CLIN. CHEM. 30/12, 1925–1932 (1984)

Identification of Polypeptides on Two-Dimensional Electrophoresis Gels by Amino Acid Composition

Gerald I. Latter, Stephen Burbeck, James Fleming, and John Leavitt

We present a method that can, in principle, provide tentative identification of a substantial proportion of the polypeptides resolvable on a given two-dimensional electrophoresis gel. Computerized microdensitometry of autoradiograms from 20 gels labeled in turn with each of the 20 common amino acids provides the data for simultaneously measuring the amino acid composition of all polypeptides of interest on the gel. These compositions are then compared with computer data bases of known protein compositions. Similarity between a known and an unknown polypeptide with comparable molecular mass indicates a potential identification, which can then be confirmed with conventional techniques. We illustrate this technique by applying it to the identification of proteins in a transformed human cell line (HuT-14).

Additional Keyphrase: *computerized microdensitometry*

A substantial fraction of the polypeptides that are expressed in a given type of cell can be resolved on a single two-dimensional (2-D) polyacrylamide electrophoresis gel (*1*). But much of the information potentially available from such gels remains undeciphered because the identity and function of so few of these polypeptides is known. For example, Leavitt et al. (*2, 3*) found 64 polypeptides that differed markedly between the diploid human fibroblast KD line and the transformed HuT-14 cell line. They could identify only three of those important polypeptides. Similarly, Bravo and Celis (*4*) found 53 human proteins sensitive to neoplastic transformation in human epithelial and fibroblast cell lines; but again only three are identified. Leavitt et al. (submitted for publication) found 49 polypeptides that differ qualitatively between untransformed and transformed NIH 3T3 cells; none have yet been identified. Only for *Escherichia coli* is the situation markedly different (*5*).

What is needed is a general method to aid in the identification of resolvable polypeptides. Migration properties, abundance, morphology (the shape and degree of resolution of the spot), and geometrical relations to nearby spots are the clues most easily available from 2-D gels. Each of these clues gives some information about the identity of the polypeptide. In rare cases these may be sufficient for identification. Actin, for example, can usually be readily identified from its appearance, abundance, and location on a gel (*6, 7*). Such cases are rare because migration properties vary according to the exact methods and materials used in producing the 2-D gels and because no large compendium of known polypeptides organized according to these migration properties is currently available. But data on the amino acid composition of peptides are widely available from a number of sources. A growing number of computer data bases (e.g., Dayhoff, NEWAT, EMBL, GenBank) contain information about amino acid sequences or about nucleotide sequences from which amino acid sequences, hence amino acid compositions, may often be derived. Compositions are obtainable for unknown polypeptides in 2-D gels as well; amino acid compositions of resolved polypeptides have been measured by excising the polypeptide from the gel and using microanalytical methods (*8*) or by computerized microdensitometry of a set of gels labeled, in turn, with each of the 20 amino acids (*9*).

The practical question is whether comparison of amino acid compositions of unknown polypeptides with those of known proteins materially aids in identification. Manabe et al. (*8*) found such comparisons helpful when they used data on the amino acid composition of serum proteins excised from 2-D gels. Their methods provide highly accurate compositions but are cumbersome if one wishes to investigate a large number of unknown polypeptides. We demonstrate in this paper that even relatively inaccurate estimates of amino acid composition, together with information on molecular mass, often provide sufficient information to match polypeptides resolved on 2-D gels with the composition and molecular mass of known polypeptides. Confirming experiments show that good matches often provide correct identification.

Materials and Methods

Overview of Computerized Measurement of Amino Acid Composition from Two-Dimensional Gels

We use computer-assisted microdensitometry to determine the amino acid composition of 11 polypeptides in HuT-14, a transformed human fibroblast cell line (*10*). The 122 polypeptides (see Figure 1) were chosen to span the range of molecular mass (approximately 22 to 200 kDa) and pI (4 to 6.5) that our gels accommodate. The composition of two of these is known exactly: the gamma-,beta actin and the mutant beta actin (*10*). An outline of the general method for obtaining amino acid compositions by computerized microdensitometry is illustrated in Figures 2 and 3.

Tissue Culture and Production of Gels

The in vitro transformed cell line HuT-14 has been described previously (*10*). High-density, subconfluent monolayer cultures in multiwell plates were incubated in 2 mL of radiolabeling medium for 24 h. A modified Eagle's Medium Selectamine Kit (GIBCO) was used to prepare the 20 radiolabeling media, each containing 10 to 50 mCi of a different ^{14}C amino acid (excepting ^{35}S methionine) per liter, from New England Nuclear, Boston, MA 02118. Radiolabeled cell monolayers were rinsed three times with cold phosphate-buffered isotonic saline, drained, and then lysed directly in O'Farrell's "lysis buffer A" (*11*). The resulting protein samples were stored at −76 °C until electrophoresis. Two replicate 2-D polyacrylamide electrophoresis gels were made from each protein sample with the ISO-DALT system from Electro-Nucleonics, Fairfield, NJ 07006. However, the second-dimension gels were thinner (1 mm instead of 1.5 mm) than is standard for the ISO-DALT system, to diminish the

Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306.

Received July 5, 1984; accepted August 16, 1984.

Fig. 1. Typical gel, showing the 122 proteins (those in boxes) whose compositions we measured
Mutant beta actin, the most acidic of the actins shown, was used as the reference protein. Its composition is known exactly

quenching effect of the gel during exposure of the autoradiograms and thereby shorten the necessary exposure times.

Autoradiography and Microdensitometry

Kodak XAR-2 film was used for auto-radiography. Exposure time varied from three to eight weeks, so that the actin standard was evenly exposed for all autoradiograms. The film was developed in a Kodak X-Omat Model 5 processor and digitized on an Optronics P-1000 film scanner at a pixel size of 200 μm. The abundance of each polypeptide of interest for each of the amino acids is measured by integrating the observed counts over the pixels in an area surrounding each spot on the gel. These areas are designated by the investigator, using an interactive graphics system. Details of this process are described elsewhere (*9*).

Calculation of Amino Acid Composition

Amino acid composition is calculated as shown in Figure 3. This calculation technique differs from that previously presented (*9*) in two ways. First, we used mutant beta actin, a single isolated protein, as the reference protein instead of the beta-,gamma actin combination. The amino acid compositions of beta and gamma actin differ slightly, so use of them combined required an assumption about their relative abundance. The second difference is in the method for combining the information from replicate gels. The previous method was to normalize the counts for each unknown polypeptide within each gel and then to average the resulting ratios across replicates. In the method used here, the two replicate integrated counts for each unknown polypeptide are summed and divided by the similarly summed counts for the reference actin (Figure 3). By summing the counts before taking the ratio we avoid the troublesome problem of averaging Cauchy distributions (*9*). If the errors across replicates are independent, then the errors are reduced in proportion to the square root of the number of replicates. Estimates of measurement error made under conditions similar to this study have shown standard errors on the order of 15% (unpublished observation). Thus with duplicate gels for each amino acid, as in this experiment, we would expect standard errors of about 10%.

The proper trade-off between reducing error and minimizing the number of gels to be analyzed depends upon the intended use for the data on amino acid composition. A 15% standard error can be reduced by 5% by using nine replicates, which is somewhat tedious but not out of the question. But the accuracy with which each amino acid is estimated is not crucial for our purpose, because the proportion of residues among the 20 amino acids is more important than an exact value for any one of them.

Search Techniques

The primary method for identifying polypeptides from the HuT-14 cell line was to compare their compositions with the

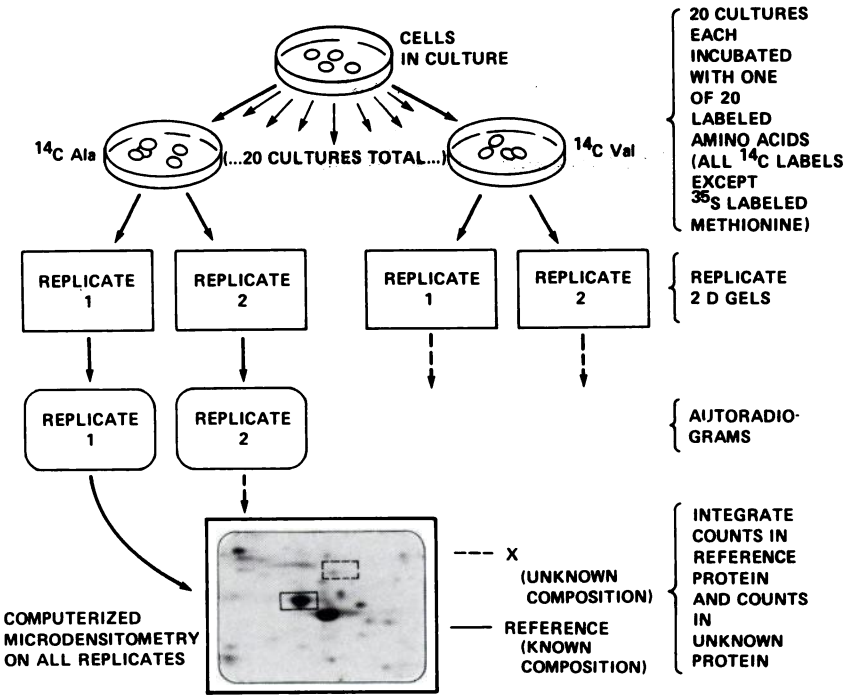

Fig. 2. Protocol from cell cultures to counts

Twenty cultures of the cell line under study are radiolabeled with each of the 20 different ^{14}C- or ^{35}S-labeled amino acids. After 2-D polyacrylamide gel electrophoresis of the samples, preferably in replicate, autoradiograms are made, and computerized microdensitometry is used to measure the relative amounts of radioactivity incorporated in the polypeptides whose composition is desired, as well as in the reference protein

compositions found in the 1984 Dayhoff protein data base (*12*). We compared each of the 120 unknown polypeptides with approximately 2300 compositions from the data base. We also compared them with over 1000 compositions collected by Kirschenbaum (*13*). Similarity in composition between an unknown polypeptide and a known protein in one of these data bases was taken as a potential identification if their molecular masses were within 20% of each other. This wide latitude for information on molecular mass was allowed because of the well-known artifacts in assessing molecular mass from migration rates in the second dimension.

One can assess the similarity of two amino acid compositions by use of various statistical procedures. We found it useful to compare polypeptides with both the correlation coefficient and the Chi-square statistic. If one or both suggest a possible match, we investigate further. The Chi-square statistic is obtained by treating the unknown composition as if it were a random sample which, on average, should match the known composition. Then a simple Chi-

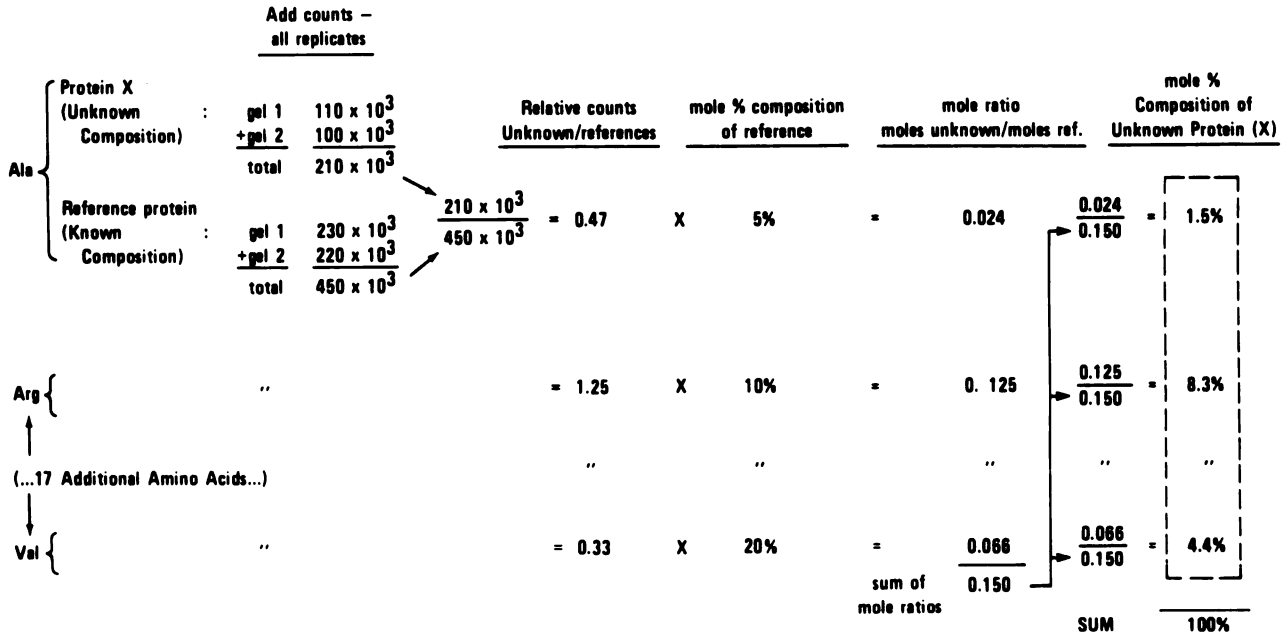

Fig. 3. Illustration of calculation of amino acid compositions by use of counts measured from autoradiograms

Counts are first summed for each protein over all replicates for each of the 20 amino acids. The ratio of counts of the unknown protein to the reference is then computed. Multiplying this by the mole percent of the reference yields the moles of unknown protein per mole of reference for each amino acid. Dividing the mole ratio of each amino acid by the sum of the mole ratios for all amino acids yields the mole percents for all 20 amino acids

square test (*14*) results in a dissimilarity estimate between 0 and 1. Because the count of the number of residues of each amino acid is not really a random sample from some population of interest, the number resulting from this test cannot be interpreted properly as the probability that the unknown polypeptide differs from the known. Nonetheless, the smaller this number the better is the match between the unknown and the known polypeptide. Most pairings give rise to a dissimilarity value of 1; any that produce .99 or less are worthy of further scrutiny.

The familiar Pearson r correlation coefficient is another useful measure of similarity. It take on values between −1 and 1; the larger the correlation coefficient the better the match. However, it is not yet clear how large r must be to indicate a match. We arbitrarily used a lower cutoff of .70, although no tentative match with a correlation of less than .80 is yet supported by any independent evidence.

We also use star diagrams in this manuscript to give a visual illustration of similarity in composition. The star diagram represents the relative number of moles of each of the 20 amino acids per 100 moles of peptide by the radial lengths of the 20 points of a star (Figure 4). The star diagrams make it clear that it is the pattern of all 20 amino acids that allows one to identify proteins rather than the exact lengths of each of the radii. If, for instance, each of the 20 radii of the non-muscle tropomyosin (box in Figure 4) were randomly lengthened or shortened by 10%, the appearance of the overall pattern would not be materially changed. The choice of the angular order in which the amino acids appear in these stars depends upon their intended use. We have simply ordered the amino acids alphabetically (according to the internationally recommended three-letter code), but various more functionally grouped orderings ought to be considered.

Results

Of the 120 unknown polypeptides, 17 have been tentatively identified (Table 1, Figure 5). To date, some evidence has been found to support the identification of seven of these tentative identifications (Table 2). The supporting evidence varies from co-migration experiments to comparisons with published gels in which polypeptides have been identified by other methods.

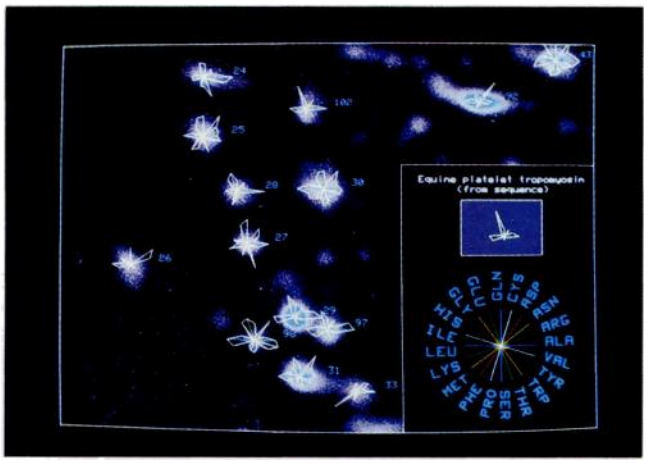

Fig. 4. View of the region of the gel below and to the left of the vimentin/tubulin area in which tropomyosins have been identified by others (*17, 18*)

Amino acid composition data are shown in the form of star diagrams drawn on top of the resolved proteins. Each of the 20 radii in the stars represents one amino acid (as identified in the reference diagram in the lower right). The length of each radial line is proportional to the mole percent of that amino acid. The star diagram in the small box represents the amino acid composition of bovine nonmuscle tropomyosin as calculated from the sequence. Visual inspection shows its similarity to that of p29, p97, and p102, all of which are in positions similar to those of proteins identified as tropomyosins by Matsumura et al. (*17*)

Table 1. Tentative Identifications of Some Polypeptides

No.	M_r, kDa	Matching protein	Sequence M_r, kDa	χ^2 test	r
p29	32	Non-muscle tropomyosin (equine)	32.8	.90	.90
p36	27	Phosphatidylcholine-transfer protein (bovine)	28.4	.85	.83
p37	24	Dihydrofolate reductase (mouse)	24.8	.76	.74
p38	26	Complement C1q subcomponent B chain (human)	27.8	.80	.87
p41 (p791)	25	α_1-Acid glycoprotein (orsomucoid, bovine)	27.2	.85	.75
p55	60	Alpha-Fl-ATPase (mitochondrial)	59.4	.96	.93
p58	57	Antithrombin III precursor (human)	60.9	.93	.90
p65	60	Leucine aminopeptidase (bovine)	60.2	.73	.93
p73	35	Lactate dehydrogenase H chain (pig)	42.4	.15	.93
p76	30	Prolactin precursor (human)	29.9	.92	.85
p96	24	C-reactive protein (rabbit)	24.4	.87	.72
p97	32	Non-muscle tropomyosin (equine)	32.8	1.0	.80
p102	37	Non-muscle tropomyosin (equine)	32.8	.95	.96
p104	52	Beta-F1-ATPase (mitochondrial)	54	.85	.94
p112 (p789)	27	PreproEGF related	(see text)		
p113 (p788)	26	PreproEGF (residues 630–880, mouse)	31.5	.35	.95
p116	50	cAMP-dependent protein kinase (type II regulatory)	52.1	.63	.93

Support for the identification of p55 and p104 as mitochondrial ATPases is based upon the fact that proteins can be shown to have mitochondrial origins by comparing gels from cultures treated with nonactin (a macrotetrolide antibiotic) with gels from untreated cultures. Nonactin inhibits translocation of cytoplasmically synthesized mitochondrial precursor proteins into the mitochondria where final processing normally takes place (*15*). The transformed HuT-14 cell line is extraordinarily sensitive to nonactin; even in very small doses it kills the cells. So we made our comparison with nonactin-treated KD cells, the normal human fibroblast from which the HuT-14 cell line was derived. Polypeptide 55 best matched the alpha subunit of F1-ATPase and its apparent molecular mass matched that of alpha F1-ATPase (*16*). P55 also disappeared after treatment with nonactin, which confirms its mitochondrial origin. The location of polypeptide 55 closely matches that of a polypeptide that has been tentatively identified as alpha F1-ATPase on the basis of its molecular mass, relative abundance, and mitochondrial origin (*15*). Two polypeptides (numbers 19 and 104) showed amino acid compositions in good agreement with that of beta-F1-ATPase. Both also have molecular masses similar to that of beta-F1-ATPase. But only polypeptide 104 disappeared after treatment with nonactin. Thus, in this case, amino acid composition and molecular mass alone were not sufficient to identify the protein. Without supporting evidence one could not choose between the two possible candidates. Anderson (*15*) tentatively identifies beta-F1-ATPase in leukocytes on the basis of molecu-

Fig. 5. Tentative identifications
All proteins numbered here have been tentatively identified except p19. Some have confirming evidence (see Table 2)

Table 2. Tentative Confirmations of Some Polypeptides

No.	Identified as	Supporting evidence
p55	Alpha-F1-ATPase (mitochondrial)	Disappeared after non-actin treatment
p104	Beta-F1-ATPase (mitochondrial)	Disappeared after non-actin treatment
p102	Non-muscle tropomyosin	Matsumura et al. (*17*), no. 2
p29	Non-muscle tropomyosin	Matsumura et al. (*17*), no. 4
p97	Non-muscle tropomyosin	Matsumura et al. (*17*), no. 5
p116	Cyclic AMP-dependent protein kinase (type II regulatory)	Comigrates with bovine variety
p65	Leucine aminopeptidase	Comigrates with bovine variety

lar mass, mitochondrial origin, and relative abundance. The location of the spot identified is near that of ours.

Support for the identification of the non-muscle tropomyosins is based upon experiments by Matsumura et al. (*17*) in which microfilaments were isolated by a new method. They have published illustrations of 2-D gels in which several non-muscle species of tropomyosin are identified. Our transformed HuT-14 cells have three polypeptides (p29, p97, and p102) that appear to match their proteins 4, 5, and 2, respectively, and these proteins have amino acid compositions very similar to that of non-muscle tropomyosin. We observe small amounts of proline (approximately one residue per molecule) in all three of these polypeptides, even though non-muscle tropomyosins reportedly have no proline. Our observation is in agreement with that of Giometti and Anderson (*18*). The assumption is that the observed proline is an artifact of amino acid reutilization: a small amount of the labeled proline is being converted to some other constituent of the polypeptide.

Polypeptide 41 matches rat α_1-acid glycoprotein precursor (Chi-square = .85, r = .75). Comigration shows that the molecular mass as estimated in the second dimension is correct, but identification by comigration is problematical because of extensive post-translational processing, which gives rise to extreme variation in pI (*19*) depending upon its neuraminic acid content. However, neuraminic acid-free α_1-acid glycoprotein has an isoelectric point >5.0 (*20*), approximately what we observe (pI = 5.2).

Our tentative identification of polypeptide 116 as cyclic-AMP-dependent protein kinase (EC 2.7.1.37) type II regulatory chain is supported by comigration. Samples of human cAMP-dependent protein kinase were not available, but radioimmunoassay experiments (*21*) show that the human varieties from breast cancer cells cross react with bovine varieties. A comigration experiment confirmed that bovine 3′5′-cyclic-AMP-dependent protein kinase migrated to the

position of p116 on the gel. Comigration experiments also help support the identification of p65 as leucine aminopeptidase [aminopeptidase (cytosol), EC 3.4.11.1].

Intragel Comparisons

In addition to the identification of polypeptides, we examined relationships among proteins on the gel. We observed several pairs of proteins with high correlations (Figure 6). For proteins 7 and 8 the correlation coefficient was .99, for proteins 1 and 3 it was .93. This degree of similarity, together with the similar phenotypic shapes and close positions on the gel, implies that the polypeptides are nearly identical. They may differ only because of post-translational processing, or they may be polymorphic variants.

Discussion

Methodological Issues

It is clear that information on amino acid composition, together with molecular mass, facilitates identification of unknown polypeptides on 2-D gels. It is equally clear that, at most, such information provides a tentative identification. With highly accurate data on amino acid composition, a very good match between an unknown polypeptide on a gel and a known protein of the same molecular mass might constitute a credible identification for many noncritical applications. Otherwise, conclusive confirming evidence or several converging lines of supporting evidence are mandatory before much can be made of the sort are tentative identifications we have obtained. Moreover, the picture as presented in Table 1 is oversimplified. Some unknown polypeptides can be more or less plausibly matched with several known polypeptides from the data base (Table 3). This may be due to the unknown polypeptide having a "common" amino acid composition or it may simply reflect some commonality between the group of known polypeptides that are similar to the unknown. But whatever the cause, it underscores the need for caution in interpreting the tentative identifications. Better accuracy in measuring the amino acid compositions can be expected to add to the power of this technique by allowing tighter criteria for assuming a tentative match and thereby cutting the number of multiple tentative identifications. This reduces the effort necessary in confirming the identifications. The rapid growth of data bases will also add to the power of the technique by increasing the number of polypeptides that can be identified if they are present in the gel.

Table 3. Example of Multiple Tentative Identifications for Protein p37 (M_r 24 kDa)

$p(\chi^2)$	r	M_r, kDa	Name
.85	.70	23.7	Protein HC—human
.92	.73	24.0	Alpha-2μ-globin precursor—rat
.97	.69	25.1	Transform. prot. p21—Kirsten murine sarcoma virus
.76	.74	24.8	Dihydrofolate reductase—mouse
.92	.70	24.8	Dihydrofolate reductase—bovine
.89	.72	24.8	Dihydrofolate reductase—pig
.85	.71	25.0	Dihydrofolate reductase—chicken

The value of data on amino acid composition in identification is simply that one can compare 20 new pieces of information in addition to that on molecular mass and perhaps abundance or morphology. The new information may converge to make a convincing identification. But are all 20 pieces of information equally useful? A principal-components analysis (*22*) of mole percentages of 2293 polypeptides in the Dayhoff data base shows that knowledge of the relative abundance of some amino acids may be more useful than others for discriminating one polypeptide from another. The principal-components analysis extracts "factors" (in the present case, linear combinations of the 20 relative amino acid abundances) that characterize the data (i.e., the mole percent compositions of the polypeptides) in a useful manner. The first factor accounts for the largest possible difference (variance) between the compositions, the second accounts for the largest possible remaining variance once that accounted for by factor one is removed, and so on. Each of the resulting factors is described as a weighted sum of the 20 amino acid abundances. Examination of these weights shows that Ala, Arg, Cys, Gly, Leu, Lys, Ser, and Pro contribute disproportionately to the variance. They have relatively large weights in the first five factors, which together account for 55.6% of the total variance. Asn, Asp, Gln, His, Ile, Met, and Phe are not weighted heavily in the first five factors but are weighted heavily in the last five factors, which together account for only 8.2% of the total variance. Met and Phe are particularly poor in discriminating power, being weighted substantially only in the last two factors. In contrast, Ala and Arg are particularly useful; Arg is weighted very strongly (−.89) in the first factor and Ala has relative large weights on all of the first five factors (.24, −.45, .51, −.46, and .40, respectively). Thus useful matches may be obtainable with partial data on amino acid composi-

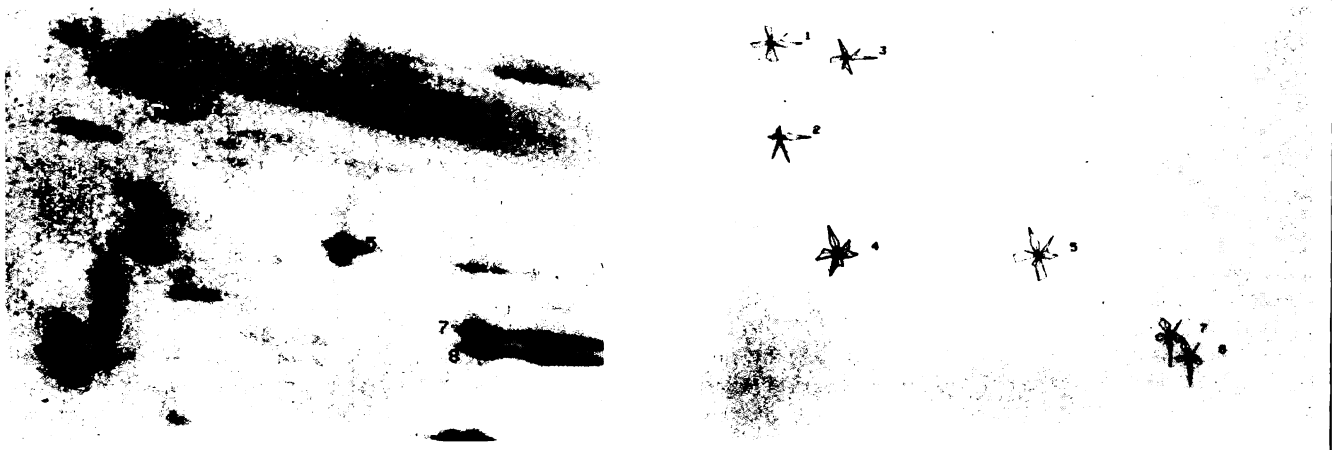

Fig. 6. Intra-gel relationships among proteins

(*a*) the proteins in the upper left corner of the gel shown in Figure 1. (*b*) star diagrams for the proteins shown in Figure 6*a* whose compositions we measured. Note the pairs of polypeptides with nearly identical pI, molecular mass, and amino acid composition. This suggests that the proteins are nearly identical

tions. However, those amino acids that superficially would seem to add little to the identification in general may, in important instances, be powerful discriminators.

To the extent that the position of a resolved polypeptide is reproducible from experiment to experiment, the effort of obtaining data on amino acid composition need not be repeated. Thus, within a given laboratory in which electrophoresis methods are carefully controlled, amino acid compositions and subsequent identification of the peptides represent an investment with continuing value. As electrophoresis methods become more standard and reproducible from laboratory to laboratory the necessity for repeating the work at each laboratory will diminish. Until that time, the compositions themselves should prove valuable for comparing gels run by different laboratories under different conditions, and would aid in cross-comparing gels from different cell lines or different species.

Biological Implications

Three of the polypeptides that differ in expression between KD and HuT-14 (*2*) appear in Table 1 (in parentheses in the first column). These and other proteins that have tentatively been identified in the HuT-14 cell line have potential implications for the understanding of this transformed human cell line.

The presence of two polypeptides that are strongly upregulated in the transformed HuT-14 cell line suggests that this cell line may be producing an excess of epidermal growth factor or a related transforming growth factor. These polypeptides (numbers 112 and 113), called p789 and p788 in previous publications (*2, 3*) are reliable markers for neoplastic transformation in human fibroblasts. The composition of p113 is very similar to that of residues 630–880 of mouse prepro-Epidermal Growth Factor (EGF) (*23*). The composition of p112 is very similar to that of p113, and the two have been shown to be closely related (*7*). Their molecular masses are about the same as that of the 28-kDa high-molecular-mass human EGF (*24*). If this identification proves to be correct then the reliability of these markers for transformation in human fibroblasts is consistent with the relationship between neoplastic transformation and a variety of oncogenes and growth factors that have an intimate relationship to EGF (*25–30*). An additional polypeptide, p116, may be related to an oncogene product. It shows a strong similarity in composition (Chi-square .63) to bovine cyclic-AMP-dependent protein kinase type II regulatory chain. The companion cAMP catalytic chain has been shown to be related to viral src gene products (*31*) and to stimulate EGF-dependent phosphorylation of the Epidermal Growth Factor Receptors (*32*).

Two other polypeptides, one of which is upregulated in the HuT-14 line, may play immunosuppressive roles. Polypeptide 96 matches rabbit C-reactive protein (Chi-square = .87, r = .72), which has been demonstrated to be immunosuppressive (*33*). Polypeptide 41 matches rat α_1-acid glycoprotein precursor (Chi-square = .85, r = .75). α_1-Acid glycoprotein concentrations are increased in the serum of acute-phase cancer patients, and it has been shown to be immunosuppressive (*34*). Confirmation of this tentative identification is necessary, because p41 appears to be the same as p791 (*2, 3*), which is upregulated by a factor of 2.5 from KD to HuT-14. But α_1-acid glycoprotein is not known to be expressed in fibroblasts. "Northern" or "Western" blot experiments are needed to show that it is in fact expressed in the HuT-14 cells.

We have also tentatively identified two polypeptides that are known to be targets of chemotherapeutic agents. Polypeptide 65 has been tentatively identified as leucine aminopeptidase. Membrane-bound leucine aminopeptidase is the target enzyme of bestatin, a chemotherapeutic agent and potential immunostimulant (*35*). Moreover, the cell-membrane-bound activity of leucine aminopeptidase is lowest in lymphocytes, higher in tumor cells, and highest in bone-marrow cells and macrophages (*35*). We have tentatively identified polypeptide 37 as tetrahydrofolate dehydrogenase (EC 1.5.1.3; dihydrofolate reductase), an enzyme that is the target for chemotherapeutic agents such as methotrexate, trimethoprim, and pyrimethamine (*36*).

Conclusive confirmation of the above tentative identifications will be useful in understanding the transformed state of the HuT-14 cell line. A careful quantitative study of the expression of these proteins in the diploid KD line vs the transformed HuT-14 line may also provide important results. The picture that is beginning to emerge from these studies, albeit a tentative one, is that in one transformed cell line there is a cytoskeletal protein (the mutant β-actin), an upregulated EGF-related growth factor, and an upregulated immunosuppressant factor (α_1-acid glycoprotein). This pattern seems to be too well coordinated to be merely a chance combination of unrelated events. It is tempting to suppose that some underlying factor has induced a coordinated "programmed" set of changes.

We acknowledge the excellent technical assistance of Edward Metz. The computer-readable data on amino acid composition collected from the publications of Kirschenbaum were kindly provided by Alex Reisner. This work was supported in part by the Koret Foundation and the Japan Shipbuilding Industry Foundation, Tokyo. We also thank the New England Nuclear Corp. for providing ^{14}C-labeled amino acids for part of this study.

References

1. Duncan R, McConkey EH. How many proteins are there in a typical mammalian cell? *Clin Chem* 28, 749–755 (1982).

2. Leavitt J, Goldman D, Merril C, Kakunaga T. Actin mutations in a human fibroblast model for carcinogenesis. *Clin Chem* 28, 850–860 (1982).

3. Leavitt J, Goldman D, Merril C, Kakunaga T. Changes in gene expression accompanying chemically-induced malignant transformation of human fibroblasts. *Carcinogenesis* 3, 61–70 (1982).

4. Bravo R, Celis JE. Human proteins sensitive to neoplastic transformation in cultured epithelial fibroblast cells. *Clin Chem* 28, 949–954 (1982).

5. Neidhardt FC, Vaughn V, Teresa AP, Bloch PL. Gene-Protein Index of *Escherichia coli* K-12. *Microbio Rev* 47, 231–284 (1983).

6. Garrels JI, Gibson W. Identification and characterization of multiple forms of actin. *Cell* 9, 793–805 (1976).

7. Leavitt J, Kakunaga T. Expression of a variant form of actin and additional polypeptide changes following chemical induced in vitro neoplastic transformation of human fibroblasts. *J Biol Chem* 255, 1650–1661 (1980).

8. Manabe T, Kiyotsugu K, Jitzukawa S, et al. A normalized protein map of human plasma and characterization of the normalized spots by microscale amino acid analysis. *Clin Chem* 28, 819–823 (1982).

9. Latter GI, Metz E, Burbeck S, Leavitt J. Measurement of amino acid composition of proteins by computerized microdensitometry of two-dimensional electrophoresis gels. *Electrophoresis* 4, 122–126 (1983).

10. Leavitt J, Bushar G, Kakunaga T, et al. Variations in beta actin accompanying incremental increases in human fibroblast tumorigenicity. *Cell* 24, 259–268 (1982).

11. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007–4021 (1975).

12. Barker WC, Hunt LT, Orcutt BC, et al. *Protein Sequence Database,* National Biochemical Research Foundation, Washington, DC, January 1984 ed.

13. Kirschenbaum, DM. A compilation of amino acid analyses of proteins XI. *Anal Biochem* 79, 470–501 (1977).

14. Ott L. *An Introduction to Statistical Methods and Data Analysis,* Duxbury Press, North Scituate, MA, 1977 pp 268–270.

15. Anderson L. Identification of mitochondrial proteins and some of their precursors in two-dimensional electrophoresis maps of human cells. *Proc Natl Acad Sci USA* 78, 2407–2411 (1981).

16. Knowles AF, Penefsky HS. The subunit structure of beef heart mitochondrial adenosine triphosphatase. *J Biol Chem* 247, 6624–6630 (1972).

17. Matsumura F, Lin Jim J-C, Yamashiro-Matsumura S, et al. Differential expression of tropomyosin forms in the microfilaments isolated from normal and transformed rat cultured cells. *J Biol Chem* 258, 13954–13964 (1983).

18. Giometti CS, Anderson NL. A variant of human nonmuscle tropomyosin found in fibroblasts by using two-dimensional electrophoresis. *J Biol Chem* 256, 11840–11846 (1981).

19. Altland K, Roeder T, Jakin HM, et al. Demonstration of α_1-acid glycoprotein (orosomucoid) by double one-dimensional slab gel electrophoresis: Evidence for intra- and interindividual variability of the microheterogeneity pattern in health and disease. *Clin Chem* 28, 1000–1010 (1982).

20. Schultze HE. Uber klinisch interessante korpereigene Polysaccharidverbindungen. *Scand J Clin Lab Invest* 10, Suppl. 31, 135–165 (1957).

21. Kapoor CL, Cho-Chung YS. Compartmentalization of regulatory subunits of cyclic adenosine 3′:5′-monophosphate-dependent protein kinases in MCF-7 human breast cancer cells. *Cancer Res* 43, 295–302 (1983).

22. Morrison DF. *Multivariate Statistical Methods,* McGraw-Hill, New York, NY, 1967, pp 221–258.

23. Burbeck S, Latter G, Metz E, Leavitt J. Neoplastic human fibroblast proteins are related to epidermal growth factor precursor. *Proc Natl Acad Sci USA* 81, 5360–5363 (1984).

24. Hirata Y, Orth DN. Epidermal growth factor (Urogastrone) in human fluids: Size heterogeneity. *J Clin Endocrinol Metab* 48, 673–679 (1979).

25. Waterfield MD, Scrace GT, Whittle N, et al. Platelet-derived growth factor is structurally related to the putative transforming protein p28sis of simian sarcoma virus. *Nature (London)* 304, 35–39 (1983).

26. Doolittle RF, Hunkapiller MW, Devare SG, et al. Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. *Science* 221, 275–277 (1983).

27. DeLarco JE, Reynolds R, Carlberg K, et al. Sarcoma growth factor from mouse sarcoma virus-transformed cells. *J Biol Chem* 255, 3685–3690 (1980).

28. Anzano MA, Roberts AB, Meyers CA, et al. Synergistic interaction of two classes of transforming growth factors from murine sarcoma cells. *Cancer Res* 42, 4776–4778 (1982).

29. Roberts AB, Anzano MA, Lamb LC, et al. Isolation from murine sarcoma cells of novel transforming growth factors potentiated by EGF. *Nature (London)* 295, 417–419 (1982).

30. Cohen S, Carpenter G, King Jr L. Epidermal growth factor-receptor-protein kinase interactions. *J Biol Chem* 255, 4834–4842 (1980).

31. Barker WC, Dayhoff MO. Viral src gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. *Proc Natl Acad Sci USA* 79, 2836–2839 (1982).

32. Ghosh-Dastidar P, Fox CF. cAMP-dependent protein kinase stimulates epidermal growth factor-dependent phosphorylation of epidermal growth factor receptors. *J Biol Chem* 259, 3864–3869 (1984).

33. Mortensen RF, Osmand AP, Gewurz H. Effects of C-reactive protein on the lymphoid system. I. Binding to thymus-dependent lymphocytes and alteration of their functions. *J Exp Med* 141, 821–839 (1975).

34. Cheresh DA, Distasio JA, Vogel CL, Lopez DM. Mitogen-induced blastogenesis and receptor mobility inhibition by breast cancer serum with elevated orosomucoid. *J Natl Cancer Inst* 68, 779–783 (1982).

35. Leyhausen G, Schuster DK, Vaith P, et al. Identification and properties of the cell membrane bound leucine aminopeptidase interacting with the potential immunostimulant and chemotherapeutic agent bestatin. *Biochem Pharmacol* 32, 1051–1057 (1983).

36. Huennekens FM, Vitols KS, Whitely JM, Neef VG. Dihydrofolate reductase. *Methods Cancer Res* 13, 199 (1976).

37. Hendricks M, Weintraub H. Tropomyosin is decreased in transformed cells. *Proc Natl Acad Sci USA* 78, 5633–5637 (1981).

38. Vandekerckhove J, Leavitt J, Kakunaga T, Weber K. Co-expression of a mutant beta-actin and the two normal beta- and gamma-cytoplasmic actins in a stably transformed human cell line. *Cell* 22, 893–899 (1980).