

Proteopathogen, una base de datos de proteínas para el estudio de la interacción *Candida albicans* – hospedador

Vital Vialás¹, Rubén Nogales-Cadenas², César Nombela¹, Alberto Pascual-Montano², Concha Gil^{1,3}

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

Existen en la actualidad repositorios públicos en la web para el almacenamiento y gestión de datos proteómicos así como bases de datos específicas de hongos. Sin embargo, ninguno de ellos está enfocado específicamente al área de investigación de hongos patógenos y su interacción con el hospedador y contiene además datos experimentales proteómicos. En este contexto presentamos Proteopathogen, una base de datos orientada a recopilar datos experimentales proteómicos y a facilitar el almacenamiento y acceso a un amplio rango de información, abarcando flujos de trabajo en Proteómica desde la descripción del experimento que lleva a la preparación de la muestra hasta los parámetros de espectrometría de masas y los péptidos que apoyan las identificaciones. Proteopathogen está actualmente enfocado hacia *Candida albicans* y su interacción con macrófagos, sin embargo, datos experimentales relativos a otras especies de hongos patógenos y células de mamíferos pueden ser adecuados para la inserción en la base de datos. Proteopathogen es públicamente accesible en <http://proteopathogen.dacya.ucm.es>.

Candida albicans es un hongo patógeno oportunista presente habitualmente en las mucosas humanas. En individuos inmunocomprometidos puede proliferar excesivamente y provocar candidiasis, una micosis muy extendida y en ocasiones fatal. En este sentido, abordar estudios proteómicos sobre la interacción de *Candida* con células del sistema inmune es clave para mejorar nuestra comprensión del proceso de infección y puede suponer un paso inicial en investigación básica para el futuro desarrollo de métodos de diagnóstico, vacunas y fármacos antifúngicos.

Respecto a la recopilación de información contenida en Proteopathogen, se han incluido datos correspondientes a tres experimentos. Los dos prime-

ros corresponden a trabajos de interacción *Candida* – macrófago [1,2] de los cuales el primero incluye 66 proteínas identificadas de *Candida* y el segundo, 38 proteínas de macrófago. El tercer estudio es un conjunto de experimentos destinados a la extracción e identificación de proteínas de membrana y proteínas con anclaje GPI [3] de *C. albicans*.

Además de la información experimental, y con el objetivo de proporcionar una visión más amplia de los datos, se recuperó información relevante de bases de datos en la web. Se obtuvieron de CGD [4] identificadores, sinónimos, ortólogos en *Saccharomyces cerevisiae*, anotaciones de Gene Ontology y anotaciones para las proteínas de *Candida*, mientras que en el caso de proteínas de macrófagos murinos, la información equivalente se obtuvo de UniProt [5] y de Mouse Genome Database [6]. También se recuperaron de KEGG [7], e información de estructuras de PDB [8]

En cuanto a la arquitectura de la aplicación, la estructura básica consiste en una base de datos MySQL gestionada por la plataforma de desarrollo web Ruby on Rails que permite crear las tablas y relaciones entre los datos, manejar las consultas y mostrar las páginas.

El contexto experimental es tratado en Proteopathogen jerárquicamente, mostrando una aproximación general del estudio, en donde se detallan autores, título de la publicación e identificador de PubMed, y dentro de esa descripción general, los experimentos concretos que dan lugar a las identificaciones.

La información sobre una proteína particular se muestra dividida en secciones (Figura 1). En la sección titulada *Protein Basic Information* se muestran el número de acceso de Uniprot, descripción de la

proteína, especie, evidencia de la existencia, nombre estándar del gen, identificadores de bases de datos específicas de organismo, ortólogos y secuencia. En la sección *Experiments* se listan aquellos experimentos en que se ha identificado la proteína en particular. Además, y en los casos que las anotaciones estén disponibles, se muestran una o más de las siguientes secciones: anotaciones de Gene Ontology con las correspondientes referencias, rutas de KEGG y CGD, e información estructural de PDB.

UniProt Accession: Q59Q59
Description: Protein described as ubiquinol-cytochrome-c reductase...
Species: *Candida albicans*
Evidence: Verified
Standard gene name: QCR2
CGD ID: CAL0003450
CGD ID: CA2026
S. cerevisiae orthology: QCR2
Synonyms: sc:orf19.10167, IPF24692.1, IPF6978.2...
Sequence:

Gene Ontology Annotations

Ontology	ID	Description	Evidence	Ref
BP	GO:0095322	mitochondrial electron transport	IEA	CGD paper
BP	GO:0009869	aerobic respiration	IEA	CGD paper
MF	GO:0003273	ubiquinol- <i>o</i> -c	IEA	PubMed
MF	GO:0003273	reductase activity	IEA	PubMed
CC	GO:0008524	membrane fraction	IDA	PubMed
CC	GO:0008578	mitochondrial respiratory chain	IEA	CGD paper

Experiments
 Experiment: Method 0. Dounce homogeniser protoplast breaking and 12-60% sucrose gradient. LC-LTD

Protein Identification parameters for QCR2, Q59Q59

PKS	Score	Coverage	PKS	Peptide Hits
8.54 E 09	42.21	14.54	39557.2	818 0 0 0 0

Peptides List supporting identification for QCR2, Q59Q59

MH	Delta m	Delta cn	Z	P pep	Xc	Sp	Peptide
1135	64195	-0.28209	0.47	2	7.25E-08	2.30	807.5 KLVATNVAQDK.T
1250	79077	-0.32267	0.71	2	2.10E-08	4.17	1080.4 KVAEQRSSGSDQK.A

KEGG Pathways
 KEGG Pathway Id: cal00190
 Name: oxidative phosphorylation

En todos los casos, las proteínas aparecen relacionadas con los experimentos a los que pertenecen de forma que se puede enlazar con la información relativa a la identificación. Estos datos comprenden, por una parte parámetros comunes a todas las proteínas identificadas en cada experimento, incluyendo la base de datos de la búsqueda, software de análisis, enzima de digestión, modificaciones fijas y variables y máximo número permitido de cortes no efectuados. Por otra parte también están presentes los datos que se refieren a cada identificación, donde se muestran la secuencia de los péptidos, su masa observada y masa teórica y la puntuación.

La interfaz pública de Proteopathogen ofrece múltiples formas de consultar el contenido. Es posible navegar por la lista de experimentos y visualizar rápidamente la lista de proteínas pertenecientes a cada uno de ellos, pero también se puede utilizar el formulario de búsqueda. Éste acepta varios tipos de identificadores y además permite realizar búsquedas por texto libre, recuperando una lista de proteínas que muestren coincidencias.

Por último, y para mejorar la interactividad con el usuario, Proteopathogen incluye un formulario para el envío de datos. Éstos serán revisados por un administrador, y una vez verificados podrán ser incluidos en la base de datos.

Referencias

- [1] Fernández-Arenas E, Cabezón V, Bermejo C, Arroyo J, Nombela C, Diez-Orejas R, et al. Integrated genomic and proteomic strategies bring new insight into *Candida albicans* response upon macrophage interaction. *Mol Cell Proteomics* 2007; 6: 460-478.
- [2] 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.
- [3] Cabezón V, Llama-Palacios A, Nombela C, Monteoliva, L, Gil C. Analysis of *Candida albicans* plasma membrane proteome. *Proteomics* 2009; 12, 9(20):4770-4786.
- [4] Arnaud MB, Costanzo MC, Skrzypek MS, Binkley G, Lane C, Miyasato SR, et al. The *Candida* Genome Database (CGD), a community resource for *Candida albicans* gene and protein information. *Nucleic Acids Res* 2005; 33:358-363.
- [5] UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acid Res* 2008; 36: 190-195.
- [6] Bult C, Eppig JT, Kadin JA, Richardson JE, Blake JA The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res* 2008;36: 724-728.
- [7] Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 2008; 36: 480-484.
- [8] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* 2000; 28: 235-242.