Some Extensions to the GELLAB Two-Dimensional Electrophoretic Gel Analysis System

P. F. Lemkin,1 L. E. Lipkin,1 and E. P. Lester2

Several important extensions to the GELLAB 2D electrophoresis gel analysis system are discussed, including concepts that lead to the generation of consistent gel maps for a particular cell line, as illustrated by some recent results for the P388D1 cell system. GELLAB is an interactive graphics computer system that facilitates (a) image accessioning, (b) spot-data extraction from the image with position and density quantitation, (c) pairing of corresponding spots between gels, and (d) spot analysis for a large number of corresponding spots (over 3000 per gel) for up to 128 gels. Multiple gels are accessioned into the system used to create a composite gel data base. These then may be partitioned into multiple classes of gels and searched for statistically significant interclass spot differences, which can be visualized in reference-lookup images of selected spot sets as well as mosaic images of particular spots in all gels. In addition to the visual feedback that is available during the gel analysis, various statistical, numerical, and other displays may be generated in the CCELP interactive program of the GELLAB system.

Additional Keyphrases: data processing • computer analysis • canonical gel maps • gel pairing • data base management

Two-dimensional electrophoresis on polyacrylamide gels (2D PAGE) (1) has proven to be an extremely powerful investigative tool for the biomedical researcher (2). In this paper we discuss several important extensions to the GELLAB system that is used for the analysis of such gels (3–10).

But before describing these extensions, we will review some of the details of GELLAB, which is a 2D PAGE image-analysis system, realized as a set of programs that are run on DEC-system-10 or DEC-system-20 computers. Briefly, 2D gels with patterns made visible as autoradiographs or photographs are assigned sequential accession numbers in the system and then converted to digital images by one of several methods, such as with a television camera or an Optronics scanner. These gel images are then analyzed, by use of the SG2DRV program (3, 6, 9). This extracts the spots from each gel and denotes each spot by a set of features including x (isoelectric point) and y (molecular mass) location, area, and integrated density. Morphologically distinctive landmark spots are identified on each gel by using either interactive computer graphics techniques (4, 6, 9) or manual observation and data entry from plots of spots for each gel (cf. Discussion). By using the gel spot sets and landmarks, all possible spots are paired by the CMPGEL program, between each gel and a representative gel, or R-gel (5, 6, 9). The composites of such paired spot sets (across all gels) are called representative spot (or R-spot) sets. These are then used to construct a composite gel (CCGAL) data base which may be analyzed in many ways. Each R-spot set, as it is formed, is assigned an arbitrary sequential numeric name. At this stage of the analysis, GELLAB, via the CCELP interactive program, permits partitioning by spot features or classes of experimental gels. Extensive statistical, numerical, and display tools for data base treatment are also provided. Up to 3000 R-spot sets for up to 128 gels may be created with the current GELLAB data-base system.

GELLAB is a continuously evolving system for 2D gel analysis. Several extensions, discussed here, have been made to the system as previously described (5–9). These extensions, taken together, result in a system that is both easier for the biologist to use and more effective and inclusive for spot analysis and accounting.

At an early stage in spot-integrated density quantitation, some correction of spot-integrated density measurements is necessary, to take "background" density into account. Several algorithms are available for estimating corrected spot-integrated density, ranging from relatively simple subtraction of the local surrounding density to globally based density corrections. Here we present a modification of the zonal notch filter (12, 13) and illustrate how it provides a better estimate of background than did our previous technique (3).

Herebefore it was possible that for a given CCGAL data base, a set of congenic spots (i.e., spots on separate gels that represent the same thing) was not accessible to analysis, if by chance the spot was not present in the gel chosen as the R-gel. Spots that do not pair with a spot in the R-gel are designated as unresolved spots, labeled with the US symbol (4, 8, 9). The set of extensions we are presenting here correct this deficiency. In this connection we introduce the concept of the extended R-spot (eR-spot) set.

This enhanced capability of dealing with a more nearly complete set of spots from gel to gel allows us to pursue the idea of a canonical gel, "C-gel" (in the sense of a standard catalog) and to approach this theoretical concept more closely than heretofore (4, 6, 9). We are now able to introduce the concept of a C-gel, an estimate of the canonical gel based on the averages of R-spot sets of replicate gels. The procedures used in creating the C-gel provide for the exclusion of spots that are poorly represented in any replicate set and of most "noisy spots" and other artifacts.

Finally, GELLAB provides a greatly expanded and explicit package of set theoretic operations, which may be applied to sets of spots delineated by particular statistical conditions or constraints. We demonstrate the utility of these operations for identifying interesting biological events in multidimensional (i.e., multiple experimental class) data bases.

1 Image Processing Section, Division of Cancer Biology and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205.
2 Department of Medicine, Section of Hematology/Oncology, University of Chicago Medical School, Box 420, 950 East 59th St., Chicago, IL 60637.
3 Address correspondence to C.F.L. at: Park Bldg., Room 417, N.I.H., Bethesda, MD 20205.
4 Nonstandard abbreviations: 2D, two-dimensional; PAGE, polyacrylamide gel electrophoresis; CCGAL, composite gel; US, unresolved spots; R-spot, representative spot; eR-spot, extended R-spot; C-gel, canonical gel; C-gel', an estimate of the canonical gel based on the R-spot sets of replicate gels; DP, distance between two paired spots mapped to the same domain; EP, extrapolated pair; ID, the mean R-set set-integrated spot density; GSP, gel segmentation file; GCF, gel comparison file; LMS', landmark set; and SRC, Search Results List.
Methods

Extension of CGL Data Base to Include Extended R-spots

Figure 1 illustrates the case in which certain spots are present in a set of gels, but not in the R-gel. As mentioned, spots that do not pair with those in the R-gel are called "unresolved spots" (US). The extended R-spot set construction handles this missing spot problem by creating eR-spot sets based on the US spots.

The initial R-spot data base is first constructed to contain all R-spots as described in references 5 and 9. It does not contain these missing US spots at the initial stage. The set of paired spot files is then scanned a second time for those US spots not in the R-gel that are "robust"—i.e., that have reasonable values for selected spot features (total area, integrated density, and density range, i.e., max-min optical density or absorbance) within the limits set by the experimenter. When a non-R-gel US spot is read from a paired spot list file, it is first tested to see whether it belongs to an existing eR-spot set. Its coordinates are then mapped to the domain of the R-gel where the nearest existing eR-spot coordinates are found. The mapping function is simply the linear landmark transformation for mapping a landmark spot in the gel to the same landmark in the R-gel. The particular landmark transformation used for a given US spot is that of the (nearest) landmark set to which the US spot belongs.

For example, given a spot's centroid (density weighted center) and landmark (x,y) coordinates for its (nearest) landmark set k in both the R-gel and in gel g, the linear transformation to the domain of the R-gel for this gel g spot is defined as follows. Let (x,y)LM[k] be the landmark coordinates of gel g landmark k. Then the mapped centroid is defined in vector notation,

\[
(x,y)_{R-gel} = (x,y)_g + (x,y)_{LM[R-gel,k]} - (x,y)_{LM[g,k]}
\]

The (x,y) is then tested to determine whether the spot lies within DP limits of the nearest eR-spot in the R-gel, DP being defined as the distance between two paired spots mapped to the same domain (i.e., R-gel), (cf. 4, 8, 9). If the US spot meets this test, it is then put into that eR-spot set. If not, a new eR-spot set is created along with an extrapolated pair (EP) spot label with zero density for the R-gel itself. Note that the number names denoting eR-spots are treated the same as those of normal R-spots for all purposes in analyzing a CGL data base. However, eR-spots have only EP and US spot-pairing labels.

Extension of CGL Data Base to Generate and Use the C-gel

One can compute an estimate of a C-gel (5-9) from the CGL data base under particular conditions. This estimate, called the C-gel, is usually derived from a set of replicate gels. By "replicate" we mean 2D gel patterns produced from the same sample or from parallel tissue cultures, run concurrently under the same conditions. Having created such a CGL data base, feature sizing as described earlier (5, 9) is used to discriminate spots from noise. R-spot sets that meet this sizing test are called "C-spot" sets. For example, requiring that 80% of the gels contain the given R-spots and that the coefficient of variation (CV) of R-spot set density be less than some small value would find robust C-spot estimates in most of the gels. The C-gel is defined to have the same accession number as the R-gel in this replicate gel, set, but the accession number (3) extension (i.e., decimal fraction part) is defined to be "9" (e.g., R-gel accession number 250.2 would have the derived C-gel accession number 250.9). (Otherwise, the accession number extension is usually used to distinguish among different radiographic exposures for a given gel.)

A composite (synthetic) GSF file for a set of replicate gels is then produced, in which the mean R-spot set centroid (x,y) mapped onto the R-gel is used as well as the mean R-spot set-integrated spot density D' [background-corrected original density measurement in optical density units (3)] and the area. The standard deviations of these variables and the number of gels per R-spot set are also entered into this file for later use in building (and for performing statistical tests with) a new CGL data base that is based on the C-gel. Spots are renumbered sequentially from 1 where only those spots meeting the feature-sizing criteria are included. This procedure uniquely defines a set of C-gel spots. Figure 2 illustrates the mapping of replicate gels onto the R-gel during creation of the C-gel.

During C-gel generation, the landmark data base (4, 9) is searched for matching (R-gel, gel g) entries for all gels g—i.e., matched against the same given R-gel. (The landmark data base consists of a list of landmark sets for various R-gel/gel-g pairs. A particular landmark set of paired spots for each gel g paired with R-gel typically contains 10 to 25 pairs of specific spot (x,y) coordinates for the two gels). A modified duplicate
of each matched landmark set is inserted into the landmark data base file, where the R-gel name is changed to that of the C-gel, and the mean centroids of the C-spot landmark spots are substituted for the corresponding R-gel landmark spots. The C-gel analysis typically proceeds as follows.

**Constructing the Initial C-gel**

1. Accession a data base consisting of replicate control gels.
2. Landmark all gels to one of the replicate gels (i.e., define one as the R-gel). Segment all gels to get the spot feature lists in the gel-segmentation files (GSP) (3, 6, 9).
3. Build a CGL data base (5, 6, 9) from these replicate gels by pairing these gels to produce gel-comparison files, or GCF (6, 9). These GCF files are then used to build the CGL data base consisting of R-spot and ERF-spot sets.
4. Pick R-spot set feature sizing parameters in order to recognize robust R-spot sets to be defined as C-spot sets.
5. Create the C-gel synthetic GSP and LMS data base files as detailed above.

**Analyzing New Experimental Gels in the Context of the C-gel**

1. Accession new experimental gels.
2. Landmark these to C-gel (visualized in and using the R-gel for landmarking purposes).
3. Segment (i.e., extract spot features for) these gels into a set of GSP files. Then, pair spots on these gels with the previously generated C-gel GSP file, resulting in a set of GCF files.
4. Build and analyze a CGL data base by using these new GCF files.

Once the CGL data base is constructed, it has the following characteristics, illustrated in Figure 3: (a) all C-gel spots are normal R-spot sets (i.e., not ERF-spots); (b) all experimental gel US spots are in ERF-spot sets (i.e., spots missing in C-gel); (c) all paired (paired with C-gel synthetic spots) experimental spots are in the normal R-spot sets; (d) any C-gel spots not in the experimental gels are normal R-spot sets consisting of one US spot (i.e., spots missing in the experimental gel); (e) the C-gel may be used later with additional replicate gels to build a new C-gel estimate. This iterative process gradually may lead to better and better estimates of the canonical gel; and (f) if all gels of a particular domain are merged together, then a C-gel result that represents those spots found in all of the gels, or at least in a given percentage of them (as defined by the user).

When the C-gel data are read back into the CGL data base during the creation of the experimental data base, the standard deviations of C-spot (x centroid, y centroid, density, area) and the number of gels per C-spot set estimate are saved for each C-spot. Their mean values now become the C-spot estimates for these features. These features have been passed from the synthetic GSP file through the synthetic GCF files now being used to create the CGL data base. These data (means, standard deviations, and number of gels per C-spot set) are of course available for use in any statistical test with the experimental gels that later is performed in CGLBP.

To summarize, a single partition exists whereby missing spots in either gel may easily be identified as being US spots in R-spot or ERF-spot sets corresponding to spots missing in the experimental or C-gel spots, respectively. By definition, all ERF-spot set numbers are those higher than the last normal R-spot set number—either of which may be analyzed by the user at any time in a given CGL data base.

A special case occurs when a data base is built by use of the C-gel and the original replicate gels. To see how well one did in constructing the C-gel, one can build a CGL data base with it and with the original replicate gels. Any ERF-spot sets that occur are US spots where the original R-spot or ERF-spot sets were sized out during the creation of C-gel. There can be no US spots in the un-extended R-spot sets.

**Use of Sets in GELLAB**

GELLAB permits manipulation of sets of gel names, first mentioned in ref. 5. These may be defined and manipulated with the usual set theoretic operations of union, intersection, and non-commutative set subtraction (11). This is useful when one is investigating many gels consisting of several observational classes or in redefining the working set of gels by subset name(s). The working set of gels is defined to be a subset of the set of all gels in the current CGL data base. Similarly, lists of selected R-spots may be manipulated. Such a list is called the Search Results List (SRL) and is generally the result of applying some statistically driven search over the R-spot CGL data base (5, 8, 9). The tools available on which search strategies may be based include multi-class t-, F-, and rank-order tests as well as R-spot set feature sizing against user-specified limits. The particular subset of gels considered for a given search may be altered by the user specifying the current "working set of gels" before the search. Because each element of the SRL appears only once, it may be treated as a set, e.g.,

\[ \{ R\text{-spot}, R\text{-spot}, \ldots, R\text{-spot}\} \]

An example of an R-spot set (cf. Figure 5) is:

\[ 48, 58, 60, 162, 220, 223, 230, 231, 233, 592 \]

SRL subsets may thus be defined as the current SRL (possibly the result of the last search), explicitly by the user or as the result of one of the usual set theoretic operations (union, intersection, non-commutative subtraction). Any SRL subset may be used to redefine the current (i.e., the "working") SRL.

The set of R-spots so defined may be used for all of the operations available for the SRL (5, 8, 9). These include: (a) creation of data files for statistical analysis with a non-GELLAB statistical package such as SPSS (21), generation of R-map images or plots where spots of interest are labeled on the image, and (or) generation of mosaic images or plots consisting of panels surrounding a particular R-spot for all gels in the data base; (b) computing the ratio histogram of selected spots (the ratio is of the mean densities of two classes); (c) printing the rank-order table (plotting spot densities for each of the

---

**Fig. 3.** Example of the occurrence of spots between the synthetic C-gel and a set of experimental gels (G1 to Gn). Spots A, B, and C are R-spots found in C-gel. Spot D is an ERF-spot not found in C-gel. Therefore, spot C would be a spot found only in the experimental gels.
gels as a function of R-spot; and (d) normalizing or re-normalizing the CGL data base by use of the total D’ of R-spots in the SRL for each gel.

Use of Search Results List Set Operations in Analyzing the Data Base

Two experimental designs are given here to illustrate the application of set theory considerations to SRL subsets.

Example 1. In analyzing a multi-dimensional gel data base (i.e., several different classes of observations of experimental gels), spots found in one pairwise classification might be compared with spots found in another. For example, (Control vs Treatment 1) vs (Control vs Treatment 2). Spot differences found in the gels after both treatments are the intersection of the two searches. Spots found only in Treatment 1 are the set difference computed by removing Treatment 2 spots that are also present after Treatment 1. Spots found after either treatment are the set union.

Example 2. In a time-sequence or drug-concentration experiment, one might ask questions about which spots first appear or increase, increase some more, and then decrease or even disappear—signaling gene expression under a particular set of conditions. The CV for R-spot set density would be large at the transition zones—and possibly large in between, depending on the shape of the curve for spot density. The case where the CV is stable in the growth curve can easily be detected with GEELAB by searching for spots for which the CVs are small. Changing spots can be found by searching with a high CV value constraint (R-spot sets with a large CV but a small mean D’ can be ignored to help decrease the false-positive rate). Taking this last case, one could find the sequence of gels with such variable spots as follows.

(1) Given k gels (or gel experimental classes, possibly with multiple samples in each class) in the sequence, find the intersection of SRL subsets for each two sequential classes, with the constraint that a spot is considered if its CV exceeds some specified value. The resulting SRLs are saved as k – 1 SRL subsets

\[ S_1, S_2, \ldots, S_{k-1} \]

That is,

\[ S_i = SRL_i \cup SRL_{i+1} \]

---

Fig. 4. R-map image of C-gel' for eight replicate gels of a P388D1 mouse macrophage-like cell line
The replicates were from eight tissue-culture bottles. The spots included in the C-gel' are marked (in white, here) with "*" and were those present in at least six of the eight gels.
Table 1. Two-class Ratio Histogram Table of Selected R-Spot Sets from the t-Test Search at .95 Significance *

Found 10 R-spots, mean SD/mn (for entire R-spot set) = .35 ± .11

<table>
<thead>
<tr>
<th>m2/m1</th>
<th>R-spot sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.40</td>
<td>162 231</td>
</tr>
<tr>
<td>1.45</td>
<td>233</td>
</tr>
<tr>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>1.60</td>
<td>48 592</td>
</tr>
<tr>
<td>1.65</td>
<td>223 230</td>
</tr>
<tr>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>58 220</td>
</tr>
<tr>
<td>1.80</td>
<td></td>
</tr>
<tr>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>2.15</td>
<td>60</td>
</tr>
<tr>
<td>2.20</td>
<td></td>
</tr>
</tbody>
</table>

* The ratio m2/m1 is computed from the mean density values of the two classes T9 and T24 (class 1 and 2, respectively) for each R-spot set selected.

(2) For each S<sub>j</sub>, compute the sequential intersection S<sub>iw</sub> as follows:

\[ S_{iw} = \text{UNION } S_j \]

\[ j = i \text{ to } i + (w - 1). \]

Thus S<sub>iw</sub> is the intersection of a sequence of subsets starting at S<i>1</i> subset i and looking forward through w subsets. This sequential intersection may be computed manually in GELLAB by using a composition of the UNION S<i>1</i> subset operation on two sets at a time. It may also be more easily specified using the SEQUENTIAL INTERSECTION operator, which takes two arguments: starting S<i>1</i> subset number i and number of subsets width w. A further extension of the sequential intersection operator computes all of the S<sub>iw</sub> for g S<i>1</i> subsets and for i ranging from 1 to (g - 1) - (w - 1)). By applying this last operation for several sizes of w (i.e., w = 2, 3, ..., g - 1), any occurrence of spots appearing in any S<i>j</i> through S<i+w−1</i> for any interval w is detected.

Use of the Zonal Notch Filter for Estimating Gel Background Density

In previous papers (3, 14-18, 22), various methods have been used for estimating gel image background density. Knowing this local background density value enables one to correct the measured integrated density D to an integrated density D' by using the transformation [valid for small maximum picture element (pixel) optical-density values]:

\[ D' = D - (\text{mean background density per pixel} \times \text{total spot area, in pixels}). \]

In the new method described here, information about extracted spots is taken into account when background density is estimated. The technique involves use of the zonal notch filter (12, 13) by considering pixel-density values that lie within a particular pixel density range (i.e., "zone") to be the background region. The method, as modified here, re-defines this density zone as that part of the gel image where there are no spots. This definition has the advantage over the classic zonal threshold algorithm of being dependent on subject domain (i.e., spot context). Both major streaks and slowly varying subtle background shading are considered to be background. At each pixel in the image, an average of N × N neighborhood pixel density is computed for those pixels in the square that are not part of extracted spots. Processing proceeds line by line, left to right, adding the next right-most column sum and subtracting the left-most minus one column sum (keeping track of the number of non-spot pixels per window). Thus the N × N pixel average need not be computed explicitly at each pixel. Whether a pixel is a background or resolved-spot pixel is assessed by testing against the propagated central core image (3, 6, 9) previously generated during gel spot segmentation. The zonal notch filter value at the centroid position of a spot is used as the estimate of its background density. An example of the application of the zonal notch filter is shown in the Results.

Fig. 5. R-maps of (a) set of landmark spots used in pairing of the data base and (b) selected R-spots in a set of P388D1 macrophage-like cells that showed .95 significance in mean density difference in particular spots over time as defined by the t-test.

The two classes are T9 and T24 (24 h later).
Results

The following results are those we found for differences over time in the P388D1 mouse macrophage-like cell line (8, 23). This cell line is extensively used in our laboratory to study the cytotoxicity of durable fibers such as asbestos, aluminum oxide, and other fibers. The cells were grown in tissue culture in L-leucine-deficient Fischer's medium (GIBCO no. 80-0039) with 50 μCi of 14C-labeled L-leucine added per culture bottle. Initially half of the 16-culture set was harvested at time T₀ and the other half 24 h later (time T₀). The 100 g/L polyacrylamide gels were run in an Ampholine pH range of 3–10 according to the O'Farrell technique (1), by David Goldman and Carl Merril of NIMH. One-week exposures of the autoradiographs were used.

Figure 4 shows a C-gel R-map image of the eight T₀ replicate gels with all spots marked that were included in the C-gel. The sizing constraints were that an R-spot had to be in six of the eight replicate gels in order to be considered a C-spot.

We computed a two-class ratio histogram showing mean densities for the R-spot sets found by the t-test search of the P388D1 CGL data base at 95% significance. For each R-spot set resulting from the search, the mean density m₄ for spots in class 2 (T₂₄) gels is divided by the mean density m₁ of spots class 1 (T₀) gels. The R-spot set name (R-spot number) is entered in the histogram table. Table 1 shows the histogram data for the above t-test.

Figure 5a illustrates some of the R-spot sets found to differ statistically significantly with time during one of the t-test searches. Figure 5b illustrates the set of landmark spots, labeled with letter names, used in spot-pairing between the gels. Figure 6 illustrates mosaic images of some of the spots found in the t-test. Figure 6a shows a spot (landmark D) where the polypeptide decreased with time. Figure 6b through 6d illustrate polypeptides that increased with time. In Figure 6d the spot was in fact missing in two of the gels; the spots were extrapolated in these gels (in the upper leftmost two panels) by use of the CGSPLP program. Such extrapolated spots can be used to aid in verifying that the missing spots were either not there, were very light (i.e., below detectability with the visualization system), or were mis-paired.

Figure 7a illustrates an original P388D1 T₀ gel image, 7b its segmented (i.e., extracted spots) image, and 7c the background image constructed by subtracting the segmented images.
Fig. 7. Example of the modified zonal notch filter used to estimate gel image background density
(a) original image of a P566D1 Tg gel; (b) segmented spot image of a, showing spots extracted; (c) difference image computed by subtracting b from a; (d) zonal notch filter of size 16 X 16 pixels computed for c, where zero (white) pixels are ignored in computing the average; (e) zonal notch filter of size 32 X 32; and (f) original image a, less 32 X 32 zonal notch filter e
Fig. 8. Example of an R-map plot, generated by QELLAS program DWRMAP, of T9-P388D1 gel used for landmarking.

Notice that the spots with labels are the largest (i.e., darkest) spots in the image. The numbers in the table on the right are the spot (CC number and x, y, and D') values, in order of integrated spot density D'. The CC no. is the connected number of the component or spot relative to a particular gel segmentation. "Landmarking" two gets from these plots is done by manually tabulating a list of CC numbers for corresponding landmark spots. These lists can then be used to update the landmark data base by using the QELLAS program LMSEDIT.
shows the result of subtracting the 32 x 32 zonal notch filter background image from the original image (Figure 7a) resulting in minimized background variation.

**Discussion**

With such extensions to GELLAB as 

- eR-spot sets and the 
- C-gen, it is now possible to investigate experimental gels more reliably than before and to contrast these experimental gels with a more stable estimate of the canonical gel map than the R-gel. Because the C-gen can be iteratively computed as new replicate control gels are produced, successively better estimates of the C-gen can be obtained, if the biochemistry involved and gel-production technology are stable and reproducible. GELLAB can currently accommodate up to 128 gels. This large capacity should not prevent one from considering the possibilities for decreasing the number of replications, both of controls and experimental gels required for statistical validation. In addition to the prospect of improved biochemical technique, leading to increased precision, another interesting facet is apparent. The C-gen construction allows, with appropriate (still to be determined in detail) statistical treatment, the iterative improvement of the parameters of the C-gen—our approach to the model or canonical gel.

By thinking of selected spots in terms of sets and set operations, the biologist more easily can handle time- or dose-dependent variables in experiments involving sequences of gels. In such cases, each step in the sequence is defined as a different class, and the problem is recast as a problem in multiple-class analysis.

There still are problems in the reliable detection and accurate measurement of very light, very fuzzy, or noisy spots. These difficulties are reflected in false positive or false-negative results. Although it is possible with this system to measure most spots that are visible to the human eye, in practice the uncertainty involved in evaluating such marginal spots precludes their consideration in an automated quantitative system. That is not to say that their qualitative consideration should be ignored or that improvements (through the use of replicate gels) cannot be obtained.

In addition, there is occasional mis-pairing. However, recent improvements in pairing algorithms (19) as well as those in progress in this laboratory should reduce this type of error.

Such is the generality of the various stages of analysis in GELLAB that, with little effort, data prepared by non-GELLAB programs could be used in succeeding stages of analysis with GELLAB. For example, alternative derived procedures (e.g., manually measured or other segmentation and pairing algorithms) could be used in preparing paired-spot lists. These in turn can be read by GELP to construct a data base. At this point, the analysis would continue as with GELLAB-prepared data.

GELLAB currently quantitates spots in terms of their total integrated density, which is expressed in optical density units relative to a given neutral-density step-wedge standard that is scanned along with the gels (it is visible at the bottom of some of the figures). However, with relatively minor changes in protocol (incorporation of radioactivity standards into the autoradiographs) and in some of the programs, results could be obtained in counts per minute. This is one of the changes we plan for 1982.

**Current Status of the GELLAB System**

GELLAB runs on the IPS DECSystem-2020 (under the TOPS-10 monitor) where picture input/output is provided by special hardware added to the system. However, in order to generalize and export GELLAB to other laboratories without special hardware, we have written several magnetic tape (mattape) and file-conversion programs, and these are being used to facilitate image acquisition and image transfer on other DECsystem computers. In particular: a mattape package exists to read Optronics scanner images produced by a PDP11 RT11 mattape system and convert them to GELLAB picture file format; a general-purpose image-conversion program converts various picture-file formats to or from the GELLAB picture-file format.

Because GELLAB is written in the SAIL, programming language (24), it runs only on DECSys-tem-10 or DECSystem-20 computers (Digital Equipment Corp., Maynard, MA 01752). However, these medium-to-large main-frame time-shared computers currently are available at many places in this country, including some where time may be rented.

The hardware-independent portion of GELLAB has been supplied to the DCRT NIH computer center TOPS-10 system. In addition, the major portion of GELLAB has been supplied to the University of Chicago, running there on a DECsystem-2060 under the TOPS-20 operating system. Additional work is in progress to further image-exchange and graphics for the TOPS-20 version.

It is possible to use graphics plots for R-maps, mosaics, and landmarking rather than grey-scale images. One can thus use GELLAB without expensive raster grey-scale display hardware, although such grey-scale displays are the preferred mode of operation. Figure 8 illustrates an R-map of the gel-segmentation file produced by the gel-spot segmenter. Spots are represented as circles, the size of which is proportional to the spot's integrated density. By finding corresponding spots between R-maps for several gels, landmarking may be performed manually instead of by using the interactive grey-scale display hardware (22).

The current version of GELLAB is experimental and will soon be released to selected TOPS-10 DECsystem sites. Interested parties having access to a DECSys-tem-10 or -20 computer should contact the senior author for information regarding obtaining a free mattape copy of the runtime programs.

The constant help afforded by Morton Schultz, Bruce Shapiro, and Earl Smith, our colleagues in the Image Processing Section, has been invaluable. Marta Wade, also of the IPS, ran the tissue cultures on the F388D1 cell line. Our collaborators Snorri Thorsteinsson and Peter Wirth of NCI, and Carl Merrill and David Goldman of NIH have provided stimulating ideas and critical evaluation of the methodology as it has developed as well as running the 2D gels for the F388D1 experiments.

**References**

20. Goldman, D., and Merrill, C., Personal communication.