

Proteomica: concetti e prospettive

P. Pucci

*Dipartimento di Chimica Organica e Biochimica,
Università di Napoli Federico II e CEINGE Biotecnologie Avanzate*

Il termine proteoma è utilizzato oggi per indicare la procedura di identificazione di un elevato numero di proteine in miscele molto complesse provenienti da organelli cellulari, intere cellule o addirittura organismi. I moderni studi sul proteoma possono essere divisi essenzialmente in due settori principali, il proteoma di espressione che tende alla definizione qualitativa e quantitativa dell'aumento e/o diminuzione dei livelli di proteine, e il proteoma funzionale che tenta di identificare componenti di compartimenti cellulari, complessi multiproteici e vie di trasduzione del segnale. Entrambi questi approcci si basano sul frazionamento delle miscele di proteine mediante elettroforesi bidimensionale su gel di poliacrilamide (2D-gel) e sulla successiva identificazione delle bande proteiche mediante tecniche di spettrometria di massa (2D-MS). Questa procedura, infatti, è ottimale per l'analisi globale di proteine in quanto si è dimostrata in grado di separare, quantificare ed identificare migliaia di proteine presenti in un singolo gel (1-3).

L'identificazione di bande proteiche da 2D-gel si ottiene essenzialmente mediante spettrometria di massa MALDI-TOF (Matrix-Assisted Laser Desorption Ionization-Time Of Flight). La banda proteica è scissa dal gel, digerita *in situ* con opportuni enzimi proteolitici e la miscela peptidica risultante è direttamente analizzata mediante MALDI-MS. L'identificazione delle proteine viene effettuata utilizzando i valori di massa accurati dei peptidi determinati mediante MALDI-MS. Questi valori, infatti, insieme ad altri parametri quali il tipo di proteasi usata per l'idrolisi o il peso molecolare della proteina presunto dal gel di SDS, vengono introdotti in alcuni programmi di ricerca disponibili in rete (ProFound, Mascot, MS-Fit, etc.). I valori di massa registrati sugli spettri sono paragonati con quelli provenienti dalla digestione teorica di tutte le proteine della banca dati, consentendo l'identificazione delle proteine. Quando questa procedura non forni-

sce dati definitivi, vengono utilizzate metodologie di spettrometria di massa tandem basate sulla ionizzazione ad electrospray (ES-MS/MS) per ottenere informazioni anche parziali di sequenza di uno o più frammenti proteolitici. E' stato dimostrato che le informazioni ottenute dalla sequenza parziale di un solo peptide sono spesso sufficienti ad identificare la proteina in banca dati.

In genere gli studi di proteomica di espressione sono indirizzati all'analisi delle differenze del pattern di espressione in cellule anormali (i.e. tumorali, stimolate da trattamenti farmacologici, etc.) in paragone con quelle normali, attraverso l'identificazione di bande proteiche che compaiono (o scompaiono) o che sono presenti a livelli quantitativi diversi. Nelle applicazioni biomediche, questo approccio comparativo viene di solito utilizzato per identificare proteine il cui livello di espressione aumenta o diminuisce in seguito all'insorgere di una patologia e che possono quindi essere utilizzate come marker specifici per la diagnosi o a scopo terapeutico.

Gli studi di proteomica funzionale hanno come obiettivo finale la comprensione a livello molecolare delle funzioni di una cellula vivente. Poiché ormai è nota l'intera sequenza del genoma di vari organismi, il numero di proteine la cui funzione è ancora sconosciuta è cresciuto enormemente negli ultimi tempi. Nel tentativo di ottenere informazioni sull'attività biologica di singole proteine, è particolarmente importante l'analisi del proteoma funzionale definito come lo studio delle fluttuazioni spaziali e temporali dei componenti e dei processi molecolari che avvengono in una cellula vivente. Infatti, in una cellula molti processi non vengono controllati soltanto dall'abbondanza relativa delle varie proteine, ma anche dalla regolazione transiente della loro attività, localizzazione cellulare e associazione con altri componenti. In particolare, è oggi chiaro che un gran numero di proteine è presente nella cellula sotto forma di complessi multiproteici;

ne consegue che la comprensione delle funzioni biologiche di queste proteine è legata all'identificazione dei loro partners molecolari (4).

Procedure basate sull'analisi proteomica possono fornire un contributo basilare all'identificazione dei componenti dei complessi multiproteici. Se è noto che una particolare proteina fa parte di un complesso, essa può essere espressa in forma ricombinante modificata con una specifica marcatura (tag) e l'intero complesso può quindi essere purificato dall'intero estratto cellulare mediante tecniche basate sulla cromatografia di affinità utilizzando l'opportuno ligando (anti-tag) immobilizzato su un supporto insolubile. Alternativamente, la proteina è marcata con un epitopo per il quale è disponibile un buon anticorpo (FLAG, HA, Myc, etc.), è sovra espressa nella cellula ed il complesso viene immunoprecipitato con l'anticorpo specifico per l'epitopo. In entrambi i casi, le proteine costituenti il complesso isolato sono frazionate mediante SDS-PAGE ed identificate utilizzando la spettrometria di massa secondo la procedura descritta.

Recentemente il nostro gruppo di ricerca ha sviluppato una nuova strategia specifica per l'identificazione dei componenti proteici che *in vivo* interagiscono, anche in modo transiente, con una particolare proteina. Tale strategia rappresenta un efficace metodo alternativo alle procedure di immunoprecipitazione o sull'approccio di biologia molecolare basato sulla tecnica del doppio ibrido. La filosofia del metodo consiste nell'utilizzo di sistemi di espressione di proteine commercialmente disponibili per produrre una proteina ricombinante (od un dominio proteico) recante una specifica modificazione (tag) che possa fungere da esca per pescare i suoi specifici partners molecolari all'interno dell'intero estratto cellulare. La proteina esca può essere espressa come ibrido di fusione con la Glutathione-S-transferasi (GST-) o con

l'epitopo FLAG, o con una coda di poli-His o modificata covalentemente con la biotina. In tutti i casi, l'esca può essere immobilizzata su un supporto solido come ad esempio particelle di agarosio (cromatografia di affinità) derivatizzato con un appropriato ligando anti-tag (glutathione, anticorpo anti-FLAG, ioni Nichel, Streptavidina, etc.). La stessa strategia può anche essere applicata all'identificazione di proteine che interagiscono con DNA/RNA utilizzando uno specifico oligonucleotide come esca.

Una volta immobilizzata, l'esca viene incubata con l'intero estratto cellulare o, quando è il caso, con estratti da organelli specifici. La proteina (o l'oligonucleotide) stabilisce interazioni non covalenti con partners specifici presenti nell'estratto cellulare, mentre le proteine non tratteneute saranno eluite durante il lavaggio. I componenti proteici specificamente riconosciuti dall'esca sono quindi eluiti e frazionati mediante SDS-PAGE. Le bande proteiche visualizzate sul gel sono digerite proteoliticamente o con metodi chimici *in situ* e le miscele di peptidi risultanti sono analizzate mediante MALDI-MS, consentendo l'identificazione delle proteine secondo la procedura descritta in precedenza.

Bibliografia

1. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Ann Rev Biochem* 2001; 70: 437-73.
2. Patterson SD, Aebersold R. Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* 1995; 16: 1791-814.
3. Yates JR 3rd. Mass spectrometry and the age of the proteome. *J Mass Spectrom* 1998; 33: 1-19.
4. Pawson T, Scott JD. Signaling through scaffold, anchoring, and adaptor proteins. *Science* 1997; 278: 2075-80.