

Technical report

GelMap - A novel software tool for the creation and presentation of proteome reference maps

Christina Rode¹, Michael Senkler¹, Jennifer Klodmann¹, Traud Winkelmann², Hans-Peter Braun¹

¹Institute of Plant Genetics, Leibniz Universitaet Hannover, Herrenhaeuser Str. 2, D-30419 Hannover, Germany

²Institute of Floriculture and Woody Plant Science, Leibniz Universitaet Hannover, Herrenhaeuser Str. 2, D-30419 Hannover, Germany

rode@genetik.uni-hannover.de

Phone: +049 511 762 3264

Fax: +049 511 762 3608

Keywords: Two-dimensional gel electrophoresis, public software, proteome data annotation, the plant proteomic platform

Abstract

Protein separation by two-dimensional gel electrophoresis is of central importance for proteomics. Upon combination with systematic protein identifications by mass spectrometry, large data sets are routinely generated in several proteome laboratories which can be used as „reference maps“ for future analyses of analogous biochemical fractions. Here we present GelMap, a novel software tool for the creation, presentation and evaluation of proteomic reference maps. Variable frames are introduced to group proteins into functional categories on three levels or into categories according to differential abundance during comparative proteome analyses. The software is easy to handle as it only requires uploading of two digital files. An additional file including detailed information on all proteins can be combined with the primary map. The capacity of GelMap for proteome annotation and evaluation is illustrated by two proteome projects based on protein separation by 2D IEF / SDS PAGE or 2D Blue native / SDS PAGE.

Introduction

Two-dimensional gel electrophoresis represents a basic tool in protein biochemistry. Since its invention by O'Farrel [1] and Klose in 1975 [2] it extensively is used for high resolution protein separations. Combined with broad-scale protein identification by mass spectrometry it was the starting point of the "proteomics" era in 1994 [3]. Two-dimensional gels with coupled information on the identity of the separated proteins are referred to as "reference maps" because they allow protein identification of analogous protein fractions by simple "spot pattern" comparison. From the very first beginning, software tools for online presentation of reference maps were developed. First solutions were based on simple gel images including numbered spots. On the pages of the gels, a link is offered to a table on spot identity. In some cases, the tables offer links to external databases (for example see http://www.genetik.uni-hannover.de/2d_ief.html; [4]). Later, software tools were developed which offer interactive, "clickable" gel images like the Make2D-DB II package provided by ExPASy [5] or SpotLink developed by Hajduch et al. 2005 [6]. Currently, about seventy 2D PAGE databases are listed on the WORLD-2DPAGE portal at ExPASy (<http://expasy.org/ch2d/2d-index.html>) including reference maps for various animal and plant tissues as well as for microorganisms. Most of these reference maps are realized on the basis of the approaches described above, e.g. (i) the tooth pulp 2-D reference map (http://www.biochem.oulu.fi/proteomics/teeth_gel.html); (ii) the yeast mitochondrial proteome map (http://www.biochem.oulu.fi/proteomics/ymp_gel.html) and (iii) the *Medicago* 2-D reference map (<http://www.mtproteomics.com/spot2.php?spot=92>).

Technically more advanced "clickable gels" offer an interactive surface which directly links protein spots to specific additional information provided on extra web pages, e.g.: (i) HEART-2DPAGE (<http://userpage.chemie.fu-berlin.de/~pleiss/bild1.html>), and (ii) *Arabidopsis thaliana* germination proteome (<http://seed.proteome.free.fr/map1.htm>).

Maps created using the Make2D-DB II package offer an additional pop-up window for each identified spot. Thereby, rapid access on spot identity is provided. In other cases, protein information is provided in a separate window, e.g. the human breast cancer serum project (http://www.dosac.unipa.it/cgi bin/2d/2d_view_map.cgi?map=BC_SERUM&ac=all).

Here, we present a new tool for the online annotation of 2D reference maps which for the first time allows assignment of proteins to functional classes, protein complexes or categories of differential abundance during comparative proteome projects. The non commercial software tool is easy to handle and offers application of various filters which allow classification of proteins to categories defined by the user.

Basics of GelMap

GelMap (www.gelmap.de) is a web based software tool. It therefore does not depend on the availability of a specific operation system. This offers unlimited compatibility and even allows that reference maps can be viewed on a smart phone. The creation of a reference map using GelMap is easy. Only a JPEG-version of a gel scan and a table containing x and y coordinates of all protein spots and spot specific information are necessary. By clicking on a spot, a pop-up window appears displaying spot specific characteristics including external links to databases like KEGG or UniProt (Figure 1). Using the filter options, different sets of data can be indicated, e.g. assignment of proteins to metabolic pathways like glycolysis or amino acid metabolism (Figure 2). The “physiological function” filters operate on three levels: i) the “protein name” level, ii) the “metabolic context” level and iii) a further “superior physiological context” level (see Table 1). Physiological clustering of all proteins identified on the reference map is visualised in a separate field on the right side of the map. Protein names and IDs can be searched using the "protein ID" and "protein search" filters. Users can choose whether their maps are freely-available or password protected. GelMap is a non-commercial tool, thus open for everybody.

Technical details of GelMap

The GelMap site is generated dynamically using PHP scripting language. Map creation is realised using a multipart-form wizard. From there, data is parsed and stored in MySQL (Oracle, Redwood Shores, USA) database tables. The map images are generated on-the-fly and can be saved with all currently shown spot markers. “Protein search” and “ID search” functions are incremental (search while you type) and are programmed using Ajax technology with JSON communication. Using the Apache (Apache Software Foundation, Forest Hill, USA) rewrite engine, all maps can be accessed by a short URL: [http://gelmap.de/\[ID\]](http://gelmap.de/[ID]).

The Plant Proteomics Platform (TPPP)

TPPP (gelmap.de/tppp) was established as a forum for the scientific community to use the GelMap software tool. Even though TPPP is focussed on plant proteomics, it can be used by all groups working with gel based techniques. A reference map of any tissue can be designed via GelMap and presented by a link set on a homepage independent from the TPPP framework.

Applications of GelMap on TPPP

The first 2D reference map on TPPP was published by Rode et al. 2011 [7] for embryos of the ornamental crop *Cyclamen* (<http://www.gelmap.de/cyclamen>). This map includes 247 protein

spots identified by mass spectrometry. All identified spots are labelled by circles on the reference gel. By mouse hovering over the gel protein names of the indicated spots pop-up. A pop-up window appears upon clicking on a single spot. Spot specific information is indicated like protein name, function and accession (see Figure 1). Furthermore, links to external databases are provided. Clicking "more details" in the pop-up window forwards the user to a list of identified peptides, sequence coverage as well as reliability scores (e.g. Mascot score). If a spot includes more than one protein this information also is given at this stage. Thus, all primary data can be evaluated by the external users. All 99 unique proteins identified by Rode et al. (2011) were clustered according to their physiological functions in 11 categories. The categories are listed at the right side of the reference gel in a separate field. The total number of proteins assigned to each category is indicated to provide a rapid overview on the size of a functional group. Subcategories are obtained by clicking on categories. For instance, by clicking on "carbohydrate metabolism", only the proteins involved in this process are marked on the map and two subcategories become visible, "glycolysis" and "oxidative pentose phosphate pathway". By clicking on "glycolysis", which was set to be a subcategory of "carbohydrate metabolism", only glycolytic enzymes are indicated. By clicking on "glyceraldehyde 3-phosphate dehydrogenase", a protein involved in glycolysis, only glyceraldehyde 3-phosphate dehydrogenases are labelled on the map (Figure 2). Rode et al. (2011) compared the proteomes of zygotic and somatic embryos of *Cyclamen*. By using the "regulation filter" at the right side of the map only those proteins are indicated which differ in abundance between the two compared proteomes by more than factor 1.5 or 3 (Figure 2). Both directions of comparisons are offered using either the zygotic or the somatic proteome fraction as the reference.

Presentation of further proteome projects by GelMap is under preparation. For instance, a GelMap based presentation of the mitochondrial proteome of *Arabidopsis thaliana*, which was analysed by 2D Blue native (BN) / SDS PAGE, was created (<http://www.gelmap.de/arabidopsis>). During BN / SDS PAGE, native protein complexes are resolved on the first gel dimension, which are dissected into their subunits upon electrophoresis on the second gel dimension in the presence of SDS [8]. As a result, subunits of individual protein complexes form vertical rows of spots on the 2D gel. BN / SDS PAGE therefore allows assignment of proteins to protein complexes, an information which nicely can be visualized by GelMap presentation. The 2D BN / SDS reference map of the *Arabidopsis* mitochondrial proteome currently includes 200 identified protein spots. The map gives an overview on the "protein complex proteome" of *Arabidopsis* mitochondria which

consists of the five complexes of the mitochondrial oxidative phosphorylation (OXPHOS) system and several other protein complexes. For GelMap annotation, the major categories to the right of the gel image are functional processes like “Oxidative phosphorylation” or “Citric acid cycle”. Subcategories are based on protein complexes assigned to all the main categories (Note: assignment of proteins to protein complexes was not possible in all cases). Finally, subunits of protein complexes are given on the third category level. To give an example: by clicking on the first subcategory of “oxidative phosphorylation”, which is "complex I", all subunits are marked on the gel which form part of this complex.

Further projects are under way in our laboratory and will be presented on TPPP in the short future. The TPPP-platform also is open for all external users as well. Make your own project and set a link on your homepage or display your map in a HTML inline frame.

How to create your own reference map using the GelMap software tool

The creation of a digital reference map via GelMap is easy. Two data files must be uploaded into the program: i) the image-file, which is a scan of the reference gel and ii) the coord-file, which is the core of the reference map containing coordinates of the indicated spots as well as the information appearing in the pop-up window and utilised by the filters. Optionally, a details file, containing supplementary spot information can be added. The **Image-file** must be a .jpg-file. We recommend to add molecular mass standards and a frame to the image file using an image editing program (e.g. Photoshop, Adobe Systems, San Jose, USA). The **Coord-file** must be a character separated text-document. However, we recommend to create it via an Excel-table (Microsoft) and subsequently save and upload this table as a .txt-file. The coord-file should contain at least three columns: The x- and y coordinates as well as a spot number. The coordinates should match the image-file, therefore they can be exported from a gel evaluation software (e.g. Delta2D, Greifswald, Germany). Additionally, there are further optional columns which can be used to indicate further spot characteristics. All columns available on GelMap and their description are listed in Table 1. Columns which are not necessary for your reference map are simply not considered. An example of a coord file document is given in table 2. The **Details-file** can contain any kind of spot specific information. We recommend to include peptide sequences and scores produced by mass spectrometry as well as scores occurred during protein identification. The details file must be a character separated .txt-document and contain an ID column, so the program can link this table correctly. For uploading these three files to GelMap we have designed an interface on the GelMap home page (<http://www.gelmap.de/create>). A name for the reference map and the project can be specified. Additionally, a password to view the map and a second password to

delete the map are to be set. The map will be displayed in the "view available maps" area of the GelMap home page (<http://www.gelmap.de/list>) and can be linked to the own web page. A more detailed protocol for creating a digital reference map is available on the GelMap home page in the "How To" area (<http://www.gelmap.de/howto>). Please note, that we are improving GelMap on weekly basis. Thus, changes in the uploading protocol can occur. Please check our website for any updates.

Final Conclusions

GelMap is represents a user-friendly tool offered on a professional graphical user interface for high end creation, publication and evaluation of protein reference maps. Filters can be used to assign proteins on three levels to functional categories or other criteria like differential abundance during comparative proteome projects. As a result, GelMap offers capabilities so far not possible with existing 2D reference map software tools. In our opinion GelMap will be helpful for future projects in gel-based proteomics.

Table 1 Columns available in GelMap

Column	Description
ID	The spots's ID (e.g. a running number)
X	Vertical position of the spot in gel
Y	Horizontal position of the spot in gel
TITLE	Protein name (e.g. „enolase“)
TYPE	Categorisation of the protein like physiological function (e.g. enolase + glyceraldehyde 3-phosphate dehydrogenase = “glycolysis”).
METATYPE	Meta categorisation of the physiological function (e.g. glycolysis + oxidative pentose phosphate pathway = “carbohydrate metabolism”).
ACCESSION CLASS	The protein's accession number in an external Database (e.g. "B9R9N6")
ORGANISM	EC Classification. Adds a link to the ExPASy proteomics server. e.g. "4.2.1.11"
REGULATIONZS	Scientific name of the organism. e.g. " <i>Arabidopsis thaliana</i> "
REGULATIONSZ	When comparing two tissues, you can enter the ratio here.
SIG	When comparing two tissues, you can enter the reciprocal ratio here.
MWPICALC	Indication if the spot is significantly higher abundant in one of the tissues compared (enter "significant")
ORIGIN	Calculated molecular weight and isoelectric point e.g. "24.30/8.95"
ACCSRC	Origin of the protein (e.g. tissue, subcellular fraction)
	Database for ACCESSION lookups: (left blank) = Uniprot; "N"=NCBI; "T"=Tair,

Columns labelled in pink have to be included in the coord-file, which will be uploaded to GelMap. Information added to the light grey-labelled columns can be filtered in the reference map by using the filter buttons. Information added to the blue-labelled columns will be linked to external databases.

Table 2 Example of a coord-file for uploading to GelMap

ID	X	Y	TITLE	TYPE	METATYPE	ACCESSION	CLASS	ACCSRC
1	235	234	actin 1	cytoskeleton	cellular processes	A7YVW7	-	
2	746	1502	enolase	glycolysis	carbohydrate metabolism	B9R9N6	4.2.1.11	
3	279	341	glyceraldehyde 3-phosphate dehydrogenase	glycolysis	carbohydrate metabolism	B9RAL0	1.2.1.12	
4	184	605	anti-oxidant 1	uncharacterized	uncharacterized	AT1G66240	-	T

The headings of this table are explained in table 1.

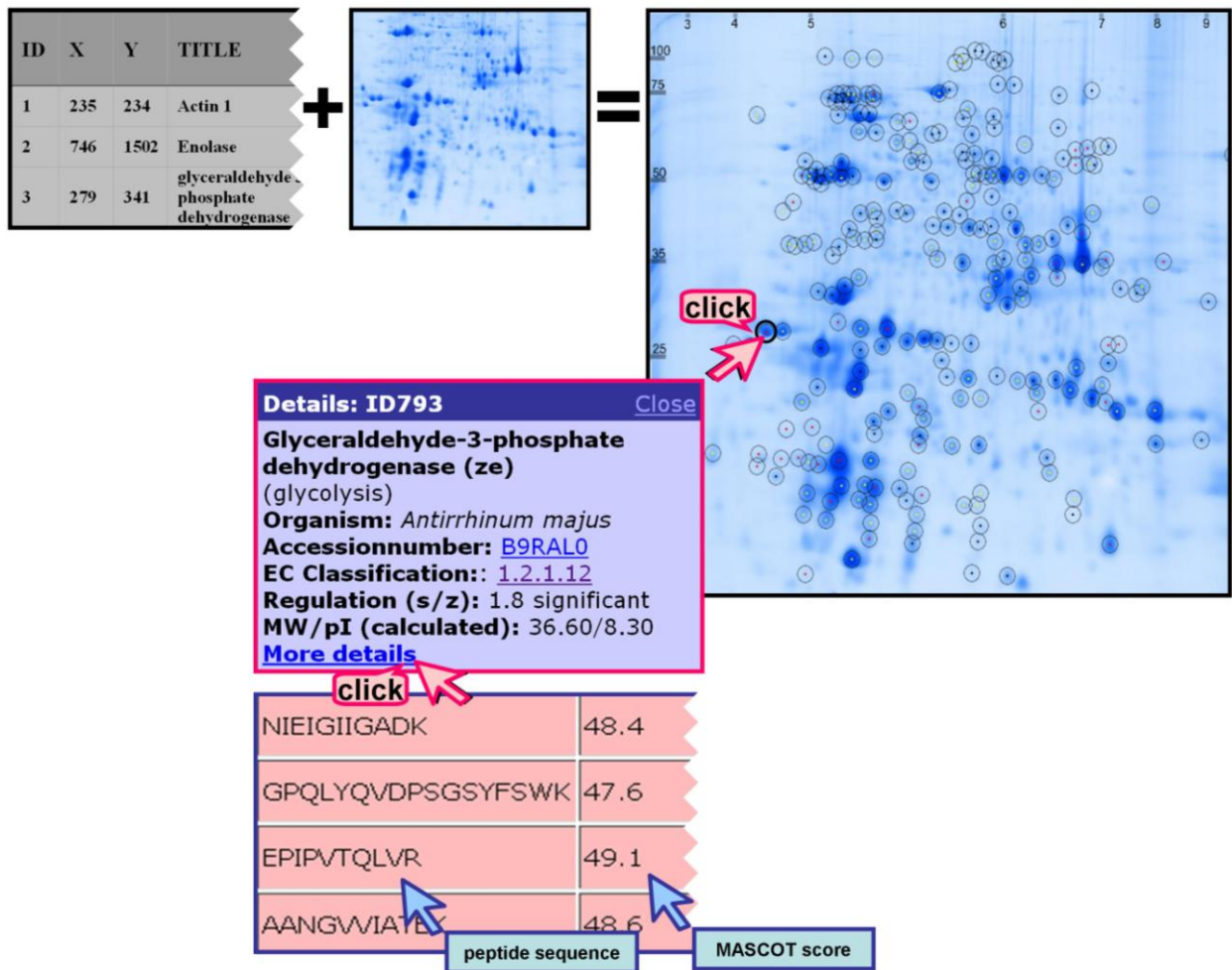


Figure 1: Creation and use of a digital proteome reference map provided by GelMap

After uploading a coord file (grey box at the upper left side of the figure) and an image-file of the reference gel (small gel at the upper left side of the figure) the digital reference map is created on-the-fly by the GelMap software. A screen shot of a reference map is shown (big gel at the right side of the figure). By clicking on a labelled spot, a pop-up window appears, displaying spot specific characteristics including external links to databases. By clicking the "More detail" link in the pop-up window the supplementary data can be explored, like demonstrated here for peptide sequences and the corresponding MASCOT scores of the protein of interest.

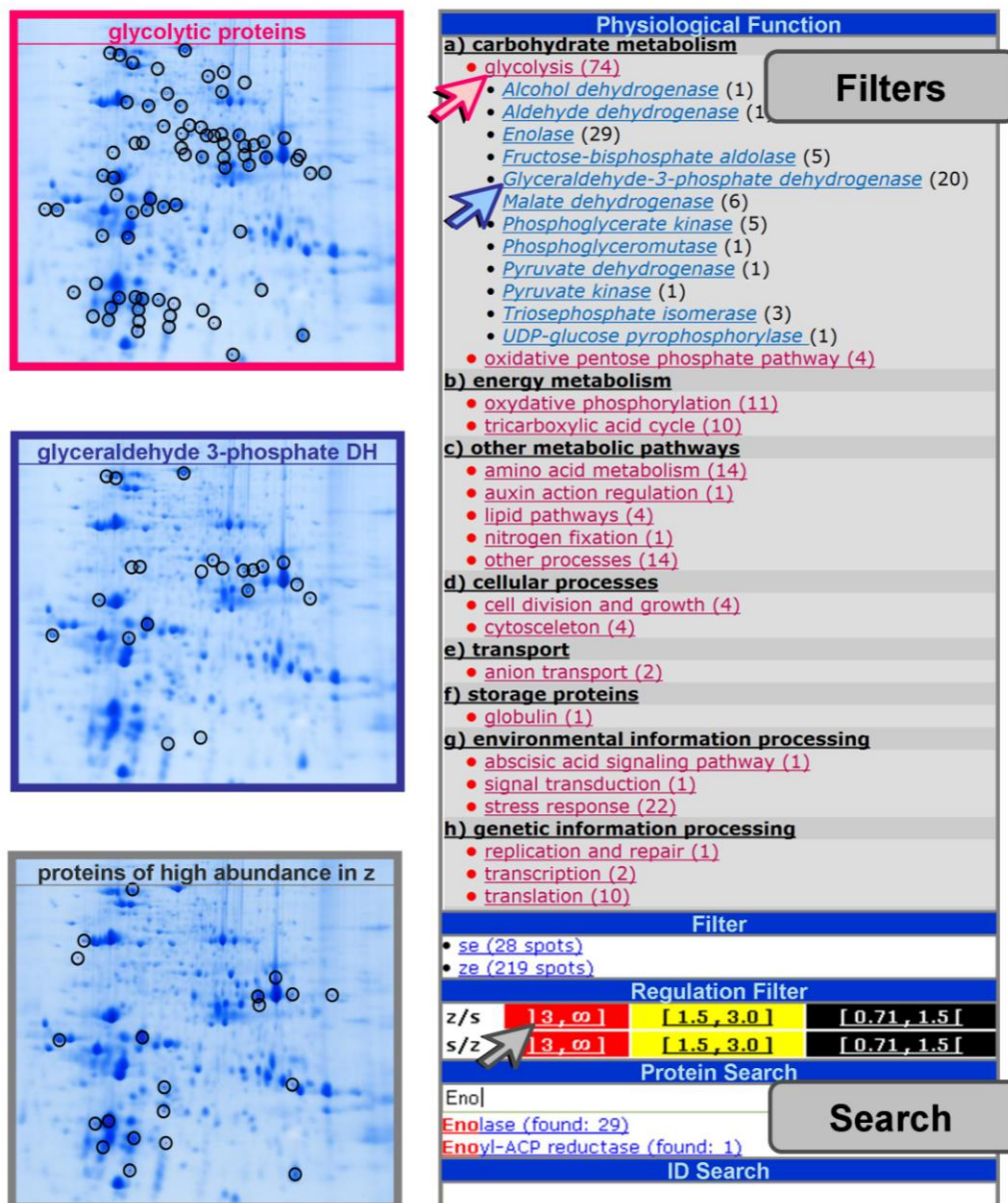


Figure 2: Illustration of the filter functions of GelMap

The Filter Window: At the right side of the figure the filter window of a digital proteome reference map designed via GelMap is shown. All proteins identified in the reference gel were clustered according to their physiological functions on three levels: i) protein name (blue letters, only visible when clicking on the superior category); ii) metabolic context (pink letters) and iii) superior physiological context (black letters, a-h). Using the filter "Physiological function": By clicking on "glycolysis" (pink arrow) all enzymes involved in this pathway which were identified in the gel become visible within this category in the filter window and labelled by black circles in the reference gel (gel framed in pink). By clicking on "glyceraldehyde 3-phosphate dehydrogenase" (blue arrow), only glyceraldehyde 3-phosphate dehydrogenase (glyceraldehyde 3-phosphate DH) are shown on the map (gel framed in blue). By using the "Regulation filter" only proteins are selected, which e.g. are more than three fold higher abundant in the zygotic embryos proteome (grey arrow). These spots will be labelled in the reference gel (gel framed in grey). "Protein search" and "ID search" functions are incremental (search while you type) and are located at the bottom part of the filter window. By typing e.g. "Eno" in the protein search field all available enzymes including this letter combination appear below. After clicking on the protein of interest it will be marked on the gel as by red arrow head. Abbreviations: DH = dehydrogenase; s = somatic embryo; z = zygotic embryo

References

- [1] O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975;250:4007-4021.
- [2] Klose J. Protein mapping by combined isoelectric focusing and electrophoresis in mouse tissues. A novel approach to testing for induced point mutations in mammals.; *Human Genetik* 1975;26:231-243.
- [3] Wilkins M. Proteomics data mining. *Exp Rev Prot.* 2009; 6:599-603
- [4] Krufft V, Eubel H, Jänsch L, Werhahn W, Braun HP. Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol* 2001;127:1694-1710.
- [5] Mostaguir K, Hoogland C, Binz PR, Appel RD. The Make 2D-DB II package: Conversion of federated two-dimensional gel electrophoresis databases into a relational format and interconnection of distributed databases. *Proteomics* 2003;3:1441-1444
- [6] Hajduch M, Ganapathy A, Stein JW, Thelen JJ. A systematic proteomic study of seed-filling in soybean: establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiol* 2005;137:1397-1419
- [7] Rode C, Gallien S, Heintz D, Van Dorsselaer A, Braun HP, Winkelmann T. Enolases - storage compounds in seeds? Evidence from a proteomic comparison of zygotic and somatic embryos of *Cyclamen persicum* Mill. *Plant Mol Biol* 2011; 75:305-319
- [8] Wittig I, Braun HP and Schägger H. Blue-Native PAGE. *NATURE Protocols* 2006;1:418-428.

Internet Databases

ExPASy: <http://expasy.org/>

WORLD-2DPAGE List presented by ExPASy: <http://expasy.org/ch2d/2d-index.html>

All named web pages were visited last on march 2011.

Acknowledgments

The authors would like to thank the DFG (German research foundation, project number WI2002/3-1) for financial support.

Figure 1

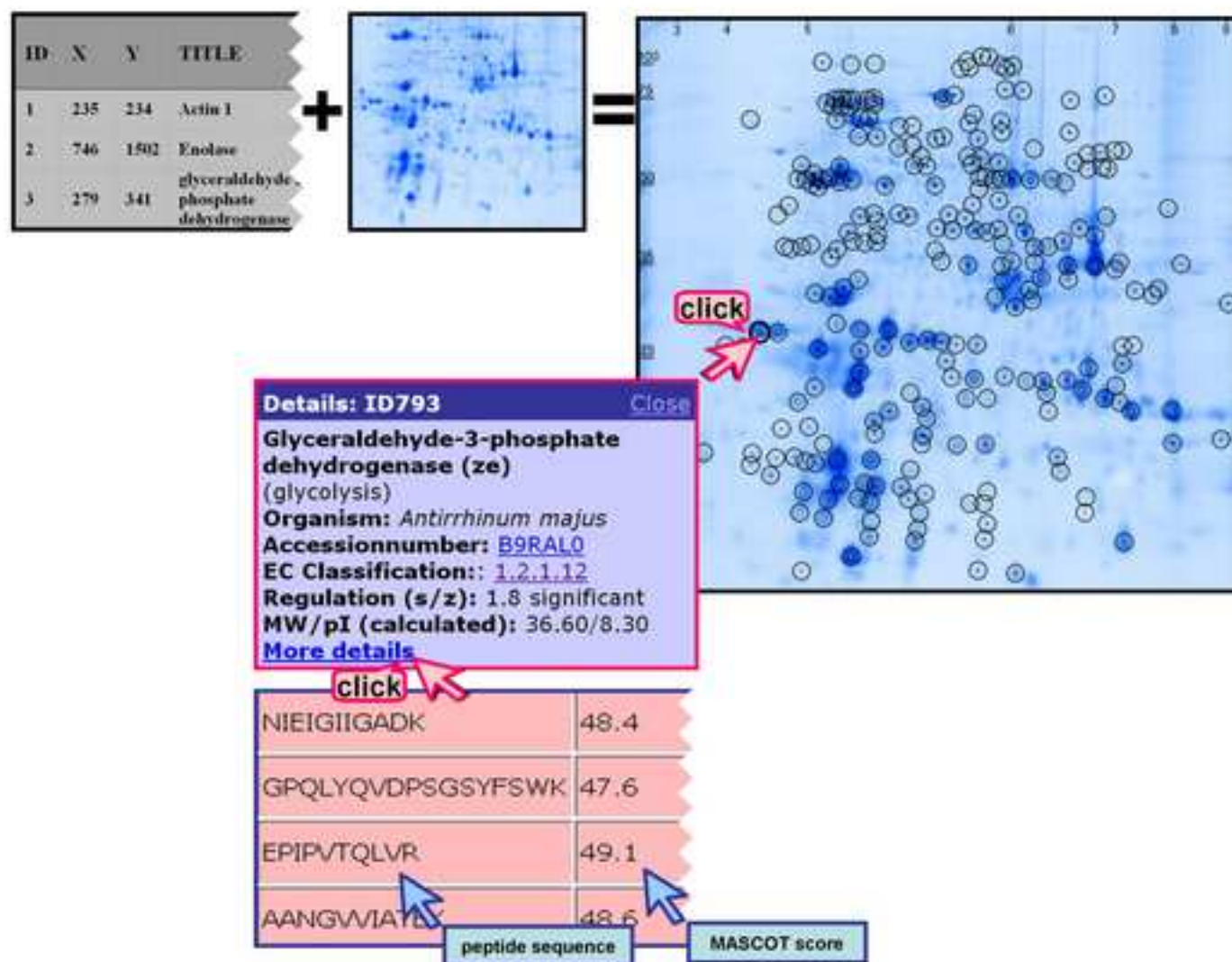


Figure 2

