Estimation of the initial velocity of enzyme-catalysed reactions by non-linear regression analysis of progress curves

Ronald G. DUGGLEBY

Department of Biochemistry, University of Queensland, St. Lucia, Queensland 4067, Australia

(Received 1 October 1984/3 January 1985; accepted 29 January 1985)

Most methods for studying the kinetic properties of an enzyme involve the determination of initial velocities. When the reaction progress curve shows significant curvature due to depletion of the substrate, accumulation of inhibitory products or instability of the enzyme, estimation of the initial velocity is a subjective and inexact process. Two methods have been suggested [Cornish-Bowden (1975) *Biochem. J.* **144**, 305–312; Boeker (1982) *Biochem. J.* **203**, 117–123] that attempt to eliminate this subjective element. The present study offers a third alternative, which is based on fitting a reparameterized form of the integrated Michaelis–Menten equation to the progress curves by non-linear regression. This method yields estimates and standard errors of the initial velocity and of the time to reach 50% reaction. No prior knowledge of the apparent product concentration at zero time or infinite time is required, since both of these quantities are also estimated from the data. It is shown that this method yields reliable estimates of the initial velocity under a wide range of circumstances, including those where the two previously published methods perform poorly.

The determination of the kinetic properties of enzymes by analysis of entire reaction progress curves is a well-known technique. It has attracted attention because it appears to offer the opportunity to evaluate both the apparent Michaelis constant (K_a) and the apparent maximum velocity (V_m) by fitting the integrated Michaelis–Menten equation (eqn. 1) to a single progress curve:

$$V_m t = y_t - K_a \cdot \ln(1 - y_t/y_\infty) \quad (1)$$

(The symbol y_t denotes the concentration of product formed by catalysis, which will differ from the total concentration of product when this compound is added to the system. The meaning of y_∞ depends on the context; in systems that do not undergo progressive enzyme inactivation, it denotes the concentration of product formed at infinite time and will be equivalent to the initial substrate concentration for irreversible reactions only. When there is significant enzyme inactivation, this latter definition is used, although the reaction will cease before the substrate is exhausted.) Despite this attraction, the method is not gained wide acceptance, and a variety of reasons can be advanced for this fact. One of these reasons is the natural reluctance to move away from the well-established technique of initial-velocity

analysis. This reluctance is probably justifiable, for, as Cornish-Bowden (1975) has shown, relatively small systematic errors can lead to wildly inaccurate values of V_m and K_a . For example, a 5% error in y_∞ can result in 300% errors in V_m and K_a . From a detailed analysis of several situations, Cornish-Bowden (1975) has argued persuasively that V_m and K_a be used merely as vehicles for calculating v_0 , the initial velocity. Under his proposal, each possible pair of experimental observations from a progress curve are combined to calculate a V_m and K_a that are then used to calculate a v_0 . The median of this series of values of v_0 is then taken to be an estimate of the true initial velocity, $\bar{v}_0$. His procedure is referred to below as the median- v_0 method.

More recently, Wharton & Szawelski (1982) have presented a similar case for using each progress curve to evaluate a single quantity, in this case the time to achieve 50% reaction. This half-time method appears to be more vulnerable to systematic errors in y_∞ than the median- v_0 method, and these authors recommend that reactions be allowed to proceed for a prolonged period so that y_∞ may be measured reliably.

A second source of systematic error in progress curves occurs when the enzyme is not stable under

assay conditions. Although there is a simple method for detecting instability (Selwyn, 1965), the half-time method cannot be used when instability is significant. The non-linear regression method of Duggleby & Morrison (1977) is also inappropriate in these circumstances. By contrast, the median- v_0 method performs quite well in the face of quite substantial (up to 25%) enzyme inactivation.

It has been noted (Duggleby & Morrison, 1977) that a common problem in progress-curve analysis is that the concentration co-ordinate may be displaced by an unknown amount (y_0) from the origin. This difficulty was not addressed by either Cornish-Bowden (1975) or by Wharton & Szawelski (1982), but was considered by Boeker (1982). Like Cornish-Bowden (1975), she argued that each progress curve be used solely to determine v_0 , and proposed that this be done by extrapolating a plot of $(y_t - y_0)/t$ versus $y_t - y_0$ to the ordinate. This procedure is referred to below as the extrapolated- v_0 method. Boeker (1982) offered no guidance as to exactly how y_0 is to be determined.

One possible approach to the estimation of v_0 is to fit a polynomial to the data. This appears attractive, since it does not require a knowledge of y_0 , y_∞ or even the underlying theoretical relationship between y_t and t . However, as Cornish-Bowden (1975) has discussed, this approach has not proved to be reliable.

In the present paper a non-linear regression method for analysing progress curves is described. The method, which does not require prior knowledge of either y_0 or y_∞ , is based on a reparameterization of eqn. (1) in such a way that v_0 and t_1 , rather than V_m and K_a , are the fitted parameters. The method bears a superficial resemblance to the procedure described by Duggleby & Morrison (1979), in which several velocities are calculated at points along a progress curve, but that method is not applicable to systems in which the enzyme is unstable. It is shown that the present method gives accurate values for v_0 over a wider range of conditions than either the median- v_0 method or the extrapolated- v_0 method.

Theory

The differential form of eqn. (1) as it applies to progress curves may be written as:

$$\frac{dy_t}{dt} = \frac{1}{1/V_m + k/(y_\infty - y_t)} \quad (2)$$

where k represents the ratio K_a/V_m . The initial velocity, v_0 , is given by eqn. (3):

$$v_0 = \frac{1}{1/V_m + k/y_\infty} \quad (3)$$

Combining eqns. (2) and (3) by eliminating V_m yields eqn. (4):

$$\frac{dy_t}{dt} = \frac{y_\infty}{y_\infty/v_0 + ky_t/(y_\infty - y_t)} \quad (4)$$

If eqn. (4) is integrated and t_1 defined by setting $y_t = y_\infty/2$, k can be eliminated to yield eqn. (5), which relates t , y_t and the three parameters v_0 , t_1 and y_∞ :

$$t = \frac{y_t}{v_0} + \frac{2v_0 t_1 - y_\infty}{v_0(1 - \ln 4)} \cdot \left[\frac{y_t}{y_\infty} + \ln \left(1 - \frac{y_t}{y_\infty} \right) \right] \quad (5)$$

Methods

Error-free data that are described by eqn. (1) were calculated in one of two ways. In the first of these, values of V_m , K_a and y_∞ were chosen, and t was then calculated from eqn. (1) for a series of equally spaced values of y_t . The second method was similar, except that y_t was calculated for a series of equally spaced values of t ; this involved solving an implicit equation, which was done by using the Newton-Raphson procedure (see Duggleby & Morrison, 1977).

Theoretical progress curves for situations in which the enzyme undergoes first-order inactivation with a rate constant of k_i were calculated by solving eqn. (6) by using the Newton-Raphson method:

$$V_m[1 - \exp(-k_i t)]/k_i = y_t - K_a \cdot \ln(1 - y_t/y_\infty) \quad (6)$$

Simulated data for both eqn. (1) and eqn. (6) were calculated from the theoretical curve; each y_t value was perturbed by adding to it a normally distributed pseudo-random number with a mean of zero and the required standard deviation. These perturbations were calculated by using the algorithm described by Brent (1974).

Eqn. (5) may be fitted to progress-curve data by non-linear regression. In the work described in the present paper, this was achieved by using a BASIC program, which may be applied to a wide range of non-linear regression problems and is described elsewhere (Duggleby, 1984). All points were equally weighted for this analysis, and the initial estimates of the parameters were close to their true values. Experience has shown that the program will tolerate quite poor initial estimates and still converge to the same final values; it just takes longer to get there. Such an analysis yields best estimates and standard errors of the three parameters v_0 , t_1 and y_∞ . For progress curves that are displaced from the origin, eqn. (5) is rewritten by replacing y_t and y_∞ with $y_t - y_0$ and $y_\infty - y_0$ respectively and treating y_0 as a fourth parameter

to be estimated from the data. A listing of the computer program (DNRP53) and a user's manual are available on request from the author.

The median- v_0 method involved the solution of eqn. (1) for all possible pairs of points from a progress curve, as outlined by Cornish-Bowden (1975). {It should be noted that the original description of this method contains an algebraic error in eqn. (2) of Cornish-Bowden (1975); the second logarithmic term should read $\ln[p_\infty/(p_\infty - p_i)]$.} The extrapolated- v_0 method involved linear regression of y_i/t against y_i to determine the intercept on the ordinate. The data were truncated, where necessary, to exclude points

for which y_i exceeded the value corresponding to 50% utilization of the substrate.

Results

Error-free data

In an initial study, error-free data were calculated from eqn. (1) for a range of conditions. Eqn. (5) was fitted to the data to obtain values for v_0 , t_i , y_0 and y_∞ . Table 1 summarizes the differences between these values and the expected values, as well as the corresponding error in v_0 as determined by the extrapolated- v_0 method. Table 2 summarizes an analysis of error-free data under various

Table 1. *Errors in the estimation of v_0 , t_i , y_0 and y_∞ when eqn. (5) is fitted to error-free data*

Theoretical curves were calculated from eqn. (1), by using $V_m = 1.0$ and $y_\infty = 1.0$. The value of K_a was adjusted to give an initial velocity of 0.1, 0.4 or 0.75. Ten y_i values were calculated so as to be equally spaced in time, up to a maximum y_i of 10%, 50% or 90% of y_∞ . The range of the data, in terms of the percentage of y_∞ of the first and last point, as well as the fraction of t_i corresponding to the last point are reported. Eqn. (5) was fitted to the data and the differences between found and expected values are reported as parts per million of the expected values except for y_0 where the error is given as parts per million of y_∞ . The values in parentheses are the corresponding errors in v_0 , which were obtained by using the procedure described by Boeker (1982).

True v_0	Range of the data		Errors in the values determined			
	Percentage reaction	Fraction of t_i	t_i	y_0	y_∞	v_0
0.1	1-10	0.156	-74	+1	+40	-26 (+230)
	7-50	1.000	-100	+2	-72	-21 (+8920)
	20-90	3.209	-3	-1	0	0 (+14880)
0.4	1-10	0.168	-20	+1	+10	-15 (+263)
	6-50	1.000	+71	-1	+60	+10 (+9970)
	16-90	2.828	0	0	-2	-3 (+12623)
0.75	1-10	0.185	-19	0	+10	-5 (+161)
	5-50	1.000	+1	0	0	-5 (+6432)
	12-90	2.281	-1	0	-2	-3 (+7637)

Table 2. *Effect of enzyme inactivation on the estimation of v_0 by three methods*

Theoretical data were calculated from eqn. (6) with $V_m = 1.0$ and an initial substrate concentration equal to $2K_a$, so that $v_0 = \frac{2}{3}$. Ten y_i values were chosen so as to be equally spaced in concentration, up to a maximum of 10% or 50% of the initial substrate concentration, and the time to reach each y_i value was calculated. The rate constant for enzyme inactivation (k_i) was chosen so as to achieve the desired extent of inactivation (5%, 10%, 30% or 90%) at the end of each progress curve. Values for v_0 were calculated by fitting eqn. (5) to the data, by the median- v_0 method (assuming y_∞ to be equal to the initial substrate concentration), or by the extrapolated- v_0 method. Results are presented as the percentage error in the estimate of v_0 .

Percentage reaction	Percentage inactivation	Percentage error in v_0		
		Eqn. (5) method	Median- v_0 method	Extrapolated- v_0 method
10	5	-0.004	0.000	+0.025
	10	+0.001	+0.043	+0.039
	30	+0.003	+0.761	+0.221
	90	+0.016	+11.938	+3.952
50	5	-0.030	-0.077	+0.742
	10	-0.056	-0.122	+0.754
	30	-0.046	+0.118	+0.922
	90	+0.448	+9.216	+4.297

Table 3. *Estimation of v_0 by three methods from simulated data*

A progress curve was calculated by using eqn. (1) assuming $V_m = 1.0$, $K_a = 1.0$ and an initial substrate concentration of 2.0, so that the expected value of v_0 is $\frac{2}{3}$. Ten y_i values were chosen so as to be equally spaced in concentration, up to a maximum of half the initial substrate concentration, and the corresponding time was determined. A simulated curve was calculated by adding to each y_i value a normally distributed pseudo-random number with a mean of zero and a standard deviation of 0.025. The initial velocity was determined by fitting eqn. (5) to the data, by the median- v_0 method and by the extrapolated- v_0 method. The means and standard deviations of 100 such simulations are given. The upper and lower quartiles are also shown, these being the values of v_0 beyond which one-quarter of estimates fall. In addition, the number of estimates of v_0 that are above the true value is given for each method. The final row of the Table reports the number of simulations for which each method was best, in the sense that it gave an estimate of v_0 closest to the true value.

	Eqn. (5) method	Median- v_0 method	Extrapolated- v_0 method
Mean	0.66980	0.66457	0.66271
Standard deviation	0.02938	0.04605	0.07252
Quartiles			
Upper	0.69083	0.69302	0.70838
Lower	0.64715	0.63676	0.60449
Number greater than v_0	53	45	46
Number closest to v_0	56	30	14

Table 4. *Estimation of v_0 by three methods from simulated data in which there is progressive enzyme activation*
Conditions and procedures were as described in Table 3, except that eqn. (6) was used to calculate the theoretical curve by using a k_i value of 0.531555, which results in 90% enzyme inactivation when 50% of the substrate is utilized.

	Eqn. (5) method	Median- v_0 method	Extrapolated- v_0 method
Mean	0.66422	0.73229	0.66513
Standard deviation	0.06969	0.11614	0.13097
Quartiles			
Upper	0.70197	0.79507	0.74275
Lower	0.61697	0.65872	0.55882
Number greater than v_0	44	73	46
Number closest to v_0	52	30	18

conditions of enzyme inactivation. Each progress curve was calculated by using eqn. (6) and was analysed by fitting eqn. (5) to the data, by the median- v_0 method and by the extrapolated- v_0 method; the values shown are the percentage errors in v_0 as calculated by the three methods.

Simulated data

In order to mimic the effect of experimental error, simulated progress curves were calculated from exact theoretical data by adding to each y_i value a normally distributed pseudo-random number with a mean of zero and a standard deviation of 0.025 (1.25% of the initial substrate concentration). Each simulated progress curve was used to calculate v_0 by the three methods noted above. Table 3 summarizes the results of 100 such simulations for which eqn. (1) is the underlying theoretical relationship. A similar analysis of 100 simulations of progress curves that exhibit progressive enzyme inactivation (eqn. 6) is summarized in Table 4.

A third set of 100 simulations was designed to test the effect of outliers. Simulated progress curves were calculated as in Table 3, except that one randomly selected point from each curve was perturbed by adding to it a normally distributed pseudo-random number with a mean of zero and a standard deviation of 0.1; this error is 4 times that in the remaining nine points. Fitting eqn. (5) to these data gave a mean v_0 of 0.66704 with a standard deviation of 0.04963, whereas the median- v_0 method gave 0.64816 ± 0.05273 . The latter method gave an estimate of v_0 that was closer to the true value of $\frac{2}{3}$ in 41 of these 100 simulations, whereas eqn. (5) was better in the remaining 59.

Discussion

The initial study (Table 1) was directed towards determining how well v_0 , t_4 , y_0 and y_∞ were determined when eqn. (5) was fitted to error-free data. In all cases, v_0 and y_0 were within a few parts per million of the expected value. The errors in t_4

and y_{∞} were also very small, and in no case did they exceed 100 parts per million (i.e. 0.01%).

It is not possible to make a direct comparison between the results in Table 1 and the median- v_0 method, since this method requires an assumption with regard to the value of y_{∞} . If y_{∞} is known exactly, the median- v_0 method should yield values in the precise agreement with the expected values, and this was found to be so within a few parts per million. Such discrepancies that were detected are simply a reflection of rounding errors in the data and calculations. The results from the median- v_0 method were affected by small changes in the assumed value of y_{∞} . For example, changing y_{∞} by 2% in the case where the true $v_0 = 0.1$ and the range of the data covered 7–50% reaction resulted in an error of 850 parts per million in v_0 , a value that is consistent with the results found by Cornish-Bowden (1975) under similar conditions.

The extrapolated- v_0 method (Boeker, 1982) is relatively free of assumptions with regard to y_{∞} and could be compared directly with the v_0 values obtained by fitting eqn. (5) to the data. Such a comparison is included in Table 1, from which it is seen that the extrapolated- v_0 method gave much larger errors, up to 14880 parts per million (1.5%). This is in general agreement with Boeker (1982), who found errors up to 0.9%.

Cornish-Bowden (1975) has examined the situation where there is progressive enzyme inactivation by assuming that V_m undergoes a first-order decrease with time. Table 2 shows v_0 values that were found under these conditions, as calculated by using eqn. (5), the median- v_0 method and the extrapolated- v_0 method. In all cases but one (50% reaction, 90% inactivation) the v_0 value from eqn. (5) differed from the true value by less than 0.06%. Even in the one case where eqn. (5) performed most poorly, the error was less than 0.5%; this error was substantially smaller than that obtained by using either the median- v_0 method (9.2%) or the extrapolated- v_0 method (4.3%).

The results presented in Tables 1 and 2 may not provide a realistic assessment of the various methods for estimating v_0 , as they were obtained by using data that were error-free. Consequently the three methods were compared by using simulated data, which were calculated for the case in Table 2 where eqn. (5) performed most poorly (50% reaction, 90% inactivation). Each simulated progress curve was used to calculate v_0 by fitting eqn. (5) to the data, by the median- v_0 method and by the extrapolated- v_0 method (Table 4).

Fitting eqn. (5) to the data gave a mean v_0 that was within 0.4% of the true value ($\bar{v}_0 = \frac{2}{3}$), and all the simulations gave values that were clustered tightly together (standard deviation = 10.5% of $\bar{v}_0$). By contrast, the median- v_0 method appeared to

overestimate the initial velocity by 9.8%. Further, the values were much more widely scattered, with a standard deviation of 17.4%, nearly double that obtained by using eqn. (5). A similar scattering was seen in the results obtained by using the extrapolated- v_0 method, although the mean value was very close to the true value. In order to confirm that the larger standard deviations seen with the median- v_0 and extrapolated- v_0 methods were due to a wider scattering of the values, rather than a few particularly aberrant values, the upper and lower quartiles are also given in Table 4. These values confirm that eqn. (5) gives v_0 values that are more tightly clustered than those given by the median- v_0 method, which itself gives values showing a smaller dispersion than those given by the extrapolated- v_0 method.

It is evident that, for this particular situation, the median- v_0 method is biased towards high values, and, indeed, 73 of the 100 simulations yielded values that were greater than the true initial velocity. By contrast, the other two methods gave approximately equal numbers of high and low values [44 were high for eqn. (5) and 46 were high for the extrapolated- v_0 method]. One final way to compare these three methods is to count the number of times each was best, in the sense that it gave a v_0 value closest to the true initial velocity. If all three methods were equally good, each would be closest to $\bar{v}_0$ in one-third of the simulations; in fact, eqn. (5) was best in 52 instances, compared with 30 and 18 for the median- v_0 and extrapolated- v_0 methods respectively. Thus eqn. (5) performed best more often than did the other two methods combined, a result that has a probability of occurring by chance of less than 0.0001 if the three methods are equally good. These results leave little doubt of the superiority of eqn. (5) as a means of estimating v_0 .

The examples summarized in Tables 2 and 4 concern systems in which an enzyme exhibits progressive inactivation, a situation in which the integrated Michaelis–Menten equation will not accurately describe the reaction progress curves. Thus the observation that the median- v_0 method and the extrapolated- v_0 method have some deficiencies under conditions of enzyme instability might be considered to be an unfair assessment of their merits.

Both Cornish-Bowden (1975) and Boeker (1982) have described the performance of their methods under a variety of circumstances, but neither of these workers has examined the effect of simulated experimental error on their procedures when the integrated Michaelis–Menten equation is obeyed. For this reason, 100 sets of simulated data were calculated and v_0 was determined by the three procedures (Table 3). All the methods performed

reasonably well, with a mean v_0 within 0.6% of the true initial velocity; the principal difference between the three methods was in the variation that was observed. The standard deviation was smallest for the results obtained by using eqn. (5), somewhat larger for the median- v_0 method, and considerably larger for the extrapolated- v_0 method. A similar pattern was evident from examination of the upper and lower quartiles. For all three methods, there were roughly equal numbers of high and low v_0 values. The superiority of eqn. (5) was most clearly revealed by comparing the three methods for each individual simulation; eqn. (5) gave a value closest to the true v_0 in 56 of the 100 simulations, compared with 30 for the median- v_0 method and 14 for the extrapolated- v_0 method. This result is similar to that found for the data of Table 4.

The median- v_0 method might be expected to be more tolerant of outliers than is eqn. (5), as the latter depends on the assumption that errors in the data are normally distributed. This expectation was confirmed to the extent that the presence of outliers increased the standard deviation of the v_0 values obtained by using eqn. (5) from 0.0294 to 0.0496. By contrast, the median- v_0 method was barely affected by outliers. In spite of this, eqn. (5) still performed somewhat better than the median- v_0 method in that it gave a v_0 value that was closer to the true value in 59 of the 100 simulations.

Finally, it should be realized that the estimation of v_0 from a progress curve will constitute the first stage of a kinetic analysis; these initial velocities will be analysed subsequently by fitting to them the appropriate rate equation. In this process, it is

useful to have a measure of the accuracy of each v_0 value, so that the analysis may be weighted appropriately.

The extrapolated- v_0 method can yield standard errors if, as was done here, the data are analysed by fitting a straight line to them, although Boeker (1982) has cautioned against such an analysis. Nevertheless, the only alternative that Boeker (1982) offers is to draw a straight line through the data by eye, which will yield neither an objective estimate of v_0 nor any indication of the associated error. Cornish-Bowden's (1975) median- v_0 method, at least as he described it, offers no measure of the error associated with v_0 , although the ranking method discussed by Porter & Trager (1977) can be applied to yield a confidence interval corresponding to a standard error. In addition to its other advantages outlined above, fitting eqn. (5) to a progress curve by non-linear regression has the virtue that it yields a standard error of v_0 directly.

References

- Boeker, E. A. (1982) *Biochem. J.* **203**, 117–123
- Brent, R. P. (1974) *Commun. ACM* **17**, 704–706
- Cornish-Bowden, A. (1975) *Biochem. J.* **149**, 305–312
- Duggleby, R. G. (1984) *Comput. Biol. Med.* **14**, 447–455
- Duggleby, R. G. & Morrison, J. F. (1977) *Biochim. Biophys. Acta* **481**, 297–312
- Duggleby, R. G. & Morrison, J. F. (1979) *Biochim. Biophys. Acta* **568**, 357–362
- Porter, W. R. & Trager, W. F. (1977) *Biochem. J.* **161**, 293–302
- Selwyn, M. J. (1965) *Biochim. Biophys. Acta* **105**, 193–195
- Wharton, C. W. & Szawelski, R. J. (1982) *Biochem. J.* **203**, 351–360