Quantitative Analysis of Electrophoretograms: A Mathematical Approach to Super-Resolution

D. A. AGARD, R. A. STEINBERG, AND R. M. STROUD

Department of Biochemistry & Biophysics, University of California School of Medicine, San Francisco, California 94143

Received September 2, 1980

A mathematical method is presented for the quantitative analysis of overlapping spots or bands taken from digitized gel patterns. The procedure is applied to both one- and two-dimensional gel electrophoretic separations.

Gel electrophoresis techniques, both one and two dimensional, have become the cornerstone of modern biological and biochemical research. They provide a rapid means of separating and analyzing macromolecules from mixtures of a few components or from whole cell extracts. Ultimately, realization of the full potential of these techniques requires that methods be developed for the accurate quantitative analysis of electrophoretograms. In the past few years several investigators have begun to develop automated techniques for quantitative analysis (1–3). This task is especially difficult in the case of two-dimensional gel patterns because of their enormous complexity. The general procedure involves first digitizing the gel electrophoretic data and then analyzing the digital image. The methods currently in use require that spots be rather well resolved for accurate quantitation. Presented here is a mathematical procedure for enhancing the apparent gel resolution: thereby separating overlapping bands or spots. As a digital representation of the gel is also required, this approach is a natural adjunct to analysis procedures already developed.

The method employed is a modified Jansson–VanCittert (4) constrained iterative deconvolution scheme. By contrast with peak-fitting approaches, no knowledge concerning either the number or the location of peaks is required; the result is thus completely unbiased. The amount of increased resolution that can be obtained with this procedure depends on how well the band or spot migration properties of the gel can be characterized. Two examples will be considered. The first is an autoradiograph of a two-dimensional gel where it was desired to quantitate two partially unresolved spots (a shoulder on a broad peak). For this example, only a limited resolution increase was required to separate the two components and hence little attempt was made to characterize the behavior of this gel system. For the second example a one-dimensional agarose gel of restriction enzyme fragments was chosen. As this problem requires a great enhancement of resolution, considerable care was taken to establish the migration behavior of bands on the gel. The approach presented here is a general one, and can be used on data from virtually any electrophoretic or chromatographic procedure.

THEORY OF THE METHOD

Inherent in any given gel electrophoresis system is an intrinsic resolution limitation that causes bands or spots to have some
shape instead of appearing as lines or points. Mathematically (using the one-dimensional case for simplicity), the gel pattern can be considered as a convolution of an ideal profile $\rho(x)$ with some smearing function $s(x)$ that serves to broaden and overlap the perfect peaks in a manner to give the observed profile $o(x)$:

$$o(x) = \int_{-\infty}^{\infty} \rho(x - u)s(u) \, du \quad [1]$$

This is illustrated in Fig. 1. What is desired is some way to reverse this convolution, thereby deconvoluting $o(x)$ to yield the ideal profile $\rho(x)$. This requires some knowledge of the smearing function $s(x)$. For the two-
Fig. 2. Autoradiographs of two-dimensional gel patterns for two different points along a time course for regulatory subunit labeling: the phosphorylated regulatory subunits are indicated by arrows in (A) and (B). Panel A shows the pattern from cells labeled for 10 min with [35S]methionine; outlined regions in (A) and (B) were scanned on a computer-controlled densitometer and are displayed in panels C and D. The marked regions in (C) and (D) were chosen for further analysis.
Fig. 3. The one-dimensional tracing derived from the region marked in Fig. 2C is shown as a heavy line in (A). The resultant profile after 50 cycles of deconvolution is depicted as the fine line in A. The starting (heavy line) and deconvoluted (fine line) profiles for the region outlined in Fig. 2D are shown in (B). The scales for the axes are arbitrary.

dimensional example \( s(x) \) was approximated by a gaussian function, while for the one-dimensional experiment \( s(x) \) was experimentally determined. Although many techniques exist for accomplishing this deconvolution operation (see Ref. (5), the best seems to be a modification of the Janssen–VanCittert iterative constrained deconvolution technique (4).

The presence of noise in the observed profile \( o(x) \) can lead to an inaccurately deconvoluted profile. This technique uses the knowledge that \( \rho(x) \) must be everywhere positive to help solve this problem. That is,
since one knows that negative optical densities are physically meaningless (although they are mathematically allowed), this information can be inserted into the deconvolution procedure. The result is a greatly reduced sensitivity to noise. Similar approaches have been employed in our laboratory for the solution of the one-dimensional phase problem in X-ray scattering of membranes (6,7) and in the solution of the X-ray crystal structure of the snake neurotoxin α-bungarotoxin (8).

The basic Jansson–VanCittert algorithm (after Frieden (5)) is

(i) \( o^k(x) = \rho^k(x) * s(x) \),

(ii) \( \rho^{k+1}(x) = \rho^k(x) + \gamma(x)[o(x) - o^k(x)] \),

(iii) \( k = k + 1 \),

and

\[ \gamma(x) = 1 - A \left| o^k(x) - A \right|, \]

where \( A \) is a constant and \( * \) denotes the convolution operation of Eq. 1. In Eq. 2 the starting point for the refinement is \( k = 0 \) and \( \rho^0(x) = o(x) \). The current guess \( \rho^k(x) \) is smeared by convolution with the gel blurring function \( s(x) \) (step i). A new guess, \( \rho^{k+1}(x) \), to the correct \( \rho(x) \) is then generated by comparison of the broadened \( \rho^k(x) (= o^k(x)) \)
with the observed gel profile \( o(x) \) in step ii. If \( o^k(x) \) is greater than \( o(x) \), as occurs when \( \rho^k(x) \) is too broad, \( \rho^{k+1}(x) \) will be sharpened by subtracting an amount proportional to \( o^k(x) - o(x) \). The applied correction is modulated by \( \gamma(x) \); when \( o^k(x) \) approaches either the lower or upper boundary limits (\( \phi, A/2 \)), the magnitude of the correction approaches zero. The entire procedure is iterated until there is no significant difference between \( o^k(x) \) and \( o(x) \). At this point, \( \rho^k(x) \) becomes an equivalent to \( \rho(x) \). The \( \gamma(x) \) function is used to apply both non-negativity and maximum positivity constraints. In practice the parameter \( A \) is set to twice some reasonable approximation of the maximum peak height (accurate selection of this parameter is not required).

The original deconvolution method was modified in two important ways: the convolution operation indicated in step i is actually performed as a multiplication in Fourier transform space, thereby speeding the algorithm about 100-fold, and the sensitivity to noise was further reduced by smoothing \( \rho^k(x) \) every 10 iterations. The convolution theorem (see, for example, Ref. 9) shows that the convolution operation of step i is equivalent to Fourier transforming both \( \rho^k(x) \) and \( s(x) \) according to Eq. [3] and taking the inverse Fourier transform of their product as in Eq. [4].
$P^k(X) = \int_{-\infty}^{\infty} p^k(x)e^{2\pi i x^k}dx$, \hspace{1cm} [3]

$S(X) = \int_{-\infty}^{\infty} s(x)e^{2\pi i x^k}dx$. \hspace{1cm} [4]

Concomitant with increasing the resolution of the gel profile occurs an increase in the apparent amount of noise. Thus there must be a trade-off between the degree of resolution enhancement and the maximum noise level that can be tolerated in the resultant profile. For this application, the deconvoluted profile $\rho(x)$ should be a smooth function since small ripples and sharp spikes are more likely to result from errors in the data than from true features. No provision was made for this in the original deconvolution scheme. A simple way of accomplishing this added noise suppression is to smooth $\rho(x)$ by coupling adjacent points.
\[ \rho'(x) = \alpha \rho(x) + \beta[\rho(x - 1) + \rho(x + 1)] \\
+ \gamma[\rho(x - 2) + \rho(x + 2)], \quad [5] \]

where \( \alpha, \beta, \gamma \) are adjustable constants and \( \alpha + 2\beta + 2\gamma = 1 \), thus for each point in the profile \( \rho(x) \) a new value is computed based on the weighted sum of it, its first \( \rho(x - 1), \rho(x + 1) \) and second neighbors \( \rho(x - 2), \rho(x + 2) \). For maximal efficiency this smoothing operation should only be applied after every 5–10 cycles of the deconvolution scheme. In the examples presented here, \( \alpha, \beta, \gamma \) were chosen to correspond to a limited gaussian function, and smoothing was applied every 10 cycles.

Quantitative resolution of overlapping spots on a two-dimensional gel. As part of a study of the metabolism of regulatory sub-units of cyclic AMP-dependent protein kinase in mouse lymphoma cells,\(^2\) it was

deemed necessary to determine the time course of regulatory subunit phosphorylation in the presence and absence of an analog of cyclic AMP. Spots corresponding to phosphorylated and nonphosphorylated forms of the regulatory subunit had been identified in two-dimensional polyacrylamide gel patterns using cyclic AMP-affinity purification of the species from cells labeled with $[^{35}\text{S}]$methionine (11), but these spots are not well resolved from contaminating species in gels of whole cell extracts ((11) Figs. 2A, B). Regulatory subunits from control cells could be affinity purified before running gels to allow quantitation of radioactivity in phosphorylated and nonphosphorylated forms, but such a procedure rests on the untested assumption that recovery of the two forms after purification will be equal. Regulatory subunits from cells stimulated with analogs or inducers of cyclic AMP do not bind efficiently to the affinity resin (R.A.S., unpublished results) rendering prepurification impracticable for such samples. The experiment was therefore performed by running gels of crude extracts of cells prelabeled with $[^{35}\text{S}]$methionine and chased for various times in nonradioactive media with or without dibutryl cyclic AMP. Radioactivity in phosphorylated and nonphosphorylated forms of kinase regulatory subunit was quantified using the deconvolution method described above to resolve these spots from contaminants. For comparison, replicate samples from the culture chased without dibutryl cyclic AMP were affinity purified and the regulatory subunit species were quantified by microdensitometry of two-dimensional gel autoradiograms as previously described (11).

Figures 2A and B show representative two-dimensional gel patterns of crude extracts from $[^{35}\text{S}]$methionine-labeled lymphoma cells. Regions of interest (enclosed in rectangle) were selected visually and scanned on a Syntax AD-1 flatbed auto-
densitometer using an aperture of $54.5 \times 109 \ \mu \text{m}$. Half-tone computer representations of scanned areas from two gels are shown in Figs. 2C and D. From these digital images, areas containing the regulatory subunit spots were selected (boxes in Figs. 2C and D) and converted to one-dimensional profiles by integrating the optical densities across the width of the rectangle and scanning from top to bottom. The resultant one-di-
menisonal profiles were smoothed using the rapid piecewise spline method of Savitsky (12). These profiles showing the contaminating spot as either the larger peak (early time point) or the shoulder (late time point) became the starting point ($o(x)$) for the iterative deconvolution procedure (thick lines, Figs. 3A and B, respectively). Resolving the shoulders into separate peaks is a rather simple problem and as such it was not necessary to carefully characterize the smearing function $s(x)$. A gaussian function was assumed for this purpose. The results after 50 cycles of deconvolution are shown as thin lines in Figures 3A and B. Following decon-
volution, it was a simple matter to separate the two peaks and to quantify the areas under each. The total procedure required less than 5 min on a small minicomputer. Figure 4 shows a comparison of the regulatory sub-
unit phosphorylation data obtained from gels of crude cell extracts using the deconvolution method with that obtained by micro-
densitometry of gels of affinity-purified ma-
terial. On the whole, the agreement be-
tween the two sets of data is quite good, and we have preliminary data suggesting that the tendency for the purified subunits to give higher phosphorylation ratios at early time points reflects a purification bias in favor of holoenzyme associated subunits (D.A.A. and R.A.S., unpublished results).

Separation of overlapping bands on a one-
dimensional gel. This experiment was de-
signed to test the limits of mathematically enhanced resolution in a system where the correct answer was known. A mixture of HindIII restriction fragments from digests of PM2 and SV40 DNA was run on an ethidium bromide agarose gel (13) under
nonresolving conditions (Fig. 5). The gel pattern was digitized and smoothed as before. Gels run under resolving conditions (not shown) indicated that the broad peaks I, II, III were each composed of two bands. The PM2 digest alone was also run on the same gel under nonresolving conditions (Fig. 6). Several of the bands from this lane were known to contain only a single molecular species and these were used to characterize the smearing function $s(x)$ for the gel. The deconvolution method used here assumes that $s(x)$ is invariant across the gel; because of diffusion effects the bands broaden as they migrate down the gel. A plot of half-width at half-height of the reference bands as a function of migration distance is shown in Fig. 7. Using this information, the $x$ axis was expanded and contracted so that all of the reference bands would be equally broad. One of these bands was chosen for the smearing function $s(x)$. The $x$ axis of the profile shown in Fig. 5 was modified in a similar manner to give the starting function $o(x)$. The results after 200 cycles of constrained deconvolution are shown in Fig. 8. Although it was apparent from the gel trace as well as the gel itself that bands II and III are separated, before processing it was not possible to ascertain that these bands were each composed of two components. After processing, the doublets were well resolved. The shoulder on the first broad peak was also resolved into its component parts. Although nonresolving conditions were utilized to set up this test, similar improvements are possible when trying to separate overlapping peaks from gels run under optimum resolving conditions.

**DISCUSSION**

A general procedure has been presented for the resolution enhancement of profiles from gel electrophoretograms. The method is fast, simple to use, and can be set up readily on any computing system. The enhanced resolution provided by this technique allows quantitative analysis of overlapping peaks. This can be important when the gel system used is at the limits of its resolution capabilities. As was shown in the two-dimensional example, this procedure obviated the need for purification of the component of interest providing benefits both by a reduction of labor and elimination of uncontrolled purification biases.

The constrained deconvolution procedure described here was chosen over other methods that may be computationally less sophisticated because the resultant deconvoluted data is considerably more accurate. This is a direct consequence of employing a nonnegativity constraint to prevent spurious nega-
Fig. 8. The profile for the region of Fig. 5 containing bands I, II, and III after correction for increased band broadening as a function of migration distance is shown (thick line). The same data after 200 cycles of constrained deconvolution are shown superimposed (fine line). The doublets in bands II and III are now resolved.

tive peaks which can lead to both qualitatively and quantitatively incorrect results. Although we have developed a two-dimensional version of this procedure for image enhancement purposes, the great computational expense makes its routine use on large gel images impractical. Simpler techniques such as those utilizing convolution kernels (see Ref. (5)) could be used to sharpen the entire image for visual inspection after which areas could be selected for accurate quantitation using the approach outlined here.

It is felt that this technique represents a natural addition to existing computer-based
processing schemes. Computational methods can simplify the task of extracting quantitative information from complex gel patterns. Ultimately, such information will be of great value in the studies of cellular regulation and development where the intensity and position of critical peaks from a whole cell extract must be resolved and compared. Such an approach should also be useful in extending the usable range of DNA sequencing gels. Fortran programs are available upon request.

ACKNOWLEDGMENTS

We would like to thank M. Sure for running the restriction digest gels and R. Bornstein for assistance in preparation of this manuscript. This work was supported by the National Institutes of Health Grant GM 24485 and by the National Science Foundation PCM77-25407 (D.A.A. and R.M.S.) as well as by National Cancer Institute Grants CA 24334 and CA 14733 (R.A.S.).

REFERENCES