

Numerical Resolution of CCD-Curves

GÖRAN BLOMQUIST and SVANTE WOLD

Departments of Biochemistry and Organic Chemistry, University of Umeå, S-901 87 Umeå, Sweden

A Fortran program using correct statistical methods for resolution of counter-current distribution (CCD) curves is described. The program has been tested on CCD-curves of the isoenzymes of enolase and lactic dehydrogenase. The results show excellent reproducibility and that CCD on proteins can well be used for analytical purpose.

Partition in aqueous polymer two-phase systems is a method of great usefulness in the study of biological materials.¹ The partition behaviour depends mainly on such properties as charge and size. It is best expressed by the partition coefficient K defined as the concentration of the material in the upper phase/its concentration in the lower phase. If the K -values of two components in a sample differ sufficiently a couple of extractions are enough to achieve a satisfactory separation. However, for complex mixtures consisting of many components a multistage procedure (consisting of a large number of extractions) such as counter-current distribution (CCD) must be used.¹⁻³ CCD has been successfully used to study complex biological systems such as isoenzymes.⁴ The resolution of the distribution curves obtained can be accomplished by calculation of the theoretical curves for the various components and by subsequently fitting a weighted sum of these theoretical curves to the observed data. The use of a computer will considerably simplify this tedious task and, in addition, allow the use of correct statistical methods. Moreover several different curves can be computed for the same case at the same time. By comparing the different theoretical curves with the experimental curve the most likely curve can easily be selected. In this paper we describe a Fortran

computer program for resolution of CCD-curves and its application to some CCD-experiments with isoenzyme systems.

MATERIALS AND METHODS

The polymers used were Dextran T 500, batch No. 5996, Pharmacia Fine Chemicals, Uppsala, Sweden, and trimethyl aminopoly(ethylene glycol) (TMA PEG). The TMA PEG⁵ was a generous gift from Dr. Göte Johansson. Dextran T 500 had a weight average molecular weight of 518 000 and the molecular weight of TMA PEG was about 6000. Enolase was prepared from baker's yeast by the method of Malmström⁶ and was bought under the name Specialjäst from Jästbolaget, Sollentuna, Sweden. Lactic dehydrogenase (LDH) was prepared from pig liver. 500 g fresh pig liver was homogenized in 1 liter 10 mM potassium phosphate buffer, pH 7.3, in a Turmix mixer. The homogenate was centrifuged at 9000 g for 45 min and the supernatant was used as liver extract.

All chemicals used were of analytical grade. The water used was distilled twice in a quartz apparatus.

Enzyme assays. Enzyme activity was measured as change in absorbance at a suitable wavelength. All measurements were carried out at 20°C using a Unicam SP-800 spectrophotometer with tempered 1 cm quartz cuvettes.

Enolase was measured according to the method of Warburg and Christian,⁷ and lactic dehydrogenase according to the method of Neilands.⁸ No absolute activities were calculated.

Phase system. The phase system used contained 6.6 % dextran, 6.4 % TMA PEG and 5 mM potassium phosphate buffer, pH 6.5. In the experiments with lactate dehydrogenase 50 mM glycine-NaOH buffer pH 8.5 was used.

Counter-current distribution. The CCD experiments were carried out in the thin-layer counter-current distribution apparatus described by Albertsson.¹ Each chamber contained

0.7 ml upper phase and 0.7 ml lower phase. The enzyme mixture to be analysed was dissolved in 1.4 ml phase systems and loaded in chamber 0. Settling time used was 8 min and shaking time 30 s. 60 transfers were completed and the temperature was $22 \pm 1^\circ\text{C}$.

After the CCD, 0.7 ml water was added to each chamber to obtain one phase and each chamber was analysed for enzyme activity.

COMPUTATION

1. *Equations.* The theoretical concentration in the *i*:th tube of a component with its maximum in the *R*:th tube is²

$$\eta_i = [60!/i!(60-i)!] \frac{K^i}{(K+1)^{60}} \epsilon_R = T_{iR} \epsilon_R \quad (1)$$

with the partition coefficient *K* measured by

$$K = (R + 0.5)/(60 - R + 0.5) \quad (2)$$

Eqns. 1 and 2 are formulated for 60 transfers; in other cases the number 60 is changed appropriately. It must be noted that the model specified by eqns. 1 and 2 applies only to tubes 8–53.²

In order to completely specify the concentrations (η) of one component in the different tubes (60 in number), it is sufficient to specify the value of the parameter *R* (the tube number with maximal concentration of the component) and the relative concentration ϵ_R . Hence the resolution of a CCD curve involves (a) the selection of the number of peaks (*N*) and thereafter (b) estimation of the 2*N* parameters ($\epsilon_k, R_k, k=1,2, \dots, N$) in the following eqn. (3) so that the curve fit is optimized.

$$y_i = \sum_{k=1}^N T_{iR_k} \epsilon_k + e_i \quad (3)$$

The residuals (e_i) describe the deviations between theoretical (η) and observed concentration (*y*) values. These residuals contain both errors of measurement and “model errors” due to simplifications inherent in the model (eqns. 1 and 2).

2. *Criteria of goodness of fit.* In the investigation of several CCD curves, it was found that the ordinary least squares criterion (eqn. 4 with all weights $w_i=1$) was inadequate.

$$U = \sum_{i=1}^{60} e_i^2 w_i \quad U = \min \quad (4)$$

This inadequacy was interpreted as due to imperfections in the model for the peak shape (eqns. 1 and 2) in areas far from the peak maximum (*R* in eqn. 2). To compensate for this nonideal behaviour and, in addition, to decouple the peaks, we have instead used a weighted least squares criterion (eqn. 4) with the weights defined by

$$w_i = 0.1 + \sum_{k=1}^N e^{-0.5 \Delta_{ki}} \quad (5)$$

In eqn. 5 the summation is made over the *N* peaks (with index *k*) and Δ_{ki} denotes the distance between the *i*:th tube and the maximum of the *k*:th peak (R_k). Hence, the weight value for the *i*:th tube changes as different combinations of the peak maxima are tried in the iterative computational procedure. The weights have the values from 0.1 to ca. 1.5 at a peak maximum with two other maxima at the closest allowed distance (two points below and above, respectively).

In the computer program, however, options are provided also for the use of standard criteria, eqn. 4 with $w_i=y_i$ if the user feels that his data are behaving ideally.

3. *Optimization strategy.* The estimation of the parameters R_k and $\epsilon_k, k=1, 2, \dots, N$; is a nonlinear problem corresponding to the minimization of the criterion *U* in eqn. 4. The problem can be linearized by the specification of the R_k parameters since this makes the weights in eqn. 5 defined as well as the *T*-values in eqn. 3. The strategy employed in the present computer program is based on this linearization and consists of a systematic variation of the set of R_k -parameters until the minimum of *U* in eqn. 4 is reached. For each new set *U* and the corresponding relative activity parameters ($\epsilon_k, k=1, 2, \dots, N$) are computed by linear least squares. The following scheme of variation of the R_k -values has been used and found efficient.

1. Specify the number of peaks (*N*) and the starting values of the peak maxima ($R_k, k=1, 2, \dots, N$). This is made at the input.
2. Compute *U* (eqn. 4) for this set of R_k values.
3. Set *j*=1. This is an index which specifies the particular peak location which is currently sub-optimized.
4. Increase R_j with 1. Compute *U*.

5. Test if this new U is smaller than the last U (in step 2, 5, or 6). If "yes" save the new U -value and go to step 8.

6. Decrease R_j with 2, compute U .

7. Test if this new U is smaller than the last U (in step 2, 5, or 6). If "yes" save the new U -value and go to step 8. If "no" reset R_j to original value (add 1).

8. Increase j with 1. If j larger than N , go to step 9, otherwise to step 4.

9. If any change has been made, that is $R_j + 1$ or $R_j - 1$ has given better fit for at least one j , go back to step 3 and start a new cycle. (If the number of iterations already made is larger than a limit, usually 25, the procedure is terminated).

10. Output. (Number of iterations, parameter values, observed and calculated activity values for $i=1$ to 60, residuals and corresponding plot).

Some additional restrictions have been found practical. First, the peak maxima are not allowed to get closer than two tubes apart. Secondly, the peaks are not permitted to be closer than two tubes from the edges of the measured data.

The procedures has been programmed in Fortran for the CD3300 computer at the University of Umeå, Sweden. Program listings and a manual are available on request.

RESULTS AND DISCUSSIONS

As an example a comparison between experimental and theoretical curves of a CCD-experiment with enolase from baker's yeast is

shown in Fig. 1. Fig. 1a shows the theoretical curve (\circ), and the experimental curve (\square) when two components are assumed in the sample. In Fig. 1b three components are assumed. In Fig. 1a there is a gap between the sum of the two theoretical curves and the experimental curve around tube 32. However, when three components are assumed the two curves fit well except for tubes 49–55 which indicate a fourth component. It has previously been shown that enolase from baker's yeast can be separated into three components when subjected to CCD or electrophoresis.^{5,6}

A more complex case to analyse is a CCD-experiment of the LDH isoenzyme system. As is well known LDH from different sources is composed of up to five isoenzymes which can be separated on gel electrophoresis. In a CCD-experiment with 60 transfers the different LDH isoenzymes overlap and make the numerical resolution of the experimental curve very time consuming. By use of the present computer program the calculation of the theoretical curves is simple and the comparison between theoretical curves and experimental curves (assuming different numbers of isoenzyme components) can easily be undertaken. In Fig. 2a a comparison of the experimental curve (\square) and the theoretical curves (\circ) of a CCD-experiment assuming 4 components of LDH in the liver extract shows a good agreement. However, when 5 components of LDH are assumed (Fig. 2b) there is an excellent agreement between the experimental curve (\square) and the sum of the theoretical curves ($-$).

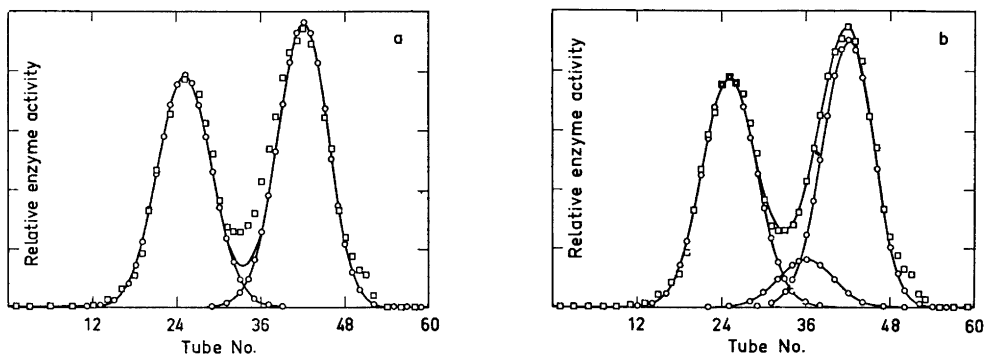


Fig. 1. Counter-current distribution of enolase from baker's yeast (\square) represents experimental curve (\circ) represents theoretical curves and full line the sum of theoretical curves. In Fig. 1a two components are assumed while in Fig. 1b three components are assumed in the isoenzyme sample.

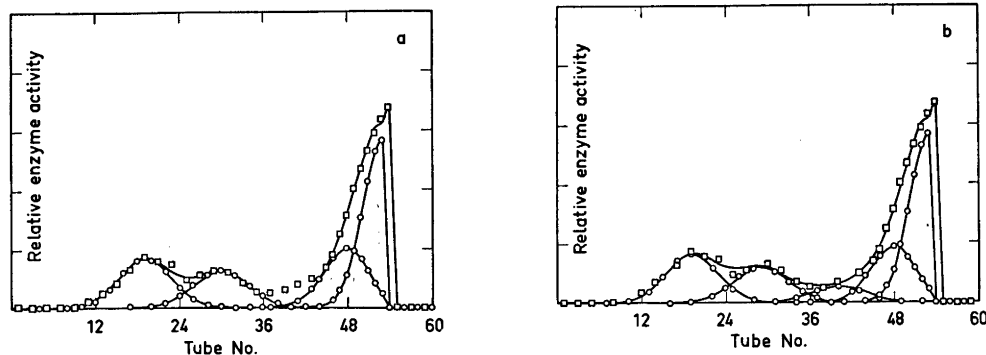


Fig. 2. Counter-current distribution of lactic dehydrogenase from pig liver. (\square) represents experimental curve (\circ) represents theoretical curves and full line the sum of theoretical curves. In Fig. 2a 4 components are assumed while in Fig. 2b 5 components are assumed in the isoenzyme sample.

To study the reproducibility of CCD experiments and of the plotting of theoretical curves three CCD-experiments were performed at different times with the same enzyme preparation. The enzyme was stored at -25°C between the experiments which were carried out at 6 month intervals. Table 1 shows the

position and amount of the different components of yeast enolase when subjected to CCD in the three experiments. The table indicates a very small variation in position of the three components from experiment to experiment. The variations are only ± 1 tube which is within experimental error. The small

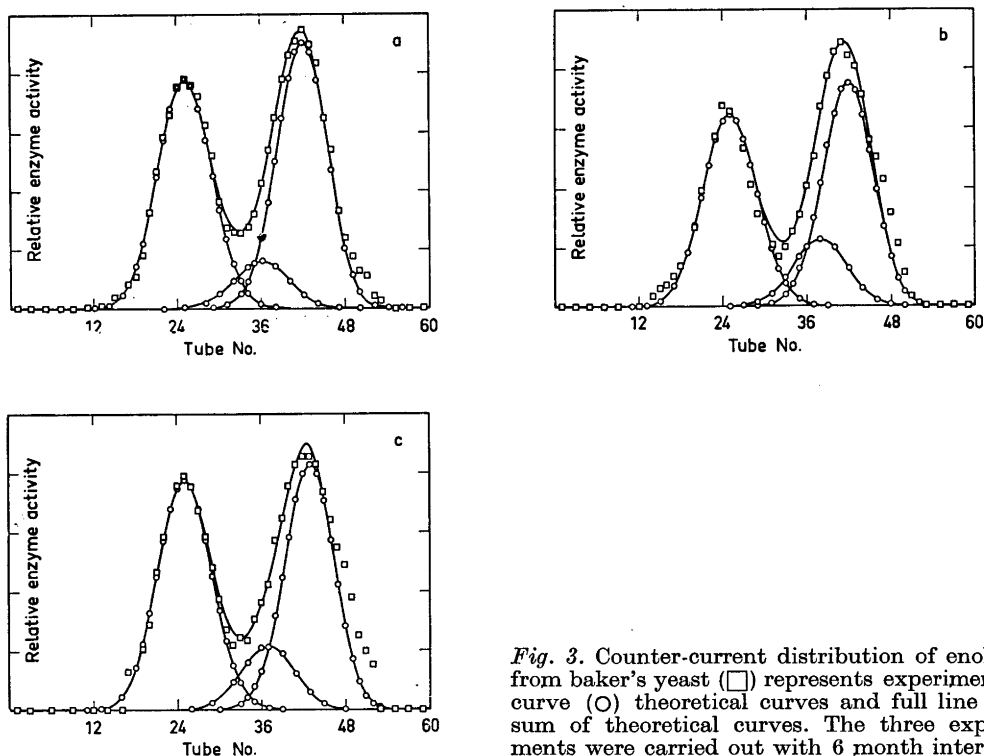


Fig. 3. Counter-current distribution of enolase from baker's yeast (\square) represents experimental curve (\circ) theoretical curves and full line the sum of theoretical curves. The three experiments were carried out with 6 month interval.

Table 1.

		CCD-experiment		
		1	2	3
Position of peak No.	1	25	25	25
	2	36	38	37
	3	44	42	43
Percentage of total enzyme activity under peak No.	1	43	43	41
	2	9	12	14
	3	48	45	45

differences in the percentage of total enzyme activity under each peak in the three experiments may be explained by changes in the isoenzyme ratios due to storage.

In Fig. 3 (a, b and c) the three experiments are plotted. The reproducibility is very good and shows the value of the method for analytical purposes. Fig. 3 also shows that a fourth component may be to the right of the third peak. From gel electrophoresis enolase from baker's yeast is known to consist of three components.⁹ It is never the less possible that a fourth component may be present in the enzyme preparation. To investigate this, further experiments with different enzyme preparations are necessary.

The computer program has been found to be excellent in the cases tested. It is not valid for calculations at the ends of the plot as can be seen in Fig. 2. To do further studies on the peak to the far right, one can change pH to make the *K*-value lower. The peak is then transferred to an area in which the computer program is valid and a theoretical resolution of the peak can be obtained.

Acknowledgement. We thank professor Per-Åke Albertsson for his interest and valuable criticism during the course of this work.

REFERENCES

1. Albertsson, P. Å. *Partition of Cell Particles and Macromolecules*, 2nd Ed., Almquist and Wiksell, Stockholm and Wiley, New York 1971.
2. Hecker, E. *Verteilungsverfahren im Laboratorium*, Verlag Chemie, Weinheim/Bergstr. 1955.
3. Craig, L. C. In Alexander, P. and Block, R. C., Eds., *A Laboratory Manual of Analytical Methods of Protein Chemistry*, Pergamon, Oxford 1960, Vol. I, p. 121.
4. Shanbhag, V., Blomquist, G., Johansson, G. and Hartman, A. *FEBS Lett.* 22 (1972) 105.
5. Johansson, G., Hartman, A. and Albertsson, P. Å. *Eur. J. Biochem.* 33 (1973) 379.
6. Malmström, B. C. *Arch. Biochem. Biophys.* 70 (1957) 58.
7. Warburg, O. and Christian, W. *Biochem. Z.* 310 (1942) 348.
8. Neilands, J. B. *Methods Enzymol.* 1 (1955) 449.
9. Pfeleiderer, G., Neufahrt-Kreilung, A., Kaplan, R. W. and Fortnagel, P. *Biochem. Z.* 346 (1966) 269.

Received September 3, 1973.