

CAPITULO 11

INTERACCIÓN PROTEÍNA-PROTEÍNA

Luciana Rodriguez Sawicki y Natalia Bottasso Arias

Introducción

Las interacciones proteína-proteína (IPP) son intrínsecas a todos los procesos celulares, incluyendo la expresión génica, crecimiento celular, proliferación, asimilación de nutrientes, morfología, motilidad, comunicación intercelular y apoptosis, entre otros. Pero las células responden a un gran número de estímulos, por lo tanto la expresión celular es un proceso dinámico, aquellas proteínas que cumplen funciones específicas pueden no estar siempre expresadas o activadas. Asimismo, no todas las células son iguales, y muchas proteínas se expresan de modo dependiente del tipo celular. Estas características propias de las proteínas hacen que la investigación de las IPP sea compleja, especialmente cuando se trata de entender la función de una proteína en su contexto biológico adecuado (Golemis, 2002).

Dentro de los aspectos críticos requeridos para entender la función de una proteína se incluyen:

- Secuencia y estructura de la proteína – se usa para descubrir motivos que ayuden a predecir la función.
- Historia evolutiva y secuencias conservadas – identifica residuos regulatorios clave.
- Perfil de expresión – revela la especificidad del tipo celular y cómo la expresión es regulada.
- Modificaciones post-traduccionales – fosforilación, acilación, glicosilación, y ubiquitinación sugieren localización, activación y/o función.
- Interacciones con otras proteínas – la función puede extrapolarse al conocer la función de sus ligandos.

- Localización intracelular – puede ayudar a inferir la función de la proteína. Hasta los años 90, el análisis de la función de una proteína se focalizaba en su estudio de modo individual. Sin embargo la gran mayoría de las proteínas interactúan con otras al momento de cumplir su función, por ende, las mismas deben ser estudiadas en el contexto de sus compañeros interactuantes para comprender de modo integral el rol que desempeñan.

El estudio de las IPP requiere de la utilización de conocimientos y técnicas de diversas áreas como biología, genética, bioquímica, biofísica y biología molecular. Comprender completamente la naturaleza de estos procesos es uno de los principales objetivos en la era de la proteómica. El conocimiento completo de las redes de interacción proteica no solo es importante para comprender los procesos celulares y sus implicancias desde un punto de vista biológico; sino también por los fines aplicados que de él se desprenden, como son el tratamiento de muchas enfermedades y desórdenes que tienen base en defectos en IPP, diseño de drogas, tratamientos anti-virales y contra otros patógenos, etc.

El descubrimiento de nuevos compañeros interactuantes para una proteína de interés puede proveer evidencia significativa sobre su rol funcional. La identificación de proteínas de unión con un rol celular definido normalmente lleva a la investigación en nuevas direcciones. Se han desarrollado múltiples técnicas para la identificación de IPP. Generalmente, se utiliza una proteína “cebo” de interés para buscar en un pool de proteínas celulares posibles interacciones, acoplado con algún modo de detección. Una técnica comúnmente utilizada, que se detallará a lo largo de este capítulo, es el sistema de doble híbrido de levadura. Esta técnica utiliza la activación transcripcional de un gen reportero en la levadura como lectura de IPP.

En este capítulo se comentarán, en primer lugar, aspectos generales de la interacción entre proteínas, como la existencia de dominios de interacción, los distintos aspectos que caracterizan una interacción proteica, la diversidad de estas interacciones, etc. Después se mencionarán y desarrollarán diversas técnicas utilizadas para la detección, *screening*, de interacciones entre proteínas. Asimismo, se comentarán brevemente las herramientas de

predicción *in silico* actualmente disponibles, las cuales poseen un alto valor predictivo que luego debe contrastarse con la experimentación. A continuación, se detallarán metodologías que sirven para la confirmación de las IPP tanto *in vitro* como *in vivo*.

El presente capítulo no pretende ser más que una introducción (por cierto muy limitada) a la identificación, confirmación y caracterización biológica de la interacción entre proteínas. El lector interesado puede profundizar en cada uno de los temas tratados en la vasta y creciente bibliografía existente. Vale decir que la selección de los temas tratados es de alguna manera “forzada”, porque sería realmente muy difícil abarcar y resumir todos los aspectos de las IPP en un solo trabajo, esto implicaría mencionar solo cuestiones superficiales y muy generales de cada tema, en detrimento de aspectos esenciales.

Tipos de interacción proteína-proteína.

Las IPP, fundamentalmente, podemos clasificarlas en estables o transitorias, y ambos tipos pueden ser tanto fuertes como débiles. Las interacciones estables son aquellas asociadas con proteínas que son purificadas como complejos de múltiples subunidades, siendo las subunidades de estos complejos idénticas o diferentes. La hemoglobina y la RNA polimerasa son ejemplos de interacciones entre múltiples subunidades que forman complejos estables (Golemis, 2002).

Las IPP que controlan la mayoría de los procesos celulares se espera sean de carácter transitorio. Como el nombre lo indica, las interacciones transitorias son de naturaleza temporal y típicamente requieren de condiciones que promuevan la interacción, tales como fosforilación, cambios conformacionales o localización en un área discreta de la célula. Este tipo de interacciones pueden ser tanto débiles como fuertes, rápidas o lentas. Mientras están en contacto con sus compañeros interactuantes, las proteínas que interactúan transitoriamente se encuentran involucradas en un amplio rango de procesos celulares, incluyendo modificación de la proteína, transporte, plegado, señalización, y ciclo celular.

Las proteínas interactúan entre sí mediante una combinación de interacciones hidrofóbicas, fuerzas de van der Waals, y puentes salinos en dominios de unión específicos de cada proteína. Estos dominios pueden ser pequeñas o grandes superficies de unión y puede tratarse de unos pocos péptidos de longitud o expandirse a lo largo de cientos de aminoácidos, y la fuerza de unión está influenciada por el tamaño del dominio de unión. Un dominio de superficie común que facilita la estabilización de IPP son las cremalleras de leucina, las cuales consisten en α -hélices en cada proteína que se unen entre sí paralelamente a través de interacciones hidrofóbicas entre los residuos de leucina regularmente espaciados en cada α -hélice proyectada en las cadenas peptídicas adyacentes. Debido a su ajustado empaquetamiento molecular, las cremalleras de leucina proveen uniones estables en el caso de complejos multiprotéicos (por ej. el heterodímero Fos/Jun y su unión al DNA), aunque todos los dominios de este tipo no se unen de modo idéntico debido a los residuos no-leucina en la α -hélice reduciendo el empaquetamiento molecular y en consecuencia, la fuerza de la interacción.

Los dominios homólogos Src (*Proto-oncogene tyrosine-protein kinase Src*), SH2 (*Src Homology 2*) y SH3 (*Src Homology 3*), son ejemplos de dominios de unión transitorios que unen secuencias de péptidos cortos y se encuentran comúnmente en proteínas involucradas en la señalización. El dominio SH2 reconoce secuencias peptídicas que poseen residuos de tirosina fosforilados, comúnmente indicativos de activación proteica. El dominio SH2 juega un rol fundamental en la señalización de factores de crecimiento. La fosforilación de residuos de tirosina en el receptor luego de la unión del ligando recluta efectores *downstream* que reconocen estos residuos mediante sus dominios SH2. Los dominios SH3 reconocen secuencias peptídicas ricas en prolina y son comúnmente usados por quinasas, fosfolipasas y GTPasas para identificar proteínas blanco. Aunque tanto SH2 como SH3 generalmente se unen a estos motivos, la especificidad en las distintas interacciones proteicas está dada por los aminoácidos vecinos a los respectivos motivos.

Más allá de las diferencias entre los distintos dominios y las especificidades de ligando; todas las interacciones implican la habilidad de un dominio de

interacción plegado de reconocer un motivo particular en otro péptido. En este punto cabe comentar dos características importantes de los dominios de interacción: su relativa versatilidad, y el hecho que frecuentemente varios dominios de interacción diferentes se encuentran unidos covalentemente en una misma cadena polipeptídica, generando una proteína que puede mediar múltiples IPP distintas.

Efectos biológicos de las IPP.

El resultado de 2 o más proteínas que interactúan con un objetivo funcional puede demostrarse de diversos modos. Los efectos medibles de las interacciones entre proteínas se pueden resumir del siguiente modo:

- Alterar las propiedades cinéticas de las enzimas, que puede ser el resultado de cambios sutiles en la unión al sustrato o efectos alostéricos.
- Permitir canalizar un sustrato mediante su movimiento a través de dominios o subunidades, resultando en un producto final.
- Crear un nuevo sitio de unión, típicamente para pequeñas moléculas efectoras.
- Inactivar o destruir una proteína.
- Cambiar la especificidad de la proteína por su sustrato a través de su interacción con diferentes compañeros de unión, por ejemplo, exhibiendo una nueva función que ninguna proteína evidencia sola.
- Actuar de modo regulatorio en un evento *upstream* o *downstream*.

Métodos experimentales para el análisis de IPP: niveles de confianza.

Los métodos experimentales disponibles para detectar interacciones entre las proteínas tienen distintos niveles de resolución (Tabla 11.1) y pueden clasificarse en cuatro categorías. La primera categoría comprende el análisis atómico, donde la interacción es detectada por métodos muy sensibles como la

difracción de rayos-X 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 SPR (del inglés *Surface Plasmon Resonance*). 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 (Torre Russis, 2003). Algunos métodos experimentales abarcan más de un nivel de resolución como se observa en la Tabla 11.1.

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	
Y2H		X	X	
Cromatografía de afinidad (Pull Down)		X	X	
SPR		X	X	
Micro-arreglo de proteínas		X	X	
Resonancia paramagnética electrónica	X	X	X	
Cromatografía de filtración por gel		X	X	
Espectrometría de masas		X	X	
Enlazamiento molecular (crosslinking)		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
Ensayo de bloqueo con Ac. monoclonales			X	X
Ensayos de adhesión				X
Knock-out				X
Co-expresión transitoria				X

Tabla 11.1. Con una cruz (X) se indican los niveles de resolución de cada uno de los métodos de análisis de IPP.

A lo largo del presente capítulo sólo detallaremos algunas de las técnicas experimentales usadas al momento de realizar un ensayo de *screening* de IPP, y algunas técnicas de confirmación *in vivo*. Nos pareció importante mostrar la presente tabla dado que realiza una enumeración bastante completa de la multiplicidad de técnicas disponibles para el estudio de IPP desde diversos puntos de vista.

11.1. Estudio *in silico* de interacciones proteína-proteína

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 así como la información publicada en la literatura científica. A continuación se expone una revisión acotada de algunas de las bases de datos disponibles, centrando la atención en el tipo de datos que contienen y cómo acceder a ellas. Es importante conocer la existencia de estas herramientas al momento de adentrarse en el estudio de las IPP de un determinado sistema.

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 (Xenarios, 2001).

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 (Bader, 2003).

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 (Ng, 2003).

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. 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 (Ji, 2003).

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 (Zanzoni, 2002).

WISTdb (Worm Interaction Sequence Tag database). Contiene el mapa de las interacciones proteicas del nematodo *C. elegans* obtenidas por doble-híbrido (Stein, 1999).

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* (Sanchez, 1999).

MYGD (MIPS Yeast Genome Database), <http://mips.gsf.de>. MYGD contiene información detallada sobre el genoma de *S. cerevisiae*, y las secuencias aminoacídicas 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 (Mewes, 2002).

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 (Costanzo, 2001).

PIM, <http://www.hybrigenics.fr>. Contiene los resultados del análisis por doble-híbrido del genoma de *H. pylori* (Rain, 2001).

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. 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”) (Eilbeck, 1999).

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. 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. Sin embargo, existen casos en los que proteínas homólogas poseen diferentes modos de interacción. Por ejemplo, la enzima nucleótido difosfato quinasa se presenta en la naturaleza en forma de hexámero o tetrámero dependiendo de la especie. Las sialidasas virales forman tetrámeros, mientras que las bacterianas y de tripanosoma aparecen como monómeros de proteínas multidominio.

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. A partir de ese año y hasta el 2002, han aparecido diferentes trabajos sobre el tema, de varios autores. De estos trabajos se ha derivado un conjunto de reglas 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 capítulo.

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 subunidades 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.

El método desarrollado por Pellegrini *et al.* (Pellegrini, 1999), denominado perfiles filogenéticos (del inglés “*phylogenetic profiles*”), se basa en la coevolución 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.* (Huynen, 2000) 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.* (Marcotte, 1999) y Enright *et al.* (Enright, 1999) 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. El método desarrollado por Marcotte y colaboradores (Marcotte, 1999), conocido como “*Rosetta stone method*”, utiliza la información anotada en las bases de datos Pfam (Bateman, 2002) y ProDom (Corpet, 2000) 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 (Gallet, 2000) 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.

Bock y Gough (Bock, 2001) utilizan la información contenida en 70 heterocomplejos 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 (Wojcik, 2001) 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.

Pazos y Valencia (Pazos, 2001; Pazos, 2002) 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.

Lu *et al.* (Lu, 2002) 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.

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

- <http://www.biochem.ucl.ac.uk/bsm/PP/server/>
- maine.ebi.ac.uk:8000/services/allfuse/
- <http://www.sanger.ac.uk/Software/Pfam/>
- gpcr.biocomp.unibo.it/predictors
- www.russell.embl.de/interprets
- <http://cbm.bio.uniroma2.it/iSPOT>

El objetivo de esta sección del capítulo fue pensado para nombrar y explicar brevemente algunas de las múltiples herramientas actualmente existentes, bases datos y métodos de predicción de IPP, que complementan las metodologías experimentales desarrolladas. La sumatoria de la información proveniente de múltiples fuentes tiene como finalidad colaborar para completar las redes de interacción de proteínas en los diversos sistemas de estudio.

11.2. Técnicas de screening.

Sistema de doble híbrido en levaduras: “Y2H (Yeast two Hybrid) System”.

Esta técnica fue primero descrita en 1989 por Fields y Song (Fields, 2009). Durante estos años se ha enriquecido y dado lugar a diversas variantes. Se trata de una técnica ampliamente utilizada ya que permite de una forma relativamente simple, descubrir y comprobar interacciones entre proteínas en un entorno más cercano a las condiciones fisiológicas, a diferencia de ensayos *in vitro*. El fundamento de la técnica implica dos dominios funcionales de un factor de transcripción que pueden ser separados físicamente, los cuales en proximidad uno con el otro gracias a proteínas interactuantes, cada una fusionada a uno de los dominios del factor de transcripción, reconstituyen la capacidad de éste de regular la expresión génica (activando genes reporteros

que varían de acuerdo al sistema y la cepa de levadura utilizados) (Westermarck, 2013).

Aplicaciones:

- Identificación de nuevas proteínas interactuantes
- Confirmación de proteínas candidatas
- Reconocimiento de dominios de interacción dentro de una proteína

Este sistema aprovecha la flexibilidad de los factores de transcripción de levaduras: Gal4 y LexA. El factor de transcripción Gal4 se encuentra dividido en dos proteínas: DNA-BD (dominio de unión de ADN) y DNA-AD (dominio de activación de ADN). Estos dos dominios son incapaces de interactuar entre sí o de activar la expresión génica por sí solos. Los mismos se fusionan por separado a dos plásmidos diferentes (Gietz, 1997):

1) Plásmido para proteína cebo (*bait*) que contiene gen para la expresión de triptófano (permitirá una posterior selección de levaduras en un medio carente de este aminoácido). En éste se insertará en marco la construcción *DNA-BD:proteína bait*.

2) Plásmido para proteína presa (*prey*) que contiene gen para la expresión de leucina (permitirá una posterior selección de levaduras en un medio carente de este aminoácido). En éste se insertará en marco la construcción *DNA-AD:proteína prey*. Las proteínas *prey* en un ensayo de tamizaje (*screening*) se incorporarán como una biblioteca de cDNA, la cual se preparará a partir de ARNm del organismo, línea celular o tejido de interés a probar. Cuando se trata de un ensayo de confirmación, la proteína *prey* será el cDNA de aquella identificada por experimentos previos como candidata.

En la Figura 11.1 se muestra brevemente la metodología de screening Y2H. Al comienzo se tienen dos cepas haploides de levadura *Saccharomyces cerevisiae*, cada una contiene uno de los plásmidos mencionados anteriormente. Se cultivan en medios carentes de triptófano o leucina, respectivamente. Luego se mezclan colonias de los dos tipos de cepas y se cultivan en medio sin triptófano ni leucina. Producto del apareamiento entre ambas cepas se generarán zigotos diploides capaces de crecer en el medio que carezca de ambos aminoácidos. En las células diploides se expresará

tanto DNA-BD: proteína *bait* como DNA-AD: proteína *prey*. Si las proteínas bait y prey interactúan, la construcción DNA-BD:*bait* unida a la secuencia de activación *upstream* (UAS) del promotor de GAL1 se fusionará con la construcción DNA-AD:*prey* y se estimulará la transcripción de genes reporteros *downstream* del promotor (por ej.: HIS3, lacZ, etc.). Esto permitirá distinguir las colonias que presenten interacción de proteínas *bait:prey* a través de medios selectivos, por ejemplo carentes de histidina o con actividad beta galactosidasa volviéndose azules en presencia de 5-bromo-4-cloro-3-indolil-b-D-galactopiranosido (X-Gal) (Matchmaker™ Gold Yeast Two-Hybrid System User Manual Clontech, 2010).

Sistema de doble híbrido en levadura

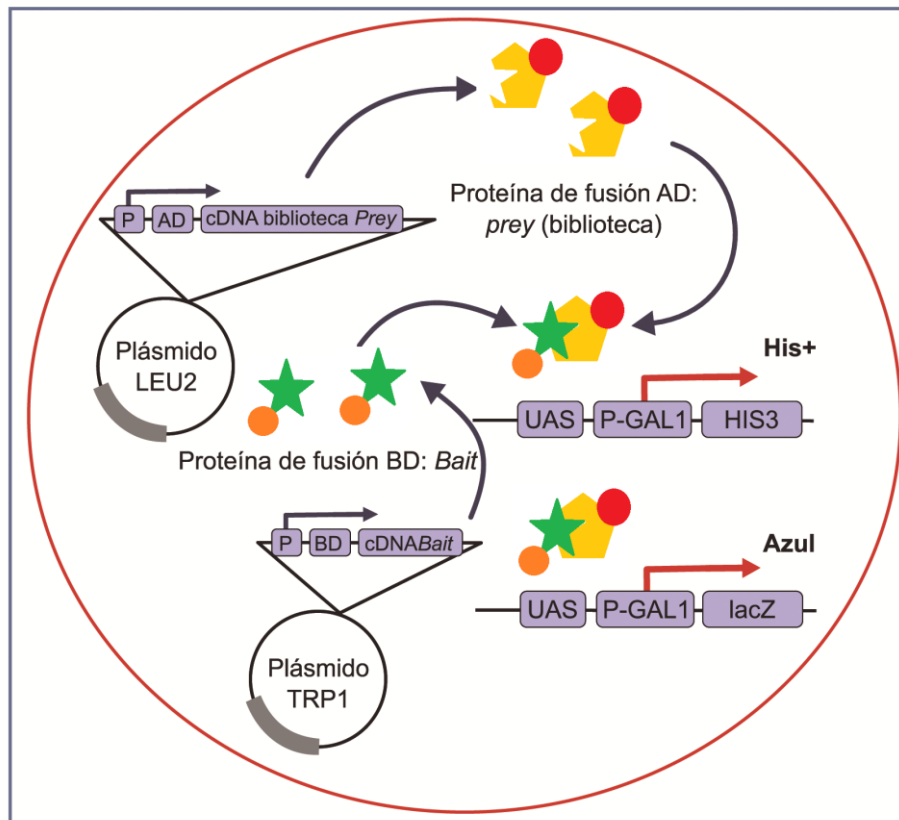


Figura 11.1. Diagrama del Sistema de Doble Híbrido en levadura. *P*: promotor, *AD*: dominio de activación, *BD*: dominio de unión, *UAS*: secuencia de activación río arriba, *lacZ/HIS3*: genes reporteros, *Plásmido LEU2*: para biblioteca prey con expresión de leucina, *Plásmido TRP1*: para proteína bait con expresión de triptófano.

Se deben incluir en el ensayo:

1) Control positivo: dos construcciones (DNA-BD:proteína X y DNA-AD:proteína Y) cuyas proteínas es sabido por literatura que interactúan. Deben dar colonias positivas en medios selectivos y activación de genes reporteros.

2) Control negativo: dos construcciones (DNA-BD:proteína X y DNA-AD:proteína Z) que es sabido por literatura que no interactúan. Deben dar negativo para la activación de genes reporteros.

Una vez identificada una proteína candidata, se puede repetir el ensayo realizando la fusión de su cDNA a DNA-AD. También es posible identificar cuál es el dominio de interacción con la proteína *bait* mediante mutaciones puntuales y deleciones en el cDNA de la proteína *prey* identificada.

Tipos de falsos positivos posibles (Gietz, 1997):

1) Productos de fusión DNA-AD:*prey* que activen la transcripción de genes reporteros sin interactuar con DNA-BD:*bait*.

2) Activación de genes reporteros en presencia de cualquier fusión DNA-BD que no es con la proteína *bait* de interés.

3) Activación de genes reporteros en presencia de un vector con DNA-BD vacío.

Debido a que puede existir la presencia de falsos positivos los resultados obtenidos mediante este screening deben ser complementados y confirmados por otras técnicas como: ensayos de *pull-down*, *Mammalian Two Hybrid*, *Yeast three Hybrid* (para complejos ternarios), Coimmunoprecipitación, FRET-FLIM, SPR, etc.

Limitaciones de la técnica y soluciones posibles (Westermarck, 2013):

1) Las levaduras no expresan per se tirosina kinasas (Tyr-K), por lo que en principio no se podría aplicar a proteínas que requieran fosforilaciones. Esto se puede solucionar generando la expresión ectópica de Tyr-K en levaduras.

2) Las proteínas utilizadas como *bait* no pueden exhibir actividad de transactivación en el sistema de genes reporteros.

3) La interacción tiene lugar en el núcleo con lo que, por ej., proteínas integrales de membrana no pueden ser estudiadas. Existen para solucionar este inconveniente técnicas como el *Split-ubiquitin system* que modifica el sistema de doble híbrido en levadura. Brevemente la molécula de ubiquitina se

divide en dos mitades y estas se fusionan a las proteínas *bait*:factor de transcripción y *prey*. Si interactúan, la molécula de ubiquitina se reconstituye, una deubiquitinasa en la levadura corta el fragmento de ubiquitina:factor de transcripción de la proteína *bait* y ese fragmento puede activar en el núcleo la transcripción de genes reporteros.

Ensayos de pull down.

Implica métodos *in vitro* utilizados para determinar interacciones físicas entre dos o más proteínas. Los ensayos de *pull down* se utilizan tanto para *confirmación* como para *identificar* nuevas proteínas interactuantes. El requerimiento mínimo para estos ensayos es la disponibilidad de una proteína *bait* purificada y, en general, marcada. La cual se utilizará para capturar proteínas interactuantes (*prey*) (Thermo Scientific Pierce Protein Interaction Technical Handbook Version 2. 2010).

Los ensayos *pull down* son una forma de purificación por afinidad y son muy similares a la inmunoprecipitación (ver más adelante), excepto que una proteína *bait* es utilizada en vez de un anticuerpo. Pueden proveer una mayor resolución y selectividad que otros ensayos basados en anticuerpos.

La proteína *bait* marcada es capturada en una matriz por la cual es afín, lo cual genera un soporte de afinidad secundario para purificar otras proteínas que interactúan con la proteína *bait*.

Para ensayos de confirmación se pueden utilizar sistemas de expresión en células de diversos orígenes, proteínas purificadas, productos de reacciones de transcripción/traducción.

Las fuentes proteicas para *screening* de proteínas interactuantes se encuentran generalmente en una mezcla compleja considerada como el ambiente nativo de la proteína *bait*. Cualquier lisado celular o fluido biológico en donde la proteína *bait* es normalmente expresada, constituye una fuente apropiada de proteínas *prey*.

La marca de la proteína *bait* puede ser generada por marcación *in vitro* de proteína purificada (por ej. sulfo-NHS-Biotina) o por expresión de proteínas recombinantes fusionadas a una etiqueta (por ej. oligoHis/poliHis o glutatión S-transferasa, GST).

Aquellas *interacciones estables* entre proteínas son más simples de aislar por métodos físicos como estos ensayos dado que el complejo proteico no se desensambla con el tiempo (poseen una baja constante de disociación). A ese tipo de complejos se los puede someter a lavados con *buffers* (soluciones amortiguadoras) de elevada fuerza iónica con el fin de eliminar falsos positivos por interacciones inespecíficas. Dependiendo de la constante de disociación que presente el complejo, se pueden ajustar las condiciones del ensayo (como el pH, concentración de sales, etc.).

En cambio las proteínas de *interacciones transitorias o lábiles* son más difíciles de identificar utilizando métodos físicos. Esto se debe a que el complejo puede disociarse durante el ensayo. Como en general se dan en proteínas de transporte y enzimas, suelen requerir la incorporación de cofactores y sustratos energéticos. Este tipo de interacciones pueden ser fortalecidas con la incorporación de *crosslinkers* (ver más adelante) que unen covalentemente el complejo previo al ensayo de *pull down*.

Para identificar una nueva interacción, la proteína *prey* desconocida debe estar presente en cantidades suficientes para permitir la visualización de la interacción por el método de detección elegido. Lisados celulares radiomarcados con ^{35}S constituyen una fuente frecuentemente utilizada de proteína para estos ensayos.

Por otro lado se puede eluir selectivamente las proteínas *prey*, manteniendo la proteína *bait* inmovilizada. Se puede lograr mediante un gradiente creciente de concentración de sales o disminuyendo el pH. Mediante esta técnica se evita la desnaturalización proteica y puede brindar información sobre la fuerza iónica relativa.

Las muestras eluidas pueden analizarse mediante SDS-PAGE y Western Blot o autorradiografía. Para luego aislar la banda de interés del gel de poliacrilamida

y realizar sobre esa muestra una *identificación por espectrometría de masa* (MS).

Controles de los ensayos pull-down (Thermo Scientific Pierce Protein Interaction Technical Handbook Version 2. 2010):

Control negativo: soporte de afinidad sin tratar con proteína bait. Sirve para identificar falsos positivos debido a uniones inespecíficas entre el soporte y las proteínas de la muestra.

Control sin muestra: soporte de afinidad tratado con proteína *bait*, sin incorporar la muestra (proteínas *prey*). Elimina falsos positivos por unión inespecífica a la marca de la proteína *bait*. Funciona también como control positivo para verificar que el soporte captura adecuadamente la proteína *bait* marcada.

Dominio SH2 (del inglés Src Homology 2)

Está técnica permite la medición de eventos transitorios como los estados de fosforilación y señalización celular. A diferencia de la co-inmunoprecipitación (tratada más adelante), es independiente de los anticuerpos que en el caso de fosfoproteínas suelen presentar una elevada señal de fondo y baja recuperación de proteínas.

SH2 es un dominio de proteínas estructuralmente muy conservado en la oncoproteína Src y en muchas otras proteínas críticas para la señalización celular desde receptores asociados a tirosina quinasas. Los dominios SH2 sirven para separar proteínas fosforiladas de mezclas complejas como lisados celulares. Reconocen y unen residuos específicos de tirosina fosforilada (pY) presentes en muchas proteínas de señalización. Estas interacciones, SH2-pY, son cruciales para la transmisión de la señal desde el receptor hacia el resto de las moléculas de la vía (Thermo Scientific Pierce Protein Interaction Technical Handbook Version 2. 2010).

Se encuentran conformados por alrededor de 100 aminoácidos y su unión a las proteínas blanco depende de la fosforilación de residuos específicos de tirosina en secuencias aminoacídicas consenso.

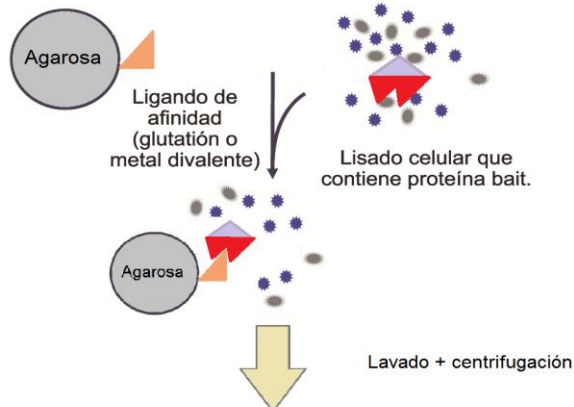
Fusión a GST, Biotina y poli-His

Esta técnica utiliza la afinidad de *GST* (glutación S transferasa) por el glutatión (unido a un soporte o matriz) para purificar proteínas interactuantes de una mezcla proteica. La proteína *bait* fusionada a GST es expresada y purificada de bacterias. La misma es incubada junto con la mezcla proteica a probar ya sea en el formato *batch* (seguida de centrifugación) o columna (con posterior elución). El complejo puede ser eluido de la matriz con glutatión libre (no unido, en solución) o con buffer en condiciones reductoras y desnaturizantes (ver Figura 11.2). Es una técnica especialmente útil para probar interacciones en solución que pueden no ser detectadas en un ensayo como Far Western Blot, realizado sobre membranas de nitrocelulosa o PVDF.

Las proteínas marcadas con *poli-histidina* (6 residuos Histidina) se unen a una resina de agarosa derivatizada con ácido nitrilo triacético (NTA-Agarosa por sus siglas en inglés) cargada con iones divalentes (níquel, cobalto, zinc, cobre, hierro). Las proteínas contaminantes pueden ser removidas con lavados apropiados, según las condiciones del ensayo. La proteína marcada podrá luego ser eluida mediante un quelante soluble (Imidazol o EDTA, siendo el primero más selectivo). Generalmente la marca de 6-His se coloca por biología molecular en el extremo N- o C-terminal, siempre y cuando sea un sitio expuesto y flexible. Se pueden incorporar en esta marca sitios de corte únicos para una proteasa, pudiendo eliminarse la cola de His luego de la elución o puede utilizarse como herramienta para despegar la proteína de la matriz. Una proteína que interacciona con aquella marcada con poli-His puede ser eluida selectivamente siempre que las proteínas contaminantes que se unen a la matriz no se liberen de la resina bajo esas condiciones de elución (ver Figura 11.2).

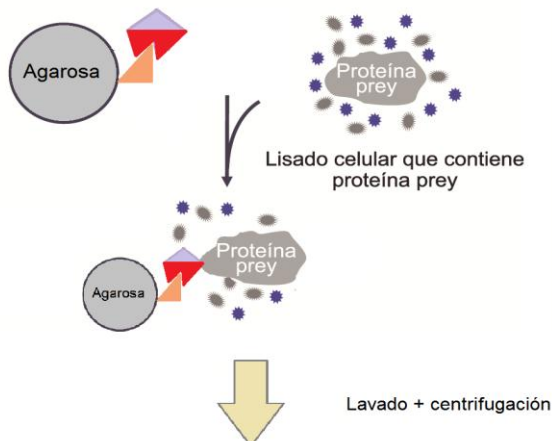
En el caso de *proteínas biotiniladas*, cuya marca es también lograda por biología molecular, se unen por afinidad a una matriz que contiene estreptavidina. Luego se coloca una solución que contiene potencialmente proteínas interactuantes con la proteína biotinilada. Las proteínas *prey* son eluidas mediante una solución de bajo pH.

Inmovilizar la proteína de fusión bait con GST o poliHis del lisado celular.

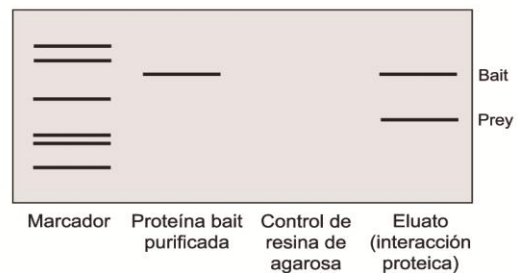


Ensayo de *Pull-Down* utilizando GST o PoliHis

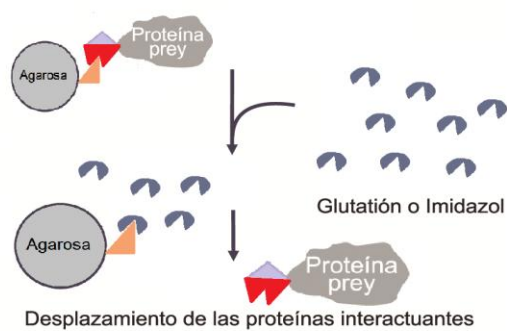
Unir la proteína prey a la proteína bait inmovilizada



Analizar la interacción proteína-proteína mediante SDS-PAGE.



Eluir complejos de interacción proteína-proteína.



= GST o polihistidina

Figura 11.2. Esquema del sistema de pull down utilizando una proteína bait marcada con GST o polihistidina.

En general todos estos métodos requieren del análisis por *espectrometría de masa* con el fin de descubrir cuáles son las proteínas que fueron eluidas en cada tipo de ensayo de pull-down.

Ventajas respecto al Far Western Blot (descrito a lo largo del capítulo): La proteína *bait* es incubada con sus potenciales proteínas interactuantes en un ambiente más cercano a las condiciones nativas, por lo tanto aumentando la eficiencia de las interacciones (Thermo Scientific Pierce Protein Interaction Technical Handbook Version 2. 2010).

Crosslinking (unión covalente por compuestos químicos).

En los sistemas biológicos, cuando dos o más proteínas tienen afinidad entre sí se produce un acercamiento y unión entre las mismas. En general estas uniones son transitorias y ocurren brevemente para promover un evento de señalización o reacción metabólica dada. Con la finalidad de poder estudiar estos complejos proteicos se busca capturarlos o congelarlos para estudiar cuales son las proteínas involucradas y cómo interactúan. Estas uniones creadas artificialmente en el laboratorio son de tipo covalente y garantizan la estabilidad del complejo para permitir su aislamiento y caracterización.

La reacción que llevan a cabo los *crosslinkers* es sobre grupos nucleofílicos ya sea en los extremos o en la cadena lateral. Con sólo pocas excepciones, son las cadenas laterales de los aminoácidos polares ionizables (arginina, tirosina, lisina, cisteína, histidina, ácido aspártico y glutámico) las que son susceptibles al ataque

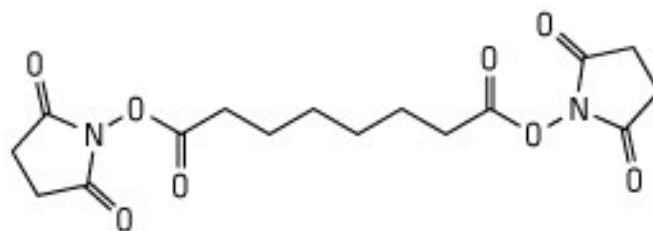
de los *crosslinkers*. No todos los residuos de este tipo son igualmente reactivos, sino que depende de la localización del mismo en la proteína (accesibilidad, polaridad, interacciones con otros residuos o con el solvente, etc.). Como consecuencia, el *crosslinking* debe ser considerado como un procedimiento empírico, debido a las diferencias en reactividad de los grupos involucrados (Golemis, 2002).

Los más comúnmente utilizados son bifuncionales ya que contienen dos grupos reactivos que reaccionan con dos cadenas laterales para generar un complejo unido covalentemente. La parte que funciona como puente entre dichos grupos reactivos se le llama espaciador. El mismo delimita la distancia entre los grupos reactivos y confiere otras propiedades (por ej. un disulfuro escindible). Además el espaciador contribuye a la geometría y a la solubilidad del *crosslinker*. Este brazo espaciador puede ir entre cero (por ej. carbodiimidas: EDC) y varios átomos de longitud (por ejemplo: DSS, ABH, etc.). Los compuestos homobifuncionales (Figura 11.3) contienen grupos funcionales idénticos a ambos lados del espaciador mientras que en los heterobifuncionales son diferentes. Estos últimos tienen ciertas ventajas como permitir el ataque a distintos grupos nucleofílicos proteicos y permitir dividir la etapa de *crosslinking* en dos pasos separados, en diferentes condiciones. Un ejemplo de ellos son los *crosslinkers* fotoactivables que combinan reactividad química con grupos que se activan frente a la exposición lumínica de longitud de onda adecuada (por ej. NHS-Diazirina). Estos últimos limitan los artefactos generados por uniones inespecíficas pero disminuyen el rendimiento, además pueden reaccionar con el solvente, “quencheando” el intermediario fotoreactivo. A su vez existen *crosslinkers* con marca isotópica que incorporan deuterio en posiciones determinadas lo cual permite marcar las proteínas además de unirlos covalentemente. De esta manera se simplifica la identificación por MS. Los *crosslinkers* son utilizados generalmente para dos fines en la deducción de las interacciones proteicas (Phizicky, 1995):

- 1) Para identificar la estructura de un complejo aislado.
- 2) Para detectar proteínas interactuantes en extractos celulares, células o purificados parciales.

En cuanto a la determinación de la arquitectura del complejo aislado, se utilizan *crosslinkers* del tipo RSSR' que contienen un enlace disulfuro escindible por agentes reductores. Se lo deja actuar formando un aducto P-RSSR'-P' (siendo P y P' las proteínas del complejo), con posterior fraccionamiento por electroforesis en SDS-PAGE sin agentes reductores. Luego se corre una segunda dimensión del gel previo tratamiento con agentes reductores que

corten el enlace S-S. Aquellas especies no unidas por el *crosslinker* no modifican su peso molecular (PM) respecto a la primer dimensión, mientras que las que fueron unidas por el *crosslinker* darán bandas de menor PM correspondientes a las de las especies P y P' puras. Cuando en la detección se utilizan anticuerpos específicos para las especies involucradas se podría evitar la segunda dimensión del gel.



DSS
Disuccinimidil Suberato
PM 368.34
Espaciador 11.4 Å

Figura 11.3. Estructura del DSS como ejemplo de *crosslinker* homobifuncional.

Para la detección de proteínas interactuantes existen dos aproximaciones:

In vivo: Se consigue mediante la utilización de *crosslinkers* hidrofóbicos capaces de atravesar la membrana celular seguido de inmunoprecipitación de la proteína *bait* (lo cual incrementa la sensibilidad de esta última técnica).

In vitro: se basa en la adición de una proteína pura a un sistema proteico complejo al cual se le agrega un *crosslinker* determinado. La posterior detección del complejo se simplifica en gran medida cuando la proteína es marcada antes del *crosslinking*, ya que de esta manera existe sólo una fuente de material marcado para detectar (poliHis, GST, isótopos radiactivos, etc.)

La simple adición de *crosslinkers* homo- o heterobifuncionales a las suspensiones celulares o a sus homogenatos causará la formación de gran cantidad de conjugados proteicos, no todos estarán involucrados en interacción proteica real, algunas simplemente resultan unidas por encontrarse en la vecindad al momento de reaccionar con el *crosslinker*. El desafío de esta

técnica radica en el análisis de los resultados una vez que los complejos han sido aislados.

Coinmunoprecipitación.

El fundamento de esta técnica se basa en la generación de un lisado celular al cual se le agrega anticuerpo, se precipita el antígeno, se lava y las proteínas unidas son eluidas y analizadas por SDS-PAGE y Western Blot. El antígeno utilizado para obtener el anticuerpo puede ser una proteína purificada o un péptido sintético acoplado a un transportador. El anticuerpo puede ser mono- o policlonal. Alternativamente, si la proteína tiene una marca incorporada como c-Myc, GST o poliHis, existen anticuerpos comerciales disponibles para detectarla. Este procedimiento puede realizarse en *batch* o en columna, en ambos casos la resina generalmente contiene proteína A o G, las cuales unen la fracción Fc de anticuerpos tipo IgG (Phizicky, 1995).

La inmunoprecipitación y la coinmunoprecipitación comparten el mismo principio, sin embargo en esta última el anticuerpo al precipitar el antígeno, co-precipita la/s proteína/s interactuante/s con la proteína contra la que está dirigido. Es posible que esta técnica no sea útil cuando el antígeno o proteína *bait* presenta interacciones transitorias con otras proteínas del entorno celular, las cuales serían arrastradas con los lavados previos a la elución final. Es en estos casos que la incorporación de *crosslinkers* sería útil para la captura de complejos transitorios.

Se asume usualmente que las proteínas co-precipitadas están relacionadas en cuanto a su función a nivel celular, pero esto debe ser verificado y demostrado por otros métodos.

Uno de los inconvenientes de esta técnica es que suelen existir interferencias en el análisis del gel producidas por bandas del anticuerpo que se eluye junto con el complejo proteico. Una solución en este sentido es la inmovilización *Directa* a través de los grupos amino superficiales del anticuerpo a un soporte de agarosa activado con N-hidroxi succinimida (NHS). En la técnica *Indirecta*

se utiliza un *crosslinker* como DSS para unir químicamente el anticuerpo a la proteína A o G de la resina. En el mismo sentido si se tiene un anticuerpo biotinilado puede ser unido a una resina de agarosa con estreptavidina inmovilizada. De esta forma el anticuerpo se une fuertemente a la matriz y no es arrastrado en las condiciones de elución del antígeno (Thermo Scientific Pierce Protein Interaction Technical Handbook Version 2. 2010).

En la coimmunoprecipitación tradicional el complejo inmune se forma en solución antes de precipitarlo con proteína A o G inmovilizada. En cambio cuando se utiliza anticuerpo inmovilizado la formación del complejo inmune y su precipitación ocurre en un único paso.

Al momento de analizar los datos obtenidos es importante definir si esta interacción ocurre *in vivo* y qué importancia tiene a nivel celular. Con el fin de verificar la interacción es importante determinar los siguientes ítems (Phizicky, 1995):

- 1) Confirmar que la proteína co-precipitada se obtiene por el anticuerpo contra la proteína bait/antígeno: el suero policlonal no debe poseer clones que unan directamente otras proteínas del extracto celular. Se puede lograr “purificando” el suero al pasarlo por una columna que contenga la proteína *bait* inmovilizada.
- 2) Determinar que el anticuerpo contra la proteína *bait*/antígeno no reconozca directamente a la/s proteína/s *prey*: se puede lograr descartar tal efecto repitiendo el ensayo con un anticuerpo específico para la proteína *prey*, en cuyo caso se debe obtener el mismo complejo *bait:prey*. Esta técnica puede ser utilizada como una herramienta de confirmación.
- 3) ¿La interacción es mediada por una tercera proteína? Será necesario aplicar MS para descubrirlo.
- 4) Determinar si la interacción ocurre en la célula o es un artefacto de la lisis celular: serán necesarios estudios de co-localización y mutagénesis puntual que perturbe el sitio de unión.

Esta técnica presenta las siguientes *ventajas*:

- 1) Detecta interacciones en una mezcla proteica compleja como es un lisado celular crudo.

2) El antígeno/*bait* y las proteínas interactuantes se encuentran en la misma concentración relativa que en la célula.

3) Los complejos se encuentran en su estado natural para la co-precipitación.

4) Las proteínas se encuentran en su estado postraducciona correspondiente.

A la vez presenta sus *desventajas*:

1) Las proteínas co-precipitadas no necesariamente interactúan directamente, pueden ser parte de complejos mayores.

2) No es tan sensible como otros métodos, por ej. la cromatografía de afinidad a proteína (*pull down*, ver más arriba), debido a que la concentración de antígeno es menor. Esto se puede evitar agregando exceso de antígeno al lisado celular. Otra forma es, como se mencionó previamente, agregando un *crosslinker* en un paso previo a la precipitación.

Far Western Blot.

En el ensayo clásico, una proteína *bait* (marcada o detectable mediante un anticuerpo específico) es utilizada para sondear y detectar la/s proteína/s *prey* en una membrana. La muestra que contiene una mezcla compleja de proteínas es separada por SDS-PAGE o PAGE nativo (colocando un patrón de peso molecular de tipo Rainbow, GE) y luego se transfiere a una membrana. En este formato la/s proteína/s *prey* se vuelve/n accesible/s para ser incubada/s con la proteína *bait*. Luego de la transferencia la membrana es bloqueada y luego se le agrega una solución con proteína *bait* purificada. Un sistema de detección específico para la proteína *bait* es utilizado para identificar las bandas correspondientes (Thermo Scientific Pierce Protein Interaction Technical Handbook Version 2. 2010).

Las proteínas desnaturalizadas pueden no ser capaces de interactuar (falso negativo) o generar interacciones apócrifas (falso positivo). Esto no implica que la identificación de interacciones reales no sea posible sino que deben tomarse recaudos como el uso de controles y una validación apropiada.

El tipo de membrana (por ej. nitrocelulosa o PVDF) utilizado para transferir las proteínas es crítico, debido a que algunas se unen selectiva o preferentemente a un tipo de membrana en particular. La tasa de transferencia proteica es inversamente proporcional a su peso molecular. Para el análisis es esencial que al menos el dominio de interacción de las proteínas en la membrana no esté desnaturalizado. Para lo cual si se desnaturalizó en la transferencia debe ser capaz de plegarse nuevamente para dejar el sitio de interacción intacto. Generalmente un porcentaje significativo de la población de proteínas se recompone luego de la remoción del SDS.

Las interacciones proteicas detectadas dependen de varios factores: la naturaleza de las proteínas, el pH, la concentración de sales y la presencia de ciertos cofactores en la incubación con la proteína *bait*. Algunas interacciones pueden incluso requerir la presencia de otras proteínas. Las condiciones necesarias deben ser mantenidas a lo largo de todo el proceso para preservar la interacción hasta el momento de la detección.

El método de detección puede ser:

- 1) Directo cuando la proteína *bait* tiene marca radiactiva
- 2) Indirecto con un anticuerpo específico para la proteína *bait*
- 3) Indirecto con un anticuerpo contra la marca (GST, poliHis, etc.) de la proteína *bait*
- 4) Indirecto cuando una proteína *bait* está marcada con biotina, mediante el agregado de una enzima (HRP, *horseradish peroxidase*, o PAL: fosfatasa alcalina) unida a estreptavidina.

Es aconsejable correr dos geles idénticos, una vez finalizada la electroforesis, uno de los geles se transfiere y se procede con el Far Western Blot como se indicó previamente y al otro se le realiza Tinción de plata o Coomassie coloidal para poder observar todas las proteínas presentes y corroborar que aquellas detectadas en la membrana también se encuentren presentes en el gel sin transferir.

Es importante incluir controles que permitan distinguir interacciones proteicas verdaderas de artefactos inespecíficos. Con este fin se puede incorporar en la incubación una proteína no relevante a los fines del estudio en vez de la

proteína *bait*. Otro control negativo sería realizar el procedimiento completo, obviando la incorporación de la proteína *bait*. Estos controles deben ser llevados a cabo en paralelo al tratamiento de la muestra.

Existen alternativas al procedimiento clásico (Figura 11.4):

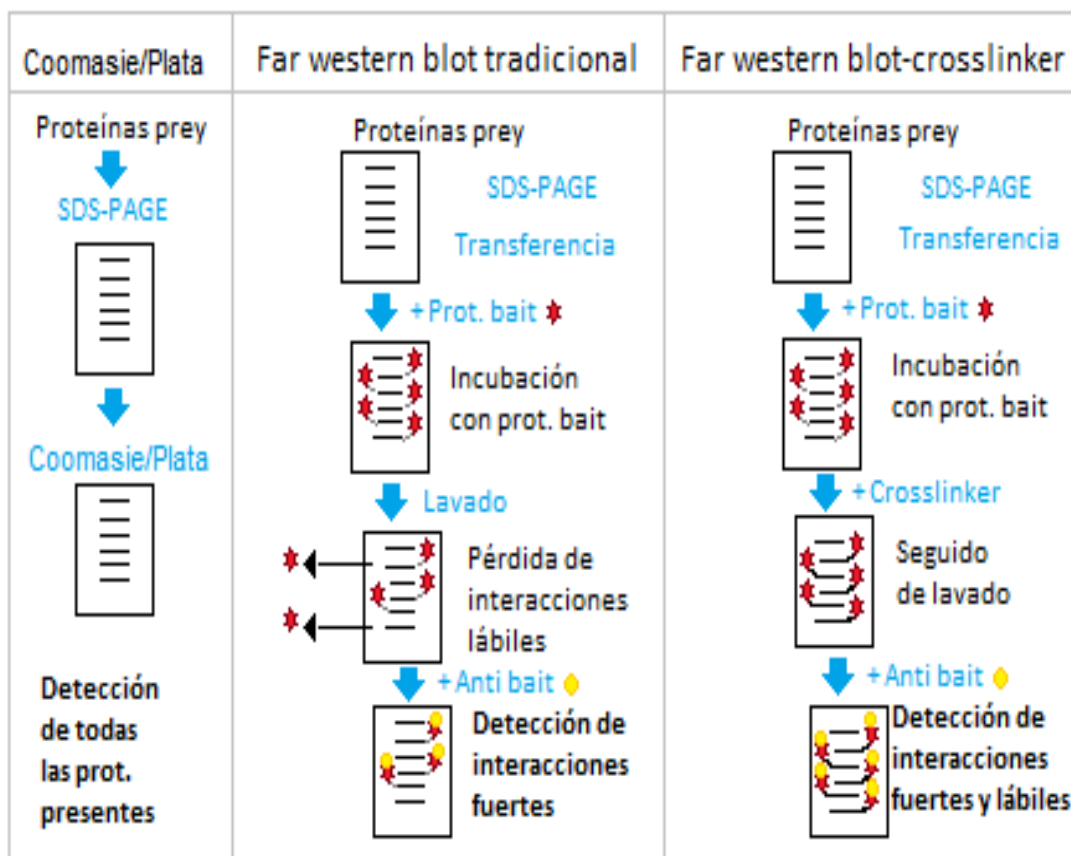


Figura 11.4. Comparación de una técnica de detección general de proteínas (Coomassie o Tinción de Plata) vs. Far Western Blot. Ventajas de la utilización de crosslinkers.

- 1) Far western blot con *crosslinker* (Sato, 2011): Cuando se sospecha que las interacciones proteicas son transitorias, se puede incorporar un paso en el que se agrega un *crosslinker* (por ej. EDC) el cual cimente las proteínas *bait* y *prey* para que en los posteriores lavados no se pierdan y queden sin ser detectadas.
- 2) Far western blot en gel: se procede como en la técnica clásica salvo que no se transfiere a ninguna membrana sino que la interacción se lleva a cabo incubando el gel con proteína *bait*. Para esto es necesario remover el SDS luego de la electroforesis mediante isopropanol 50% lo cual va a permitir el plegamiento correcto de las proteínas. La desventaja es que son necesarias

mayores cantidades de proteínas *bait* y *prey* para que la interacción pueda ser detectada.

11.3. Técnicas de confirmación *in vivo*.

Förster Resonance Energy Transfer.

El uso de herramientas bioinformáticas así como las diversas técnicas de screening previamente explicadas, pueden darnos como resultado un par de proteínas candidatas a interactuar entre sí. El paso siguiente en el estudio de IPP es la confirmación *in vivo* de esta interacción. Una vez establecida la interacción *in vivo*, puede estudiarse la constante de asociación de las proteínas, mapear las áreas de interacción entre las mismas, etc. El objetivo de esta sección del capítulo consiste en explicar dos técnicas de microscopía útiles para la comprobación *in vivo* de IPP.

La microscopía confocal (ver Capítulo 13 Microscopía) es una técnica ampliamente utilizada para el estudio de los componentes intracelulares así como sus asociaciones entre sí. Esta técnica permite el estudio *in vivo* de IPP, y la colocalización es un prerrequisito para que dos especies moleculares interactúen. Una proteína promedio como la albúmina (67 kDa) mide 3x8 nm aproximadamente. El límite de resolución teórico de un microscopio confocal en el eje XY es de ~200 nm. La colocalización de dos proteínas que permite evidenciar esta metodología posee un margen de error muy alto. Dos proteínas marcadas por fluoróforos distintos pueden encontrarse a más de 100 nm entre sí, y sin embargo, mediante la simple observación de las mismas por microscopía, colocalizar en el plano XY. El límite de resolución del microscopio no es lo suficientemente pequeño para aportar información válida que nos permita corroborar la interacción entre dos proteínas que colocalizan.

Afortunadamente, otras técnicas que pueden ser empleadas en células vivas pueden proveer esta información, como FRET (del inglés *Förster resonance*

energy transfer). El fenómeno de FRET implica la transferencia de energía desde un fluoróforo dador a un fluoróforo aceptor adyacente que posee una superposición espectral significativa y orientación apropiadas (Figura 11.5) (Lakowicz, 1999). El fenómeno de FRET es un *quencher* de la fluorescencia muy eficiente (Fricker, 2006). El término *Quenching* de la fluorescencia hace referencia a cualquier proceso que produzca una disminución en la intensidad de la fluorescencia emitida por una determinada sustancia. Una gran variedad de procesos pueden provocar la desactivación de la fluorescencia, entre ellos, la transferencia de energía; en el caso de FRET, desde el dador (que se *quenecha*) al aceptor. El fenómeno de FRET puede detectarse espectralmente como una disminución (*Quenching*) en la emisión del dador con un incremento proporcional en la emisión del aceptor sensibilizado.

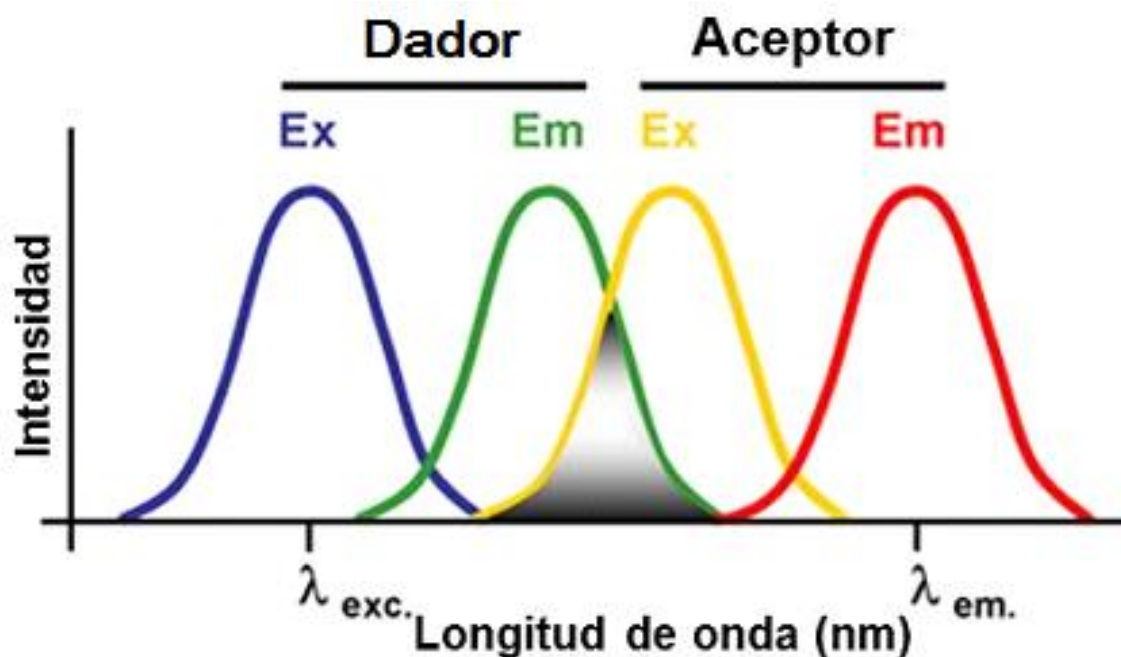


Figura 11.5. La superposición espectral entre el espectro de emisión del dador (espectro verde) y el espectro de excitación del aceptor (espectro amarillo), sumado a la correcta orientación de los fluoróforos en el espacio permiten la transferencia de energía, Esta transferencia de energía es evidenciada como una disminución de la intensidad de fluorescencia del espectro del dador, y un aumento de la intensidad de fluorescencia del espectro del aceptor.

En capítulos anteriores se mencionó la siguiente ecuación de la teoría de Förster:

$$KT(r) = \frac{1}{\tau_D} \left(\frac{R_0}{r} \right)^6 \quad \text{Ec. 11.1}$$

donde KT es la velocidad de decaimiento del estado excitado del dador (en ausencia de aceptor), transferencia de energía de Förster, τ_D es la vida media del donante, R_0 es el radio crítico de Förster, donde la transferencia de energía y el decaimiento espontáneo del estado excitado del dador son igualmente probables, y r es la distancia entre el donante y el aceptor. De esta ecuación se deduce que la eficiencia de FRET disminuye con la inversa de la sexta potencia de la separación entre los fluoróforos. La distancia a la que la energía de transferencia disminuye un 50% (conocida como radio de Förster) es alrededor de 3-6 nm para un par común dador-aceptor.

Uno de los muchos factores que pueden influenciar la medida del tiempo de vida de una población de fluoróforos es la proximidad a una segunda molécula fluorescente con las propiedades espectrales que permitan a ésta absorber la energía de emisión del primer fluoróforo a través de FRET. El decaimiento del tiempo de vida (τ) del primer fluoróforo (dador) puede ser usado como una medida de la interacción FRET con el segundo fluoróforo (aceptor) (Figura 11.6). Esta medida se conoce como FLIM (del inglés *Fluorescence Lifetime Imaging Microscopy*) (Lleres, 2007). La transferencia de energía entre dador y aceptor sólo puede ocurrir cuando las dos moléculas se encuentran muy cercanas (~ 1 a 10 nm), consistente con el hecho de estar molecularmente en contacto. Entonces, debido a esta dependencia a la proximidad, FRET provee un ensayo que permite detectar interacciones entre 2 proteínas que se encuentran marcadas con un dador y aceptor FRET respectivamente (Lleres, 2007).

Para obtener medidas cuantitativas, particularmente la asociación de proteínas en estado estacionario, es necesario tener en cuenta una serie de fenómenos que pueden interferir en la determinación. Se deben realizar controles para corregir el *background*, autofluorescencia (fluorescencia debida a sustancias

presentes en la muestra que son fluorescentes, pero no han sido marcadas intencionalmente para ser observadas), sangrado de los espectros (los canales de detección de la emisión a veces no son lo suficientemente estrechos para detectar sólo la emisión del aceptor excitado, se corre el riesgo de tener algo de señal proveniente de la cola del espectro de emisión del dador), fotoblanqueo (destrucción fotoquímica del fluoróforo, puede suceder por la exposición prolongada a longitudes de onda de excitación), y diferente sensibilidad de los fluoróforos al ambiente.

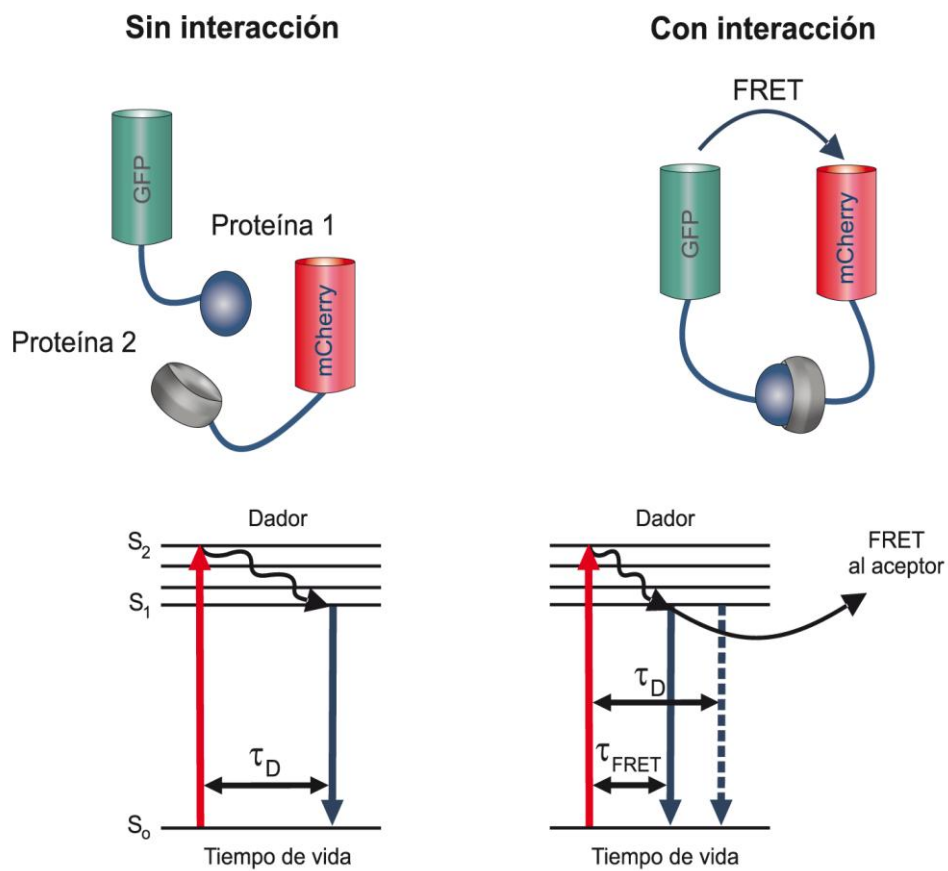


Figura 11.6. La proteína 1 se encuentra marcada con la proteína fluorescente verde GFP mientras que la proteína 2 se encuentra marcada con la proteína fluorescente roja mCherry. GFP-mCherry son capaces de formar un par FRET. El diagrama de Jablonski en la situación sin interacción (ver cap. Fluorescencia) muestra el tiempo de vida τ_D del dador GFP. Cuando las proteínas 1 y 2 interaccionan, GFP y mCherry están lo suficientemente cerca y correctamente orientadas para que la transferencia de energía entre el dador GFP y el aceptor mCherry se produzca. El diagrama de Jablonski en la situación con interacción muestra la disminución del τ_D del donador cuando se produce la transferencia de energía, ahora llamado τ_{FRET} .

Para llevar a cabo una medida FLIM-FRET, las dos proteínas de las que se pretende estudiar su interacción deben estar marcadas cada una con un fluoróforo adecuado. El modo más conveniente de obtener este par de proteínas marcadas in vivo es expresar cada una de ellas fusionada a su respectiva proteínas fluorescentes (ver Capítulo 2 Fluorescencia). Las dos proteínas fluorescentes usadas como marca deben poseer las propiedades espectrales apropiadas que les permitan funcionar como par FRET. Un par FRET muy comúnmente utilizado consiste en usar EGFP (*Enhanced-Green Fluorescent Protein*) como dador y mCherry (*monomeric Cherry red fluorescent variant*) como aceptor (Figura 11.7). Otra combinación que puede ser usada como par FRET es CFP (*Cyan Fluorescent Protein*, como dador) y YFP (*Yellow Fluorescent Protein*, como aceptor). Idealmente, el experimento de FLIM-FRET debería ser llevado a cabo en líneas celulares que expresen de modo estable ambas proteínas de fusión. Sin embargo, dado el hecho que FLIM-FRET es generalmente insensible a la concentración relativa de los fluoróforos, es posible llevar a cabo el protocolo en células que expresen de modo transitorio una o ambas proteínas de fusión (Lleres, 2007).

En la Figura 11.7 se muestra un ejemplo de una medida de FLIM-FRET que permite evidenciar la asociación del par FRET GFP-mCherry. GFP es el fluoróforo dador y mCherry el fluoróforo aceptor. En la Figura 11.7A, células HeLa fueron cotransfectadas con las proteínas GFP y mCherry, ambas expresadas de modo independiente. En el primer recuadro, se observa la medida del τ_D (tiempo de vida media del dador, GFP). El siguiente recuadro muestra el τ_{FRET} . En el caso que el fenómeno FRET se estuviera dando, τ_{FRET} debería ser menor que el τ_D . Dado que τ_{FRET} es igual al τ_D , en esta situación, GFP y mCherry coexpresadas de modo independiente en células HeLa no interactúan, se encuentran entre sí a distancias mayores que las necesarias para que ocurra el fenómeno FRET. El tercer recuadro muestra el porcentaje de eficiencia FRET. En el caso A, la eficiencia FRET es nula.

En la Figura 11.7B, se repite el ensayo, esta vez transfectando las células HeLa con una construcción de GFP unida a mCherry por 17 aminoácidos. En el primer recuadro se muestra la medida del τ_D , 2.11ns en este caso (como

previamente mencionamos, el τ de un fluoróforo se ve influenciado por su entorno, la presencia de mCherry modifica el τ_D de GFP). En el segundo recuadro, se muestra la medida del τ_{FRET} . En este caso, el τ_{FRET} es de 1.95ns, menor que el τ_D , indicando que el fenómeno FRET está ocurriendo.

En la Figura 11.7C, las células se transfectaron con una construcción de GFP unida a mCherry por sólo 7 aminoácidos. En el primer recuadro el τ_D es 2.05ns mientras que el τ_{FRET} que muestra el segundo recuadro es 1.58ns. La mayor cercanía entre GFP y mCherry aumenta la eficiencia FRET (tercer recuadro) así como disminuye el τ_{FRET} .

El rango de interacciones moleculares comprobados mediante FRET está creciendo e incluye la dimerización de factores de transcripción o receptores, formación de dominios lipídicos, interacciones entre subunidades en un único complejo proteico funcional o “metaboloma”, asociación de proteínas regulatorias o de señalización.

Correlación de espectroscopía de fluorescencia (FCS).

La caracterización cuantitativa de interacciones biomoleculares es de fundamental importancia para comprender los mecanismos celulares. En los últimos años, dos desarrollos han revolucionado el modo según el cual estas interacciones pueden ser medidas: el descubrimiento de proteínas autofluorescentes y la posibilidad de detectar estas moléculas con una sensibilidad de una única molécula. La óptica confocal ha sido la mejora en microscopía que nos permite, hoy en día, detectar moléculas con una sensibilidad de una única molécula por campo (Langowski, 2008).

La difusión de partículas fluorescentes dentro y fuera del volumen de detección (aprox. 1fl) hace que la intensidad de fluorescencia en los detectores fluctúe aleatoriamente. La correlación de espectroscopía de fluorescencia (FCS) mide y analiza estas fluctuaciones, a través de la información colectada sobre la movilidad de las moléculas que cruzan el haz de láser (Figura 11.8). FCS permite la medida de interacciones biomoleculares. Cuando un ligando

fluorescente se une a una macromolécula, su movilidad se verá restringida por la presencia de su compañero de interacción de mayor tamaño (Fricker, 2006). En biología celular, la movilidad que hemos mencionado está caracterizada por el coeficiente de difusión D expresado normalmente en unidades de $\mu\text{m}^2\text{s}^{-1}$. La medida de D revelará eventualmente interacciones entre el ligando fluorescente y macromoléculas de mayor tamaño, siempre y cuando D cambie significativamente.

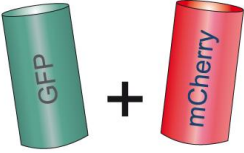
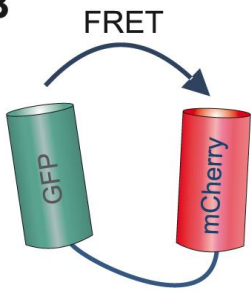
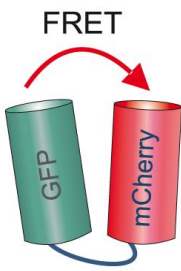
	Tiempo de vida medio del dador τ_D	τ_{FRET}	Eficiencia FRET (%)
A  GFP + mCherry	2.34 ns	2.34 ns	0 %
B  FRET 17 Aminoácidos	2.11 ns	1.95 ns	20 %
C  FRET 7 Aminoácidos	2.05 ns	1.58 ns	60 %

Figura 11.7. Medida *in vivo* de FLIM-FRET. Células HeLa vivas que coexpresan EGFP libre y mCherry libre (A). (B) Células expresando EGFP unida directamente a mCherry por un linker de 17 aminoácidos. (C) Expresión de EGFP unida a mCherry por 7 aminoácidos en células HeLa. Se calcularon los tiempos de vida y eficiencia FRET del dador (EGFP) en cada una de las 3 situaciones (Lleres, 2007).

En una primera aproximación, D es proporcional a la mayor dimensión lineal de la macromolécula, su dependencia de la masa molecular M de una proteína

globular no es muy fuerte; para que D se duplique, M debería aumentar $2^3=8$ (Ley de Fick $M=-D_x A_x t(dc/dn)$, A es la superficie normal a la dirección de la difusión en m^2 , t es el tiempo en segundos y dc/dn el gradiente de concentración). Una asociación entre dos ligandos de tamaño similar disminuiría D aproximadamente un 26%, cambio poco detectable en presencia de ruido en la señal (Langowski, 2008).

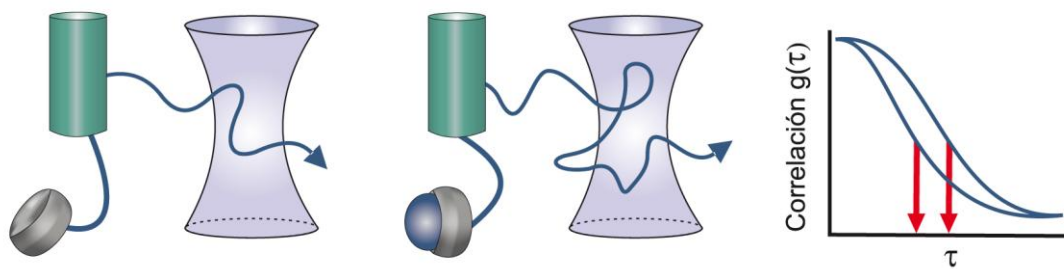


Figura 11.8. El movimiento de una proteína marcada fluorescentemente dentro y fuera del volumen de iluminación provoca fluctuaciones en la intensidad de fluorescencia detectada. Cuando una proteína se asocia a otra, las fluctuaciones debidas al pasaje dentro y fuera del volumen de detección cambian. La función de autocorrelación permite estudiar estas fluctuaciones de intensidad de fluorescencia. El decaimiento de la función de autocorrelación se relaciona con la disminución de la movilidad difusional de las especies que atraviesan el volumen de detección.

Un modo más sensible para utilizar el movimiento aleatorio de macromoléculas para detectar su interacción consiste en utilizar dos colores en la correlación cruzada de espectroscopía de fluorescencia (FCCS). En este caso, ambos componentes de la interacción son marcados con fluoróforos que son espectroscópicamente distinguibles. Si forman un complejo, entrarán y saldrán del foco del láser al mismo tiempo; si no interaccionan, sus respectivos movimientos (y fluctuaciones de fluorescencia) no estarán correlacionados. FCCS es un modo conveniente de demostrar unión entre dos macromoléculas marcadas con diferentes fluoróforos porque el complejo mostrará fluorescencia correlacionada a ambas longitudes de onda.

Debido al pequeño foco del haz láser, las medidas dentro de células vivas se han hecho posibles. Las lentes de un típico microscopio de alta resolución poseen un punto focal de 300nm de diámetro y 1,5 μ m de largo, de este modo, los procesos que implican difusión dentro de una célula u organela pueden ser comprobados de un modo posición-dependiente.

El primer dato que se obtiene en un experimento de FCS es la intensidad de fluorescencia dependiente del tiempo $F(t)$, que es proporcional al número de partículas en el volumen de observación al tiempo t (Langowski, 2008).

Función de autocorrelación.

Las fluctuaciones en la intensidad de fluorescencia se caracterizan por su función de autocorrelación $G(\tau)$, que describe el movimiento aleatorio del fluoróforo.

$$G(\tau) = \frac{\langle F(t)F(t + \tau) \rangle}{\langle F(t) \rangle^2} \quad \text{Ec. 11.2}$$

Donde $F(t)$ es la intensidad de fluorescencia dependiente del tiempo. Esta función de autocorrelación se grafica usando los datos experimentales obtenidos.

Para obtener cantidades tales como coeficientes de difusión, concentraciones, o evaluar constantes de reacción, se ajusta una función de correlación teórica a la medida $G(\tau)$, la cual está basada en un modelo que contiene estas cantidades como parámetros libres (Langowski, 2008).

$$G(\tau) = \frac{1}{cV_{\text{eff}}} \left(1 + \frac{4D\tau}{w_0^2} \right)^{-1} \left(1 + \frac{4D\tau}{z_0^2} \right)^{-1/2} + 1 \quad \text{Ec. 11.3}$$

Correlación cruzada de dos colores (FCCS).

Cuando el coeficiente de difusión D cambia suficientemente al producirse la unión/interacción de dos macromoléculas, el complejo puede distinguirse en $G(\tau)$ como una segunda especie y su concentración ser determinada. Sin

embargo, en casos donde D cambia sólo ligeramente o no cambia, como en el caso en el que un ligando no fluorescente se une a una partícula fluorescente de tamaño mucho mayor, esta estrategia no puede ser utilizada.

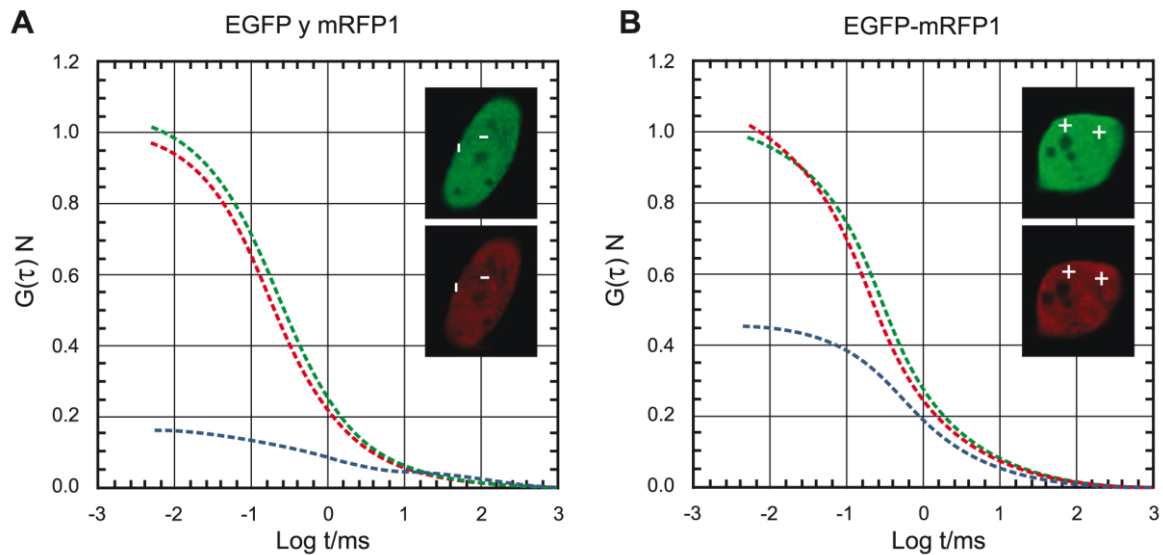


Figura 11.9. A, B- FCCS controles *in vivo* para fluoróforos independientes (A) y covalentemente unidos (B). Se muestran las amplitudes de autocorrelación normalizadas $G(\tau)$ en los canales rojo (línea roja) y en verde (línea verde) medidos en células HeLa expresando (A) EGFP y mRFP1 separadamente; (B) una construcción de ambas proteínas EGFP y mRFP1 unidas covalentemente. La función de correlación cruzada se muestra en azul. Los rombos representan la información medida, las líneas las curvas ajustadas. Imágenes insertas: imágenes confocales de las células en los canales para EGFP (verde) y mRFP1 (rojo). Las cruces blancas indican el punto de medida (Langowski, 2008).

En FCCS, la fluorescencia es detectada simultáneamente a dos longitudes de onda distintas en el mismo volumen de detección. Es simple de observar que en una mezcla de dos moléculas fluorescentes emitiendo a dos longitudes de onda diferentes sin interactuar entre ellas, las mismas difundirán independientemente y la amplitud de la función de correlación cruzada será cero (Figura 11.9A). Por el contrario, cuando la partícula está marcada con dos fluoróforos que emiten simultáneamente a las dos longitudes de onda respectivas, la función de correlación cruzada es igual a la función de autocorrelación para un único color en FCS (asumiendo eficiencias de detección iguales y superposición exacta de los volúmenes de detección para ambos canales). Este último caso se da, por ejemplo, cuando las especies fluorescentes forman un complejo (Figura 11.9B) (Langowski, 2008).

En FCCS, la cantidad de complejo formada entre dos biomoléculas marcadas fluorescentemente puede obtenerse midiendo la amplitud de la correlación cruzada.

Bibliografía

Bader GD, Betel D, Hogue CWV. BIND: the biomolecular interaction network database. *Nucleic Acids Res*; 31:248-50; 2003.

Bateman A, Birney E, Cerruti L, Durbin R, Eddy SR, *et al.* The Pfam protein families database. *Nucleic Acids Res*; 30:276-280; 2002.

Bock JR, Gough DA. Predicting protein-protein interactions from primary structure. *Bioinformatics*; 17:455-60; 2001.

Chothia C, Janin J. Principles of protein-protein recognition. *Nature*; 256:705-8; 1975.

Corpet F, Servan F, Gouzy J, Kahn D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res*; 28:267-9; 2000.

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.

Enright A, Iliopoulos I, Kyripides N, Ouzounis C. Protein interaction maps for complete genomes based on gene fusion events. *Nature*; 402:86-90; 1999.

Fricker M, Runions J, Moore I. Quantitative Fluorescence Microscopy; from art to science. *Annual Review Plant Biology*. 57:79-107. 2006.

Gallet X, Charlotiaux B, Thomas A, Brasseur R. A fast method to predict protein interaction sites from sequences. *J Mol Biol*; 302:917-26; 2000.

Gietz R.D. *et al.* Identification of proteins that interact with a protein of interest: Applications of the yeast two hybrid system. *Molecular and Cellular Biochemistry*, 172: 67–79; 1997.

Golemis E. Protein-protein interactions: A molecular cloning manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 2002.

- Hink MA, Bisseling T, Visser AJ.. Imaging protein-protein interactions in living cells. *Plant Mol. Biol.* 50:871–83; 2002.
- Howard CV, Reed MG. *Unbiased Stereology: Three-Dimensional Measurement in Microscopy*. Oxford: BIOS Sci. Publ. 1998.
- Huynen M, Snel B, Lathe W, Bork P. Predicting protein function by genomic context: quantitative evaluation and qualitative. *Genome Res*;10:1204-10; 2000.
- 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*;31:255-7. 2003.
- Kato N, Pontier D, Lam E. Spectral profiling for the simultaneous observation of four distinct fluorescent proteins and detection of protein-protein interaction via fluorescence resonance energy transfer in tobacco leaf nuclei. *Plant Physiol.* 129:931–42. 2002
- Lakowicz, J.R. Principles of Fluorescence Spectroscopy 2nd Edition. Springer, New York. 1999.
- Langowski J. Chapter 20: Protein-protein interactions determined by Fluorescence Correlation Spectroscopy. *Methods in cell biology*, vol. 85. Elsevier Inc. 2008.
- Lleres D, Swift S, Lamond AI. Detecting Protein-Protein Interactions *in vivo* with FRET using Multiphoton Fluorescence Lifetime Imaging Microscopy (FLIM). *Current Protocols in Cytometry*. 2007.
- Lu L, Lu H, Skolnick J. MULTIPROSPECTOR: an algorithm for the prediction of proteinprotein interaction by multimeric threading. *Proteins*;49:350-64. 2002.
- Marcotte EM, Pellegrini M, Ho-Leung N, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. *Science*; 285:751-3. 1999.
- Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. A combined algorithm for genome-wide prediction of protein function. *Nature*; 402:83-6. 1999.
- Matchmaker™ Gold Yeast Two-Hybrid System User Manual Clontech. PT4084-1 (PR033493) Cat. No. 630489. 2010.
- 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*;30:31-4. 2002.
- Ng SK, Zhang Z, Tan SH. Integrative approach for computationally inferring protein domains interactions. *Bioinformatics*;19:923-9; 2003.
- Pazos F, Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng*;14:609-14; 2001.

- Pazos F, Valencia A. In silico two-hybrid system for the selection of physically interacting proteins pairs. *Proteins*;47:219-27; 2002.
- 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*; 96:4285-8; 1999.
- Phizicky E. M. and Fields S. Protein-protein interactions: Methods for detection and analysis. *Microbiol Rev.* 59:94-123. 1995.
- 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.
- 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*; 27:89-94. 1999.
- Sato, Y., *et al.*, Detecting weak protein-protein interactions by modified far-western blotting, *J. Biosci. Bioeng.* 2011.
- Stein L, Thierry-Mieg J. Scriptable access to the *Caenorhabditis elegans* genome sequence and other Acedb databases. *Genome Res*;8:1308-15. 1999.
- Thermo Scientific Pierce Protein Interaction Technical Handbook Version 2. 2010.
- 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; p.265-78. 1999.
- Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*; 17:284-5. 2001.
- Torre Russis V, Valles A, Gómez R, Chinae g, Pons T. Interacciones proteína-proteína: bases de datos y métodos teóricos de predicción. *Biotecnología Aplicada*; 20:201-208; 2003.
- Westermarck J., *et al.* Identification of Protein Interactions Involved in Cellular Signaling. *Mol Cell Proteomics*. 2013.
- Wojcik J, Schächter V. Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics*;17 Suppl 1:296-305; 2001.
- 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*;29:239-41; 2001.

Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G.
MINT: a molecular interaction database. FEBS Lett;513:135-40. 2002.