 Genetic Polymorphism in Normal Human Fibroblasts as Analyzed by
Two-dimensional Polyacrylamide Gel Electrophoresis*1

Kim E. Walton, David Styer, and Eric I. Gruensteint

From the Department of Biological Chemistry, College of Medicine, and the Department of Mathematics, College of Arts and Sciences, University of Cincinnati, Cincinnati, Ohio 45267

Two-dimensional gel electrophoresis has been used to measure the degree of genetic polymorphism among the proteins of normal human fibroblasts. Autoradiographic analysis of the gel protein profiles from radioactively labeled cells allowed comparison of as many as 308 discrete polypeptides at a time. In addition, a newly developed technique for double label autoradiography was used to increase the sensitivity of the system for detection of small differences in the protein profiles of different cell lines. Only about 1.2% of the proteins of different cell lines were found to differ in their electrophoretic mobility. This corresponds to an average heterozygosity of approximately 0.6%. Previous studies of genetic polymorphism using different methods of one-dimensional electrophoretic analysis have estimated the average heterozygosity of the human population at about 6.7%. Detailed mathematical analysis shows the variation of the observed from the expected number of differences to be statistically highly significant. While the reasons for this difference are not clear, the observation of low levels of genetic polymorphism on two-dimensional gels should enhance the usefulness of this technique for detection of altered proteins in inherited disease.

Genetic polymorphism occurs within a population of animals of the same species when a given gene locus is found to produce at least two different alleles each having a frequency greater than 1%. Among the human population, the ABO blood groups are one good example. Detection of polymorphic proteins within a population is most commonly accomplished by the electrophoretic analysis of proteins on gels. If the protein being studied is an enzyme, its position on the gel can often be determined by locating the enzyme activity with a specific enzyme stain. Since the 1966 landmark studies of Harris (1), Johnson et al. (2), and Lewontin and Hubby (3), enzymes have been examined for genetic polymorphism in over 100 different species ranging from protozoa to mammals (4). In man alone, over 70 different gene loci have been examined of which approximately 30% have been found to be electrophoretically polymorphic (5). Quantitative evaluations of polymorphism are usually expressed as the degree of heterozygosity, and, for the human population, the average heterozygosity is estimated to be 6.7% (5). Recently, a new technique for protein analysis has become available in which as many as 1000 proteins can be simultaneously analyzed for electrophoretic mobility. Developed by Patrick O’Farrell (6), the technique involves a two-dimensional polyacrylamide gel in which the proteins are resolved in the first dimension by isoelectric focusing and in the second dimension by SDS-gel electrophoresis. Thus, the position of a protein in the first dimension is determined exclusively by its isoelectric point while its position in the second dimension is a function only of its molecular weight.

We have utilized this new two-dimensional gel technique to measure the degree of polymorphism among the proteins from five different lines of normal human fibroblasts. These lines were selected to represent a broad cross-section of the human population. Our results suggest that the previous estimate of genetic polymorphism among the human population may be too high. The low levels of polymorphism observed on our two-dimensional analyses further suggest that these gels may be very useful in detecting genetically altered proteins in individuals with inherited diseases of metabolism.

EXPERIMENTAL PROCEDURES

Materials—Earle’s minimum essential medium (catalogue No. F-11), nonessential amino acids, fetal calf serum, penicillin (10,000 units/ml), streptomycin (10,000 μg/ml), and Earle’s balanced salt solution were purchased from Grand Island Biological Co. Tricine and Triton X-100 were obtained from Sigma. L-[3,5,5-H]Arginine (1.0 mCi/ml, 50 to 30 Cl/mmol); L-[4,5-H]Histidine (1.0 mCi/ml, 60 to 80 Cl/mmol); L-[U-14C]Arginine (0.1 mCi/ml, > 270 mCi/mmol), and L-[U-14C]Histidine (0.1 mCi/ml, > 270 mCi/mmol) were purchased from New England Nuclear. Sodium dodecyl sulfate was specially pure grade from Gallard-Schlesinger. Ampholines in the pH range 5 to 7 and 3 to 10 were purchased from LKB. Urea was ultrapure grade from Schwartz/Mann. Molecular weight standards used were β-galactosidase (130,000), phosphorylase a (94,000), and deoxyribonuclease I (31,000) from Worthington Biochemical Corp.; bovine serum albumin (68,000), catalase (60,000), and ovalbumin (43,000) from Sigma; and glutamate dehydrogenase (53,000) from Boehringer Mannheim. 2,5-Diphenyloxazole (PPO) was purchased from New England Nuclear and dimethyl sulfoxide was obtained from either Sigma or Fisher Scientific Co. Ink used for blackening the gels was Flo-Master transparent black ink made by Venus Estebrook.

Cells—Five lines of normal human fibroblasts were obtained from

1 The entire mathematical analysis is presented as a miniprint immediately following this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, Md 20014. Request Document No. 78M-1156, cite author(s), and include a check or money order for $2.15 per set of photocopies.

2 The abbreviations used are: SDS, sodium dodecyl sulfate; IEF, isoelectric focusing; Me2S0, dimethyl sulfoxide; Tricine, N-Tris(hydroxymethyl)methyl]glycine.

7951
the Institute for Medical Research, Camden, N. J. The cell lines are described in Table I. Upon arrival, they were placed in a humidified incubator at 37°C and 5% CO₂. The fibroblasts were passaged into three tissue culture plates (100 x 20 mm) using a 0.04% trypsin solution and maintained in 10 ml of Earle's minimum essential medium (F-11) supplemented with 1% penicillin/streptomycin, 20 mM Tricine, 24 mM NaHCO₃, 1% (v/v) nonessential amino acids, and 10% (v/v) fetal calf serum. This medium is referred to as modified F-11.

Labeling of Cell Protein—After reaching confluence, the cells were washed once with Earle's balanced salt solution. They were then incubated with 5 ml of medium containing either 0.5 mCi of [³H]-arginine or 0.025 mCi of [¹⁴C]-arginine for 24 h. Cells were then washed with Earle's balanced salt solution. They were then incubated with 5 ml of medium containing either 0.5 mCi of [³H]-lysine or 0.025 mCi of [¹⁴C]-lysine and were then washed with 0.15 M NaCl and harvested into 2 ml of 0.15 M NaCl by scraping with a rubber policeman. Cells were pelleted by centrifugation and stored at -70°C. Radioactivity incorporated into protein was determined on the basis of hot trichloroacetic acid-precipitable activity. Protein concentration was assayed by the method of Lowry et al. (7).

Sampling Preparation—The labeled cell pellets were prepared for electrophoresis by a modification of the method of Ames and Nikaio (8). Cells were dissolved in SDS at an SDS:protein ratio of approximately 13:1, heated in boiling water for 15 s, and then Triton X-100 was added to give an SDS/Triton ratio of 1:8. Samples were then prepared for loading onto the isoelectric focusing gels by first adding amphiolines to a final concentration of 2% (i.e. 1.6% pH 5 to 7, 0.4% pH 3 to 10) and then solid urea to a final concentration of 9 M. Samples containing approximately 20 to 100 μg of protein were loaded onto gels as described below.

Single Label Polyacrylamide Gels—Proteins from each cell type were analyzed on separate two-dimensional polyacrylamide gels according to the method of O'Farrell (6). Briefly, the sample was loaded onto a 2-mm diameter IEF gel containing 2% ampholines (1.6% pH 5 to 7, 0.4% pH 3 to 10) and electrophoresed overnight. Gels were placed in O'Farrell's sample buffer containing 10% (w/v) glyceral, 5% (v/v) β-mercaptoethanol, 2.3% (w/v) SDS, and 0.0625 M Tris-HCl, pH 6.8, frozen, and stored at -70°C. Prior to analysis in the second dimension, the gels were thawed and incubated in the sample buffer for 2 h. Each IEF gel was then placed across the top of an SDS polyacrylamide slab gel (prepared according to Laemmli (9) as modified by Lu-Stourgeon and Rusch (10)) and the proteins were electrophoresed into the second dimension. Radiolabeled proteins were visualized by autoradiography.

Double Label Polyacrylamide Gels—For experiments in which [³H]-labeled protein from one cell line was compared with [¹⁴C]-labeled protein from another cell line on the same two-dimensional gel, the [³H]- and [¹⁴C]-labeled proteins were distinguished by the method for double label autoradiography of Gruenstein and Pollard (11). [³H]- and [¹⁴C]-labeled cell protein samples were mixed prior to IEF gel analysis to give a [³H]/[¹⁴C] ratio of 150:1. The sample containing a mixture of [¹⁴C]- and [³H]-labeled protein from the two cell lines was loaded onto a single IEF gel and analyzed in two dimensions as described above. The slab gel was then treated with Me₂SO to remove water and the scintillator fluor, 2,5-diphenyloxazole, was diffused into the gel as described by the Bonner and Laskey modification (12) of the procedure of Randerath (13). The gel was then dried onto a piece of filter paper under vacuum. The presence of the scintillator fluor enhances detection of [³H] by a factor of approximately 1000 while enhancing [¹⁴C] detection only about 10-fold. Autoradiographic exposure of the enhanced gel to either Kodak XR or XRP X-omat medical X-ray film at -70°C then detected [³H]-labeled proteins almost exclusively. Next, the gel is painted black with Flo-Master Marking Ink, thus screening out photons and reducing the efficiency of [³H] detection by at least 1000-fold. A second, longer exposure at room temperature is then made on fresh film revealing exclusively the [¹⁴C]-labeled proteins.

pH Profile and Molecular Weight Standards—The pH gradient for the IEF dimension was determined using an Ingold surface electrode. pH readings were taken at 5-mm intervals along the length of the IEF gel and the pH plotted as a function of distance as shown in Fig. 1A.

The seven proteins described as molecular weight standards in materials were utilized in the second dimension. These were applied to a well formed at the edge of the SDS slab gel about 0.5 cm from the end of the IEF gel. These proteins of known molecular weight were then subjected to the same electrophoresis conditions as those proteins in the IEF gel. After electrophoresis, the strip of SDS gel containing the molecular weight standards was cut from the body of the gel and stained with Coomassie blue to visualize the proteins. The results were graphed as shown in Fig. 1B.

RESULTS

Fig. 2, A to D, shows the two-dimensional gel protein profiles of four different normal human fibroblast cell lines. Arrows indicate areas of difference between that cell line and line GM43. Differences may be due either to the presence of an extra protein spot (black arrows) in cell line GM43 or to a missing (white arrows) cell line GM43 protein. Some apparent differences between protein profiles are not marked with arrows. These "missing proteins" were detected on longer exposure of the gels (Fig. 2, E and F). For example, in C and D of Fig. 2, gels of GM43 and GM38 are shown after an exposure of 1 day. Slightly more than 200 proteins can be seen. All protein spots are matched on both gels except for two which are seen in cell line GM43 but not in GM38 (as shown by the arrows). By using a longer exposure of 3 days (E and F), 300 spots are visible in the gels of these same two cell lines. There is now one more spot in GM43 not seen in GM38, and one protein spot in GM38 not in GM43. This gives a total of four differences between these two protein profiles.
Genetic Polymorphism in Cultured Fibroblasts

Fig. 2. Two-dimensional profiles: single label gel comparisons. Cells were labeled with [14C]lysine plus [14C]arginine and analyzed on separate two-dimensional polyacrylamide gels as described under “Experimental Procedures.” A to D show the protein patterns of cell lines GM975, GM316, GM43, and GM38, respectively. For purposes of analysis, all gel profiles were compared to line GM43 as a standard. White arrows indicate missing protein spots and black arrows indicate extra spots. E and F are longer autoradiographic exposures of C and D. The black and white arrows in each of these two panels indicate the differences that were found when the two cell lines were compared. E and F also illustrate the increase in protein spots as well as in protein differences that is observed when longer exposures of the protein profiles are studied.

As shown by the arrows in Fig. 2 (E and F) and by the enlargements in Fig. 3.

As a control, two gels of the same cell type, GM1652, were analyzed. Fig. 4 illustrates this comparison showing that, of the approximately 200 proteins visualized on each gel, there are four apparent differences. These differences, however, must be artifactual since the same cell line is being compared. Therefore, they must be caused by slight variations in either the individual IEF gels or in the SDS slab gels.

In order to deal with artifacts such as those demonstrated by our comparison of GM1652 with itself, we have developed a technique for double label autoradiography of 14C and 3H on polyacrylamide gels (11). In this technique, two cell lines, one labeled with 3H and one labeled with 14C, are harvested, solubilized, and the proteins are then mixed. The proteins from the two cell lines are then analyzed together on a single IEF gel which is then applied to an SDS slab gel for the second dimension. Therefore, from the point of solubilization through the final two-dimensional gel, proteins from both cell lines have been subjected to identical conditions. This double
Genetic Polymorphism in Cultured Fibroblasts

FIG. 3. Enlargements of certain areas from the gels shown in E and F of Fig. 2. Open arrows indicate missing spots and solid arrows indicate extra spots.

FIG. 4. Single label control gels. A illustrates the protein profile obtained from the cell line GM1652 labeled with [3H]arginine plus [3H]lysine. B illustrates the protein profile obtained from the same cell line, GM1652, labeled with [14C]arginine and [14C]lysine. The arrows indicate the apparent differences observed upon comparison of the protein patterns.

The double label method was used to compare GM1652 with itself and the results are shown in Fig. 5. The apparent differences seen by single label comparison are eliminated by double label comparison. The four spots which were missing (shown as white arrows in Fig. 4) are now present in Fig. 5, A and B. This technique, therefore, appears to effectively eliminate artifactual differences introduced by single label comparisons of protein profiles. In addition, the data suggest that no artifactual differences are introduced by the use of 3H- versus 14C-labeled amino acids, even though the labeled atoms on the 3H-amino acids are different from the 14C-amino acids.

Comparison of different cell types using the double label method is shown in Fig. 6. A shows the profile of 3H-labeled GM1652 and B shows the profile of 14C-labeled GM38. Of the approximately 200 spots obtained in each pattern, four differences were found (as indicated by the arrows). Since both cell types were run on the same two-dimensional gel, we conclude that these nonmatching spots result from real chemical-structural differences in the proteins of the cell lines themselves. C is an autoradiograph from a separate gel containing the same amount of 14C-labeled material as that of the double label gel but no 3H. It was exposed in the same way as the double label gel for the 3H detection of A. C therefore is a measure of the amount of 14C-labeled material simultaneously detected during the process of 3H exposure. Only a small number of 14C-labeled proteins are seen and these represent the most intensely labeled proteins which are badly overexposed in the 3H detection of A. We define this as 14C spillover. Similarly, D represents the 3H spillover during the 14C exposure. In this case there is essentially no spillover.

In both the single label and double label comparisons, differences were further confirmed by comparing the protein profiles at longer exposure times. A summary of the results for the five different cell lines appears in Table II. The average difference in protein profiles for cells compared by single label autoradiographic analysis is 1.8%. For protein profiles compared by double label autoradiographic analyses of the two-dimensional gel, the average difference is reduced to 1.2%. It is also interesting to note that the differences that have been detected by both single and double label comparisons always
FIG. 5. Double label control gels. Individual plates of GM1652 cells were labeled with either \(^{3}H\)lysine and arginine or \(^{14}C\)lysine and arginine. Aliquots from both plates were combined and analyzed on a single two-dimensional polyacrylamide gel. Exposures were then made to detect \(^{3}H\)-labeled protein (A) and \(^{14}C\)-labeled protein (B) as described under “Experimental Procedures.” C and D are autoradiographs of separate gels containing only \(^{14}C\)- or \(^{3}H\)-labeled protein, respectively.

Differences in the minor, faintly visible proteins. In none of the cases that we have examined have differences been seen in the major, dark autoradiographic spots. On the basis of estimates of 6.7% average heterozygosity among the human population, we had expected to see an average difference among our protein profiles of about 13%. This discrepancy will be considered in more detail under “Discussion.”

DISCUSSION

The system of two-dimensional polyacrylamide gel electrophoresis developed by O’Farrell is an extremely powerful tool for the comparison of large populations of proteins. In particular, the technique has great potential for the identification of genetically altered proteins responsible for inherited disorders of metabolism in the human population. In using two-dimensional gels for this purpose in humans, it is, however, not usually possible to obtain a highly inbred population and the question of interference by genetic polymorphism therefore becomes acutely important. We have examined this question by determining the number of electrophoretic differences in the two-dimensional protein profiles of fibroblast cultures from five different, normal individuals. Four of these individuals were selected to represent different ages, sexes, or races. The fifth subject (GM38) is the daughter of one of the other four (GM43). Careful analysis of the protein profiles by side-by-side comparison of parallel two-dimensional gels and by single gel comparison using the newly developed technique of double label autoradiography on two-dimensional gels has revealed an average electrophoretic difference of 1.2% of the total number of proteins resolved.

We felt it was important to compare this observed value to the number of differences that would have been expected to arise on the basis of the average heterozygosity per locus for the human population which is estimated to be approximately 6.7% (3). In order to make this comparison, it is first necessary to consider the differences between the methods which have been used for arriving at the estimated 6.7% average heterozygosity per locus and the technique of two-dimensional polyacrylamide gel electrophoresis. Average heterozygosity is generally determined for any given gene locus by examining the electrophoretic mobility of the protein product of that locus in a large number of individuals—usually in excess of 100. Since the particular protein being studied is rarely purified, detection of the protein of interest is generally accomplished by means of a specific histochemical reaction. Therefore, the vast majority of proteins studied has been enzymes for which specific stains were available. By examining the enzyme products of many different gene loci in separate experiments, it is possible by extrapolation to arrive at an estimate of the average heterozygosity per locus for the population. For this purpose, the proteins studied have been chosen to represent a wide variety of enzymatic reactions such as oxygenases, dehydrogenases, transferases, and so on.

In two-dimensional polyacrylamide gel electrophoresis analysis, on the other hand, many different proteins are visualized simultaneously while the specific function of almost all of them remains unknown. Comparisons of the protein profiles of five individuals will, therefore, result in examination of only five, rather than several hundred, examples of the product of any given gene locus. The chances of obtaining a truly repre-
Genetic Polymorphism in Cultured Fibroblasts

Double label gel comparisons. Individual plates containing GM1652 and GM38 cells were labeled with \(^{1}H\)(lysine and arginine) and \(^{14}C\)(lysine and arginine), respectively. Aliquots of both cell types were mixed together and analyzed as described in Fig. 5. A = \(^{1}H\)-labeled GM1652. B = \(^{14}C\)-labeled GM38. C and D are spillover controls as in Fig. 5.

TABLE II
Degree of variation in protein profiles for different cell lines analyzed on two-dimensional gels

<table>
<thead>
<tr>
<th>Comparisons and subjects</th>
<th>Spots matched</th>
<th>Spots unmatched</th>
<th>Unmatched spots/total spots</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single label</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38 vs. 43</td>
<td>305</td>
<td>4</td>
<td></td>
<td>1.3</td>
</tr>
<tr>
<td>975 vs. 43</td>
<td>283</td>
<td>4</td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>316 vs. 43</td>
<td>303</td>
<td>9</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>1652 vs. 43</td>
<td>191</td>
<td>3</td>
<td></td>
<td>1.6</td>
</tr>
<tr>
<td>Single label</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(^{1}H) 1652 vs. (^{14}C) 1652</td>
<td>190</td>
<td>4</td>
<td></td>
<td>2.1</td>
</tr>
<tr>
<td>Double label</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1652 vs. 1652</td>
<td>133</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Double label</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1652 vs. 38</td>
<td>211</td>
<td>4</td>
<td></td>
<td>1.9</td>
</tr>
<tr>
<td>1652 vs. 43</td>
<td>196</td>
<td>3</td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>1652 vs. 975</td>
<td>169</td>
<td>2</td>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td>1652 vs. 316</td>
<td>288</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

We do not know the reason for this apparent discrepancy,
but it seems appropriate to speculate on two possible reasons. 1) The two-dimensional gel system may be significantly less sensitive to electrophoretic variations than the other methods previously employed for the detection of genetic polymorphism. 2) There may be a significant systematic difference between the particular subset of total proteins being examined on our two-dimensional gels, and the subset of proteins previously studied using other methods.

Several studies have already demonstrated the sensitivity of two-dimensional polyacrylamide gels. In his original description of the technique, O'Farrell (6) was able to detect a single missense mutation in Escherichia coli among 1100 unaltered proteins, demonstrating that the method is sensitive to a single charge difference. Millman et al. (14) have used two-dimensional gel analysis of eukaryotic cells to detect a missense mutation in the hypoxanthine-guanine phosphoribosyltransferase of HeLa cells. Isoelectric focusing constitutes the first dimensional separation in two-dimensional polyacrylamide gel electrophoresis and Zechel (19) has demonstrated the resolving power of isoelectric focusing in his studies involving the α-subunit of RNA polymerase from E. coli. Using this technique, he was able to separate the wild type α-subunit from a mutant α-subunit which differed only by the replacement of a leucine residue with a histidine residue. These data all indicate that the ability of the two-dimensional gels to resolve proteins with modified primary structures should be at least as great as that of the systems relying on the one-dimensional electrophoretic mobility commonly used in studies on genetic polymorphism (18).

The second possibility, i.e. that two-dimensional polyacrylamide gel electrophoretic analysis is selecting a different subset of total cellular proteins from those examined by previous one-dimensional electrophoretic analysis, seems to us a more likely explanation. Perhaps the most important difference between one- and two-dimensional analytical techniques relates to the detection of proteins by enzymatic and nonenzymatic methods, respectively. In a large majority of cases, previous measurements of genetic polymorphism on one-dimensional gels have utilized specific enzymatic stains to detect the protein of interest. The two-dimensional technique requires denaturation of the proteins and relays, therefore, on the amount of radioactivity incorporated during their biosynthesis to detect the proteins. Thus, previous estimates of genetic polymorphism have been based primarily on proteins which have been selected because of their enzymatic activity. Analysis by two-dimensional gels, on the other hand, selects proteins that have taken up the most radioactivity independent of whether they are serving as enzymes. It may be, therefore, that the two-dimensional gel analysis focuses primarily on proteins present in stochiometric rather than catalytic amounts such as structural or membrane proteins. Indeed, Harris and Hopkinson (5) in 1972 suggested that, "...membrane-bound enzymes or other enzymes which are not readily solubilized from tissues in a native form may show a different incidence of polymorphism than the soluble or readily solubilized enzymes." There is, in fact, some evidence supporting the notion that such proteins may be less mutable than other cell proteins. For example, skeletal muscle actin, a nonenzymatic protein, which we would classify for these purposes as a structural protein, exhibits no polymorphism even among members of different but related species and remarkably little polymorphism among evolutionarily distant species (17). This dramatic lack of polymorphism is thought to result from the variety of different proteins with which actin must structurally interact. Furthermore, for proteins of major metabolic pathways which are likely to be present in greater than average amounts, it has also been observed that the degree of genetic polymorphism is far lower than would otherwise have been expected. The work of Cohen et al. (18) has demonstrated the presence of only 0.1% average heterozygosity per locus among enzymes of the glycolytic pathway from the brains of mice, monkeys, and humans. They suggest that this low degree of variability arises from the many regulatory interactions in which these proteins are involved, thus rendering them less tolerant to even minor alterations in primary structure. A similar situation seems to exist for human ribosomal core proteins. Bucknall et al. (19) have shown that the average heterozygosity per ribosomal core protein locus is 0.75%. This may again be due to the involvement of these proteins in complex interactions.

It may also be important to note that in no case were any of the 100 or so most intensely radioactive proteins on our gels found to vary electrophoretically among any of the cell lines studied. All of the variation was found among the fainter protein spots indicating that these proteins are either present in much smaller quantities or are turning over more slowly. If this is true, it would tend to support the suggestion that major structural and enzymatic pathway proteins are predominantly nonpolymorphic. In those instances where we did detect proteins present in the two-dimensional profile of one cell line but missing from another, it is impossible to say unequivocally that they were not there at all since they could simply be so markedly reduced in quantity as to have remained below the limits of our detection. If this is true, it, however, only serves to further reduce the degree of genetic polymorphism we are seeing to even lower levels.

As with the one-dimensional electrophoretic methods utilized by others, the two-dimensional system is likely to detect only electrophoretic allozymes. Despite the sensitivity of our isoelectric focusing dimension, enabling it to distinguish between allozymes differing by as little as a single charge, it is possible that some degree of polymorphism has been lost as a result of having denatured the proteins. Kittos et al. (20) have presented evidence for the existence of conformational polymorphism among proteins with the same primary structure. Such "conformers" might be detected in a nondenaturing one-dimensional system but lost under the denaturing conditions of our two-dimensional system. While we consider it unlikely that the presence of conformers will be found to explain the quantitative discrepancy between the expected and observed degrees of genetic polymorphism, we cannot exclude this possibility.

In addition to the implications vis-à-vis genetic polymorphism, the small amount of protein structural variation seen on two-dimensional gels from different lines of normal human fibroblasts has important consequences for the usefulness of these gels in the detection of altered proteins in disease states. In studies comparing the protein profiles of normal and pathological cells, the existence of only a small degree of polymorphism will certainly be an advantage. To some extent this advantage may be attenuated if the reason for the low levels of polymorphism we have observed is, as we have suggested, the result of focusing largely on structural proteins. In that case, many of the mutant enzymes in inherited disorders of metabolism may be silent or invisible on these gels. On the other hand, this disadvantage is probably more than offset by the ability to measure changes in nonenzymatic proteins which do not at present have any other in vitro bioassays.

Acknowledgments—We would like to thank Ms. Linda Papa for her excellent technical assistance and Prof. Harold Bell for his stimulating discussions of the mathematical correlations of one- and two-dimensional analyses of genetic polymorphism.
### Genetic Polymorphism in Cultured Fibroblasts

#### REFERENCES

#### TABLE 1

| Average retardation of stone-ground uncorrected from liquid 1, and the number of differences, (d/k), 2 | + | 105 | 102 | 101 | 100 | 99 | 98 | 97 | 96 | 95 | 94 | 85 | 80 | 75 | 70 | 65 | 60 | 55 | 50 | 45 | 40 | 35 | 30 | 25 | 20 | 15 | 10 | 5 | 0 |
| 1. Acids, phosphoric (1,4-cells) | 0.72 | 0.74 | 0.76 | 0.78 | 0.80 | 0.82 | 0.84 | 0.86 | 0.88 | 0.90 | 0.92 | 0.94 | 0.96 | 0.98 | 1.00 | 1.02 | 1.04 | 1.06 | 1.08 | 1.10 | 1.12 | 1.14 | 1.16 | 1.18 | 1.20 | 1.22 | 1.24 | 1.26 | 1.28 | 1.30 | 1.32 | 1.34 | 1.36 | 1.38 | 1.40 |
| 2. Phenylalanine, standard (1,4-cells) | 0.76 | 0.78 | 0.80 | 0.82 | 0.84 | 0.86 | 0.88 | 0.90 | 0.92 | 0.94 | 0.96 | 0.98 | 1.00 | 1.02 | 1.04 | 1.06 | 1.08 | 1.10 | 1.12 | 1.14 | 1.16 | 1.18 | 1.20 | 1.22 | 1.24 | 1.26 | 1.28 | 1.30 | 1.32 | 1.34 | 1.36 | 1.38 | 1.40 | 1.42 | 1.44 |
| 3. Phenylalanine, uncorrected (v, 1,4-cells) | 0.82 | 0.84 | 0.86 | 0.88 | 0.90 | 0.92 | 0.94 | 0.96 | 0.98 | 1.00 | 1.02 | 1.04 | 1.06 | 1.08 | 1.10 | 1.12 | 1.14 | 1.16 | 1.18 | 1.20 | 1.22 | 1.24 | 1.26 | 1.28 | 1.30 | 1.32 | 1.34 | 1.36 | 1.38 | 1.40 | 1.42 | 1.44 | 1.46 | 1.48 | 1.50 |
| 4. Amino acids | 0.88 | 0.90 | 0.92 | 0.94 | 0.96 | 0.98 | 1.00 | 1.02 | 1.04 | 1.06 | 1.08 | 1.10 | 1.12 | 1.14 | 1.16 | 1.18 | 1.20 | 1.22 | 1.24 | 1.26 | 1.28 | 1.30 | 1.32 | 1.34 | 1.36 | 1.38 | 1.40 | 1.42 | 1.44 | 1.46 | 1.48 | 1.50 | 1.52 | 1.54 | 1.56 |

*N* Table 1 is taken from Harris and Rodi, J. (1). Table 3 and comes from a total of 200 pairs of cells, of which 10 were found to be monochromatones.

**Note:** The number of differences, (d/k), is calculated from eq. 1. The index of the differences, (d/k), is defined as the ratio of the number of differences, (d/k), to the sum of the number of differences, (d/k), and 1.

**Remarks:** (1) Instead of 1 molecule.
Genetic Polymorphism in Cultured Fibroblasts

The purpose of this supplement is to provide the mathematical background for calculating the expected number of differences between the two-dimensional protein profiles of two randomly selected individuals.

We start with the assumption that all genes have been equally well understood, let us initially assume that each gene has only two alleles. From the data of Durr and Heffner (1) we should expect to be able to make reasonable good guesses as to the number of genes with two or more alleles is relatively small. We will, nevertheless, consider the case for some case of more than two alleles.

Let

- \( n \) = the number of genes whose products appear on the plate;
- \( x_{ij} \) = the frequency of the first and second allele, respectively, of gene \( i \), \( 1 \leq i \leq n \);
- \( h_{ij} \) = the heterozygosity of gene \( i \), \( 1 \leq i \leq n \).

We assume Hardy-Weinberg equilibrium here so that the frequency of individuals homozygous for the first (second) allele, \( A_{1}A_{1} \) (\( A_{2}A_{2} \)) of gene \( i \) is \( x_{1}^{2} \) (\( x_{2}^{2} \)), and the frequency of individuals heterozygous for gene \( i \) is

\[
\text{eqn. 1) } h_{ij} = 2x_{1}x_{2}
\]

Finally, we assume that distinct genes and distinct alleles from distinct genes are on the plate.

Let \( m \) be the total number of different spots showing on the two plates combined. Then \( n \leq m \leq 2n \).

Let \( n_{0} \) be the total number of differences seen on the two plates. (One difference is where there is a spot on one plate, but not the other.) If \( h_{ij} \) is the number of differences determined by gene \( i \), \( 0 \leq h_{ij} \leq n_{0} \). Thus, for the case of \( 2n \) alleles, \( h_{ij} \) can assume values of 0, 1, or 2 only. If we study the effect of gene \( i \) on the two plates, we find that there are 8 possible situations. These are all illustrated in Figure 2b.

![Figure 2b](image)

The 8 pairings in Figure 2 have been broken down into 3 columns: I, II, III, which have exactly 1 allele; II, those which have exactly 1 difference; and III, those which have 0 differences. Then, \( h_{ij} \) is a random variable which takes the value

\[
\text{eqn. 2) } 0 \text{ with probability } x_{1}^{2}x_{2}^{2} + x_{1}x_{2}^{2}\left(2x_{1}^{2} + 2x_{1}\right)
\]

\[
\text{eqn. 3) } 1 \text{ with probability } M(2x_{1}^{2} + 2x_{1})x_{2}^{2} + x_{1}^{2}x_{2}^{2}(2x_{1}^{2} + 2x_{1}) + x_{1}^{2}x_{2}(2x_{1}^{2} + 2x_{1}) + x_{1}^{2}x_{2}^{2}(2x_{1}^{2} + 2x_{1})
\]

\[
= 2x_{1}h_{ij}
\]

\[
\text{eqn. 4) } 2 \text{ with probability } x_{1}x_{2}^{2} + x_{1}^{2}x_{2}^{2}
\]

The expected value of \( h_{ij} \) is

\[
\text{eqn. 5) } E[h_{ij}] = 2x_{1}h_{ij}
\]

Since the expected value of the sum is equal to the sum of the expected values, then

\[
E[2n] = 2\sum_{i=1}^{n} E[h_{ij}]
\]

\[
= 2nE[h_{ij}]
\]

We have, therefore, obtained the expected number of genetic differences between two randomly chosen individuals in the heterozygosity, \( h_{ij} \), of each gene. Furthermore, if we multiply and divide the last equation by \( n \),

\[
E[2n] = 2\frac{\sum_{i=1}^{n} h_{ij}}{n}
\]

\[
\text{eqn. 6) } E[2n] = 2\frac{\sum h_{ij}}{n}
\]

where \( \overline{h} \) is the average heterozygosity for all genes. Since \( \overline{h} \) is a normally determined value for many species, it would be convenient to be able to calculate \( E[2n] \) knowing only \( \overline{h} \). While we do not present the details here, it is possible to calculate the upper and lower limits of \( E[2n] \).

**Limits of Expected Number of Differences**

The upper limit of \( E[2n] \) for a fixed \( n \) will occur if all assuming that each gene has at least as many alleles when all \( h_{ij} = 2 \). The lower limit of \( E[2n] \) will occur as many \( h_{ij} = 0 \) or 1. The proof of these statements is as follows:

In order to find the maximum and minimum value of \( E[2n] = 2\sum_{i=1}^{n} h_{ij} \), given that \( n \) and \( \overline{h} \) are known, we can use algebraic manipulations to show that there is but one local extremum (2). Since \( \overline{h} = \frac{\sum h_{ij}}{n} \), it will suffice to minimize and maximize \( E[2n] = 2\sum h_{ij} \) under the constraint that \( \sum x_{1}^{2} + \sum x_{2}^{2} = \overline{h} \) in fixed conditions. At any extremum there is one real number \( x \) such that

\[
\frac{\partial E[2n]}{\partial x_{1}} = 0 \text{ and } \frac{\partial E[2n]}{\partial x_{2}} = 0
\]

But this is \( x_{1} = x_{2} = 0 \) for \( 1 \leq i \leq n \),

and thus, \( h_{ij} = 0 \).

This shows that there is but one extremum, and it follows from that since each \( h_{ij} \), equals the same constant value of \( n \), all \( h_{ij} \) must be equal to each other, and \( h_{ij} = 0 \) at this extremum. Thus, one can show that this extremum is a maximum for \( n = 0 \) and is a minimum for \( E[2n] \) under the piece constant.

Having shown that the upper interval for \( h_{ij} \) is a maximum, it follows that for any value of \( h_{ij} \) which is less than the maximum obtained, \( h_{ij} \) will be minimized when all \( h_{ij} = 0 \). If these are no more than two \( h_{ij} \) found which are either \( 0 \) or 1, we can apply the case to those which are either \( 0 \) or 1, and not obtain a smaller value for \( E[2n] \) by increasing one \( h_{ij} \) at the expense of the others. In the case of \( 0 \) and the others are \( 0 \). Thus, by an iterative process, we conclude that \( E[2n] \) will be minimized when all \( h_{ij} \) is at (one sensible one of the \( h_{ij} \) are either 0 or 1).

**Expected Number of Differences (I, II, III)**

![Expected Number of Differences](image)

Thus, in the same case there are 3 alleles rather than two some particular form, I, let the frequency of the alleles to \( x_{1}, x_{11}, \) and \( x_{2} \). Then the heterozygosity,

\[
h_{ij} = 2x_{1}x_{2} + 2x_{1}x_{2} + 2x_{1}x_{2}
\]

\[
= 1 - x_{1}^{2} - x_{2}^{2} - x_{1}x_{2}
\]

Extensive computations of the nature of those derived from Figure 2 show that

\[
E[2n] = 4x_{1}x_{2} + 2x_{1}x_{2} + 4x_{1}x_{2} + 4x_{1}x_{2} + 4x_{1}x_{2} + 4x_{1}x_{2} + 4x_{1}x_{2} + 4x_{1}x_{2}
\]

\[
= 4x_{1}x_{2} + 2x_{1}x_{2} + 16x_{1}x_{2} + 16x_{1}x_{2} + 16x_{1}x_{2} + 16x_{1}x_{2}
\]

\[
= 12x_{1}x_{2} + 24x_{1}x_{2}
\]

which simplifies to

\[
E[2n] = 12x_{1}x_{2} + 24x_{1}x_{2}
\]

Thus, \( E[2n] \) is increased over the 2 allele case by exactly \( 12x_{1}x_{2} \) which shows that allowing 3 alleles can only increase \( E[2n] \).
Genetic Polymorphism in Cultured Fibroblasts

The results of performing these calculations using the data of table 3 show that

\[ \frac{1}{(2 \times 0.5) \times 0.5} = 2.5 \times 10^{-4} \]

Now let us turn our attention to the situation where we have two experimental twodimensional gels with a total number of spots, N. By T we mean all spots on the first plate plus all spots on the second plate which are not coincident with any spot on the first plate. Since all our calculations are done on the basis of the number of genes, n, each of which is represented by either one or two spots on each plate, we can find a relationship between T, the total number of distinct spots, and n. This can be done as follows.

The total number of spots on each plate must fall between n and 2n, so that

\[ n \leq T \leq 2n \]

and

\[ \frac{2n - n}{2n} \leq \phi \leq 1 \]

Knowing that no single gel has more than 2 errors, and that, however, a crude estimate which can be derived from the evidence (b) points toward the fact that two-thirds of all lanes have not polymorphic. If we were reasonable, therefore, to bracket T between n and (7/3)n = (21/9)n = 2.3 n. Then

\[ \frac{n}{2.3 n} \leq \phi \leq \frac{2n}{2.3 n} \]

and, choosing conditions that will minimize E(exp), i.e., \( n = 150 \), we calculate that

\[ \frac{1}{(2 \times 0.5) \times 0.5} = 2.5 \times 10^{-4} \]

so that, if the average heterogeneity of the population were 0.75, we can calculate the minimum E(exp) by recalling that we have not come across any possible \( \phi \) of 0.5. This will in fact, occur when 75 genes have an average heterogeneity of 0.5, i.e., there has an average heterogeneity of 0.75 and at the other \( \phi = 0.5 \).

Thus, \( \phi = 0 + 0.25 \) and

\[ \frac{1}{(2 \times 0.5) \times 0.5} = 2.5 \times 10^{-4} \]

which, for the total of 150 gene slots represents a minimum average heterogeneity of 0.25. Consequently, we can assume a distribution of average heterogeneities similar to those described by Harris and Neuman (10), in which case E(exp) = 0.04.

We can now calculate the standard deviation for both the minimum E(exp) = 0, and the Harris and Neuman data E(exp) = 0.3 from the equation

\[ \phi = \frac{\phi}{\sqrt{\phi}} \]

where

\[ \phi = \frac{1}{(2 \times 0.5) \times 0.5} = 0.75 \]

and

\[ \phi = \frac{1}{(2 \times 0.5) \times 0.5} = 3.5 \]

For the case of minim E(exp) in which we have \( \phi = 0.5 \) for gene lost and

\[ \phi = 0.04 \]

we have a probability of 0.95 for gene lost and 0.05 for gene found.

Finally, suppose that we have a stage of two-dimensional gels at a given time all being represented on the two plates. If in both 3 or 5 are unrepresented, then that lane must give rise to at least two differences (\( \phi \geq 0.4 \)).

\[ \frac{1}{(2 \times 0.5) \times 0.5} = 2.5 \times 10^{-4} \]

and

\[ \phi = 3.5 \]

Then, in the experimental data presented earlier in this manuscript, the total number of spots seen on each two-dimensional gel is approximately 200 and the minimum value of \( \phi \) is approximately 0.3, our observed number of differences is 10, each of the 4 comparisons, at least 4 standard deviations from the expected value.
Genetic polymorphism in normal human fibroblasts as analyzed by two-dimensional polyacrylamide gel electrophoresis.

K E Walton, D Styer and E I Gruenstein


Access the most updated version of this article at http://www.jbc.org/content/254/16/7951

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/254/16/7951.full.html#ref-list-1