Automation in clinical microbiology: a new approach to identifying micro-organisms by automated pattern matching of proteins labelled with $^{35}$S-methionine

SOAD TABAQCHALI, R SILMAN*, DIANE HOLLAND*

From the Departments of Medical Microbiology and *Reproductive Physiology, St Bartholomew's Hospital Medical College, London

SUMMARY A new rapid automated method for the identification and classification of micro-organisms is described. It is based on the incorporation of $^{35}$S-methionine into cellular proteins and subsequent separation of the radiolabelled proteins by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). The protein patterns produced were species specific and reproducible, permitting discrimination between the species. A large number of Gram negative and Gram positive aerobic and anaerobic organisms were successfully tested. Furthermore, there were sufficient differences within species between the protein profiles to permit subdivision of the species. New typing schemes for Clostridium difficile, coagulate negative staphylococci, and Staphylococcus aureus, including the methicillin resistant strains, could thus be introduced; this has provided the basis for useful epidemiological studies.

To standardise and automate the procedure an automated electrophoresis system and a two dimensional scanner were developed to scan the dried gels directly. The scanner is operated by a computer which also stores and analyses the scan data. Specific histograms are produced for each bacterial species. Pattern recognition software is used to construct databases and to compare data obtained from different gels: in this way duplicate “unknowns” can be identified. Specific small areas showing differences between various histograms can also be isolated and expanded to maximise the differences, thus providing differentiation between closely related bacterial species and the identification of differences within the species to provide new typing schemes. This system should be widely applied in clinical microbiology laboratories in the near future.

The introduction of automation in clinical microbiology has lagged behind that of chemical pathology and haematology. By and large, microbiologists still rely on microbial morphology and metabolism for the classification and identification of micro-organisms, and on serotyping and phage typing for subdivision within species. In recent years various new approaches to taxonomy have been introduced. Included in these are procedures for manipulating the genetic material of the organism, such as nucleic acid transfer, DNA hybridisation techniques, and the guanine plus cytosine ratio (mole % G + C). A more definitive method for the taxonomy and comparison between micro-organisms would be to analyse the entire sequence of DNA bases in the bacterial genome. DNA fingerprinting using various endonucleases has recently been applied to some bacterial species and has shown remarkable specificity. These techniques at present, however, are time consuming and require specialised expertise, and are therefore not appropriate for use on a large scale.

An alternative approach to the analysis of nucleic acids is to measure the microbial proteins which are the products of the expression of the genetic material within each micro-organism. Comparisons of the proteins among different groups of organisms have been made using various electrophoretic methods; and such comparisons have been used for the identification and classification of various bacteria by analysing enzymic, ribosomal, cell envelope and whole cell proteins.

The most widely used method of protein separation uses electrophoresis on polyacrylamide gel (PAGE), containing sodium dodecyl sulphate (SDS). SDS-polypeptide complexes separate strictly according to
polypeptide size. Two dimensional electrophoresis and isoelectric focusing, which resolve minor differences between strains, have also been introduced. Rodwell and Rodwell used two dimensional (2-D) electrophoresis and incubation with radiolabelled protein hydrolysate to distinguish between small and large colony types of Mycoplasma mycoides sub-species mycoides and M. mycoides sub-species capri. Strains indistinguishable by DNA hybridisation were shown to have distinctly unrelated proteins by 2-D PAGE. Howard et al. analysed proteins labelled with 35S-methionine and separated on SDS-PAGE to distinguish between eight human serovars of Ureaplasma ureolyticum and strains of bovine origin. Comparing the autoradiographs, they showed that bovine isolates were distinguishable from human strains, and this confirmed serological and G + C ratios. Computer assisted numerical analysis of electrophoretic route was applied by Feltham and Sneath. Kersters and de Ley, and Sneath and Sokal. These programmes were adapted to study skin flora using an inexpensive Commodore PET 3000 series and to clarify the taxonomy of the 'viridans' streptococci.

Although Kersters and de Ley, Feltham and Sneath, and Jackman all commented on the potential of microbial protein patterns in taxonomy and on the considerable contribution to the classification and identification of many genera, they also emphasised the need for careful standardisation and computer data processing. Nevertheless, it is clear, that with the application of microbial whole cell protein electrophoresis, most bacteria can be grouped, at least to species level, by their specific protein patterns.

Whole cell protein electrophoresis, however, requires large volumes of the organism for the extraction and purification of protein. Furthermore, after electrophoresis the proteins cannot be scanned by densitometry directly without prior fixing and staining of the gels. The stains used for this purpose, such as Coomassie blue, are, on the whole, insensitive and require careful standardisation of the concentrations of the proteins to obtain optimal staining results. Densitometry of the unstained proteins has been applied and absorbance curves were generated for computer analysis, but the method requires ultra pure acrylamide to avoid measuring impurities at 280 nm, and gels containing SDS cannot be scanned for the same reason. This method, therefore, has several disadvantages and seems unlikely to be widely used in clinical laboratories.

This paper describes a new technique developed at St Bartholomew's Hospital, using 35S-methionine as substrate to produce radiolabelled proteins which can then be separated on SDS-PAGE and visualised by autoradiography. The method has the potential to identify bacteria to species and subspecies level in a single test.

### Background

A colleague in the department of reproductive physiology had been engaged in experiments entailing the translation of mRNA in vitro using a radiolabelled amino acid, the separation of the translated protein by polyacrylamide gel electrophoresis, and visualisation by autoradiography. During the course of the experiments unexpected protein patterns were repeatedly produced. The question of possible bacterial contamination was raised in discussions with the department of medical microbiology. As a first step, the antisera used in the experiments were sterilised by filtering and, at each step of the procedure, bacterial cultures were made. Filtering eradicated the extra patterns and the culture yielded two different species of Pseudomonas; Pseudomonas aeruginosa and another Pseudomonas species. To confirm that the extra protein bands were due to the bacterial contaminants, the two Pseudomonas strains were incubated with 35S-methionine and the resultant proteins were separated on SDS-PAGE. Autoradiographs showed two patterns.

### Table 1 Micro-organisms investigated by radiolabelled protein patterns

| Escherichia coli | Staphylococcus aureus |
| Klebsiella aerogenes | S. epidermidis |
| K. pneumoniae | S. saprophyticus |
| Enterobacter cloacae | S. haemolyticus |
| Proteus mirabilis | Streptococcus faecalis |
| P. morganii | S. pyogenes |
| Serratia marcescens | 'viridans' streptococci |
| Shigella boydii | S. pneumoniae |
| S. dysenteriae | Group B |
| S. sonnei | Group C |
| S. flexneri | Group G |
| Salmonella hedor | Bacteroides fragilis |
| S. typhimurium | B. thetaiotaomicron |
| S. enteritidis | B. vulgatus |
| S. montevideo | B. ovatus |
| A. cloacaevar. iwoffii | B. corrodens |
| A. Iwoffii | B. bivius |
| A. cloacae var antratus | Fusobacterium nucleatum |
| Pseudomonas aeruginosa | F. necrophorum |
| Aeromonas spp | Clostridia |
| Neisseria gonorrhoeae | 15 species |
| N. meningitidis | Candida albicans |
| N. pharyngis | C. parapsilosis |
| N. flavescens | |
Fig 1 Autoradiograph of radiolabelled protein bands from four Gram negative organisms: channels 1 and 8—E coli; 2 and 7—Pseudomonas aeruginosa; 3 and 6—Klebsiella aerogenes; 4 and 5—Proteus mirabilis.

Fig 2 Autoradiograph of radiolabelled protein bands from Neisseria and Branhamella strains: channels 1, 5, and 8—N meningitidis; 2 and 4—B catarrhalis; 3 and 6—N pharyngis; 7—N flavescens.

Fig 3 Autoradiograph of radiolabelled protein bands from: channels 1—Fusobacterium nucleatum; 2—Bacteroides fragilis; 3—B thetaotaomicron; 4—B vulgatus; 5—B ovatus.

Fig 4 Autoradiograph of radiolabelled protein bands from six E coli serotypes: channels 1 and 7—01; 2 and 12—040; 3 and 8—06; 4 and 11—075; 5 and 10—095; 6 and 9—0108.
different protein patterns. When other microorganisms were tested in the same way, each gave a distinct pattern. It was clear that the radiolabelled protein patterns generated by the different bacteria were specific to each bacterial species. It therefore seemed likely that this method could be used in clinical microbiology, particularly if the methods could be computerised. It was also important to determine whether such a technique might ultimately replace existing methods for identifying and classifying bacteria.

The use of radiolabelling

Identification of micro-organisms

A range of different bacterial species and fungi were tested. The radiolabelled protein patterns, which were visualised by autoradiography, were substantially different between the species. Fig 1 shows the autoradiograph of the proteins labelled with $^{35}$S-methionine of Escherichia coli, Pseudomonas aeruginosa, Klebsiella aerogenes and Proteus mirabilis run with coded duplicates. The differences between the strains and the duplicates are easily identified. Similarly, the different Neisseria species can be distinguished (fig 2). The five Gram negative anaerobic bacterial species also showed specific differences (fig 3). Similar experiments were performed on a wide range of other bacterial and fungal species (table 1). In each case there were sufficient differences between the protein profiles to distinguish between the species and, at the same time, there were similarities in pattern among related species to enable grouping of the genera.

Typing of micro-organisms

Having established that the protein patterns were specific for each species, the next step was to deter-
mine whether the species could be differentiated further. Various *E coli* serotypes were tested and each gave a different pattern (fig 4), showing that there were sufficient detailed differences within the species to permit subdivision or typing of these organisms. The method, therefore, was potentially useful in the typing of pathogenic organisms, for which existing typing methods were unavailable or unsatisfactory.

The method was first applied successfully to the typing of *Clostridium difficile*. Nine distinct types were identified, A–E and W–Z, based on the position of major protein bands.\(^{43,44}\) Using immunoblotting with homologous and heterologous antisera, it was shown that these major protein bands were immunogenic and specific to the type of *C difficile* (Wren B, Tabaqchali S, unpublished observations). Restriction endonuclease analysis of the DNA from these nine standard *C difficile* strains showed individual patterns for each type, confirming the reliability of this typing scheme.\(^{45}\) Further screening of *C difficile* extended these groups to 14 (fig 5). The method could also be used to type coagulase negative staphylococci,\(^{46}\) *Staphylococcus aureus*, (including the methicillin resistant *S aureus\(^{47}\) Neisseria gonorrhoeae, and *Haemophilus influenzae* type b (unpublished data).

**Epidemiological Studies**

**The epidemiology of *C difficile***

Clusters of antibiotic associated colitis (AAC) caused by *C difficile* have indicated that cross infection might be important, but there has been a lack of convincing evidence due to the unavailability of a reliable typing scheme for this organism. Various typing schemes have been suggested,\(^ {43,44,48–51}\) among which is the method described in this paper, based on the radiolabelled protein profiles (fig 5).\(^ {43,44}\) This method was applied to type strains isolated from patients and their environment from various hospitals\(^ {44}\) and from a prospective six month study of immunocompromised patients during an outbreak of AAC and diarrhoea.\(^ {52}\) Strains were also obtained during this survey from general medical patients. The results showed for the first time that nosocomial acquisition of *C difficile* and cross infection occurred among immunocompromised patients. Cross infection presented a serious clinical problem and was controlled only after strict isolation procedures were implemented. The study also showed that the development of symptoms was related to the type of *C difficile*: types E and X were associated with outbreaks of AAC; types A and B were isolated from asymptomatic mothers and neonates.\(^ {43}\)

**Methicillin resistant *S aureus* (MRSA)**

Outbreaks of infection with methicillin resistant *Staphylococcus aureus* (MRSA) have recently increased in the United Kingdom and, in particular, in London hospitals.\(^ {53}\) Conventional typing of the strain of *S aureus* causing these outbreaks has been unsatisfactory because the strain does not react with the international set of phages. It does, however, react with experimental phages and can be identified by its pattern of antibiotic resistance. A rapid simple method for typing the MRSA would be useful in detecting outbreaks and in differentiating between strains. SDS-PAGE of \(^ {35}\)S-methionine labelled cellular proteins of *S aureus* strains resistant to methicillin provided such a typing scheme for this organism.\(^ {47}\) The method was applied during an outbreak of cross infection on a cardiothoracic ward. The results showed that the outbreak strain was indistinguishable from the epidemic strain of MRSA (EMRSA) prevalent in other hospitals in the London area, and confirms the suggestion that the EMRSA forms a unique "clone", and that a single strain may be responsible for the infection in London and Australia.\(^ {54,55}\) In contrast, 44 epidemiologically separate strains of *S aureus*, which included non-epidemic strains of MRSA, other resistant strains, sensitive strains (supplied by Dr Marples, PHLS, London), and some animal strains, all gave individually distinct radiolabelled protein profiles. This method permitted rapid confirmation that an epidemic strain was responsible for the outbreaks of infection and indicated the need for urgent control measures.

**Coagulase negative staphylococci**

Coagulase negative staphylococci have emerged as important opportunistic pathogens in recent years and are particularly associated with implants and catheters. The clinical importance of repeated isolations of these organisms remains unknown because of the absence of a comprehensive typing scheme. Several methods have been applied,\(^ {56–60}\) but none has proved adequate to type all the strains isolated. Protein profiles labelled with \(^ {35}\)S-methionine were therefore introduced for the typing of coagulase negative staphylococci and compared with biotyping and antibiograms.\(^ {46}\) This typing system was found to be particularly useful in studying the epidemiology of coagulase negative staphylococci in patients with peritonitis that was associated with continuous peritoneal dialysis.\(^ {46}\) These organisms are the commonest cause of peritonitis in such patients, but it is not known, however, if repeated episodes of peritonitis are due to recurrence or re-infection. Such data can only be determined by typing all the strains isolated from these episodes. Such a study is in progress and the results will have important therapeutic implications.

The development of an automated scanning technique

The above studies were carried out using auto-
Automation in clinical microbiology

radiography of radiolabelled proteins to visualise the patterns. Although autoradiography is extremely sensitive, it is slow, taking one to five days, depending on the radioactivity in each protein band. It is also difficult for the human eye to discriminate between different electrophoretic runs, and the procedure cannot be automated. To do this it is necessary to convert the analogue pattern seen on the autoradiographs to a digital signal. One possible method is to use a densitometer to scan the autoradiographs, but this adds another step to an already elaborate procedure. The whole procedure can be simplified, however, by using a scanner to detect beta emission. Scanning, therefore, can bypass autoradiography and densitometry altogether. With such a procedure, the beta emission can be read directly from the dry gel plate and the signal entered in the computer. No such scanner was available for this, and the help of Professor Pullan was sought; it was he who devised the unique scanner which was vital to the successful development of the automated technique. At this stage Automated Microbiology Systems (AMBIS) patented the method and decided to finance the project, and in collaboration, they also undertook the development of the necessary hardware, consumables, and software.

**The Scanner**

A two dimensional scanner for beta emission, combining high resolution with high sensitivity and capable of scanning the dried gels directly, was developed. (fig 6). The scanner head consisted of 12 well separated cathode strips at right angles above 26 fine horizontal anode wires to give 312 discrete detection points. A voltage is applied between the electrodes which are protected by an aperture plate covered with a thin film of Mylar. An argon (90%) and methanol (10%) gas mixture flows into the space between the aperture plate and the cathode strips and is retained by the Mylar. Between the Mylar and the sample is the resolution plate with 312 apertures, each 3 mm long and 1 mm wide, each hole operating as a detector, through which a small element of the sample is detected. The sample moves in a 5 × 30 matrix during a scan so that beta emissions are detected from 150 discrete sites at each of the 312 apertures, and the total data are collected from 46800 sample sites. The aperture plate also mechanically protects the detector from contamination and damage.

Beta particles that pass through the Mylar and into the detection gas produce ionisation signals at both electrodes. Because each set of wires and strips is connected to delay lines, the exact location of each signal is determined by the measurement of delay times and coordinates passed on to the computer. The gas pressure and the voltage required for the operation are 15 pounds per square inch and 1650 volts, respectively.

The scanner is connected and operated by an IBM pc/AT computer. A gel to be scanned is placed on the sample platen and scanning starts after appropriate instructions have been typed into the computer. The gel and the platen move under the scanner detector head and are lifted so that they come into close contact with the resolution plate. Scanning times are usually 15 minutes to one hour, depending on the activity in the gels. The time for the detection of the radiolabelled proteins is therefore reduced from days to
minutes. Scan data are collected in the computer and can be retrieved at the end of the scan.

**Computer/VDU System**

An IBM pc/AT computer with a high quality graphics screen controls the system (fig 6), which includes the electrophoresis unit, the scanner, the printer, and also the collection, storage, and manipulation of the scan data. It has a fast program execution and data processing ability, so it is possible to scan a gel and obtain all the qualitative and quantitative data in about 30–60 minutes. The computer is connected to a printer to provide documentation of data and graphics (fig 6). Computer software has been developed to scan the gel, to examine each lane, extract and analyse the information.

**Automated Electrophoresis Unit**

A flat bed computer controlled electrophoresis unit (AMBIS EP tower) has been developed. The IBM pc/AT computer controls the electrophoresis running conditions and includes special sensors which automatically terminate the end of the run (fig 7). Precast gels were also developed to avoid interlaboratory variations in preparation.

**Retrieval and Analysis of Data**

At the end of the scan of a gel the data are retrieved on the VDU where a picture resembling the autoradiograph is presented for analysis (fig 8). Each lane

---

**Fig 7** AMBIS electrophoresis unit.

**Fig 8** Display of radiolabelled protein patterns after scanning and display on VDU. Retrieval of scan data showing one channel "boxed" and extracted data from this channel shown in histogram on the left side of the screen.
MOLECULAR WEIGHT MARKERS

Fig 9  $^{14}$C-molecular weight markers: histogram of the intensity (counts) against distance (cm). Peak 1–200,000; peak 2–92,000; peak 3–69,000; peak 4–46,000; peak 5–30,000; peak 6–14,300.

MOLECULAR WEIGHT CALIBRATION CURVE

Fig 10 Molecular weight calibration curve. Calibration curve computer drawn from markers in fig 9. Six molecular weight markers are plotted as molecular size against distance (cm).
is "boxed in" separately and the information extracted. A histogram of each extracted channel is shown, representing the density and width of each of the protein bands for the particular bacterial strain. The molecular weight standard channel included in each gel is also extracted and analysed.

**Molecular weight calibration**

Fig 9 shows a histogram representing the molecular weight markers. A standard curve can then be drawn by the computer (fig 10) by moving an arrow cursor to the apex of each standard peak and marking the position. The molecular weights of selected peaks from other histograms can then be calculated from this standard curve. Fig 11 shows examples of the histograms and the molecular weights of major peaks from strains of *E coli* run on separate gels. Good reproducibility was also obtained with other bacterial species (*K pneumoniae, P aeruginosa, P mirabilis* etc).

The determination of the molecular weights of the major protein bands provides useful information for certain bacterial species with simple protein patterns, such as *C difficile*: it can be used to differentiate between the various types. The process is more difficult when there is a more complex protein pattern (fig 12) due to a possible overlap of certain peaks in closely related species of bacteria. Fig 11 shows the molecular weights of dominant bands of *E coli*. The complexity of a pattern, however, is not a problem when it is analysed by the Fourier transformation algorithm\(^6^1\) and the Pearson product-moment correlation coefficient.\(^3^6\)

**Pattern recognition software**

A program based on the Fourier transformation of data\(^6^2\) allows comparisons to be made between different gels by transforming the channel data into a wave function and determining the sine and cosine

---

**Fig 11** Scan pattern of radiolabelled proteins from four channels of *E coli*, represented as molecular weight spectra. *A* and *B* from one gel; *C* and *D* from a second gel. The most prominent bands in each channel are marked 1, 2, and 3. Molecular weights of these peaks (\(\times 1000\)) for channels *A*, *B*, *C* and *D* respectively, are: peak 1-96.0, 96.0, 91.1, 99.5; peak 2-58.9, 59.9, 55.8, 58.0; peak 3-45.2, 45.2, 44.4, 45.2.
components of that function. The Pearson product-moment correlation coefficient provides an alternative pattern matching algorithm. These algorithms can be used for matching "unknowns" against a database to provide a potential identification system, or for cluster analysis, to enable the classification of bacteria.

**Rubberbanding**

Another feature of the software is the ability to identify defined segments of the histograms and expand these areas to occupy the full length of the channel. This facility is called "rubberbanding" and provides comparisons between gels of areas of variable molecular weights. The full length of the extracted channel is 360 pixels long. A portion of the channel is selected for comparison—for example, between pixel positions 200 and 60 (fig 13). A “rubberband” adjustment of the channel stretches that area between the selected molecular weights to show with sufficient differences to permit discrimination between the species. Furthermore, constant variations within species also occur. This property makes the establishment of new typing schemes possible.

To maximise the differences between these strains the area on the channels are marked with the pixels and then “rubberbanded.” These areas, which may be of any size, are then examined by the Fourier transformation or the Pearson product-moment algorithms. Table 2 gives an example. *Klebsiella* and *Enterobacter* had a high correlation coefficient with each other (0-67). To distinguish more clearly between these two species the channels and the histograms were visualised and the areas of maximum differences identified. These areas were marked by moving the pixel positions to encompass pixel positions 103–141 on both histograms (fig 14). They now occupy the entire channel length of 360 pixels (fig 15). This procedure is used in the normalisation of the data between different electrophoretic runs and in concentrating on areas of maximum dissimilarity between channels.

**Normalisation**

Even using the most stringent electrophoretic conditions, it is necessary to compensate for small variations in separation due to different gel batches, minor buffer changes, and running conditions. The position of the gel under the scanner head will also vary slightly from scan to scan. To normalise data it is necessary to standardise the channels. This can be achieved by using the molecular weight standard channel. Molecular weights are used as marker positions for “rubberbanding” adjustments to normalise the channel.
Fig 13 Scan data from four Gram negative organisms used as standards. A—Pseudomonas aeruginosa; B—Klebsiella pneumoniae; C—Proteus mirabilis; D—Escherichia coli. The column on the left shows the full channel profile of each organism with arrows marking the pixel positions (60 and 200) used for normalisation. The column on the right are the profiles after normalisation (rubberbanding) and are the patterns used to create the database.

Table 2 Mean (SD) correlation coefficient results using Fourier transformation

<table>
<thead>
<tr>
<th>Database</th>
<th>Salmonella (n = 5)</th>
<th>E coli (n = 5)</th>
<th>Proteus (n = 5)</th>
<th>Pseudomonas (n = 5)</th>
<th>Serratia (n = 10)</th>
<th>Klebsiella (n = 10)</th>
<th>Enterobacter (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmonella</td>
<td>0.93 (0.02)</td>
<td>0.08 (0.13)</td>
<td>0.51 (0.02)</td>
<td>0.00 (0.03)</td>
<td>0.62 (0.15)</td>
<td>0.01 (0.12)</td>
<td>0.28 (0.03)</td>
</tr>
<tr>
<td>E coli</td>
<td>0.13 (0.08)</td>
<td>0.93 (0.03)</td>
<td>0.70 (0.70)</td>
<td>-0.12 (0.04)</td>
<td>-0.16 (0.10)</td>
<td>0.39 (0.08)</td>
<td>0.40 (0.07)</td>
</tr>
<tr>
<td>Proteus</td>
<td>0.57 (0.05)</td>
<td>0.12 (0.07)</td>
<td>0.97 (0.02)</td>
<td>0.57 (0.14)</td>
<td>0.97 (0.15)</td>
<td>0.12 (0.15)</td>
<td>0.33 (0.05)</td>
</tr>
<tr>
<td>Pseudomonas</td>
<td>-0.04 (0.13)</td>
<td>-0.17 (0.04)</td>
<td>0.53 (0.12)</td>
<td>0.95 (0.00)</td>
<td>0.30 (0.16)</td>
<td>-0.24 (0.14)</td>
<td>-0.35 (0.04)</td>
</tr>
<tr>
<td>Serratia</td>
<td>0.51 (0.07)</td>
<td>0.09 (0.13)</td>
<td>0.54 (0.13)</td>
<td>0.11 (0.09)</td>
<td>0.90 (0.06)</td>
<td>0.22 (0.11)</td>
<td>0.48 (0.07)</td>
</tr>
<tr>
<td>Klebsiella</td>
<td>0.01 (0.11)</td>
<td>0.50 (0.06)</td>
<td>0.17 (0.09)</td>
<td>0.21 (0.05)</td>
<td>0.09 (0.17)</td>
<td>0.82 (0.10)</td>
<td>0.67 (0.11)</td>
</tr>
<tr>
<td>Enterobacter</td>
<td>0.22 (0.17)</td>
<td>0.33 (0.11)</td>
<td>0.42 (0.13)</td>
<td>-0.20 (0.06)</td>
<td>0.32 (0.15)</td>
<td>0.67 (0.20)</td>
<td>0.86 (0.12)</td>
</tr>
</tbody>
</table>
Fig 14 Normalised pattern for Klebsiella aerogenes above and Enterobacter cloacae below; correlation coefficient 0·67. Arrows indicate pixel area to be "rubberbanded" between 103 and 141.

lengths, so that data from one gel can be compared with data from another (table 3). The areas between these molecular weights differed among the six gels and can only be compared after normalisation. This degree of normalisation automatically compares point to point simple subtraction of raw differences between the "standard" and the "unknown" channels

Fig 15 An expansion of the areas indicated by arrows in fig 14. The resulting correlation coefficient is 0·20 (table 6).

Fig 16 Normalisation adjustment for Pearson product-moment correlation coefficient. A: column 1—full channel profile of "standard" K pneumoniae; column 2—full channel profile of "unknown" K pneumoniae; column 3—simple subtraction comparison of column 1 and 2 showing point to point differences. The resulting correlation coefficient is 0·06; B: the same as A, after normalisation between molecular weights 200 000 and 30 000. The resulting correlation coefficient is 0·60; C: the same as B, after column 2 has been moved 5 pixel positions downwards. The correlation coefficient is now 0·96. D: column 1—the same as in B—that is, K pneumoniae after normalisation; column 2—E coli, normalised between molecular weights 200 000 and 30 000; column 3—the differences between columns 1 and 2. The correlation coefficient is 0·76 and this correlation does not improve with changes in pixel movement in column 2.

Fig 17 Identification of bacteria by pattern matching of radiolabelled proteins.
difficile type

Table 3  Pixel positions used for normalising gels

<table>
<thead>
<tr>
<th>Gel No</th>
<th>Pixel positions at molecular weights:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200 000</td>
</tr>
<tr>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
</tr>
<tr>
<td>5</td>
<td>66</td>
</tr>
<tr>
<td>6</td>
<td>52</td>
</tr>
</tbody>
</table>

Table 4  Mean (SD) molecular weight values of 14 C difficile type strains

<table>
<thead>
<tr>
<th>C difficile type</th>
<th>No of estimations</th>
<th>Molecular weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Band 1</td>
</tr>
<tr>
<td>A</td>
<td>11</td>
<td>51 439 (901)</td>
</tr>
<tr>
<td>A2</td>
<td>12</td>
<td>52 328 (1084)</td>
</tr>
<tr>
<td>B</td>
<td>11</td>
<td>48 036 (372)</td>
</tr>
<tr>
<td>C</td>
<td>11</td>
<td>43 143 (884)</td>
</tr>
<tr>
<td>C2</td>
<td>11</td>
<td>44 233 (835)</td>
</tr>
<tr>
<td>D</td>
<td>7</td>
<td>46 174 (918)</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>47 341 (985)</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>47 905 (1149)</td>
</tr>
<tr>
<td>G</td>
<td>11</td>
<td>46 219 (710)</td>
</tr>
<tr>
<td>W</td>
<td>10</td>
<td>45 930 (1162)</td>
</tr>
<tr>
<td>X</td>
<td>12</td>
<td>46 457 (894)</td>
</tr>
<tr>
<td>X2</td>
<td>6</td>
<td>46 925 (793)</td>
</tr>
<tr>
<td>Y</td>
<td>9</td>
<td>44 681 (1009)</td>
</tr>
<tr>
<td>Z</td>
<td>12</td>
<td>55 961 (1180)</td>
</tr>
</tbody>
</table>

the molecular size of the major bands. It was apparent, however, that “X” and “X2” were indistinguishable and that differences between type “X” and “W” were minimal, even though they can clearly be seen on the autoradiograph. A database for the 14 types of C difficile using the molecular weights and values is in preparation.

Identification of Gram negative bacteria

Gram negative bacteria gave complex protein patterns. E coli, P aeruginosa, K pneumoniae and Proteus mirabilis were run in duplicate on three gels and scanned. The major peaks were identified and their molecular weights determined (figs 11 and 12). There was good reproducibility among the various runs of the same species on a single gel and on different gels. There was, however, some overlap of major bands among the different species, but each bacterial species has at least one protein band that is individual to it. Furthermore, the overall pattern for each organism is quite different and specific for the species (fig 12).

Although determining the molecular weights of major bands provided useful information, it was ultimately time consuming when the protein pattern was complex. The major bands had to be identified manually, making this program unsuitable for the identification of these bacteria.

IDENTIFICATION OF BACTERIA BY FOURIER TRANSFORMATION

For the identification of bacterial species with complex protein patterns, the Fourier transformation method of analysis was introduced. Seven Gram negative organisms (Salmonella typhimurium, E coli, Proteus mirabilis, P aeruginosa, Serratia marcescens, K pneumoniae and Enterobacter cloacae) were electrophoresed on gels supplied by AMB, using the flat bed automated electrophoresis unit described previously, and then scanned. The organisms were run singly on six gels, except for Serratia and Klebsiella which were run in duplicate on each gel. Full length channels were extracted, 360 pixels long, using a cursor width of 5 pixels. Then seven channels from one gel were “rubberbanded” between positions 53–213 and used as “standards” to construct a database using the fast Fourier transformation algorithm. The remaining 46 channels were regarded as “unknowns”, and after “rubberbanding” between positions 53–213 the database was searched to find the best match for that channel. The name of the best match was then produced, along with a correlation coefficient for the match. A list of the next best matches with correlation coefficients was also provided. Fig 17 shows an example of an “unknown” entered into the computer, which searches the database and provides a visual “best match” as E coli. Five to ten channels of the
same organism run on five gels were treated as “unknowns”. Each “unknown” channel showed a “best match” against its correct corresponding species in the database (table 2). A total of five E coli channels matched correctly, giving an average similarity coefficient of 0·93, a standard deviation (SD) of 0·08, and a much lower similarity coefficient with the other species. Similarly, all other “unknowns” matched correctly with their corresponding species with a mean similarity coefficient ranging from 0·82 to 0·97 (table 2).

Identification of bacteria by the Pearson product-moment correlation coefficient

The Pearson product-moment algorithm was applied to distinguish between closely related bacterial species. Channels corresponding to E coli and K pneumoniae from one gel were used using the algorithm, and 28 channels of the same two organisms run on different gels were treated as “unknowns”. Normalisation was achieved as before and the unknown “channels” were further adjusted by using 10 different pixel positions up and down to find the highest correlation coefficient (Fig 16). Table 5 shows the correlation coefficient results before and after fine adjustments.

Taking the standard K pneumoniae and comparing it with the 14 remaining K pneumoniae channels, the mean result was 0·80 with a SD of 0·04. Conversely, when E coli was used as a standard, the correlation coefficient for the other E coli strains was 0·88 (SD 0·04) and with K pneumoniae channels 0·77 (SD 0·03).

Using the algorithm, “unknown” K pneumoniae matched against their standard and were clearly differentiated from the E coli standard. Although unknown E coli correlated at 0·80 with the standard K pneumoniae (Fourier 0·46), they were clearly matched with the E coli standard because the standard deviations were low. Analysis by the Pearson product-moment, therefore, can further distinguish between organisms which had been closely matched by the Fourier transformation.

Discrimination between closely related bacterial species and the development of typing schemes

Bacterial species that are genetically closely related as shown by DNA homology studies will inevitably produce similar protein profiles, such as those between E coli and Shigella and between Klebsiella and Enterobacter. Despite the close similarity of the patterns, however, there are usually segments in the channels with sufficient differentiating characteristics between the species. Furthermore, constant variations between species also occur. New typing schemes are therefore feasible.

To maximise the differences between these strains the area on the channels are marked with the pixels and then “rubberbanded.” These areas, which may be of any size, are then examined by the Fourier transformation or the Pearson product-moment algorithms (table 5). Klebsiella and Enterobacter had a high correlation coefficient with each other (0·67). To distinguish more clearly between these two species the channels and the histograms were visualised and the areas of maximum differences identified. These areas were marked by moving the pixel positions to encompass pixel positions 103–141 on both histograms (fig 16). The areas between those pixel positions are then “rubberbanded” (fig 15). Table 6 shows the comparison of Klebsiella and Enterobacter within that closely defined area, analysed by the Fourier transformation. The correlation coefficient between Klebsiella and Enterobacter is now 0·2 (0·21) and between Enterobacter and Klebsiella —0·14 (0·21) instead of 0·67.

Good discrimination was also possible between E coli and Klebsiella pneumoniae (fig 16) and between Shigella and E coli. This procedure therefore permits discrimination between closely related bacterial species and could also be used in identifying differences within the species to provide new typing schemes.

Classification of bacteria by cluster analysis

Further computation of the data is being developed to provide a similarity matrix which can be clustered by a program like the unweighted pair group method using averages (UPGMA) and plotted as a dendrogram. This will be particularly useful in expressing similarity and dissimilarity between organisms. Similar organisms will be grouped together, and dissim-

<table>
<thead>
<tr>
<th>Challenge organisms</th>
<th>Klebsiella (n = 10)</th>
<th>Enterobacter (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Klebsiella</strong> (n = 2)</td>
<td>0·83 (0·12)</td>
<td>—0·14 (0·20)</td>
</tr>
<tr>
<td><strong>Enterobacter</strong> (n = 1)</td>
<td>0·20 (0·21)</td>
<td>0·93 (0·03)</td>
</tr>
</tbody>
</table>
ilarity will be defined as the “taxonomic distance” between the positions of each unknown in what Sneath has described as “phonetic hyperspace.”

It is clear that several approaches need to be adopted for the analysis of the data and the construction of database libraries. It is also essential to accept that different algorithms may have to be used for different taxonomic ranks. Further developments are in progress.

Conclusions

Contaminants are the scourge of the microbiologist, but very occasionally can lead the worker into exciting new discoveries. This work started with a chance contamination of an experiment and ended with the development of an automated new approach to the identification and classification of micro-organisms.

The method entailed the incorporation of $^{35}$S-methionine into bacterial proteins and subsequent separation of the radiolabelled proteins by SDS-polyacrylamide gel electrophoresis. The resultant protein patterns were species specific, permitting discrimination between the species. Sufficient differences in the protein profiles within species were also present, so that subdivision or typing of the species could be done.

The first section of the work used autoradiography to visualise the protein patterns, but this procedure was time consuming and therefore an automated approach was introduced. An automated scanner was developed to scan the gels, and the information was then stored in a computer. Scan data were then retrieved and processed using software that permitted quantitation of protein bands and construction of molecular weight standard curves. Further development of pattern recognition software allowed databases to be constructed and data obtained from different gels to be compared. This facility correctly identified duplicate “unknowns” from other gels.

This method lends itself to two areas of study. The first is a practical application whereby the construction of database libraries would be available for the identification of bacteria. Initially it would be used in research where classical methods are inadequate, as for C. difficile, MRSA, and coagulase negative staphylococci. The second application would entail the construction of dendrograms to show “natural” relations between bacteria to complement the phonetic classification.

The principle of the method as a taxonomic aid remains the same, regardless of the micro-organisms. It can be extended to viruses, protozoa, and fungi. The proteins generated represent the genetic expression of the organism and will provide patterns for new species or strains. These pattern differences form the basis of classification into different classes. Once a pattern classification system is established, unknown organisms can be identified by pattern matching.

This technique provides a general, rapid, and simple automated method for the identification of bacteria. Computer software is constantly being upgraded according to future requirements and new challenges. It is predicted that this method will soon be widely applicable in clinical microbiological laboratories.

Appendix of methods

Bacterial strains
These were isolated from clinical specimens in the department of medical microbiology at this hospital or obtained as NCTC and ATCC strains. Some strains were sent to us from other hospitals in the United Kingdom and from abroad. All strains were identified using standard microbiological techniques, checked for purity, and cultured on horse blood agar plates before being subjected to $^{35}$S-methionine incorporation and electrophoresis.

Radiolabelled compounds
L-$^{35}$S-methionine (Amersham International, United Kingdom) was acquired as an aqueous solution in 20 mM potassium acetate containing 0.1% 2-mercaptoethanol. The specific activity ranged from 600-1450 Ci/mmol.

Molecular weight markers
$^{14}$C-methylated protein markers consisted of six proteins with molecular a weight range of 14,300 to 200,000 daltons. The radioactive concentration of each protein was 0.833 Ci/ml.

Incorporation of $^{35}$S-Methionine into bacterial proteins
Methionine free Eagle’s medium (50 ul) or Difco containing 1 ul of $^{35}$S-methionine (about 10 μCi) was inoculated with a portion of one to two colonies of the pure culture from a blood agar plate. The mixture was then incubated at 37°C for 30 minutes to two hours, depending on the organism. Anaerobes were incubated in the anaerobic chamber (80% nitrogen, 10% hydrogen, 10% carbon dioxide). At the end of the incubation period an equal volume of buffer (0.125 M Tris hydrochloric acid, 10% mercaptoethanol, 4% sodium dodecylsulphate (SDS), 20% glycerol, pH 6.8) was added and the mixture boiled for two minutes. After cooling to room temperature 10 to 40 ul of the sample was then applied to a polyacrylamide gel and electrophoresed. Twelve to 18 samples could be applied to each gel.
Automation in clinical microbiology

Electrophoresis (Vertical Slab Gel Electrophoresis and Autoradiography)

The system was based on the SDS discontinuous buffer system based on the method of Laemmli. The resolving polyacrylamide gel was cast with total polymer 12.5% and crosslinker 2.6% in buffer (0.375 M Tris (hydrochloric acid), 0.1% SDS; pH 8.3). The stacking gel had a total polymer 3% and crosslinker 2.6% in buffer (0.125 M Tris hydrochloric acid, 0.1% SDS; pH 6.8). The electrode buffer (0.025 M Tris, 0.192 M glycine, 0.1% SDS; pH 8.3) was placed in the upper and lower reservoirs and electrophoresis carried out at 4°C using a constant voltage (Shandon or Biorad). The voltage was set at 60 volts for 15-20 minutes to allow the sample to migrate from the wells into the stacking gel. The voltage was then increased to 150-200 volts and electrophoresis continued for three to four hours. The run was ended when bromphenol blue, a dye marker added to the sample, reached the end of the gel.

The gels were fixed by immersion in an aqueous solution of 20% glacial acetic acid and 20% isopropanol for about three hours. After fixing they were dried using a gel slab drier (Biorad) by heating under vacuum for about one hour. The dried gels were then placed on x-ray film (Fuji Rx) between two glass plates secured by clips and left in a light proof box. The optimum exposure time was determined empirically for each gel and ranged from 18 hours to five days. The time depended on the amount of radioactivity in each protein band.

Electrophoresis (Semiautomated Horizontal Electrophoresis System)

Gels
Discontinuous SDS-PAGE gels are supplied by AMB. Cast on gel-bond (FMC) consists of a stacking gel 4 cm in length and a 12.5% acrylamide resolving gel about 13 cm long. The gel surface is protected by a sheet of Mylar. Three lines are drawn on the back of the gel-bond to standardise wick placement and sample application. The lines for the wicks are drawn at a distance of 1.5 cm from the top and bottom of the gel. The line for the sample application strip is drawn 0.5 cm below the top wick line.

Electrophoresis conditions
Electrophoresis running buffer (250 ml) is poured into each buffer tray. The plate surfaces on the electrophoretic unit (fig 7) are cleaned with methanol and cooled to 7°C. One plate at a time is covered with a thin film of n-decane (BDH) and a gel is centred on to the plate (stacking gel at the top), using more decane as necessary to remove all the air bubbles. The Mylar protection sheet is then removed from the gel. The wicks (AMB) are about 8 x 18 cm in size and four wicks are used for each gel, two for each buffer tank. The two pieces are placed with the absorbent surfaces together, offset by 2 cm at the edge. The wicks are then dipped into the buffer to ensure saturation and the 2 cm uncovered absorbent surface placed on to the gel to line up with the wick placement lines. The wick is smoothed to ensure good contact with the gel and to remove any trapped air. A sample application strip is laid carefully on to the sample application strip line, leaving a 0.5 cm gap between it and the wick. The wells in the strip hold up to 12 µl of sample, and 12 samples could be applied on one gel. Each sample is pipetted carefully into the wells. The Mylar sheet is replaced on to the gel surface, the top of the sheet lying 0.5 cm below the bottom of the sample application strip. The lid of the electrophoresis unit is then closed and a program selected on the IBM computer to give running conditions of 20 watts for 30 minutes to allow the samples to enter the gel and to stack. Power is then automatically increased to 60 watts for sample separation in the resolving gel. The temperature of the platens is programmed to remain at 7°C and the end of run sensor is activated. Under these conditions, an electrophoresis run takes about 90-100 minutes. At the end of the run the gel is lifted off the platen and immediately dried in a forced air drier, without prior fixing.

This work was supported by Automated Microbiology Systems Ltd, the Wellcome Trust, The Medical Research Council, and the Joint Research Board of St Bartholomew’s Hospital. We acknowledge the assistance of Mrs C Whittfield in typing the manuscript and preparing the figures.

References
8 Kaper JB, Bradford HB, Roberts NC, Falkow S. Molecular epi-


Automation in clinical microbiology


Requests for reprints to: Professor Soad Tabaqchali, Department of Medical Microbiology, St Bartholomew's Hospital Medical College, West Smithfield, London EC1A 7BE, England.
Automation in clinical microbiology: a new approach to identifying micro-organisms by automated pattern matching of proteins labelled with 35S-methionine.

S Tabaqchali, R Silman and D Holland

doi: 10.1136/jcp.40.9.1070

Updated information and services can be found at: [http://jcp.bmj.com/content/40/9/1070](http://jcp.bmj.com/content/40/9/1070)

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to: [http://group.bmj.com/group/rights-licensing/permissions](http://group.bmj.com/group/rights-licensing/permissions)

To order reprints go to: [http://journals.bmj.com/cgi/reprintform](http://journals.bmj.com/cgi/reprintform)

To subscribe to BMJ go to: [http://group.bmj.com/subscribe/](http://group.bmj.com/subscribe/)