Abstract

It is theorized that a comparison between the protein composition of normal versus cancerous tissue may provide a clue to the underlying cause of the cancer. Two-dimensional (2-D) gel electrophoresis provides a method of separating the proteins for each tissue into a 2-D map, called an electrophoretogram. Examination of the map for a normal tissue overlayed onto that of a cancerous tissue is critical to this research. It is thought that some cancers arise due to a defect in the DNA. This defect may then produce proteins that are also defective. These defective proteins might then go on and become responsible for the cancer. Therefore, if we can identify the proteins responsible for a particular cancer we may be able to prevent or stop the cancer by eliminating the defective protein and substituting the proper one.

The first step in the process of identifying what proteins are involved in a particular cancer is to take a normal or cancerous tissue and extract the proteins from it. One then can perform a procedure to separate the individual proteins. 2-D gel electrophoresis is one such technique and will move the most. Isoelectric focusing causes proteins to migrate through acrylamide gel in response to an electric field so that the more highly charged (acidic) proteins will move the most. SDS gel electrophoresis causes proteins to move in the other axes direction as a function of molecular weight. The net result of the 2-D gel electrophoresis is the map, or electrophoretogram, and is a kind of "fingerprint" of the protein composition of the tissue sample.

The Visage system is a commercially available system from Bioimage in Ann Arbor, Michigan that facilitates the analysis of the electrophoretogram. The system includes several components that allow the scanning of the 2-D pattern, quantification of the spot (protein) data, databasing, and terminal display. The terminal display system is based on a MASSCOMP microcomputer which can display images from the database in a variety of ways. While acceptable for visual display the system does not produce any hardcopy output, which is a serious drawback. Also, the system does not allow overlaying in a manner useful for our scientists. In addition, there are frequently over 1000 spots or proteins in a sample and a relatively small screen display can hardly provide the resolution necessary to determine differences between the proteins of cancerous and normal tissues.

The system, however, can compute various indices on each "spot" which describe its size, location, and density and it can transform the coordinates of one map to the coordinates of another based on a series of reference points. This is extremely important for comparative viewing. There is also a tape drive component to the Visage system and one can obtain the protein spot indices on the mainframe for further processing.

Where do we go from here?

It was felt that we needed to develop a mainframe-based system to do some of the important things which the Visage system did not do, such as produce hardcopy output. However, if one must write an extensive plotting program using TEMPLATE, for example, the benefits may not be worth the effort required. Therefore, some investigation was required. Interestingly, several people had already discounted SAS software as a potential solution (this notion was based on pre-Version 5 SAS software).

The specifications for the envisioned system may be broken down into two categories. One is criteria that are absolutely essential. The second includes things that are desirable but not essential.

Essential criteria include the following:

1. Superimpose (overlay) 2 plots
2. Generate a graphical measure of density
3. Draw a spot number "close" to the spot
4. Hardcopy output
5. Control size of spots and number

Things that would be nice included:

1. Produce large 36"-wide plot
2. Produce a reasonable facsimile of the original
3. Store and retrieve plots without regeneration
4. Expand selected sections of the plot
5. Select colors for display of spots and numbers

After some investigation (and several arguments) it was decided to use SAS software to develop the application. Version 5 SAS software, as it turns out, gave us virtually everything without
requiring a great deal of development time.

How Version 5 SAS Software Met the Needs

For many years, SAS software has provided tools that allowed easy data entry, manipulation, analysis and plotting. With the advent of Version 5, SAS software provides an incredible array of new features that allow even more power and flexibility to develop truly easy-to-use applications that meet many needs. And perhaps the most remarkable thing of all is the relatively small amount of development time that it takes to develop these systems.

Since this project came at a time when Version 5 was only recently released it is not surprising that people were unfamiliar with the new tools that were available to the SAS programmer: SAS/AF, the SAS/GRAPH annotation facility, and PROC GREPLAY. These remarkable tools, which in my estimation are the most important developments in SAS since the introduction of SAS/FS® or maybe even the introduction of SAS76, solved all of our problems in a reasonable amount of time.

Figure 1 shows basically how all the major pieces of the application fit together. SAS/AF is used to govern the interaction with the user and allows them easy access to all the features of the system, which include the SAS/GRAPH routines, PROC GANNO and PROC GREPLAY. Figure 2 shows a more detailed flowchart illustrating more of the functional aspects of the system. Notice that startup of the SAS software is governed by a REXX exec which allows us to operate the application by initiating SAS/AF via the PROC DISPLAY command.

SAS/AF is a remarkable tool that allows a large variety of functions. This application utilizes three SAS/AF screen types: HELP screens; MENU screens; and PROGRAM screens. PROC BUILD allows us to access the SAS/AF catalog (see figure 3), which contains all the HELP, MENU, and PROGRAM screens involved. The SAS/AF manual adequately describes the construction of these three types of screens and the reader is referred to these for more details. The most important part of SAS/AF in this application is the ability to define a user portion of the PROGRAM screen. This allows selective execution of sections of SAS code corresponding to options the user has selected. Access to an IBM 3179G allows one to create applications using color, blinking, reverse video and underline attributes. Truly visual attractive applications are possible. Another nice thing one can do with SAS/AF is to include an exit routine to bounce totally out of SAS(1) and a hidden option for a circle to represent the spot. The new routines also allow fast branching to particular sections of the system macros DCALL, DERRON, DKEY, DMSG, DCURSOR, _DALARM, DLASTC, and DCMND allows one to have a high level of control of the program screen. One can, for example, check an entered data set and make sure that it exists. If it doesn't exist the user could be notified so that they can then make a proper entry. The >> and >>> indicators allow fast branching to particular sections of SAS/AF. See references (2, 3, 4, 5) for complete documentation.

The plot program screen performs a variety of duties and functions. It takes the spot indices from the Visage system and computes a radius for a circle to represent the spot. The new annotation facility of Version 5 is then used to generate data which is used in a PROC GANNO step to plot the simulated electrophoretograms as seen in figure 4 (6,7). A great help in generating the annotation data was provided by system macros supplied by SAS Institute with the Version 5 tape. These include %DCLANNO, %SYSTEM, %CIRCLE, %FRAME, and %LABEL. Once again the versatility of the SAS/AF program screen was extremely useful in conditional execution of portions of the program in response to the user-selected options. Simulation of spot density, if desired, was effected by generating a variety of functions. This application utilizes three circles and number, gave us a method for discrimination in plot regions with high density. Two of the other main options include 1) selection of the output device and 2) entry into the graphics catalog. Plots were generated using statements of the form:

```
PROC GANNO
ANNO-ANNOTDATA;
```

The ability to produce SAS/GRAPH output on our large 36" Zeta was highly desired but only became available with the advent of SAS 5.15. This ability along with the ability to control the size of the circle and number, gave us a method for discrimination in plot regions with high density. Once available, this option has been well received and utilized.

The REPLAY program utilizes PROC GREPLAY and allows the user to select the type of output device desired as well as the type of template to be used. One can select templates to be overlay plots or produce them side by side. These predefined templates were created from GREPLAY and are stored in their own catalog. The plots are also stored in their own catalog. One can access them as in the following example:

```
PROC GREPLAY IGOUT=SAVED.PLOTS
tc=SAVED.TEMPLTS
template-overlay;
```

One can also zoom in on specific areas of a plot by defining the template in a tricky way (7,8). One basically tells the system that the plotting area is larger than it really is. Then, the only part you
end up seeing is the small area that would display
in the area of the screen had it really been as big
as you defined. Confusing? Mike Kalt's paper
(see reference 8) should clear things up.

An example of an overlay (of an expanded portion
of the simulated electrophoretogram) is shown
in figure 5. One of the plots, which might represent
a cancerous sample, was generated using the spot
number off. The other plot, which might represent
a normal sample, was generated with the spot number
on. Notice that most of the spots with numbers
have a companion spot without a number. Now
notice that spots 100 and 200 have no companion.
If this was real data, those two spots, which represent
proteins, might give a scientist insight into what
cause this particular cancer.

Summary
There are some remarkable new features that exist
in Version 5 SAS software. These can be combined
in useful and interesting ways to help solve problems
that may lead to useful medical discoveries.

Acknowledgments
SAS, SAS/GRAPH, SAS/AF, and SAS/FSP are
registered trademarks of SAS Institute, Inc., Cary,
NC, USA.

Contact Author
Michael S. Lajiness
The Upjohn Company
301 Henrietta Street
Kalamazoo, Michigan 49001

References
   2, page 72.
2. SAS Institute Inc., Changes and Enhancements
to SAS/AF Software. TR: P-149.
   Version 5. TR: P-141.
   SUGI Conference Proceedings.
5. SAS Institute Inc., SAS Communications. Spring
   the Annotate and Template Facilities. TR:
P-142.
7. C. Kelly, Enhancing SAS/GRAPH Output with
8. M. Kalt, Creative Uses of PROC GREPLAY.
Figure 1. Major Elements in the System

SAS/AF®

SAS/GRAPH®

PROC GANNO  PROC GREPLAY

Simulated Electrophoretogram
Proposed System Design

Rexx Exec

\[
\text{SAS/AF Application (batch)}
\]

Input data

permanent SAS data files

Plot Data

Access Graphics Catalog

terminal small plotter large plotter
Figure 3. Example of a SAS/AF Catalog

**SAS/AF Catalog**

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td>Help</td>
<td>General system help info.</td>
</tr>
<tr>
<td>Main</td>
<td>Menu</td>
<td>Main menu</td>
</tr>
<tr>
<td>Build</td>
<td>Program</td>
<td>Access proc build from PF24</td>
</tr>
<tr>
<td>Exit</td>
<td>Program</td>
<td>Gets you out of SAS</td>
</tr>
<tr>
<td>Plot</td>
<td>Program</td>
<td>Plotting program &amp; options</td>
</tr>
<tr>
<td>Read</td>
<td>Program</td>
<td>Data entry program</td>
</tr>
<tr>
<td>Replay</td>
<td>Program</td>
<td>Graphics catalog option</td>
</tr>
</tbody>
</table>
Figure 4. Example of a Single Simulated Electrophorogram
Figure 5. Example of Simulated Electrophoretogram Overlay