 DOI 10.1002/pmic.200900023 *Proteomics* 2009, *9*, 4664–4668

TECHNICAL BRIEF

Proteopathogen, a protein database for studying *Candida albicans* – host interaction

Vital Vialás[1], Rubén Nogales-Cadenas[2], César Nombela[1], Alberto Pascual-Montano[2] and Concha Gil[1,3]

[1] Departamento de Microbiología II, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid, Spain
[2] Departamento de Arquitectura de Computadores y Automática, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Madrid, Spain
[3] Unidad de Proteómica UCM-Parque Científico de Madrid, Facultad de Farmacia, Universidad, Complutense de Madrid, Madrid, Spain

There exist, at present, public web repositories for management and storage of proteomic data and also fungi-specific databases. None of them, however, is focused to the specific research area of fungal pathogens and their interactions with the host, and contains proteomics experimental data. In this context, we present Proteopathogen, a database intended to compile proteomics experimental data and to facilitate storage and access to a range of data which spans proteomics workflows from description of the experimental approaches leading to sample preparation to MS settings and peptides supporting protein identification. Proteopathogen is currently focused on *Candida albicans* and its interaction with macrophages; however, data from experiments concerning different pathogenic fungi species and other mammalian cells may also be found suitable for inclusion into the database. Proteopathogen is publicly available at http://proteopathogen.dacya.ucm.es

Received: January 13, 2009
Revised: June 25, 2009
Accepted: July 2, 2009

Keywords:
Candida albicans / Database / Host / MS / Microbiology / Pathogen

Candida albicans is an opportunistic pathogenic fungus, which can be found as a component of the usual flora in human mucoses. Although it does not normally cause disease in immunocompetent colonized hosts, in the case of immunosuppressed patients *Candida* cells can overproliferate and become pathogenic. Cells in yeast form (oval cells) may produce hyphae, penetrate tissues and eventually cause invasive candidiasis. At present, the frequency of this fatal opportunistic mycosis continues to be distressing and, unfortunately, solution is hindered by the reduced effectiveness and serious side effects of the few available drugs, the appearance of antifungal-drug resistance, and the lack of accurate and prompt diagnostic procedures [1].

Addressing proteomic studies involving the way *Candida* interacts with immune cells is thus essential in order to improve our comprehension of the process of infection and represents the primary step of investigation that could lead to future development of diagnosis methods, vaccines and antifungal drugs [2–5].

Experimental techniques in proteomics have quickly evolved in such a way that nowadays we have to deal with vast amounts of complex data originated by the combination of multi-dimensional separation techniques and MS analysis together with the bioinformatics software reports [6]. Existing public repositories for management and storage of proteomic data such as World 2-D PAGE [7], the Proteome Database System for Microbial Research 2-D PAGE [8], or PRIDE [9]; and fungi-specific databases such as BioBase MycoPathPD [10], Candida Genome Database (CGD) [11] or Candida DB [12] are very popular and useful tools. However, none of them deals with proteomic experimental

Correspondence: Dr. Concha Gil, Departamento de Microbiología II, Facultad, de Farmacia, Universidad Complutense, Plaza de Ramón y Cajal s/n, 28040 Madrid, Spain
E-mail: conchagil@farm.ucm.es
Fax: +34-913941745

Abbreviations: CGD, Candida Genome Database; **GO,** *Gene Ontology*; **KEGG,** Kyoto Encyclopedia of Genes and Genomes; **PDB,** Protein Data Bank

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

data related to the specific research area of fungal pathogens and their interaction with the host. In this context, we present Proteopathogen, a protein database, currently focused on the *C. albicans* – macrophage interaction model – which enables a framework for the access and submission of proteomic workflow data, from description of the experimental approaches leading to sample preparation to MS settings and identification-supporting peptides. Through its interface web site, the database can easily be queried to allow an efficient browsing through all the stored data, improving the quality of eventual analysis of MS results.

Regarding the compilation of information used to populate the database, data from three different studies were considered suitable to be present in Proteopathogen. The first two correspond to publish works relating to proteomics of the *Candida* – macrophage interaction [2, 3], where the former reports 66 different *C. albicans* identified proteins and the latter, 38 murine macrophage proteins. The third study represents an analysis of the *C. albicans* plasma membrane proteome [13]. It compiles a set of experiments aimed at extraction and identification of membrane proteins and a set of experiments intended to obtain enrichment in glycosylphosphatidylinositol-anchored surface proteins, which have been reported to be involved in cell wall biogenesis, cell–cell adhesion and interaction with the host [14].

In all cases, protein identifications lists are collected together with the pertinent experimental context specified by descriptions of the experimental approaches, MS settings and peptides supporting identification for each of the proteins (Table 1).

Along with the experimental information, and in order to provide a deeper view of the data, complementary information is retrieved from public web repositories. In the case of *C. albicans* proteins, identifiers, synonyms, aminoacid sequence of the translated open reading frame, *Saccharomyces cerevisiae* orthologs, *Gene Ontology* (GO) annotation, pathway annotations and scientific literature references were obtained from CGD [11], whereas in the case of murine macrophage proteins, the equivalent information was obtained from UniProt KnowledgeBase [15] and the Mouse Genome Database [16]. Additionally, pathways annotations were retrieved from Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database [17] and structure information from the Protein Data Bank (PDB) [18].

Concerning the architecture of the software, the back-end layer consists of a MySQL database managed by the web application development framework Ruby on Rails that sets up structure and relations of data, handles queries to the database and displays the user web-based interface.

The experimental context is addressed in Proteopathogen in a hierarchical manner, where a main general approach, which may correspond to a published article, is characterized by a description or title, authors, target species and Pubmed identifier when available; and experiments within it, are in they turn, characterized by the description of the particular experiment, the date when it was performed and number of identified proteins.

Information on one particular protein is split into several sections in Proteopathogen. *Protein Basic Information* displays the UniProt accession number, description, species, evidence for the existence, standard gene name, organism-specific database identifiers, yeast orthologs for *Candida* proteins and human orthologs for mouse proteins and sequence. The Section 2 lists experiments in which the particular protein has been identified. Where available, one or more of the following sections will be displayed as well: the table entitled GO showing GO annotations along with the pertinent scientific references, the *KEGG Pathways* and *CGD Pathways* tables rendering annotations from KEGG and CGD respectively, and *PDB*, a table specifying structural information. Where no PDB identifiers are found for *C. albicans* proteins, *S. cerevisiae* orthologs are used instead, and similarly, when a PDB identifier cannot be found for mouse proteins, the human ortholog is used.

In all cases, proteins are unambiguously related to their corresponding experiment, thus enabling a relation to the data concerning experimental parameters of identification and identification-supporting peptides. This data comprise, on the one hand, common MS settings for all proteins identified in the particular experiment, including search database, MS type, analysis software, digestion enzyme, fixed aminoacid modifications, variable modifications and maximum allowed number of miscleavages; and on the other hand, particular parameters and peptides list for each protein, including number of matched peptides, score,

Table 1. Overview of the stored data in Proteopathogen as well as their published evidences

References	Description of experimental approach	Species	♯Protein identifications
[2]	*C. albicans* differentially expressed proteins after 3 h interaction with RAW 264.7 murine macrophages. 2-D silver-stained gel. MS/MS (MALDI/TOF-TOF)	*C. albicans*	66
[3]	Proteins identified from cytoplasmic extracts of RAW 264.7 cells after 45 min interaction with *C. albicans*	*Mus. musculus*	38
[13]	Identification of Glycosyl phosphatidil inositol (GPI)-anchored membrane proteins Identification of membrane proteins	*C. albicans*	292 1273

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

observed peptide mass, calculated peptide mass, start and end coordinates, number of missed cleavages and the sequence of the peptide.

The web interface to Proteopathogen offers multiple ways to query the database. Through the *Browse Experiments* search option, a list containing all sets of experimental approaches is displayed. In its turn, one particular experiment can be browsed through all the proteins identified in it.

The *Search* form may be used in different manners. Queries for one particular protein can be performed by supplying one of the multiple supported identifiers, namely standard gene names, Candida feature name, Candida DB identifiers, CGD identifiers, MGI identifiers and UniProt accession numbers. Free text queries can be performed as well, which will retrieve a list of proteins showing coincidences in the description field of the Proteopathogen protein entry. As an additional feature, peptide sequences can also be searched for retrieving in this case, proteins in any experiment having the searched sequence in any of the identification-supporting peptides. Wild characters (''*'') and Boolean operators are supported for free text queries and for peptide sequence queries.

In order to enhance interactivity and collaboration with users, a submission form is included in the web interface to allow the upload of more proteomic experimental approaches as long as they concern the topics addressed in Proteopathogen. Sequential steps request from the user the following information: a description of the experimental context, a related protein list, MS parameters

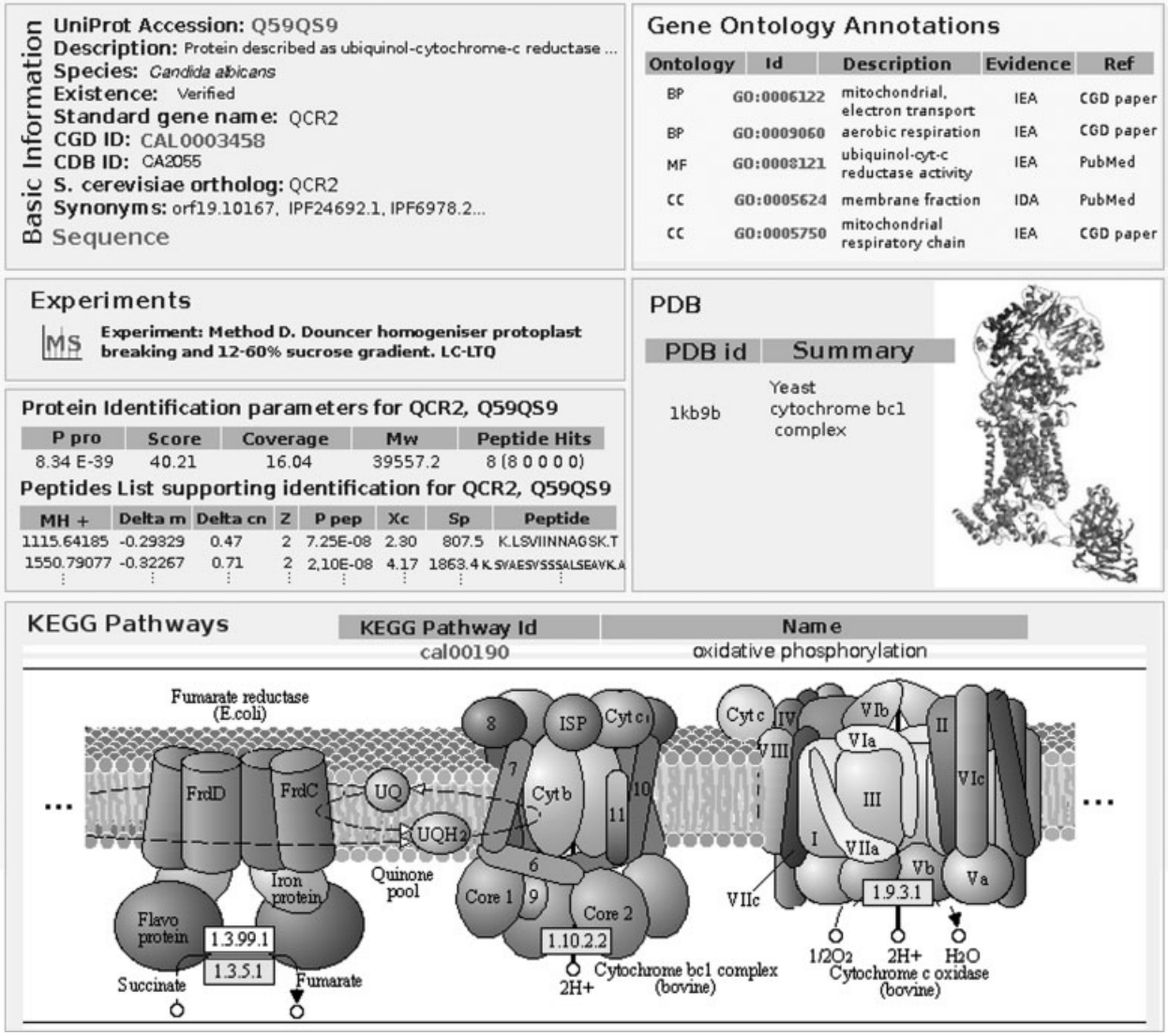

Figure 1. Use case: Search for *C. albicans* ubiquinol-cytochrome-*c* reductase QCR2. The different sections in the result comprise information on protein description and identifiers, experiments in which it has been identified, GO annotation, KEGG and CGD pathway annotation and structural information from PDB.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

and identification-supporting peptides lists. These data are subject to revision prior to eventual insertion into Proteopathogen by the database curators. Besides, the whole relational database and the MS data reports are available for download at the web site.

All the information that is retrievable from Proteopathogen when queried for one particular protein is shown in Fig. 1 for the specific case of ubiquinol-cytochrome-*c* reductase QCR2 of *C. albicans* which has been reported to show antigenic properties in human [19].

The *Protein Basic Information* section displays the Uniprot accession number, a brief description of the protein as stated at CGD, evidence for its existence, standard gene name, feature name, CGD and Candida Database identifiers, yeast ortholog gene name, synonyms and sequence.

The Section 2 lists all the experiments in which QCR2 has been identified. All of them belong to the same general approach aimed at purification of membrane proteins. In every case, the corresponding links to the MS identification parameters and supporting peptides are displayed as well. This experimental data are shown in Fig. 1 for identification of QCR2 in the experiment described as ''Method D. Douncer homogenizer protoplast breaking and 12–60% sucrose gradient. LC-LTQ''.

The section entitled *GO annotations* shows terms related to the electron transport chain, but more interestingly, it also shows an *inferred from direct assay* (IDA) annotation to the term *membrane fraction* [20], which fits to the fact that the protein is identified in five of the methods aimed at purification of membrane proteins.

KEGG Pathways table provides a link to the KEGG Pathway entry for *Oxidative phosphorylation*, and provides the feature to show in place the image corresponding to the map from KEGG. *CGD Pathways* displays an analogous link to the pathway entry at CGD that, in this case, is named *aerobic respiration (cyanide sensitive)–electron donors*.

Finally, in the *PDB* section, there are four structure images available along with links to the PDB entries, corresponding to a cytochrome bc1 complex from *S. cerevisiae*. Orthologs were used since no structure could be found for the *Candida* protein.

In conclusion, Proteopathogen represents, up to date, the first public web-based repository for proteomics data related to studies involving *C. albicans* pathogenicity and its interaction with immune system cells in the host. Moreover, it enables a framework for public access and submission of this type of data and it is intended to be more actively populated in the near future, including data from different pathogenic fungi and mammalian cells, becoming a reference database in its field. Unlike other protein identification databases, Proteopathogen is focused to a specific topic but, at the same time, includes a wide range of data including descriptions of the experimental contexts, information on proteins such as GO and pathway annotations, structural information and detailed MS parameters. Therefore, Proteopathogen will contribute to save time and facilitate analysis of proteomic workflow reports for researchers interested in this area.

The authors are grateful to César Vicente from the Computer Architecture Department, Complutense University of Madrid for his excellent technical assistance. This work was supported by BIO 01989-2006 from the Comision Interministerial de Ciencia y Tecnología (CYCIT, Spain), DEREMICROBIANA – CM from Comunidad Autónoma de Madrid, and REIPI, Spanish Network for the Research in Infectious Diseases, RD06/0008/1027 from the Instituto de Salud Carlos III. The Proteomics work was carried out in the Proteomics Unit UCM-Parque Científico, a member of the National Institute for Proteomics PROTEORED, funded by Genoma España. APM and RNC are partially supported by Spanish grants BIO2007-67150-C03-02, S-Gen-0166/2006 and PS-010000-2008-1.

The authors have declared no conflict of interest.

References

[1] Calderone, R. A. (Ed.), *Candida and Candidiasis*, ASM Press, Washington D.C 2002.

[2] Fernández-Arenas, E., Cabezón, V., Bermejo, C., Arroyo, J., *et al.*, Integrated genomic and proteomic strategies bring new insight into *Candida albicans* response upon macrophage interaction. *Mol. Cell. Proteomics* 2007, *6*, 460–478.

[3] Martínez-Solano, L., Nombela, C., Molero, G., Gil, C., Differential protein expression of murine macrophages upon interaction with *Candida albicans*. *Proteomics* 2006, *6*, 133–144.

[4] Pitarch, A., Nombela, C., Gil, C., *Candida albicans* biology and pathogenesis: insights from proteomics. *Methods Biochem. Anal.* 2006a, *49*, 285–330.

[5] Pitarch, A., Nombela, C., Gil, C., Contributions of proteomics to diagnosis, treatment, and prevention of candidiasis. *Methods Biochem. Anal.* 2006b, *49*, 331–361.

[6] Monteoliva, L., Albar, J. P., Differential proteomics: an overview of gel and non-gel based approaches. *Brief Funct. Genomic Proteomics* 2004, *3*, 220–239.

[7] Hoogland, C., Mostaguir, K., Appel, R. D., Lisacek, F., The World-2DPAGE Constellation to promote and publish gel-based proteomics data through the ExPASy server. *J. Proteomics* 2008, *71*, 245–248.

[8] Pleissner, K. P., Eifert, T., Buettner, S., Schmidt, F. *et al.*, Web-accessible proteome databases for microbial research. *Proteomics* 2004, *4*, 1305–1313.

[9] Martens, L., Hermjakob, H., Jones, P., Adamski, M. *et al.*, PRIDE: the proteomics identifications database. *Proteomics* 2005, *5*, 3537–3545.

[10] Csank, C., Costanzo, M. C., Hirschman, J., Hodges, P. *et al.*, Three yeast proteome databases: YPD, PombePD, and CalPD (MycoPathPD). *Methods Enzymol,* 2002, *350*, 347–373.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[11] Arnaud, M. B, Costanzo, M. C., Skrzypek, M. S., Binkley, G. *et al.*, The Candida Genome Database (CGD), a community resource for *Candida albicans* gene and protein information. *Nucleic Acids Res.* 2005, *33*, 358–363.

[12] Rossignol, T., Lechat, P., Cuomo, C., Zeng, Q. *et al.*, CandidaDB: a multi-genome database for Candida species and related Saccharomycotina. *Nucleic Acids Res.* 2008, *36*, 557–561.

[13] Cabezón, V., Llama-Palacios, A., Nombela, C., Monteoliva, L., Gil, C., Analysis of *Candida albicans* plasma membrane proteome. *Proteomics* 2009, *9*, in press, DOI: 10.1002/pmic.200800988.

[14] Plaine, A., Walker, L., Da Costa, G., Mora-Montes, M. *et al.*, Functional analysis of *Candida albicans* GPI-anchored proteins: roles in cell wall integrity and caspofungin sensitivity. *Fungal Genet. Biol.* 2008, *45*, 1404–1414.

[15] UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2008, *36*, 190–195.

[16] Bult, C. J., Eppig, J. T., Kadin, J. A., Richardson, J. E., *et al.*, The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.* 2008, *36*, 724–728.

[17] Kanehisa, M., Araki, M., Goto, S., Hattori, M. *et al.*, KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008, *36*, 480–484.

[18] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G. *et al.*, The Protein Data Bank. *Nucleic Acids Res.* 2000, *28*, 235–242.

[19] Pitarch, A., Abian, J., Carrascal, M., Sanchez, M. *et al.*, Proteomics-based identification of novel *Candida albicans* antigens for diagnosis of systemic candidiasis in patients with underlying hematological malignancies. *Proteomics* 2004, *4*, 550–559.

[20] Insenser, M., Nombela, C., Molero, G., Gil, C., Proteomic analysis of detergent-resistant membranes from *Candida albicans*. *Proteomics* 2006, *6*, S74–S81.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim