



Proteomica Funzionale

MARIA MONTI, STEFANIA ORRÙ,
DANIELA PAGNOZZI, PIERO PUCCI

Introduzione

Gli ambiziosi progetti che alla fine del secolo XX hanno avuto come obiettivo la determinazione della sequenza di interi genomi, incluso quello dell'uomo, hanno paradossalmente contribuito ad un rinnovato interesse per lo studio delle proteine. Infatti, ben prima del completamento del progetto Genoma Umano è apparso chiaro che, sebbene la conoscenza della sequenza dell'intero DNA umano avrebbe costituito un enorme patrimonio di informazioni, questo risultato avrebbe dovuto essere considerato un punto di partenza e non un punto di arrivo nella comprensione del funzionamento della macchina cellulare a livello molecolare (1, 2). D'altra parte i successi dei progetti di sequenziamento dei genomi hanno posto le basi per lo studio dei fenomeni biologici con metodologie globali, in generale descritte dal suffisso "omica". In questi ultimi anni, quindi, la sfida della ricerca biologica si è di nuovo rivolta alle proteine, dando luogo alla cosiddetta "era proteomica" con l'obiettivo di identificare e localizzare le proteine all'interno di un determinato compartimento cellulare, di una cellula o di un organismo così come alla definizione dei percorsi proteici in un sistema cellulare. Questi nuovi obiettivi comunque, non possono essere raggiunti facilmente poiché il livello di difficoltà intrinseco aumenta di vari ordini di grandezza nel passare dalle ricerche di tipo genomico a quelle di tipo proteomico. La natura statica del genoma infatti non può essere paragonata alle proprietà dinamiche del proteoma; i profili dell'espressione proteica cambiano diverse volte durante il ciclo cellulare e sono pesantemente influenzati da una serie di stimoli intra- ed extra-cellulari (temperatura, stress, segnali apoptotici, etc.) (3). Esistono inoltre numerosi aspetti, difficilmente decifrabili dal messaggio genico, che influenzano grandemente i meccanismi funzionali delle cellule viventi (4). Questi aspetti riguardano essenzialmente la qualità e la dinamica dei processi successivi alla trascrizione, come la maturazione dell'RNA (processo di splicing alternativo), e/o alla traduzione, come le modifiche post-traduzionali delle proteine. Queste considerazioni portano

a riformulare completamente il vecchio paradigma "un gene, una proteina" che non riflette più la reale natura di un proteoma cellulare; si può affermare invece che ad un genoma corrisponde una molteplicità di proteomi il cui limite superiore, almeno per ora, non è definibile.

La **proteomica** può essere pertanto definita come lo studio dei proteomi nella loro complessità. Il termine **proteoma**, un acronimo di **proteine** e **genoma**, è nato nel 1994 per merito del ricercatore australiano Marc Wilkins che, con tale vocabolo, voleva intendere il complemento proteico espresso da un genoma, cioè l'intero complesso dei prodotti dell'espressione di un genoma. Questo termine 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. Da queste definizioni di carattere del tutto generale discendono varie altre definizioni di proteomica che, opportunamente specificate o aggettivate, si riferiscono agli obiettivi specifici che l'approccio proteomico intende affrontare.

Gli attuali studi di proteomica sono prevalentemente focalizzati su due aree principali, da una parte la proteomica di espressione che tende alla definizione qualitativa e quantitativa dell'aumento e/o diminuzione dei livelli di proteine, e dall'altra gli studi di proteomica funzionale rivolti alla caratterizzazione dell'attività proteica, dei complessi multiproteici e dei percorsi di signalling (ovvero di trasduzione del segnale) (5). Generalmente, gli studi di proteomica di espressione sono rivolti alla definizione dei pattern di espressione proteica nelle cellule anormali (cioè cellule tumorali, cellule sottoposte allo stimolo dovuto al trattamento con un farmaco, etc.) in paragone con le cellule normali. Nelle applicazioni biomediche, questo approccio comparativo è di solito utilizzato per identificare le proteine che sono sovra- o sotto-espresse in un modo che è caratteristico di una specifica malattia per poterli successivamente utilizzare come marcatori diagnostici o come bersagli terapeutici (6). In questi studi, l'analisi affidabile delle variazioni quantitative nell'espressione proteica è cruciale. Queste variazioni sono di solito ottenute dalle intensità della colorazione delle bande proteiche sui gel, metodo questo che richiede una grossa mole di lavoro ed è particolarmente soggetto ad errore. Recentemente, sono stati ottenuti risultati migliori e più affidabili mediante l'uso di tecniche di marcatura con isotopi stabili o con coloranti fluorescenti (7, 8).

Gli approcci di proteomica funzionale hanno due obiettivi principali, la definizione della funzione biologica di proteine sconosciute e la definizione dei meccanismi cellulari a livello molecolare. Nelle cellule la maggior parte delle proteine esplica la propria funzione biologica attraverso la rapida e transiente associazione con grandi complessi

CEINGE Biotecnologie Avanzate s.c.a r.l. via Comunale Margherita 482, 80145 Napoli, Italy. Tel. 0039-0813722896, Fax 0039-0813722808, e-mail: pucci@unina.it.

Dipartimento di Chimica Organica e Biochimica, Università di Napoli Federico II, Napoli, Italy.

multi-proteici (9). La comprensione delle funzioni di una proteina così come la definizione dei meccanismi molecolari all'interno di una cellula dipende quindi dall'identificazione dei partners con cui interagisce. L'associazione di una proteina sconosciuta con partner che appartengono ad uno specifico complesso multiproteico coinvolto in un determinato meccanismo cellulare potrebbe infatti essere fortemente indicativa della sua funzione biologica (10). Inoltre, la definizione delle interazioni fra proteine all'interno della cellula potrebbe essere di grande aiuto anche nella descrizione dettagliata dei percorsi di signalling cellulare (11).

Identificazione delle proteine mediante spettrometria di massa.

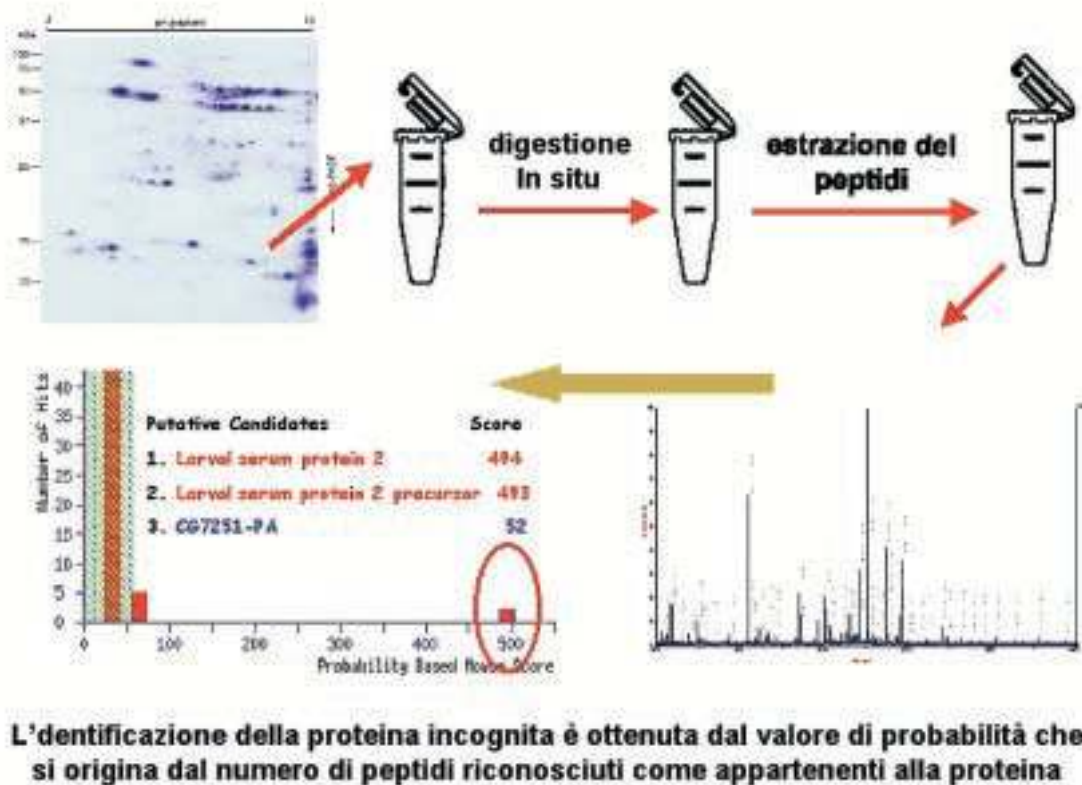
Il passaggio cruciale in tutti gli studi di proteomica consiste nell'identificazione delle proteine preventivamente separate su gel mono- o bi-dimensionali. Questa identificazione è ottenuta mediante tecniche di spettrometria di massa, grazie all'elevatissima sensibilità e all'ampio intervallo dinamico di analisi proprio di questa metodologia. Sono state sviluppate varie procedure sperimentali per l'identificazione delle proteine, che possono essenzialmente essere riassunte in due approcci alternativi e/o complementari. In entrambi i casi, le bande proteiche, la cui colorazione è condotta in condizioni da non interferire con la successiva analisi, sono escisse dal gel, trattate *in situ* con un appropriato enzima proteolitico e la miscela peptidica risultante è estratta dal gel ed è disponibile per le successive analisi di spettrometria di massa (12).

In un primo approccio, l'identificazione delle varie proteine è effettuata attraverso la procedura del *peptide mass fingerprinting* il cui principio si basa sull'osservazione

che proteine con una diversa sequenza amminoacidica, in seguito all'azione di una proteasi specifica, generano un insieme discreto di peptidi, definiti dalla loro massa, che è unico per ciascuna proteina. La miscela di peptidi è direttamente analizzata mediante spettrometria di massa MALDI-TOF (Matrix-Assisted Laser Desorption Ionization-Time Of Flight) (Fig. 1).

I valori dei pesi molecolari dei peptidi così ottenuti, insieme ad altre informazioni come la specificità dell'enzima utilizzato per l'idrolisi ed il peso molecolare della proteina approssimativamente determinato dall'analisi su gel, sono immessi in diversi programmi per la ricerca dei valori di massa (ProFound, Mascot, MS-Fit) disponibili in rete. Questi valori di massa sono paragonati a quelli originati dall'idrolisi virtuale di tutte le proteine presenti in banca dati consentendo così l'identificazione della proteina.

Alternativamente, quando la procedura del *peptide mass fingerprinting* non è sufficiente per l'identificazione della proteina, possono essere utilizzate tecniche di spettrometria di massa tandem a ionizzazione electrospray accoppiate con la cromatografia liquida (LC-ESI-MS/MS) (Fig. 2). Le miscele peptidiche prodotte mediante idrolisi *in situ* sono separate mediante analisi per HPLC capillare e le frazioni eluite dalla colonna sono direttamente introdotte nella sorgente ESI dello spettrometro di massa per la determinazione dei loro pesi molecolari. Gli ioni peptidici sono quindi simultaneamente isolati e frammentati all'interno dello spettrometro di massa e producono una serie di spettri di frammentazione da cui è possibile ottenere le necessarie informazioni sulla sequenza amminoacidica dei singoli peptidi. Queste informazioni sono quindi utilizzate per la ricerca in banca dati consentendo l'identificazione della proteina. È stato dimostrato, infatti, che le infor-



L'identificazione della proteina incognita è ottenuta dal valore di probabilità che si origina dal numero di peptidi riconosciuti come appartenenti alla proteina

Fig. 1: Rappresentazione schematica del metodo di identificazione delle proteine mediante *peptide mass fingerprinting*. L'elaborazione dei dati avviene attraverso programmi disponibili in rete i cui indirizzi sono qui riportati:

Mascot: <http://www.matrixscience.com>; **ProSpector:** <http://prospector.ucsf.edu>; **ProFound:** <http://prowl.rockefeller.edu>

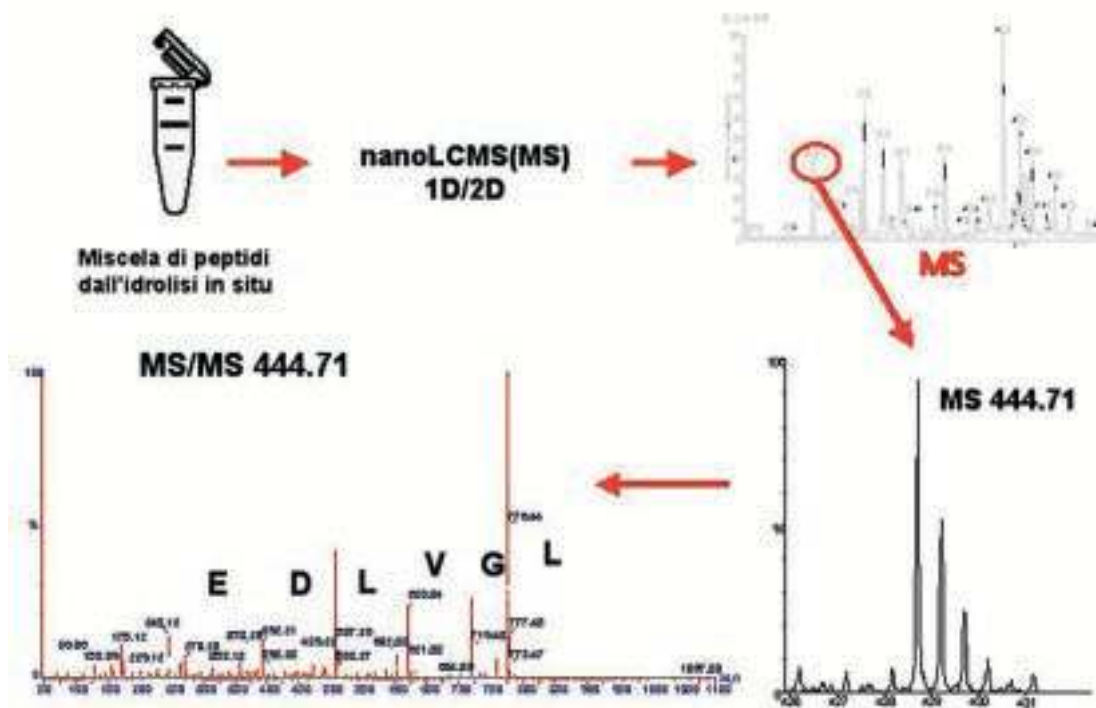


Fig. 2: Rappresentazione schematica del metodo di identificazione delle proteine mediante LC-MS/MS. L'elaborazione dei dati avviene attraverso gli stessi programmi sopra indicati.

mazioni di sequenza ottenute da due soli peptidi sono sufficienti per identificare una proteina in modo non ambiguo.

Identificazione dei partners proteici mediante approcci di proteomica funzionale.

Con l'aumento dei numerosi progetti di sequenziamento dei genomi, si è verificata una concomitante crescita esponenziale del numero di sequenze proteiche la cui funzione biologica è ancora sconosciuta. Le scienze biologiche si trovano quindi ad affrontare una sorta di situazione paradossale in cui la sequenza proteica, il corrispondente gene che codifica per esse, la sua localizzazione cromosomica ed anche il meccanismo di regolazione possono essere definiti ma rimane completamente sconosciuto il ruolo biologico che svolge la proteina all'interno della cellula.

È ormai chiaro che un grande numero di proteine partecipa alla formazione di complessi multiproteici e che la comprensione della funzione di una proteina all'interno della cellula necessita dell'identificazione dei partners con cui essa interagisce. Procedure basate sull'analisi proteomica possono fornire un contributo basilare all'identificazione dei componenti dei complessi multiproteici. La filosofia del metodo consiste nella possibilità di esprimere la proteina di interesse (esca) in forma ricombinante modificata con una specifica marcatura (tag); il complesso dei partners molecolari della proteina esca 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, che rivela un'alta efficienza di legame. Sono state svi-

luppate diverse strategie che si basano su questo semplice concetto e di seguito saranno descritti e ricapitolati alcuni dei principali approcci attualmente utilizzati negli studi di proteomica funzionale.

Strategia del "fishing for partners"

Utilizzando sistemi di espressione proteica commercialmente disponibili, la proteina esca può essere prodotta come proteina ibrida di fusione con la Glutathione-S-transferasi (*GST-fused protein*), modificata con una coda di istidine (*poly-His*), biotinilata, etc. In tutti i casi, l'esca può essere immobilizzata su un supporto solido sfruttando l'interazione specifica del "tag" con un opportuno ligando. Ad esempio, si possono utilizzare particelle di agarosio derivatizzate con glutathione per legare le proteine ibride con GST, o le resine ad ioni Nichel che interagiscono specificamente con la coda di istidine. L'uso di tutti questi marcatori di affinità può essere applicabile ad un grande numero di proteine poiché si osserva solo un minimo e trascurabile effetto sulla struttura terziaria (e quindi sull'attività della proteina esca) che potrebbe ostacolare la stabilità del complesso. Una volta immobilizzata, l'esca viene incubata con l'intero estratto cellulare o, quando è il caso, con estratti da organelli specifici allo scopo di pescare i suoi specifici partners molecolari, di qui il nome dato a questa procedura "*Fishing for Partners*". I componenti proteici trattenuti sono quindi eluiti e frazionati mediante elettroforesi in condizioni denaturanti (SDS-PAGE) e le bande proteiche identificate secondo la procedura descritta in precedenza. Uno schema di questo tipo di approccio è mostrato nella Fig. 3.

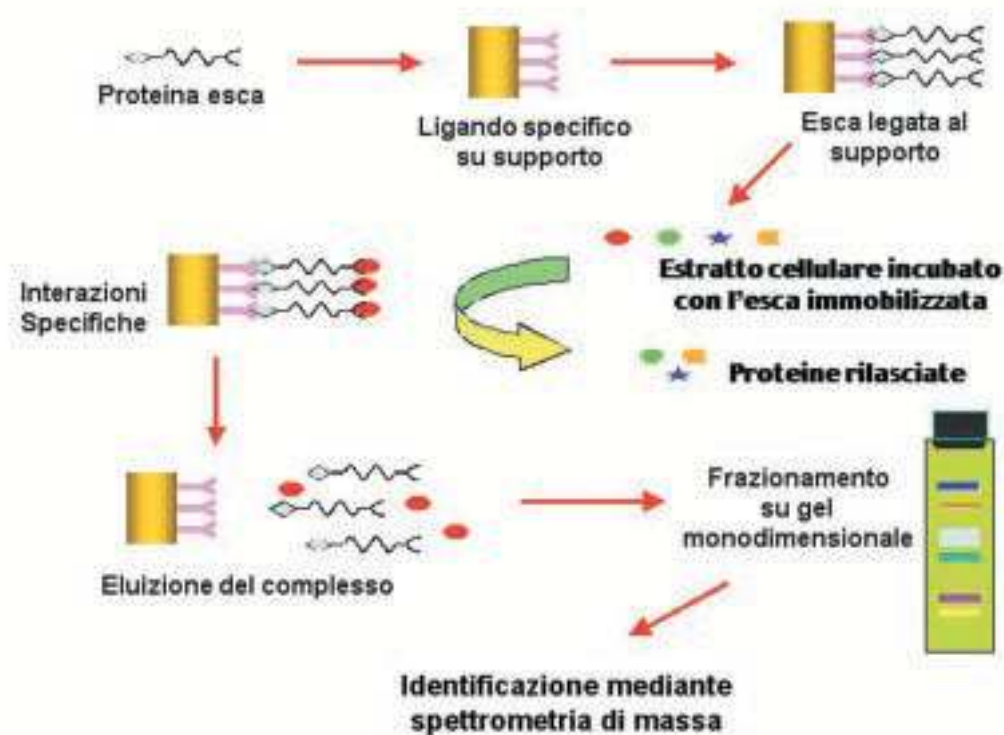


Fig. 3: Rappresentazione schematica di un esperimento di proteomica funzionale secondo la strategia definita *Fishing for Partners* essenzialmente basata sulla cromatografia di affinità, utilizzando una proteina esca con un opportuno “tag” marcatore.

Nonostante questo approccio abbia trovato una larga applicazione negli studi sulle interazioni fra proteine, esiste un elevato numero di controindicazioni. L'eccessiva contaminazione di fondo rende necessario operare dei passaggi di pre-purificazione degli estratti cellulari; tuttavia, la presenza di diverse bande identiche sia nel campione che nel controllo e le difficoltà nell'identificazione delle proteine che interagiscono in modo specifico con l'esca sottolineano immediatamente le limitazioni di questa metodologia. Inoltre, la lisi delle cellule in condizioni drastiche porta alla distruzione dell'architettura dei compartimenti cellulari, generando in principio interazioni non fisiologiche fra proteine che normalmente sono segregate in differenti compartimenti. Tuttavia, la critica maggiore verso questo approccio è che le interazioni fra l'esca ed i suoi partner si stabiliscono *in vitro* e potrebbero non essere indicative di interazioni realmente funzionali.

Il successo dell'approccio basato su tecniche di affinità dipende quindi dall'assenza di eccessive interazioni non specifiche, che a sua volta è correlata alla specificità del riconoscimento fra l'esca ed il partner. Quando la specificità di questo legame è molto alta, come nel caso delle proteine che legano il DNA, è lecito aspettarsi un basso livello di legame aspecifico. In questa particolare variante della strategia del *fishing for partners*, viene utilizzato uno specifico oligonucleotide legato ad un supporto insolubile come esca. Le proteine nucleari possono quindi essere incubate con l'esca alla ricerca di specifici partner, seguendo la strategia illustrata precedentemente.

Strategie di immunoprecipitazione

Allo scopo di superare gli inconvenienti della metodologia del *fishing for partners*, sono state introdotte strategie alternative che si basano principalmente su tecniche di immunoprecipitazione. La proteina esca è espressa marcata con un tag costituito da un epitopo peptidico per il quale è disponibile un buon anticorpo (FLAG, HA, Myc, etc.). L'esca, trasferita nella particolare linea cellulare dove si vuole condurre lo studio, viene espressa nella cellula dove formerà *in vivo* il complesso multiproteico interagendo con i suoi partners specifici. Una volta formatosi, il complesso viene immunoprecipitato utilizzando un anticorpo specifico per l'epitopo con cui è stata marcata la proteina esca. Le proteine costituenti il complesso immunoprecipitato sono quindi frazionate mediante SDS-PAGE ed identificate utilizzando metodologie di spettrometria di massa secondo la procedura descritta (Fig. 4).

Sebbene l'approccio dell'immunoprecipitazione sembri produrre dati significativi nella maggior parte dei casi, è necessario discutere alcuni aspetti negativi. Gli anticorpi utilizzati nelle metodologie di immunocolorazione non sono sempre utilizzabili nei protocolli di immunoprecipitazione che richiedono una maggiore (e quantitativa) efficacia nel riconoscimento del corretto antigene rispetto alle applicazioni di *western blot* o dei saggi ELISA. La reattività incrociata con antigeni aspecifici così come il legame non specifico agli anticorpi delle proteine, dei tag o del supporto insolubile possono portare a falsi positivi. Si raccomanda quindi

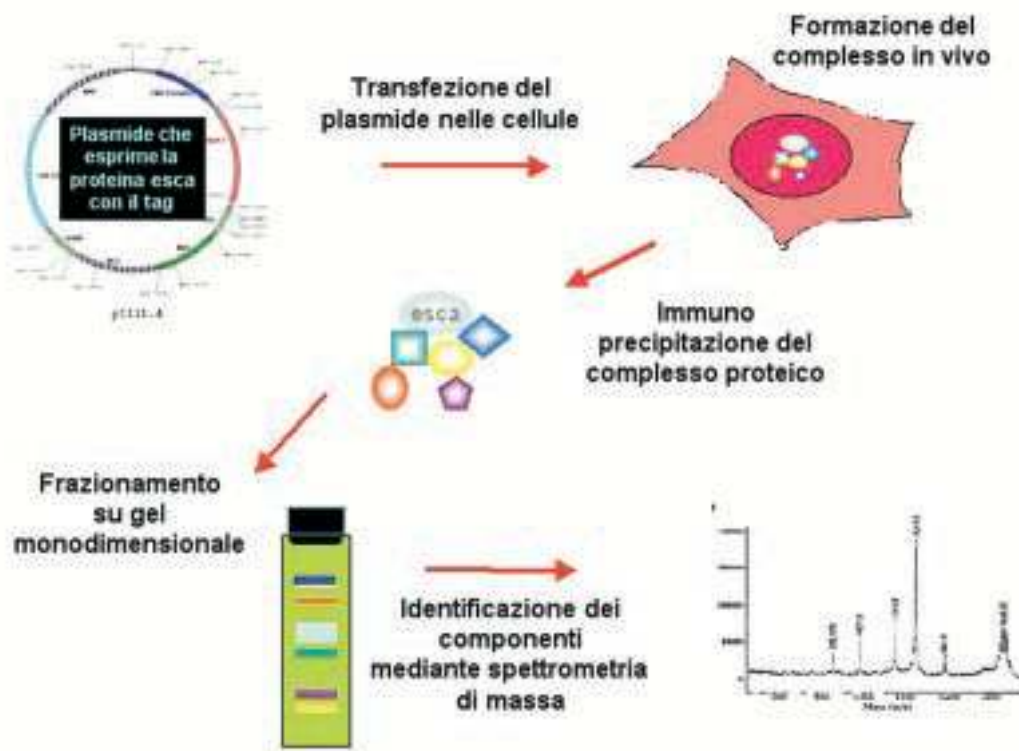


Fig. 4: Rappresentazione schematica di un esperimento di proteomica funzionale basato su tecniche di immunoprecipitazione del complesso proteico formatosi in vivo.

fortemente la pre-purificazione degli estratti cellulari proteici con anticorpi dell'animale ospite non ancora immunizzato contro lo specifico antigene.

Recentemente è nata una questione molto dibattuta circa l'uso di anticorpi specificamente diretti contro la proteina esca; questo anticorpo potrebbe competere con le proteine che interagiscono per legare gli epitopi esca portando così alla destabilizzazione delle interazioni fra le proteine ed alla conseguente dissociazione, almeno parziale, del complesso. Questi problemi sono stati risolti utilizzando le proteine marcate con tags sebbene la presenza della marcatura potrebbe influenzare la conformazione della proteina, alterando o impedendo la formazione del complesso. Un primo tentativo per risolvere questo problema è stato quello di effettuare esperimenti preliminari in cui la proteina esca può essere marcata sia all'estremità N-terminale sia a quella C-terminale; inoltre, laddove disponibile, la struttura tridimensionale dell'esca dovrebbe essere studiata minuziosamente per decidere dove sia meglio posizionare la marcatura. Infine, si dovrebbe evitare la sovra-espressione della proteina marcata nella cellula ospite poiché un'eccessiva concentrazione dell'esca potrebbe alterare i rapporti stechiometrici con i propri partners naturali portando spesso alla formazione di interazioni non specifiche o non naturali con le proteine ospiti.

Il sistema TAP-tag

150 Il sistema della duplice purificazione mediante marcatura per affinità (Tandem Affinity Purification tag system, TAP-

tag) è stato sviluppato per la purificazione dei complessi multiproteici in alte rese e in condizioni native (10). Secondo questa metodologia la proteina esca viene modificata con due marcature differenti, di solito separate da una sequenza specificamente idrolizzabile da un enzima. La proteina esca doppiamente marcata viene quindi espressa nella cellula in condizioni fisiologiche, interagisce con i suoi specifici partners intracellulari ed il complesso formato può essere purificato mediante due successivi passaggi di cromatografia di affinità. Questo sistema è estremamente flessibile e possono essere facilmente realizzate delle variazioni alla strategia originale come ad esempio la localizzazione della sequenza TAP-tag (tag cassette) sia all'estremità N-terminale che a quella C-terminale della proteina esca o l'introduzione di marcature alternative e l'ottimizzazione del sistema in diversi organismi ospite.

Nel sistema TAP-tag originalmente proposto, il gene codificante per la proteina di interesse viene modificato mediante aggiunta di un costrutto contenente due differenti "tags" di affinità, la Proteina A, specificamente riconosciuta dalle Immunoglobuline G (IgG) e un peptide capace di interagire fortemente con la Calmodulina. Tra le regioni codificanti per i due "tags" viene inserito un sito di idrolisi specifico per la Proteasi del Virus del tabacco (TEV). Il costrutto marcato può essere quindi introdotto transientemente o stabilmente nelle cellule o negli organismi di elezione. Nel migliore dei casi il vettore ricombinante dovrebbe rimpiazzare il gene endogeno della proteina naturale, sebbene questa condizione potrebbe non essere

sempre realizzabile. In tutti i casi comunque la sovra-espressione della proteina esca è evitata poiché il sistema del TAP-tag è stato specificamente disegnato per l'isolamento dei complessi multiproteici ai loro livelli naturali.

Nel primo step di purificazione, il complesso è isolato dall'intero estratto cellulare mediante interazione con particelle di agarosio derivatizzate con IgG; una volta allontanate le proteine non interagenti, il complesso sempre legato al supporto di agarosio è trattato con TEV che idrolizza il legame con la Proteina A e causa il rilascio del complesso. A questo punto il secondo passaggio di purificazione è condotto utilizzando l'affinità del secondo "tag" per la Calmodulina. Il complesso così purificato è frazionato nei suoi componenti mediante SDS-PAGE e le proteine partecipanti vengono identificate come descritto in precedenza. Questa metodologia di purificazione basata su un doppio passaggio di affinità riduce in maniera significativa la possibile presenza di proteine contaminanti non specifiche, diminuendo così sia il rumore di fondo non specifico sul gel di poliacrilammide sia la possibile presenza di falsi positivi. Questo metodo è stato originariamente sviluppato nei lieviti ed ha trovato un vasto impiego nella caratterizzazione dei complessi multiproteici nel *Saccharomyces cerevisiae*. Tuttavia, sono state sviluppate le condizioni ottimizzate per un generico uso della strategia TAP-tag.

Conclusioni e prospettive.

L'ambizioso progetto della mappa del proteoma umano (indirizzo web: <http://www.hupo.org>) rappresenterà la massima grande sfida scientifica della proteomica di espressione. Occorre sottolineare che, comunque, proprio per la natura dinamica del proteoma, il proteoma umano non dovrà risultare in una catalogazione delle differenti forme di espressione dei circa 40.000 geni dell'uomo ma piuttosto dovrà proporsi di rilevare le sottili relazioni esistenti tra le condizioni fisio-patologiche dell'individuo e la realizzazione del proprio programma dell'espressione genica. A tale scopo risulterà cruciale la messa a punto di metodi per un'ancora più rapida analisi di miscele complesse di proteine, accanto allo sviluppo di metodologie bioinformatiche in grado di gestire, archiviare e "dare un senso" all'enorme ammasso di dati prodotti.

Parallelamente, con il crescente aumento dei progetti di sequenziamento dei vari genomi che si apprestano ad arri-

vare a conclusione, la comprensione della funzione delle proteine e la definizione dei meccanismi cellulari a livello molecolare costituirà sempre più una delle maggiori esigenze della moderna biologia. Gli approcci di proteomica funzionale disponibili oggi si sono dimostrati essere utili per la rivelazione dei partners che interagiscono con le proteine bersaglio. Le prospettive future in questo tipo di studi saranno rivolte all'adattamento di questi approcci allo studio di sistemi *in vivo* attraverso la produzione di modelli animali in cui viene espressa la forma marcata della proteina di interesse. Le analisi di tipo proteomico dei complessi proteici che si formano *in vivo* riveleranno l'identità dei singoli componenti e se questi possano differire da un territorio all'altro. Si può confidare che attraverso la continua integrazione e l'impegno di varie componenti culturali del panorama scientifico sarà possibile affrontare con successo queste sfide con la finalità della comprensione della biologia come sistema.

Bibliografia

1. Taylor S.W., Fahy E., Ghosh S.S., Trends Biotechnol 2003, **21**, 82-88.
2. Patterson S.D., Aebersold R.H., Nat Genet 2003, **33**, 311-323.
3. Godovac-Zimmermann J., Brown L.R., Mass Spectrom Rev, 2001, **20**, 1-57.
4. Fields, S., Science, 2001, **241**, 1221-1224
5. Godovac-Zimmermann J., Brown L.R., Curr Opin Mol Ther 2003, **5**, 241-249.
6. Banks R.E., Dunn M.J., Hochstrasser D.F., et al. Lancet 2000, **356** 1749-1756.
7. Alban A., David S.O., Bjorkestén L., et al. Proteomics 2003, **3**, 36-44.
8. Gygi S.P., Rist B., Gerber S.A., Turecek F., Gelb M.H., Aebersold, R. Nat Biotechnol 1999, **17**, 994-999.
9. Alberts B. Cell, 1998, **92**, 291-294.
10. Gavin A.C., Bosche M., Krause R., Grandi P., Marzioch M., Bauer A., Schultz J., Rick J.M., Michon A.M., Cruciat C.M., Remor M., Hofert C., Schelder M., Brajenovic M., Ruffner H., Merino A., Klein K., Hudak M., Dickson D., Rudi T., Gnau V., Bauch A., Bastuck S., Huhse B., Leutwein C., Heurtier M.A., Copley R.R., Edelman A., Querfurth E., Rybin V., Drewes G., Raida M., Bouwmeester T., Bork P., Seraphin B., Kuster B., Neubauer, G., Superti-Furga, G. Nature, 2002, **415**, 141-147.
11. Lewis T.S., Hunt J.B., Aveline L.D. et al. Mol Cell 2000, **6**, 1343-1354.
12. Mann, M., Hendrickson, R. C., Pandey, A., Annual Review of Biochemistry, 2001, **70**, 437-473.