On the Analysis of Solute Uptake in Circulating Diffusing Systems by a Curve Fitting Procedure

Robert Diamond
M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH

A method is presented of analysing the time course of observed solute concentration in a circulating system in terms of the reaction of the solute with a stationary bed, using the laws of diffusion. Indications are that an accuracy of about 1% is achieved.

In an accompanying paper[1] Dryland and Sheppard describe a system used for the synthesis of polypeptides.[2] The essential characteristics of this system are that it includes a column containing a gel to which is bound an oligopeptide to which a further amino acid residue is to be added. The (chemically modified) addend is introduced into the system and circulates in a closed loop, passing through the column several times. As it does so it is irreversibly removed from solution by covalently bonding to the bound peptide and, in addition, spreads by diffusion in the column, and, to a lesser extent, in other parts of the circulation loop.

The loop includes an u.v. absorption cell at which the concentration of the addend may be continuously monitored. The addend passes through this cell before its first passage through the column, so that the time-integral of the first peak on the concentration-time curve is a measure of the amount of reagent added to the system. Subsequently the concentration-time curve at the observation cell shows a series of peaks which become progressively broader by diffusion and lower by the combined effects of diffusion and irreversible uptake in the column. After a few cycles these peaks overlap each other to a large extent. It is the purpose of this paper to show how such a concentration-time curve may be analysed to give reliable measures of the uptake of reagent in the presence of diffusion.

Theoretical Background.—Let z be a one-dimensional positional co-ordinate measured along the axis of the column from a fixed origin and let ξ similarly be a one-dimensional positional co-ordinate but measured relative to a moving frame of reference so that at time t

$$\xi = z - Ut \tag{1}$$

where U is the velocity of the frame of reference. U May take any value and is not limited to the case of equality with the flow velocity, V, of the solvent. Let C be the concentration of reagent in the solution then

$$\left(\frac{\partial C}{\partial \xi}\right)_t = \left(\frac{\partial C}{\partial z}\right)_t, \left(\frac{\partial^2 C}{\partial \xi^2}\right)_t = \left(\frac{\partial^2 C}{\partial z^2}\right)_t$$
$$\left(\frac{\partial C}{\partial t}\right)_\xi = \left(\frac{\partial C}{\partial z}\right)_t\left(\frac{\partial z}{\partial t}\right)_\xi + \left(\frac{\partial C}{\partial t}\right)_z = U\left(\frac{\partial C}{\partial z}\right)_t + \left(\frac{\partial C}{\partial t}\right)_z. \tag{2}$$

Then, following Houghton,[3] we consider first the equation of continuity at a point in the solution moving with the solution, then $U = V$ and Fick's second law becomes

$$\left(\frac{\partial C}{\partial t}\right)_\xi = E\left(\frac{\partial^2 C}{\partial \xi^2}\right)_t - \frac{1}{r}\left(\frac{\partial B}{\partial t}\right)_z \tag{3}$$

in which E is the axial diffusion coefficient, B is the number of adsorbed moles per unit volume of packed absorbent, and r is the ratio volume of solvent/volume of absorbent. The last term represents exchange between the solution and the fixed absorbent. With the foregoing results this is

$$\left(\frac{\partial C}{\partial t}\right)_z + V\left(\frac{\partial C}{\partial z}\right)_t = E\left(\frac{\partial^2 C}{\partial z^2}\right)_t - \frac{1}{r}\left(\frac{\partial B}{\partial t}\right)_z. \tag{4}$$

In general, therefore, for any U equations (2) and (4) give

$$\left(\frac{\partial C}{\partial t}\right)_\xi - (U - V)\left(\frac{\partial C}{\partial z}\right)_t = E\left(\frac{\partial^2 C}{\partial \xi^2}\right)_t - \frac{1}{r}\left[\left(\frac{\partial B}{\partial t}\right)_\xi - U\left(\frac{\partial B}{\partial z}\right)_t\right]. \tag{5}$$

We now suppose, again following Houghton, that

$$B = B_0 + K_1 C + K_2 C^2. \tag{6}$$

Here B_0 represents irreversibly bound material, K_1 represents the usual linear absorption, and K_2 may be assumed to be small but, if negative, may be used to represent the onset of saturation. We then have

$$\left(\frac{\partial B}{\partial t}\right)_\xi = \left(\frac{\partial B_0}{\partial t}\right)_\xi + (K_1 + 2K_2 C)\left(\frac{\partial C}{\partial t}\right)_\xi$$
$$\left(\frac{\partial B}{\partial z}\right)_t = \left(\frac{\partial B_0}{\partial z}\right)_t + (K_1 + 2K_2 C)\left(\frac{\partial C}{\partial z}\right)_t \tag{7}$$

which, in equation (5), gives, after some rearrangement,

$$\left(\frac{\partial C}{\partial t}\right)_z + \left(\frac{\partial C}{\partial \xi}\right)_t\left[\frac{V - UP}{P}\right] = \frac{E}{P}\left(\frac{\partial^2 C}{\partial m^2}\right)_t - \frac{1}{rP}\left(\frac{\partial B_0}{\partial t}\right)_z \tag{8}$$
$$P = 1 + (K_1 + 2K_2 C)/r$$

which is true for any frame velocity U and its corresponding ξ. In particular, if we choose

$$U = V/Q \tag{9}$$
$$Q = 1 + K_1/r$$

and define

$$\lambda = 2K_2/Q \tag{10}$$
$$D = E/Q \tag{11}$$

then equation (8) becomes

$$\left(\frac{\partial C}{\partial t}\right)_\xi - \frac{UC\lambda}{1+\lambda C}\left(\frac{\partial C}{\partial \xi}\right)_t = \frac{D}{1+\lambda C}\left(\frac{\partial^2 C}{\partial \xi^2}\right)_t - \frac{1}{(r+K_1)(1+\lambda C)}\left(\frac{\partial B_0}{\partial t}\right)_z \quad (12)$$

If there is no saturation, $K_2 = 0, \lambda = 0$, and if B_0 is constant then this reduces to

$$\left(\frac{\partial C}{\partial t}\right)_\xi = D\left(\frac{\partial^2 C}{\partial \xi^2}\right)_t \quad (13)$$

which describes simple diffusion in the co-ordinate system ξ, *i.e.* moving with a velocity and apparent diffusion coefficient given by equations (9) and (11). Note that D, U, and λ are all independent of B_0 so that equation (13) holds for any constant B_0. B_0 Does, however, have a time derivative, particularly in the early stages, the influence of which we assess later.

If $\lambda \neq 0$ then asymmetries arise in the peaks because solute is swept past a saturated bed with velocity V rather than the reduced velocity U, so that high C regions of solution appear to travel faster than low C regions. This effect is studied in detail by Houghton.[3] Asymmetric peaks are seen in our system, but for quite different reasons. The observed curves can be accounted for satisfactorily assuming $K_2 = \lambda = 0$, and no further consideration will be given to the non-linear case.

A general solution to equation (13) is

$$C(\xi,t) = \int_{-\infty}^{\infty} \frac{C(\xi',0)}{2\sqrt{\pi D t}} e^{-(\xi-\xi')^2/4Dt} d\xi' \quad (14)$$

Suppose the reagent is initially confined to a narrow range of ξ' values and may be described by

$$C(\xi',0) = \frac{N}{a} e^{-\pi\xi'^2/a^2} \quad (15)$$

for some small positive a, then equation (14) gives

$$C(\xi,t) = \frac{N}{\sqrt{a^2+4\pi D t}} e^{-\pi\xi^2/(a^2+4\pi Dt)} \quad (16)$$

in which

$$AN = A\int_{-\infty}^{\infty} C(\xi,t)d\xi = A\int_{-\infty}^{\infty} C(\xi',0)d\xi' \quad (17)$$

represents the total solute present, the integration being done at constant time, A being the cross sectional area of solution in the column. In equation (15), a is the width of the peak measured from the peak centre to where C is $e^{-\pi}$ ($\sim 1/20$) of its maximum value, and equation (16) shows that the square of this width rises linearly with time. If the initial distribution is not Gaussian, equation (16) becomes an increasingly good description as $4\pi Dt$ becomes large compared with the square of the initial width.

To assess the influence of the time dependence of B_0 [equation (12)] we suppose $(\partial B_0/\partial t)_z$ is proportional to the prevailing concentration in the solution and to the number of unreacted sites, *i.e.* to $C(F - B_0)$, where F is the final value of B_0 attained at infinite time. We also recognize that N itself is time-dependent and that $(N - N_\infty)$ is a measure of $(F - B_0)$. With this approach, equation (13) should then be modified to read

$$\left(\frac{\partial C}{\partial t}\right)_\xi = D\left(\frac{\partial^2 C}{\partial m^2}\right)_t - GC(N - N_\infty) \quad (18)$$

where G is a constant.

Equation (16) is then still a solution to equation (18) when time dependence of N is recognised because the left-hand side of equation (18) now has an additional term equal to $\frac{C}{N}\frac{dN}{dt}$ and equation (18) is satisfied if

$$\frac{C}{N}\frac{dN}{dt} = -GC(N - N_\infty) \quad (19)$$

i.e. if

$$N = \frac{N_0 N_\infty}{N_0(1 - e^{-N_\infty Gt}) + N_\infty e^{-N_\infty Gt}} \quad (20)$$

The approach adopted has been to treat equation (16) as describing the shapes of the observed peaks, with an independently measured value of N for each peak which may then be compared with equation (20).

Experimental

Reagent is initially admitted to the system in a highly confined aliquot of high concentration, which passes first through the observation cell and then cycles through the reaction column and the observation cell alternately. Effectively, therefore, the cell exists at fixed points on the z scale of

$$z = z_0 + mL \quad (21)$$

where m is an integer ≥ 0 and L is the effective loop length. The observed concentration is therefore

$$C(t) = \sum_{m=0}^{\infty} \frac{N(m)}{\sqrt{a^2+4\pi Dt}} e^{-\pi(z_0+mL-Ut)^2/(a^2+4\pi Dt)} \quad (22)$$

where $AN(m)$ is the total reagent in the m^{th} peak. Let t_m be the time at which the m^{th} peak passes, *i.e.*

$$t_m = (z_0 + mL)/U = T_0 + mT \quad (23)$$

and let τ_m be the time, relative to t_m of an observation in the vicinity of t_m

i.e., $$t = t_m + \tau_m \quad (24)$$

then equation (22) becomes

$$C(t) = \sum_m \frac{N(m)\, e^{-\pi\tau_m^2/[\alpha+\beta\tau_m+\gamma m]}}{U[\alpha+\beta\tau_m+\gamma m]^{\frac{1}{2}}} \quad (25)$$

in which

$$\alpha = \frac{a^2}{U^2} + \frac{4\pi D z_0}{U^3}$$

$$\beta = 4\pi D/U^2 \quad (26)$$

$$\gamma = 4\pi DL/U^3$$

Thus the observed concentration-time curve is expected to consist of a number of peaks centred on values of t_m given by equation (23), with squared widths rising linearly with slope β within one peak, and rising stepwise with increment γ from peak to peak.

The analysis consists in fitting a curve of the form of equation (25) to the observed curve by the method of least squares in which the measured parameters are the $N(m)$ values which measure the remaining reagent and the five system parameters T_0, T, α, β, and γ which, though of little interest in themselves, must be correctly measured if the associated $N(m)$ values are to be meaningful. In practise however, the first peak, for which $m = 0$, relates to material which has not yet passed through the column even once and it usually appears to have a rather smaller β value, *i.e.* is more nearly symmetrical, than is typical of all subsequent peaks. Furthermore, the modification to the circuit tubing which occurs after the reagent is admitted to the system means that the interval T between the $m = 0$ and $m = 1$ peaks is slightly larger than the interval between subsequent pairs of peaks. Accordingly, it has been found beneficial to allow the $m = 0$ peak to have parameters α, β, and T_0 of its own, independent of the closely similar values which characterise the entire progression of the remaining peaks. Thus the software measures a total of eight layout parameters plus the coefficients $N(m)$.

At the late end of the time range, peaks are affected by the leading fringes of peaks whose centres are outside the observed range. All such fringes are taken fully into account in the analysis, the only assumption being that there is no further uptake of reagent beyond the end of the observed range, so that all $N(m)$ values are assumed equal for the last peak whose centre is in range and all later peaks. Every peak is considered at every data point provided only that its corresponding exponential in equation (25) exceeds e^{-8}.

In addition to the curve-fitting process, the peaks are integrated by ordinate summation between minima with the minima located by a change of sign of the first derivative except that such reversals are ignored if within 10 readings of the previous maximum or within 30 of the previous minimum. Ordinate sums between these minima are alternative measures of N and will be denoted $N'(m)$. Peak maxima are similarly located as a preliminary step from which initial, approximate, measures of T_0 and T are obtained and supplied to the curve fitting process to ensure rapid convergence. For a peak which is completely resolved from neighbouring peaks, especially if it is high and narrow, $N'(m)$ is a more accurate measure of its integral than is $N(m)$, and these circumstances certainly apply to $N(0)$. Accordingly, all N' values are available in the output, but, in particular, $N(m)$ values are expressed as percentages $[N(m)/N'(0)] \times 100$. This also renders irrelevant the U in the denominator of equation (25) and the A in equation (17). The later peaks invariably overlap one another to an increasing extent, and for these there is no doubt that $N(m)$ is a better measure of their integrals than is $N'(m)$.

Figure 1 is an example of an analysis done by this means. The figure shows a continuous undulatory trace which is the observed curve of $C(t)$ *versus* t scaled so that the highest observed value coincides with the 100% mark. Beneath it, dotted, are the individual components of the fitted curve corresponding to discrete m values, and to the terms in the summation equation (25). Their sum coincides with the observed curve so precisely that, on this scale, they cannot be told apart except at the right-hand foot of the first peak and on the right-hand side of the last peak where points plotting the calculated total can just be seen. The dotted line which fluctuates around the horizontal meridian of the figure is the error (observed C − calculated C) on a scale in which one unit on the vertical axis is three times the r.m.s. error, which in this case is 4.74% of the final mean level* or 1.01% of the height of the first peak. The r.m.s. error in the region after the second peak is only one eighth as great. The solid bent-line plots values of $N(m)/N'(0)$ as a percentage, for which the vertical scale runs from 50 to 110% by 10% intervals. The sloping straight dotted lines are plots of the quantity

$$\alpha + \beta\tau_m + \gamma m \qquad (27)$$

[equation (25)] for each m, the horizontal range of each line indicating the time-range over which the corresponding peak is taken into account. The first of these lines is independent of the others, as already explained, but the remaining ones have α, β, and γ in common. Equations (26) show that, ideally,

$$\gamma = \beta T \qquad (28)$$

which, if true exactly, would mean that these straight lines would run into one another end to end. Since part of the circuit is capillary tubing and a pump is present, it was considered inadvisable to assume that equation (28) would hold exactly and β, γ, and T are independently measured. The fact that the results are approximately in agreement with equation (28) suggests that the visible asymmetry of the early peaks is being correctly measured and correctly applied to the later peaks whose interpretation depends to some extent on the validity of these measures.

Table 1.

	$N'(m)/N'(0)$ %		$N(m)/N'(0)$ %	
m	Obs.	Calc.	Obs.	Calc.
0	100.0	100.0	101.72	100.0
1	78.3	79.05	78.75	78.86
2	75.5	74.78	76.77	76.67
3	75.1	73.70	76.41	76.38
4	74.3	73.41	76.47	76.30
5	73.3	73.33	76.21	76.30

This example yields the results given in Table 1. The observed values are obtained in the manner described above and the calculated values are from equation (20) with suitable values of G and N_∞.

The discrepancy at peak 0 is attributable to the steep sided character of this peak which, in the analysis, has an analytical form attributed to it which may not be justified before diffusion has operated on it to bring about this form. The remaining figures are free of this objection, yet show an increasing divergence between observed $N(m)$ and observed $N'(m)$. The reason for this is apparent from Figure 1 where it can be seen that the calculated sixth peak ($m = 5$) comes just to the right of the observed peak and of the calculated total, the difference being attributable to the fringes of the fifth and seventh peaks which contribute in this region. Since peak integrals in this region must be equal to the mean height of the observed curve multiplied by the peak separation, the $N'(m)$ values are slightly less than the $N(m)$ values because their peak separations (controlled only by the minima of the observed curve) are slightly less. It follows that the proper determination of T, and of β and γ which influence it, are important for measurement of $N(m)$, especially for large m. If the work of Figure 1 is repeated without measuring β but by setting it through equation (28) from the measured values of γ and T (with appropriate adjustments in the methods of measuring both γ and T) then $N(5)/N'(0)$ is reduced from 76.21 to 75.94% with smaller differences elsewhere and the r.m.s. error, σ, between observed and calculated curves rises from 4.74 to 4.80% of the final mean level. While the

* By 'final mean level' we mean the last peak integral divided by the peak separation.

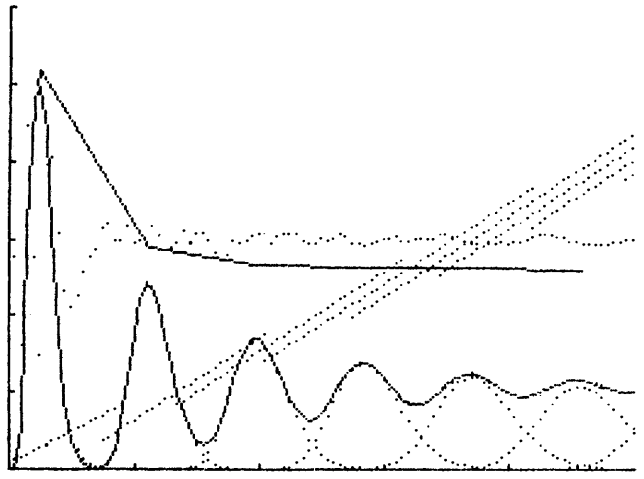

Figure 1. An example of a concentration-time curve with its interpretation. For explanation and description see text. The vertical scale has four interpretations as follows. For the curve and its components the range is 0—120% of first peak height. For the error curve it is from -9σ to $+9\sigma$. For $N(m)/N'(0)$ it is from 50 to 110%. For $[\alpha + \beta\tau_m + \gamma m]$ values it is from 0 to 12 000. The square root of this quantity gives the prevailing effective width of each peak on the horizontal scale which runs from 0 to 500 units.

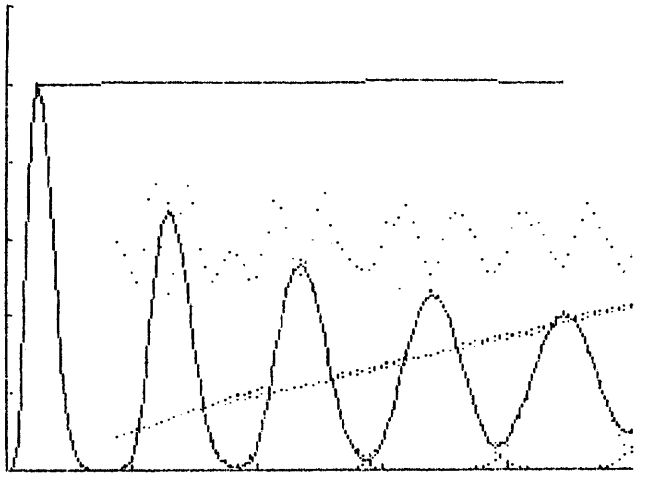

Figure 2. An example of a calibration experiment for which the peak integrals should be constant. All scales are as for Figure 1.

difference of 0.27% between these two values of $N(5)/N'(0)$ is probably not significant, their difference of about 2.5% from $N'(5)/N'(0)$ certainly is significant, with the $N'(m)$ values being systematically in error because their measurement does not take the peak asymmetry and overlap into account. Analyses are usually performed with β independently measured. Comparison with equation (28) then provides some check on the overall performance of the system.

As a further indication that $N(m)$ values are superior to $N'(m)$ values, $N(m)$ ($m \neq 0$) in this example is consistent with equation (20) within 0.1% whereas agreement between $N'(m)$ and equation (20) is nearly nine times worse, because equation (20) cannot model the lingering decline of $N'(m)$ values unless the early decline is more gradual.

Figure 2 shows the results of a calibration experiment in which the gel in the column was replaced by washed sand so that there is no uptake of reagent in the column and all peaks should

Table 2.

		$N(m)/N'(0)$ %					
m	$\frac{N'(m)}{N'(0)}$ %	$i = 5$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
0	100.0	—	101.46	102.20	102.55	102.66	102.44
1	99.3	100.15	100.20	100.32	100.73	100.07	100.05
2	99.5	100.35	100.62	100.77	100.68	100.62	100.47
3	99.7	100.52	100.47	100.57	100.45	100.52	100.56
4	100.1	100.26	100.23	100.20	100.27	100.23	100.24
σ%		3.68	5.25	5.70	5.92	5.98	5.69

Table 3.

m	$N'(m)/N'(0)$ %	$N(m)/N'(0)$ %
0	100.0	101.12
1	100.9	98.5
2	101.7	101.46
3	101.1	101.5
4	100.5	101.45
5	99.8	100.65
6	100.8	99.91
7	98.5	99.09
8	100.0	98.73
9	97.8	98.43
10	99.8	98.49

integrate to 100%. In this instance the option of beginning the analysis at the second peak ($m = 1$) has been taken, and the results are listed in the first three columns of Table 2. The r.m.s. error of the fitted curve is 3.68% of the final mean level. It is evident from the plotted error that in this instance the residual error is systematic rather than random and that, typically, the error changes sign four times within the width of one peak. This is not an indication that peak widths are mismeasured, in which case there would be two such sign reversals per peak, but rather, it is an indication that the analysis has detected a small amount of kurtosis in the shapes of the observed peaks, which is to be expected if the concentration profile of the reagent at time zero is rectangular. Normally, with a gel in the column, kurtosis is not detectable because the peaks are broader.

The software permits up to 500 uniformly spaced readings of $C(t)$ to be processed. The $N'(m)$ measures always begin with $m = 0$ and use every reading, but the curve fitting part of the program may use every j^{th} reading beginning with the i^{th}. All the foregoing results have used every fifth reading out of a possible 500, beginning with the fifth. In the last five columns of Table 2 results are given for the same data with the five possible values of i. These five determinations are equivalent but independent in the sense that no observation contributes to the results in more than one column, thus providing a good consistency check.

Finally, in Table 3 we list the results from a similar experiment in which 11 peaks are present in the analysed range. These show that, because of the increased steepness of the early peaks, that the first two are not well measured by curve fitting, that in the middle of the range $N'(m)$ and $N(m)$ values are closely similar, but that towards the end of the range the $N'(m)$ values are appreciably more scattered than are the $N(m)$ values.

The process has been implemented in Basic on a Hewlett-Packard HP85. Equal weights are given to all observations and the least-squares normal equations are solved by the conjugate gradient method without searches, which is equivalent to full matrix inversion. Typically, some four to five cycles provide

convergence. The normal equations are solved in a quarter or less of the data-scan time required to set them up.

References

1 A. Dryland and R. C. Sheppard, *J. Chem. Soc., Perkin Trans. 1*, 1986, preceding paper.
2 R. C. Sheppard, *Chem. Br.*, 1983, **19**, 402.
3 G. M. Houghton, *J. Phys. Chem.*, 1963, **67**, 84.

Received 21*st May* 1985; Paper 5/854