What Happens When the Wrong Equation is Fitted to Data?

Simon Brown[1], Noorzaid Muhamad[2], Kevin C Pedley[3], David C Simcock[3]

[1]School of Human Life Sciences, University of Tasmania, Launceston, Tasmania 7250, Australia,
[2]Universiti Kuala Lumpur, Royal College of Medicine Perak, 30450 Ipoh, Perak, Malaysia,
[3]Institute of Food, Nutrition and Human Health, Massey University, Palmerston North, New Zealand

*Simon.Brown@utas.edu.au, noorzaid@rcmp.unikl.edu.my,
K.C.Pedley@massey.ac.nz, D.C.Simcock@massey.ac.nz*

Abstract. A common variant of the Michaelis-Menten model of enzyme kinetics involves inhibition by excess substrate. This phenomenon is known as substrate inhibition and the mathematical description of it requires an inhibition constant (K_i) as well as the usual kinetic parameters (K_m and V_{max}). Fitting the 3-parameter substrate inhibition expression to data that might reasonably be described by the 2-parameter Michaelis-Menten model yields biased estimates of K_m and V_{max}. Numerical simulations demonstrate that the extent of the bias is related to the magnitude of the estimated K_i. The quality of the data is particularly important in determining the size of K_i and, therefore, in the bias of the other parameters. Consideration of the residuals, statistical justification of the inclusion of extra parameters and reporting of the estimated values should be matters of routine. The estimates of K_m and V_{max} obtained from a three-parameter substrate inhibition model can only be compared with the corresponding estimates from the two-parameter Michaelis-Menten model with caution.

Keywords: distribution, enzyme kinetics, simulation, substrate inhibition.

1 INTRODUCTION

We have deliberately adapted the title of a paper published in 1978 by Ellis and Duggleby [1] because we have noticed the use of substrate inhibition (SI) models of enzyme kinetics that are not justified by the data [2-4]. This involves analysing data using a three-parameter model rather than the usual two-parameter Michaelis-Menten model (Figure 1). In general, no statistical analysis is given to substantiate the significance of the added parameter which might support the use of such models. In at least some cases, even the most cursory analysis would demonstrate that the extra parameter is not statistically justified.

Several other justifications might be made for analysing data that do not exhibit SI as though they do. These include that
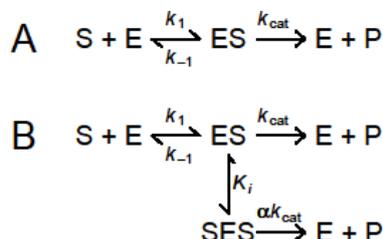
Figure 1. The mechanism of (A) a Michaelis-Menten enzyme [5] and (B) an enzyme exhibiting substrate inhibition [6]. In each case S, E and P are substrate, enzyme and product, respectively and ES and SES are the enzyme-substrate and substrate-enzyme-substrate complex, respectively. The k_i are rate constants, α is a factor reflecting the relative rate of P release from SES and K_i is the dissociation constant of SES.

    (i)   they represent a condition in which SI is eliminated as is the case for *ent*-copalyl diphosphate synthase [3],

    (ii)   related enzymes exhibit SI in the conditions employed, as is the case for IMP dehydrogenase [2], and

    (iii)   related enzymes may exhibit SI in some conditions, as is the case glutamate dehydrogenase [4].

Of these examples, Prisic and Peters [3] have the strongest justification because they report very pronounced SI of *ent*-copalyl diphosphate synthase (E. C. 5.5.1.13) in the presence of Mg^{2+}, but the activity is considerably reduced and SI is effectively eliminated when Mg^{2+} is absent. Obviously, these authors wished to compare the kinetic of the enzyme in various conditions, and it might be argued that this necessitates the treatment of the data using an SI model. The latter two situations provide much weaker justification. For example, IMP dehydrogenase (E. C. 1.1.1.205) does exhibit SI in some protozoa (such as *Cryptospiridium parvum* [7]), but not in others (such as *Leishmania donovani* [8]), which makes the assumption of SI in the *Toxoplasma gondii* enzyme [2] suspect. Similarly, glutamate dehydrogenase (GDH, E. C. 1.4.1.3) can exhibit non-Michaelis-Menten kinetics, including substrate inhibition [9]. While the behaviour of the enzyme can be complicated [10-12], it does depend on the conditions employed. For example, Rife and Cleland [13] pointed out that SI by α-ketoglutarate is not apparent in the bovine enzyme at low NH_3 concentrations.

We have worked on the nitrogen metabolism, and especially the GDH, of the nematode parasite *Teldorsagia circumcincta* for some time [14, 15]. Recently, our own kinetic data [15] have been confirmed using a recombinant enzyme [4]. Unfortunately, these authors assumed that the enzyme exhibited SI, despite the absence of any indication that this was the case, without providing any statistical support [16, 17] for their implicit suggestion that the extra parameter was justifiable and without even providing a complete set of parameter estimates [4]. To make these comparisons it is necessary to know how the assumption of SI impacts on the estimates of K_m and V_{max}.

These observations prompted us to ask three questions. First, how large might the unreported parameter be? Second, how might the use of an inappropriate model

distort the estimates of the reported parameters? Third, how the experimental error in the measurements influences the parameter estimates obtained. Here, we outline the relevant theory and then address each of these questions in turn using numerical experiments.

2 THEORY

The Michaelis-Menten reaction [5] in which an enzyme (E) converts a substrate (S) to a product (P) involves the transient formation of an enzyme-substrate complex (ES) (Figure 1A). The rate of P formation is given by

$$v_M(s) = \frac{V_{max}\, s}{K_m + s} \qquad (1)$$

where s is the concentration of S, V_{max} is the asymptotic value of v_M at high s and K_m is sometimes referred to as the 'affinity' of the enzyme for S [18]. Equation (1) describes a rectangular hyperbola in which v_M increases with s, $v_M = 0.5 V_{max}$ when $s = K_m$ and approaches V_{max} as s approaches infinity. While there is some argument as to whether any enzyme functions according to this model [19], it does provide a simple means of characterising the kinetic behaviour of many enzymes.

However, the Michaelis-Menten model really does not describe the mechanism of the many enzymes that exhibit SI [9]. In such cases, the initial rise in v with increasing s is followed by a decline in v as s is increased further (Figure 2). A number of mechanisms could give rise to this behaviour, but a simple model in which S may bind to a second site on the enzyme, forming a ternary SES complex (Figure 1B) and thereby eliminating (or at least reducing) the activity of the enzyme is commonly described [6]. This results in a modified Michaelis-Menten equation

$$v_S(s) = \frac{V_{max}\, s}{K_m + \left(1 + \dfrac{s}{K_i}\right)s} \qquad (2)$$

in which K_i is the dissociation constant of SES (Figure 1B). It is obvious from (2) that the maximum v_S is

$$v_S\left(\sqrt{K_m K_i}\right) = \frac{\sqrt{K_i}}{2\sqrt{K_m} + \sqrt{K_i}} V_{max} \qquad (3)$$

and that $v_S(K_m) \leq v_S((K_m K_i)^{1/2})$. It is not possible to put any useful bounds on the relative magnitudes of K_m and K_i. The concentrations at which $v_S = 0.5 v_S((K_m K_i)^{1/2})$ are
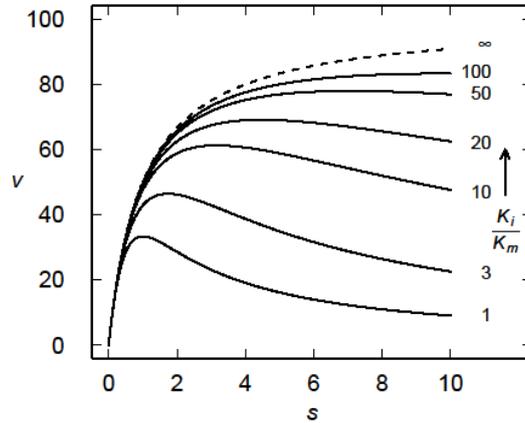
Figure 2. Kinetics of the Michaelis-Menten reaction (dashed curve) and substrate inhibition model (solid curves) for several different values of K_i/K_m (as indicated). The dashed curve was calculated using (1), but is equivalent to (2) if $K_i/K_m \rightarrow \infty$, and the other curves were calculated using (2). In all cases $V_{max} = 100$ units and $K_m = 1$ unit.

$$\frac{s_{0.5}}{K_m} = \frac{1}{2}\sqrt{\frac{K_i}{K_m}}\left(\sqrt{\frac{K_i}{K_m}} + 4 \pm \sqrt{\left(\sqrt{\frac{K_i}{K_m}} + 2\right)\left(\sqrt{\frac{K_i}{K_m}} + 6\right)}\right). \qquad (4)$$

While the two values of $s_{0.5}/K_m$ given by (4) exhibit increase with K_i/K_m, they exhibit distinct behaviours (Figure 3). As K_i/K_m approaches infinity, the smaller value approaches 1 whereas the larger value increases monotonically (Figure 3A), reflecting the asymmetry of (2) around the maximum as is apparent in Figure 2. As K_i/K_m approaches infinity, the maximum v_S approaches V_{max} (3) and $v_S(K_m)$ approaches $0.5V_{max}$, but around $K_i/K_m = 1$ the maximum v_S equals $v_S(K_m)$ (Figure 3B).

3 COMPUTATIONAL METHODS

Experimental data were simulated in R [20] using (1) and a normally distributed random error term (ε; mean $= 0$, standard deviation $= \theta$) was added to each datum

$$v(s) = v_M(s) + \varepsilon(0, \theta), \qquad (5)$$

where $s = (0, 0.5, 1, 2, 3, 5, 7.5, 10)$, $V_{max} = 100$ and $K_m = 1$. The error term was calculated using the `rnorm` function in R and values of θ as specified. Equation (2) was fitted to the eight simulated data points by least squares nonlinear regression to obtain estimates of V_{max}, K_m and K_i. This process was repeated to obtain n simulated replicates yielding a total of $8n$ data points and n replicates of the parameter estimates. The mean absolute error (Δ) was estimated from these $8n$ points using
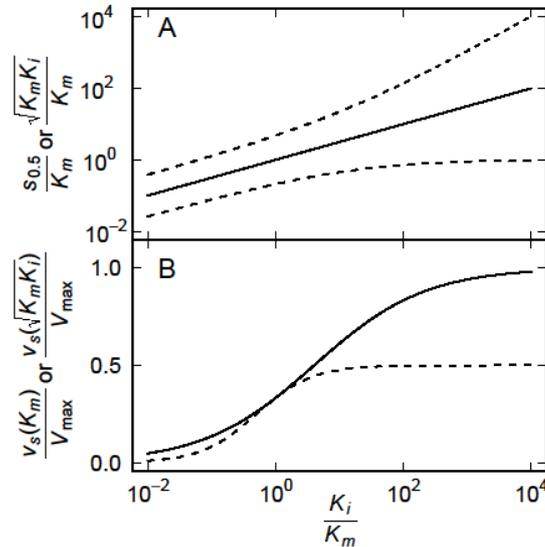
Figure 3. Dependence on K_i/K_m of (A) the concentration of S yielding maximum v_S (solid line) and the concentrations at which v_S is half of the maximum values (dashed curves, (4)), and of (B) the maxium v_S (solid curve, (3)) and the value of $v_S(K_m)$ (dashed curve). Note that $(K_m K_i)^{1/2}/K_m = (K_i/K_m)^{1/2}$, but the form used is intended to emphasise the connection with (3).

$$\Delta = \frac{1}{8n} \sum_{j=1}^{n} \sum_{i=1}^{8} \left| \hat{v}_S(s_i) - v_M(s_i) \right|, \tag{6}$$

where $\hat{v}_S$ is the fitted value of (2). The magnitude of Δ can be compared with $V_{max} = 100$.
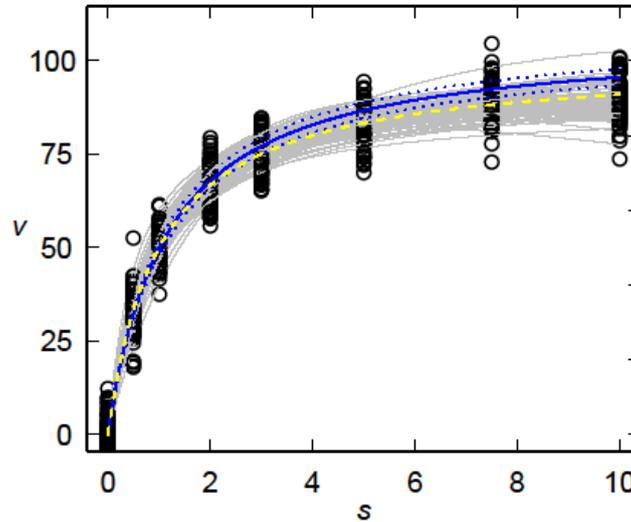
   We have previously demonstrated the value of a confidence band for (1)

$$\frac{\varepsilon}{v_M} = \sqrt{\left(\frac{\varepsilon_V}{V_{max}}\right)^2 + \frac{\varepsilon_K^2}{(K_m + s)^2}} \tag{7}$$

[21] and the same approach yields a confidence band (ε) for (2)

$$\frac{\varepsilon}{v_S} = \sqrt{\left(\frac{\varepsilon_V}{V_{max}}\right)^2 + \frac{\varepsilon_K^2 + (s/K_i)^4 \varepsilon_I^2}{\left(K_m + s\left(1 + \frac{s}{K_i}\right)\right)^2}} \tag{8}$$

where ε_V, ε_K and ε_I are the errors associated with V_{max}, K_m and K_i, respectively. As for (2) and (1), (8) tends towards (7) if K_i is large.

Figure 4. Fits (grey curves) of (2) to simulated data (○) calculated from (5) assuming $\theta = 3$. The average parameter values (K_m, V_{max} and K_i) obtained from the fits to each of 100 simulated datasets were used in (2) to calculate the expected curve (solid blue curve) and the associated 95% confidence band (dotted blue curves) were calculated from (8) using the corresponding error estimates (ε_K, ε_V and ε_I). The dashed yellow curve is the Michaelis-Menten curve (1) on which the simulation was based ($V_{max} = 100$ units, $K_m = 1$ unit).

4 NUMERICAL EXPERIMENT

Some representative fits of (2) to simulated data (5) are shown in Figure 4. It is clear that the simulated experimental variation leads to a considerable range of behaviour, including examples in which there is a clear indication of a decline in v with increasing s (see the lowest grey curve for $s = 8-10$ units in Figure 4). Using the averages of the $n = 100$ estimated values of V_{max}, K_m and K_i in (2) yields an overestimate of v compared with v_M, even when the 95% confidence interval (8) is taken into account (Figure 4).

What is not apparent from Figure 4 is the range of parameter estimates obtained even with $\theta = 3$. A larger number of simulations yields distributions of the parameters (Figure 5). Clearly, there is a strong linear relationship between the estimated values of K_m and V_{max} (Figure 5A) and both parameters have unimodal distributions (Figures 5, A and B). However, a significant proportion of the simulations yielded parameter estimates that were much larger than the actual values ($V_{max} = 100$ units, $K_m = 1$ unit) and a smaller proportion of the estimates were smaller (Figure 5, A and B). The estimates of K_i had a bimodal distribution (Figure 5C) in which about half of the values were large (about 10^6) and the remainder were 10^0-10^3. A K_i of about 10^6 is sufficient to render (2) essentially indistinguishable from (1), but those values that are less than 10^2 (and even 10^3) yield curves quite distinct from (1), as is clear from Figure 2. However, high K_i values tended to be

associated with estimates of K_m and V_{max} close to the correct values (Figure 5, B and C). Smaller values of K_i (say $< 10^2$) were associated with overestimates of both K_m and V_{max} (Figure 5, B and C).

Increasing the simulated experimental error by increasing the standard deviation (θ) of the error term in (5) yields a corresponding increase in the variation in the parameter estimates (Figure 6). These distributions are similar to those shown in Figure 5, but the range of variation in both V_{max} and K_m increases exponentially with increasing Δ (Figure 6). The bimodal distribution of K_i is clear because the mean is largely unaffected by changes in Δ, whereas the median declined with increasing Δ (Figure 6C).

5 DISCUSSION

The unjustified use of (2) rather than (1) can distort the estimates of V_{max} and K_m obtained. In general, this tends to yield overestimates of V_{max} and K_m, especially if experimental error is sufficient to result in estimates of $K_i < 10^3$, as is illustrated in Figure 2. It is a simple matter to determine whether the introduction of an extra parameter is statistically justified [16, 17], but this tends not to be done [2, 4] and it is difficult to assess without access to the data. Irrespective of this, if (2) is fitted to data, it is important that all three parameter estimates (and the associated error) are published because then (8) can be used in conjunction with (7) to assess the viability of (1) and (2). Unfortunately, the parameter estimates are not always given [4]. In these circumstances, it is possible only to assess the data by eye, in which case systematic patterns in the residuals are a valuable indication [1].
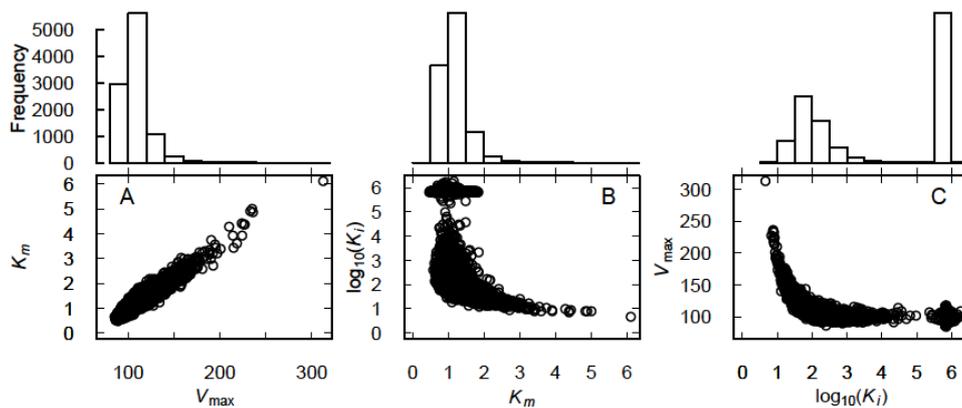
Figure 5. Distributions and relationships between the parameter estimates ($\circ$) obtained from fitting (2) to each of 10000 simulations of (5). (A) The relationship between the estimates of V_{max} and K_m and the distribution of V_{max} (upper histogram). (B) The relationship between the estimates of K_m and $\log_{10}(K_i)$ and the distribution of K_m (upper histogram). (C) The relationship between the estimates of V_{max} and $\log_{10}(K_i)$ and the distribution of $\log_{10}(K_i)$ (upper histogram). It was assumed that $V_{max} = 100$ units, $K_m = 1$ unit and $\theta = 3$.
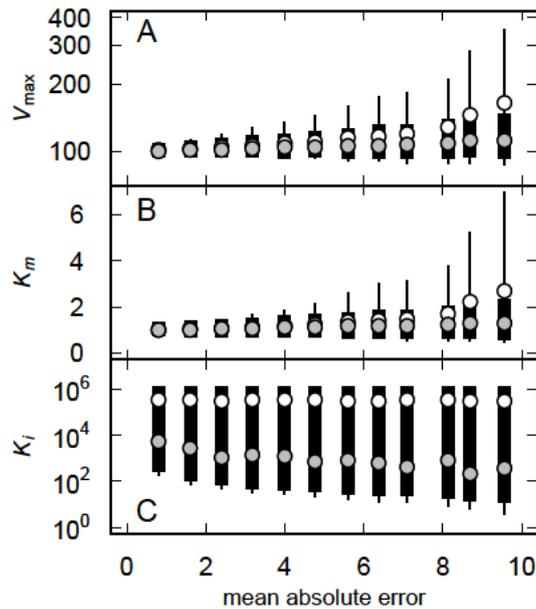
Figure 6. Effect of error on the distributions of the estimates of V_{msc} (A), K_m (B) and K_i (C) obtained by fitting (2) to data simulated using (5). In each case the solid black column represents the 25th-75th percentiles, the line extends from the 5th to the 95th percentile and the open ($\circ$) and grey ($\bullet$) circles represent the mean and median, respectively.

In passing, we suggest that the inappropriate use of SI models is partly promoted by their availability in popular software packages such as Prism, which was used in two of the three examples we outlined in the introduction [2, 4]. Another common source of bias in parameter estimates [23, 24] is the use of the double-reciprocal plot proposed by Lineweaver and Burk in 1934 [22]. We speculate that the widespread use of software of this sort promotes the continued application of such transformations despite the statistical arguments in favour of the use of nonlinear regression [25] and the availability of the necessary software [20].

The K_m or the V_{max} obtained from (1) are not necessarily comparable with those obtained from (2) because the inclusion of K_i can distort the estimates. Consequently the parameter estimates obtained from the models shown in Figure 1 (1-2) can only be compared with caution, which justifies the approach adopted by Prisic and Peters [3]. However, if (2) is used, then all of the parameter estimates

should be reported and, if possible, the statistical significance of the third parameter (K_i) should be demonstrated [16, 17].

REFERENCES

1. Ellis, KJ and Duggleby, RG, "What happens when data are fitted to the wrong equation?" Biochemical Journal 1978; 171: 513-517.

2. Sullivan, WJ, Jr, Dixon, SE, Li, C, Striepen, B and Queener, SF, "IMP dehydrogenase from the protozoan parasite *Toxoplasma gondii*", Antimicrobial Agents and Chemotherapy 2005; 49: 2172-2179.

3. Prisic, S and Peters, RJ, "Synergistic substrate inhibition of ent-copalyl diphosphate synthase: a potential feed-forward inhibition mechanism limiting gibberellin metabolism", Plant Physiology 2007; 144: 445-454.

4. Umair, S, Knight, JS, Patchett, ML, Bland, RJ and Simpson, HV, "Molecular and biochemical characterisation of a *Teladorsagia circumcincta* glutamate dehydrogenase", Experimental Parasitology 2011; 129: 240-246.

5. Michaelis, L and Menten, ML, "Die Kinetik der Invertinwirkung", Biochemische Zeitschrift 1913; 49: 333-369.

6. Haldane, JBS, "Enzymes", Longmans, Green and Co., London. 1930.

7. Umejiego, NN, Li, C, Riera, T, Hedstrom, L and Striepen, B, "*Cryptosporidium parvum* IMP dehydrogenase. Identification of functional, structural, and dynamic properties that can be exploited for drug design", Journal of Biological Chemistry 2004; 279: 40320-40327.

8. Dobie, F, Berg, A, Boitz, JM and Jardim, A, "Kinetic characterization of inosine monophosphate dehydrogenase of *Leishmania donovani*", Molecular and Biochemical Parasitology 2007; 151: 11-21.

9. Reed, MC, Lieb, A and Nijhout, HF, "The biological significance of substrate inhibition: a mechanism with diverse functions", BioEssays 2010; 32: 422-429.

10. Hudson, RC and Daniel, RM, "L-Glutamate dehydrogenases: distribution, properties and mechanism", Comparative Biochemistry and Physiology 1993; 106B: 767-792.

11. Kurganov, BI, "New approach to analysis of deviations from hyperbolic law in enzyme kinetics", Biokhimiya 2000; 65: 898-909.

12. Smith, TJ and Stanley, CA, "Untangling the glutamate dehydrogenase allosteric nightmare", Trends in Biochemical Sciences 2008; 33: 557-564.

13. Rife, JE and Cleland, WW, "Kinetic mechanism of glutamate dehydrogenase", Biochemistry 1980; 19: 2321-2328.

14. Muhamad, N, Brown, S, Pedley, KC and Simpson, HV, "Kinetics of glutamate dehydrogenase from L_3 *Ostertagia circumcincta*", New Zealand Journal of Zoology 2004; 31: 97.

15. Muhamad, N, Simcock, DC, Pedley, KC, Simpson, HV and Brown, S, "The kinetics of glutamate dehydrogenase of *Teladorsagia circumcincta* and the lifestyle of the parasite", Comparative Biochemistry and Physiology 2011; 159B: 71-77.

16. Akaike, H, "A new look at the statistical model identification", IEEE Transactions on Automatic Control 1974; 19: 716-723.

17. Bardsley, WG, McGinlay, PB and Wright, AJ, "The F test for model discrimination with exponential functions", Biometrika 1986; 73: 501-508.

18. Briggs, GE and Haldane, JBS, "A note on the kinetics of enzyme action", Biochemical Journal 1925; 19: 338-339.

19. Hill, CM, Waight, RD and Bardsley, WG, "Does any enzyme follow the Michaelis-Menten equation?" Molecular and Cellular Biochemistry 1977; 15: 173-178.

20. Ihaka, R and Gentleman, R, "R: a language for data analysis and graphics", Journal of Computational and Graphical Statistics 1996; 5: 299-314.

21. Brown, S, Muhamad, N, Pedley, KC and Simcock, DC, "A simple confidence band for the Michaelis-Menten equation", International Journal of Emerging Sciences 2012; 2: 238-246.

22. Lineweaver, H and Burk, D, "The determination of enzyme dissociation constants", Journal of the American Chemical Society 1934; 56: 658-666.

23. Dowd, JE and Riggs, DS, "A comparison of estimates of Michaelis-Menten kinetic constants from various linear transformations", Journal of Biological Chemistry 1965; 240: 863-869.

24. Blunck, M and Mommsen, TP, "Systematic errors in fitting linear transformations of the Michaelis-Menten equation", Biometrika 1978; 65: 363-368.

25. Johnson, ML, "Why, when, and how biochemists should use least squares", Analytical Biochemistry 1992; 206: 215-225.