MIAPEGelDB, a web-based submission tool and public repository for MIAPE gel electrophoresis documents

Xavier Robin[a, b], Christine Hoogland[✉, a], Ron D. Appel[a, c] and Frédérique Lisacek[a]

[a]Proteome Informatics Group, Swiss Institute of Bioinformatics, Geneva, Switzerland
[b]Biomedical Proteomics Research Group, Department of Structural Biology and Bioinformatics, University of Geneva, Switzerland
[c]Computer Science Department, University of Geneva, Switzerland

July 2008

Abstract

The HUPO Proteomics Standards Initiative (PSI) defines standards for data representation in proteomics to facilitate data exchange and comparison, and quality assessment. A set of minimum reporting requirements, called MIAPE (for Minimum Information About a Proteomics Experiment) is provided to ensure consistency of data set annotation. Like the MIAME reporting requirements for transcriptomics, it is anticipated that journal editors will soon require such annotation for published data sets, simplifying further mining of data. Therefore, tools for data entry and public repositories for long-term storage will be needed.

MIAPEGelDB is a public repository and a web-based data entry tool for documents conforming to the MIAPE gel electrophoresis guidelines. It aims to guide authors through the publication of the minimal set of information for their proteomics experiments using a clear, sequential interface. After publication by their author, documents in MIAPEGelDB can be viewed in HTML or plain text formats, and further used through stable URL links from remote resources. MIAPEGelDB is accessible at: miapegeldb.expasy.org.

Citation: Robin X., Hoogland C., Appel R. D. and Lisacek F. (2008). MIAPEGelDB, a web-based submission tool and public repository for MIAPE gel electrophoresis documents. *Journal of Proteomics* 71 (2) p. 249–251. PMID: 18590991. DOI: 10.1016/j.jprot.2008.06.005.

Abbreviations: GelML: Gel Markup Language; MIAME: Minimum Information About a Microarray Experiment; MIAPE: Minimum Information About a Proteomics Experiment; PSI: Proteomics Standards Initiative; XML: Extended Markup Language.

Keywords: Bioinformatics, Proteomics Standards Initiative, MIAPE guidelines, Gel electrophoresis, Public repository, Web services.

Submitted: January 15th, 2008; Revised: June 6th, 2008; Accepted: June 9th, 2008.

Introduction

For several years, the HUPO Proteomics Standards Initiative (PSI) has been developing standards for data representation and reporting guidelines for proteomics (Taylor *et al.*, 2006). They consist of MIAPE modules (Taylor *et al.*, 2007) specifying the minimum information that must be reported together with the experimental data, controlled vocabularies to describe data and protocols, and finally standard file formats to encode data and metadata. Meanwhile, several journals have defined their own internal guidelines to address the lack of such finalised public guidelines (Wilkins *et al.*, 2006, Bradshaw *et al.*, 2006; Celis, 2004). Unfortunately, multiple guidelines lead to inconsistent annotation of data sets. As a consequence, computational methods cannot be efficiently applied across collection of data sets published in different journals.

Assuming that the MIAPE standard will ultimately prevail, it remains a challenge for researchers who are willing to submit compliant data. Information required by MIAPE is certainly quite comprehensive but its proper submission is tedious and time-consuming. Directions appear complex, and since not all

✉ Corresponding author. Swiss Institute of Bioinformatics, CMU, 1 Michel-Servet, CH-1211 Geneva 4, Switzerland. Tel.: +41 22 379 5828; fax: +41 22 379 5858; e-mail address: Christine.Hoogland@isb-sib.ch

software tools handle MIAPE-compliant annotation (Taylor, 2006), the inclusion of the required information in a document is too often considered laborious by researchers. Besides, researchers also need a public location to make this data available (Prince *et al.*, 2004).

Currently, no proteomic database supports MIAPE annotation. In the field of gel electrophoresis most databases do not even publish used protocols and/or references, with the notable exception of the well-known SWISS-2DPAGE database (Hoogland *et al.*, 2004). Since MIAPE guidelines are likely to replace existing journal-defined guidelines when they will have reached a state acceptable by the proteomics community, then software tools and public repositories will have to be set to handle such data annotation.

The MIAPE gel electrophoresis module defines reporting guidelines for n-dimensional gel electrophoresis. "The reporting guidelines cover gel manufacture and preparation, running conditions, visualization techniques such as staining, the method of image capture and a technical description of the image obtained. They do not explicitly cover sample preparation, but do require the recording of which samples were loaded onto a gel. They do not include spot detection or other analyses of gel images, nor do they include protein identification procedures" (Gibson *et al.*, 2008).

To address these needs, we built MIAPEGelDB, a web-based data entry tool that eases and speeds up submission and includes a public repository for documents conforming to the MIAPE gel electrophoresis guidelines (v. 1.4 of January 10, 2008). MIAPEGelDB aims to guide authors through the publication of the minimal set of information for their gel electrophoresis experiments using a clear, sequential web-based interface.

Material and Methods

We used the Catalyst perl web framework (catalystframework.org) to build the web interface. Submitted data are stored in a PostgreSQL relational database (postgresql.org). The scheme of the database is not detailed here and can be found at miapegeldb.expasy.org/technical. Pages are served by an Apache server using the mod_perl module. Version 1.4 (10th January, 2008) of the MIAPE gel electrophoresis module submitted to *Nature Biotechnology* (Gibson *et al.*, 2008) and the SWISS-2DPAGE protocols (expasy.org/ch2d/protocols) were applied to build and test the application.

Results

Documents in MIAPEGelDB are grouped by 'Experiment', which allows associating related gels according to specific needs (standard protocol, time course, comparative analysis, etc.). The edition of a MIAPE gel electrophoresis document typically begins with the creation of such an 'Experiment', providing a name and a description that briefly describe the context of the experiment. This information can be searched later on as text. Then the gel electrophoresis protocol data entry actually starts.

The interface was designed to guide authors sequentially through the whole process of data submission. Related information is grouped together on separate pages, corresponding more or less to each section of the MIAPE document. When a submission page is validated, the author is automatically transferred to the next one. The flow of successive pages depends on information previously entered, thus it can be different for specific workflows such as 1-D or 2-D gels. In each case, the input sequence proposed by the tool is based on the most common workflow. However, any step of the data submission can be reached individually from the document parent page, where the status of each step is specified with a small icon (done or to do) (Figure 1C). Data submission is saved at each step allowing data entry to be postponed to a later time. During the editing process the document is only accessible to its author through password protection; it can thus be shared with colleagues before publication.

When a document is completed, its author makes it public, that is, accessible to anyone with no user restriction. In any case, documents for each experiment or gel are available through a unique URL, which will remain the same irrespective of a private or public status. When similar experiments or protocols are performed, data submission can be speeded up by duplicating yet existing document pages, which then serve as templates.

Even though MIAPEGelDB performs syntax checks (such as units or values conformity) data consistency cannot be extensively verified, in particular for text content. Therefore, it should be noted that doc-

uments produced and stored in MIAPEGelDB may not be fully MIAPE-compliant. Similarly, the quality assessment of data annotation in MIAPEGelDB is left to readers, reviewers and journal editors. This issue was also raised in the ArrayExpress database for RNA microarrays, where several data sets were found to be insufficiently annotated (Brazma & Parkinson, 2006) and were not complying with MIAME guidelines (Brazma *et al.*, 2001). In order to minimise this shortcoming we introduced controlled vocabularies whenever available (for example for measurement units, typical equipments or compounds). The use of pre-defined terms is meant to help end-users in remaining consistent and possibly prevent typing errors.

Finally, anyone (i.e. unregistered or not logged-in users) can view all published documents (Figure 1A). A search box is available on each page of MIAPEGelDB and for an input term, a full-text search is performed over gel documents and experiment names and description. An advanced search page is also available, which allows refining the search to specific fields of experiments or documents. Results are then listed using the name and description defined by the author.

Documents stored in MIAPEGelDB can be used for cross-linking with a proteomic database or a LIMS, or for further reference within an article. Text documents can be parsed and mined to extract information. The search function can be used for instance to retrieve gels performed using the same sample or experimental techniques. Documents are produced in plain text and HTML formats (Figure 1B). An XML format will be offered shortly, as the Gel Markup Language (GelML) specification is now finalised (psidev.info/index.php?q=node/254). Further mid-term developments will also allow uploading documents either in XML or plain text format.

Obviously, MIAPEGelDB will need to evolve in parallel with the MIAPE guidelines. The update mechanism is not yet chosen, but we have already foreseen three different possibilities:

- Lock all existing documents in a read-only state corresponding to the MIAPE version in use when they were created, then update database schema and interface according to the new version of the MIAPE document.

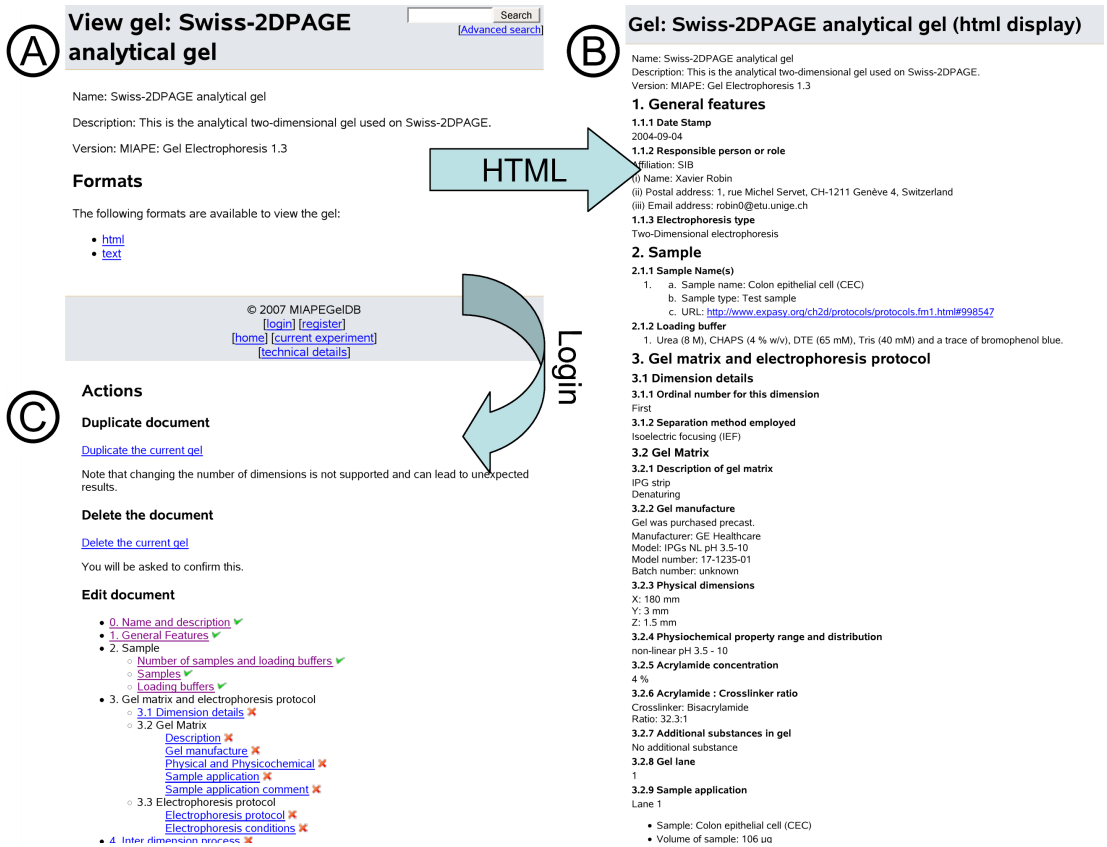

Figure 1: Workflow of MIAPEGelDB. (A) Following navigation, search, or an external link from elsewhere such as SWISS-2D-PAGE, gel information can be viewed in HTML (B) or plain text format (not shown). Through the web interface, registered users access further actions such as duplicating, deleting or editing their own documents. (C) A summary shows which parts of the document are already completed and which still need to be filled-in. Authors can use it to continue data entering from where it was last stopped.

- Duplicate part or whole of the application and database schema, so as to have one separate copy for each version of the MIAPE document; both versions would coexist independently. Obviously this duplication would be transparent to users.
- Update the database, adding or removing tables according to the changes in the MIAPE guidelines, then upgrade existing documents to the new version.

The mechanism will be selected upon the estimated impact of the modifications in the MIAPE guidelines.

Finally MIAPEGelDB should be easily extended to other MIAPE modules such as Gel [image] Informatics or Sample preparation and handling, when they are ready.

Conclusion

MIAPEGelDB is accessible at miapegeldb.expasy.org. It should support the dissemination of MIAPE guidelines while encouraging users to generate MIAPE-compliant data. However, standards will not be adopted without strong incentive measures from journal editors and/or funding organisations (Taylor, 2006).

The definition of similar standards for microarrays (MIAME (Brazma *et al.*, 2001)) and the subsequent setting up of public repositories have made mining of published data possible and led to discoveries (see references in (Prince *et al.*, 2004)). Such data mining could be performed on proteomic data were they to be consistently annotated.

Acknowledgements

This work is partially supported by the ProDaC (Proteomics Data Collection) Coordinated Action funded by the European Commission (more information at www.fp6-prodac.eu). The authors would like to thank Jean-Charles Sanchez, Alirezah Vazzadeh and Pierre-Alain Binz for their feedback about the application, Laurent Dami, Khaled Mostaguir and people from the Swiss Institute for Bioinformatics for technical advice. We are grateful to the anonymous reviewers for their valuable suggestions to improve the paper.

References

Bradshaw R. A., Burlingame A. L., Carr S., and Aebersold R. (2006). Reporting protein identification data: The next generation of guidelines. *Molecular & Cellular Proteomics* 5 (5) p. 787–788. PMID: 16670253. DOI: 10.1074/mcp.E600005-MCP200.

Brazma A., Hingamp P., Quackenbush J., *et. al.* (2001). Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nature Genetics* 29 (4) p. 365–371. PMID: 11726920. DOI: 10.1038/ng1201-365.

Brazma A. and Parkinson H. (2006). ArrayExpress service for reviewers/editors of DNA microarray papers. *Nature Biotechnology* 24 (11) p. 1321–1322. PMID: 17093465. DOI: 10.1038/nbt1106-1321.

Celis J. E. (2004). Gel-based proteomics: What does MCP expect? *Molecular & Cellular Proteomics* 3 (10) p. 949. PMID: 15475425.

Gibson F., Anderson L., Babnigg G., *et. al.* (2008). Guidelines for reporting the use of gel electrophoresis in proteomics. *Nature Biotechnology* 26 (8) p. 863–864. PMID: 18688234. DOI: 10.1038/nbt0808-863.

Hoogland C., Mostaguir K., Sanchez J.-C., *et al.* (2004). SWISS-2DPAGE, ten years later. *Proteomics* 4 (8) p. 2352–2356. PMID: 15274128. DOI: 10.1002/pmic.200300830.

Prince J. T., Carlson M. W., Wang R., et al. (2004). The need for a public proteomics repository. *Nature Biotechnology* 22 (4) p. 471–472. PMID: 15085804. DOI: 10.1038/nbt0404-471.

Taylor C. F. (2006). Minimum reporting requirements for proteomics: A MIAPE primer. Proteomics 6 (S2) p. 39–44. PMID: 17031795. DOI: 10.1002/pmic.200600549.

Taylor C. F., Hermjakob H., Julian R. K. Jr., *et al.* (2006). The work of the human proteome organisation's proteomics standards initiative (HUPO PSI). *OMICS: A Journal of Integrative Biology* 10 (2) p. 145–151. PMID: 16901219. DOI: 10.1089/omi.2006.10.145.

Taylor C. F., Paton N. W., Lilley K. S., *et. al.* (2007). The minimum information about a proteomics experiment (MIAPE). *Nature Biotechnology* 25 (8) p. 887–893. PMID: 17687369. DOI: 10.1038/nbt1329.

Wilkins M. R., Appel R. D., Van Eyk J. E., *et. al.* (2006). Guidelines for the next 10 years of proteomics. *Proteomics* 6 (1) p. 4–8. PMID: 16400714. DOI: 10.1002/pmic.200500856.