PERGAMON

Progress in Biophysics & Molecular Biology 80 (2002) 3–22

Progress in Biophysics & Molecular Biology

www.elsevier.com/locate/pbiomolbio

Review

Human proteomic databases: a powerful resource for functional genomics in health and disease

Pavel S. Gromov[a,*], Morten Østergaard[b], Irina Gromova[a], Julio E. Celis[a]

[a] *Institute of Cancer Biology and Danish Centre for Human Genome Research, The Danish Cancer Society, Strandboulevarden 49, DK-2100, Copenhagen, Denmark*

[b] *Department of Medical Biochemistry, University of Aarhus, Aarhus, Denmark*

Abstract

Decoding of the genome information in terms of regulation and function will be the next great challenge in the life sciences in this millennium and indeed, today we are experiencing a rapid explosion of technology for the high throughput expression analysis of genes and their products (functional genomics). In particular, the field of proteomics is booming as proteins are often the functional molecules and represent important targets for the pharmaceutical industry. The proteomic technology is complex, and comprises a plethora of state-of-the-art techniques to resolve, identify and detect their interacting partners, as well as to store and communicate protein information in comprehensive two-dimensional polyacrylamide gel electrophoresis (2D PAGE) databases. Besides annotating the genome, these databases will offer a global approach to the study of gene expression both in health and disease. Here, we review the current status of human 2D PAGE databases that we are systematically constructing for the study of bladder cancer and skin ageing.

Keywords: Human genome; Proteome; Databases; Diseases

Contents

1. Introduction 4

2. Making a 2D PAGE protein database 5
   2.1. 2D PAGE—the core separation technology 5
   2.2. Image analysis 6

3. Human 2D PAGE databases 7

Abbreviations: 2D, two-dimensional; PAGE, polyacrylamide gel electrophoresis

*Corresponding author. Tel.: +45-35-25-73-54; fax: +45-35-25-77-21.

E-mail address: psg@cancer.dk (P.S. Gromov).

0079-6107/02/$ - see front matter
PII: S0079-6107(02)00005-6

3.1. The human keratinocyte database 7
3.1.1. Protein identification 9
3.1.2. Entries, annotations and links 10
3.2. 2D protein databases for the study of bladder cancer 15
3.3. 2D gel and blot galleries 17

4. Concluding remarks 17

Acknowledgements 18

References 18

1. Introduction

Undoubtedly, the Human Genome Project has paved the way to the revolution in biology and medicine that we are experiencing today (Lander et al., 2001; Venter et al., 2001). Decoding of the genome information in terms of regulation and function will be the next great challenge in the life sciences in this millennium (Abbott, 1999). Indeed, today we are experiencing a rapid explosion of technology for the high throughput expression analysis of genes and their products (functional genomics). In particular, the field of proteomics is booming as proteins are often the functional molecules and represent important targets for the pharmaceutical industry. In a way, the discovery process is being accelerated in part by the prospect of translating basic findings into clinical applications (translational research).

Historically, biochemical studies of proteins have been focussed mainly on the analysis of single molecules. Today, however, thanks to proteomic technologies it is possible to analyse complex biological samples and to even contemplate the possibility of embarking on the study of whole proteomes. The protein universe of a cell is extremely complex and highly diversed and any such endeavour must take in consideration the large dynamic range of protein expression, post-translation modifications, protein–protein interactions as well as functional aspects. Obviously, this is a daunting task that will need to be tackled on a collaborative basis at the international level.

As the Human Proteome Project is set in place (Peters et al., 2001; Lander et al., 2001; Venter et al., 2001), there will be an urgent need to store information in proteomic databases that for the time being use two-dimensional polyacrylamide gel electrophoresis (2D PAGE) as the core technology. These databases will gather qualitative and quantitative protein behavioural data generated in various cell types, tissues and fluids and are expected to annotate the human genome. A list of various 2D PAGE databases available in the Internet can be found in the World-2D PAGE Index: http://www.expasy.ch/ch2d/2d-index.html. Recently, several companies: Celera (Service, 2000; www.celera.com), Oxford GlycoSciences (www.ogs.com), GeneProt (www.gene-prot.com), have independently launched a global study of the human proteome with the aim of identifying and cataloguing proteins and their isoforms expressed in human tissues and fluids. This new generation of databases will help researchers to further understand fundamental cellular processes in cell as well as to select key targets for drug development programmes. The purpose of this article is to review the current state of our human 2D PAGE databases (http//

biosun.biobase.dk) which are being systematically built for the global analysis of the human proteome both in health and disease.

2. Making a 2D PAGE protein database

2.1. 2D PAGE—the core separation technology

The impact of the proteomic technologies for large-scale monitoring of protein expression depends very much on the number of proteins to be resolved. Recent data released by the Human Genome Consortium together with Celera Cenomics estimated that the human genome may contain about 30 000 protein coding genes (Lander et al., 2001; Venter et al., 2001), a number that it is far lower than formerly anticipated. However, due to alternative RNA splicing, overlapping of transcription units and *trans*-splicing RNA (Labrador et al., 2001), the actual number of messenger species in eukaryotic genomes far exceed the number of coding units. Current estimates based on the various assemblies of expressed sequence tags (ESTs) suggest that the average number of mRNA species might be two to three times higher yielding a putative number of 70 000–90 000 distinct primary translation products (Wheeler et al., 2001; Liang et al., 2000). In addition, many proteins have isoforms that arise from post-translational processing and modifications.

Data from a few laboratories, including our own, have shown that only a fraction of the whole genome is expressed at the protein level in a given cell type (Duncan and McConkey, 1982; Celis et al., 1991a, b; Celis et al., 1998). With a few exceptions, single human cell types may express in the range of 5000–6000 different primary translation products plus their modified variants, which can be extensive in some particular cases (Gooley and Packer, 1997). As much as 80–90% of these polypeptides may represent housekeeping proteins (components of metabolic pathways, cytoarchitecture elements, etc.) that are expressed, albeit in variable amounts, by all cell types (Celis et al., 1991a, b, 1998).

The core technology for making proteomic databases is high-resolution 2D PAGE. This technique is considered today as the method with the highest resolution for the separation of complex protein mixtures such as those present in cells, tissues or fluids. (Görg et al., 2000; Hanash, 2000). 2D PAGE coupled with mass spectroscopy offer today the most widely accepted format to study protein expression and identity (Corthals et al., 2000; Fey and Larsen, 2001; Naaby-Hansen et al., 2001). Currently, several alternative approaches are being developed to achieve high-throughput, that include capillary isoelectric focusing (Shen et al., 2000), multiple-capillary liquid chromatography (Shen et al., 2001), and large-scale peptide or protein chips and arrays (Figeys and Pinto, 2001; Emili and Cagney, 2000; Borrrebaeck, 2000; Holt et al., 2000). None, however, has the resolution and sensitivity of 2D PAGE, which displays thousands of proteins at a time and allows the quantitative analysis of any single protein in relation to the rest.

Proteomic research in our laboratory has been mainly based on the use of carrier ampholytes as originally described by O'Farrell (1975, 1977) and Klose (1975). Even though gels run with carrier ampholytes are difficult to reproduce, we have standardised the technology to a level that is possible to obtain reproducible separations of nearly 4000–5000 [^{35}S]-methionine labelled

polypeptides from whole human cell extracts using broad pH gradients (Celis et al., 1995, 1998, 2000; see also procedures and videos in http://biobase.dk/cgi-bin/celis). Clearly, this number is still short of the total number of proteins that may be present in an eukaryotic cell. Missing polypeptides are either not resolved by the pH gradient (too acidic or too basic), do not enter the gel due to solubilisation problems and/or size, or are present in too low abundance to be detected with the current detection procedures.

Considerable improvements in 2D gel technology have been made in the last years in an effort to overcome some of the limitations mentioned above (Görg et al., 2000; Hanash, 2000 and references therein). In particular, immobilised pH gradients (IPGs: wide and narrow pH gradients) provide more reproducible focussing patterns, avoid some of the problems associated with carrier ampholytes (e.g. cationic drift), allow higher protein loading, and offer enhanced resolution in the first dimension. Görg and colleagues have improved considerably the separation of very basic proteins (Görg, 1999) using IPGs of very basic pH range and isoelectric pre-fractionation has been used to enrich for proteins with basic pIs (Zuo and Speicher, 2000). Appropriate extraction procedures in combination with pre-gel fractionation have been utilised to enhance the separation and visualisation of integral membrane proteins (Santoni et al., 2000). Additional developments include also the use of very narrow pH gradients (approximately 0.001 pH/cm), fluorescent dyes (Steinberg et al., 1996,) as well as the advent of various automated 2D gel matching softwares (Smilansky, 2001 and references therein).

2.2. Image analysis

A systematic analysis of the human proteome with the aim of establishing 2D protein databases requires in addition to a reproducible 2D gel system, computer-assisted technology to scan the gels, make synthetic images, assign numbers to individual spots and match gel spots (Pleißner et al., 1999; Panek and Vohradsky, 1999; Smilansky, 2001). In addition, one needs functions to enter and retrieve qualitative and quantitative information. The first step in making a database is to prepare a synthetic image (digital form of the gel image) of the gel (autoradiogram, Coomassie blue or silver-stained gel, fluorogramm) chosen as a standard or master reference. This can be done with laser scanners, charge couple devices (CCD), array scanners, television cameras, rotating drum scanners and multiwhire chambers. Various softwares for the analysis of 2D protein gel images and processing of data have been developed, they include: PDQUEST (BioRad), MELANIE 3 (Swiss Institute of Bioinformatics), ProXPRESS (Perkin-Elmer), PHORETIX (Nonlinear Dynamics Ltd), Z3 (Compugen) and several others, that provide different levels of automation and speed.

In our workstation, autoradiograms are scanned with a laser scanner and the data are analysed using the PDQUEST II software running on a SPARK station computer from SUN Microsystems, Inc. The scanner measures intensity in the range of 0–2.0 absorbance. Today more accurate quantitative data can be obtained by using a phosphorimager. Steps in image analysis include: image acquisition, initial smoothing, background subtraction, final smoothing, spot detection and fitting of ideal Gaussian distributions to spot centres. Spot intensity is calculated as the integration of a fitted Gaussian. If calibration strips containing individual segments of known amount of radioactivity are used, it is possible to merge multiple exposures of the sample image into a single data image of greater dynamic range. Once the synthetic image is

created it can be stored on disk and displayed directly on the monitor for editing. To our knowledge, none of the softwares available on the market provide full automatic analysis of 2D gel images. The editing process is time consuming and can only be performed by operators that have extensive experience in the analysis of 2D gel images. Functions that can be used to edit the images include: cancel (for example to erase scratches that may have been interpreted as spots by the computer; cancel streaks or low-dpm spots), combine (sometimes a spot may be resolved into several closely packed spots), restore, uncombine and add spot to the gel.

Each polypeptide is assigned a number by the computer, a fact that facilitates the entry and retrieval of qualitative and quantitative information for any given spot in the gel (Celis et al., 1995, 1998). The standard image can be matched automatically by the computer to other standard or reference gels provided a few landmark spots are given automatically or manually as reference to initiate the process. Proteins are matched according to their gel position and additional means to verify their relatedness are needed before one can take full advantage of the data. Once a standard map of a given protein sample is made, one can enter qualitative and quantitative information to establish a reference or master database (see, for example, http:/biobase.dk/cgi-bin/celis). Categories or entries are created so as to gather information on physical, chemical, biochemical, physiological, genetic, architectural as well as biological properties of proteins. In general, entries reflect the type of biological problem that is being studied using the database approach.

3. Human 2D PAGE databases (http://biobase.dk/cgi-bin/celis)

As a result of a long-term and systematic effort to analyse the human proteome in health and disease, we have established several comprehensive 2D proteomic databases for the study of global cell regulation. The databases, which focus on skin biology and bladder cancer, gather protein data on non-cultured cell types such as keratinocytes (Celis et al., 1992, 1995) and transitional and squamous cell carcinomas of the bladder (Celis et al., 1999b), as well as urine (Rasmussen et al., 1996).

3.1. The human keratinocyte database

The human keratinocyte 2D PAGE database (Celis et al., 1995, http://biobase.dk/cgi-bin/celis) is the largest of its kind available today. It currently list 3629 cellular (2315 isoelectric focusing, IEF; 956 non-equilibrium pH gradient electrophoresis, NEPHGE), and 358 externalised polypeptides (IEF). Fig. 1 shows the synthetic master 2D PAGE images (IEF and NEPHGE)) of non-cultured human keratinocyte proteins as depicted in the World Wide Web (http://biobase.dk/cgi-bin/celis). Proteins flagged with a redcross correspond to polypeptides that have been identified using the various technologies described below under protein identification. The database offers extensive search and listing facilities for data retrieval as well as various options for image manipulation.

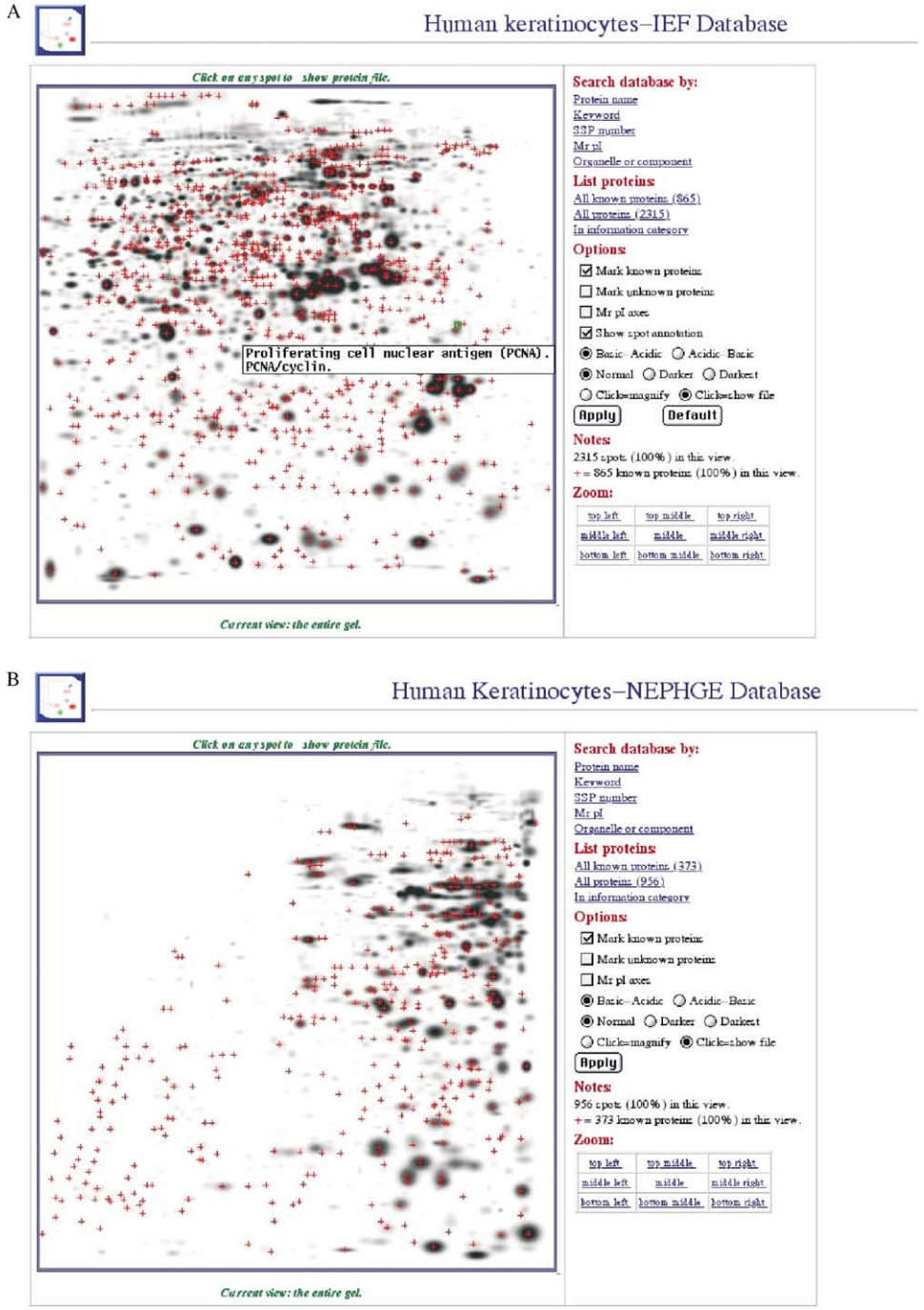

Fig. 1. Master synthetic images of human keratinocyte proteins separated by IEF (A) and NEPHGE (B) 2D PAGE as depicted in the World Wide Web. Proteins flagged with a red cross correspond to known proteins. PCNA/cyclin is indicated with a green mark as an example. By clicking on any spot, it is possible to open a file that contains protein information as well as links to other related WWW sites.

3.1.1. Protein identification

Proteins resolved on the 2D gel/blots have been identified using a combination of techniques that include mass spectrometry (Wilm, 2000; Roepstorff, 2000; Mann et al., 2001), 2D PAGE Western blotting (Celis and Gromov, 2000) and Edman degradation of internal peptides (Bauw et al., 1989; Vandekerckhove et al., 1990). Today, however, only some of these procedures are being used and are briefly discussed below.

3.1.1.1. Mass spectrometry. Mass spectrometry is today the technique of choice for protein identification as it requires picomoles levels of the proteins. There are several formats of this technology, namely: MALDI-TOF, Q-TOF, electrospray ionisation and several others that have been developed during the last years having in mind sensitivity and high-throughput analysis (Wilm, 2000; Roepstorff, 2000; Mann et al., 2001 and references therein). Peptides can also be sequenced using mass spectrometry (MS–MS), a fact that greatly facilitates the identification of unknown proteins as well as cloning. Recently, Gygi and colleagues described a novel approach based on the isotope-coded affinity tag (ICAT) labelling of proteins that highly enhanced the potential of mass spectrometry for accurate quantification and identification of individual proteins in complex mixtures (Gygi et al., 1999).

In our laboratory, we use MALDI-TOF mass spectrometry for peptide mass fingerprinting as well as for partial peptide sequencing (MS–MS). Routinely, we are able to identify proteins in total cell lysates in the range of 10–20 fmol per spot (up to 1 ng).

3.1.1.2. 2D PAGE Western blotting of low-abundance proteins. The detection and identification of low-abundance proteins in proteome studies is currently considered as a major bottleneck in large-scale proteomics. Protein levels in cells are often predicted by examining the codon bias value of the particular gene as this parameter reflects the probability of selectively utilising certain codons (Futcher et al., 1999). Thus, it has been estimated that proteins exhibiting codon bias values of <0.2 may be expressed in approximately 10 000 protein copies per cell and are considered as low-abundance proteins that cannot be detected without pre-run enrichment. Recently, Fey and Larsen (2001) resolved about 10 000 [^{35}S] methionine-labelled proteins from yeast including low-abundance proteins with the codon bias of <0.2. Even though the authors used broad pH range gels and a protein load of $<100\,\mu g$ they were able to apply MS–MS to identify low-level proteins (Fey and Larsen, 2001).

In our laboratory, we have used immunoblotting coupled with enhanced chemoluminiscence (ECL, Amersham) detection to reveal components that are present in as few as 500 molecules per cell. The latter is exemplified in Fig. 2 which shows an ECL developed 2D gel Western blot of crude keratinocyte extract reacted with antibodies raised against $p21^{H-ras}$, a protein that is known to be expressed in about 20 000 molecules per cell (Scheele et al., 1995). The corresponding area of the ^{35}S-methionine autoradiogram is shown for comparison.

The amplitude of the intensities of the various $p21^{H-ras}$ spots shown in Fig. 2 indicated that the low-abundance product marked with an arrow may be present in no more than 500 molecules per cell. This technology reveals in addition the extent of the modification(s) and can effectively complement mass spectrometry and metabolic labelling with specific radioactive precursors to determine the nature of the modifications. Given the rapid progress in phage display strategies for

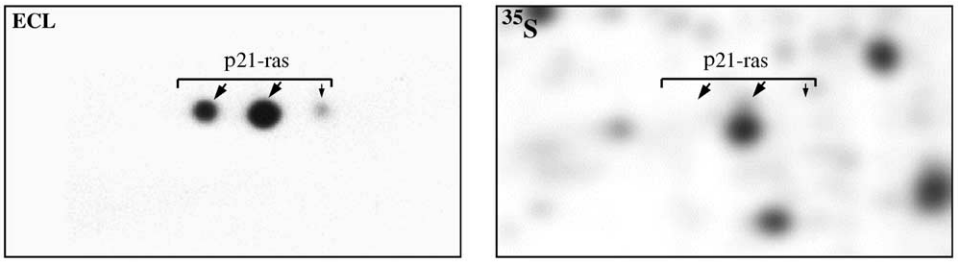

Fig. 2. 2D gel ECL immunodetection of $p21^{H-ras}$ and their modified variants expressed in human keratinocytes. Left: 2D gel Western blot of human keratinocyte proteins reacted with a monoclonal antibody against $p21^{H-ras}$ and developed using the ECL procedure. Right: A corresponding autoradiograph of [^{35}S]methionine-labelled proteins obtained from the same blot (see left). The positions of the $p21^{H-ras}$ variants are indicated with arrows.

preparing thousands of highly specific antibodies, we believe that ECL immunodetection will play an important role in the identification of low-abundance proteins.

3.1.1.3. Protein overlay. Protein-targeting interactions play a central role in most biological processes. Their detection and analysis in vitro can provide important information on specificity, affinity, and structure–function relationships. Protein blot overlay assays (also known as ''Far-Western'', ''Western–Western'', ''ligand'', or ''affinity'' blotting) are very powerful techniques for detecting and analysing proteins or protein motifs involved in cellular-targeting processes. These methods are based on the principle that proteins (or protein fragments) resolved by electrophoresis and transferred to an immobilising matrix (nitrocellulose or nylon membrane) can be probed with putative binding partners followed by subsequent detection of the complexes formed. Many probes can be used as putative binding partners, including metal ions, nucleotides, nucleic acids, hormones, proteins, antibodies, viruses and cells.

Overlay procedures have been successfully applied to a wide range of protein–protein interactions, such as kinase anchoring (Hausken et al., 1998) and ligand protein interactions using Ca^{+2} (Hoffmann et al., 1998), GTP (Gromov and Celis, 1998), and nucleic acids (Dejgaard and Celis, 1998). Fig. 3 illustrates the 2D gel mapping of Ca^{+2} binding proteins and small GTP-binding on nitrocellulose blots. Using this approach about 40 small GTP-binding proteins have been detected and recorded in the human keratinocyte database.

Drawbacks of the procedure include: (i) weak binding of some proteins to the immobilising matrix; (ii) loss of reactivity due to denaturing effects and (iii) some protein complexes are usually formed in specific microenvironments that are difficult to reproduce in in vitro studies. Despite of these limitations, protein overlay coupled with the 2D PAGE can be effectively used for confirming suspected interactions as well as for identifying new binding partners.

3.1.2. Entries, annotations and links

As mentioned above, categories or entries are created so as to gather information on physical, chemical, biochemical, physiological, genetic, architectural as well as biological properties of proteins. In general, they reflect the type of biological problem that is being studied using the

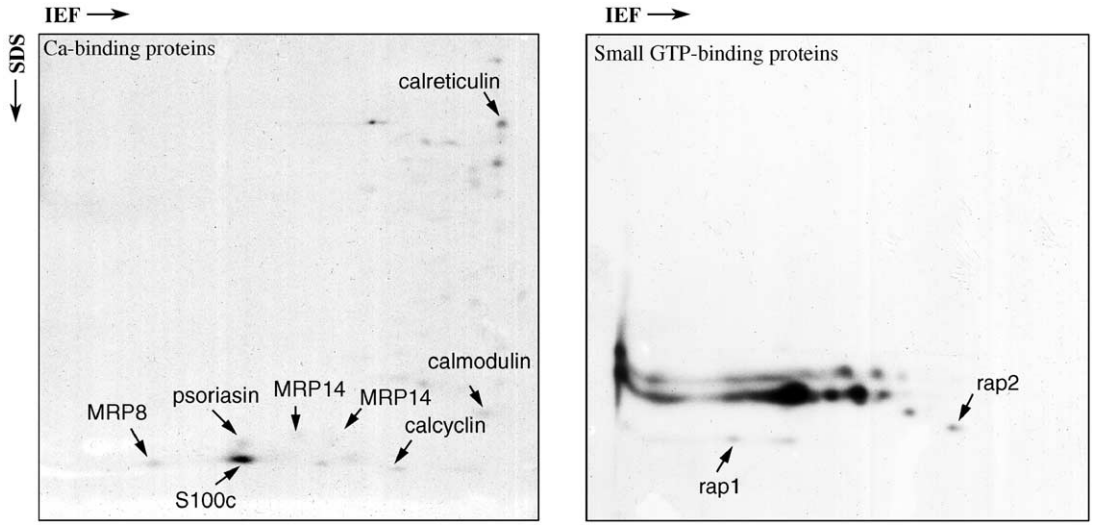

Fig. 3. Protein blot overlay of Ca-binding proteins (left image) and small GTP-binding proteins (right image) expressed in human keratinocytes. Proteins were identified by mass spectrometry and by comparison with the master 2D-gel protein image.

database approach. Some of these entries are also available in other databases accessible at the Web site.

At present, about 100 information categories are available in the World Wide Web version of the IEF keratinocyte database, including protein name, cellular localization, proteins differentially regulated in differentiated keratinocytes, proteins expressed in other cell types, heat shock proteins, proteins affected by interleukins (1-α, 1-β, 2, 3, 6, 7 and 8), proteins affected by interferons (α, β and γ), proteins differentially regulated in keratinocytes treated with various chemicals (calcium, okadaic acid, phorbol myristate acetate, retinoic acid, staurosporin), transformation sensitive proteins, levels in foetal human tissues, etc.

Categories for post-translational modifications like "phosphoryl" and "glycosyl" include data for those proteins that are either phosphorylated or glycosylated. Co- and post-translationally modified variants can be distinguished from the primary translated products by in vivo labelling of the proteins with the corresponding isotope-labelled metabolite followed by 2D PAGE. Once a radiolabelled ligand is covalently attached to the protein, it can be detected on a gel or in a blot using autoradiography or phosphorimage. Metabolic radiolabelling is carried out by incubating cells with the radiolabelled compound under appropriate conditions. Fig. 4 illustrates two-dimensional profiles of phosphorylated (phosphoproteome, Fig. 4A), glycosylated (glycoproteome, Fig. 4B), palmitoylated and myristoylated (lipoproteome, Fig. 4C and D), farnesylated and geranylgeranylated proteins (lipoproteome, Fig. 4E and F) from human AMA cells, a cell type for which the comprehensive proteome database was established in our laboratory in 1991 (Celis et al., 1991b).

By superimposing the 2D gel pattern of chemically modified protein isoforms with the 2D gel master [^{35}S]-methionine image, it is possible to pinpoint the position of protein variants on the master image. Post-translational modifications in most cases lead to changes in molecular weight

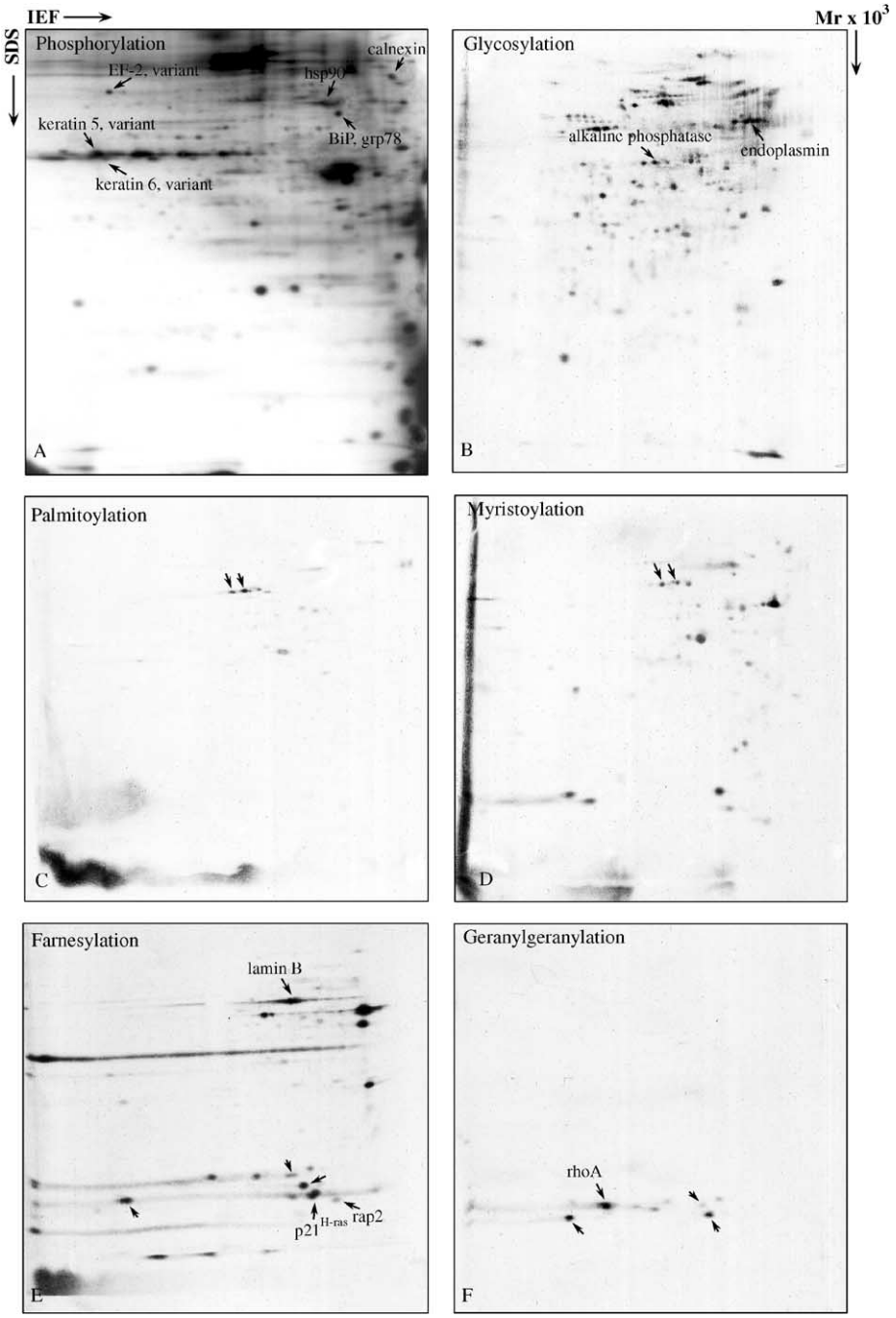

Fig. 4. Two-dimensional gel (IEF) fluorographs of transformed epithelial human amnion cells (AMA) labelled with (A) [^{32}P]-orthophosphate (phosphoproteome), (B) [^{3}H]-mannose (glycoproteome), (C) [^{3}H]-palmitate (lipoproteome), (D) [^{3}H]-myristate (lipoproteome), (E) [^{3}H]-farnesyl (lipoproteome) and (F) [^{3}H]-geranylgeranyl (lipoproteome). Several of the identified proteins are indicated with arrows. Arrowheads indicate proteins that are modified by two types of ligands: palmitate and myristate (C and D), and farnesyl and geranylgeranyl (E and F).

and/or pI and the use of very narrow IPGs (up to 0.001 pH units) may allow resolving protein variants that differ slightly in their electric charge. The exact type of modification can be then characterised by mass spectrometry which is the most powerful tool for determining the nature of the modified isoforms (Larsen and Roepstorff, 2000 and references therein).

In general, nearly all of the information contained in the database has been gathered from experiments performed in our laboratory. Files for known proteins contain links to a subset of Medline (http://www.ncbi.nlm.nih.gov/PubMed/), Swiss-Prot (http://expasy.hcuge.ch/sprot/sprot-top.html) and PDB (http://www.embl-heidelberg.de/pdb/). Other links include OMIM (http://www. ncbi.nlm.nih.gov/Omim/), GeneCards (http://bioinformatics.weizmann.ac.il/cards), UniGene (http://www.ncbi.nlm.nih.gov/UniGene/index.html) and other Web sites such as CySPID (cytoskeletal protein database; http://paella.med.yale.edu/~panzer/cytoskdb/index.html), metabolic pathways (compiled by KEGG; http://www.genome.ad.jp/kegg/), the cytokine explorer (http://kbot.mig.missouri. edu:443/cytokines/explorer.html), histology images (http://biosun.biobase.dk/~pdi/jecelis/micrographs.html), etc. In the future, as new databases and related Web sites become available, it will be possible to navigate throughout various databanks containing complementary information on nucleic acid and protein sequences, genome mapping, diseases, protein structure, post-translational modifications, antibodies and cellular localization of the antigen, signalling pathways, histology, etc.

Information obtained on any given polypeptide, known or unknown, can be easily retrieved from the database by clicking on the corresponding spot, in this case the proliferating cell nuclear antigen (PCNA) (Fig. 5).

Only a small fraction of the file is shown. PCNA, also called PCNA/cyclin, was the first protein discovered by the 2D gel technology for which it was possible to guess a function by performing simple cell biology experiments (Celis et al., 1984, 1987; Celis and Bravo, 1984; Zuber et al. 1989). This protein showed S-phase specific expression (Fig 6), and patterns of immunostaining using autoantibodies colocalised with autoradiographic patterns of tritiated thymidine incorporation (Fig. 7).

These data suggested that PCNA/cyclin most likely played a role in DNA replication, a suggestion that was later proved correct (Miura, 1999; Warbrick, 2000 and references therein).

Functions to query the databases include search by name, protein number or keywords, molecular weight and pI, as well as organelle or cellular component (Fig. 1, right-side panel). Also, all proteins recorded in a given entry can be easily queried. As an example, Fig. 8 shows a list of all proteins recorded in the keratinocyte database that can be retrieved when choosing the entry "autoantigen". By clicking on the protein name it is possible to display its position on the 2D image.

Using other functions listed in Fig. 1, it is possible to flip (important as gels are presented and analysed in different orientations by various laboratories) and enlarge the master synthetic image, as well as to retrieve a list of known proteins and categories available.

To facilitate the comparison between keratinocyte proteins run with carrier ampholytes and IPG's, we provide in the gel gallery section of the Web site IPG gel images that can be compared directly to the master keratinocyte image. One hundred known polypeptides have been selected for reference (Fig. 9). The images can be retrieved in both orientations: basic to acidic and acidic to basic.

Spot information: SSP 9226

Proliferating cell nuclear antigen (PCNA). PCNA/cyclin.; Mr = 37.4 kDa ; pI = 4.4

For more information, see also MEDLINE or the gopher servers GenBank, Swiss-Prot, PIR, and PDB
Note: Netscape may not return from the Gopher Menu page using 'Go Back'.
Use the Window menu History command to return from the Gopher search.

Category	Description	Annotation Entry	Link
genecards	Link to GeneCards.		GeneCards
humanchrom	Human Chromosome.		Human/Mouse Homology
omim	Link to OMIM.		OMIM
swissprot	Link to Swiss-Prot.		SWISS-PROT
protname	Protein Name.	Proliferating cell nuclear antigen (PCNA). PCNA/cydin.	
autoantigen		Autoantigen.	
bfgf	Keratinocyte Proteins Affected by bFGF.	Not affected. J.E.Celis, unpublished observation.	
cellcycle		Increase synthesis during S-phase.	
cellsignetdatab		Listed in CSNDB.	Cell Signalling Network Database (CSNDB)
cellulloc	Cellular Localization.	Nucleus.	
cellulrole	Cellular Role According to EGAD.	DNA replication.	
credinv	Credit to Investigator that Aided the Identification.	R.Bravo and J.E.Celis, Aarhus.	
difregdiffker	Proteins Differentially Regulated in In Vitro Differentiated Keratinocytes.	Downregulated. Ratio differentiated/proliferating = 0.7. E.Olsen et al., Electrophoresis 16, 2241, 1995.	
distrexker	Distribution in Triton Extracted Keratinocytes.	Triton cytoskeleton and supernatant. J.E.Celis, unpublished observation.	
donotcomigrate	Known Human Keratinocyte Proteins that do not Comigrate with Known Mouse Prot.	Proliferating cell nuclear antigen.	
expblasqcellcar	Expressed in bladder Squamous Cell Carcinomas.	Expressed.	
expeosinoph	Expressed in Eosinophils.	Expressed.	
expfehuliver	Expressed in Fetal Human Liver.	Expressed.	
expfehupancreas	Expressed in Fetal Human Pancreas.	Expressed.	
expfehuspleen	Expressed in Fetal Human Spleen.	Expressed.	
expfehutongu	Expressed in Fetal Human Tongue.	Expressed.	
expfehuureter	Expressed in Fetal Human Ureter.	Expressed.	
expfibro	Expressed in Fibroblasts.	Expressed. J.E.Celis, unpublished observation.	
exphelacells	Expressed in Hela Cells.	Expressed (high). J.E.Celis, unpublihed observation.	
explymph	Expressed in Lymphocytes.	Expressed.	
expneutroph	Expressed in Neutrophils.	Expressed.	
expurothcell	Expressed in Urothelial Cells.	Expressed. J.E.Celis et al., unpublished observation.	
genemap	Gene Map.	20p12.	
genename	Gene Name.	PCNA.	
il1alpha	Keratinocyte Proteins Affected by Interleukin 1 Alpha.	Downregulated. J.E.Celis, unpublished observation.	
il1beta	Keratinocyte Proteins Affected by Interleukin 1 Beta.	Downregulated. J.E.Celis, unpublished observation.	
interfalpha	Proteins Affected by Interferon Alpha.	Not affected. J.E.Celis, unpublished observation.	Signaling pathway
interfgam	Keratinocyte Proteins Affected by Interferon Gamma.	Downregulated. J.E.Celis, unpublished observation.	Signaling pathway
levelsk14	Protein Levels in SV40 Transformed K14 Keratinocytes.	Ratio transformed/normal proliferating = 7.76. J.E.Celis and E.Olsen, Electrophoresis 15, 309, 1995.	
marker		Proliferating cells. R.Bravo et al., Exp. Cell Res.136, 311, 1981. S-phase cells,J.E.Celis and A.Celis, Proc. Nat. Acad. Sci. 82, 3262, 1985.	
methodident	Method of Identification.	2D gel electrophoresis and 2D immunoblotting	
okadaicacid	Keratinocyte Proteins Affected by Okadaic Acid.	Not affected. O.K.Vintermyr and J.E.Celis, unpublished observation.	
percentotprot	Percentage of Total 14C-Labeled Protein Recovered From IEF and NEPHGE Gels.	0.028%. J.E.Celis and M.Ostergaard, unpublished observation.	
pma	Keratinocyte Proteins Affected by PMA.	Downregulated. J.E.Celis, unpublished observation.	
prosite	Link to Prosite.	PDOC00265	PROSITE
protaffpsor	Proteins Affected in Psoriatic Skin.	Upregulated. J.E.Celis et al., unpublished observation.	Psoriasis Foundation
search	Protein Name.	Proliferating+cell+nuclear+antigen (PCNA). PCNA/cydin.	
unigene		proliferating cell nuclear antigen	

This page produced using custom web software based on PDQUEST.

Gels analyzed using PDQUEST from Bio-Rad (formerly PDI).

Fig. 5. File for the proliferating cell nuclear antigen (PCNA). Only some of the entries available in the Internet version of the database are presented.

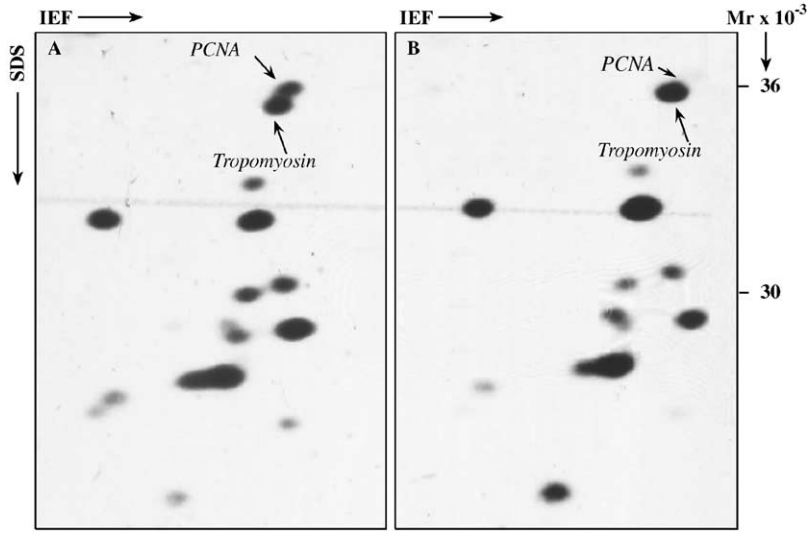

Fig. 6. Synthesis of [^{35}S]methionine-labelled PCNA in synchronised HeLa cells. Synchronised cells were obtained by mechanical detachment of mitotic cells. (A) S-phase (B) G1-phase. The positions of PCNA and a tropomyosin are indicated with arrows. From Celis et al. (1987).

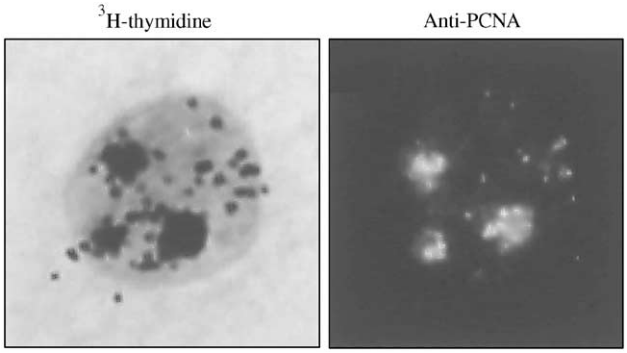

Fig. 7. Immunofluorescence (PCNA antibodies) and autoradiographic analysis ([^{3}H]-thymidin incorporation) of asynchronous transformed human amnion (AMA) cells. AMA cells were levelled with [^{3}H]-thymidine (30 min, 2 μCi/ml) fixed with methanol and probed with anti-PCNA antibodies. The immunofluorescence pictures (right) were taken prior to autoradiography (left). From Madsen and Celis (1985).

3.2. 2D protein databases for the study of bladder cancer

2D PAGE databases of TCC and SCC proteins have been established with the long-term goal of unravelling the molecular mechanisms underlying tumour progression (Celis et al., 1996, 1999a, b). So far, nearly 700 tumours of various grades and stages have been analysed by 2D gel

(lvl 1) Human keratinocytes-IEF Database

Proteins in category 'autoantigen'

(Total of 32 proteins listed. Page 1 of 1)

Line	Protein name	SSP	Mr	pI
1	Calreticulin precursor,Non muscle analog of calsequestrin.	9401	60.7	4.16
2	Calreticulin,Non muscle analog of calsequestrin.	9509	63.1	4.15
3	hsc 70, hsc 73, clathrin uncoating ATPase, beta internexin, thermin.	6504	65.6	5.30
4	hsc 70,clathrin uncoating ATPase,beta internexin,thermin variant.	6505	66.3	5.21
5	hsp 28,hsp 27 related.	6111	30.0	5.50
6	hsp 28,hsp 27 variant.	5102	28.3	5.54
7	hsp 28,hsp 27.	4110	28.3	5.97
8	hsp 90,hsp 83.	8612	89.3	4.81
9	hsp 90.	8611	89.6	4.90
10	hsp 90.	7613	89.8	5.17
11	hsp 90.	7614	89.9	5.11
12	hsp 90.	7617	90.0	4.99
13	hsp 90.	7619	90.0	4.95
14	hsp 90.	7616	90.3	5.01
15	hsp 90.	7618	90.6	4.96
16	hsp 90.	6608	91.1	5.23
17	Ku antigen (86 kDa), nuclear factor IV ,primatin.	5602	85.6	5.68
18	Ku antigen (86 kDa), nuclear factor IV, primatin variant.	5624	85.9	5.53
19	Ku antigen (86 kDa), nuclear factor IV, primatin variant.	5601	85.8	5.60
20	Ku antigen (86 kDa), nuclear factor IV, primatin variant.	5623	85.9	5.55
21	Lamin B variant.	6521	65.2	5.04
22	Lamin B.	7510	65.1	5.09
23	nucleolar protein B23, numatrin, nucleophosmin variant.	8332	38.2	4.58
24	Nucleolar protein B23,numatrin, nucleophosmin variant.	8331	38.1	4.60
25	Nucleolar protein B23,numatrin, nucleophosmin.	8207	37.8	4.64
26	Proliferating cell nuclear antigen (PCNA) variant. PCNA/cydin variant.	9229	37.5	4.36
27	Proliferating cell nuclear antigen (PCNA). PCNA/cydin.	9226	37.4	4.42
28	Ribosomal P protein.	9005	14.4	4.04
29	Ribosomal P protein.	4218	36.6	5.68
30	Ribosomal P protein.	4202	36.8	5.85
31	SS-A/Rho variant, Sjogren syndrome antigen 2 variant.	3425	52.5	5.99
32	SS-A/Rho, Sjogren syndrome antigen 2.	4409	52.4	6.03

Show page: 1

Fig. 8. Proteins recorded in the keratinocyte database that can be retrieved from the category "autoantigen". Only a fraction of the file is shown.

protein profiling. So far, several biomarkers have been identified which might be valuable for a more accurate classification of superficial lesions and may facilitate the identification of individuals at risk. We have shown that the expression of a number of proteins including adypocyte-type FABP, GST-μ, PDGH and keratin 13 strongly correlates with a particular step of TCC progression (Celis et al., 1996, 1999a). In addition, several differentiation markers (keratin 10, keratin 14, PA-FABP, stratifin, galectin 7 and psoriasin) have been found to be associated with SCC differentiation (Østergaard et al., 1997; Celis et al. 1999b). We are also in the process of establishing a comprehensive 2D gel database of urine proteins with the aim of identifying protein markers that may prove invaluable for the non-invasive follow-up of patients, as well as for early detection and staging.

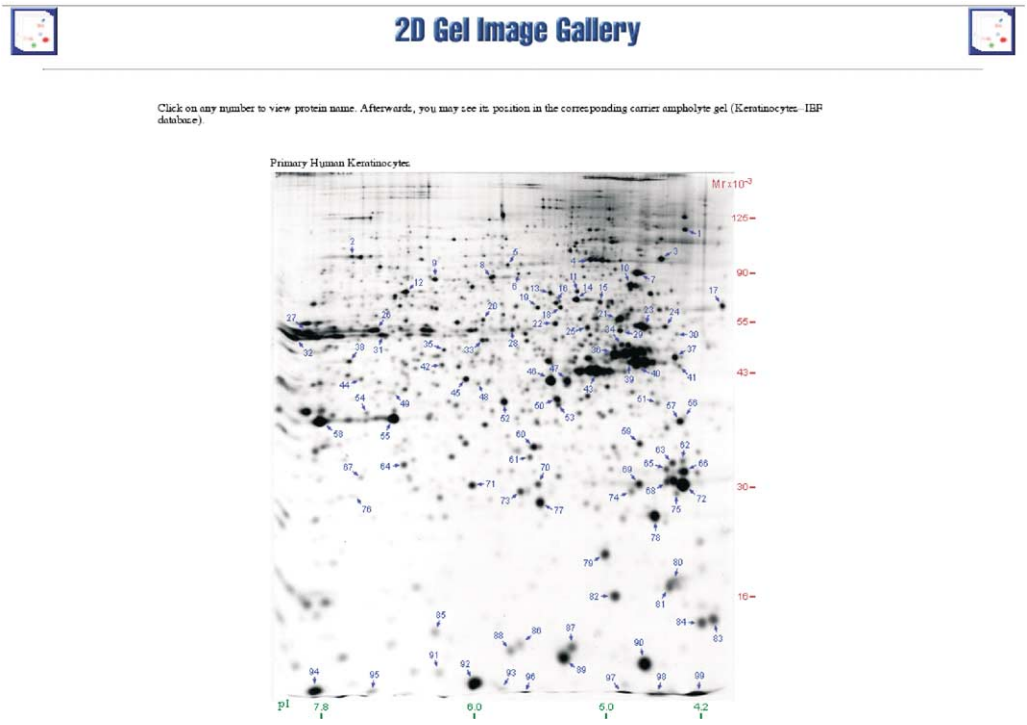

Fig. 9. File showing a 2D gel autoradiograph of [^{35}S]methionine-labelled human keratinocytes proteins resolved in broad-range IPG 3-10. One hundred proteins are indicated as reference.

3.3. 2D gel and blot galleries

To facilitate the comparison of protein expression profiles between various human tissues and cells, we have included in the Web site an extensive gallery of reference 2D gels that include cultured and non-cultured cell types, normal human tissues and tumours, fluids, subproteomes from organelles and fractions, post-translational modifications as well as comparisons (normal vs. SV40 transformed keratinocytes, psoriatic vs. normal keratinocytes, differentiated vs. normal keratinocytes, keratinocytes treated by various interferons, interleukins and growth factors, etc.). We also provide a gallery of 2D gel immunoblots that underline the value of the 2D gel technology to determine antibody specificity, as well as to reveal post-translational modifications. In addition, we present a ZOO-PLANT gallery that contains proteome profiles of cell types from various species. Protocols and videos are included to facilitate and reproduce the data displayed.

4. Concluding remarks

Recent advances in protein identification technologies, in particular mass spectrometry, have made possible the establishment of proteomic 2D PAGE databases that link protein and DNA

mapping and sequence information and that offer an effective route to drug discovery by pinpointing signalling pathways and components that are deregulated in particular diseases. With the integrated approach offered by 2D PAGE databases, it is now possible to reveal and identify phenotype-specific proteins, sequence them, assign partial protein sequences to genes for which the full DNA sequence and the chromosome location are known, and to study the regulatory properties and function of groups of proteins that are co-ordinately expressed in a given biological process. Human 2D PAGE databases are expected to provide an integrated picture of the expression levels and properties of the thousands of protein components of organelles, pathways, and cytoskeletal systems, both in health and disease and are expected to address problems that cannot be approached by DNA analysis, namely, relative abundance of the protein product, post-translational modification, subcellular modification, turnover, interaction with other proteins as well as functional aspects.

There are still many additional challenges that will have to be addressed before complete proteomic projects can be implemented to the study of diseases. These include, to name a few, the heterogeneity of biopsy material, the need to develop better image analysis systems for supporting gel comparisons, quantitations and databasing, determination of interacting partners, functional aspects, as well as the luck of procedures for identifying and functionally characterising targets that lie in pathway of disease. The latter shortcoming is of paramount importance to the pharmaceutical industry, as the identification of disease disregulated targets alone is not sufficient to start a drug screening process. Laser capture microdissection (LCM), a recently developed technique, provides an ideal method for rapid and reliable extraction of cells (or groups of cells) from a given tissue section in which the exact morphologies of both the captured cells and the surrounding tissue are preserved. Recently, it has been shown that proteins could be recovered from laser captured microdissected tissue in a form that is suitable for 2D PAGE analysis and mass spectrometry (Lawrie et al., 2001; Moskaluk, 2001). To that end LCM coupled with 2D gel proteomics would have a considerable impact for authentic proteomic profiling and databasing of human tissues both in health and disease.

Clearly, we have gone a long way from the establishment of the first manual proteomic databases (Bravo et al., 1981), and we expect that these resources will be instrumental to the life sciences in this millennium.

Acknowledgements

This work was supported by grants form the Danish Biotechnology Programme, the Danish Cancer Society and the Danish Centre for Molecular Gerontology.

References

Abbott, A., 1999. A post-genomic challenge: learning to read patterns of protein synthesis. Nature 402 (6763), 715–720.

Bauw, G., Van Damme, J., Puype, M., Vandekerckhove, J., Gesser, B., Ratz, G.P., Lauridsen, J.B., Celis, J.E., 1989. Protein-electroblotting and -microsequencing strategies in generating protein data bases from two-dimensional gels. Proc Natl Acad Sci USA 86 (20), 7701–7705.

Borrrebaeck, C.A., 2000. Antibodies in diagnostics—from immunoassays to protein chips. Immunol Today 21 (8), 379–382.

Bravo, R., Bellatin, J., Celis, J.E., 1981. [^{35}S]-methionine labelled polypeptides from HELA cells. Coordinates and percentage of some major polypeptides. Cell Biol Int Rep. 5 (1), 93–96.

Celis, J.E., Bravo, R., 1984. Synthesis of the nuclear protein cyclin in growing, senescent and morphologically transformed human skin fibroblasts. FEBS Lett 165 (1), 21–25.

Celis, J.E., Gromov, .P., 2000. High-resolution two-dimensional gel electrophoresis and protein identification using western blotting and ECL detection. EXS 88, 55–67.

Celis, J.E., Bravo, R., Larsen, P.M., Fey, S.J., 1984. Cyclin: a nuclear protein whose level correlates directly with the proliferarive state of normal as well as transformed cells. Leukemia Res. 8 (2), 143–157.

Celis, J.E., Madsen, P., Nielsen, S., Rasmussen, H., 1987. Expression of cyclin (PCNA) and the phosphoprotein dividin are late obligatory events in the mitogenic response. Anticancer Res 7 (4A), 605–616.

Celis, J.E., Rasmussen, H.H., Leffers, H., Madsen, P., Honore, B., Gesser, B., Dejgaard, K., Vanderkerkchove, J., 1991a. Human cellular protein patterns and their link to genome DNA sequence data: usefulness of two-dimensional gel electrophoresis and microsequencing. FASEB J. 5 (8), 2200–2208.

Celis, J.E., Leffers, H., Rasmussen, H.H., Madsen, P., Honore, B., Gesser, B., Dejgaard, K., Olsen, E., Ratz, G.P., Lauridsen, J.B., et al., 1991b. The master two-dimensional gel database of human AMA cell proteins: towards linking protein and genome sequence and mapping information (update 1991). Electrophoresis 12 (11), 765–801.

Celis, J.E., Rasmussen, H.H., Madsen, P., Leffers, H., Honore, B., Dejgaard, K., Gesser, B., Olsen, E., Gromov, P., Hoffmann, H.J., et al., 1992. The human keratinocyte two-dimensional gel protein database (update 1992): towards an integrated approach to the study of cell proliferation, differentiation and skin diseases. Electrophoresis 13 (12), 893–959.

Celis, J.E., Rasmussen, H.H., Gromov, P., Olsen, E., Madsen, P., Leffers, H., Honore, B., Dejgaard, K., Vorum, H., Kristensen, D.B., 1995. The human keratinocyte two-dimensional gel protein database (update 1995): mapping components of signal transduction pathways. Electrophoresis 16 (12), 2177–2240.

Celis, J.E., Ostergaard, M., Basse, B., Celis, A., Lauridsen, J.B., Ratz, G.P., Andersen, I., Hein, B., Wolf, H., Orntoft, T.F., Rasmussen, H.H., 1996. Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. Cancer Res 56 (20), 4782–4790.

Celis, J.E., Østergard, M., Jensen, N.A., Gromova, I., Rasmussen, H.H., Gromov, P., 1998. Human and mouse proteomic databases: novel resources in the protein universe. FEBS Lett 430 (1–2), 64–72.

Celis, J.E., Ostergaard, M., Rasmussen, H.H., Gromov, P., Gromova, I., Varmark, H., Palsdottir, H., Magnusson, N., Andersen, I., Basse, B., Lauridsen, J.B., Ratz, G., Wolf, H., Orntoft, T.F., Celis, P., Celis, A., 1999a. A comprehensive protein resource for the study of bladder cancer: http://biobase.dk/cgi-bin/celis. Electrophoresis 20 (2), 300–309.

Celis, J.E., Celis, P., Østergaard, M., Basse, B., Lauridsen, J.B., Ratz, G., Rasmussen, H.H., Orntoft, T.F., Hein, B., Wolf, H., Celis, A., 1999b. Proteomics and immunohistochemistry define some of the steps involved in the squamous differentiation of the bladder transitional epithelium: a novel strategy for identifying metaplastic lesions. Cancer Res 59 (12), 3003–3009.

Celis, J.E., Kruhoffer, M., Gromova, I., Frederiksen, C., Østergaard, M., Thykjaer, T., Gromov, P., Yu, J., Palsdottir, H., Magnusson, N., Orntoft, T.F., 2000. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. FEBS Lett 480 (1), 2–16.

Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., Sanchez, J.C., 2000. The dynamic range of protein expression: a challenge for proteomic research. Electrophoresis 21 (6), 1104–1115.

Dejgaard, K., Celis, J.E., 1998. Two-dimensional Northwestern blotting. In: Celis, J.E., Carter, N., Hunter, T., Shotton, D., Simons, K., Small, J.M. (Eds.), Cell Biology. A Laboratory Handbook., Vol. 4. Academic Press, New York, pp. 482–486.

Duncan, R., McConkey, E.H., 1982. How many proteins are there in a typical mammalian cell? Clin Chem 28 (4), 749–755.

Emili, A.Q., Cagney, G., 2000. Large-scale functional analysis using peptide or protein arrays. Nature Biothechnology 18 (4), 393–397.

Fey, S.J., Larsen, P.M., 2001. 2D or not 2D? Curr Opin Chem Biol 5 (1), 26–233.

Figeys, D., Pinto, D., 2001. Proteomics on a chip: promising developments. Electrophoresis 22 (2), 208–216.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., Garrel, J.I., 1999. A sampling of the yeast proteome. Mol Cell Biol 19 (11), 7357–7368.

Görg, A., 1999. IPG-Dalt of very alkaline proteins. Methods Mol Biol 112, 197–209.

Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W., 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21 (6), 1037–1053.

Gooley, A.A., Packer, N.H., 1997. The importance of protein co- and post-translational modifications in proteome projects. In: Wilkins, M.R., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (Eds.), Proteome research: new frontiers in functional genomics (principles and practice). Springer, Berlin, pp. 65–91.

Gromov, P., Celis, J.E., 1998. Blot overlay assay for the identification of GTP-binding proteins. In: Celis, J.E., Carter, N., Hunter, T., Shotton, D., Simons, K., Small, J.M. (Eds.), Cell Biology. A Laboratory Handbook., Vol. 4. Academic Press, New York, pp. 454–457.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of the complex protein mixtures using isotpe-codded affinity tags. Nature Biotech 17 (10), 994–999.

Hanash, S.M., 2000. Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. Electrophoresis 21 (6), 1202–1209.

Hausken, Z.E., Coghlan, V.M., Scott, J.D., 1998. Overlay, ligand blotting and band-shift techniques to study kinase anchoring. In: Clegg, R.A. (Ed.), Protein Targeting Protocols. Methods in Molecular Biology, Vol. 88. Humana Press, Totowa, NJ, pp. 47–64.

Hoffmann, H.J., Gromov, P., Celis, J.E., 1998. Calcium overlay assay. In: Celis, J.E., Carter, N., Hunter, T., Shotton, D., Simons, K., Small, J.M. (Eds.), Cell Biology. A Laboratory Handbook., Vol. 4. Academic Press, New York, pp. 450–453.

Holt, L.J., Bussow, K., Walter, G., Tomlinson, I.M., 2000. By-passing selection: direct screening for antibody-antigen interactions using protein arrays. Nucleic Acids Res 28 (15), E72.

Klose, J., 1975. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26 (3), 231–72243.

Labrador, M., Mongelard, F., Plata-Rengifo, P., Baxter, E.M., Corces, V.G., Gerasimova, T.I., 2001. Protein encoding by both DNA strands. Nature 409 (6823), 1000.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M.,

Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., Szustakowki, J., 2001. Initial sequencing and analysis of the human genome. Nature 409 (6822), 860–921.

Larsen, M.R., Roepstorff, P., 2000. Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. Fresenius J Anal Chem 366 (6–7), 677–690.

Lawrie, L.C., Curran, S., McLeod, H.L., Fothergill, J.E., Murray, G.I., 2001. Application of laser capture microdissection and proteomics in colon cancer. Mol. Pathol. 54 (4), 253–258.

Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L., Quackenbush, J., 2000. Gene index analysis of the human genome estimates approximately 120,000 genes. Nat Genet 25 (2), 239–240.

Madsen, P., Celis, J.E., 1985. S-phase patterns of cyclin(PCNA) antigen staining resemble topographical patterns of DNA synthesis. A role for cyclin in DNA replication? FEBS Lett 193 (1), 5–11.

Mann, M., Hendrickson, R.C., Pandey, A., 2001. Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem. 70, 437–473.

Miura, M., 1999. Detection of chromatin-bound PCNA in mammalian cells and its use to study DNA excision repair. J Radiat Res 40 (1), 1–12.

Moskaluk, C.A., 2001. Microdissection of histologic sections. Manual and laser capture microdissection techniques. In: Powel, S.M. (Ed.), Colorectal cancer. Methods and protocols. Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 1–13.

Naaby-Hansen, S., Waterfield, M.D., Cramer, R., 2001. Proteomics—post-genomic cartography to understand gene function. Trends Pharmacol Sci 22 (7), 376–384.

O'Farrell, P.H., 1975. High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250 (10), 4007–4021.

O'Farrell, P.Z., Goodman, H.M., O'Farrell, P.H., 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell 12 (4), 1133–1141.

Østergaard, M., Rasmussen, H.H., Nielsen, H.V., Vorum, H., Orntoft, T.F., Wolf, H., Celis, J.E., 1997. Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. Cancer Res 57 (18), 4111–4117.

Panek, J., Vohradsky, J., 1999. Point pattern matching in the analysis of two-dimensional gel electropherograms. Electrophoresis 20 (18), 3483–3491.

Peters, K.F, Menaker, T.J, Wilson, P.L, Hadley, D.W., 2001. The Human Genome Project: an update. Cancer Nurs 24 (4), 287–292.

Pleißner, K.-P., Hoffmann, F., Kriegel, K., Wenk, C., Wegner, S., Sahlström, A., Oswald, H., Alt, H., Fleck, E., 1999. New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases. Electrophoresis 20 (4–5), 755–765.

Rasmussen, H.H., Orntoft, T.F., Wolf, H., Celis, J.E., 1996. Towards a comprehensive database of proteins from the urine of patients with bladder cancer. J Urol 155 (6), 2113–2119.

Roepstorff, P., 2000. MALDI-TOF mass spectrometry in protein chemistry. EXS 88, 81–97.

Santoni, V., Molloy, M., Rabilloud, T., 2000. Membrane proteins and proteomics: un amour impossible? Electrophoresis 21 (6), 1054–1070.

Scheele, J.S., Rhee, J.M., Boss, G.R., 1995. Determination of absolute amounts of GDP and GTP bound to Ras in mammalian cells: comparison of parental and Ras-overproducing NIH 3T3 fibroblasts. Proceedings of the National Acad Sci USA 92 (4), 1097–1100.

Service, R.F., 2000. Proteomics. Can Celera do it again? Science 287 (5461), 2136–2138.

Shen, Y., Berger, S.J., Anderson, G.A., Smith, R.D., 2000. High-efficiency capillary isoelectric focusing of peptides. Anal Chem 72 (9), 2154–2159.

Shen, Y., Tolic, N., Zhao, R., Pasa-Tolic, L., Li, L., Berger, S.J., Harkewicz, R., Anderson, G.A., Belov, M.E., Smith, R.D., 2001. High-throughput proteomics using high-efficiency multiple-capillary liquid chromatography with on-line high-performance ESI FTICR mass spectrometry. Anal Chem 73 (13), 3011–3021.

Smilansky, Z., 2001. Automatic registration for images of two-dimensional protein gels. Electrophoresis 22 (9), 1616–1626.

Steinberg, T.H., Haugland, R.P., Singer, V.L., 1996. Application of SYPRO orange and SYPRO red protein gel stains. Anal Biochem 239 (2), 238–245.

Vandekerckhove, J., Bauw, G., Vancompernolle, K., Honore, B., Celis, J.E., 1990. Comparative two-dimensional gel analysis and microsequencing identifies gelsolin as one of the most prominent downregulated markers of transformed human fibroblast and epithelial cells. J Cell Biol 111 (1), 95–102.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R, Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science. 291 (5507), 1304–1351.

Warbrick, E., 2000. The puzzle of PCNA's many partners. Bioessays 22 (11), 997–1006.

Wheeler, D.L., Church, D.M., Lash, A.E., Leipe, D.D., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Tatusova, T.A., Wagner, L., Rapp, B.A., 2001. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 29 (1), 11–16.

Wilm, M., 2000. Mass spectrometric analysis of proteins. Adv Protein Chem. 54, 1–30.

Zuber, M., Tan, E.M., Ryoji, M., 1989. Identification of proliferating cell nuclear antigen (cyclin) in DNA replication in living cells. Mol Cell Biol. 9 (1), 57–66.

Zuo, X., Speicher, D.W., 2000. A method for global analysis of complex proteomes using sample prefractionation by solution isoelectrofocusing proir to two-dimensional electrophoresis. Anal Biochem 284 (2), 266–278.