3862–3865 *Nucleic Acids Research, 2003, Vol. 31, No. 13*
DOI: 10.1093/nar/gkg536

JVirGel: calculation of virtual two-dimensional protein gels

Karsten Hiller, Max Schobert, Claudia Hundertmark[1], Dieter Jahn* and Richard Münch

Institut für Mikrobiologie, Technische Universität Braunschweig, Spielmannstrasse 7, D-38106 Braunschweig, Germany and [1]Institut für Software, Abteilung Informationssysteme, Technische Universität Braunschweig, Mühlenpfordtstrasse 23, D-38106 Braunschweig, Germany

Received February 14, 2003; Revised March 12, 2003; Accepted March 26, 2003

ABSTRACT

We developed JVirGel, a collection of tools for the simulation and analysis of proteomics data. The software creates and visualizes virtual two-dimensional (2D) protein gels based on the migration behaviour of proteins in dependence of their theoretical molecular weights in combination with their calculated isoelectric points. The utilization of all proteins of an organism of interest deduced from genes of the corresponding genome project in combination with the elimination of obvious membrane proteins permits the creation of an optimized calculated proteome map. The electrophoretic separation behaviour of single proteins is accessible interactively in a Java™ applet (small application in a web browser) by selecting a pI/MW range and an electrophoretic timescale of interest. The calculated pattern of protein spots helps to identify unknown proteins and to localize known proteins during experimental proteomics approaches. Differences between the experimentally observed and the calculated migration behaviour of certain proteins provide first indications for potential protein modification events. When possible, the protein spots are directly linked via a mouse click to the public databases SWISS-PROT and PRODORIC. Additionally, we provide tools for the serial calculation and visualization of specific protein properties like pH dependent charge curves and hydrophobicity profiles. These values are helpful for the rational establishment of protein purification procedures. The proteomics tools are available on the World Wide Web at http://prodoric.tu-bs.de/proteomics.php.

INTRODUCTION

Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) is currently one of the most comprehensive techniques for separating thousands of proteins simultaneously with high reproducibility. The completion of numerous genome sequencing projects offers the rapid identification of these protein spots via N-terminal amino acid sequence determination or mass spectrometry in combination with peptide mass fingerprinting (1). The complex pattern of protein spots usually observed is mainly the result of the individual separation behaviour of each protein dependent on its unique molecular weight (MW) and isoelectric point (pI). However, the degree of visibility of individual proteins of the pattern relies on their cellular concentration, the amount of closely separating proteins and the separation quality of the chosen pH/MW-range of the employed electrophoretic system. Calculation of the theoretical MW and the pI from the amino acid composition of a protein (2,3), in combination with the physical principles of the employed electrophoretic separation procedures, allows determination of the approximate position of protein spots during 2D PAGE. After determination of these two parameters it is possible to construct and visualize a virtual 2D gel. This may serve both as a reference map for protein spot identification and for the choice of the optimal pH gradient in order to characterize proteins of interest.

Users are asked to cite this article when publishing results which have been obtained with the tools described here.

MATERIALS AND METHODS

The World Wide Web server for the JVirGel tool is Apache 2.0 (http://www.apache.org) on a Linux SuSE 8.0 operating system (http://www.suse.com). The web interface and parts of the calculations were programmed by use of PHP 4.3 technology (http://www.php.net). The graphical output for charge curves and hydrophobicity profiles was realized with PHP and the GD library module (http://www.boutell.com/gd). The virtual two-dimensional (2D) gel applet was developed with Java™ SDK 1.4 (http://java.sun.com).

Theoretical pIs and charge curves were calculated according to Skoog *et al.* (2). Algorithms and hydropathy index for calculating hydrophobicity plots were taken from Kyte and Doolittle (4). Prediction of transmembrane helices was done using TMHMM 2.0 (5) on the web server (http://www.cbs.dtu.dk/services/TMHMM). Non-redundant complete

*To whom correspondence should be addressed. Tel: +49 5313915801; Fax: +49 5313915854; Email: d.jahn@tu-bs.de

Nucleic Acids Research, Vol. 31, No. 13 © Oxford University Press 2003; all rights reserved

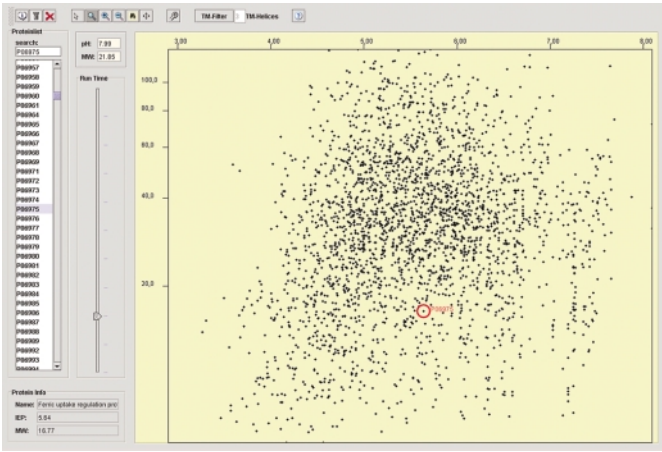

Figure 1. Screenshot of the virtual 2D protein gel applet. The image shows the proteome of *Escherichia coli* in a pH range of 3.0–8.0 U. The Fur (ferric uptake regulator) protein is marked.

proteome sets were downloaded from the Proteome Analysis database (6) at the European Bioinformatics Institute (http://www.ebi.ac.uk/proteome). 2D-PAGE experiments were performed as described before (7).

DESCRIPTION OF THE WEB INTERFACE AND TOOLS PROVIDED

The program JVirGel accepts FASTA formatted protein sequences. There is the choice of either copying and pasting the amino acid sequence data into a textbox, of uploading a FASTA file or selecting a whole proteome out of a selection list. Currently six prokaryotic (*Bacillus subtilis*, *Escherichia coli*, *Helicobacter pylori*, *Listeria monocytogenes*, *Pseudomonas aeruginosa* and *Staphylococcus aureus*) and six eukaryotic (*Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Homo sapiens*, *Mus musculus* and *Saccharomyces cerevisiae*) proteomes are accessible. When selecting a whole proteome, it is possible to filter out potential membrane proteins. After selection of the protein sequences, the program offers three different analytical modes: (a) a virtual 2D gel as Java™ applet; (b) a virtual 2D gel as clickable HTML document; and (c) a serial calculation of MW, pI, pH-dependent charge curves and hydrophobicity profiles.

The first mode (a) is the default selection of the software package and starts the applet embedded in a web page. For this feature it is necessary that a Java™ runtime environment 1.3 or 1.4 is installed on the operating system. The applet was successfully tested for Windows, Linux and MacOS X platforms using Netscape 4.7–7.0, Internet Explorer 5.0–6.0 and Opera 7. Please note, that some older browser versions, such as Netscape 4.7, may have problems in some environments. After starting the applet, all imported proteins are listed on the left-hand side of the screen. On the right-hand side a graphical view of the virtual gel is shown (Fig. 1). Using the slider function, it is possible to change the timescale for the electrophoresis of the gel dynamically. This allows the distinct resolution of various regions of the gel for the detection of very small proteins or the resolution of high molecular weight proteins. In addition the software offers a zooming in and out function to specifically enlarge pH-/MW-ranges of interest. Selecting a protein in the list by mouse clicking automatically tags the corresponding protein on the gel image. Alternatively clicking on a spot marks the corresponding protein in the list. These features enable the user to identify an unknown spot and to estimate the location of selected known proteins on the virtual gel. Potential membrane proteins, which are usually not accessible to routinely employed proteomics techniques, were identified by using TMHMM (5). If a whole preoteome was selected by the user, it is possible to exclude these proteins from the virtual gel by choosing the number of predicted transmembrane domains. Database links to the SWISS-PROT database (8) and PRODORIC database (9) are provided by a mouse click on single spots. If user-defined sequences are pasted or uploaded, database links are searched according to the given FASTA identifier. PRODORIC offers additional information about the regulation of the corresponding gene if a prokaryotic proteome was chosen.

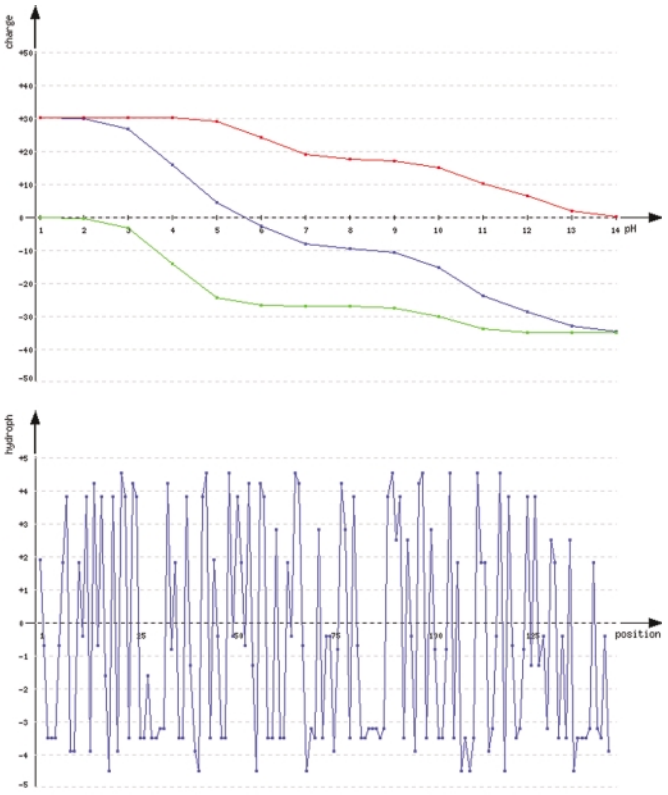

Figure 2. Screenshots of charge curves and hydrophobicity plots of the Fur (ferric uptake regulator) protein of *Escherichia coli*.

The second mode (b) offers a Java™-independent version of a virtual 2D gel, which is not dynamic and limited in function. However, it even runs in the web browser of every computer with less capacity.

In the third mode (c), MW, pI, pH-dependent charge curves and hydrophobicity profiles are calculated and displayed as a list or graphical output (Fig. 2). Depending on the sequence input, a huge number of proteins or even whole proteomes can

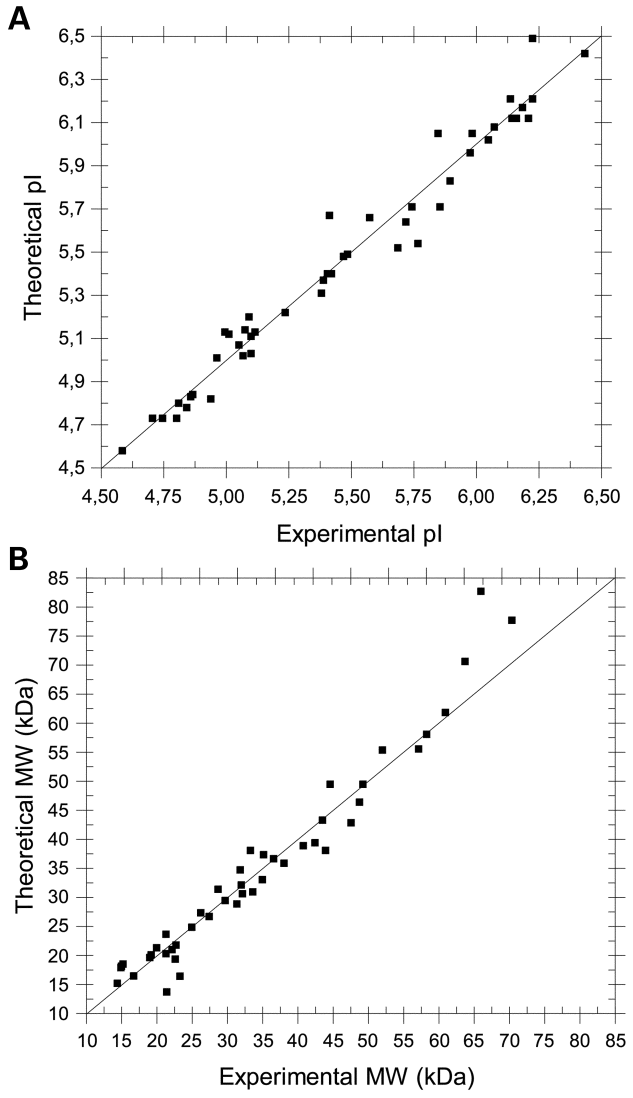

Figure 3. Comparison between calculated and experimentally obtained pI and MW values for 47 randomly selected proteins from *Pseudomonas aeruginosa.* 2D PAGE experiments were performed as described before (7). The proximity of the data points to the 45° line is a measure of agreement between the experimental and calculated values. The correlation coefficient between these values is 0.98 for both pI and MW. The identity of the employed proteins is given in the Supplementary Material.

be calculated serially. The charge curves for individual proteins offer information about pH-dependet change of charge which provides a basis for the establishment of protein purification protocols via methods like ion-exchange chromatography (10). Hydrophobicity profiles offer an overview of hydrophobic or hydrophilic properties of polypeptide chains (4), thus providing initial clues on protein characteristics such as folding, solubility or the prediction of transmembrane helices.

ACCURACY OF VIRTUAL 2D GELS

The accuracy of a virtual 2D gel relies on the algorithm used for the calculation of the theoretical MW and pI. The direct correlation with experimental data shows some system-dependent but informative limitations. The calculation of the molecular weights of denaturated proteins is quite precise unless the employed gene codes for a native protein with a signal peptide. However, the MW of a signal peptide relative to the MW of the mature protein is in most cases negligible. The calculation of the pI is based on the charged amino acid composition of a protein. The algorithm assumes that each ionizable residue is completely independent of the others. Therefore, this theoretical approach cannot take into account any protein-specific differences in the pI of charged amino acid residues resulting from the interaction with neighbouring residues. Moreover, the surface exposition of charged amino acids is not addressed by the used algorithm. Protein modifications resulting in additional charge, such as a bound phosphate group, which usually changes the overall pI of a protein are not included. Usually, the deviation in pIs for these reasons do not exceed a value of ±1 pH unit (2,3). Nevertheless, reliable pI and MW calculations are achievable as the comparison of calculated and experimentally obtained values for 47 randomly selected proteins from *P.aeruginosa* shows (Fig. 3 and Supplementary Material).

FUTURE PROSPECTS

We are planning to increase the number of preselected proteomes in the future. A stand-alone application for generating virtual 2D protein gels with extended features is currently being developed and will be published elsewhere.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Nicole Quäck and Martin Eschbach for providing experimental proteomics data. We also thank Jörgen Haneke for technical support and Dr Barbara Schulz for critical proof-reading of the manuscript. This work was funded by the German Bundesministerium für Bildung und Forschung (BMBF) for the Bioinformatics Competence Center 'Intergenomics' (Grant No.031U110A/031U210A).

REFERENCES

1. Gevaert,K. and Vandekerckhove,J. (2000) Protein identification methods in proteomics. *Electrophoresis*, **21**, 1145–1154.
2. Skoog,B. and Wichman,A. (1986) Calculation of the isoelectric points of polypeptides from the amino acid composition. *Trends Anal. Chem.*, **5**, 82–83.
3. Patrickios,C.S. and Yamasaki,E.N. (1995) Polypeptide amino acid composition and isoelectric point. *Anal. Biochem.*, **231**, 82–91.
4. Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
5. Sonnhammer,E.L.L., Heijne,G.V. and Krogh,A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Int. Conf. Intell. Syst. Mol. Biol.* AAAI Press, Montreal, Canada, pp. 175–182.

6. Pruess,M., Fleischmann,W., Kanapin,A., Karavidopoulou,Y., Kersey,P., Kriventseva,E., Mittard,V., Mulder,N., Phan,I., Servant,F. and Apweiler,R. (2003) The Proteome Analysis database: a tool for the in silico analysis of whole proteomes. *Nucleic Acids Res.*, **31**, 414–417.
7. Marino,M., Hoffmann,T., Schmid,R., Mobitz,H. and Jahn,D. (2000) Changes in protein synthesis during the adaptation of *Bacillus subtilis* to anaerobic growth conditions. *Microbiology*, **146**, 97–105.
8. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
9. Münch,R., Hiller,K., Barg,H., Heldt,D., Linz,S., Wingender,E. and Jahn,D. (2003) PRODORIC: Prokaryotic database of gene regulation. *Nucleic Acids Res.*, **31**, 278–280.
10. Wang,N.W. (1990) Ion exchange in purification. *Bioprocess. Technol.*, **9**, 359–400.