



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

Michaelis-Menten models are used to analyze quasi steady state data when the enzymes concerned do not exhibit substrate inhibition, substrate activation, or any other complicating features such as allosterism. The one site version can be extended to cover the case of an enzyme with several kinetically differing but independent sites, or mixtures of isoenzymes.

Example 1: Substrate varied mode

From the main SIMFIT menu select [A/Z], open program **mmfit**, select the substrate-varied option, and view the default test file `mmfit.tf4` which has the following data.

<i>S</i>	<i>v</i>	<i>se(v)</i>
0.21759	0.20273	0.0054324
0.21759	0.20050	0.0054324
0.21759	0.19241	0.0054324
0.39440	0.31925	0.015018
0.39440	0.34123	0.015018
0.39440	0.31252	0.015018
0.71490	0.50336	0.011163
0.71490	0.48104	0.011163
0.71490	0.49241	0.011163
1.2958	0.67103	0.018464
1.2958	0.70535	0.018464
1.2958	0.67639	0.018464
2.3488	0.90847	0.015994
2.3488	0.93885	0.015994
2.3488	0.91501	0.015994
4.2575	1.1107	0.021537
4.2575	1.1439	0.021537
4.2575	1.1035	0.021537
7.7172	1.3639	0.048544
7.7172	1.2947	0.048544
7.7172	1.3882	0.048544
13.988	1.6565	0.042217
13.988	1.5894	0.042217
13.988	1.5785	0.042217
25.355	1.6468	0.078963
25.355	1.7954	0.078963
25.355	1.6748	0.078963
45.959	1.8712	0.029314
45.959	1.8568	0.029314
45.959	1.8148	0.029314

The columns contain data in the following format.

1. **Column 1:** *S*, the non-negative substrate concentration which must be in non-decreasing order.
2. **Column 2:** *v*, the non-negative initial rate measured for the concentration in column 1.
3. **Column 3:** *se*, the positive sample standard deviation of the replicate rate measurements. This column can be omitted or set to 1 if unweighted regression is required.

To illustrate the functionality of the `SimFIT` program `mmfit` we shall fit a one site model followed by a two site model (or mixture of two isoenzymes) and see if any improvement in fit can be supported by statistical analysis. The two models are as follows.

$$f_1(S) = \frac{V_{max}S}{K_m + S}$$

$$f_2(S) = \frac{V_{max_1}S}{K_{m_1} + S} + \frac{V_{max_2}S}{K_{m_2} + S}$$

To fit these two models choose to start fitting at order 1 and end fitting at order 2, using the further default settings, to obtain the following results tables.

Table 1: For best-fit 1:1 Michaelis-Menten function f_1

Number	Parameter	Value	Std. Error	Lower95%cl	Upper95%cl	<i>p</i>
1	V_{max}	1.7861	0.040866	1.7024	1.8698	0.0000
2	K_m	1.9734	0.097463	1.7738	2.1731	0.0000

Parameter correlation matrix for model f_1

1	
0.8185	1

Table 2: For best-fit 2:2 Michaelis-Menten function f_2

Number	Parameter	Value	Std. Error	Lower95%cl	Upper95%cl	<i>p</i>
1	V_{max_1}	1.0254	0.10377	0.81211	1.2387	0.0000
2	V_{max_2}	1.0290	0.13352	0.75455	1.3035	0.0000
3	K_{m_1}	9.7460	2.8652	3.8566	15.636	0.0022
4	K_{m_2}	1.0433	0.11698	0.80283	1.2837	0.0000

Predicted maximum rate (i.e. apparent V_{max}) = 2.0544

Predicted half saturation point (i.e. apparent K_m) = 3.1811

Parameter correlation matrix for model f_2

1			
-0.9568	1		
-0.8573	0.9638	1	
-0.9631	0.9810	0.9088	1

In order to determine if a significant improvement in fit has resulted we need to consider the following questions.

1. Are the parameters well-determined with both fits ?
2. Does the residuals analysis indicate satisfactory fits ?
3. Does the F test for excess variance support model f_2 in preference to f_1 ?
4. Can the best-fit curves be seen to differ when plotted against the data ?
5. Does the graphical deconvolution display convincing evidence that both components of f_2 are contributing to the overall fit ?

The results displayed in Tables 1 and 2 show that both models fit well with parameters that differ significantly from zero. Table 3 indicates that an excellent fit has resulted for model f_2 and Table 4 supports the conclusion that there is statistical evidence that model f_2 should be accepted as explaining the data better than model f_1 . This is then further emphasized by the graphical displays showing the data with best-fit curves for f_1 and f_2 , and the deconvolution of the f_2 fit into the two contributing components.

Table 3: Goodness of fit for model f_2

Analysis of residuals: $WSSQ$	28.293
$P(\chi^2 \geq WSSQ)$	0.3442
$R^2, cc(theory, data)^2$	0.9963
Largest Absolute relative residual	5.88%
Smallest Absolute relative residual	0.25%
Average Absolute relative residual	2.28%
Absolute relative residuals in range 0.1-0.2	0.00%
Absolute relative residuals in range 0.2-0.4	0.00%
Absolute relative residuals in range 0.4-0.8	0.00%
Absolute relative residuals > 0.8	0.00%
Number of negative residuals (m)	13
Number of positive residuals (n)	17
Number of runs observed (r)	22
$P(\text{runs} \leq r : \text{given } m \text{ and } n)$	0.9957
5% lower tail point	10
1% lower tail point	9
$P(\text{runs} \leq r : \text{given } m \text{ plus } n)$	0.9959
$P(\text{signs} \leq \text{least number observed})$	0.5847
Durbin-Watson test statistic	2.5008 >2.5, -ve serial correlation?
Shapiro-Wilks W statistic	0.9678
Significance level of W	0.4806
Akaike AIC (Schwarz SC) stats	6.2425 (11.847)
Verdict on goodness of fit: incredible	

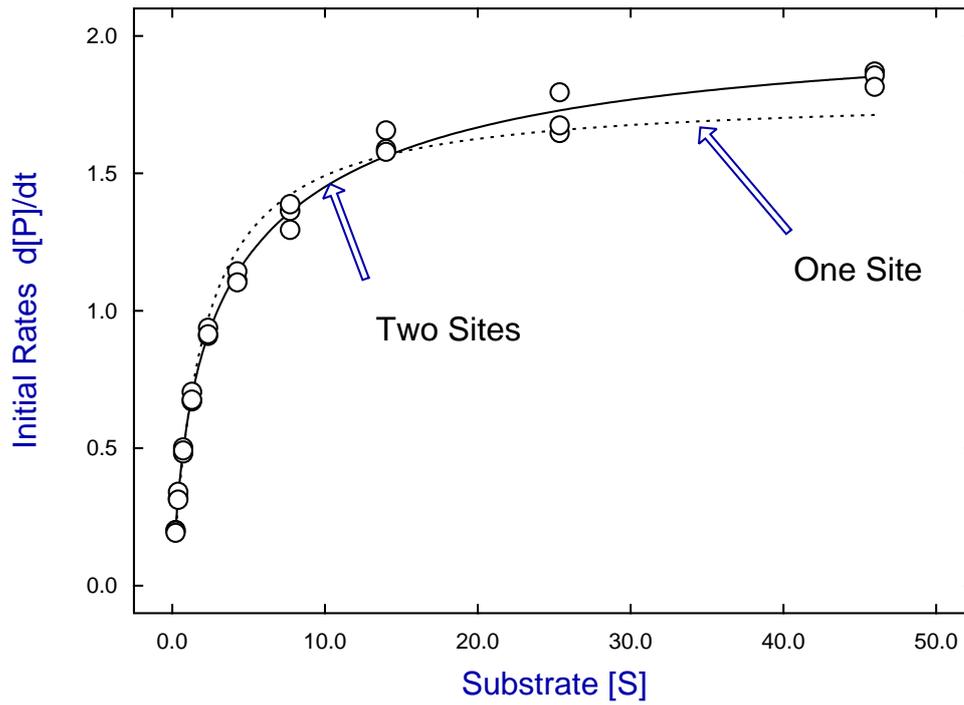
Table 4: F test results for model f_2 against f_1

$WSSQ$ previous	257.18
$WSSQ$ current	28.293
Number of parameters previous	2
Number of parameters current	4
Number of x values	30
Akaike AIC previous	68.458
Akaike AIC current	6.2425, ER = 3.2346E+13
Schwarz SC previous	71.260
Schwarz SC current	11.847
Mallows' C_p	210.34, $C_p/2 = 105.17$
Numerator degrees of freedom	2
Denominator degrees of freedom	26
F test statistic (FS)	105.17
$P(F \geq FS)$	0.0000
$P(F \leq FS)$	1.0000
5% upper tail point	3.3690
1% upper tail point	5.5263

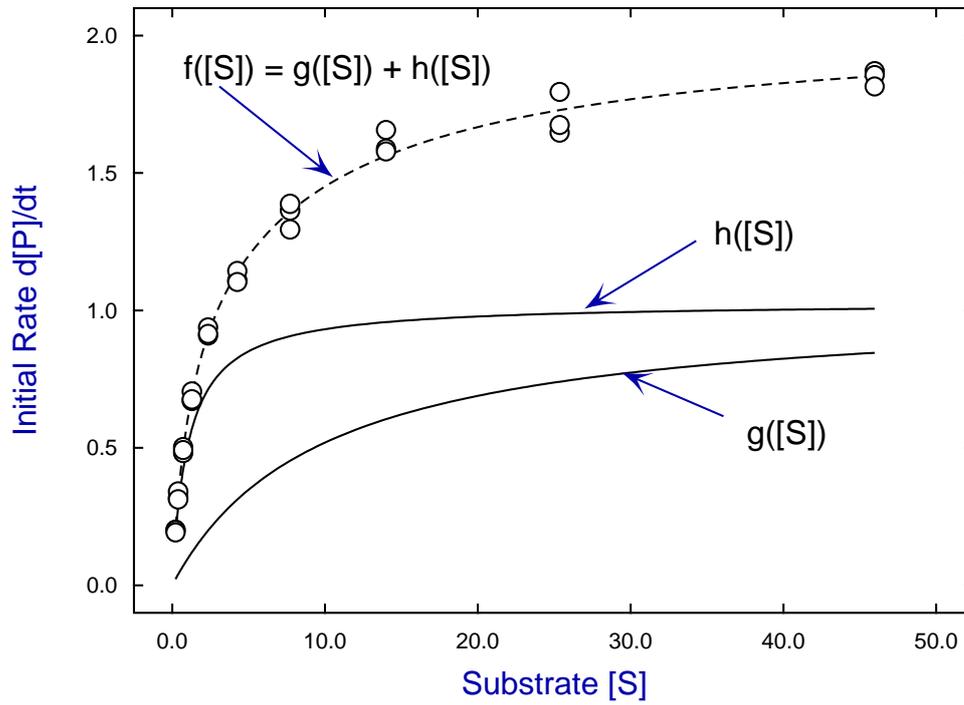
Conclusion based on F test

Reject previous model at 1% significance level
 There is strong support for the extra parameters
 Tentatively accept the current best fit model

Fitting 1 and 2 Site Michaelis-Menten Models



Deconvolution: 2 Michaelis-Mentens



Example 2: Isotope displacement mode

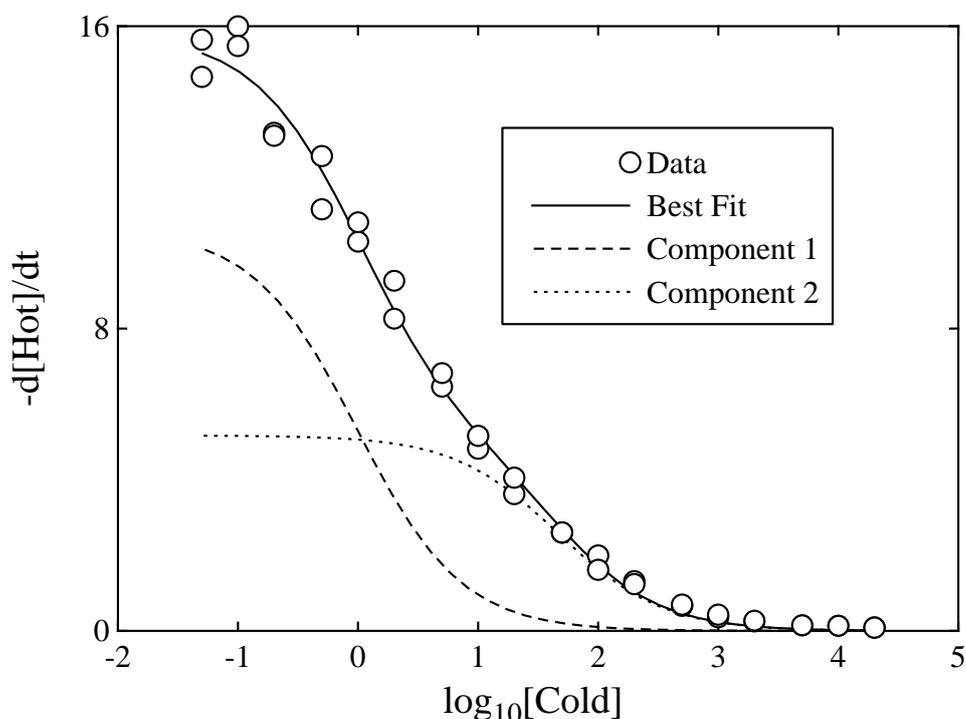
When there is no appreciable kinetic isotope effect, that is, the binding or kinetic transformation process is the same whether the substrate is labeled or not, this allows experiments in which labeled ligand is displaced by unlabeled ligand, or where the flux of labeled substrate is inhibited by unlabeled substrate. Since the ratios of labeled ligand to unlabeled ligand in the bound state, free state, and in the total flux are equal, a modified form of Michaelis-Menten equations can be used to model the binding or kinetic processes. For instance, suppose that total substrate, S say, consists of labeled substrate, $[Hot]$ say, and unlabeled substrate, $[Cold]$ say. Then the flux of labeled substrate for $k \geq 1$ active sites will be given by

$$-\frac{d[Hot]}{dt} = \frac{V_{max_1}[Hot]}{K_{m_1} + [Hot] + [Cold]} + \frac{V_{max_2}[Hot]}{K_{m_2} + [Hot] + [Cold]} + \dots + \frac{V_{max_k}[Hot]}{K_{m_k} + [Hot] + [Cold]}.$$

So, if $[Hot]$ is kept fixed and $[Cold]$ is regarded as the independent variable, then program **mmfit** can be used to fit the resulting data. In other words, cold substrate is being used as a competitive inhibitor of the flux of hot substrate in such experiments.

Using the isotope displacement option in program **mmfit** with the default test file `hotcold.tf1` establishes that two sites is a statistically significant improvement over one site, and leads to the following deconvolution plot to display the best-fit curve together with the separate components.

Isotope Displacement Kinetics



Note that an important difference between using **mmfit** in this mode rather than in straightforward kinetic mode is that the kinetic constants are modified in the following sense: the apparent V_{max} values estimated are actually the true values multiplied by the concentration of labeled substrate, while the apparent K_m values estimated are the true ones plus the concentration of labeled substrate.

Where the actual concentration of $[Hot]$ is known it is possible to fit such data in a more satisfactory and discerning manner by using SIMFIT program **qfit**, where the $[Hot]$ can be input as a fixed constant term so that the actual kinetic constants can be estimated rather than the apparent ones mentioned above.

Theory

Quasi steady-state enzyme kinetics is actually based on the assumption that the substrate concentration remains constant, i.e. $dS/dt = 0$, while the initial rate of product production $dP/dt \geq 0$ is measured. Although it is a contradiction of nomenclature it is a widely used short hand convention nevertheless that an initial rate $v(S)$ can be defined as a flux from substrate S into product P as follows

$$v(S) = \frac{dP}{dt} = -\frac{dS}{dt}$$

and, in the case of $k \geq 1$ independent active sites, the appropriate model equation is

$$v(S) = \frac{V_{max_1} S}{K_{m_1} + S} + \frac{V_{max_2} S}{K_{m_2} + S} + \dots + \frac{V_{max_k} S}{K_{m_k} + S}.$$

In bygone times before the advent of computers, experimentalists had to fit such equations by plotting in transformed spaces, such as the Lineweaver-Burke double reciprocal plot, and then extrapolating to estimate slopes and intercepts, but thankfully this era is long since gone. However, this does not mean that fitting such an equation by constrained weighted least squares is a simple process. It is not. In fact the case with $k = 1$ is trivial, the case with $k = 2$ is reasonable, but the cases $k > 2$ require data that is very extensive and accurate, and where the parameters are sufficiently distinct to allow model discrimination. For this particular model that requires V_{max_i} values to be similar, but K_{m_i} to be distinct.

Program **mmfit** performs the following steps.

1. The v values are first weighted using $w_i = 1/se_i^2$, or used unweighted if all $se_i = 1$.
2. Using the ranges of S_i and v_i the data are transformed into internal coordinates of order unity.
3. Possible starting estimates are calculated for the parameters based on the internal coordinates, and then these are altered by adding pseudo-random perturbations until an approximate minimum value for the weighted sum of squares is located.
4. The parameters are then transformed into internal coordinates that will hopefully be of order unity to stabilize the optimization.
5. From these random starting estimates the lowest and highest possible limits are calculated, then constrained optimization is performed by the quasi-Newton technique.
6. The internal parameters are transformed back into user-space, and the Hessian is estimated at the solution point then inverted to calculate the parameter covariance matrix.
7. The order of parameters is permuted so that the subscripts for $i = 1, 2, \dots, k$ refer to best-fit parameters in the order $V_{max_1} \leq V_{max_2} \leq \dots \leq V_{max_k}$. This is to allow retrospective comparison of fits to alternative data sets.
8. The apparent (overall) V_{max} is calculated as the sum of the V_{max_i} and the apparent (overall) K_m is calculated numerically.
9. Analysis of the residuals is performed together with numerous statistical procedures to ascertain goodness of fit, parameter reliability, and model discrimination.
10. Results tables and graphs are then provided.

Program **mmfit** allows users to control the random search for starting estimates and the technique to be used for calculating the gradient vector, and should the cases with $k > 2$ be required, users can perform extensive random searches to obtain starting estimates that can be input retrospectively for manual starts. If these steps do not succeed it is time to try the SIMFIT advanced curve-fitting program **qfit**.