

Interacciones proteína-proteína: bases de datos y métodos teóricos de predicción

Victor de la Torre Russis^{1,2}, Alfredo Valles^{1,2}, Raúl Gómez¹, Glay Chinae¹, ✉ Tirso Pons¹

¹Centro de Ingeniería Genética y Biotecnología (CIGB), PO Box 6162, Habana 10600, Cuba.
Tel.: +(537)-271 60 22, ext. 1133; Fax: +(537)-271 47 64 or +(537)-33 60 08;
E-mail: tirso.pons@cigb.edu.cu ²Centro Nacional de Bioinformática (BIOINFO), Habana 10200, Cuba.

RESUMEN

Las interacciones proteína-proteína son eventos críticos para muchos procesos celulares que se extienden desde la formación de estructuras macromoleculares y complejos enzimáticos, hasta la regulación y la transducción de señales. Comprender estas interacciones en detalle, proporciona información útil sobre el mecanismo molecular de funcionamiento de la célula, y ayuda en el diseño de fármacos más específicos y en la ingeniería de procesos celulares. En el presente artículo se revisaron los avances recientes en el estudio de las interacciones entre proteínas, la creación de bases de datos, la utilización combinada de sus anotaciones y el desarrollo de métodos teóricos de predicción. El período revisado abarcó desde el año 1975 hasta el 2003.

Palabras claves: bases de datos, doble-híbrido, micro-arreglo de ADN, micro-arreglo de proteínas, espectrometría de masas

Biotecnología Aplicada 2003;20:201-208

ABSTRACT

Protein-protein interactions: databases and prediction methods. Protein-protein interactions are critical for virtually every cellular process, ranging from the formation of macromolecular and enzymatic complexes to the protein regulation and signal transduction. Detailed knowledge of the protein interaction networks should provide new insights into the structure and properties of the cellular systems, and it might enable the design of specific drugs and the engineering of cellular processes. In the present article we review the experimental techniques available for the study of protein-protein interaction, the development of new databases, the combined use of their complementary information, and the different computational methods for the prediction of protein interactions.

Keywords: databases, yeast two-hybrid, DNA microarray, protein chips, mass spectrometry

Introducción

Las interacciones entre proteínas desempeñan un papel fundamental en varios aspectos de la organización estructural y funcional de la célula [1, 2]. Estas interacciones pueden ser clasificadas en tres tipos: *i*) las que ocurren entre dominios de una misma cadena polipeptídica, *ii*) las que ocurren entre dominios de diferentes cadenas polipeptídicas en proteínas multiméricas, y *iii*) las que ocurren de manera transiente en complejos formados entre proteínas independientes [3].

En la actualidad se dispone de las secuencias de aminoácidos codificadas por 130 genomas secuenciados completamente (<http://igweb.integratedgenomics.com/GOLD/>; <http://www.expasy.ch/sprot/>), de un gran número de datos obtenidos a través de micro-arreglos de ADN (del inglés "DNA microarray"), e información de gran valor sobre interacciones proteína-proteína obtenida mediante los métodos doble-híbrido (del inglés "yeast two-hybrid"), espectrometría de masas, micro-arreglo de proteínas, cristalografía, inmunoprecipitación y mediciones por BIAcore. La combinación de estos métodos experimentales y su utilización a gran escala en *Caenorhabditis elegans* [4], *Saccharomyces cerevisiae* [5-8] y *Helicobacter pylori* [9] ha dado lugar a la creación de un amplio conjunto de bases de datos, al tiempo que la visión sobre la función de las proteínas ha evolucionado.

Desde el punto de vista clásico, la función de una proteína, denominada frecuentemente función molecular, está determinada por la acción local de una sola molécula proteica; mientras que la visión post-genómica

define la función de la proteína como su enrejado de interacciones con otras moléculas (Figura 1). El término más utilizado para describir este punto de vista es función contextual o función celular.

Del análisis combinado de los datos experimentales sobre interacciones entre proteínas, se ha comprendido que la complejidad de los organismos no depende del número de proteínas codificadas en sus genomas, sino del número de interacciones entre ellas [10]. Por ejemplo, la planta *Arabidopsis thaliana*, el nemátodo *C. elegans*, la mosca *Drosophila melanogaster* y el hombre poseen un número similar de genes (25,498; 19,099; 14,100 y 30,000-35,000; respectivamente). Por ello, constituye un desafío el desarrollo de métodos teóricos para predecir los detalles a cerca de las

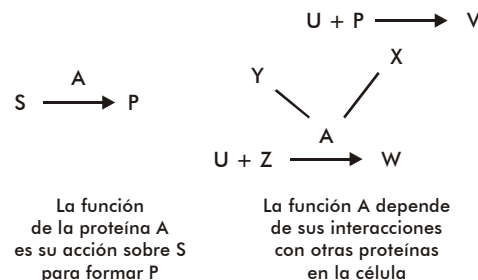


Figura 1. Evolución del significado de la función de las proteínas. La visión clásica se muestra a la izquierda y la visión post-genómica a la derecha.

interacciones específicas entre proteínas a nivel atómico, y la aplicación de dichos métodos a escala genómica [11, 12]. Una vez dilucidadas las redes de interacciones o “interactoma” en sistemas celulares modelos, será posible el diseño de fármacos más específicos y la ingeniería de procesos celulares.

En conexión con esta problemática, se celebró en Junio del 2001 en Charleston, Carolina del Sur, la primera conferencia sobre modelación de interacciones entre proteínas a escala genómica (<http://reco3.ams.sunysb.edu/conference/>). Esta conferencia incluyó sesiones sobre el acoplamiento proteína-proteína (del inglés “protein-protein docking”), la relación estructura-función, las interacciones proteína-moléculas pequeñas, y la identificación de interacciones en vías metabólicas [13]. Adicionalmente, al igual que los experimentos para la evaluación a ciegas de los métodos de predicción de estructura tridimensional de proteínas (CASP y CAFASP), se ha creado un nuevo experimento denominado CAPRI (del inglés “Critical Assessment of Predicted Interactions”) para evaluar los métodos de acoplamiento proteína-proteína. Más información sobre CAPRI puede encontrarse en el sitio de internet <http://capri.ebi.ac.uk>.

Este artículo revisa los avances alcanzados en el estudio de las interacciones proteína-proteína, y presta especial atención a las anotaciones en las bases de datos, su utilización, y el desarrollo de métodos teóricos para su predicción.

Métodos experimentales para el análisis de interacciones proteína-proteína: Niveles de confianza

Los métodos experimentales disponibles para detectar interacciones entre las proteínas tienen distintos niveles de resolución (Tabla 1) y pueden clasificarse en cuatro categorías [14]. La primera categoría comprende el análisis atómico, donde la interacción es detectada por métodos muy sensibles como la difracción de rayos-X (DRX) que aporta información específica sobre los átomos y residuos que participan en la interacción. La segunda categoría agrupa a los métodos que detectan la interacción de forma directa, por ejemplo, experimentos de doble-híbrido y mediciones por BIAcore. La tercera categoría agrupa a los métodos de detección de complejos mediante técnicas de inmunoprecipitación o de espectrometría de masas. La cuarta categoría comprende bioensayos realizados a escala celular, por ejemplo, la proliferación celular estimulada por la interacción ligando-receptor. La segunda y tercera categorías, a diferencia de la primera, indican cuales proteínas están formando parte del complejo en un instante de tiempo pero no revelan el detalle químico de la interacción. Algunos métodos experimentales abarcan más de un nivel de resolución como se observa en la Tabla 1.

Además, recientemente se ha descrito en la literatura científica el uso de la tecnología de micro-arreglo de proteínas para el estudio de las interacciones proteína-proteína de levaduras [15-17] y la metodología de purificación por afinidad en “tandem” (TAP, del inglés “tandem affinity purification”) [5, 6, 18]. Ambas técnicas detectan interacciones físicas entre proteínas, de manera directa, y según la Tabla 1, ellas pueden ser incluidas en las categorías dos y tres.

Tabla 1. Métodos experimentales para el estudio de las interacciones entre proteínas.

Método	Nivel de resolución			
	atómico	directo	complejos	celular
Rayos-X y RMN	X	X		
Ensayos de competencia		X		
ELISA		X	X	
Ensayos de retardación en gel		X	X	
Doble-híbrido		X	X	
Cromatografía de afinidad		X	X	
BIAcore		X	X	
micro-arreglo de proteínas		X	X	
Resonancia paramagnética electrónica	X	X	X	
Cromatografía de filtración en gel		X	X	
Espectrometría de masas (TAP)		X	X	
Enlazamiento molecular (cross-linking)		X	X	
Co-inmunoprecipitación		X	X	
Co-sedimentación		X	X	
Sedimentación en gradiente de sacarosa		X	X	
Co-purificación			X	
Microscopía electrónica			X	
Inmunofluorescencia				X
Inmunolocalización				X
Inmunotinción				X
FRET	X	X	X	X
Ensayos de bloqueo con anticuerpos monoclonales			X	X
Ensayos de adhesión				X
Knock-out				X
Co-expresión transiente				X

FRET: energía fluorescente transferida por resonancia; ELISA: ensayo enzimático de inmunoabsorción; RMN: resonancia magnética nuclear; Knock-out: experimentos de delección de genes.

El uso eficiente de los datos sobre la interacción entre proteínas requiere de una evaluación crítica de su exactitud, redundancia (cuando se describe la interacción entre un par específico de proteínas, utilizando iguales condiciones experimentales) y complementariedad (cuando se describe la interacción entre un par específico de proteínas, por varios métodos con enfoques diferentes o por un solo método utilizando diferentes condiciones experimentales). Desdichadamente, la correspondencia entre un conjunto amplio de anotaciones de interacciones proteína-proteína, sólo se ha investigado recientemente [19]. Una posible explicación es que la comparación de los datos experimentales es difícil pues se obtienen en diferentes condiciones, poseen diferentes formatos y se requiere de un sistema de anotaciones como referencia. Según estimados de von Mering y colaboradores, más de la mitad de las interacciones informadas por los métodos de análisis masivos (análisis realizados para todas las proteínas codificadas en un genoma) son ilegítimas o falsos-positivos [19].

Es importante notar, además, que las proteínas interactúan entre ellas con un amplio intervalo de afinidades y escalas de tiempo. Algunas interacciones fisiológicas son transientes y débiles, por consecuencia, la detección de dichas interacciones está frecuentemente en el margen de error de la observación, lo cual provoca que puedan aparecer interacciones no fisiológicas como falsos-positivos [8, 9]. Por otro lado, muchas de las interacciones experimentales informadas en la literatura científica son inconsistentes con los datos tridimensionales (3D) obtenidos para complejos macromoleculares, por ejemplo, RNA-polimerasa, Arp2/3 y proteasoma, o con información adicional [11, 20].

Para aumentar el nivel de confianza en las interacciones detectadas experimentalmente, y documentadas en diferentes bases de datos, se debe observar la misma interacción por diferentes métodos [21]. Nótese que de alrededor de 80,000 interacciones descritas para proteínas de levadura, por métodos diferentes, solamente 2,400 son descritas por más de un método [19]. Otro enfoque para validar la interacción combina la información que brindan métodos experimentales independientes con datos 3D obtenidos por DRX [11]. Una lista de interfases proteína-proteína presentes en estructuras resueltas por DRX está disponible en la base de datos SPIN-PP, accesible en el sitio <http://trantor.bioc.columbia.edu/cgi-bin/SPIN/>.

Siguiendo esta línea de razonamiento, la integración de diferentes bases de datos y/o diferentes métodos, contribuirá a validar las anotaciones en las bases de datos [22]. Un ejemplo de ello lo constituye el método propuesto por Tong *et al.* [23] que consiste en combinar la metodología de las bibliotecas de fagos (del inglés "phage display") con la metodología de doble-híbrido. Primeramente, realizan una búsqueda de secuencias consenso en proteínas de levadura capaces de unir dominios de reconocimiento de péptidos mediante una biblioteca de fagos. Como resultado de esta búsqueda se obtiene un enrejado de interacciones que conecta a proteínas que poseen una secuencia consenso con aquellas que poseen dominios de reconocimiento específico para esas secuencias consenso, a la vez, define sitios de unión en algunas de las proteínas. Un segundo enrejado de interacciones se obtiene de la búsqueda de proteínas que interaccionan con los dominios de reconocimiento de las secuencias consenso, utilizando el método de doble-híbrido. Este método ha sido aplicado en la búsqueda de proteínas que interactúan con dominios SH3.

Bases de datos sobre interacciones entre proteínas

Para documentar y describir las interacciones entre proteínas, se ha desarrollado un grupo de bases de datos que incluyen toda la información obtenida por los diferentes métodos experimentales descritos en la Tabla 1 [14], y de la información publicada en la literatura científica. A continuación se expone una revisión de las bases de datos, centrando la atención en el tipo de datos que contienen y cómo acceder a ellas. En algunos casos se hacen comentarios sobre el volumen de información contenido en la base de datos, en el momento en que se escribe el presente manuscrito.

DIP (Database of Interacting Proteins), <http://dip.doe-mbi.ucla.edu>

Documenta interacciones moleculares, proteína-proteína, determinadas experimentalmente por los métodos de doble-híbrido, inmunoprecipitación, ensayos de bloqueo con anticuerpos monoclonales y anotaciones extraídas de la literatura científica, entre otros. DIP está implementada como una base de datos relacional formada por cuatro tablas. Una primera con información sobre las proteínas, una segunda sobre interacciones proteína-proteína, una tercera que describe detalles sobre los experimentos, y la cuarta tabla contiene la lista de todas las referencias en

Medline. DIP permite la representación visual de las interacciones así como navegar a través de la red de interacciones proteicas mediante una interfase WWW [24, 25]. En el momento de escritura de este manuscrito, el número de proteínas descritas en DIP ascendía a 6,978 y el número de interacciones a 18,260.

BIND (Biomolecular Interaction Network Database), <http://bind.ca>

Documenta interacciones moleculares que incluyen proteína-proteína, proteína-RNA, proteína-DNA, proteína-moléculas pequeñas, complejos moleculares, y vías metabólicas. Los datos experimentales obtenidos por doble-híbrido, espectrometría de masas, cristalografía, bibliotecas de fagos, y las anotaciones complementarias se encuentran almacenadas en forma de objetos en el formato *ASN.1* (Abstract Syntax Notation.1; <http://www.oss.com/asn1/index.html>), utilizado por el Centro Nacional para la Información Biotecnológica (NCBI), y pueden ser interrogados a través de una interfase WWW [26]. Actualmente, BIND documenta 15,141 interacciones, y 1,306 complejos.

InterDom (Database of interacting domains), <http://InterDom.lit.org.sg>

InterDom combina datos sobre fusiones de dominios, interacciones entre proteínas, complejos proteicos, e información de la literatura científica. Esta base de datos es una compilación de interacciones tentativas entre dominios de proteínas, obtenidas por métodos teóricos, que puede ser utilizada para la anotación y validación "*in silico*" de las interacciones detectadas experimentalmente. La estrategia fundamental es utilizar múltiples métodos teóricos y datos experimentales independientes, para predecir interacciones entre dominios de proteínas y asignar mayor puntuación a aquellas donde coincida la mayor cantidad de evidencias. Los datos están organizados en una base de datos relacional en el sistema MySQL, y puede ser interrogada con una combinación de Perl, PHP, Java, y HTML. InterDom contiene anotadas 15,058 posibles interacciones entre dominios, en su versión 1.1 con fecha 15 de febrero, 2002. Los autores de esta base de datos sugieren además, un sistema de puntuación para inferir las interacciones entre dominios de proteínas [27].

KDBI (Kinetic Data of Bio-molecular Interactions database), <http://xin.cz3.nus.edu.sg/group/kdbi/kdbi.asp>

Documenta datos experimentales sobre la cinética de interacción entre proteínas, proteína-ARN, proteína-ADN, proteína-ligando, ARN-ligando, ADN-ligando, y reacciones descritas en la literatura. KDBI contiene actualmente 8,273 entradas, que involucran 1,380 proteínas, 143 ácidos nucleicos, y 1,395 moléculas pequeñas. Además, posee enlaces a referencias en MEDLINE y a estructuras 3D depositadas en las bases de datos PDB y NDB. Los parámetros cinéticos anotados incluyen constantes de asociación, disociación, orden de la reacción, constantes de inhibición, de afinidad, constante catalítica, todos ellos pueden ser interrogados mediante una interfase WWW [28].

MINT (Molecular INTERaction database),
<http://cbm.bio.uniroma2.it/mint/index.html>

MINT contiene información sobre complejos moleculares, interacción entre dominios, modificaciones enzimáticas de las proteínas interactuantes, constantes de unión y otros datos cinéticos. Toda la información contenida en MINT es curada de manera manual por expertos, y está organizada en una base de datos relacional en el sistema SQL [29]. Actualmente, MINT tiene documentadas 4,568 interacciones que pueden ser interrogadas mediante una interfase WWW escrita en lenguaje HTML codificado (PHP; <http://www.php.net>).

WISTdb (Worm Interaction Sequence Tag database)

Contiene el mapa de las interacciones proteicas del nemátodo *C. elegans* obtenidas por doble-híbrido; proyecto guiado por el Dr. Marc Vidal del Instituto Dana Farber, de la escuela de medicina de Harvard en Boston, Massachusetts [4]. Las anotaciones han sido organizadas utilizando el sistema de manipulación de bases de datos orientado a objetos, ACEDB [30], y puede ser interrogada mediante su interfase ACEBrowser [31].

ASEdb (Alanine Scanning Energetics Database), <http://www.asedb.org>

Contiene mediciones experimentales de la afinidad con que se unen dímeros de proteína-proteína, proteína-ácido nucleico, y proteína-pequeñas moléculas, en los que se ha mutado una cadena lateral por alanina. En casos en que la estructura 3D esta disponible, contiene áreas de superficies de la cadena lateral mutada y enlaces a la base de datos PDB. ASEdb es una base de datos relacional útil para el análisis de la contribución de un aminoácido a la energética de la interacción. ASEdb contiene 2,919 mutantes, 1,953 poseen datos estructurales para el monómero y 580 para el dímero [32].

FlyNets,
http://gifts.univ-mrs.fr/GIFTS_home_page.html

Contiene las interacciones proteína-ADN, proteína-ARN y proteína-proteína, descritas en la mosca *D. melanogaster* [33]. FlyNets está compuesta por dos partes: *i*) FlyNets-base es una base de datos especializada en interacciones moleculares que ocurren durante el proceso de formación del patrón embriogénico; *ii*) FlyNets-list es una base de datos más general que documenta cualquier interacción molecular publicada, que ocurra en la mosca. Todos los datos están organizados en una colección de ficheros con formato de hipertexto, semejante a la estructura de la base de datos GenBank, y pueden ser interrogados a través de una interfase WWW.

MYGD (MIPS Yeast Genome Database),
<http://mips.gsf.de>

MYGD contiene información detallada sobre el genoma de *S. cerevisiae*, y las secuencias aminoácidas codificadas. Además, contiene información complementaria sobre la función de los genes y proteínas, complejos proteicos, fenotipo de los mutantes, interacciones proteína-proteína, y vías metabólicas. Toda la información está anotada en ficheros con formato

de hipertexto y puede ser interrogada a través de una interfase WWW [34].

BioKnowledge Library,
<http://www.proteome.com>

BioKnowledge es una base de datos relacional que contiene información específica sobre las proteínas. La información es recopilada de la literatura científica e incluye: función bioquímica, función y localización celular, interacciones genéticas y con otras proteínas, regulación, dominios y patrones de secuencia característicos. Todas estas características; concernientes a los organismos modelos *S. cerevisiae*, *C. elegans*, *Schizosaccharomyces pombe*; y otros organismos como *Candida albicans* y *Homo sapiens*. BioKnowledge es curada manualmente por expertos, y la información sobre cada proteína está organizada en formato HTML, con una página WWW para cada molécula [35].

PRONET, <http://pronet.doubletwist.com>

Es un proyecto no-académico fundado por Doubletwist™ y Myriad™. Esta base de datos contiene información sobre interacciones proteína-proteína y datos producidos por métodos propios [24].

CURAGEN, <http://www.curagen.com>

Contiene información sobre interacciones proteicas en levadura, obtenidas en los laboratorios del Dr. S Field y de la compañía Curagen. Posee herramientas para visualizar la red de 957 interacciones descritas, y los datos están accesibles en el sitio Web de la compañía para usuarios con objetivos académicos [7].

PIM, <http://www.hybrigenics.fr>

Contiene los resultados del análisis por doble-híbrido del genoma de *H. pylori*. En el estudio se describieron 1,465 interacciones proteicas anotadas en la base de datos [9].

BRITE (Biomolecular Relations in Information Transmission and Expression),
<http://www.genome.ad.jp/brite/>

Es una base de datos de relaciones binarias y comparación de grafos que involucran genes y proteínas. Contiene interacciones proteína-proteína obtenidas por doble-híbrido en *S. cerevisiae* y *H. pylori*, y por otras extraídas de la literatura científica; interacciones en vías metabólicas; información sobre similitud de secuencia; y relaciones obtenidas a través de micro-arreglo de ADN. BRITE, puede ser interrogada a través de una interfase WWW.

INTERACT,
<http://www.cl.cam.ac.uk/~wb204/GD99/>

Es una base de datos orientada a objetos, que contiene información sobre las interacciones entre proteínas descritas para los organismos *S. cerevisiae*, *C. elegans*, *Human*, y *Escherichia coli*. Cada interacción es representada como un grafo, con las proteínas individuales en los vértices y la interacción entre pares de ellas formando los bordes [36]. La información contenida en esta base de datos puede ser interrogada a través de una interfase WWW, que incluye lenguaje para modelado en realidad virtual (VRML; del inglés "Virtual Reality Modelling Language").

Métodos teóricos de predicción de interacciones entre proteínas

En el análisis funcional de un genoma, un paso crítico es la comprensión de las interacciones entre las proteínas codificadas [37]. Por ello se ha desarrollado un conjunto de métodos teóricos para abordar el problema de la predicción de interacciones entre proteínas a partir del análisis de sus secuencias aminoacídicas, de información de los genomas y de datos cristalográficos de complejos macromoleculares. De manera general, todos los investigadores utilizan como hipótesis que si se conoce el modo de interacción entre dos proteínas o dos dominios de proteínas, cualquier interacción que pueda ocurrir entre proteínas homólogas involucra contactos del mismo tipo [3]. Sin embargo, existen casos en los que proteínas homólogas poseen diferentes modos de interacción. Por ejemplo, la enzima nucleótido difosfato quinasa (EC 2.7.4.6) se presenta en la naturaleza en forma de hexámero o tetrámero dependiendo de la especie [38]. Las sialidasas (EC 3.2.1.18) virales forman tetrámeros, mientras que las bacterianas y de tripanosoma aparecen como monómeros de proteínas multidominio [39]. Una revisión reciente sobre el tema de predicción de interacciones entre proteínas fue publicada por Valencia y Pazos [40].

El reconocimiento proteína-proteína está determinado por las propiedades físicas y químicas de la interfase, la cual ha sido caracterizada en términos de su geometría (tamaño, forma y complementariedad) y de su naturaleza química (tipo de grupo químico y aminoácido, hidrofobicidad, interacciones electrostáticas y enlaces por puente de hidrógeno). El análisis de las superficies de contacto entre proteínas o interfase tuvo sus inicios con el trabajo de Chothia y Janin en 1975 [41]. A partir de ese año y hasta el 2002, han aparecido diferentes trabajos sobre el tema, de varios autores [42-58]. De estos trabajos se ha derivado un conjunto de regularidades útiles para los métodos de predicción de interacción proteína-proteína.

La utilización de los métodos de acoplamiento proteína-proteína, para predecir las regiones de interacción y la estructura de complejos proteicos, es un tema que merece una descripción detallada y se aleja del objetivo del presente artículo. Revisiones recientes sobre este tema fueron publicadas por Halperin *et al.* [59], y Smith y Sternberg [60].

Varios investigadores han utilizado la información contenida en los genomas (la fusión de dominios, la conservación del orden génico, la distribución filogenética), para predecir interacciones entre proteínas. Si dos cadenas polipeptídicas son sub-unidades de una misma proteína o complejo, sus genes están frecuentemente adyacentes en el genoma, y se expresan y regulan de manera conjunta al nivel de ADN. Dandekar *et al.* [61] utiliza la información que brinda la conservación del orden de genes ortólogos y la arquitectura de los operones en 6 genomas de procariotas y 3 de archaeas. Con este criterio, se identificaron para cada genoma alrededor de 100 proteínas involucradas en interacciones físicas. Overbeek *et al.* [62] y Lathe *et al.* [63] aplicaron este criterio a genomas poco relacionados. El método desarrollado por Pellegrini *et al.* [64], denominado perfiles filogenéticos (del inglés "phylogenetic profiles"), se basa en la co-

evolución de proteínas que interactúan, dicho de otra manera presencia-ausencia de genes ortólogos en genomas diferentes. Si el patrón de proteínas ortólogas (presencia-ausencia) se conserva en organismos de la misma especie, se debe probablemente a que una de las proteínas no puede ejercer su función sin la otra. Huynen *et al.* [65] aplicaron este método al genoma de *Mycoplasma genitalium*, prediciendo 34% de los pares como interactuantes y un 29% adicional pertenecientes a la misma vía metabólica o proceso funcional. Marcotte *et al.* [66, 67] y Enright *et al.* [68] desarrollaron métodos basados en la fusión de dominios (del inglés "fused domain approach"). Los métodos utilizan la hipótesis de que cuando dos proteínas A y B contienen dominios homólogos a dominios diferentes de una tercera proteína en otro organismo, pero A y B no son homólogas entre sí, entonces A y B interactúan. Enright y colaboradores utilizan BLAST [69] para determinar ortología entre proteínas y encuentran 64 casos de genes fusionados en cuatro genomas procariotas. El método desarrollado por Marcotte y colaboradores, conocido como "Rosetta stone method", utiliza la información anotada en las bases de datos Pfam [70] y ProDom [71] para determinar homología entre dominios de proteínas individuales.

La limitación de los métodos descritos radica en que sólo son aplicables a genomas secuenciados completamente, para tener la certeza de la ausencia de genes específicos; por otro lado son aplicables solamente en bacterias, donde el orden de los genes es una característica relevante; y por último dependen de la calidad del alineamiento múltiple de las proteínas en estudio.

Gallet y colaboradores [72] desarrollaron un método simple de predicción que identifica secuencias de aminoácidos, denominadas dominios de unión al receptor, basado en el análisis de sus propiedades hidrofóbicas. Las secuencias de aminoácidos más propensas a interactuar, se obtuvieron de un análisis estadístico de las frecuencias de aparición de los aminoácidos en sitios de interacción conocidos. El método identificó el 95% de los residuos involucrados en la interacción ADN-proteína y el 83% en los dominios de unión a calcio. Sprinzak y Margalit [73] utilizaron un enfoque similar para identificar patrones de secuencia característicos, observados como regularidad, en pares de proteínas cuya interacción ha sido comprobada de forma experimental. Una de las proteínas posee un patrón de secuencia característico definido en la clasificación InterPro [74], y la otra proteína que interacciona un segundo patrón característico. A este conjunto de pares de patrones característicos se les denominaron patrones de secuencia correlacionados. El método fue probado en una base de datos de interacciones, comprobadas de manera experimental, en la levadura *S. cerevisiae*.

Bock y Gough [75] utilizan la información contenida en 70 hetero-complejos extraídos de la base de datos PDB, para entrenar un algoritmo de clasificación denominado "Support Vector Machine (SVM)" capaz de reconocer y predecir interacciones, basado solamente en la identidad de secuencia de aminoácidos. Wojcik y Schächter [76] combinan métodos de búsqueda de similitud de secuencia aminoacídica y agrupamiento de secuencias (del inglés "clustering") con la información contenida en las bases de datos acerca de dominios en

proteínas que interactúan, para predecir mapas de interacción proteína-proteína. Este método se ha empleado para predecir el mapa de interacción de proteínas en la bacteria *Escherichia coli*, utilizando la información previa del mapa de interacción de *H. pylori*.

Pazos y Valencia [77, 78] combinan información de secuencias aminoacídicas e información de genomas para calcular las distancias evolutivas entre proteínas que pertenecen a familias relacionadas y para calcular la correlación de mutaciones entre posiciones de un alineamiento múltiple de secuencias. Esta idea se basa en observaciones previas que indican una correspondencia entre los árboles filogenéticos de proteínas asociadas (ejemplo: ligando-receptor); para una revisión sobre este tema, vea el artículo de Fraser *et al.* [79]. El método propuesto por Pazos y Valencia se ha aplicado a gran escala, utilizando 67,000 pares de proteínas de *E. coli*, de los cuales 2,742 fueron predichos correctamente [77]. Un método similar al propuesto por Pazos y Valencia, ha sido desarrollado por Ramani y Marcotte [80], quienes alinean árboles filogenéticos, representados por matrices, para definir proteínas específicas involucradas en la interacción. Este método se ha aplicado a más de 18 familias de proteínas.

Lu *et al.* [81] desarrollaron un método denominado MULTIPROSPECTOR que extiende la metodología de compatibilidad secuencia-estructura (del inglés "threading") a la predicción de estructura cuaternaria, y que consta de dos fases. Una primera fase ejecuta el método de "threading" PROSPECTOR, para generar un conjunto de estructuras 3D potenciales de la proteína en estudio. En una segunda fase, re-evalúan la compatibilidad de las estructuras 3D generadas con las estructuras molde que forman parte de complejos cristalográficos determinadas experimentalmente. El método fue evaluado utilizando un conjunto de 40 homodímeros, 15 heterodímeros y 69 monómeros, explorados contra una selección de 2,478 estructuras representativas de la base de datos PDB. El método predijo correctamente 36 homodímeros, 15 heterodímeros y 65 monómeros. Además, se evaluaron todas las posibles interacciones para proteínas de levadura, de un total de 2,865 predicciones utilizando MULTIPROSPECTOR, 1,138 están documentadas en la base de datos DIP, en comparación con PSI-BLAST que predice 1,781 interacciones, de ellas 1,215 documentadas en DIP.

Zhou y Shan [82] y Fariselli *et al.* [83] utilizan redes neuronales entrenadas en conjuntos de proteínas que interactúan, específicamente perfiles de secuencia y accesibilidad al solvente de residuos vecinos, para predecir sitios de interacción proteína-proteína. Estos métodos son capaces de predecir correctamente el 69-73% de los residuos que componen la interfase.

Aloy y Russell [84, 85] utilizan las estructuras cristalográficas de complejos moleculares para evaluar mediante un potencial empírico, cuándo una proteína homóloga al par de proteínas interactuantes conserva el mismo mecanismo de interacción. Por último, Bader y Hogue [86] han aplicado la teoría de grafo para detectar regiones densamente pobladas en enrejados de interacciones proteína-proteína, y de esta manera predecir complejos moleculares. El algoritmo desarrollado por estos investigadores, MCODE (del inglés "molecular complex detection"), fue evaluado

utilizando las anotaciones sobre las interacciones en *S. cerevisiae*.

Una lista de métodos de predicción con sitios WWW, se muestra a continuación:

- <http://www.biochem.ucl.ac.uk/bsm/PP/server/> [47].
- maine.ebi.ac.uk:8000/services/allfuse/ [68].
- <http://www.sanger.ac.uk/Software/Pfam/> [70].
- gpcr.biocomp.unibo.it/predictors [83].
- www.russell.embl.de/interprets [84, 85].
- <http://cbm.bio.uniroma2.it/iSPOT> [87].

Los autores de esta revisión han desarrollado un método para la inferencia de interacciones entre proteínas que utiliza el modelo de cadenas ocultas de Markov (HMM) y la combinación de las anotaciones contenidas en las bases de datos PFAM, DIP y SWISS-PROT/TrEMBL. Este método permite además, una anotación automática acerca de la función contextual o función celular de las proteínas en estudio, así como de los dominios funcionales que las componen, lo que resulta de gran utilidad para complementar y enriquecer experimentos de análisis funcional de genomas, proteómica y proyectos de secuenciación. El método con todas sus particularidades conformó la Tesis de Diploma de Victor de la Torre Russis, defendida en la Universidad de La Habana en el año 2003 y está disponible a través de una interfase WWW en los sitios de internet <http://www.bioinfo.cu/iPPI.html> y <http://bio.cigb.edu.cu/iPPI/iPPI.html>.

Actualización

Fraser *et al.* 2003 proponen que las interacciones proteína-proteína tienden a disminuir la velocidad a la cual ellas evolucionan [88]. Las proteínas que interactúan acumulan menos cambios de aminoácidos en su secuencia, debido posiblemente, a restricciones estructurales.

Sprinzak y colaboradores proponen un método de cómputo para evaluar la exactitud de los datos experimentales sobre interacciones proteína-proteína obtenidos por el método de doble-híbrido [89]. Estos investigadores proponen que el 50% de las interacciones proteína-proteína identificadas en la levadura *S. cerevisiae* corresponden a casos positivos y que el número total de interacciones para este organismo ("interactoma") es de 10,000-16,600.

Ofran y Rost utilizaron una red neuronal para predecir la región de contacto en las interacciones proteína-proteína [90]. El método sólo requiere de la información de la secuencia aminoacídica de las proteínas que interactúan. Estos investigadores concluyeron que el 70% de sus predicciones son correctas y que predicen correctamente al menos un sitio de interacción en 66 de un total de 333 complejos proteicos estudiados (20% de los casos).

Goffard *et al.* 2003 desarrollaron el método IPPRED para la inferencia de interacciones entre proteínas basado en la similitud de secuencias aminoacídicas [91]. IPPRED utiliza el programa BLAST [69] y los datos experimentales documentados en BIND [26] y en la literatura científica para inferir las interacciones entre proteínas. Además, IPPRED está disponible a través de una interfase WWW en el sitio de internet <http://cbi.labri.fr/outils/ippred/>.

von Mering *et al.* 2003 implementaron la predicción de asociaciones entre proteínas en el método

STRING [92]. STRING no utiliza la información experimental sobre interacciones proteína-proteína anotada en las bases de datos referidas en este artículo; por el contrario utiliza información sobre la fusión de dominios, la conservación del orden génico y la distribución filogenética, contenida en los genomas secuenciados completamente. La nueva versión del método está disponible a través de una interfase WWW en el sitio de internet <http://www.bork.embl-heidelberg.de/STRING/>.

Conclusiones

La anotación funcional detallada para un genoma secuenciado completamente requiere del estudio de las interacciones entre sus componentes. De aquí se desprende la importancia y el enorme esfuerzo para desarrollar métodos experimentales y teóricos más precisos para el estudio de las interacciones entre pro-

teínas. Una vez descritas estas interacciones o “*interactoma*” será posible el diseño de fármacos más específicos y la ingeniería de procesos celulares.

La combinación del mayor número posible de datos experimentales y teóricos, sobre interacciones entre proteínas, proporcionará información más completa y detallada de los sistemas biológicos. Especial cuidado debe prestarse a las anotaciones en las bases de datos, pues de manera general son la fuente para el desarrollo de métodos teóricos de predicción. El uso combinado de las anotaciones en bases de datos que se complementen entre sí, aumentará el nivel de confianza en dichas anotaciones y permitirá una curación y validación detallada.

Agradecimientos

Los autores agradecen a Liliana Basabe y a la Dra. Lila Castellanos la lectura crítica del manuscrito y las oportunas sugerencias.

- Pawson T, Nash P. Protein-protein interactions define specificity in signal transduction. *Genes Dev* 2000;14:1027-47.
- Pawson T, Raina M, Nash P. Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett* 2002;513:2-10.
- Teichmann SA, Murzin AG, Chothia C. Determination of protein function, evolution and interactions by structural genomics. *Curr Opin Struct Biol* 2001;11:354-63.
- Walhout A, Sordella R, Lu X, Hartley J, Temple G, Brasch M, et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 2000;287:116-22.
- Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;415:141-7.
- Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002;415:180-3.
- Uetz P, Giot L, Cagney G, Mansfield TD, Judson RS, Knight JR, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000;403:623-7.
- Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* 2001;98:4569-74.
- Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, et al. The protein-protein interaction map of *Helicobacter pylori*. *Nature* 2001;409:211-5.
- Claverie JM. Gene number. What if there are only 30,000 human genes? *Science* 2001;291:1255-7.
- Edwards AM, Kus B, Jansen R, Greenbaum D, Greenblatt J, Gerstein M. Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet* 2002;18:529-36.
- Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein-protein interactions. *Genome Res* 2002;12:37-46.
- Vajda S, Vakser IA, Sternberg MJE, Janin J. Modeling of protein interactions in genomes. *Proteins* 2002;47:444-6.
- Xenarios I, Eisenberg D. Protein interaction databases. *Curr Opin Biotechnol*, 2001; 12:334-9.
- MacBeath G, Schreiber SL. Printing protein as microarrays for high throughput function determination. *Science* 2000;289:1760-2.
- Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, et al. Global analysis of protein activities using proteome chips. *Science* 2001;293:2101-5.
- Zhu H, Klemic JF, Chang S, Bertone P, Casamayor A, Klemic KG, et al. Analysis of yeast protein kinases using protein chips. *Nature Genet* 2000;26:283-9.
- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nature* 1999;17:1030-2.
- von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 2002;417:399-403.
- Kemmeren P, van Berkum NL, Vilo J, Bijma T, Donders R, Brazma A, et al. Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol Cell* 2002;9:1133-43.
- Schwikowski B, Uetz P, Fields S. A network of protein-protein interactions in yeast. *Nature Biotechnol* 2000;18:1257-61.
- Gerstein M, Lan N, Jansen R. Integrating interactomes. *Science* 2002;295:284-7.
- Tong AH, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, et al. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* 2002;295:321-4.
- Xenarios I, Fernández E, Salwinski L, Joyce-Duan X, Thompson MJ, Marcotte EM, Eisenberg D. DIP: The database of interacting proteins: 2001 update. *Nucleic Acids Res* 2001;29:239-41.
- Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 2002;30:303-5.
- Bader GD, Betel D, Hogue CWV. BIND: the biomolecular interaction network database. *Nucleic Acids Res* 2003;31:248-50.
- Ng SK, Zhang Z, Tan SH. Integrative approach for computationally inferring protein domains interactions. *Bioinformatics* 2003;19:923-9.
- Ji ZL, Chen X, Zhen CJ, Yao LX, Han LY, Yeo WK, et al. KDBI: kinetic data of bio-molecular interactions database. *Nucleic Acids Res* 2003;31:255-7.
- Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G. MINT: a molecular interaction database. *FEBS Lett* 2002;513:135-40.
- Thierry-Mieg J, Thierry-Mieg D, Stein L. ACEDB: The ACE database manager. In: S. Letovsky, editor. *Bioinformatics, Databases and Systems*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1999. p.265-78.
- Stein L, Thierry-Mieg J. Scriptable access to the *Caenorhabditis elegans* genome sequence and other Acedb databases. *Genome Res* 1999;8:1308-15.
- Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* 2001;17:284-5.
- Sánchez C, Lachaize C, Janody F, Bellon B, Röder L, Euzenat J, et al. Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an internet database. *Nucleic Acids Res* 1999; 27:89-94.
- Mewes HW, Frishman D, Güldener U, Mannhaupt G, Mayer K, Mokrejs M, et al. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res* 2002;30:31-4.
- Costanzo MC, Crawford ME, Hirschman JE, Kranz JE, Olsen P, Robertson LS, et al. YPDTM, PombePDTM and WormPDTM: model organism volumes of the BioKnowledgeTM library, an integrated resource for protein information. *Nucleic Acids Res* 2001;29:75-9.
- Eilbeck K, Brass A, Paton N, Hodgman C. INTERACT: an object oriented protein-protein interaction database. *ISMB* 1999;7:87-94.
- Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. Protein function in the post-genomic era. *Nature* 2000;405:823-6.
- Morera S, Lebras G, Lascu I, Lacombe ML, Veron M, Janin J. Refined X-ray structure of *Dictyostelium discoideum* nucleoside diphosphate kinase at 1.8 Å resolution. *J Mol Biol* 1994;243:873-90.

39. Crennell SJ, Garman EF, Laver W, Vimr E, Taylor GL. Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. *Proc Natl Acad Sci USA* 1993;90:9852-6.
40. Valencia A, Pazos F. Computational methods for the prediction of protein interactions. *Curr Opin Struct Biol* 2002;12:368-73.
41. Chothia C, Janin J. Principles of protein-protein recognition. *Nature* 1975;256:705-8.
42. Janin J, Chothia C. The structure of protein-protein recognition sites. *J Biol Chem* 1990;265:16027-30.
43. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. *J Mol Biol* 1993;234:946-50.
44. Young L, Jernigan RL, Covell DG. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci* 1994;3:717-29.
45. Janin J. Principles of protein-protein recognition from structure to thermodynamics. *Biochimie* 1995;77:497-505.
46. Janin J. Protein-protein recognition. *Prog Biophys Mol Biol* 1996;64:145-66.
47. Jones S, Thornton JM. Principles of protein-protein interaction. *Proc Natl Acad Sci USA* 1996;93:13-20.
48. Jones S, Thornton JM. Analysis of protein-protein interaction sites using surface patches. *J Mol Biol* 1997a;272:121-32.
49. Jones S, Thornton JM. Prediction of protein-protein interaction sites using surface patches. *J Mol Biol* 1997b;272:133-43.
50. Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* 1997;10:999-1012.
51. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci* 1997;6:53-64.
52. Keskin O, Bahar I, Badretdinov AY, Piitsyn OB, Jernigan RL. Residue frequencies and pairing preferences at protein-protein interfaces. *Protein Sci* 1998;7:2578-86.
53. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol* 1999;285:2177-98.
54. Norel R, Perrey D, Wolfson HJ, Nussinov R. Examination of shape complementarity in docking of unbound proteins. *Proteins* 1999;36:307-17.
55. Scheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interaction. *Curr Opin Struct Biol* 2000;10:153-9.
56. Valdar WSJ, Thornton JM. Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins* 2001;42:108-24.
57. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins* 2001;43:89-102.
58. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. *Proteins* 2002;47:334-43.
59. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 2002;47:409-43.
60. Smith GR, Sternberg MJ. Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol* 2002;12:28-35.
61. Dandekar T, Snel B, Huynen M, Bork P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* 1998;23:324-8.
62. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* 1999;96:2896-901.
63. Lathe WC, Snel B, Bork P. Gene context conservation of a higher order than operons. *Trends Biochem Sci* 2000;25:474-9.
64. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* 1999;96:4285-8.
65. Huynen M, Snel B, Lathe W, Bork P. Predicting protein function by genomic context: quantitative evaluation and qualitative. *Genome Res* 2000;10:1204-10.
66. Marcotte EM, Pellegrini M, Ho-Leung N, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. *Science* 1999a;285:751-3.
67. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. A combined algorithm for genome-wide prediction of protein function. *Nature* 1999b;402:83-6.
68. Enright A, Iliopoulos I, Kyrpidis N, Ouzounis C. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 1999;402:86-90.
69. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403-10.
70. Bateman A, Birney E, Cerruti L, Durbin R, Ewinger L, Eddy SR, et al. The Pfam protein families database. *Nucleic Acids Res* 2002;30:276-280.
71. Corpet F, Servan F, Gouzy J, Kahn D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res* 2000;28:267-9.
72. Gallet X, Charleatoux B, Thomas A, Brasseur R. A fast method to predict protein interaction sites from sequences. *J Mol Biol* 2000;302:917-26.
73. Sprinzak E, Margalit H. Correlated sequence-signatures as markers of protein-protein interaction. *J Mol Biol* 2001;311:681-92.
74. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 2001;29:37-40.
75. Bock JR, Gough DA. Predicting protein-protein interactions from primary structure. *Bioinformatics* 2001;17:455-60.
76. Wojcik J, Schächter V. Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics* 2001;17 Suppl 1:296-305.
77. Pazos F, Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng* 2001;14:609-14.
78. Pazos F, Valencia A. In silico two-hybrid system for the selection of physically interacting proteins pairs. *Proteins* 2002;47:219-27.
79. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. *Science* 2002;296:750-2.
80. Ramani AK, Marcotte EM. Exploiting the co-evolution of interacting proteins to discover interaction specificity. *J Mol Biol* 2003;327:273-84.
81. Lu L, Lu H, Skolnick J. MULTIPROSPER: an algorithm for the prediction of protein-protein interaction by multimeric threading. *Proteins* 2002;49:350-64.
82. Zhou HX, Shan Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins* 2001;44:336-43.
83. Fariselli P, Pazos F, Valencia A, Casadio R. Prediction of protein-protein interaction sites in heterocomplexes with neural networks. *Eur J Biochem* 2002;269:1356-61.
84. Aloy P, Russell RB. Interrogating protein interaction networks through structural biology. *Proc Natl Acad Sci USA* 2002;99:5896-901.
85. Aloy P, Russell RB. InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics* 2003;19:161-2.
86. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2.
87. Brannetti B, Zanzoni A, Montecchi-Palazzi L, Cesareni G, Helmer-Citterich M. iSPOT: a web tool for the analysis and recognition of protein domain specificity. *Comparative and Functional Genomics* 2001;2:314-8.
88. Fraser HB, Wall DP, Hirsh AE. A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC Evolutionary Biol* 2003;3:11.
89. Sprinzak E, Sattath S, Margalit H. How reliable are experimental protein-protein interaction data? *J Mol Biol* 2003;327:919-23.
90. Ofra Y, Rost B. Predicted protein-protein interaction sites from local sequence information. *FEBS Lett* 2003;544:236-39.
91. Goffard N, Garcia V, Iragne F, Groppi A, de Daruvar A. IPPRED: server for proteins interactions inference. *Bioinformatics* 2003;19:903-4.
92. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;31:258-61.

Recibido en Mayo de 2003. Aprobado en Agosto de 2003.