COMPUTER DATABASES FOR TWO-DIMENSIONAL ELECTROPHORESIS

T. Toda, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

Two-dimensional (2-D) electrophoresis is an unrivalled technique of the highest capacity to separate many thousands of proteins in complex specimens such as a crude extract of cells. The development of computer systems for 2-D gel image analysis has promoted the construction of comprehensive databases for 2-D electrophoresis. The recent advance in the Internet computer network has allowed us to share useful pieces of information located in many 2-D electrophoresis databases on the world-wide web (WWW) via the Internet. And a new field of research named proteomics has been growing based on 2-D electrophoresis databases.

2-D Electrophoresis

2-D electrophoresis, established by O'Farrell in 1975, is a combined technique of isoelectric focusing (IEF) and sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) for achieving the finest resolution of proteins in crude specimens. Protein molecules of different isoelectric points (p*I*) are separated at the first stage of IEF, and components of different molecular masses are resolved in the second-dimensional SDS-PAGE.

The IEF in a mobile pH gradient, which is automatically generated by lining up carrier ampholytes in the order of p*I*, is performed in the original protocol of O'Farrell's 2-D electrophoresis. However, there is still a problem with the conventional IEF method in the drift of the pH gradient toward the cathode. Righetti and colleagues have devised an improved method of IEF which runs on an immobilized pH gradient (IPG) to prevent cathodic drift. The IPG-IEF is adopted in most current protocols of 2-D electrophoresis for constructing computer databases.

Protein Detection

Proteins separated on a gel plate are detected by autoradiography if they are labelled with a radioisotope. Nonradioactive proteins are visualized by silver or dye staining depending on the protein amount on the gel. The protein profile on the 2-D gel reports useful information about qualitative and quantitative states of proteins in the specimen. Most computer databases for 2-D electrophoresis have generally been constructed on such visualized 2-D gel images.

2-D Gel Image Analysis

Computer analysis of the 2-D gel images was developed after O'Farrell's paper because 2-D gel images were too complicated to be analysed manually. Lipkin and Lemkin of the National Institutes of Health in the USA made their GELLAB system on a DEC minicomputer in 1980, and independently of this, Anderson *et al.* constructed their original TYCHO system in 1981. Garrels and Franza prepared a prototype software for 2-D gel image analysis on a compact Hewlett Packard desktop computer in 1979, and then developed it into the QUEST system for an UNIX-based minicomputer in 1983. The PDQUEST, which is a commercially available software package for SUN UNIX workstations, was a revision of the QUEST system. Many other software packages, such as Melanie II and GELLAB II, have been made commercially available. Those Macintosh- and Windows-based software packages have allowed 2-D gel image analysis to be performed easily for database construction.

Western Blotting

Western blotting is an immunochemical technique for detecting specific proteins. Protein spots separated by 2-D electrophoresis are transferred on to a nitrocellulose or polyvinylidene difluoride (PVDF) membrane by electrotransfer blotting. After blocking, the membranes are treated with a specific antibody and followed by a second enzyme-linked antibody. Localization of specific protein on the membrane is visualized by an enzyme reaction. The results of spot protein identification by Western blotting are also included in 2-D electrophoresis databases.

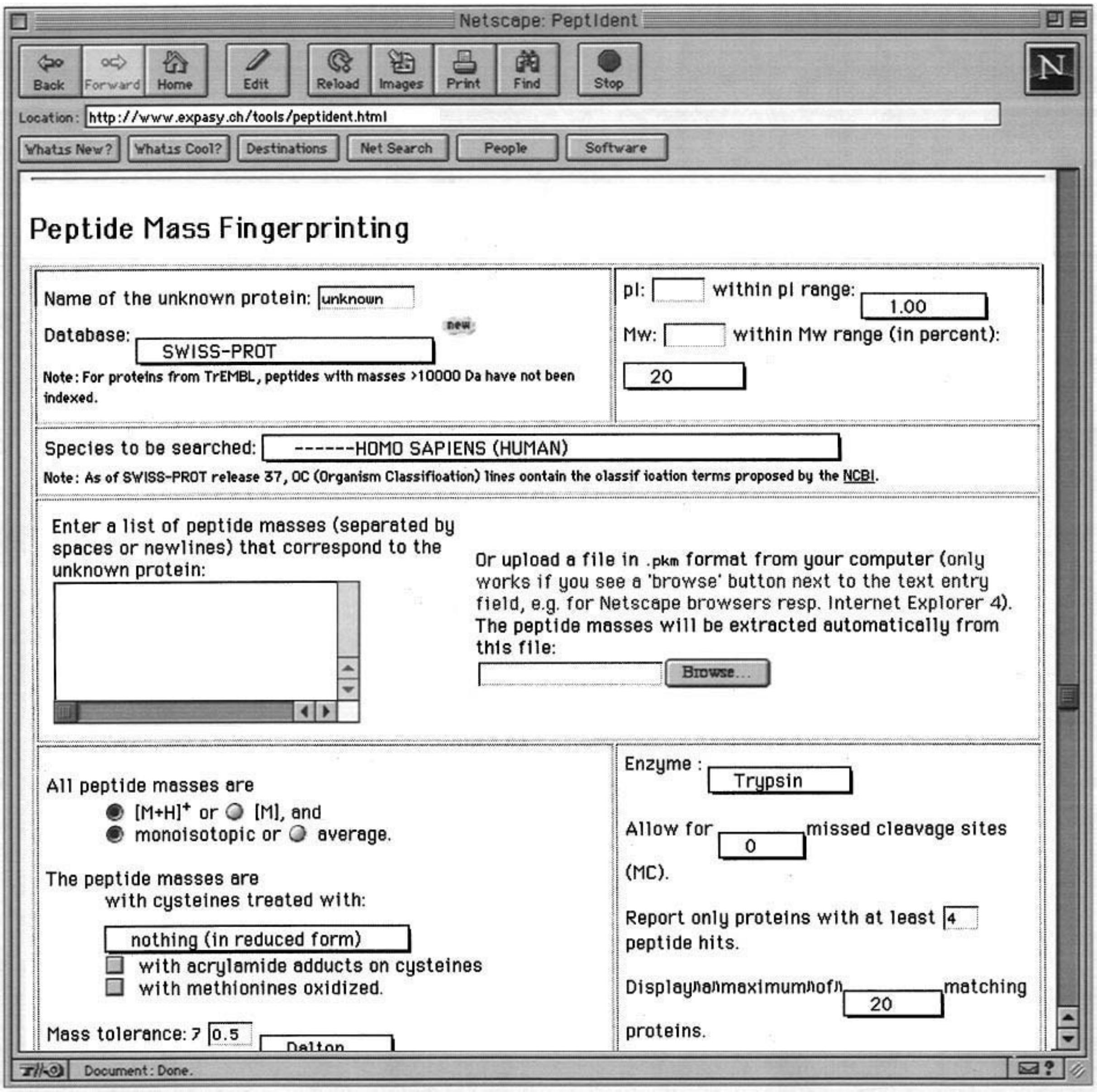
Microsequencing

The N-terminal peptide sequences of proteins are directly determined with an automatic peptide sequencer after electrotransfer blotting on to a PVDF membrane, if the N-terminal α-amino group is free.

Proteins that are blocked at the N-terminus are often cleaved with endopeptidase. The digests are separated by reversed-phase high performance liquid chromatography (HPLC) and then subjected to sequencing. Protein databases on the Internet may answer a query on the homology of the amino acid sequence for identification of the protein if the database includes an entry of the protein itself or family gene products. The sequence data and the results of homology search are both valuable information for constructing 2-D electrophoresis databases.

Mass Spectrometry

Mass spectrometry is another powerful technique for identification of proteins separated by 2-D electrophoresis. The peptide mass fingerprinting of endopeptidase digests is extensively used for primary assignment of the protein. The ExPASy Molecular Biology server (http://www.expasy.ch/) offers useful proteomic tools such as PeptIdent, which helps us in peptide mass fingerprinting works for protein identification (**Figure 1**).



Location: http://www.expasy.ch/tools/peptident.html

Peptide Mass Fingerprinting

Name of the unknown protein: unknown

Database: SWISS-PROT

Note: For proteins from TrEMBL, peptides with masses >10000 Da have not been indexed.

pI: within pI range: 1.00

Mw: within Mw range (in percent): 20

Species to be searched: ------HOMO SAPIENS (HUMAN)

Note: As of SWISS-PROT release 37, OC (Organism Classification) lines contain the classification terms proposed by the NCBI.

Enter a list of peptide masses (separated by spaces or newlines) that correspond to the unknown protein:

Or upload a file in .pkm format from your computer (only works if you see a 'browse' button next to the text entry field, e.g. for Netscape browsers resp. Internet Explorer 4). The peptide masses will be extracted automatically from this file:

All peptide masses are
● [M+H]$^+$ or ○ [M], and
● monoisotopic or ○ average.

The peptide masses are
with cysteines treated with:
nothing (in reduced form)
☐ with acrylamide adducts on cysteines
☐ with methionines oxidized.

Mass tolerance: ? 0.5 Dalton

Enzyme : Trypsin

Allow for 0 missed cleavage sites (MC).

Report only proteins with at least 4 peptide hits.

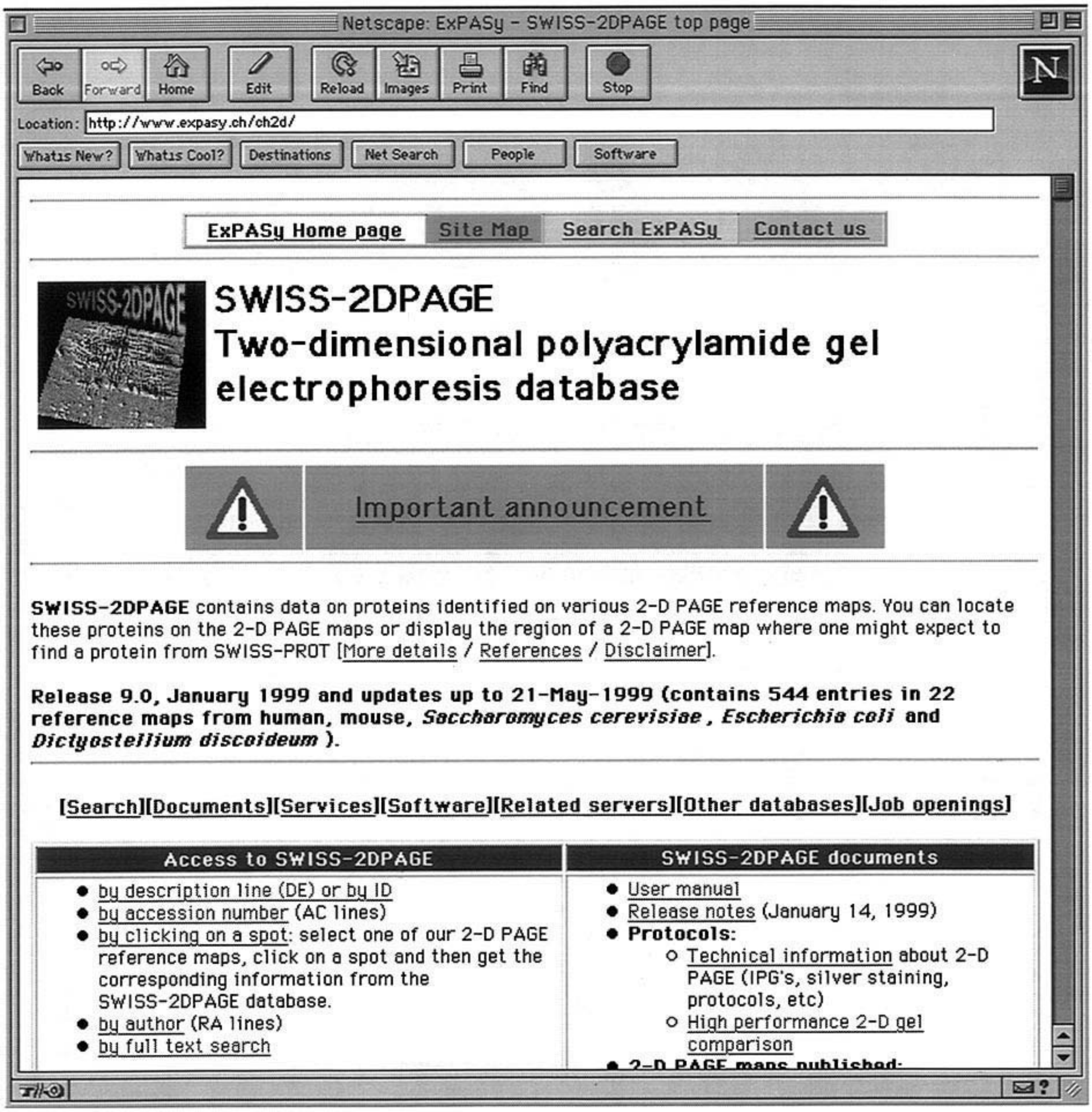
Displayıaımaximumıofı 20 matching proteins.

Figure 1 Proteomic analysis tool PetIdent in the ExPASy Molecular Biology server. The URL is http://www.expasy.ch/tools/peptident.html.

History of 2-D Electrophoresis Databases

The 2-D electrophoresis database, reported by Lipkin and Lemkin in 1980, was only for multiple 2-D gel image analyses on a stand-alone minicomputer. In their GELLAB system, a composite gel (CGEL) database was made by extracting spot data from multiple gels and merging them into a representative gel. The primary database consisted of the lists of corresponding spots, their associated properties and interrelations. The human serum protein 2-D electrophoresis database for clinical use, reported by Anderson and his co-workers in 1981, was also only for personal use on an offline computer. Human keratinocyte 2-D electrophoresis databases, made by Celis *et al.* on a UNIX workstaion with PDQUEST software, were offered to many research groups commercially, but not made accessible on the Internet.

Appel and his co-workers constructed their SWISS-2DPAGE database (**Figure 2**) on the ExPASy molecular biology server of the Swiss Geneva University Hospital in 1993. It was the first regular 2-D electrophoresis database on the WWW, and it has been accessible on the Internet.



Location: http://www.expasy.ch/ch2d/

SWISS-2DPAGE
Two-dimensional polyacrylamide gel electrophoresis database

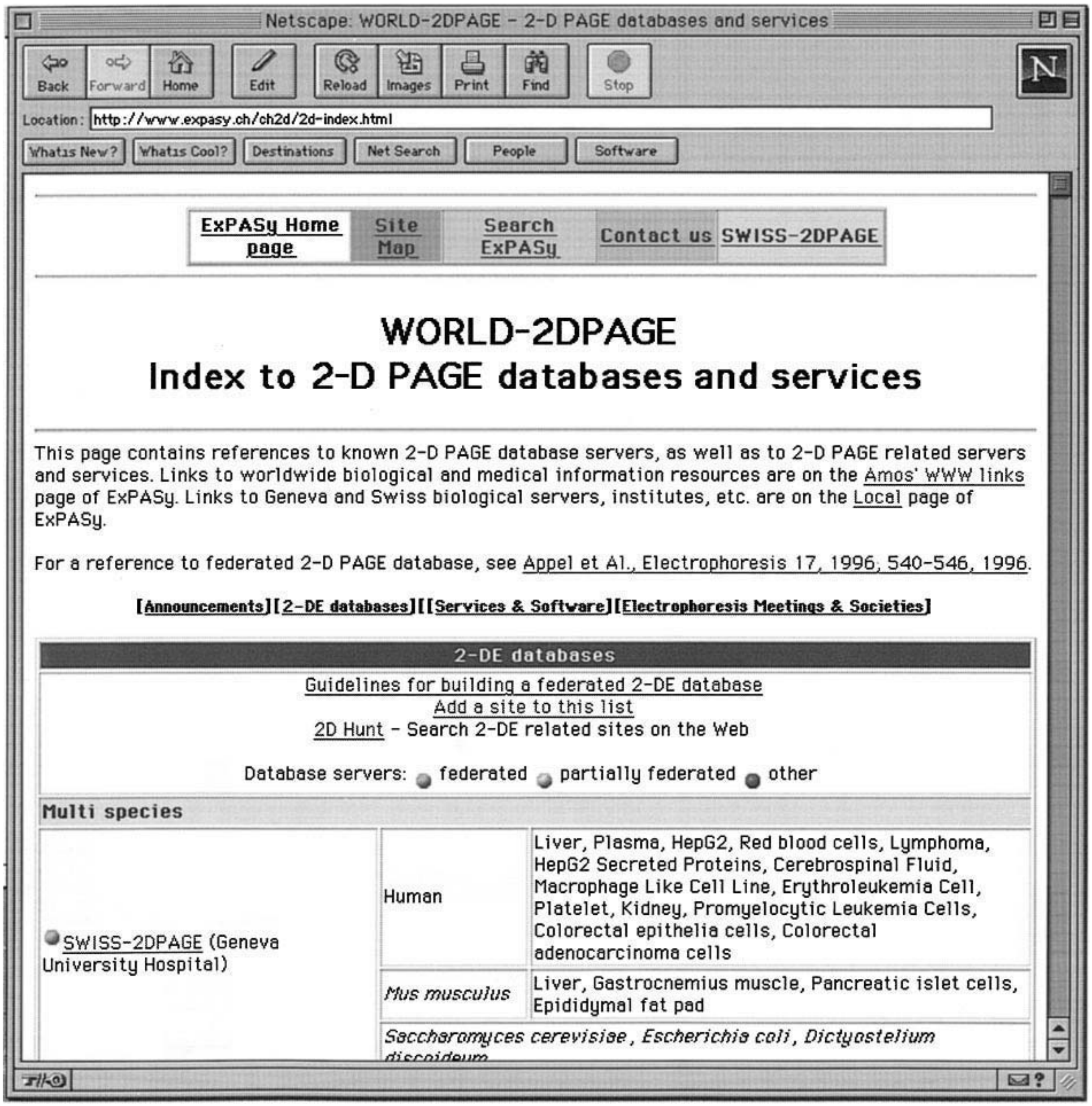
SWISS-2DPAGE contains data on proteins identified on various 2-D PAGE reference maps. You can locate these proteins on the 2-D PAGE maps or display the region of a 2-D PAGE map where one might expect to find a protein from SWISS-PROT [More details / References / Disclaimer].

Release 9.0, January 1999 and updates up to 21-May-1999 (contains 544 entries in 22 reference maps from human, mouse, *Saccharomyces cerevisiae*, *Escherichia coli* and *Dictyostellium discoideum*).

[Search][Documents][Services][Software][Related servers][Other databases][Job openings]

Access to SWISS-2DPAGE	SWISS-2DPAGE documents
● by description line (DE) or by ID ● by accession number (AC lines) ● by clicking on a spot: select one of our 2-D PAGE reference maps, click on a spot and then get the corresponding information from the SWISS-2DPAGE database. ● by author (RA lines) ● by full text search	● User manual ● Release notes (January 14, 1999) ● **Protocols:** ○ Technical information about 2-D PAGE (IPG's, silver staining, protocols, etc) ○ High performance 2-D gel comparison ● **2-D PAGE maps published:**

Figure 2 The SWISS-2DPAGE top page in the ExPASy Molecular Biology server of the Swiss Geneva University Hospital. The URL is http://www.expasy.ch/ch2d/.



WORLD-2DPAGE
Index to 2-D PAGE databases and services

This page contains references to known 2-D PAGE database servers, as well as to 2-D PAGE related servers and services. Links to worldwide biological and medical information resources are on the Amos' WWW links page of ExPASy. Links to Geneva and Swiss biological servers, institutes, etc. are on the Local page of ExPASy.

For a reference to federated 2-D PAGE database, see Appel et Al., Electrophoresis 17, 1996, 540-546, 1996.

[Announcements][2-DE databases][[Services & Software][Electrophoresis Meetings & Societies]

2-DE databases

Guidelines for building a federated 2-DE database
Add a site to this list
2D Hunt - Search 2-DE related sites on the Web

Database servers: ● federated ● partially federated ● other

Multi species

● SWISS-2DPAGE (Geneva University Hospital)	Human	Liver, Plasma, HepG2, Red blood cells, Lymphoma, HepG2 Secreted Proteins, Cerebrospinal Fluid, Macrophage Like Cell Line, Erythroleukemia Cell, Platelet, Kidney, Promyelocytic Leukemia Cells, Colorectal epithelia cells, Colorectal adenocarcinoma cells
	Mus musculus	Liver, Gastrocnemius muscle, Pancreatic islet cells, Epididymal fat pad
	Saccharomyces cerevisiae, *Escherichia coli*, *Dictyostelium discoideum*	

Figure 3 The WORLD-2DPAGE home page for indexing to 2-D PAGE databases and services throughout the world. The URL is http://www.expasy.ch/ch2d/2d-index.html.

WORLD-2DPAGE home page (**Figure 3**) presented at http://www.expasy.ch/ch2d/2d-index.html in the ExPASy server offers a convenient link to many 2-DE databases in the world.

The databases listed in **Table 1** were indexed in the WORLD-2DPAGE page for link as of 15 June, 1999.

Some of the databases have been built following the guidelines for the federated 2-D electrophoresis database recommended by Appel *et al*. Rules for the database federation are summarized in **Table 2**.

World-Wide Web for 2-D Electrophoresis Database

The WWW is a database server system which was initiated in the European Centre for Nuclear Research on the basis of Internet protocol technology. The WWW allows browsers on a computer to access information stored on remote servers through the Internet. One of the main features of WWW documents is the hypertext structure; another significant feature is the function of a clickable image map.

Table 1 A partial list of 2-D electrophoresis databases indexed in the WORLD-2DPAGE home page

SWISS-2DPAGE (Geneva University Hospital)	http://www.expasy.ch/ch2d/
SIENA-2DPAGE: 2D-PAGE database at Department of Molecular Biology, University of Siena, Italy	http://www.bio-mol.unisi.it/2d/2d.html
2-DE maps at LSB Corp	http://WWW.LSBC.COM/2dmaps/patterns.htm
Human and mouse 2-D PAGE databases:	
Danish Centre for Human Genome Research	http://biobase.dk/cgi-bin/celis
RAT HEART-2DPAGE (German Heart Institute, Berlin)	http://gelmatching.inf.fu-berlin.de/ ~ pleiss/2d/
HEART-2DPAGE (German Heart Institute, Berlin)	http://userpage.chemie.fu-berlin.de/ ~ pleiss/dhzb.html
HSC-2DPAGE (Heart Science Centre, Harefield Hospital)	http://www.harefield.nthames.nhs.uk/nhli/protein/
PDD: Protein Disease Database (NIMH–NCI)	http://www-lecb.ncifcrf.gov/PDD/
PPDB: PhosphoProtein DataBase	http://www-lecb.ncifcrf.gov/phosphoDB/
Molecular Anatomy Laboratory at Indiana University, Columbus	http://iupucbio1.iupui.edu/frankw/molan.htm
Human Colon Carcinoma Protein Database at JPSL, Ludwig Institute for Cancer Research, Melbourne, Australia	http://www.ludwig.edu.au/jpsl/jpslhome.html
TMIG-2DPAGE: Age-related Protein Database at Tokyo Metropolitan Institute of Gerontology	http://proteome.tmig.or.jp/2D/
UCSF 2D PAGE (A375 cell line)	http://rafael.ucsf.edu/2DPAGE/home.html

NIMH-NCI, National Institute of Mental Health and National Cancer Institute.

Hypertext Cross-Reference

Links from the 2-D electrophoresis database to related databases, such as GenBank DNA database in NCBI Entrez, SWISS-PROT protein database or SWISS-3DIMAGE three-dimensional structure databases, through hypertext cross-references is a common function of most databases for 2-D electrophoresis. The function is achieved with anchor tags in hypertext mark-up language (HTML) and executive scripts for common gateway interface (CGI). For example, in TMIG-2DPAGE, the data entry for human nm-23 carries the following hypertext cross-reference to the corresponding SWISS-PROT (P15531) and NCBI GenBank (X75598) entries respectively.

SWISS-PROT

```
<A HREF = “http://www.expasy.ch/cgi-bin/
sprot-search-ac?P15531” TARGET = “_blank”>
P15531 </A>
```

GenBank

```
<A HREF = “http://www3.ncbi.nlm.nih.gov/
htbin-post/Entrez/ query?db = n&form =
6&dopt = g&uid = X75598” TARGET =
“_blank”> X75598 </A>
```

Table 2 Guidelines for building a federated 2-DE database

1. Individual entries in the database must be accessible from remote by keyword search
2. The database must be linked to other databases through active hypertext cross-references
3. A main index has to be supplied that provides a means of querying all databases through one unique entry point
4. Individual protein entries must be accessible through clickable images
5. 2-DE analysis software, been designed for use with federated 2-DE database, must be able to access directly individual entries in any federated 2-DE database

For full details, see Appel RD *et al.* (1996) in *Electrophoresis* 17: 540–546.

Clickable Image Map

Clickable image map is a function of CGI installed in the WWW server software for achieving an active link from a location on an image to a corresponding file. All federated 2-D electrophoresis databases on the WWW have 2-D gel images for the clickable image map. The content of each protein data entry is displayed on the monitor screen when the mouse button is clicked on the corresponding spot on the 2-D gel image.

Setting up a 2-D Electrophoresis Database on the WWW

To set up a 2-D electrophoresis database, software for WWW server, such as Apache httpd, must be working on a UNIX or LINUX server computer connected to the Internet. The function of clickable image map is included in the Apache modules if they are properly installed. To set up a new 2-D electrophoresis database, prepare the following four types of files first, according to the *Apache Reference Manual*, which is available in the Apache directory.

1. A hypertext file for the image map top page (***.html).
2. A 2-D gel image file in the GIF format for the clickable image map (***.gif).

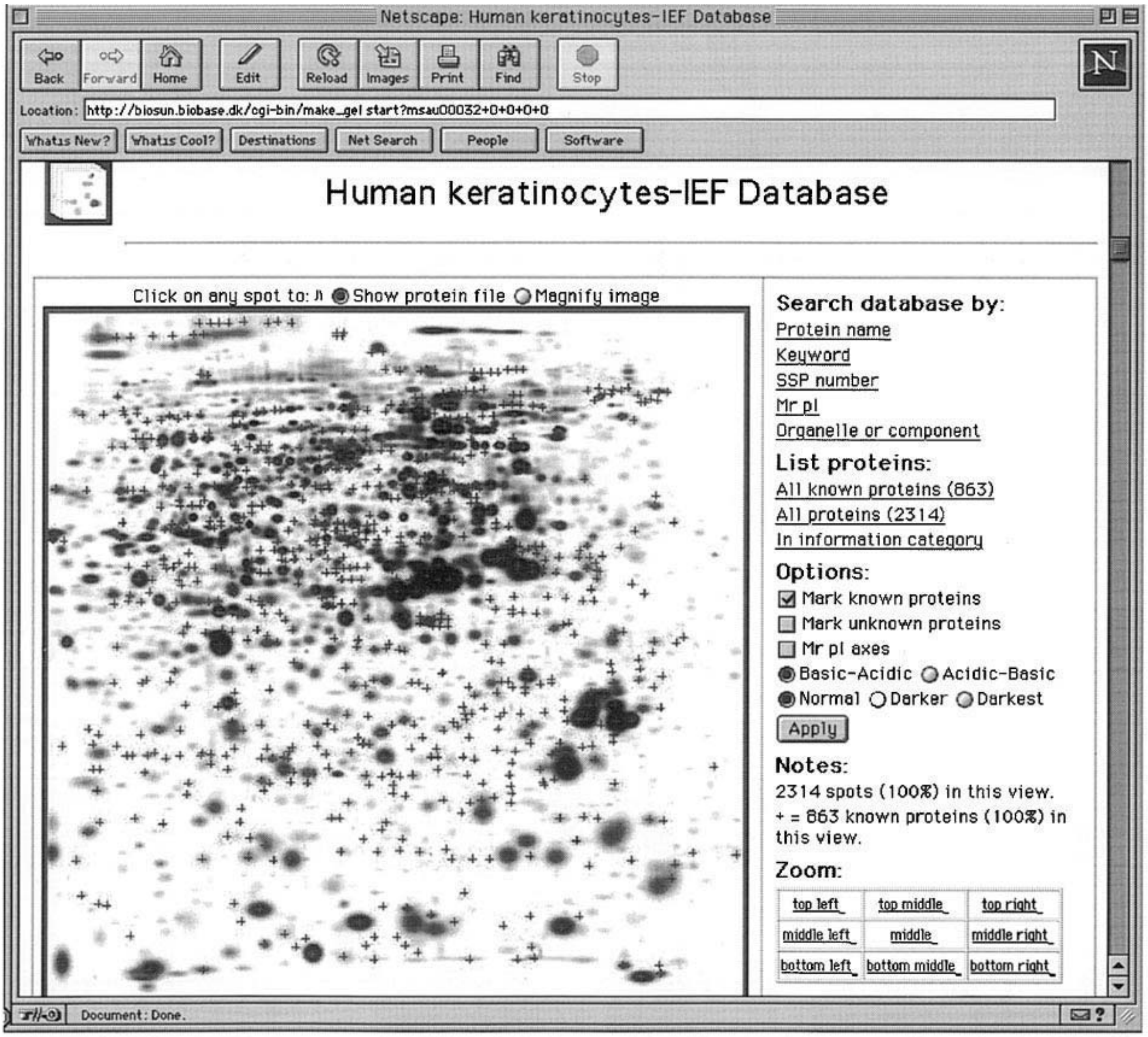
3. A map file for directing the jumping destination (***.map).
4. Text files of protein data entries as destinations for links (***.html).

When these files are placed in the appropriate directories in the WWW server, the basal activity of the 2-D electrophoresis database starts running.

2-D Electrophoresis Databases on the WWW

SWISS-2DPAGE

The SWISS-2DPAGE is a fully implemented 2-DE federated database. The ExPASy WWW server is home to five locally maintained databases (SWISS-PROT, SWISS-2DPAGE, PROSITE, SWISS-3DIMAGE and ENZYME). Besides the database service, the server offers special tools for proteomic analysis, such as PeptIdent for peptide mass fingerprinting. SWISS-PROT serves as a searchable main index, prepared according to rule 3, and has cross-references to the other four databases. SWISS-2DPAGE contains data on protein identified on various 2-D PAGE reference maps. The hypertext cross-references as of 15 June 1999 include links to SWISS-PROT, YPD database of yeast *Saccharomyces* and the ECO2DBASE *E. coli* database. The X-ray crystallography Protein Data Bank (PDB), the Mendelian Inheritance in Man data bank (OMIM), the G-protein-coupled receptor database (GCRDb) and the



Location: http://biosun.biobase.dk/cgi-bin/make_gel start?msau00032+0+0+0+0

Human keratinocytes-IEF Database

Click on any spot to: ● Show protein file ○ Magnify image

Search database by:
Protein name
Keyword
SSP number
Mr pI
Organelle or component

List proteins:
All known proteins (863)
All proteins (2314)
In information category

Options:
☑ Mark known proteins
☐ Mark unknown proteins
☐ Mr pI axes
● Basic-Acidic ○ Acidic-Basic
● Normal ○ Darker ○ Darkest
Apply

Notes:
2314 spots (100%) in this view.
+ = 863 known proteins (100%) in this view.

Zoom:

top left	top middle	top right
middle left	middle	middle right
bottom left	bottom middle	bottom right

Figure 4 Human keratinocytes 2-D PAGE database in the Danish Centre for Human Genome Research. The URL is http://biosun.biobase.dk/cgi-bin/make_gel.start?msau00032+0+0+0+0.

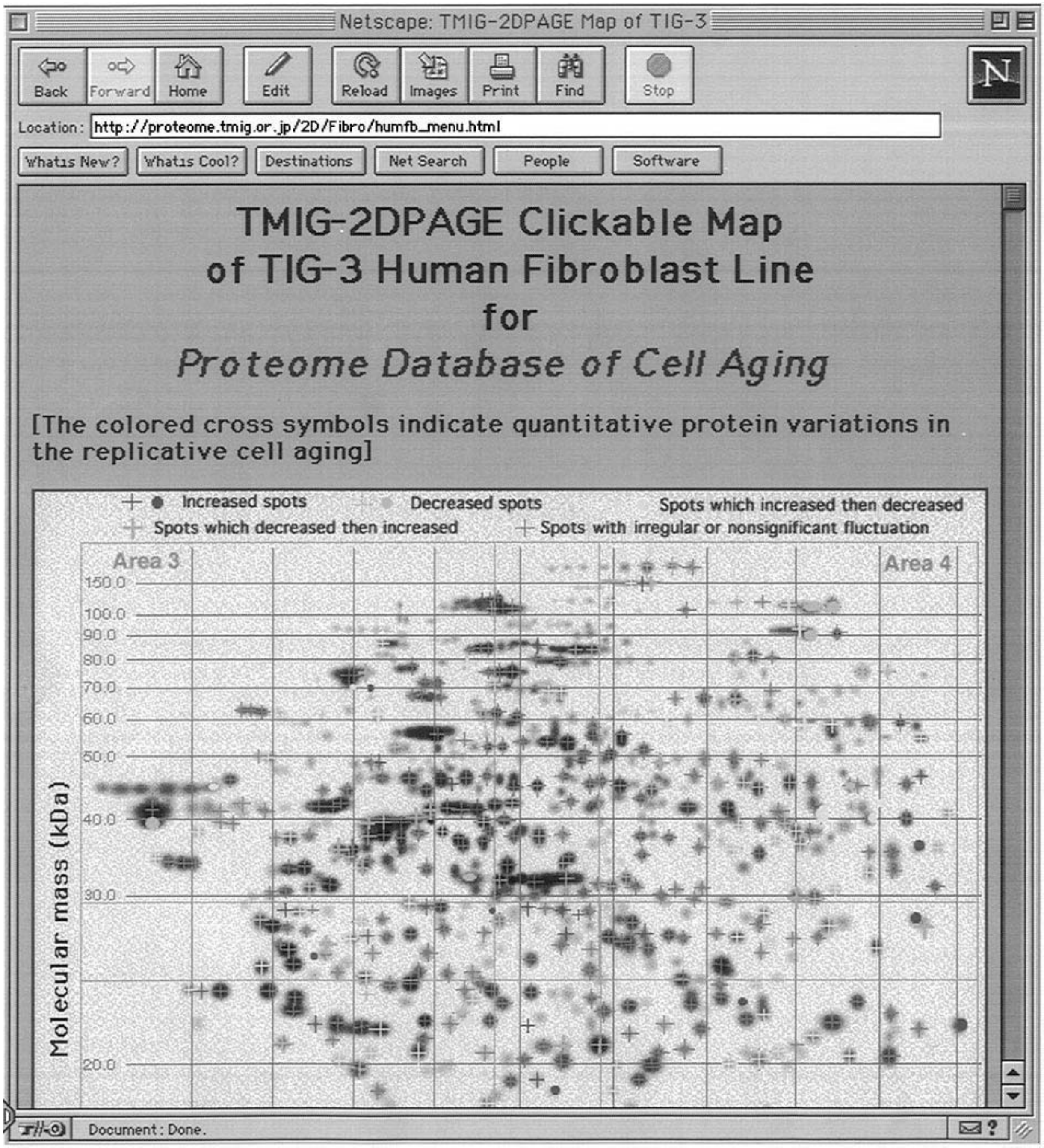
database of yeast (*Saccharomyces cerevisiae*) genes coding for proteins (LISTA) are also linked through SWISS-PROT.

Human 2-D PAGE Databases of the Danish Centre For Human Genome Research

The human 2-D PAGE databases at the University of Aarhus, which were developed for functional genome analysis in health and disease, contain data on proteins identified on various reference maps. Databases of human keratinocytes (**Figure 4**) are served for the study of skin biology, and those of transitional cell carcinomas are for the study of bladder cancer.

Heart-2DPAGE

The Heart-2DPAGE at the German Heart Institute in Berlin is a human myocardial 2-D protein database that implements the first four rules of the



http://proteome.tmig.or.jp/2D/Fibro/humfb_menu.html

TMIG-2DPAGE Clickable Map
of TIG-3 Human Fibroblast Line
for
Proteome Database of Cell Aging

[The colored cross symbols indicate quantitative protein variations in the replicative cell aging]

Figure 5 TMIG-2DPAGE clickable map of TIG-3 human fibroblast line. The URL is http://proteome.tmig.or.jp/2D/Fibro/humfb_menu.html.

federated 2-D database. It contains data on the human heart ventricle and atrium. The data are accessible both by keyword search on the protein name and by mouse-clicking on two images of 2-D. Entries provide data related to the isoelectric point, the molecular mass and the amino acid sequence of each spot protein.

PDD Protein Disease Database

The PDD Protein Disease Database on the WWW server is a part of the National Institute of Mental Health and National Cancer Institute (NIMH-NCI) Protein–Disease Database Project for correlating diseases with proteins observed in serum, cerebrospinal fluid, urine and other common human body fluids based on biomedical literature. The PDD database includes the data of quantitative and qualitative protein variations with disease states, answering questions on protein patterns found in common body fluids with respect to disease conditions in the literature.

TMIG-2DPAGE

TMIG-2DPAGE is a database of human proteins involved in the mechanisms of cell ageing. It includes two clickable image maps for data entries of proteins in normal cell ageing, and for those in the disease state of Werner's syndrome patients. On the TMIG-2DPAGE clickable map of TIG-3 human fibroblast line, quantitative variations of proteins observed in the process of replicative cell ageing are demonstrated with coloured crosses (**Figure 5**).

As a mouse button is clicked on a protein spot, the corresponding data entry is displayed on the browser. Each data entry contains information on age-related protein variation, physicochemical properties, references and active links to both SWISS-PROT and NCBI Entrez nucleotide sequence databases.

Keyword Search of 2-D Electrophoresis Databases

The 2DHunt on the ExPASY server (http://www.expasy.ch/ch2d/2DHunt/) is a convenient search engine for WWW sites of 2-D electrophoresis databases. The site list for 2DHunt search is periodically created by the site retrieval robot supplied from the Marvin (Multi-Agent Retrieval Vagabond on Information Networks). Although some irrelevant sites may sometimes be hit, the search engine is helpful for many researchers to find specific 2-D electrophoresis databases of their own interests.

Further Reading

Anderson NL, Taylor J, Scandore A *et al.* (1981) The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clinical Chemistry* 27(11): 1807–1820.

Appel RD, Hoogland C, Bairoch A and Hochstrasser DF (1999) Constructing a 2-D database for the World Wide Web. *Methods in Molecular Biology* 112: 411–416.

Celis JE, Gromov P, Stergaard M *et al.* (1996) Human 2-D PAGE databases for proteome analysis in health and disease: http://biobase.dk/cgi-bin/celis. *FEBS Letters* 398(2–3): 129–134.

Garrels JI and Franza Jr BRF (1989) The REF52 protein database. *Journal of Biological Chemistry* 264(9): 5283–5298.

Hoogland C, Sanchez JC, Tonella L *et al.* (1998) Current status of the SWISS-2DPAGE database. *Nucleic Acids Research* 26(1): 332–333.

Lemkin PF (1997) The 2DWG meta-database of two-dimensional electrophoretic gel images on the Internet. *Electrophoresis* 18(15): 2759–2773.

Lipkin LE and Lemkin PF (1980) Data-base techniques for multiple two-dimensional polyacrylamide gel electrophoresis analyses. *Clinical Chemistry* 26(10): 1403–1412.

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* 250(10): 4007–4021.

Righetti PG, Gianazza E and Bjellqvist B (1983) Modern aspects of isoelectric focusing: two-dimensional maps and immobilized pH gradients. *Biochemical and Biophysical Methods* 8(2): 89–108.

Toda T and Kimura N (1998) TMIG-2DPAGE: a new concept of two-dimensional gel protein database for research on aging. *Electrophoresis* 18(2): 344–348.

Toda T and Ohashi M (1995) Review: Current status and perspectives of computer-aided two-dimensional densitometry. *Journal of Chromatography A* 698(1): 41–54.

Vesterberg O and Svensson H (1966) Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. IV. Further studies on the resolving power in connection with separation of myoglobins. *Acta Chemica Scandinavica* 20(3): 820–834.

CONTINUOUS ION EXCHANGE USING POWDERED RESINS

See III / POWDERED RESINS: CONTINUOUS ION EXCHANGE