Review

Current status and perspectives of computer-aided two-dimensional densitometry

T. Toda[a],[*], M. Ohashi[b]

[a]*Department of Molecular Biology, Tokyo Metropolitan Institute of Gerontology, 35-2 Sakaecho, Itabashi-ku, Tokyo 173, Japan*

[b]*Clinical Laboratory, Cancer Institute Hospital (Japanese Foundation for Cancer Research), 1-37-1 Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan*

Abstract

Two-dimensional electrophoresis is a powerful technique for demonstrating cellular changes in protein synthesis and post-translational modification. The method of computer-aided two-dimensional densitometry was developed to facilitate the comparison of gels. However, powerful minicomputers were required to execute the sophisticated programs in the early stages of development. Recent advances in workstations have given rise to commercially available desktop systems on which the full functions of two-dimensional densitometry including spot matching and database construction are available. Thus two-dimensional gel protein databases of various cell types have been accumulating, and some of them are accessible through ExPASy server on the Internet computer network.

Contents

[*] Corresponding author.

1. Introduction

High-resolution two-dimensional electrophoresis, reported by O'Farrell [1] in 1975, had a great impact because the resolution of denatured polypeptides on his gels was so excellent as to demonstrate the majority of proteins expressed in a single cell type. The high reproducibility of separation allowed us to find physiologically significant proteins that appear as unmatched spots between a sample gel and its corresponding control gel. O'Farrell's report stimulated many groups to develop the method of computer-aided two-dimensional densitometry for quantitative and qualitative analysis of protein spots on the gel. Bossinger et al. [2] of the University of California at San Diego (UCSD), Lemkin et al. [3] of the National Institutes of Health (NIH) and Anderson and co-workers [4,5] of the Argonne National Laboratory (ANL) developed systems for two-dimensional densitometry using minicomputers that already existed in their institutions. Garrels [6] of the Salk Institute constructed a desktop system using a Hewlett-Packard computer. The desktop system was well designed to perform automatic detection and quantification of spots on such a small computer; however, the extended functions of spot matching and database construction were not accessible with this system because of the inadequate memory. The NIH group [7,8] and ANL group [9] independently developed sophisticated software for automatic spot matching and database construction. Garrels et al. [10,11] transcribed his original programs into QUEST software for a minicomputer-based system on which automatic spot matching and database construction were achieved. The software was then made available for commercial distribution under the name PDQUEST through PDI in 1985. Bio Image software, which is another commercial software developed from the foundation studies of Robert et al. [12], also reached a comparable level to PDQUEST software. Thus the full functions of two-dimensional densitometry, including spot matching and database construction, because more accessible to biochemists and molecular biologists. Subsequently, publications on the method itself decreased whilst and reports on the

applications [13–20] increased. We have now reached a new stage in applications of two-dimensional densitometry. Architectures of hardware and algorithms of software that have been developed for two-dimensional densitometry are reviewed in this paper. Additionally, the quantitative aspects of two-dimensional densitometry and the perspective of two-dimensional gel protein databases are discussed.

2. Quality of two-dimensional gel

Sufficient resolution of protein spots on a two-dimensional gel pattern is a key factor in making the most of two-dimensional densitometry. Even though most software for two-dimensional densitometry splits interior spots in a multi-spot area automatically or interactively, reliable quantification is expected only for discrete spots. When accurate quantification of protein spots is required for two-dimensional densitometry, special care must be taken that the protein loading does not exceed the range of linearity of the detection method. In the case of autoradiography, the exposure of films should be well controlled so as not to exceed the saturation density level of the film.

When gel patterns are compared by two-dimensional densitometry to screen out the unmatched spots among them, high reproducibility is required of the two-dimensional gel patterns. Reproducible electrophoresis may reduce errors in automatic spot matching. Isoelectric focusing on a ready-made thin-layer gel strip of an immobilized pH gradient is desirable for reproducible isoelectric focusing in the first-dimensional separation. This might be especially effective in the inter-laboratory comparison of gel patterns obtained in independent laboratories by different technicians with various skills.

3. Quantitative aspects of two-dimensional densitometry

Applications of two-dimensional densitometry in the field of molecular biology tend to place more importance on the function of spot match-

ing and database construction. However, accuracy in optical quantification is a fundamental quality of densitometry. The determination of proteins in solution by a photometric measurement is theoretically based on the Lambert–Beer law:

$$A = \varepsilon \, lc$$

where A, ε, l and c are the absorbance, the molar absorptivity, the length of the light path and the molar concentration, respectively. The law is not directly applicable to densitometric measurements of stained proteins on a gel plate or an autoradiogram on a film, because the sample is not in the form of a solution and the molar absorptivity does not come directly from the physico-chemical properties of the protein molecules themselves. However, the determination of the amount of protein in the spots on a two-dimensional gel pattern is not impossible provided that the absorbance of the revealed protein spots is proportional to the amount of protein in a unit of pixel area. The reliability of densitometry depends primarily on the quality of the optical system used for image acquisition. For example, when a video camera of a charge-coupled device (CCD) is used for image acquisition, the difficulty in achieving even illumination all the camera view is a problem. This problem can be overcome by using a "blank image subtraction method" as discussed in a previous paper [21]. In brief, the absorbance of the light is determined according to the

$$A = -\log \frac{I}{I_0}$$

where A, I_0 and I are the absorbance, the intensity of the incident light and the intensity of the transmitted light, respectively. Even if the incident light is not even, the ratio I/I_0 is not affected by the place-to-place variance in brightness of the incident light. I_0 and I are actually measured as the intensities of an electric signal from the CCD camera when a sample is absent and present in the camera view, respectively. The effect of the "blank subtraction" is simply understood if the former equation is transformed into

$$A = \log I_0 - \log I$$

The subtraction of logarithmically converted "sample image data" ($\log I$) from logarithmically transformed "blank image data" ($\log I_0$) gives the absorbance (A). However, the absorbance is not directly related to the amount of protein if the densitometry is applied to coloured protein spots. Ultraviolet imaging densitometry of unstained gels, reported by Yamamoto et al. [22], brought a new approach to direct determination of proteins according to the physico-chemical properties of the protein molecules themselves. Calibration of absorbance levels using a standard film of authentic protein spots is helpful in determining the absolute amounts of proteins on a sample gel plate.

4. Hardware for two-dimensional densitometry

The minimum requirement of the hardware for image analysis including two-dimensional densitometry is a combination of a powerful computer with an optical reader for image acquisition and a video terminal to display the image. Computer-aided image analysis has been under investigation since the beginning of the computer age. However, it remained difficult to accomplish real-time image processing on a standard multi-purpose computer system because of the lack of image input/output facilities. A rotating drum scanner was developed for reading X-ray diffraction films in the early stages of X-ray crystallography. Stage scanning optical readers were developed in the field of astronomy to analyse telescopic photographs. A real-time picture processor (RTPP) in combination with a microscopic photometer and a grey-level CRT display was established for use in histology [23]. Thus the early generation of two-dimensional densitometers was based on such earlier image processing facilities.

Garrels [6] assembled his original hardware with use of a Zeiss microscopic photometer and a Hewlett-Packard desktop computer (Fig. 1). He realized the two-dimensional densitometry in his unique "line analysis" algorithm on the small desktop computer without extra memory extension. Lipkin and Lemkin [7,8] built up their two-dimensional gel analysis system on the
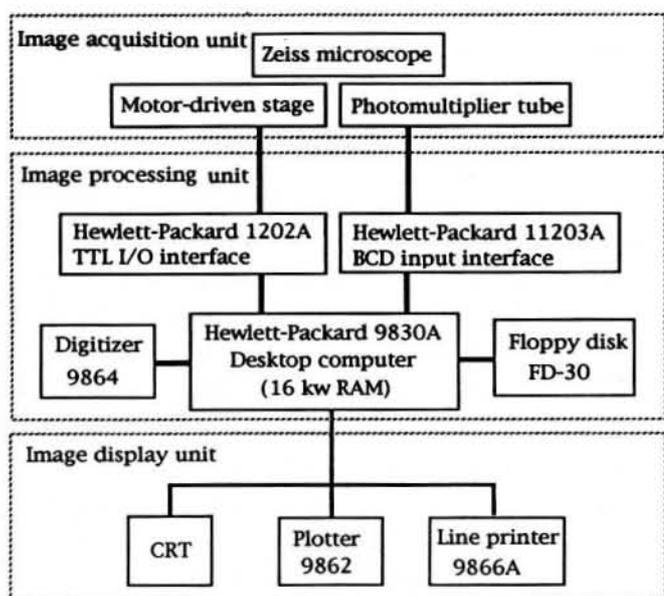
Fig. 1. Hardware architecture of the desktop computer-based system reported by Garrels [6].

Fig. 3. Hardware architecture of the TYCHO system [24].

RTPP system (Fig. 2). Anderson and co-workers [24,25] developed the PDP11-based TYCHO system (Fig. 3) for promoting their "molecular anatomy" project. Garrels et al. [10,11] ultimately established their minicomputer-based QUEST system on the basis of his original
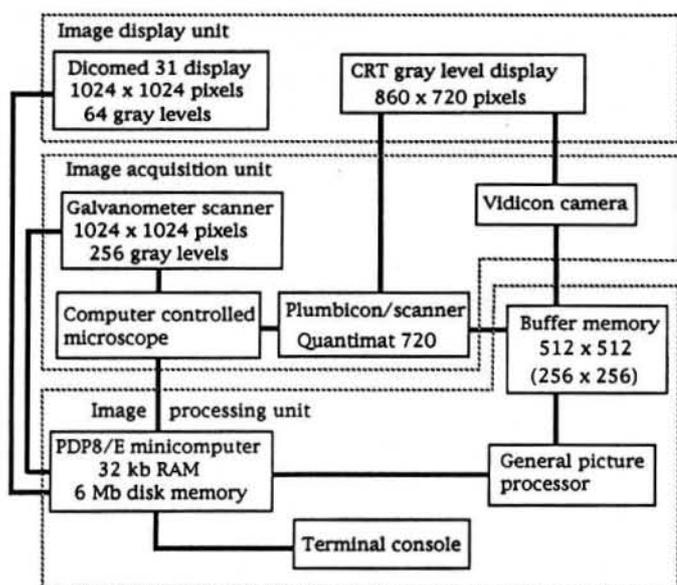
algorithm. They recompiled the QUEST software to run on an UNIX workstation-based commercial system (Fig. 4).

Many other groups including ours, to whom minicomputers were inaccessible, tried to access the two-dimensional densitometry by assembling the hardware system with a personal computer and an economical video camera. We [21] constructed the FIMC hardware with an 8-bit microprocessor, 4 pages of frame image memory, the

Fig. 2. Hardware architecture of the minicomputer-based GELLAB system [3,7] that was constructed on the RTPP system [23].

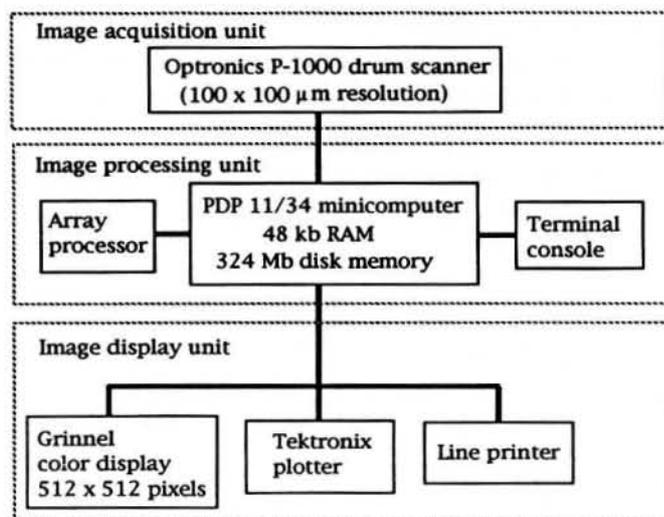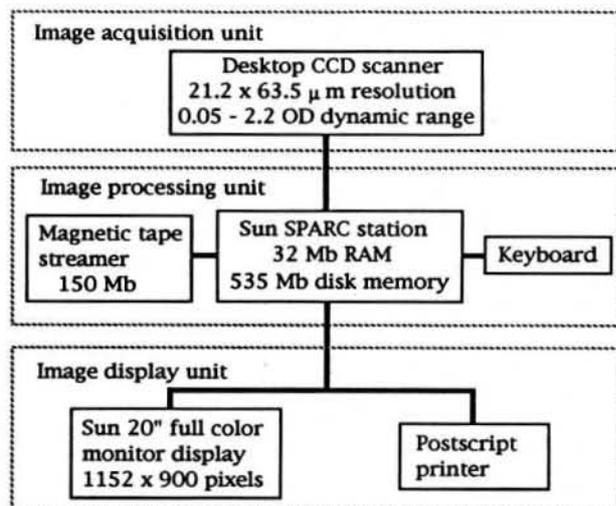Fig. 4. Hardware architecture of a typical UNIX workstation-based system on which the commercial PDQUEST software can run.

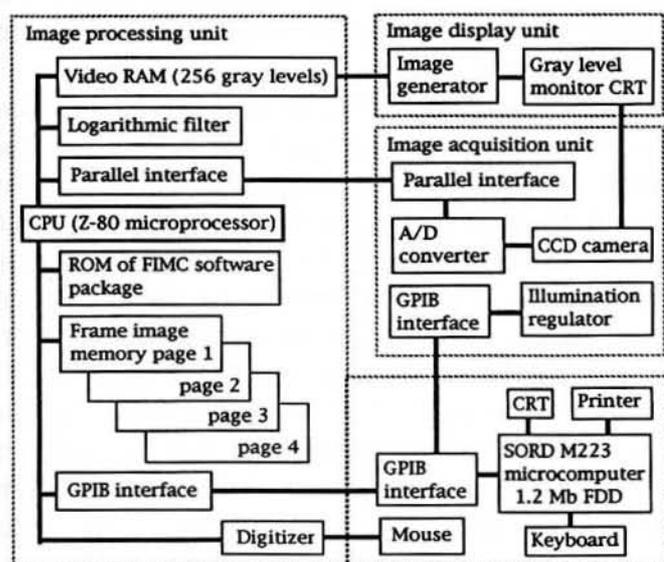Fig. 5. Hardware architecture of the microcomputer-based FIMC system [14].

memory. The results were displayed on a 256 grey level CRT screen through the video RAM.

4.1. Image acquisition unit

Rotating drum scanners were used for acquiring autoradiograms in the early generations of minicomputer-aided two-dimensional densitometers. The resolution was about 150×100 μm, which was sufficient for reading spots on a large two-dimensional gel pattern prepared according to O'Farrell's original method. Pixel data transferred from the scanner in 8-bit unsigned integers were written on a magnetic tape. Kronberg et al. [26,27] used a scanning-microscope photometer for reading fine gel patterns.

The adaptation of a CCD camera allowed real-time image acquisition, although it gave rise to another problem, namely achieving an "absolutely even illumination" on the field of camera view. A desk-top scanner with a line array CCD is a standard optical reader which is connected to a host computer through a GPIB interface. Maytin et al. [28] introduced the laser scanner, which is expected to be a high quality optical reader. Typical specifications of three types of optical readers, which are now commercially available, are summarized in Table 1. Visualization of protein spots by staining or autoradiography is required for the above types of visible-light optical readers. The method of ultraviolet imaging densitometry reported by Yamamoto et al. [22] offered a new concept of densitometry of unstained protein.

video RAM; the logarithmic filter, the GPIB interface and the CCD camera (Fig. 5). The FIMC software package was prepared on the SORD microcomputer using the assembler language, and the machine codes were written in the ROM on the FIMC hardware. The interactive two-dimensional densitometry was executed by giving parameters to the BASIC main program on the SORD microcomputer which controlled the image processing unit through GBIP interface. Each two-dimensional gel was put on the computer-controlled transilluminator and the image was acquired with the CCD camera. The pixel data were handled in the frame image

Table 1
Typical specifications of commercially available optical readers

Parameter	CCD scanner	CCD camera	Laser scanner
Light source	Fluorescent lamp	Cathode grid lamp	Helium–neon laser
Optical sensor	Line array CCD	Matrix array CCD	Photomultiplier
Colour separation	Filter (red, green, blue)	IR filter (400–700 nm)	633 nm
Linear absorbance range	0–2	0–3	0–4
A/D conversion	8 bits (256 levels)	12 bits (4096 levels)	12 bits (4096 levels)
Maximum gel size	30×40 cm (scan area)	27×27 cm (lamp house size)	22×25 cm (scan area)
Maximum resolution	21.1×63.5 μm	1024×1024 pixels per frame	40 μm (beam diameter)
Reading speed	8 min per field	<1 s per frame	1 min per line

4.2. Data processing unit

When absorbance values ranging from 0 to 2.5 are digitized in 8-bit binary integers, the pitch is about 0.01 absorbance. This means that the sensitivity is about 0.01 absorbance. In most two-dimensional densitometers, the pixel data are preserved in the random access memory. A megabyte of memory is required to preserve a frame of 1024×1024 pixel data in 8-bit binary integers. The resolution of the image data correspond to ca. 25×25 μm when a full image of a 25×25 cm gel is acquired. The large memory occupation for image storage was a heavy burden even on multi-purpose minicomputer systems until the RTPP was developed. The extension of the frame image memory in the RTPP unit released the host computer from the large memory occupation. Bossinger et al. [2] and Anderson et al. [24] adopted another method of temporary image storage on magnetic memory media. Image processing was performed by a batchwise execution of programs afterwards. In the desktop system of Garrels [6], pixel data were processed line-by-line immediately after each scan and the resulting parameters of peaks were stored in the core memory. The method of line-by-line analysis did not require a large memory space for data storage.

The appearance of advanced UNIX workstations changed the situation for computer-aided two-dimensional densitometry. They have a sufficiently large memory space to preserve the raw image data. The execution of programs for the extended two-dimensional densitometry, which were developed for minicomputer-aided systems, is accessible on desktop workstations. The speed of image processing also approaches the level of previous minicomputers.

4.3. Image display unit

Two-dimensional gel images are displayed on a grey level monitor or on a color monitor CRT during and after the image processing. The 256 gray level image on a 1024×1024 pixel monitor is desirable for the real display of a high-resolution two-dimensional gel pattern, but the 64 gray level image on a 860×720 pixel monitor is enough for the interactive image processing on the screen practically.

Before computer-aided image processing was intended, the grey level monitor fitting for displaying fine images had been unavailable on any general multipurpose computer. The processed images were displayed in the binary bit map image or in a contour image or in a contour image in the early stages of development. Carman et al. [23] introduced the Dicomed 31 display into their RTPP system for achieving the gray level display. Anderson and co-workers [24,25] utilized the Grinnel color display in their TYCHO system. We realized the gray level display on our microcomputer-aided FIMC system by installation of the video RAM and grey image generator.

Both gray level display and full color display are commercially available on most microcomputer systems and UNIX workstations at present. High-quality image printers and postscript laser printers are generally used to get a hard copy of the processed image. Color printers are also available.

5. Software for two-dimensional densitometry

The image processing of two-dimensional densitometry proceeds step by step (Fig. 6). Most software for two-dimensional densitometry is a package of program modules that work in each step. The minimal package for primitive two-dimensional densitometry from the image acquisition through the spot quantitation is available on both minicomputers and microcomputers. Program modules for extended activities after spot quantitation were chiefly prepared for minicomputers, but some of them works on UNIX workstations as well. The compatibility of software is sustained by the availability of the same C language compiler.

Various software for two-dimensional densitometry has been developed on different environments of hardware by independent groups. Bossinger et al. [2] encoded their source programs for

```
┌─────────────────────────────────────────┐
│  Image acquisition and pixel data storage │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│     Noise reduction for image smoothing   │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│         Background subtraction            │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│  Spot detection and boundary determination │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│  Spot quantitation and density calibration │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│      Spot matching and gel comparison     │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│          Database construction            │
└─────────────────────────────────────────┘
```
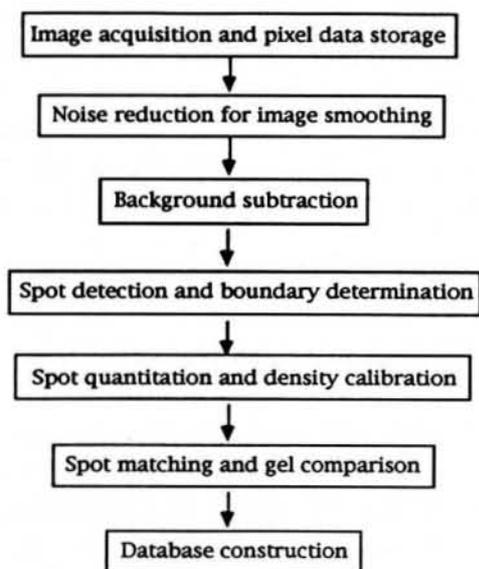
Fig. 6. General procedure of two-dimensional densitometry.

two-dimensional gel analysis in C language. Lemkin et al. [3] established their original SAIL language, and they wrote the GELLAB program in the original language. Garrels [6] developed his original software for two-dimensional densitometry on a Hewlett-Packard desktop computer in BASIC language first. Subsequently, he and his co-workers rewrote the advanced QUEST software for the UNIX minicomputer system in C language. We wrote subroutine programs to execute real-time image processing on our microcomputer-based FIMC system in assembler language. The main program for the interactive performance of two-dimensional densitometry was written in extended BASIC language. The subroutine programs were "called" from the main program with FIMC commands.

5.1. Image acquisition

An analogue signal from a CCD image sensor is amplified and converted into an 8–12-bit binary integer through the A/D converter. The main task of the program for image acquisition is to receive the digitized values from the optical reader and write them in the RAM or on a magnetic memory medium. The software for a scanning-microscope photometer includes a subroutine program to control the stage movement. Most software for a CCD camera includes a program to control the light source. When the pixel data from the optical reader are "noisy" because of the lamp fluctuation or instability of the image sensor, taking the average of the pixel data by repetitive scanning is effective in reducing the "noise". If the raw pixel data received from the optical reader relate to intensities of the light transmittance, the data are subjected to a logarithmic conversion and a binary negation to obtain the relative absorbance. The theoretical considerations for the spectrometric quantification of proteins on a gel plate have already been mentioned. When the pixel data are still noisy, the noise should be removed by "smoothing" before stepping to the next "background subtraction" stage. On most systems for two-dimensional densitometry, signals from the image acquisition unit indicate intensities of transmitted light (I value). When the signals are logarithmically converted first and then digitized into 8-bit binary integers ranging from 0 to 255, the step size is 0.01 AU in the full range of 2.5 AU. Therefore a byte of memory is enough to store a pixel data in optical density level. However, more bits of memory is required if the signals from the image acquisition unit are digitized before logarithmic conversion. Fig. 7 shows the results of 8-bit binary digitization of optical intensity values. In this case, the intensity of incident light (I_0) is defined as the maximum of transmitted light $(I_{max} = 255)$. The step size is
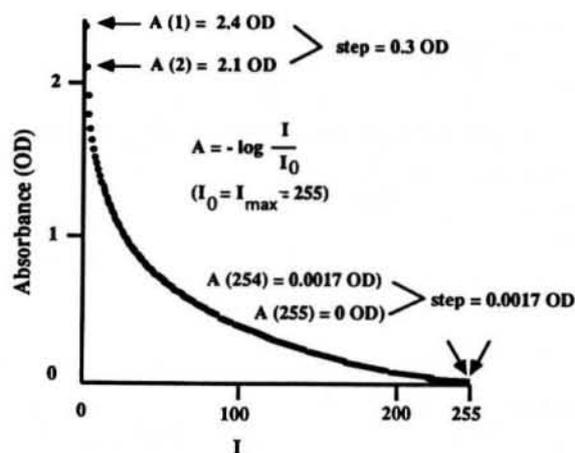

Fig. 7. Absorbance levels obtained from the 8-bit binary digitized intensities of transmitted light (I values). "OD" = Absorbance units.

small (0.0017 AU) in the low absorbance range, but large (0.3 AU) around 2.4 AU. The step sizes and the dynamic ranges given by 8-, 10- and 12-bit binary digitization of I values are summarized in Table 2. Twelve-bit binary digitization of transmitted light intensity is enough to achieve the smaller step sizes than 0.1 AU in the dynamic range of 0–3 AU. After logarithmic conversion, the data may be stored in a 8-bit binary memory.

When insufficient memory space is available to preserve the raw pixel data, compression of the data to parameters about positions and height of peaks saves memory space. Garrels [6] introduced his algorithm of "line-by-line" peak detection to overcome the insufficient memory space of his desktop computer-aided system. As the optical data from single line scanning were received from the scanner, each peak of ab-

Table 2
Dynamic range and steps of absorbance levels given by 8-, 10- and 12-bit binary digitization of I values

---

(I) 8-bit binary digitization of I value ($I_{max} = 255$)

$$\frac{A(1) = 2.047}{A(2) = 2.106} \quad >\text{step} = 0.301 \text{ Au}$$

$$\frac{A(254) = 0.0017}{A(255) = 0} \quad >\text{step} = 0.0017 \text{ Au}$$

(II) 10-bit binary digitization of I value ($I_{max} = 1023$)

$$\frac{A(1) = 3.010}{A(2) = 2.709} \quad >\text{step} = 0.301 \text{ Au}$$

$$\frac{A(1022) = 0.00042}{A(1023) = 0} \quad >\text{step} = 0.00042 \text{ Au}$$

(III) 12-bit binary digitization of I value ($I_{max} = 4095$)

$$\frac{A(1) = 3.612}{A(2) = 3.311} \quad >\text{step} = 0.301 \text{ Au}$$

$$\frac{A(4) = 3.010}{A(5) = 2.913} \quad >\text{step} = 0.097 \text{ Au}$$

$$\frac{A(4094) = 0.00011}{A(4095) = 0} \quad >\text{step} = 0.00011 \text{ Au}$$

$$A(I) = -\log\frac{I}{I_0}, \quad I_0 = I_{max}$$

sorbance was detected and integrated instantaneously. The raw data used were discarded and the processed results were kept in the core memory. The procedure of his two-dimensional gel analysis which was based on the "line-by-line" detection are summarized in Fig. 8.

5.2. Smoothing

Contamination of photometric signals with noises is inevitable. The noises should be reduced before proceeding to the next step, because the noises disturb the proper execution of programs for the background subtraction and the spot boundary determination. Taking an average of pixel values at every position by the repeated image acquisition is effective for reducing "white noise" of a CCD camera. However, noises from other causes cannot be removed by the repetitive acquisition method, and the method is also inapplicable to scanners. The noises in an acquired image should be reduced with an area processing filter which is a matrix of weighting parameters for replacing each pixel value with the average of neighboring pixel values and itself. Patton and Tempst [29] examined the
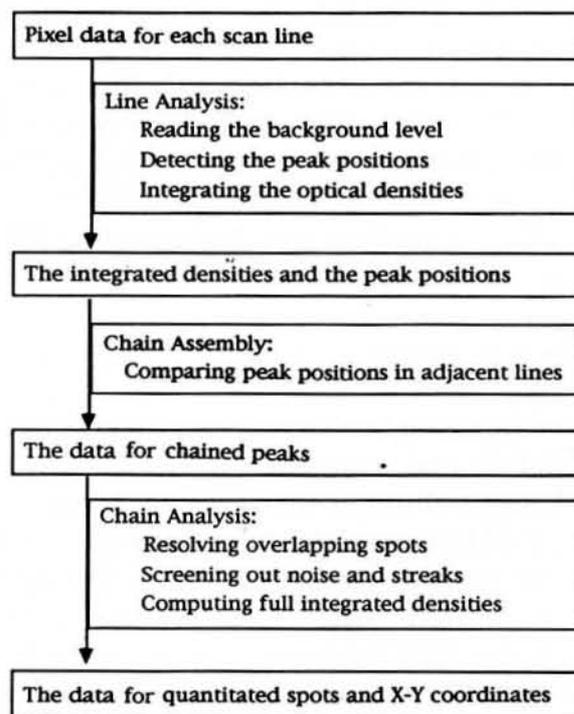

Fig. 8. A procedure of the "line-by-line" analysis on the desktop computer-based system reported by Garrels [6].

effect of various area processing filters upon the noise reduction and the spot enhancement. They reported that the 7×7 least squares filter significantly improved detection of low abundance polypeptides while having only minimal effects on the high abundance polypeptides. The 7×7 Gaussian and 3×3 median filters also improved detection of low abundance polypeptides but reduced the integrated area of the high-abundance polypeptides.

5.3. Background subtraction

Two-dimensional gel images appear often with an irregular "background", which should be subtracted before spot quantification. Staining of polyacrylamide gradient gels does not remove the gradient background completely. Minor spots adjacent to a major spot appear as shoulder peaks near the bottom of the high peak. Tailing of spots is another cause of the irregular background.

Many algorithms of background subtraction have been applied to two-dimensional densitometry. Garrels [6] defined the background in his "line-by-line" analysis as the most recent density level when no peak was detected. Another two-dimensional densitometer system preserves a frame of image data in the RAM or on the hard disk memory. The program for background subtraction processes the data in the memory after completion of the image acquisition. The simplest method of background subtraction is an "even subtraction" in which the minimum or percentage minimum level of pixel data is defined as the absolute background level, and the value is subtracted from all pixel data evenly. A histogram of the number of readings at each absorbance level helps to determine the absolute background level to be subtracted. A more careful method is "local background subtraction", in which the background level to be subtracted is determined in a local area given by an interactive mouse/keyboard operation. The most effective subtraction is made by the so-called "rolling disk" method. The algorithm of the method is based on tracing the centre of an imaginary disk of a given diameter which is

rolled on the underlying surface of the three-dimensional plot of absorbance data. The trace of the centre displays the three-dimensional profile of the background. Although the rolling disk method was recognized as the best for noise reduction, it was not applied to desktop computer-equipped systems for a long time because it required too much CPU time to be performed on earlier microprocessors of low MIPS. Recent advances in processor chips have realized the algorithm in PDQUEST software on a desktop workstation. However, it should be kept in mined that broad peaks of low heights are often erased by the background subtraction unless a large enough diameter is given to the rolling disk.

5.4. Spot detection and boundary determination

The manual tracing of a spot edge with a mouse is possible only for a small number of spots on a simple gel pattern. Automatic spot detection is convenient for processing hundreds of spots on a high-resolution two-dimensional gel pattern. Automatic spot detection is available in advanced commercial software.

Binarization of pixel data to make a temporary "bit map image" is the easiest way to demonstrate the area of spots to be quantified. The bit corresponding to a point is unity if the grey level is higher than a given threshold and zero otherwise. The sensitivity of spot detection is determined by the threshold which is given by a system operator. The image data are compressed by a factor of eight in the binarization. The binarization is suitable for small microcomputers in which only limited memory space is available for image processing. Contour display is another extended algorithm of binarization. A superimposition of multiple outlines of a spot obtained by repetitive binarization at varying threshold levels gives rise to the contour image.

Garrels [6] introduced a more sophisticated algorithm of spot detection into his original two-dimensional densitometer equipped with a desktop computer. His algorithm of "line-by-line" analysis is summarized in Fig. 8. Each peak of the one-dimensional density profile was detected

and integrated immediately after every scan. The second step of his protocol was "chain assembly", in which peaks on successive scan lines were matched and assembled into a "chain". Overlapping spots were resolved, and noise and streaks were removed by a Gaussian curve fitting in the next "chain analysis".

Taylor et al. [5] introduced the least-squares method to analyze two-dimensional Gaussian peaks on their autoradiograms produced by use of the ISO DALT system. Kronberg et al. [27] recommended another method of implicit modeling of spots for detecting convex areas and natural boundaries. The Laplacian spot finder utilized in Bio Image software is another non-parametric algorithm to determine spot edges. The background is determined based on regions where the second derivative of density is zero. Spot boundaries are determined by evaluating 36 radians generated from the spot center. The zero cross of the second derivative of the density levels is determined to generate the spot boundary. The method quantitates non-Gaussian spots more accurately than parametric methods.

In any case, the automatic determination of boundaries in a multi-spot area, which may come from incomplete resolution, is difficult. Interactive manual "splitting" of the fused spots is necessary on this occasion. Display of a perspective view is helpful for judging the correctness of the splitting.

5.5. Spot quantification

The main task of spot quantification software is to sum the pixel data within a closed boundary drawn automatically or manually. The reference to a standard density film of known absorbance levels placed in the same field of scanning is the simplest and best way to determine the absolute absorbance level. The number of pixels per unit area (mm^2) needs to be given if the absolute determination of the integrated absorbance (absorbance mm^2) is required. The pixel number per mm^2 is constant in an ordinary CCD scanner. It varies in a CCD camera system by zooming the field of view. For a CCD camera, the reading of a scale bar or a size marker is

convenient. However, the comparison of the absolute absorbance values of corresponding spots beyond gels is almost meaningless because the apparent integrated absorbance of spots is affected by the reproducibility of the sample preparation and gel electrophoresis. Relative quantification by referring to the integrated absorbances of the internal standard spots is of practical use in general studies.

5.6. Spot matching and gel comparison

Two-dimensional electrophoresis may display full efficiency especially when it is used for detecting minor differences in protein profiles among samples. Therefore, the computer-aided two-dimensional densitometer is expected to perform not only the simple quantification of spots but also spot matching among samples. A direct superimposition or the "flicker" display of rapidly alternating images on a screen, which was introduced by Lipkin and Lemkin [3], is effective for spot matching, provided that the two-dimensional gel patterns are absolutely reproducible. However, the actual gel patterns could not remove small distortions which might be the result of the irregular distribution of carrier ampholytes by local heat and/or a non-reproducible gel gradient. Taylor et al. [9] at ANL reported an advanced algorithm for the $X–Y$ coordinate stretching procedure to achieve the best fitting of spots on an "object" pattern to corresponding spots on a "master" pattern by a local neighbourhood analysis. The spot identification software successively assigned a "master spot number" (MSN) to the matched spot on the objective pattern. Also in 1981, Lemkin and Lipkin [8] reported an algorithm of "multiple two-dimensional gel analysis" at the 3rd International Conference on Electrophoresis at Charleston. The method of "point-by-point correspondence" in small regions of three near-neighbour landmarks accomplished complete spot pairing among many gel patterns. Advanced software including the activity of spot matching is commercially available on a workstation-based system.

Miller et al. [30] developed their "landmark

triangle" method on the contour maps. The triangle matching for coordinate transformation is adopted to Bio Image softwares. Garrels et al. established another procedure of the semi-automatic spot matching in their QUEST software that has been available in the commercial PDQUEST software. Gels are grouped into "matchsets" by the operator before matching begins. One gel of the matching begins. One gel of the matchset is chosen as a reference gel and is displayed in a window of the workstation screen. Each of the other gels is displayed opposite the reference gel, and apparently matched spots are marked by the operator as landmarks. For each close neighbour of a landmark in the reference gel, the x and y displacements relative to the landmark are noted and search is made for a spot in another selected gel with the same displacement relative to the corresponding landmark. When such a spot is found, the computer judges them to be matched. After all neighbours of landmarks have been tested for possible matches, the new matches are used in turn as starting points from which matches propagate. The procedure continues recurrently until no new matches are made. When some spots in every match are left unmatched, the spot adding algorithm finds spots that have matched patterns in some but not all gels. Each of these spots is then added at the position predicted by the neighbour match algorithm into each of the gels where it was not detected.

5.7. Database management

Anderson [4] first published his idea of "making molecular catalogues by two-dimensional electrophoresis" in 1979. Two years later, Lemkin and Lipkin [8] reported "multiple two-dimensional gel analysis" in which the database management algorithm had been established. An imaginary R-gel of representative spots (R-spots) was created by merging spots from original gels into the R-gel. The two-dimensional gel protein database so obtained serves as a master image for spot matching with unknown spots on a sample gel. The software of such database management was further developed and matured in

the QUEST system by Garrels et al. [10,11] and made available for commercial distribution in the PDQUEST software.

It is plausible that a single database is sufficient for each inbred strain of animal, because all proteins expressed in the animal are defined as the gene coding. However, a two-dimensional gel protein database should be constructed for each tissue or cell type in practice, because the gene expression and the post-translational modification are tissue specific. The first step in the development of a two-dimensional gel protein database is to run two-dimensional electrophoresis on a protein sample extracted from a specific cell type of a defined strain of animal under reproducible electrophoretic conditions. The applicability of the database depends on the method and the condition of electrophoresis used. A suitable method of protein detection should be chosen for each purpose of the database. Dye staining and silver staining are the most convenient methods for demonstrating major protein profiles in clinical investigations. Autoradiography of isotopically labelled proteins demonstrates proteins which are newly synthesized during the labelling period. Western blotting is suitable for detecting specific proteins.

The second step is to acquire the gel image and to perform the image processing, including smoothing, background subtraction, spot detection and spot quantification. The third step is to create a master image by merging spots into a cumulative image file. The fourth step is to register the information about a spot of protein identified by various biochemical methods.

The immunochemical demonstration of protein by Western blotting is a first choice of method to identify a protein spot if the specific antibody is available. Electrophoretic transfer to a PVDF membrane and subsequent peptide micro-sequencing are another choice for identifying the protein in a discrete spot on a two-dimensional gel pattern. Further information might be found in the DNA and protein sequence databases if the information has been already registered in them. The accumulation of large amounts of information in the two-dimensional gel protein database will help to identify

protein spots in sample gels in the subsequent studies.

6. Conclusions and perspectives

Projects for establishing computer-aided two-dimensional densitometry started shortly after O'Farrell's report [1]. Lemkin et al. [3], Bossinger et al. [2], Anderson et al. [24], Kronberg et al. [26] and Merril and Goldman [31] independently constructed their own systems for two-dimensional gel analysis using a DEC minicomputer. Garrels at the Salk Institute took other steps using his Hewlett-Packard desktop computer system to access two-dimensional densitometry. Rotating drum scanners, scanning microscope photometers and video cameras were adopted as the image readers in those systems.

At the beginning of computer-aided two-dimensional densitometry, the technique was a kind of privilege for the members and collaborators of the laboratories where the original system was developed. Some researchers, to whom no minicomputer was available, needed to make their own two-dimensional densitometer system using a microcomputer, as Garrels [6] did. However, the users of the microcomputer-aided systems had to endure the functional limitations with primitive quantification of spots because of the small core memory and the low MIPS. Further sophisticated functions of spot matching and database management were not realized on any microcomputer-based systems. This might be the major reason why Garrels and his co-workers changed to a minicomputer-aided system. The situation was absolutely changed by the appearance of SPARC/UNIX workstations at reasonable prices. The full functions of the software developed on minicomputer systems are completely available on the compact desktop workstation of the Solaris open windows system.

The resolution and the quantitative reliability of two-dimensional densitometry depend primarily on the hardware for image acquisition. A rotating drum scanner and a scanning microscope photometer were preferentially used in the minicomputer-aided systems because of their high quality at that time. However, a high-resolution CCD video camera or a high-quality desktop CCD scanner is now commonly used in workstation-based systems because they suit real-time performance.

Calibration of absorbance levels by reading a film of standard grey levels should be carried out on any type of optical reader to ensure linearity in quantification. Automatic spot detection and interior spot splitting are convenient, but minor spots are often neglected by the software if the setting of the sensitivity threshold is not optimum. Therefore, careful observation with the eyes is essential to check the appropriateness of automatic spot detection and splitting. The greatest advantages of the workstation-based
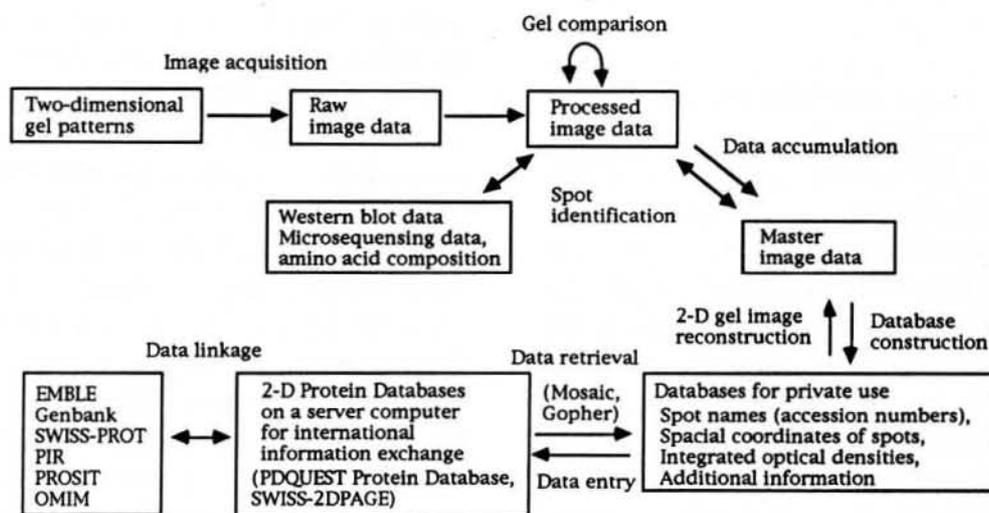

Fig. 9. The current status of two-dimensional densitometry and the resulted protein databases in the DNA/protein researches.

Table 3
Two-dimensional gel protein databases

Type	Source	Authors	Ref.
Plasma	Human	Anderson et al.	*Electrophoresis*, 12 (1991) 883–906
		Golaz et al.	*Electrophoresis*, 14 (1993) 1223–1231
Cerebrospinal Fluid	Human	Yun et al.	*Electrophoresis*, 13 (1992) 1002–1013
Liver	Rat	Anderson et al.	*Electrophoresis*, 12 (1992) 907–930
	Mouse	Giometti et al.	*Electrophoresis*, 13 (1992) 970–991
	Human	Hochstrasser et al.	*Electrophoresis*, 13 (1992) 992–1001
		Hughes et al.	*Electrophoresis*, 14 (1993) 1216–1222
Amniotic cells	Human	Celis et al.	*Electrophoresis*, 11 (1990) 989–1071
			Electrophoresis, 12 (1991) 765–801
Keratinocytes	Human	Celis et al.	*Electrophoresis*, 12 (1991) 802–872
			Electrophoresis, 13 (1992) 893–959
			Electrophoresis, 14 (1993) 1091–1198
Fibroblasts	Human	Celis et al.	*Electrophoresis*, 11 (1990) 1072–1113
REF52 cells	Rat	Garrels et al.	*Electrophoresis*, 11 (1990) 1114–1130
Escherichia coli		VanBogelen et al.	*Electrophoresis*, 11 (1990) 1131–1166
			Electrophoresis, 12 (1991) 955–994
			Electrophoresis, 13 (1992) 1014–1054

systems are the availability of "spot matching" and "database management" software. Although the reproducibility of electrophoretic separation has been markedly improved by using ready-made Immobiline dry gel strips and ready-made sodium dodecyl sulfate polyacrylamide gel, the simple superimposition of gel patterns is insufficient for complete spot matching. Automatic spot matching by referring to "three near-neighbour landmarks", which is available in most of the commercial software operating on a UNIX workstation, is convenient for differential protein analysis.

The hardware and the software systems for two-dimensional densitometry came to maturity by 1985. Subsequently, the number of reports on the applications of the two-dimensional densitometry increased. The protein databases (two-dimensional densitometry) are expected to be linked to the integrated DNA/protein sequence databases (Fig. 9). To assume the quality of the two-dimensional gel protein databases, identification of polypeptide spots in the gel patterns is critical. The immunochemical determination of

specific polypeptides on PVDF membranes [32,33] is effective to identify polypeptide spots on the gel. Many groups have been absorbed in establishing two-dimensional gel protein databases for various types of tissues and animals. The results of their efforts have been reported as revisions of the database annually in *Electrophoresis*, issue 11 or 12, since 1990. Some of the databases are shown in Table 3. A workshop on two-dimensional gel protein databases was held at the Committee on Data for Science and Technology (CODATA) Secretariat in Paris in 1992 [34]. The databases are expected to be in the public domain like the DNA databases of GenBank and EMBL. Recently, the SWISS-PROT database was extended to cover a two-dimensional gel protein database (SWISS-2DPAGE) by Appel et al. [35], and the databases are accessible through the ExPASy server on the Internet. The details of the Internet service were reported at the international working conference "2D Electrophoresis: from Protein Maps to Genomes", held in Sienna in October 1994. The continuous accumulation of

substantial databases on two-dimensional gel protein maps is in prospect.

References

[1] P.H. O'Farrell, J. Biol. Chem., 250 (1975) 4007–4021.
[2] J. Bossinger, M.J. Miller, P. Vo, E.P. Geiduschek and N.-H. Xuong, J. Biol. Chem., 254 (1979) 7986–7998.
[3] P. Lemkin, C. Merril, L. Lipkin, M. Van Keuren, W. Oertel, B. Shapiro, M.Wade, M. Schultz and E. Smith, Comput. Biomed. Res., 12 (1979) 517–544.
[4] N.G. Anderson, Nature, 278 (1979) 122–123.
[5] J. Taylor, N.L. Anderson, B.P. Coulter, A.E. Scandora and N.G. Anderson, in B.J. Radola (Editor), Electrophoresis '79, Walter de Gruyter, Berlin, 1980, pp. 329–339.
[6] J.I. Garrels, J. Biol. Chem., 254 (1979) 7961–7977.
[7] L.E. Lipkin and P.F. Lemkin, Clin. Chem., 26 (1980) 1403–1412.
[8] P.F. Lemkin and L.E. Lipkin, in R.C. Allen and P. Arnaud (Editors), Electrophoresis '81, Walter de Gruyter, Berlin, 1981, pp. 401–411.
[9] J. Taylor, N.L. Anderson and N.G. Anderson, in R.C. Allen and P. Arnaud (Editors), Electrophoresis '81, Walter de Gruyter, Berlin, 1981, pp. 383–400.
[10] J.I. Garrels, J.T. Farrar and I.C.B. Burwell, in J.E. Celis and R. Bravo (Editors), Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications, Academic Press, New York, 1984, pp. 37–91.
[11] J.I. Garrels, J. Biol. Chem., 264 (1989) 5269–5282.
[12] K.V. Robert, H. Hon, S.B. Alan and W. Sammons, in R.C. Allen and P. Arnaud (Editors), Electrophoresis '81, Walter de Gruyter, Berlin, 1981, pp. 371–381.
[13] J.E. Celis, G.P. Ratz and A. Celis, Leukemia, 1 (1987) 707–717.
[14] B.R. Franza, L.C. Sambucetti, D.R. Cohen and T. Curran, Oncogene, 1 (1987) 213–221.
[15] L.B. Epstein, J. Interferon Res., 7 (1987) 487–495.
[16] M.H. Beresini, M. Lempert and L.B. Epstein, J. Immunol., 140 (1988) 485–493.
[17] B.E. Sumpio, A.J. Banes, M. Buckley and G. Johnson, J. Vasc. Surg., 7 (1988) 130–138.
[18] D.W. Ballard, E. Bohlein, J.W. Lowenthal, Y. Wano, B.R. Franza and W.C. Greene, Science, 241 (1988) 1625–1654.
[19] R. Goodman and A.S. Henderson, Proc. Natl. Acad. Sci. U.S.A., 85 (1988) 3928–3932.
[20] J.I. Garrels and B.R. Franza, Jr., J. Biol. Chem., 264 (1989) 5299–5312.
[21] T. Toda, T. Fujita and M. Ohashi, Electrophoresis, 5 (1984) 42–47.
[22] H. Yamamoto, M. Nakatani, K. Shinya, B.-H. Kim and T. Kakuno, Anal. Biochem., 191 (1990) 58–64.
[23] G. Carman, P. Lemkin, L. Lipkin, B. Shapiro, M. Schltz and P. Kaiser, J. Histochem. Cytochem., 22 (1974) 732–740.
[24] N.L. Anderson, J. Taylor, A.E. Scandora, B.P. Coulter and N.G. Anderson, Clin. Chem., 27 (1981) 1807–1820.
[25] J. Taylor, N.L. Anderson, A.E. Scandora, Jr., K.E. Willard and N.G. Anderson, Clin. Chem., 28 (1982) 861–866.
[26] H. Kronberg, H.G. Zimmer and V. Neuhoff, Electrophoresis, 1 (1980) 27–32.
[27] H. Kromberg, Z.H.-G. Zimmer and V. Neuhoff, in R.C. Allen and P. Arnaud (Editors), Electrophoresis '81, Walter de Gruyter, Berlin, 1981, pp. 413–422.
[28] E.V. Maytin, E.N. Levin and R.R. Anderson, Anal. Biochem., 194 (1991) 284–294.
[29] W.F. Patton and P. Tempst, Electrophoresis, 14 (1993) 650–658.
[30] M.J. Miller, P.K. Vo, C. Nielsen, E.P. Geiduschek and Ng.-H. Xuong, Clin. Chem., 28 (1982) 867–875.
[31] C.R. Merril and D. Goldman, Clin. Chem., 28 (1982) 1015–1020.
[32] P.J. Matsudaira, Biol. Chem., 262 (1987) 10035–10038.
[33] G.I. Tous, J.L. Fausnaugh, O. Akinyosoye, H. Lackland, P. Winter-Cash, F.J. Victoria and S. Stein, Anal. Biochem., 179 (1989) 50–55.
[34] R.J. Simpson, A. Tsugita, J.E. Celis, J.I. Garrels and H.W. Mewes, Electrophoresis, 13 (1992) 1055–1061.
[35] R.D. Appel, J.C. Sanchez, A. Bairoch, O. Golaz, M. Miu, J.R. Vargas and D.F. Hochstrasser, Electrophoresis, 14 (1993) 1232–1238.