ProMoST (Protein Modification Screening Tool): a web-based tool for mapping protein modifications on two-dimensional gels

Brian D. Halligan[1],*, Victor Ruotti[1], Weihong Jin[1], Scott Laffoon[3], Simon N. Twigger[1] and Edward A. Dratz[2,3]

[1]Bioinformatics Research Center and [2]Biotechnology and Biomedical Engineering Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53213, USA and [3]Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, USA

ABSTRACT

ProMoST is a flexible web tool that calculates the effect of single or multiple posttranslational modifications (PTMs) on protein isoelectric point (pI) and molecular weight and displays the calculated patterns as two-dimensional (2D) gel images. PTMs of proteins control many biological regulatory and signaling mechanisms and 2D gel electrophoresis is able to resolve many PTM-induced isoforms, such as those due to phosphorylation, acetylation, deamination, alkylation, cysteine oxidation or tyrosine nitration. These modifications cause changes in the pI of the protein by adding, removing or changing titratable groups. Proteins differ widely in buffering capacity and pI and therefore the same PTMs may give rise to quite different patterns of pI shifts in different proteins. It is impossible by visual inspection of a pattern of spots on a gel to determine which modifications are most likely to be present. The patterns of PTM shifts for different proteins can be calculated and are often quite distinctive. The theoretical gel images produced by ProMoST can be compared to the experimental 2D gel results to implicate probable PTMs and focus efforts on more detailed study of modified proteins. ProMoST has been implemented as cgi script in Perl available on a WWW server at http://proteomics.mcw.edu/promost.

INTRODUCTION

A large fraction of proteins are modified after synthesis (post-translational modification or PTM) and these modifications may control the protein activity (1), location in the cell (2), protein lifetime (3) and protein or nucleic acid binding partners (4). Many of these PTMs can be observed as distinct protein isoform spots on two-dimensioinal (2D) gels. When proteins have been identified on 2D gels by mass spectrometry, it has been found that there are two or three protein isoforms made from each gene on average in eubacteria (5) and archae bacteria (6), and ten or more separable protein isoforms resulting from single eukaryotic genes are often found (7–9), although not all proteins show evidence for isoforms. Mass spectrometry is the most generally useful method for identifying PTMs, but antibody methods are effective if suitable reagents are available.

The most frequently occurring protein PTM appears to be phosphorylation, which is controlled by a balance of protein-specific kinase and phosphatase enzymes that are typically regulated by environmental stimuli or internal cellular programs. A detailed characterization of the sites of phosphorylation is very difficult to accomplish by mass spectrometry, and the quantitation of protein levels (10), protein phosphorylation or other PTMs by mass spectrometry requires isotopic internal standard approaches (11), although relative quantitation can be obtained with a variety of differential isotope labeling technologies (12,13). Phosphorylation replaces neutral hydroxyl groups on serines, threonines or tyrosines with negatively charged phosphates with pKs near 1.2 and 6.5. Thus, at pHs greater than about 7.5 phosphates add two negative charges to proteins; near pH 6.5 they add one-and-a-half negative charges, and below about pH 5.5 phosphates add a single negative charge. Depending on the isoelectric point (pI) of the unmodified protein and the number of other titratable groups in the protein, adding a phosphate (or any other PTM) has a smaller or larger effect on the pI of the modified protein.

The 2D gel patterns of the PTM isoforms are of great value because the relative amount of each isoform can be rapidly

*To whom correspondence should be addressed. Tel: +1 414 456 8838; Fax: +1 414 456 6595; Email: Halligan@mcw.edu

© 2004, the authors

determined from staining intensity on the gel, and it is difficult to obtain this information in any other way. Examples of gel shifts due to protein phosphorylation that have been characterized on 2D gels and by mass spectrometry, and examples of gel shifts with changes in the redox state of low pK thiols, are discussed later in this paper to illustrate these points.

The calculation of the pI for proteins and the relationship between the pI value and migration in the electro-focusing dimension has been previously explored (14,15). There are a number of programs and web services that calculate pI values for proteins (16–18), but ProMoST additionally allows the user to calculate values for proteins with a wide range of PTMs, including those entered by the investigator, and provides a graphical representation of the gel shifts due to PTMs for an unlimited number of proteins.

MATERIALS AND METHODS

Algorithms

The ProMoST web service calculates a number of different quantities for each amino acid sequence input. The amino acid composition of the sequence is calculated for internal use in determining charge and pI values, but is also available for display. The monoisotopic or average isotopic molecular mass of the protein is calculated by summing the high-precision mono or average isotopic masses of its amino acids and adding the mono or average isotopic mass of one water molecule, corresponding to an H at the N-terminal end and a OH group at the C-terminal end of the molecule. This value is used in generating the pseudo-2D gel plot as the molecular weight (MW). For PTMs that add significantly to the mass of the protein, such as phosphorylation, the mass of the modification is also considered.

The charge on the protein at a particular pH value is calculated by determining the sum of the partial charges for all the charged amino acids and modifications, using the standard equations (19):

charge of type N (negative) groups

$$C_N = \sum_{i=1}^{n} \frac{-1}{1 + 10^{pK_{N_i} - pH}}, \qquad \textbf{1}$$

charge of type P (positive) groups

$$C_P = \sum_{i=1}^{p} \frac{1}{1 + 10^{pH - pK_{P_i}}}, \qquad \textbf{2}$$

total charge

$$C_T = C_N + C_P. \qquad \textbf{3}$$

In our calculation of the charge on the protein, we use specific values for the ionizable groups on each of the amino acids. Additionally, unlike other methods that use a fixed value for the pK of the N-terminal amino and C-terminal carboxyl groups of the protein, we use specific pK values, dependent on the amino acids present at the ends. Furthermore, we also adjust the pK values for the side chains of the terminal amino acids to a specific value for that position. The charge of

proteins with PTMs is calculated in the same manner, adding the pK values for the modifications for each of the PTMs or removing the pK values for modifications that remove ionizable side chain groups. All of the pK values used in the charge calculations can be changed by the user with the 'Advanced' interface option. Additionally, the user can define additional PTMs by specifying the pK values for the modification.

The most important value calculated by the ProMoST program is the pI of the unmodified and modified versions of the protein. The pI is defined as the pH value at which the positive and negative charges on the protein are balanced and the net charge is zero. To determine the pH value at which the net charge is zero, an initial pH value of 7 is tested and the net charge on the protein calculated. Depending on the sign of the charge on the protein, a delta value of 3.5 is added or subtracted from the initial pH value of 7 and the charge on the protein recalculated at the new pH. The process of dividing the delta pH value in half and changing sign as needed is reiterated until a net protein charge of less than 0.002 is obtained. This 'binary search' method rapidly converges on an accurate value for the protein pI and is much simpler than solving the equations required in determining the pI value exactly (D.Tabb, http://fields.scripps.edu/DTASelect/20010710-pI-Algorithm.pdf).

RESULTS AND DISCUSSION

Purpose

The purpose of the ProMoST web service is to provide an easy method to examine the effect of different protein modifications on the relative position of migration of the protein on 2D (isoelectric focusing/SDS–PAGE) gels. In this way, an investigator can visually determine if the pattern of spots observed on a 2D gel corresponds to those determined by theoretical calculation for various test modifications. Furthermore, by inputting FASTA format sequence data, the effect of sequence changes due to mutations or allelic variation on predicted gel shift migration of the protein can also be examined.

The ProMoST program also allows for the flexible alteration of pK values for modified and unmodified amino acids, as well as for the addition of new protein modifications. In this way, if an investigator has reason to believe that an alternative value for the pK for a particular amino acid or modification needs to be adopted, all of the pK values are available to be changed, using the 'Advanced' interface. Additionally, it is possible to define new PTMs, in addition to the predefined PTMs. The interface allows for the examination of multiple instances of the same PTM, e.g. a protein with one, two, three, four or five phosphorylated tyrosine residues. This ability makes it easy for 2D gel users to compare the trains or pairs of spots observed on the gel with the patterns of spots predicted for different PTMs.

Audience

The prime audience of this program is biochemists and others performing 2D gel electrophoresis of proteins who are interested in protein modification, or readers of journal articles wishing to evaluate the consistency of published data with possible PTMs. This program is expected to be of significant value to the 2D gel-based proteomics community.

Overview

The ProMoST web service provides an HTML interface to a Perl-based cgi program that calculates protein pI and molecular mass values. The interface allows the user to choose the standard pK values for charged amino acids and modifications or to alter the values. The program takes protein accession numbers, names or sequence as input and produces tables of values for modified and unmodified proteins. It also has a graphic output of theoretical 2D gels.

Input

Two forms of the web interface are available to accept protein information from the user. The 'Simple' interface gives the user a choice of pasting or entering the protein information in a text box or uploading a file containing protein sequence information (Figure 1). The protein information can consist of a list of accession numbers or protein names, or protein sequences in FASTA format. The accession numbers or protein names are used by the program to obtain

Figure 1. ProMoST simple interface. The ProMoST simple interface allows the user to either paste in or upload amino acid sequence data in FASTA format or accession numbers for a single protein or group of proteins. The simple interface also allows for the user to select PTMs or define new PTM values. PTMs can be applied singly or multiply to each protein. The user can also choose various output and formatting options.

the sequence data from a local copy of the NCBI nr protein database.

In addition to the normal charged amino acids, values for common protein modifications (deamidation and phosphorylation) are also included. The user is able to specify the number of each modification that is to be considered. Thus, it is possible to examine the effects of a single phosphotyrosine or phosphoserine/phosphothreonine or a series of up to 10 phosphotyrosines or phosphoserines/phosphothreonines on the same protein molecule.

To extend these capabilities, the user can also add the name and pK values for up to three additional user-defined protein modifications and the user can choose to block either the N-, C- or both terminal ends of the protein. In addition to the protein information and the pK values, the user can also specify a wide range of output options.

The standard pK values for the charged amino acids (internal, C-terminal and N-terminal) are presented by the web interface as a series of text boxes in the Advanced interface (Figure 2). The user can thereby examine and change any of the default pK values and also has the ability to exclude any of the charged amino acids from the pI calculation, as would be required if the residue were modified to an uncharged state.

Output

The output of the program is divided into two sections: the input data and the calculated results (Figure 3). The user can opt to have the input data displayed in the form of the actual input accession number/protein name, the deduced accession number, the sequence read from the database or the composition of the protein. Any or all of these options can be active at the same time.

There are three main output modes, all of which can be used at the same time. Data can be displayed to the screen, or optionally, it can be either displayed or saved as a text file or saved as an Excel format file. The screen display takes the form of an HTML table (Figure 3, top). The user has the option to choose from different columns of data. The molecular mass choices include the monoisotopic mass, the average isotopic mass, both or neither calculated molecular mass. The protein information can be displayed as the input accession numbers the deduced accession numbers or sequence description. The calculated pI is optional. The table also shows which modifications are active for each line in the table.

Data can also be sent to either a tab-delimited text file or to an Excel format file. The files can be viewed on the screen either with the browser (text files) or with Excel (Excel files). By using the browser 'Save link as' option, the user can directly save the text or Excel file to their computer.

A graphic gel image output is available (Figure 3, bottom). The user can specify the molecular mass and pI range of the gel, as well as the physical size of the gel image. Proteins are plotted to the gel as ovals at the location of their calculated molecular mass and pI. The ovals are color coded for the modification. The parent, unmodified protein is plotted as an open oval and in the case of multiple proteins on the same plot, it is labeled with a protein index number that

Figure 2. ProMoST advanced interface. The ProMoST advanced interface allows the user to change the pK values used in the charge calculations. In addition to allowing for alternative values, this feature also allows the user to remove titratable groups, simulating what might be done by chemical modification of the protein.
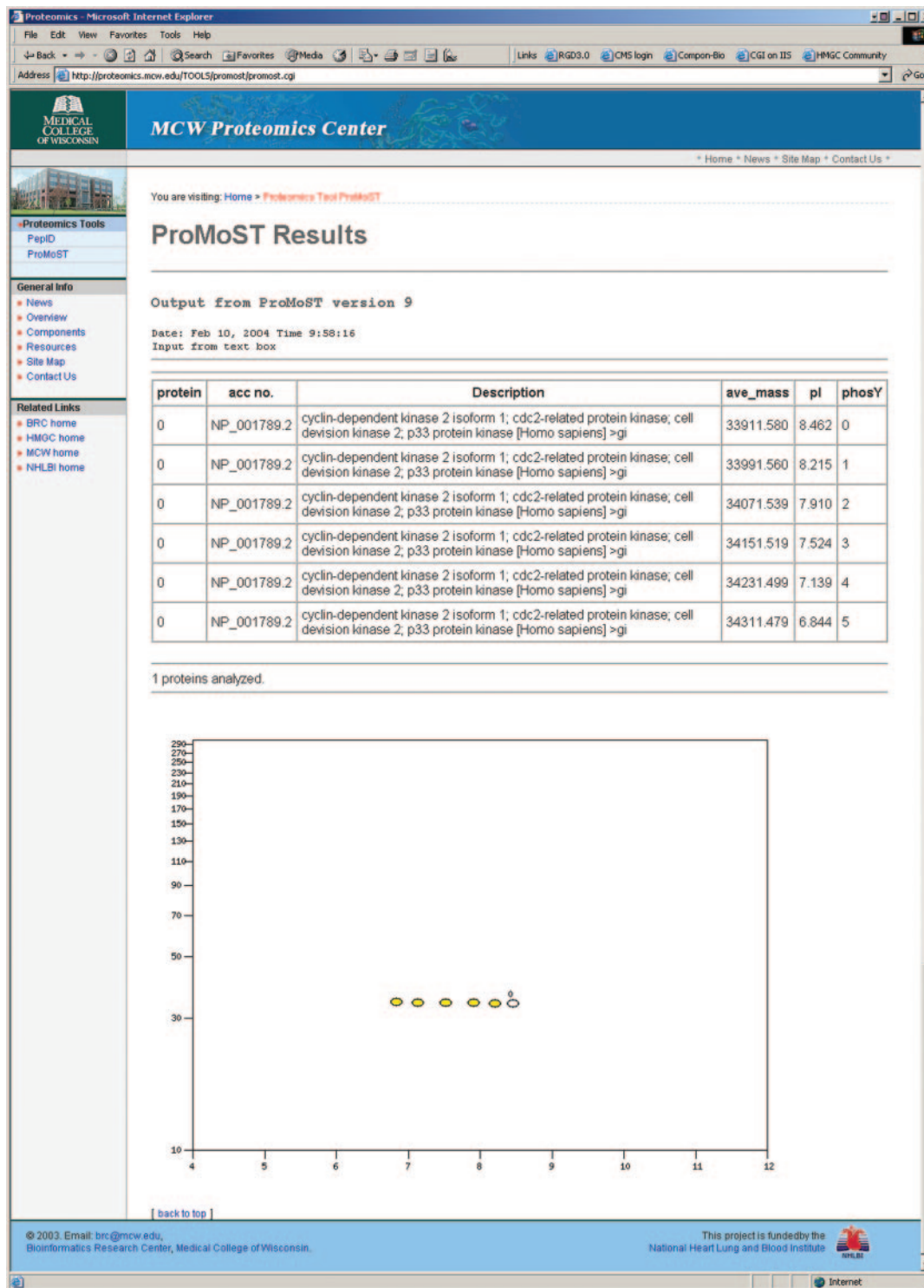
Figure 3. ProMoST results output. The human CDK2 protein is used to illustrate the tabular and graphic output from the ProMoST program. The predicted migration of the parent protein is indicated as an open oval and the predicted migration of the phosphorylated proteins with one to five phosphotyrosines are indicated by yellow filled ovals. The vertical axis is molecular mass in kDa and the horizontal axis is pI values.

matches the table or file of values. Comparison with experimental data is facilitated if the pattern of the pIs and MWs of the unmodified and modified proteins are printed out on transparencies at the same dimensions as the experimental gels.

Three examples are briefly described, where advantage has been taken of 2D gels to track the relative amounts of protein isoforms under different biological conditions. These examples suggest that this technique can be extremely powerful,

especially to enhance understanding of complex regulatory systems. Cyclin-dependent kinase 2 (CDK2) plays a central role in controlling major events in the cell cycle in complex activation and inactivation processes that depend on site-selective phosphorylation reactions. Roger *et al*. (7) were able to follow the relative stoichiometry of six different phosphorylated forms of CDK2 under different stimulus conditions, were able to implicate a previously unrecognized form and to alter the conception of the sequence of regulatory phosphorylation events. Figure 3 shows the predicted 2D gel spot profile for 0–5 phosphates added to CDK2, which agrees quite well with the experimental data. Stathmin is a phosphoprotein that is regulated by cell surface receptors and is involved in control of the cell cycle. Zugaro *et al*. (8) showed that there were at least 14 different isoforms of stathmin that can be distinguished on 2D gels. Muller *et al*. (9) were also able to isolate the different phosphorylated isoforms of stathmin from 2D gels and confirm the phosphorylation sites, using mass spectrometry. The state of stathmin phosphorylation controls microtubule stability and influences mitotic spindle assembly (20). Another example of the utility of isoelectric point shifts are low pK thiols which are being recognized as critical components in redox signaling in cells (21), and the resulting sulfenic, sulfinic or sulfonic acid forms can be detected by protein pI shifts (22,23).

Future directions

While we have focused on shifts due to PTMs, this tool will also compute and display virtual gels of complete proteomes, combinations of proteomes or sub-proteomes from a list of accession numbers similar to a recently published tool, JVirGel (24), and will be faster due to the efficient manner of pI calculation. In the current implementation, the program does not check whether or not there are the appropriate residues and motifs for each of the PTMs plotted. In the future, ProMoST could be extended to recognize potential PTM motif sites in proteins and propose which proteins might have isoforms, additionally this could also be extended to potential proteolytic processing, such as signal sequences. Once proteins are identified from 2D gels, differences between predicted and observed pI and MW could be used to flag those proteins as possible candidates for more detailed study. The development of multiplex differential fluorescent dye detection technology (25,26) is overcoming past limitations in gel-to-gel reproducibility of 2D gels and leading to expanding use for comparison of protein differences in experiment and control samples. Gel staining methods and fluorescent dyes that are specific for particular PTMs are being developed (27–30); these are expected to lead to expanded use of 2D gels for analysis of complex regulatory patterns in systems biology, and ProMoST is expected to be useful in this analysis.

ACKNOWLEDGEMENTS

REFERENCES

1. Manning,G., Plowman,G.D., Hunter,T. and Sudarsanam,S. (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.*, **27**, 514–520.
2. Chiu,V.K., Silletti,J., Dinsell,V., Wiener,H., Loukeris,K., Ou,G., Philips,M.R. and Pillinger,M.H. (2003) Carboxyl methylation of ras regulates membrane targeting and effector engagment. *J. Biol. Chem.*, **279**, 7346–7352.
3. Boatright,K.M. and Salvesen,G.S. (2003) Caspase activation. *Biochem. Soc. Symp.* 233–242.
4. Khidekel,N. and Hsieh-Wilson,L.C. (2004) A 'molecular switchboard'-covalent modifications to proteins and their impact on transcription. *Org. Biomol. Chem.*, **2**, 1–7.
5. Hecker,M. (2003) A proteomic view of cell physiology of *Bacillus subtilis*—bringing the genome sequence to life. *Adv. Biochem. Eng. Biotechnol.*, **83**, 57–92.
6. Babnigg,G. and Giometti,C.S. (2003) ProteomeWeb: a web-based interface for the display and interrogation of proteomes. *Proteomics*, **3**, 584–600.
7. Coulonval,K., Bockstaele,L., Paternot,S. and Roger,P.P. (2003) Phosphorylations of cyclin-dependent kinase 2 revisited using two-dimensional gel electrophoresis. *J. Biol. Chem.*, **278**, 52052–52060.
8. Zugaro,L.M., Reid,G.E., Ji,H., Eddes,J.S., Murphy,A.C., Burgess,A.W. and Simpson,R.J. (1998) Characterization of rat brain stathmin isoforms by two-dimensional gel electrophoresis-matrix assisted laser desorption/ionization and electrospray ionization-ion trap mass spectrometry. *Electrophoresis*, **19**, 867–876.
9. Muller,D.R., Schindler,P., Coulot,M., Voshol,H. and van Oostrum,J. (1999) Mass spectrometric characterization of stathmin isoforms separated by 2D PAGE. *J. Mass Spectrom.*, **34**, 336–345.
10. Barnidge,D.R., Dratz,E.A., Martin,T., Bonilla,L.E., Moran,L.B. and Lindall,A. (2003) Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal. Chem.*, **75**, 445–451.
11. Gerber,S.A., Rush,J., Stemman,O., Kirschner,M.W. and Gygi,S.P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci., USA*, **100**, 6940–6945.
12. Gygi,S.P., Rist,B., Griffin,T.J., Eng,J. and Aebersold,R. (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J. Proteome Res.*, **1**, 47–54.
13. Goshe,M.B. and Smith,R.D. (2003) Stable isotope-coded proteomic mass spectrometry. *Curr. Opin. Biotechnol.*, **14**, 101–109.
14. Bjellqvist,B., Hughes,G.J., Pasquali,C., Paquet,N., Ravier,F., Sanchez,J.C., Frutiger,S. and Hochstrasser,D. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023–1031.
15. Bjellqvist,B., Basse,B., Olsen,E. and Celis,J.E. (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis*, **15**, 529–539.
16. Womble,D.D. (2000) GCG: the Wisconsin Package of sequence analysis programs. *Meth. Mol. Biol.*, **132**, 3–22.
17. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
18. Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D. and Bairoch,A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784–3788.
19. Sillero,A. and Ribeiro,J.M. (1989) Isoelectric points of proteins: theoretical determination. *Anal. Biochem.*, **179**, 319–325.
20. Cassimeris,L. (2002) The oncoprotein 18/stathmin family of microtubule destabilizers. *Curr. Opin. Cell Biol.*, **14**, 18–24.
21. Poole,L.B., Karplus,P.A. and Claiborne,A. (2004) Protein sulfenic acids in redox signaling. *Annu. Rev. Pharmacol. Toxicol.*, **44**, 325–347.
22. Chevallet,M., Wagner,E., Luche,S., van Dorsselaer,A., Leize-Wagner,E. and Rabilloud,T. (2003) Regeneration of peroxiredoxins during recovery after oxidative stress: only some overoxidized peroxiredoxins can be reduced during recovery after oxidative stress. *J. Biol. Chem.*, **278**, 37146–37153.

23. Seo,M.S., Kang,S.W., Kim,K., Baines,I.C., Lee,T.H. and Rhee,S.G. (2000) Identification of a new type of mammalian peroxiredoxin that forms an intramolecular disulfide as a reaction intermediate. *J. Biol. Chem.*, **275**, 20346–20354.

24. Hiller,K., Schobert,M., Hundertmark,C., Jahn,D. and Munch,R. (2003) JVirGel: calculation of virtual two-dimensional protein gels. *Nucleic Acids Res.*, **31**, 3862–3865.

25. Tonge,R., Shaw,J., Middleton,B., Rowlinson,R., Rayner,S., Young,J., Pognan,F., Hawkins,E., Currie,I. and Davison,M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, **1**, 377–396.

26. Patton,W.F. (2002) Detection technologies in proteome analysis. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.*, **771**, 3–31.

27. Jaffrey,S.R., Erdjument-Bromage,H., Ferris,C.D., Tempst,P. and Snyder,S.H. (2001) Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. *Nat. Cell Biol.*, **3**, 193–197.

28. Aulak,K.S., Miyagi,M., Yan,L., West,K.A., Massillon,D., Crabb,J.W. and Stuehr,D.J. (2001) Proteomic method identifies proteins nitrated in vivo during inflammatory challenge. *Proc. Natl Acad. Sci., USA*, **98**, 12056–12061.

29. Patton,W.F. and Beechem,J.M. (2002) Rainbow's end: the quest for multiplexed fluorescence quantitative analysis in proteomics. *Curr. Opin. Chem. Biol.*, **6**, 63–69.

30. Steinberg,T.H., Agnew,B.J., Gee,K.R., Leung,W.Y., Goodman,T., Schulenberg,B., Hendrickson,J., Beechem,J.M., Haugland,R.P. and Patton,W.F. (2003) Global quantitative phosphoprotein analysis using multiplexed proteomics technology. *Proteomics*, **3**, 1128–1144.