate starting parameters from the extremes of the data set, undertake a random search in an attempt to improve these, then perform a preliminary fit in the constrained L_1 to refine the starting estimates. It scales the model fitted so that the internal parameters will be of order unity at the start of the optimization, then performs weighted nonlinear regression by the quasi-Newton method with parameters constrained to be non-negative. Even so, it must be realized that distinguishing order 2 from order 1 will usually be straightforward, distinguishing order 3 from order 2 will demand very accurate data over a wide range, and distinguishing order 4 from order 3 will usually only succeed if the data have special features.

The test files

Many of the test files distributed with the SimFIT package contain exact data. There are two reasons for this.

1. Exact data sets generated by program **makdat** have known parameters so, by fitting these it is possible to observe how accurately SimFIT can estimate the known parameters.

2. Such exact data sets can be corrupted by adding random error using program **adderr** to simulate experimental data.

The recommended procedure for analyzing your own data is to simulate using the suspected model and anticipated error, then fit to see how likely it will be for the fitting program to support your working hypothesis.

Example 1

In particular `rffit.tf6` contains exact data for a 4:4 rational function, and fitting 3:3 then 4:4 yields these parameter estimates.

Number	Parameter	Value	Std. Error	Lower95%cl	Upper95%cl	p	
1	α_0	1.0982E-05	6.0976E-06	-2.4385E-06	2.4403E-05	0.0991	*
2	α_1	2.0997E+00	3.8507E-04	2.0988E+00	2.1005E+00	0.0000	
3	α_2	2.4562E-04	1.4423E-03	-2.9289E-03	3.4201E-03	0.8679	*
4	α_3	1.0493E-01	7.4973E-05	1.0476E-01	1.0509E-01	0.0000	
5	α_4	1.0982E-13	9.3838E-08	-2.0654E-07	2.0654E-07	1.0000	*
6	β_1	4.0251E-04	7.3125E-04	-1.2070E-03	2.0120E-03	0.5930	*
7	β_2	1.0022E+00	1.5814E-03	9.9868E-01	1.0056E+00	0.0000	
8	β_3	1.0000E-11	1.2096E-05	-2.6623E-05	2.6623E-05	1.0000	*
9	β_4	2.4982E-03	1.9153E-06	2.4940E-03	2.5024E-03	0.0000	

Now this test data was generated as the sum of two dead-end substrate inhibition models namely

$$
\begin{aligned}
v(S) &= \frac{2S}{1 + S^2} + \frac{0.1S}{1 + 0.0025S^2} \\
&= \frac{2.1S + 0.105S^3}{1 + 1.0025S^2 + 0.0025S^4}
\end{aligned}
$$

so, pointing out that numbers less than about 10^{-4} of the largest parameter estimate are effectively zero, the estimates are almost identical to the values used to generate the data, which supports the fact that program **rffit** is capable of determining the parameters of a 4:4 function given highly accurate and extensive data. The next table illustrates the strong support for the extra parameters in the 4:4 fit compared to the 3:3 fit, while the graph illustrates the reason: a positive rational function of order n:n can have at most $n - 1$ positive turning points so the 3:3 function can never model the three positive turning points of the 4:4 function.

2

WSSQ-previous	7.8928E+01
WSSQ-current	7.2902E-05
Number of parameters-previous	7
Number of parameters-current	9
Number of *S*-values	20
Akaike *AIC*-previous	4.1456E+01
Akaike *AIC*-current	-2.3244E+02
Schwarz *SC*-previous	4.8426E+01
Schwarz *SC*-current	-2.2348E+02
Mallows C_p	1.1909E+07
Numerator degrees of freedom	2
Denominator degrees of freedom	11
F test statistic (*FS*)	5.9546E+06
$P(F \geq FS)$	0.0000
$P(F \leq FS)$	1.0000
5% upper tail point	3.9823E+00
1% upper tail point	7.2057E+00

$ER = 2.9947E+59$

$C_p/M_1 = 1.7013E+06$

Conclusion based on *F* test
Reject previous model at 1% significance level
There is strong support for the extra parameters
Tentatively accept the current best fit model

Best Fit 3:3 and 4:4 Functions
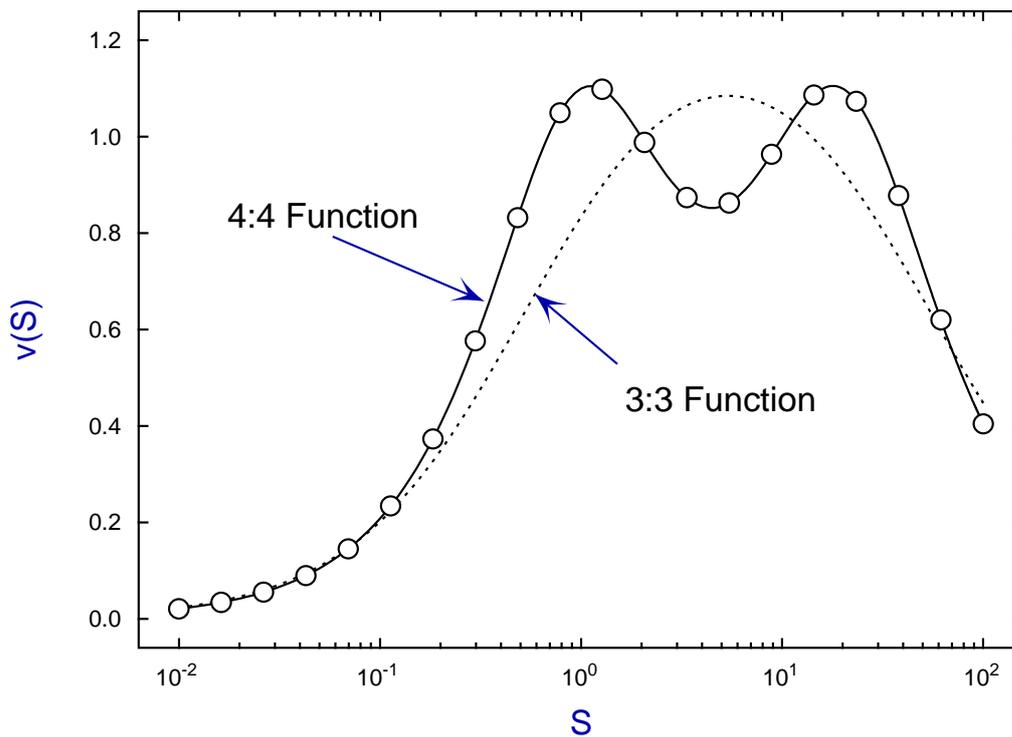
Note that the data are in a geometrical progression, that is, for k points x_i distributed equally on a log scale between end points A and B we would have $x_1 = A, x_2 = \lambda x_1, x_3 = \lambda x_2, \ldots, x_k = \lambda x_{k-1} = \lambda^{k-1} x_1 = B$

where $\lambda = (B/A)^{1/(k-1)}$. When fitting models like saturation functions, exponentials, or rational functions, a logarithmic scale is optimal for model discrimination and parameter determination [1].

Example 2

Test file `rffit.tf7` was obtained from `rffit.tf6` using **adderr** to generate triplicates with 7.5% relative error and standard errors calculated from replicates. Using starting estimates equal equal to one, but with α_0 and α_n suppressed gives reasonable parameter estimates as shown by the next table and graph.

Number	Parameter	Value	Std. Error	Lower95%cl	Upper95%cl	p	
2	α_1	2.1330E+00	2.7697E-02	2.0775E+00	2.1886E+00	0.0000	
3	α_2	2.1863E-01	3.1074E-01	-4.0464E-01	8.4189E-01	0.4848	*
4	α_3	1.0866E-01	2.0534E-02	6.7470E-02	1.4984E-01	0.0000	
6	β_1	0.0000E+00	0.0000E+00	0.0000E+00	0.0000E+00	1.0000	f
7	β_2	1.2320E+00	3.3648E-01	5.5711E-01	1.9069E+00	0.0006	
8	β_3	1.1784E-03	1.2893E-02	-2.4682E-02	2.7039E-02	0.9275	*
9	β_4	2.5835E-03	3.5474E-04	1.8719E-03	3.2950E-03	0.0000	

Best Fit 3:3 and 4:4 Functions
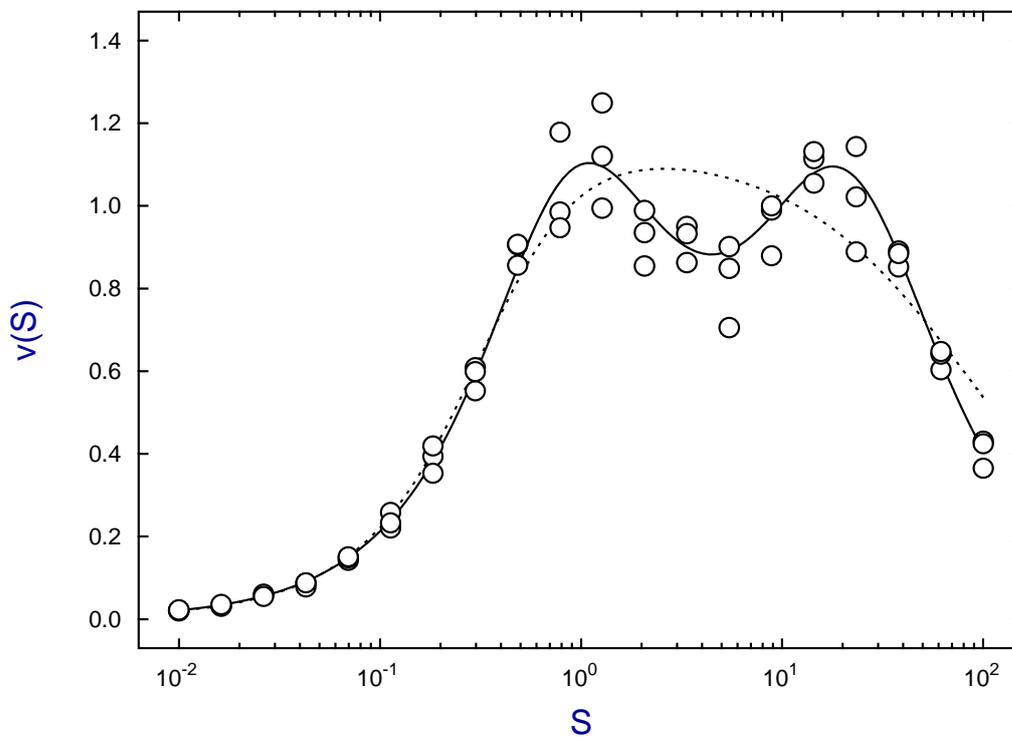
In general for *n:n* functions with $n > 2$ the following conclusions should be noted.

1. The data must be very accurate and span a large range of substrate concentration.

2. Data should be collected using a geometric progression between end points.

3. Program **rffit** must be run many times to compare the fit with random starting estimates with the fit from user-supplied starting estimates, or with all equal to one as with this example fitting `rffit.tf7`.

4. Model discrimination may work for $n \le 3$ but parameter estimates may be poor for $n \ge 3$.

Theory

All kinetic schemes that can be devised for non-aggregating enzymes that have zero rate at zero substrate concentration lead to the following quasi steady state rate equation

$$v(S) = \frac{\alpha_1 S + \alpha_2 S^2 + \cdots + \alpha_n S^n}{\beta_0 + \beta_1 S + \beta_2 S^2 + \cdots + \beta_n S^n}$$

where $\alpha_i \geq 0$ and $\beta_i \geq 0$, and the only difference between models is the way that the coefficients α_i, β_i are expressed as functions of the rate constants. Some simple facts about this function are as follows.

- The order n can be as large as the subscript k in any enzyme substrate complexes ES_k.

- Given any order n with any nonnegative values for α_i, β_i it is possible to define a possible enzyme scheme with these values [2].

- The $v(S)$ curve is only sigmoidal if $\alpha_2 \beta_0 > \alpha_1 \beta_1$ [3].

- A order n positive rational function can have at most $n - 1$ turning points and $2n - 2$ points of inflexion in the first quadrant [4].

- If $\alpha_n = 0$ the $v(S)$ curve descends asymptotically to zero from a final turning point.

- If $\alpha_n > 0$ the $v(S)$ curve approaches a horizontal asymptote $v(\infty) = \alpha_n/\beta_n$.

- If $\alpha_n > 0$ the $v(S)$ curve descends from a final turning point if $\alpha_n \beta_{n-1} < \alpha_{n-1} \beta_n$.

As the main idea in fitting such a rational function to experimental data is to fix a minimum order n to suggest a possible enzyme mechanism, some facts about parameter estimation and discrimination should be noted.

- Any attempt to fit higher order models requires accurate data over a large range.

- Experimental points should have the S values in a geometric progression.

- Statistical techniques for model discrimination perform very well when distinguishing the case $n = 2$ from $n = 1$ but rapidly deteriorate for $n > 2$ [5],[6],[7].

- Fitting cases with $n > 2$ should be investigated by comparing the results with all starting estimates equal to one with those from several random searches.

- When the order n has been decided and good starting estimates have been located it is time to use the more advanced program **qnfit**.

Limiting cases

Before nonlinear regression became accepted as the only meaningful way to fit enzyme kinetic data, much use was made of fitting straight lines to extreme substrate concentrations in order to estimate parameter values. This has given rise to much confusion.

It might be thought that a satisfactory approximation to a $n{:}n$ function at low substrate concentration would be the 1:1 Michaelis-Menten function

$$v(S) = \frac{\alpha_1 S}{\beta_0 + \beta_1 S}.$$

However this cannot be sigmoidal and $v(S)$ curves can be sigmoidal, so the best 1:1 function to approximate low substrate values is

$$v(S) = \frac{\alpha_1^2}{\alpha_1 \beta_0 + (\alpha_1 \beta_1 - \alpha_2 \beta_0)S},$$

as this is the equation of the asymptotes approached at low substrate concentration in all the transformed plots such as double-reciprocal, Scatchard, etc. used in enzyme kinetics [8].

Irrespective of which technique is used to fit the data for low S the only parameters that can be estimated are the apparent kinetic constants

$$V_{max} = \alpha_1^2/(\alpha_1\beta_1 - \alpha_2\beta_0), \text{ and } K_m = \alpha_1\beta_0/(\alpha_1\beta_1 - \alpha_2\beta_0).$$

Again, it might be thought that a satisfactory approximation to a $n{:}n$ function at high substrate concentration would be the n:n Hill equation

$$v(S) = \frac{\alpha_n S^n}{\beta_0 + \beta_n S^n}.$$

However $v(S)$ curves can have turning points and the Hill equation is monotonic, so the best 1:1 function to approximate high substrate values is

$$v(S) = \frac{\alpha_n^2}{(\alpha_n\beta_{n-1} - \alpha_{n-1}\beta_n) + \alpha_n\beta_n S},$$

as this is the equation of the asymptotes approached at high substrate concentration in all the transformed plots such as double-reciprocal, Scatchard, etc. used in enzyme kinetics [8].

Irrespective of which technique is used to fit the data for high S the only parameters that can be estimated are the apparent kinetic constants

$$V_{max} = \alpha_n/\beta_n, \text{ and } K_m = (\alpha_n\beta_{n-1} - \alpha_{n-1}\beta_n)/(\alpha_n\beta_n).$$

Another issue concerns reduction in degree by cancelation of common factors between numerator and denominator which can occur with some enzyme mechanisms containing cycles which, after using the principle of microscopic reversibility, lead to zero values for one or more of the Sylvester eliminants [9]. In the extreme case of factorization down to order 1:1, one possible expression for the reduced equation would be

$$v(S) = \frac{\alpha_1\alpha_n}{\alpha_n\beta_0 + \alpha_1\beta_n S}.$$

To the extent that fitting a Michaelis-Menten equation to a $n{:}n$ function is justified because statistical evidence does not support $n > 1$, the two parameters that can be estimated are the apparent kinetic constants

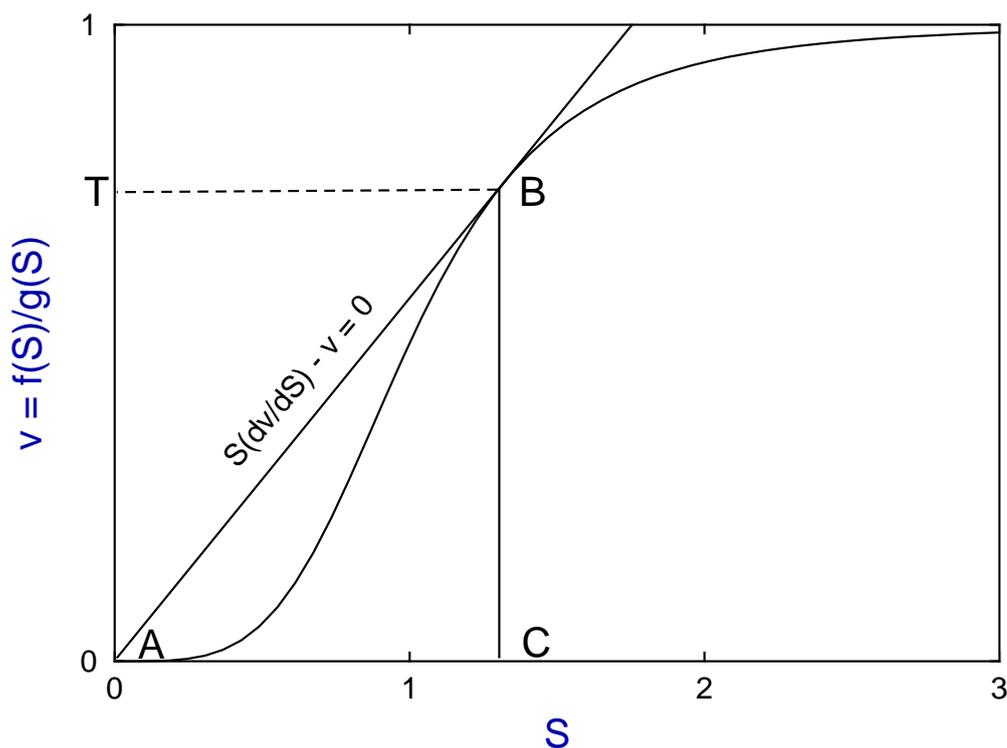$$V_{max} = \alpha_n/\beta_n, \text{ and } K_m = \alpha_n\beta_0/(\alpha_1\beta_1).$$

Sigmoidicity

A positive rational function will be sigmoid if

$$\alpha_2\beta_0^2 - \alpha_1\beta_0\beta_1 - \alpha_0\beta_0\beta_2 + \alpha_0\beta_1^2 > 0,$$

but in the usual case $\alpha_0 = 0$ it is possible to define satisfactory measures of sigmoidicity, which can be explained by reference to the next figure [10].

6

Sigmoidicity of Positive Rational Functions

The point labeled C is the first positive root of

$$S(dv/dS) - v = 0$$

and the point labeled $T = v(C)$ is the v coordinate where the tangent from the origin touches the curve. Consider then the expressions

$$\Delta_1 = \frac{T}{\max(v),\ \text{for}\ S \geq 0}$$

$$\Delta_2 = \frac{Area(ABC)}{\int_0^C v(S)\ dS}$$

where Δ_1 and Δ_2 both increase as sigmoidicity increases. It can be shown that, for fractional saturation functions of order n, the following inequality applies
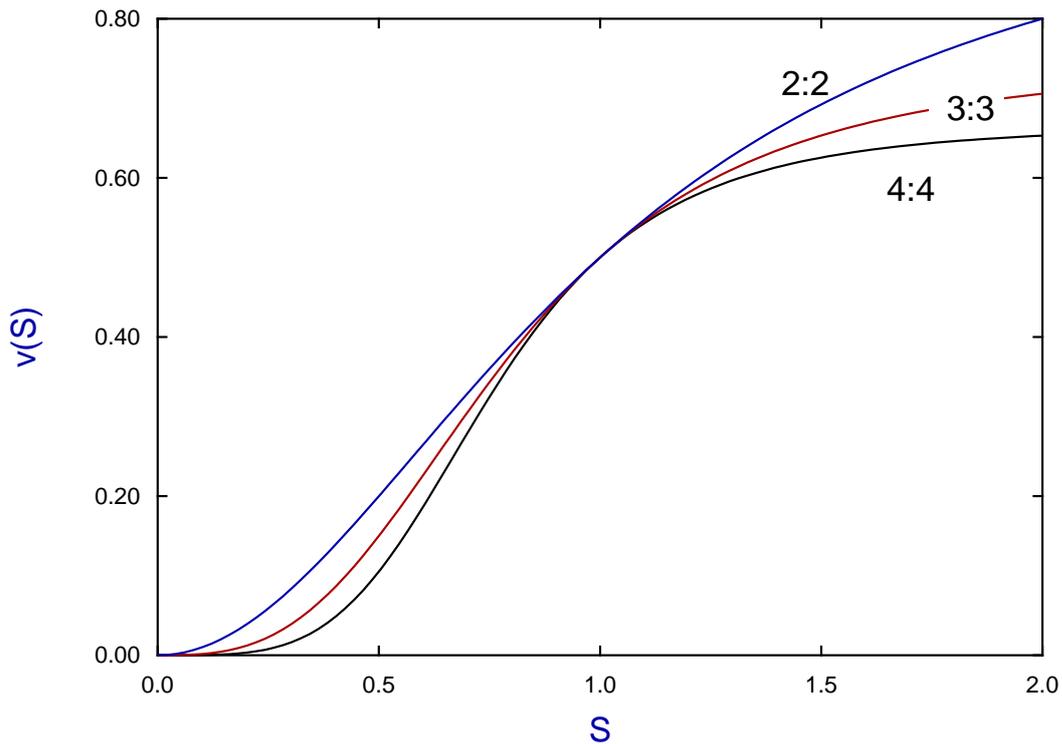
$$T \leq \frac{n-1}{n},$$

while, at least for the cases $n = 2, 3$, and $n = 4$ with some qualification, the positive rational function curve of maximum sigmoidicity is the equation

$$v = \frac{S^n}{1 + S^n}$$

shown in the next graph for the normalized functions $x^2/(1 + x^2)$, $(3/2)x^3/(1 + 2x^3)$, and $2x^4/(1 + 3x^4)$.

7

v(S) Curves with Maximum Sigoidicity

Substrate inhibition

Characterization of the shapes of rational functions in the first quadrant has been achieved for $n \leq 3$, and the most distinctive feature is the number of turning points [11]. All that can be said is that analysis depends on the signs and magnitudes of determinants such as

$$D_i = \begin{vmatrix} \alpha_i & \alpha_{i-1} \\ \beta_i & \beta_{i-1} \end{vmatrix}$$

leading to the necessary and sufficient condition $D_n < 0$ for a final turning point, but only the necessary but not sufficient condition $D_i D_{i-1} < 0$ for the maximum number.

It is possible to establish a criterion for the maximum steepness of descent from a final maximum, which is the most commonly encountered type of substrate inhibition, and to present maximal examples for low degree cases. Unfortunately this is, like sigmoidicity, only possible for $n < 4$ and, with some reservations for $n = 4$. That is because, for $n > 4$ the number of possible curve shapes due to multiple turning points creates an intractable situation.

At least for the cases $n = 2, 3$ and $n = 4$ with some qualification, the positive rational function curve showing maximum substrate inhibition is the equation

$$v = \frac{S}{1 + S^n}$$

shown in the next graph for the normalized functions $150x/(1 + x^2)$, $225x/(2 + x^3)$, and $300x/(3 + x^4)$ [12].

8

v(S) Curves with Maximum Substrate Inhibition
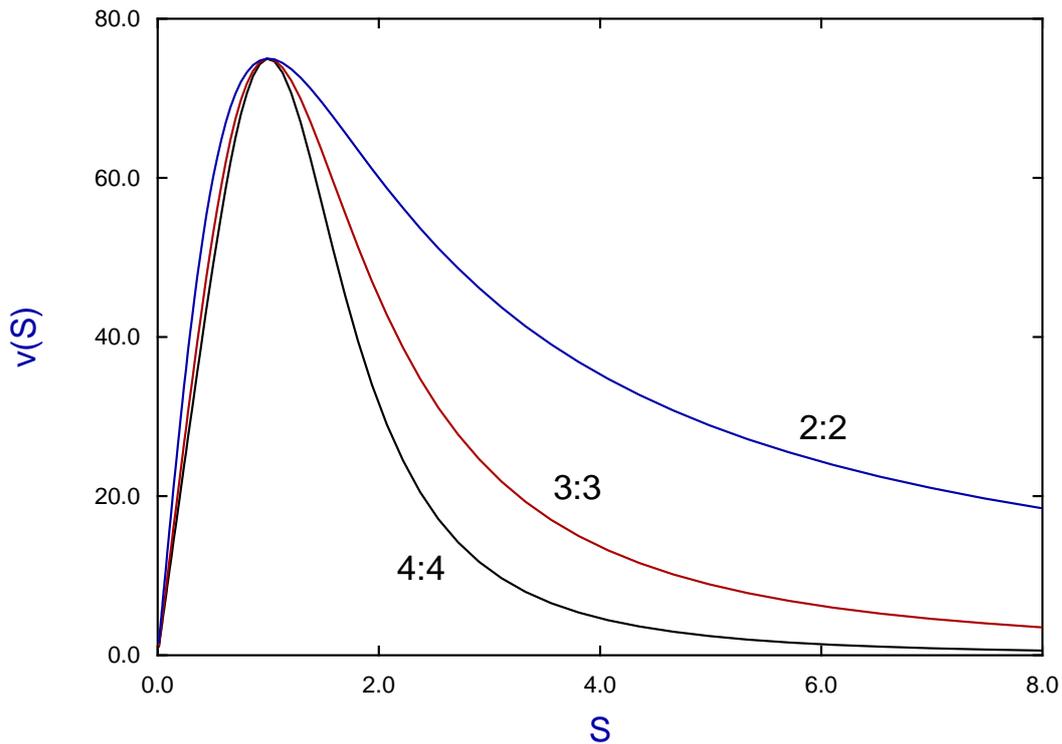
References

[1] 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

[2] Simple enzyme kinetic mechanisms that can give all possible velocity profiles with chemically reasonable rate constant values.
Bardsley,W.G. (1983) *J. theor. Biol.* **104**, 485-491

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

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

[5] The use of non-linear regression analysis and the F test for model discrimination with dose response curves and ligand binding data.
Bardsley,W.G. & McGinlay,P.B. (1987) *J. theor. Biol.* **126**, 183-201

[6] Conditions when statistical tests for model discrimination have high power. Some examples from pharmacokinetics, ligand binding, transient and steady-state enzyme kinetics.
Bardsley,W.G., McGinlay,P.B. & Roig,M.G. (1987) *Biophys. Chem.* **26**, 1-8

[7] 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

[8] The determination of positive and negative co-operativity with allosteric enzymes and the interpretation of sigmoid curves and non-linear double reciprocal plots for the MWC and KNF models.
Bardsley,W.G. & Waight,R.D. (1978) *J. theor. Biol.* **70**, 135-156

[9] The reduction in degree of allosteric and other complex rate equations using Sylvester's dialytic method of elimination.
Bardsley,W.G. (1977) *J. theor. Biol.* **67**, 121-139

[10] 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

[11] The probability of obtaining complex kinetic curves for enzyme mechanisms with cubic terms in the pseudo-steady state rate equations.
Wardell,J.M., Bardsley,W.G., Kavanagh,J.P. & Wood,R.M.W. (1982) *J. theor. Biol.* **95**, 465-487

[12] 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