

UNIVERSITA' DEGLI STUDI DI NAPOLI

FEDERICO II

Dottorato di ricerca in Scienze Chimiche

XXIII Ciclo



**Identificazione di complessi proteici coinvolti in specifici processi
cellulari**

Tutore

Prof. Pietro Pucci

Dottoranda

Marianna Cozzolino

Coordinatore

Prof. Lucio Previtiera

أولئك الذين خبرة تزيد المعرفة ؛
الذي يزيد من السذاجة والخطأ

Chi fa esperienze accresce il sapere;

Chi è credulo aumenta l'errore.

Proverb. Arab. Erpen.57.

Indice

A. Riassunto	7
<i>A.1. Studio del “traffico” molecolare di SUMF1</i>	8
<i>A.2 Studio dell’interattoma del fattore trasrizzionale CBX7</i>	9
<i>A.3 Messa a punto di strategie di pre-frazionamento di complessi proteici</i>	10
1. Introduzione	12
<i>1.1 Il passaggio dall’era genomica all’era proteomica</i>	12
<i>1.2 Gen-“omica”, gen-“omica” funzionale e prote-“omica”</i>	14
<i>1.3 Determinazione delle interazioni proteina-proteina mediante approcci di Proteomica</i>	15
<i>1.3.1 La Proteomica Funzionale</i>	16
<i>1.3.1.1 Approcci basati su Affinity</i>	18
<i>1.3.1.2 Strategie di immunoprecipitazione</i>	20
<i>1.4 La Spettrometria di massa: tecnica di elezione per gli studi di Proteomica</i>	23
<i>1.4.1 Le metodologie in spettrometria di massa</i>	23
<i>1.4.2 La strumentazione</i>	25
<i>1.5 Database pubblici per le interazioni in proteina-proteina</i>	26
<i>1.6 Verifica delle interazioni</i>	27
<i>1.6.1 Co-IP</i>	27
<i>1.6.2 La microscopia confocale</i>	28
<i>1.7 Che cosa suggerisce oggi la Proteomica Funzionale</i>	29
<i>1.8 Le interazioni proteina-proteina nelle malattie umane</i>	30
<i>1.9 Scopo del progetto di tesi</i>	32
Bibliografia Introduzione	34
2. Studio del “traffico” molecolare di SUMF1	39
2.1 Introduzione	39
<i>2.1.1 Le Solfatasi</i>	39
<i>2.1.2 Le solfatasi umane</i>	41
<i>2.1.3 Modifiche Post-traduzionali delle Solfatasi</i>	43
<i>2.1.4 SUMF1</i>	44
2.2 Materiali e Metodi	45
<i>2.2.1 Materiali</i>	45

2.2.2 Metodi	46
2.2.2.1 <i>Colture Cellulari</i>	46
2.2.2.2 <i>Estrazioni delle proteine</i>	46
2.2.2.3 <i>Isolamento dei complessi proteici mediante tecniche di immunoprecipitazione</i>	46
2.2.2.4 <i>Precipitazione metanolo e cloroformio</i>	48
2.2.2.5 <i>SDS-PAGE</i>	48
2.2.2.6 <i>Western Blotting</i>	49
2.2.2.7 <i>SDS-PAGE preparativa</i>	49
2.2.2.8 <i>Colorazione</i>	49
2.2.2.9 <i>Deglicosilazione in situ</i>	50
2.2.2.9.1 <i>Rimozione delle bande proteiche dal gel e decolorazione</i>	50
2.2.2.9.2 <i>Riduzione ed alchilazione delle proteine in situ</i>	50
2.2.2.9.3 <i>Deglicosilazione delle proteine presenti nel gel</i>	50
2.2.2.9.4 <i>Lavaggi delle bande dopo deglicosilazione</i>	50
2.2.2.10 <i>Idrolisi in situ</i>	51
2.2.2.10.1 <i>Digestione in situ delle proteine presenti nel gel</i>	51
2.2.2.10.2 <i>Estrazione dei peptidi</i>	51
2.2.2.11 <i>Identificazione</i>	51
2.2.2.11.1 <i>Analisi mediante Matrix Assisted Laser Desorption Ionization (MALDI)</i>	51
2.2.2.11.2 <i>Analisi mediante LCMS/MS</i>	51
2.2.2.11.3 <i>Identificazione mediante Mascot</i>	52
2.3 Risultati	53
2.3.1 <i>SUMF1 è secreta e reinternalizzata nel Reticolo Endoplasmatico (RE)</i>	53
2.3.2 <i>Identificazione delle proteine che controllano la funzione ed il traffico di SUMF1</i>	58
2.3.3 <i>Il legame di PDI con SUMF1 è mediato sia da interazioni covalenti che non covalenti</i>	66
2.3.4 <i>Il legame di ERGIC-53 con SUMF1 è mediato da N-glicani</i>	68
2.3.5 <i>L'interazione di ERp44 con SUMF1 è mediata da tioli</i>	70

2.3.6 <i>PDI, ERGIC-53 ed ERp44 modulano co-operativamente la ritenzione e la secrezione di SUMF1 e la sua attivita' enzimatica</i>	73
2.3.7 <i>Distribuzione subcellulare di SUMF1</i>	75
2.4 Conclusione	77
Bibliografia SUMF1	81
3. Studio dell'interattoma del fattore trascrizionale CBX7	85
3.1 Introduzione	85
3.1.1 <i>CBX7 e la carcinogenesi tiroidea</i>	85
3.1.2 <i>CBX7 e i geni Polycomb</i>	86
3.1.3. <i>Pre-frazionamento di complessi proteici</i>	90
3.1.3.1 <i>Gradiente di ultracentrifugazione</i>	91
3.1.3.2 <i>Cromatografia ad esclusione molecolare</i>	92
3.2 Materiali e Metodi	93
3.2.1 <i>Materiali</i>	93
3.2.2 <i>Metodi</i>	93
3.2.2.1 <i>Colture Cellulari</i>	93
3.2.2.2 <i>Estrazioni delle proteine nucleari</i>	93
3.2.2.3 <i>Isolamento dei complessi proteici mediante tecniche di immunoprecipitazione</i>	94
3.2.2.4 <i>Prefrazionamento dei complessi proteici</i>	95
3.2.2.4.1. <i>Prepurificazione dei complessi mediante Gel Filtration</i>	95
3.2.2.4.2 <i>Prepurificazione dei complessi mediante Centrifugazione zonale in gradiente di densita'</i>	95
3.2.2.5 <i>Identificazione</i>	97
3.2.2.5.1 <i>Analisi mediante LCMS/MS</i>	97
3.2.2.5.2 <i>Identificazione mediante Mascot</i>	97
3.3. Risultati	98
3.3.1. <i>Identificazione delle proteine che interagiscono con CBX7</i>	98
3.3.1.1. <i>Caratterizzazione dei complessi molecolari coinvolgenti CBX7 in vivo</i>	105
3.3.1.2 <i>CBX7 inibisce l'attivita' di HDAC2</i>	106

3.3.1.3 <i>CBX7 lega il promotore del gene dell'E-cadherin</i>	107
3.3.1.4 <i>CBX7 regola positivamente il promotore di E-cadherin</i>	109
3.3.1.5 <i>I livelli di espressione di CBX7 ed E-cadherin sono correlati nei carcinomi tiroidei umani</i>	110
3.3.2 <i>Separazione dei complessi di CBX7 mediante tecniche di Frazionamento</i>	111
3.3.2.1. <i>Separazione mediante centrifugazione in gradiente di densita'</i>	111
3.3.2.2 <i>Separazione mediante Gel Filtration</i>	116
3.4 Conclusioni	121
Bibliografia CBX7	125

A. Riassunto

Una completa descrizione della complessa rete di meccanismi cellulari e l'uso di network per predire la maggior parte dei comportamenti cellulari sono i principali obiettivi della system biology. Un ruolo chiave nella biologia contemporanea è svolto dalla proteomica funzionale che si propone di fornire delucidazioni sulle funzioni delle proteine e sulla definizione dei meccanismi cellulari a livello molecolare. Il conseguimento di questi obiettivi dipende principalmente dall'identificazione di proteine all'interno di complessi funzionali in vivo. L'isolamento delle proteine che interagiscono tra loro è effettuato mediante procedure di cromatografia di affinità o di immuno-precipitazione nelle quali l'esca proteica ed i suoi specifici *partners* possono essere isolati utilizzando uno specifico ligando immobilizzato su supporti solidi insolubili. Questi approcci conducono all'identificazione di molte proteine che appartengono a complessi distinti dotati di funzioni biologiche diverse. Attribuire l'appartenenza di ogni proteina ad un specifico complesso costituisce un ingente problema che può essere solo parzialmente risolto usando database di interazione proteina-proteina ed informazioni di letteratura.

È possibile che lo sviluppo delle metodologie di pre-frazionamento per separare complessi proteici preservando le loro interazioni native rappresenti una soluzione essenziale per il futuro della proteomica funzionale. La pre-purificazione dei complessi può essere intrapresa solamente in condizioni native sulla base delle loro caratteristiche chimico-fisiche, come la grandezza e dimensione (cromatografia di gel-filtrazione), la densità (ultracentrifugazione), ecc. In seguito al pre-frazionamento, il complesso associato ad una specifica funzione biologica può essere isolato usando tecniche di purificazione per affinità.

Approcci di proteomica funzionale capaci di descrivere singole proteine appartenenti a complessi coinvolti in specifiche funzioni cellulari avrà un forte impatto su futuri studi della system biology.

In particolare il progetto di tesi di dottorato si è articolato nei seguenti punti:

- ✓ Studio del "traffico" molecolare di SUMF1.
- ✓ Studio dell'interattoma del fattore trascrizionale CBX7.
- ✓ Messa a punto di strategie di pre-frazionamento di complessi proteici.

✓ *A.1. Studio del “traffico” molecolare di SUMF1*

SUMF1 è una proteina del reticolo endoplasmatico (RE) capace di attivare tutte le solfatasi mediante la modifica post- traduzionale di una cisteina presente nel loro dominio catalitico in formilglicina. Questa unica modifica è essenziale per l'attività delle solfatasi. Infatti, mutazioni nel gene di SUMF1 rendono inattive tutte le solfatasi causando una grave malattia, la Multiple Sulfatase Deficiency (MSD), che colpisce diversi organi e tessuti, soprattutto quello nervoso. Studi precedenti hanno dimostrato che SUMF1 può essere secreta nel mezzo di coltura cellulare e poi reinternalizzata da altre cellule che la localizzano nuovamente nel RE, dove attiva le solfatasi endogene. Tuttavia, i pathways molecolari attraverso cui SUMF1 è secreta dalle cellule o è reinternalizzata dal mezzo sono ancora in gran parte sconosciuti. Per delucidare il meccanismo che controlla il trafficking e la funzione di SUMF1, sono stati identificati i partners della proteina mediante esperimenti di immunoprecipitazione seguiti da analisi con spettrometria di massa.

In particolare la proteina SUMF1, coniugata al peptide FLAG, è stata espressa in una linea cellulare umana. Tali cellule sono state poi lisate ed il complesso o i complessi sovramolecolari coinvolgenti l'esca, sono stati purificati mediante cromatografia di affinità. Quest'ultima prevede il riconoscimento specifico del peptide FLAG, coniugato all'esca, mediante un anticorpo legato covalentemente ad un supporto solido di agarosio. Le miscele proteiche, dell'immunoprecipitato e di un opportuno controllo, costituito dall'immunoprecipitato derivante da cellule non sovraesprimenti la proteina esca coniugata con il FLAG, sono state poi frazionate mediante elettroforesi monodimensionale e il gel colorato con Comassie colloidale. Le bande proteiche contenenti i campioni sono state escisse dal gel, ridotte, alchilate e digerite in situ con tripsina. La miscela di peptidi risultante è stata sottoposta a esperimenti di LCMSMS che abbinano procedure di cromatografia liquida capillare con analisi mediante spettrometria di massa tandem. Mediante quest'ultimo approccio è stato possibile ottenere informazioni relative alla sequenza di uno o più frammenti proteolitici e quindi risalire all'identificazione delle proteine che interagiscono specificamente con SUMF1. I dati ottenuti hanno consentito di identificare alcuni partners molecolari della proteina esca, quali PDI, ERp44 and ERGIC-53, consentendo di definire il percorso molecolare di SUMF1 durante il processo di secrezione.

✓ *A.2 Studio dell'interattoma del fattore trasizionale CBX7.*

Nel secondo anno di Dottorato l'attività scientifica è stata indirizzata all'impiego delle metodologie di proteomica funzionale per l'identificazione dei partners molecolari del fattore trascrizionale Cbx7 (Chromobox protein homologue 7), una proteina di 251 amminoacidi che possiede un chromodomain nella regione N-terminale. Cbx7 è una proteina appartenente alla famiglia delle chromobox e membro del polycomb repressive complex 1 che, insieme con il polycomb repressive complex 2, inibisce l'espressione di specifici geni. La forma murina di Cbx7 può associarsi facoltativamente all'eterocromatina e al cromosoma X inattivo, suggerendo un eventuale ruolo nella repressione della trascrizione genica. E' stato recentemente dimostrato che il gene *cbx7* è drasticamente down-regolato in sei diverse linee cellulari con carcinoma tiroideo. Inoltre, l'analisi dell'espressione di Cbx7 in un numero elevato di campioni di carcinoma tiroideo rivela una progressiva riduzione dei livelli di Cbx7 che risulta quindi correlato con il grado di malignità delle neoplasie tiroidee. Altri studi correlano addirittura la bassa espressione di Cbx7 con una ridotta sopravvivenza in pazienti con carcinoma al colon. Il ripristino dell'espressione di Cbx7 nelle cellule con cancro tiroideo riduce la loro velocità di crescita, indicando che Cbx7 gioca un ruolo critico nella regolazione della proliferazione delle cellule tiroidee trasformate.

Per delucidare il meccanismo attraverso cui la ridotta concentrazione del fattore Cbx7 è coinvolta nella carcinogenesi sono stati isolati ed identificati i complessi proteici che la coinvolgono. In particolare la proteina Cbx7, coniugata al peptide V5, è stata sovraespressa in una linea cellulare umana. Tali cellule sono state poi lisate ed il complesso o i complessi sovramolecolari coinvolgenti l'esca, sono stati purificati mediante immunoprecipitazione grazie ad un anticorpo specifico per il peptide V5 legato covalentemente ad un supporto solido di agarosio. Le miscele proteiche, dell'immunoprecipitato e di un opportuno controllo, costituito da cellule non esprimenti la proteina esca coniugata con il V5, sono state poi frazionate mediante elettroforesi monodimensionale e le bande proteiche contenenti i campioni, dopo colorazione con Comassie colloidale, sono state escisse dal gel, ridotte, alchilate e digerite in situ con tripsina. La miscela di peptidi risultante è stata sottoposta a esperimenti di LCMSMS che abbinano procedure di cromatografia liquida capillare con analisi mediante spettrometria di massa tandem. Mediante quest'ultimo approccio è stato possibile ottenere informazioni relative alla sequenza di uno o più frammenti proteolitici e quindi risalire all'identificazione delle proteine che interagiscono specificamente con Cbx7.

I dati ottenuti hanno consentito di definire alcuni nuovi partners molecolari della proteina Cbx7. Tra di essi è stata studiata in particolare l'interazione con l'istone deacetilase 2 (HDAC2). Tale proteina

gioca un ruolo chiave nella trasformazione delle cellule neoplastiche e nella down-regolazione dell'espressione di E-cadherin, evento frequente durante la progressione del tumore.

E' stato dimostrato che Cbx7 interagisce con HDAC2 e ne inibisce l'attività. Inoltre entrambe le proteine legano il promotore dell'E-cadherin, ma esercitano funzioni diametralmente opposte. Infatti, il legame di Cbx7 decresce l'effetto di inibizione esercitato da HDAC2 sull'espressione di E-cadherin. Quindi, la riduzione dei livelli di espressione di Cbx7 in cellule tumorali potrebbe comportare una ridotta capacità della cellula di regolare positivamente l'espressione di E-cadherin, inibita da HDAC2, e quindi correlare con un fenotipo maligno in pazienti affetti da malattie neoplastiche.

✓ **A.3 Messa a punto di strategie di pre-frazionamento di complessi proteici.**

L'esperimento di proteomica mediante il quale si è cercato di delucidare il meccanismo attraverso cui il fattore Cbx7 controlla la proliferazione cellulare, ha fornito una lunga lista di putativi interattori del fattore trascrizionale, indicando chiaramente che Cbx7 è coinvolta in numerosi complessi probabilmente dotati di diversa funzione biologica. Come in altre applicazioni di proteomica, il compito più impegnativo in questi esperimenti è l'interpretazione dei dati e in particolare il tentativo di assegnare i singoli componenti proteici allo specifico complesso cui essi appartengono, risalendo quindi alla funzione biologica dell'intero complesso. Nonostante la ricchezza delle informazioni fornite da banche dati di proteine e database di interazioni proteina-proteina e la disponibilità di strumenti software, la deduzione di un significato biologico dalla lista degli interattori rimane ancora un problema di difficile soluzione. Queste difficoltà potrebbero essere superate dall'impiego di tecniche di pre-frazionamento dell'estratto proteico di partenza in grado di preservare le interazioni native tra le proteine. In tal modo si tenderebbe ad isolare i singoli complessi funzionali ridimensionando la complessità dei dati di proteomica funzionale.

La pre-purificazione dei singoli complessi coinvolgenti CBX7 è stata realizzata sulla base delle loro caratteristiche chimico-fisiche, sfruttando le dimensioni e la forma o la densità diverse dei vari complessi. L'estratto nucleare totale è stato quindi preventivamente sottoposto a frazionamento mediante cromatografia ad esclusione molecolare o ultracentrifugazione in gradiente di glicerolo.. Le diverse frazioni ottenute da queste pre-purificazioni sono state analizzate utilizzando un anticorpo anti-V5, per individuare quelle contenenti Cbx7. Le frazioni positive all'immunorivelazione sono state selezionate, mentre le altre sono state scartate, ed i complessi multiproteici sono stati singolarmente isolati mediante immunoprecipitazione. La diminuzione della complessità dell'estratto proteico nucleare con queste procedure ha comportato una notevole diminuzione delle proteine non specifiche abbassando notevolmente il rischio di identificazione di falsi positivi. Inoltre, tutti i partners localiz-

zati nella stessa frazione appartengono verosimilmente allo stesso complesso funzionale, rendendo così l'interpretazione dei dati molto più semplice.

In particolare, dall'esperimento di ultracentrifugazione in gradiente di glicerolo sono stati identificati una serie di componenti del complesso PRC1 quali BMI-1, hPH1, hPH2, hPH3, RING1B, in ottimo accordo con quanto riportato in letteratura. Il complesso PRC1 è formato da almeno un rappresentante di ciascuna delle quattro famiglie di proteine altamente conservate che ne costituiscono il core: Polycomb (CBX2/4/6/7/8), Polyhomeotic (HPH1/2/3), Posterior sex combs (PSC1-6) e Sex combs extra (RING 1A/1B). In aggiunta alle quattro subunità del core di PRC1, è stato riportato che anche altre proteine associano con il complesso come BCOR, che è capace di intervenire nell'ubiquitinazione di H2A.

La cromatografia ad esclusione molecolare, invece, ha portato all'identificazione di proteine coinvolte nel processing dell'RNA, molte delle quali appartenenti al complesso di splicing, suggerendo che CBX7 sia strettamente connessa con l'RNA. Sta diventando, infatti, sempre più evidente che i processi di regolazione della cromatina, di trascrizione e splicing sono strettamente connessi. Sebbene sia noto che CBX7 si associa alla cromatina in maniera RNA dipendente, non è chiaro se quest'interazione sia diretta o indiretta. Il ritrovamento di proteine che legano l'RNA potrebbe far supporre che il legame sia indiretto mediato da tali proteine.

Capitolo I

Introduzione

1.1 Il passaggio dall'era genomica all'era proteomica

Negli ultimi decenni ci sono state un gran numero di rivoluzioni in biologia molecolare. Quella più recente e che ha richiesto un enorme sforzo è stata il completamento della sequenza del genoma umano. Oltre alle informazioni inerenti alla stessa sequenza, il genoma fornisce ai ricercatori un inventario di geni con i quali perseguire indagini globali riguardo al funzionamento delle cellule e degli organismi. La genomica e le altre discipline "omiche" forniscono, infatti, un'infrastruttura per la tradizionale ricerca biologica nella comprensione delle funzioni dei singoli geni e i meccanismi di specifici processi cellulari [1]. Il genoma umano, tuttavia, permette inoltre di predire le proteine che possono essere potenzialmente generate, ma non dove e quando o a che livello; non può informarci sul tipo di cellule in cui le proteine saranno espresse o in quali fasi di sviluppo o di differenziazione questo accadrà. Né può prendere in considerazione l'enorme diversificazione di strutture che risultano da splicing, inserzione di geni, delezioni, ricombinazioni e altri tipi di riarrangiamenti. La struttura genica da sola ci dice poco circa la funzione fisiologica delle proteine poiché ignora le modifiche co- e post-traduzionali cui le proteine sono soggette. La regolazione dei processi cellulari richiede una miriade di comandi, positivi e negativi, per mantenere sotto controllo le reazioni che prendono luogo ed essere sicuri che nessun evento cruciale avvenga a un tempo o in un luogo inappropriato. Noi sappiamo oggi che il segnale principale che orchestra queste reazioni si basa sulla modifica delle proteine [2]. Il motivo è un'evidenza palese della non validità del dogma centrale secondo cui a un gene corrisponde una proteina.

Da una prospettiva evuzionistica, l'apparente complessità osservata a livello di un organismo non necessariamente deve riguardare le dimensioni del suo genoma.

Basti pensare, infatti, che l'uomo possiede 25.000 geni, più o meno quanti ne ha il vermicciattolo *Caenorhabditis elegans*, il cui corpo è costituito di solo mille cellule, e decisamente meno di quanti ne abbia una piantina alta una diecina di centimetri, la *Arabidopsis thaliana*, organismi dalla struttura e dal comportamento decisamente meno complessi dei nostri. La complessità di un organismo quindi potrebbe essere correlata alla capacità delle sue proteine di interagire di continuo fra di loro a formare complessi funzionali.

Le funzioni biologiche non sono solo regolate da un genoma statico, ma anche da una popolazione dinamica di proteine mediante un gioco di regolazione di geni e di proteine con influenze extracellulari [1]. Funzioni cellulari vitali come la replicazione del DNA, la trascrizione e la traduzione del-

l'mRNA richiedono l'azione coordinata di un certo numero di proteine che vengono assemblate in una matrice di complessi multiproteici di composizione e struttura diversa. Allo stesso modo, i processi biologici sono orchestrati e regolamentati mediante reti di *signalling* dinamici che connettono stimoli chimici o fisici a specifiche molecole effettrici. L'analisi dei complessi proteici e delle reti d'interazione proteina-proteina sono quindi di centrale importanza nella ricerca biologica dell'era postgenomica [3].

Anche se le proteine sono i principali effettori del messaggio genomico, esse provvedono solo ad una parte del messaggio. Infatti, esse raramente agiscono da sole. Complesse funzioni cellulari sono quasi sempre il risultato dell'azione coordinata di molte proteine che agiscono in assemblaggi molecolari o pathway per realizzare una determinata attività. Si pensa che gli schemi forniti dalle reti d'interazione proteica possano contribuire alla nostra comprensione sui rapporti tra genomi e fenotipi.

1.2 Gen-“omica”, gen-“omica” funzionale e prote-“omica”

Negli ultimi anni, una sconcertante e talvolta fuorviante terminologia è sorta in seguito gli esperimenti biologici su larga scala. La Fig. 1.1 da una visione semplificata delle principali categorie [1].

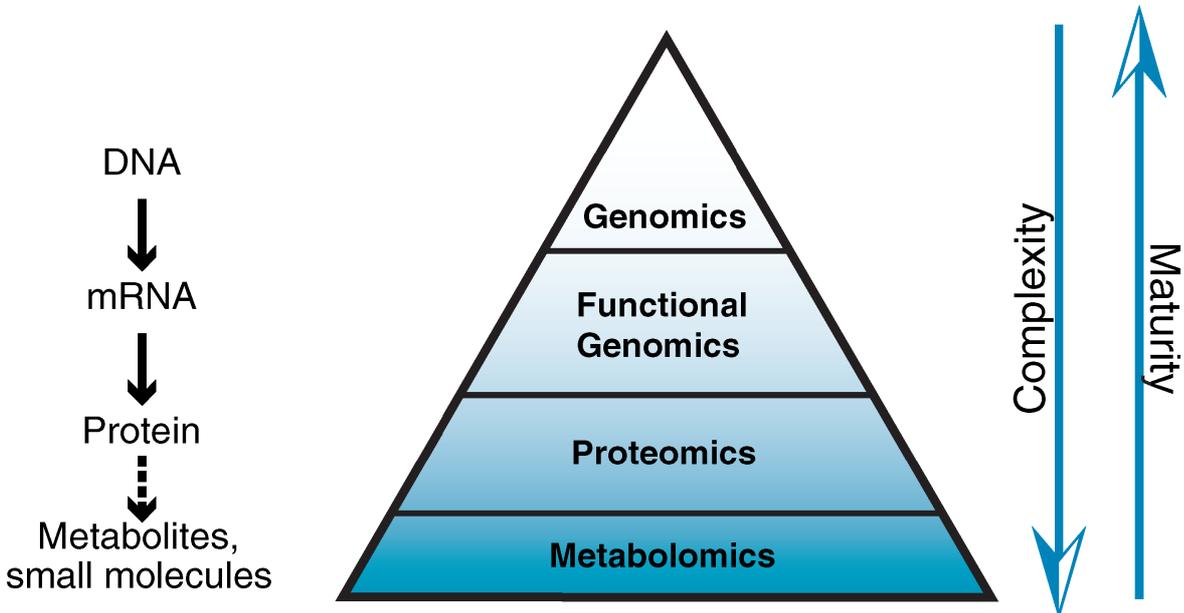


Fig. 1.1. Relazione schematica delle diverse discipline “omiche” in relazione al flusso di informazione dal genoma attraverso i trascritti alle proteine e piccole molecole.

Muovendo dalla genomica alla proteomica, la complessità aumenta drasticamente mentre la maturità della tecnologia diminuisce.

La *genomica* mira a determinare la sequenza cromosomica lineare di un organismo modello, così come le differenze nelle sequenze tra gli individui. Annotazione di genomi, compresa la definizione delle sequenze codificanti e sequenze di regolazione, fanno comunque parte di quella che è definita genomica. Il livello successivo verso il basso procede con i trascritti [RNA messaggero (mRNA)]. Non vi è necessariamente una relazione uno-a-uno tra la sequenza del genoma e il trascritto; le cellule usano lo splicing alternativo per generare diversi trascritti da un singolo gene e tutta una serie di meccanismi per controllare la produzione, lo stato della trascrizione, e la degradazione dell'mRNA. Inoltre per la misura globale dei livelli di trascrizione in miriadi di condizioni, la genomica funzionale comprende diverse strategie in cui è eliminato il messaggio e sono osservati gli effetti cellulari conseguenti. Questa parte della *genomica funzionale*, quando eseguita su organismi interi, è anche chiamata *fenomica* perché è osservato il fenotipo di un organismo. Ci sono molti altre strategie di genomica funzionale. Queste sono tutte eseguite a livello di oligonucleotidi, piuttosto che a livello di proteine.

Le proteine, i vettori principali dell'attività biologica, rappresentano il successivo livello di complessità. La funzione delle proteine dipende dalla precisa sequenza di aminoacidi, dalle modifiche (soprattutto quelle di regolazione, come la fosforilazione), dalla struttura tridimensionale (3D), dalla concentrazione proteica, dall'associazione con altre proteine, e dall'ambiente extracellulare. Di conseguenza, la *proteomica* è volta a stabilire la struttura di una proteina, le modifiche, la localizzazione e le interazioni proteina-proteina in aggiunta ai livelli di espressione proteica. Attualmente la spettrometria di massa (MS) è la tecnologia più versatile per poter misurare direttamente le proteine endogene [4]. Tecnologie non basate sulla spettrometria di massa in proteomica coinvolgono sistemi basati su array (ad esempio, il saggio del doppio ibrido per le interazioni proteina-proteina) o su strumenti strutturali e di imaging. Molte proteine sono enzimi o regolano la funzione degli enzimi che in ultima analisi determinano il livello di piccole molecole (metaboliti). Una visione globale della cellula richiede che siano misurate e modellate anche i metaboliti. La corrispondente disciplina su larga scala, che è la più avanzata nel campo dell'ingegneria e delle analisi di flusso del metabolita, è talvolta chiamato *metabolomica*. Approcci sperimentali in metabolomica spesso coinvolgono la spettrometria di massa, come in proteomica.

Queste definizioni non si devono però considerare assolute. Se da un lato infatti vi è generale un accordo sulla nomenclatura e la delimitazione di queste discipline, dall'altro diversi approcci sperimentali in queste categorie spesso si sovrappongono. Pertanto nessuno degli approcci sopra citati può da solo delucidare una complessa funzione biologica. Piuttosto, ogni disciplina con la sua peculiarità può fornire un contributo a chiarire un quadro al quanto complesso.

1.3 Determinazione delle interazioni proteina-proteina mediante approcci di Proteomica

Le proteine quasi mai lavorano completamente da sole. Sono spesso i membri di complessi multiproteici, tanto che è persino difficile parlare della loro funzione o della loro struttura 3D al di fuori di questo contesto, come ben illustrato dalla struttura ai raggi X del ribosoma [5]. A parte la formazione di complessi multiproteici stabili, le proteine associano temporaneamente con i loro targets per modificarli, regolarli mediante effetti sterici, o farli traslocare in diversi compartimenti cellulari. Tutti questi meccanismi sono intimamente connessi alla funzione delle proteine e i partners con cui interagiscono sono tra i componenti più utili nella comprensione della biologia di una proteina.

Le interazioni proteiche sono state studiate per decenni mediante immunoprecipitazioni e Western blotting o microsequenziamento delle proteine, ma è da pochi anni che i ricercatori hanno cominciato ad affrontare la problematica dell'intero "interattoma", cioè la totalità delle interazioni proteina-proteina in una proteoma.

Il proteoma è stato originariamente definito come l'intero corredo proteico espresso da un genoma. Tuttavia, questa definizione non tiene conto che il proteoma è un'entità molto dinamica che cambia in base allo stato cellulare e l'ambiente extracellulare. Pertanto, la definizione di proteoma dovrebbe specificare che esso è il corredo proteico di una cellula in un determinato periodo di tempo specificato, compreso dell'insieme di tutte le isoforme della proteina e delle modifiche proteiche. Anche se la proteomica tenta di fare uno sforzo su larga scala, gran parte del suo fascino risiede nella sua capacità di focalizzare i suoi strumenti su popolazioni selezionate di proteine in circostanze specifiche, contribuendo direttamente al problema funzionale e meccanicistico [1].

1.3.1 La Proteomica Funzionale

La capacità di ogni organismo di sopravvivere dipende da diversi complessi meccanismi cellulari che soddisfano tutte le funzioni necessarie per la vita delle cellule. Inoltre, le cellule devono essere in grado di modificare i modelli di espressione genica in risposta ai segnali extra e intracellulari al fine di garantire che il sottoinsieme appropriato di geni è espresso nel momento giusto [6, 7]. In passato, i meccanismi cellulari e di trasduzione del segnale sono stati pensati come un insieme di processi isolati o percorsi lineari dove le funzioni biologiche si susseguivano una dopo l'altra in un modello tempo-dipendente. Oggi, tuttavia, è ben noto che i meccanismi biologici fondamentali sono esercitati da un gran numero di proteine raccolte insieme per formare temporaneamente o transientemente grandi complessi funzionali [8], o pathways cellulari [9]. Una volta svolte le loro funzioni, questi complessi dissociano formando singoli componenti liberi di associare ad altri complessi per svolgere attività distinte. Per complicare ulteriormente il quadro, meccanismi diversi e molti pathways di signalling sono in grado di interagire l'uno con l'altro in modo da formare una rete complessa, che integra contemporaneamente segnali extracellulari e intracellulari. Questi processi dinamici costituiscono i network molecolari attraverso cui le informazioni di signalling fluiscono all'interno della cellula [10,11]. Una descrizione completa di questo complesso intrico di vie di trasduzione del segnale e l'uso di network per predire l'intera gamma di comportamenti cellulari sono gli obiettivi principali della system biology [12, 13]. Un ruolo chiave nel campo delle scienze biologiche contemporanee e future può essere svolto dalla proteomica funzionale, un settore emergente nel campo della ricerca proteomica indirizzata verso il chiarimento delle funzioni proteiche e la definizione dei meccanismi cellulari a livello molecolare [14, 15]. Una descrizione completa dei processi cellulari è infatti strettamente dipendente da una chiara definizione dei complessi multi proteici che vi prendono parte e delle singole componenti proteiche coinvolte in queste entità funzionali. L'associazione di una proteina sconosciuta con partners appartenenti ad uno specifico complesso di proteine coinvolto in un meccanismo particolare potrebbe fortemente suggerire la sua funzione biologica

[16]. Inoltre, una descrizione dettagliata dei pathways di signalling cellulare potrebbe trarre grande beneficio dalla delucidazione delle interazioni proteina-proteina in vivo.

L'isolamento di interi complessi multi-proteici può essere svolto con approcci basati sul concetto di affinità biologica (affinity) (esperimenti di "pull down" e "TAP-Tag"), nelle quali uno specifico ligando molecolare viene utilizzato per legare la proteina di interesse ed i suoi partners proteici [17-19]. In alternativa, le strategie di immunoprecipitazione possono essere usate per isolare sia la proteina endogena [20] sia una versione ricombinante dell'esca marcata con un epitopo "tag", approfittando della disponibilità di diversi sistemi anti tag dotati di elevata efficacia di legame [21]. In entrambi i casi è necessario prevedere tutta una serie di esperimenti di controllo a causa del grande numero di possibili falsi positivi.

I componenti proteici dei complessi funzionali vengono eluiti, frazionati mediante SDS-PAGE ed identificati con tecniche LC-MS/MS [22, 23]. Questi approcci di solito restituiscono come risultato lunghe liste di interattori identificati in un singolo esperimento di proteomica funzionale [24, 25]. È ormai chiaro, infatti, che una singola proteina può essere coinvolta in diversi complessi, ognuno dotato di una propria specifica funzione biologica. Pertanto, il tentativo di razionalizzare l'elenco degli interattori all'interno di un quadro biologico significativo costituisce attualmente un grave problema in proteomica funzionale. Anche la semplice verifica di tutta la serie di proteine che interagiscono con la proteina esca con differenti procedure sperimentali, ad esempio co-immunoprecipitazioni, etc, è spesso impossibile a causa del numero veramente elevato di candidati. Due approcci sono normalmente utilizzati per gestire tali elenchi, la strategia più semplice consiste nella scelta di uno o pochi interattori tra le varie proteine candidate in conformità a dati precedenti o informazioni di letteratura. L'interazione di queste proteine con l'esca può essere confermata e investigata funzionalmente [26-28]. Un secondo approccio più impegnativo e dispendioso in termini di tempo mira ad assegnare ogni proteina ad un complesso funzionale specifico, un compito che corrisponde a risolvere un gran numero di puzzle di "cento pezzi" in un unico tempo e con gli stessi pezzi [29]. Come accennato in precedenza, infatti, ogni proteina potrebbe appartenere simultaneamente a diversi complessi funzionali. Questo tremendo problema può essere solo parzialmente risolto utilizzando database d'interazione proteina-proteina e informazioni di letteratura, ma, quando ha successo, potrebbe essere indicativo di nuove, non svelate funzioni biologiche della proteina esca [30]. L'obiettivo finale di questo secondo approccio sarebbe il raggiungimento di una visione globale funzionale dell'interattoma della proteina esca che deve poi essere inserita in un adeguato studio di system biology [31, 32].

Nonostante il continuo aumento dei database di interazioni proteina e strumenti software [33, 34], l'associazione delle singole proteine ad uno specifico complesso proteico costituisce tuttora il "collo

di bottiglia" di qualsiasi approccio di proteomica funzionale. Lo sviluppo di diverse strategie a monte basate sul pre-frazionamento degli estratti proteici per separare i singoli complessi proteici preservando le loro interazioni native potrebbe quindi rappresentare uno strumento essenziale per il futuro della proteomica funzionale.

1.3.1.1 Approcci basati su Affinity

L'identificazione di proteine che interagiscono in complessi stabili o anche transienti mediante procedure di affinity sfrutta l'affinità intrinseca di un ligando specifico per la proteina d'interesse. La molecola ligando è covalentemente legata ad un supporto insolubile (di solito sfere di agarosio/Se-farosio) e usata come esca per legare selettivamente la proteina di interesse ed i suoi partners ("pull down"). Teoricamente non ci sono limitazioni nella scelta delle molecole di legante che può essere impiegato in esperimenti di pull-down. Numerosi esempi sono stati riportati utilizzando substrati modificati e/o inibitori di proteine selezionate con specifiche attività enzimatiche [35-37]. In altri casi sono stati utilizzati anche farmaci nativi e/o modificati indirizzati verso specifici target proteici [38]. A questo proposito, una questione particolarmente interessante che deriva da questa strategia consiste nella possibilità di individuare nuove molecole bersaglio precedentemente sconosciute per un farmaco specifico. In una variante più diffusa della strategia di "fishing", la proteina di interesse (esca) può essere espressa con un tag adeguato, come ad esempio fusa alla GST-, con una coda di poli-His etc, ed immobilizzata sul supporto insolubile approfittando della disponibilità di vari sistemi anti-tag dotati di alta efficienza di legame. L'esca immobilizzata può quindi essere utilizzata per pescare i propri partner specifici all'interno di un intero estratto cellulare [21]. Questi sistemi di "tag affinity" forniscono un'applicabilità generale, con un gran numero di proteine e un effetto minimo sulla struttura terziaria e l'attività biologica delle esche, evitando l'instabilità dei complessi. Il successo di un approccio basato sull'affinity dipende dalla mancanza di eccessive interazioni aspecifiche, che a sua volta è collegata alla specificità del riconoscimento esca-partners. Non sorprende quindi che la strategia di fishing ha trovato grande applicazione nell'isolamento di proteine leganti acidi nucleici, in cui la specificità di legame è molto elevata, in studi su complessi di trascrizione e/o di traduzione o sulla macchina dello splicing. Anche oligonucleotidi di sintesi che comprendono specifiche sequenze di coppie di base o frammenti di DNA o RNA amplificati per PCR possono essere immobilizzati covalentemente su supporti inerti disponibili in commercio e utilizzati per l'isolamento delle proteine coinvolte nel legame specifico della sequenza nucleotidica [39-41]. Anche un tag di biotina può essere aggiunto a oligonucleotidi sintetici; l'immobilizzazione dell'esca sul supporto solido sfrutterà quindi il vantaggio dalla sua alta affinità con le molecole di streptavidina [42] (Fig. 1.2).

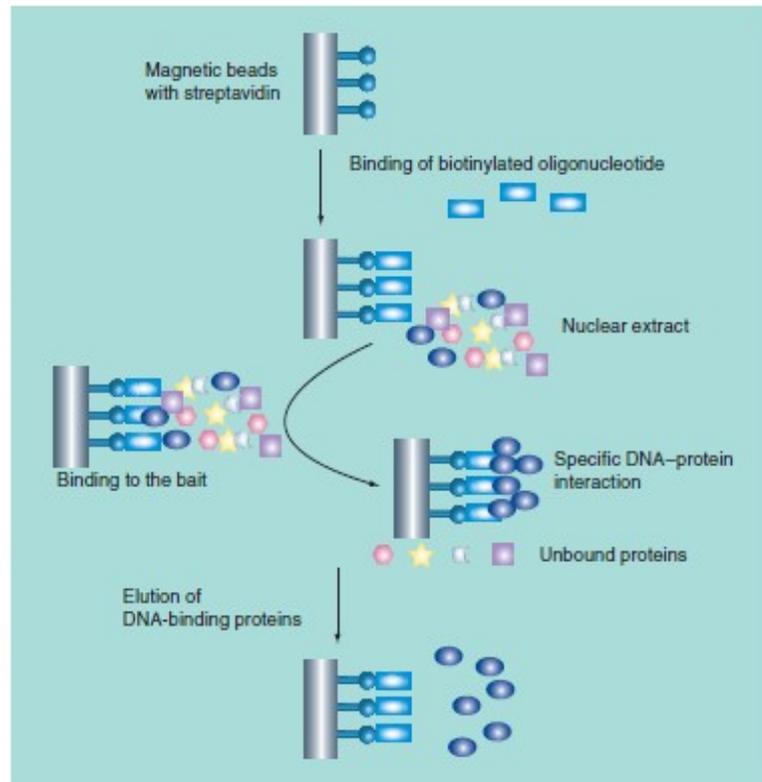


Fig. 1.2.: Strategia di isolamento di proteine che legano il DNA.

In tutti i casi l'estratto proteico, totale o frazionato è incubato con la proteina immobilizzata. Le componenti proteiche espressamente riconosciute dall'esca possono quindi essere eluite e frazionate mediante SDS-PAGE. Le bande proteiche sono escisse dal gel, digerite con enzimi proteolitici e le miscele peptidiche derivanti analizzate mediante tecniche LC-MS/MS, che portano alla identificazione delle proteine attraverso la ricerca in database [43]. Questa sequenza sperimentale, SDS-PAGE/ digestione in situ (su gel)/LC-MS/MS, di solito è preferibile ad altri metodi esistenti, perché l'SDS-PAGE, anche se dotata di una risoluzione molto bassa, è abbastanza sensibile da permettere l'analisi di piccole quantità di complessi proteici isolati attraverso procedure di affinity e/o immunoprecipitazione (vedi sotto). Quindi, poiché una singola banda da gel monodimensionale spesso può contenere numerose proteine diverse, le metodologie LC-MS/MS, essendo in grado di analizzare miscele di peptidi molto complesse, sono abitualmente utilizzate per l'identificazione delle proteine. Gavin et al. [31] e Krogan et al. [32] hanno usato una versione leggermente diversa della strategia di purificazione per affinity nota come TAP-Tag (Tandem Affinity Purification) per l'isolamento di numerosi complessi multiproteici in cellule di lievito. Questo metodo utilizza due diversi tag per la stessa proteina di solito intervallati da una sequenza linker che può essere idrolizzata da un enzima. Il sistema TAP Tag originale consiste in due domini di legame alle IgG della proteina A da *Staphylococcus aureus* (Prot. A) e un peptide legante la calmodulina (CBP), separati da un sito di idrolisi

specifico per la proteasi del virus del mosaico tabacco, TEV. Il vettore ricombinante sostituisce il gene endogeno wild-type, evitando così l'over-espressione della proteina esca e consentendo il recupero di complessi proteici espressi al loro proprio livello naturale. Benché ampiamente utilizzata, la cattura dei complessi per affinità soffre di un enorme numero di inconvenienti. Oltre a una serie di problemi tecnici, durante la preparazione del lisato cellulare, l'organizzazione dei compartimenti subcellulari è distrutta e le proteine che sono normalmente segregate in organelli diversi possono entrare in contatto, generando interazioni specifiche, ma non fisiologiche. Ancora più importante, in questi esperimenti tutte le interazioni hanno luogo sulle beads derivatizzate in un ambiente non cellulare, con diverse condizioni, come ad esempio il rapporto molare tra l'esca e i suoi partners proteici, in gran parte alterate.

1.3.1.2 Strategie di immunoprecipitazione

Diverse strategie basate su tecniche di immunoprecipitazione sono state introdotte per l'isolamento dei complessi proteici al fine di superare la maggior parte degli inconvenienti proposti dagli approcci basati sull'affinità [44, 45]. Queste procedure essenzialmente si possono suddividere in due categorie, immunoprecipitazione diretta della proteina esca endogena da estratti cellulari o produzione *in situ* di una forma modificata della proteina esca seguita da immunoprecipitazione con anticorpi anti-tag. L'approccio classico di immunoprecipitazione della proteina endogena consiste nel legare l'anticorpo diretto contro la proteina esca su beads di agarosio e/o sefarosio derivatizzate con la proteina A-(o G). Dato che la porzione di anticorpo riconosciuta dalla proteina A (o G) è diverso dal sito di legame dell'antigene, l'anticorpo immobilizzato è ancora in grado di immunoprecipitare la proteina bersaglio e può quindi essere utilizzato per isolare i suoi complessi nativi. In alternativa, l'esca proteica è immunoprecipitata dall'intero estratto proteico utilizzando una forma solubile di anticorpi in opportune condizioni sperimentali al fine di evitare la dissociazione dei complessi. Una volta che si è verificato lo specifico legame antigene-anticorpo, le beads derivatizzate con Proteina A-(o G) sono aggiunte al campione per "pescare" i complessi presenti nell'estratto proteico. In entrambi i casi, dopo diversi lavaggi, gli interattori dell'esca possono essere eluiti dalle beads in condizioni denaturanti e le proteine identificate come indicato in precedenza (Fig. 1.3).

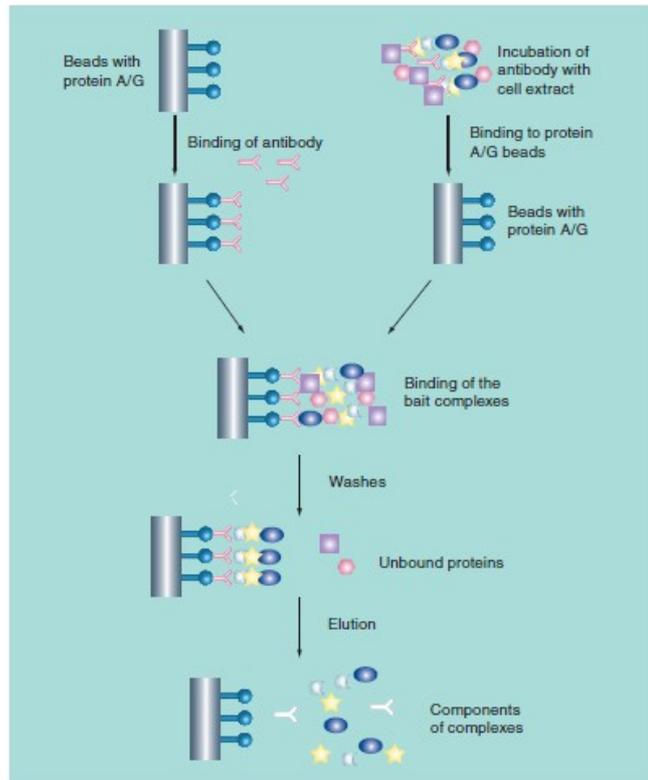


Fig. 1.3.: Strategia di immunoprecipitazione endogena.

Quando non sono disponibili anticorpi adatti per questa procedura, la proteina bersaglio è espressa con un piccolo epitopo peptidico contro il quale esistono buoni anticorpi (come FLAG, V5, c-myc, HA, etc) che può essere posizionato all’N- o al C-terminale. (In questo caso è previsto un effetto, se pur minimo, sulla struttura terziaria e l’attività biologica delle proteine). La proteina cui è stato aggiunto il tag, viene poi trasfettata in una appropriata linea cellulare e immunoprecipitata utilizzando anticorpi anti-tag commercialmente disponibili e covalentemente coniugati alle beads di agarosio [21]. In seguito alla rimozione delle proteine non specifiche, gli interattori possono essere eluiti per competizione con una forte concentrazione della forma libera del peptide tag. L’immunoprecipitazione endogena è sempre da preferirsi ogni volta che la proteina esca è espressa a livelli sufficienti da permettere di isolare una quantità relativamente soddisfacente di complessi. Tuttavia, il secondo approccio è certamente dotato di una serie di vantaggi consistenti in rese di immunoprecipitazione più elevate e in un basso livello di contaminazione del campione, diminuendo la possibile presenza di falsi positivi. Essendo l’anticorpo legato covalente con le beads, l’eluizione degli interattori specifici può essere eseguita in condizioni blande; inoltre, il campione finale non è contaminato dalle catene di anticorpi, aumentando così le possibilità di identificazione delle proteine. La sovraespressione della proteina nelle cellule ospite deve però assolutamente essere evitata in quanto l’alta concentrazione dell’esca altera il rapporto stechiometrico con i suoi partners naturali portando spesso alla formazione di interazioni non specifiche e/o non fisiologiche. La resa di purificazione è più ele-

vata per le affinity rispetto alla procedura di immunoprecipitazione che è strettamente dipendente dalla performance degli anticorpi. Inoltre, il legame degli anticorpi specifici con epitopi nella struttura della proteina nativa potrebbe competere con il legame di eventuali interattori con le stesse regioni, determinando in tal modo la dissociazione di alcuni partners proteici [45].

Quando si confrontano esperimenti di affinity con quelli di immunoprecipitazione, molto spesso si osserva una scarsa coincidenza nei risultati. Una scarsa sovrapposizione tra i complessi di proteine del lievito è stato ad esempio trovato in due studi che hanno utilizzato il metodo TAP-Tag e la procedura di immunoprecipitazione con l'epitopo FLAG seguita dall'identificazione in MS dei componenti isolati [46,47]. Anche se sorprendenti, questi risultati non sono senza precedenti. Utilizzando l'approccio del TAP-Tag, Gavin et al. [31] e Krogan et al. [32] sono stati in grado di purificare e identificare in totale più di 3000 proteine diverse (cioè il 47% del proteoma di *S. Cerevisiae*) [48]. Tuttavia, quando i dati sono stati confrontati, solo un migliaio di proteine sono state trovate essere comuni ad entrambe le serie di dati. Inoltre, solo sei complessi erano identici tra i due studi. Quasi 200 dei complessi riportati da Gavin e collaboratori [31] non hanno condiviso le proteine con i complessi trovati da Krogan et al. [32]. Oltre agli effetti degli strumenti di bioinformatica e di diversi possibili differenze nella purificazione delle proteine con il TAP-Tag, questi risultati sconcertanti potrebbero riflettere le dinamiche dei processi cellulari. Le proteine transientemente associate in complessi funzionali per svolgere un determinato processo poi rapidamente dissociano e ciascuna di esse è libera di formare altri complessi coinvolti in meccanismi diversi.

1.4 La Spettrometria di massa: tecnica di elezione per gli studi di Proteomica

La capacità di identificare le proteine senza ambiguità è un prerequisito essenziale per la loro indagine funzionale. Come le proteine derivate dallo stesso gene possono essere in gran parte identiche, e differire solo in piccoli ma rilevanti dettagli funzionali, gli strumenti per l'identificazione delle proteine devono non solo identificare un gran numero di proteine, ma anche essere in grado di differenziare tra quelle più simili. E' diventato chiaro, in seguito al completamento del progetto genoma umano, che la complessità dell'organismo è generata più da un proteoma complesso che da un genoma complesso. Queste informazioni possono essere generate grazie alla spettrometria di massa, un approccio che identifica le proteine mediante analisi parziale dei loro peptidi derivati da digestione. La Spettrometria di Massa (MS) è attualmente il metodo di scelta per il sequenziamento dei peptidi perché è sensibile e consente di routine l'identificazione di peptidi che sono presenti a livelli di femtomole. La MS è anche rapida, tant'è che il sequenziamento dei singoli peptidi può essere completato entro centinaia di millisecondi, e migliaia di peptidi possono quindi essere identificati in un'unica analisi; permette anche la caratterizzazione delle modifiche dei peptidi (comprese naturalmente le modifiche post-traduzionali, come ad es. la fosforilazione). La MS può anche essere adattata per misurare quantitativamente l'abbondanza dei peptidi e non richiede la pre-esistente conoscenza delle proteine da analizzare. Avanzamenti nel processamento dei campioni e nella strumentazione sono andati di pari passo con lo sviluppo di software che richiamano automaticamente le informazioni di sequenza dagli spettri di massa acquisiti e forniscono delle validazione statistiche relative alla precisione delle sequenze determinate.

Anche se i singoli gruppi di ricerca utilizzano differenti metodi per l'analisi dei complessi proteici mediante MS, i principi di base sono essenzialmente gli stessi.

1.4.1 Le metodologie in Spettrometria di massa

L'identificazione di una proteina precedentemente purificata, o di proteine da miscele complesse, separate solitamente mediante elettroforesi su gel SDS-PAGE mono e bidimensionali, può essere effettuata mediante la cosiddetta procedura del mass fingerprint. Il principio fondamentale su cui esse si basano sfrutta l'idea che il pattern proteolitico di ciascuna proteina sia univocamente dettato dalla specifica sequenza amminoacidica, e che pertanto risulti identificativo della proteina in esame [49].

Questo metodo prevede l'idrolisi chimica o enzimatica direttamente sulle bande proteiche escisse da gel e la successiva analisi di massa dei peptidi rilasciati, mediante MALDI/MS. I pesi molecolari ottenuti sono introdotti in opportuni programmi [50], sviluppati per l'identificazione di proteine, attraverso ricerche in banche dati telematiche. L'insieme dei dati raccolti, il mass fingerprint per l'ap-

punto, viene confrontato con una digestione “virtuale” effettuata tramite software di tutte le proteine contenute nelle banche dati genomiche. La determinazione accurata della massa di un numero sufficiente di peptidi identificherà una e una sola proteina fra quelle presenti nelle banche dati. In altre parole, solo una proteina potrà dare origine all’insieme dei peptidi aventi le masse osservate.

La risposta può essere ambigua e non univoca se una serie di parametri non sono soddisfatti: numero di peptidi non sufficiente, presenza di contaminanti peptidici dovuti alla manipolazione del campione, scarsa risoluzione elettroforetica. In tal caso sono utilizzate tecniche di spettrometria di massa tandem accoppiate alla cromatografia liquida capillare. Le miscele peptidiche prodotte dall’idrolisi chimica o enzimatica sono separate mediante HPLC capillare e le frazioni eluite dalla colonna sono direttamente introdotte nella sorgente ESI dello spettrometro. Gli ioni isolati da un primo analizzatore passano in una camera di collisione. Gli ioni frammento prodotti sono selezionati da un secondo analizzatore. I frammenti carichi prodotti danno informazioni in merito alla struttura primaria del peptide selezionato. La strumentazione scelta per questo tipo di indagine prevede l’utilizzo di un accoppiamento tra due analizzatori separati da una camera di collisione oppure di una trappola ionica. In tal caso le banche dati sono interrogate con dati relativi non solo alle masse peptiche, ma anche alla loro sequenza. E’ stato dimostrato che due soli peptidi sono sufficienti per identificare una proteina in modo non ambiguo [51].

Con questo metodo, i peptidi vengono frammentati mediante collisione con gas inerti (Collision-Induced Dissociation, CID); in appropriate condizioni i peptidi si frammentano in corrispondenza dei legami fra gli amminoacidi (legami peptidici); l’insieme dei frammenti risultanti dalla CID di un singolo peptide, poniamo ad esempio, di dieci amminoacidi, sarà costituito dallo stesso peptide che avrà subito la perdita di uno, due, tre amminoacidi e così via. La differenza di massa fra il peptide originario, il frammento di nove amminoacidi, quello di otto, di sette, ecc., corrisponde al peso molecolare di ciascun amminoacido e quindi fornisce la sequenza del peptide, sia pure parziale. L’informazione di sequenza in alcuni casi può confermare l’identificazione mediante mass fingerprint, le conferisce grande affidabilità e può essere utilizzata per ulteriori ricerche, in altri casi come detto può fornire informazioni che diversamente non si sarebbero potute ottenere. Analogamente, lo studio delle modificazioni posttraduzionali delle proteine viene effettuato misurando le variazioni di massa di peptidi dovute all’aggiunta di uno o più gruppi chimici quali, ad esempio, il gruppo fosfato nel caso della fosforilazione.

I profili proteomici di miscele di proteine sono ottenuti misurando il peso molecolare di un campione complesso di proteine. In questo caso non si ha la formazione di un solo “picco” di massa, cioè la misurazione di un singolo peso molecolare, ma di numerosissimi picchi, ciascuno corrispondente a una singola proteina. Sebbene in questo caso non sia possibile determinare direttamente l’identità

delle proteine coinvolte, il profilo complessivo o alcune sue parti possono essere dotate di una specificità tale da risultare diagnostici, ad esempio, di una specie batterica, dello stato funzionale di una cellula o di una patologia in atto.

1.4.2 La strumentazione

La spettrometria di massa si impone come tecnica elettiva per le identificazioni proteiche nel panorama scientifico proteomico. Ciò è dovuto allo sviluppo di strumentazioni innovative e all'accumulo di una notevole quantità di informazioni in merito a sequenze geniche e proteiche, raccolte in banche dati. La spettrometria di massa si avvale di numerose e differenti strumentazioni, caratterizzate da vantaggi e limiti. Gli spettrometri di massa si compongono di sorgenti, necessarie alla ionizzazione e volatilizzazione del campione in esame, analizzatori che separano gli ioni in base al rapporto massa/carica dell'analita, ed un rivelatore che determina il numero di ioni per ogni rapporto massa/carica trovato. Se in passato il suo impiego era limitato all'analisi di molecole di piccole e medie dimensioni, oggi è una tecnica fondamentale per lo studio delle macromolecole biologiche. Il limite, dovuto ai sistemi di analisi, risiedeva nella difficoltà di trasferire biomolecole polari cariche, di alto peso molecolare, allo stato gassoso. Nuove tecniche di ionizzazione hanno aperto la strada allo studio delle macromolecole biologiche mediante spettrometria di massa: electrospray ionization (ESI) e matrix-assisted laser ionization (MALDI) [52-54].

Ambedue le tecniche di ionizzazione sono accoppiate ai seguenti analizzatori: time of flight (TOF), quadrupoli (Q), ion trap (IT). Le configurazioni comunemente selezionate in base alle esigenze sperimentali sono: MALDI TOF, ESI IT, ESI Q, ESI Q-TOF.

Nella tecnologia MALDI il campione è analizzato in fase solida. La tecnica di ionizzazione, inducendo la formazione preferenziale di ioni monocarica, consente l'analisi diretta di miscele complesse. Il campione in esame co-cristallizza in presenza di una matrice, un composto aromatico acido, che media il trasferimento di energia da un raggio laser (UV o raramente IR) al campione stesso. Il sistema aromatico della matrice è eccitato dal laser e, tornando allo stato fondamentale, trasferisce energia al campione con essa cocristallizzato.

Detta energia è sufficiente a promuovere la formazione di uno ione monocarica MH^+ e ad indurre il desorbimento dal cristallo originario. A definire una tecnica di spettrometria di massa non è solo il metodo di ionizzazione scelto, ma anche il tipo di analizzatore, in cui gli ioni prodotti sono separati in base al rapporto massa/carica. Ad una sorgente MALDI è accoppiato un analizzatore a tempo di volo (TOF), fisicamente costituito da un tubo d'acciaio, di lunghezza compresa tra 1 e 2 metri. Il principio su cui si realizza la separazione sfrutta il fatto che gli ioni generati in sorgente, avendo

energia cinetica $1/2mv^2$, si muovono con velocità inversamente proporzionali alla radice quadrata della massa e pertanto raggiungeranno il rivelatore in tempi successivi, in ordine di massa crescente. Nella tecnologia ESI, il campione è analizzato in fase liquida. Un flusso di gas inerte in una camera di ionizzazione, ad alta temperatura e a pressione inferiore a quella atmosferica, favorisce l'evaporazione del solvente, che contiene l'analita carico. L'evaporazione del solvente porta ad una drastica diminuzione del volume: le forze repulsive tra molecole cariche prevalgono rispetto alle forze di tensione superficiale, determinando una repentina nebulizzazione. Un'opportuna differenza di potenziale indirizza gli ioni verso l'analizzatore.

Durante il processo di ionizzazione, le molecole proteiche e/o i peptidi vengono protonati da un numero variabile di protoni, in corrispondenza di residui basici o all'estremità N-terminale. Si forma una popolazione di residui a carica multipla $(M+nH)^{n+}$ che, separati da un analizzatore, danno luogo a distribuzioni di tipo gaussiano dei segnali m/z , ciascuna delle quali può essere trasformata su scala di massa reale da un algoritmo matematico.

Poiché l'analita è in fase liquida, la tecnica di ionizzazione descritta è accoppiata con una di frazionamento a monte: la cromatografia liquida. Questo tipo di strumentazione è generalmente configurato con un analizzatore quadrupolare o con una trappola ionica. Il primo consta di quattro barre metalliche con geometria cilindrica, alle quali è applicato un campo elettrico costante. Il movimento dello ione dipende da un campo elettrico oscillante a radio-frequenze. Solo lo ione con un definito rapporto m/z percorrerà, ad un dato valore del campo elettrico, una traiettoria stabile che gli consente l'uscita dal quadrupolo. Nella trappola ionica l'idea di base è opposta. E' un quadrupolo chiuso, al cui interno gli ioni si muovono con una traiettoria stabile, per effetto di un campo elettrico. Gli ioni con un dato rapporto m/z sono selezionati dal campo elettrico che, rendendo instabile la loro traiettoria all'interno dell'analizzatore li spinge ad uscire verso il detector.

1.5 Database pubblici per le interazioni proteina-proteina

Le proteine e le loro interazioni sono sempre state studiate una per una. Decadi di approcci "una per volta" hanno fornito informazioni funzionali per solo il 5-10% di tutti i geni predetti [55]. Più recentemente, la delucidazione del pieno repertorio di geni in diversi organismi e i rilevanti progressi nel campo della genomica funzionale e della proteomica hanno spianato la strada per le analisi delle interazioni proteina-proteina. Con lo sviluppo di queste metodologie su larga scala, è stata generata una lista di potenziali target.

Pertanto è nata la necessità di cercare in letteratura per scoprire se alcune delle proteine identificate come partners di una determinata proteina esca fossero già state identificate da altri ricercatori. Così sono stati sviluppati una serie di database che integrano i dati delle interazioni proteina-proteina

provenienti da diverse fonti, quali interazioni fisiche o genetiche, con dati provenienti dalla letteratura. Di solito i ricercatori iniziano questo tipo di approfondimento di ricerca con PubMed. Tuttavia esistono database pubblici molto utili, quali ad es. HRPD, IntAct e BIND.

La “Human Protein Reference Database” (HRPD) rappresenta una piattaforma centralizzata per visualizzare ed integrare le informazioni inerenti l'architettura di dominio, le modificazioni post-traduzionali, le reti di interazione e di associazione per ogni proteina del proteoma umano. Tutte le informazioni in HRPD sono estratte manualmente dalla letteratura da biologi esperti che leggono, interpretano e analizzano i dati pubblicati. IntAct è un database d'interazioni proteina-proteina disponibile liberamente in rete. Tutte le interazioni sono estrapolate dalla letteratura e da comunicazioni provenienti direttamente dagli utenti. Interazioni che derivano da studi in larga scala di solito non sono recuperate mediante ricerche con l'utilizzo di parole chiave in PubMed, mentre è facile trovarle in Intact. BIND invece è un database progettato per memorizzare le descrizioni complete di interazioni, complessi molecolari e pathways.

1.6 Verifica delle interazioni

Dopo aver identificato i ligandi della proteina esca isolati in seguito ad un esperimento di immunoprecipitazione o di affinity, un potenziale partner deve passare tutti gli esperimenti di controllo. Tuttavia, l'interazione partner-esca potrebbe non essere diretta, ma mediata da un partner di interazione comune oppure potrebbe essere semplicemente non fisiologica. E' quindi importante verificare il legame con i metodi supplementari. In alcuni casi può essere utilizzato il metodo iniziale. Ad esempio, se è stato usato un approccio con una proteina taggata, le interazioni individuate possono essere sistematicamente verificate taggando ciascuna delle proteine interagenti e ripetendo la procedura.

1.6.1 Co-IP

Uno dei metodi più comunemente utilizzati per la verifica delle interazioni proteina proteina è la co-immunoprecipitazione (Co-IP).

In un esperimento tipico, come descritto dettagliatamente in precedenza, i complessi dell'esca sono catturati da un lisato proteico utilizzando un anticorpo specifico. L'anticorpo è poi immobilizzato usando o la proteina A o la proteina G legate a beads di agarosio o sefarosio. Dopo aver lavato le beads, l'esca e le proteine associate all'esca sono eluiti. Le proteine legate possono quindi essere identificate mediante Spettrometria di massa.

Per confermare un interattore così identificato, si ripete l'esperimento e dopo l'eluizione si saggia la presenza del partner nell'eluato mediante l'utilizzo di un anticorpo specifico anti-partner. Se anche

questo saggio risulta essere positivo, si ripete l'esperimento per una terza volta, immunoprecipitando la proteina partner e poi rivelando l'esca per Immunoblotting.

Esperimenti di Co-IP di solito generano un significativo background ed è quindi importante condurre in parallelo controlli negativi. Questi esperimenti possono essere effettuati in diversi modi: (i) possono essere eseguite co-IP da linee cellulari o tessuti che esprimono le loro proteine endogene. Il vantaggio di questo approccio è che sono studiati complessi di proteine endogene. Pertanto sono evitati eventuali effetti artificiali di tag di affinità o sovraespressione. Lo svantaggio è che sono richiesti anticorpi molto specifici. (ii) È anche possibile utilizzare cellule transfettate con un plasmide codificante per una proteina esca taggata. Negli esperimenti di co-IP può essere utilizzato un anticorpo diretto contro il tag (invece che contro la proteina esca). Un vantaggio di questo approccio è che si può essere relativamente sicuri che l'anticorpo diretto contro il tag è specifico e non una reazione aspecifica con altre proteine. Inoltre, le proteine taggate possono essere spesso eluite per competizione mediante incubazione con peptidi, o altre piccole molecole. Tali eluizioni specifiche spesso riducono la quantità di proteine contaminanti nell'eluato. (iii) In alternativa, si possono eseguire esperimenti di co-IP utilizzando cellule transfettate con versioni taggate di due putativi partner di interazione [56].

1.6.2 La microscopia confocale

La colocalizzazione è definita come la presenza di due o più differenti molecole che risiedono nello stesso compartimento cellulare. Se le proteine interagiscono *in vivo*, esse dovrebbero colocalizzare o, almeno, si dovrebbe visualizzare una sovrapposizione di distribuzione all'interno della cellula. La localizzazione intracellulare di due (o più) proteine può essere studiata al microscopio confocale.

A tal proposito, le cellule sono dapprima transfettate con plasmidi di espressione che codificano la prima proteina fusa a un tag specifico e la seconda proteina interagente fusa a un altro tag. Le cellule sono poi fissate e incubate con anticorpi primari diretti contro il tag della prima e della seconda proteina, rispettivamente. Infine il campione è incubato con anticorpi secondari marcati con diversi fluorofori. Sfruttando il fatto che i due fluorofori mostrano diversi massimi di emissione, si può monitorare la localizzazione intracellulare delle proteine: se le due proteine colocalizzano, anche le sonde fluorescenti saranno colocalizzate. Questo genera un'immagine al microscopio confocale che contiene entrambi i contributi di colori. La sovraespressione e / o il tag potrebbero fare in modo che la proteina non localizzi correttamente. Pertanto, alternativamente all'espressione di proteine taggate transientemente, si può studiare la localizzazione delle proteine endogene. Questo, tuttavia, richiede la disponibilità di anticorpi primari altamente specifici diretti contro le proteine di interesse.

Un ulteriore problema con l'individuazione di proteine endogene è che i componenti del complesso proteico potrebbero non essere espressi sufficientemente nelle linee cellulari studiate [56].

1.7 Che cosa suggerisce oggi la Proteomica Funzionale?

L'assemblaggio di proteine per formare complessi funzionali è un processo dinamico che avviene al momento opportuno. Il complesso transiente dissocia subito dopo che il meccanismo biologico è stato completato e i singoli componenti possono partecipare alla formazione di altri complessi guidati da segnali specifici. Inoltre, i processi cellulari si verificano in luoghi ben definiti all'interno della cellula e quindi i complessi funzionali assemblano al momento giusto e nel posto giusto per svolgere correttamente una specifica funzione. Esperimenti di proteomica funzionale volti a descrivere il network di processi cellulari dovrebbero quindi prendere in considerazione con attenzione la distribuzione nel tempo e la localizzazione intracellulare di ogni specifico meccanismo che si vuole esaminare [57].

La Biologia moderna sta vivendo una situazione paradossale in cui molto spesso la sequenza della proteina, il gene corrispondente codificante e anche i meccanismi di regolazione concernente l'espressione genica sono noti, ma il ruolo biologico della proteina non è per niente chiaro. Quest'apparente paradosso suggerisce che vi è la necessità di cambiare il nostro approccio alla comprensione del funzionamento della cellula. Dobbiamo comprendere che l'attività di una proteina e il suo ruolo biologico potrebbero essere due aspetti molto diversi. L'attività di una proteina descrive la proteina stessa, sia essa un enzima, un trasportatore, un recettore, uno chaperone, uno scaffold, ecc., ma questa attività può essere esercitata in processi biologici completamente diversi a seconda della necessità della cellula. La funzione biologica di una proteina è quindi legata al processo specifico a cui la proteina sta in realtà partecipando. Di conseguenza, poiché la stessa proteina è in grado di assemblarsi con partner diversi per formare diversi complessi funzionali, ne consegue che una proteina può avere una sola attività, ma diverse funzioni biologiche.

È relativamente facile prevedere che la proteomica funzionale avrà un ruolo fondamentale negli studi futuri della system biology. Una descrizione accurata di componenti proteici all'interno di complessi funzionali è un pre-requisito fondamentale per comprendere il funzionamento della cellula e per svelare la complessa rete di pathways di signalling a livello molecolare. L'interpretazione dei dati di proteomica costituisce attualmente un grave problema, poiché non è sempre possibile assegnare esattamente una proteina identificata ad uno specifico complesso riducendo così notevolmente l'impatto di questi studi. Esperimenti specifici possono essere progettati per dare un significato funzionale ai risultati della proteomica. Vi è molto spazio per la creatività nel collegare strategie

biologiche, cellulari e molecolari, con il potente approccio della proteomica per risolvere le questioni che non potevano essere affrontate in precedenza.

Un ulteriore obiettivo da perseguire in futuro, sarà il tentativo di effettuare esperimenti di proteomica funzionale in tempo reale nei sistemi in vivo mediante la produzione di modelli animali recanti una versione taggata dell'esca proteica. Animali omozigoti, quando vitali, forniranno i tessuti e / o le cellule progenitrici per l'isolamento di complessi funzionali che si formeranno in vivo, in modo da rivelare l'identità dei loro singoli componenti ed affrontare la questione del loro funzionamento reale *in vivo*.

1.8 Le interazioni proteina-proteina nelle malattie umane

“La vita è una relazione tra molecole, non la proprietà di ogni molecola. Così è per la melattia, che mette in pericolo la vita”, scrissero Zuckerkandl and Pauling (1962) nel loro capitolo su ‘*Molecular disease, evolution and genic heterogeneity*’. Sono trascorsi circa 50 anni e ancora siamo lontani dal capire i meccanismi molecolari della maggior parte delle malattie, nonché le interazioni molecolari che si stabiliscono in organismi sani e malati [58]. Infatti, le proteine non funzionano in modo isolato, ma piuttosto nella cellula, interagiscono prevalentemente con altre proteine, ma anche con altre molecole come il DNA, l'RNA e altre piccole molecole. Così, gli studi delle proteine e delle loro interazioni sono essenziali per capire il loro ruolo all'interno della cellula. In genere, il termine ‘*protein interaction*’ comprende una vasta gamma di eventi, come complessi transitori e stabili, oltre che interazioni fisiche e funzionali. Le interazioni proteiche sono coinvolte nel *signaling*, nei network metabolici etc. Una migliore comprensione delle interazioni proteiche, sia con altre proteine o con DNA, RNA, molecole piccole o di membrana, potrebbe rivelare il meccanismo molecolare dei processi che portano alle malattie, processi estremamente complessi, così come lo sono le proteine e le interazioni coinvolte in esse.

Mutazioni all'interfaccia di interazione delle proteine (o siti correlati, ad esempio siti attivi, i siti di legame allosterico) evidentemente potrebbero disturbare l'interazione della proteina. In realtà, le proteine sono continuamente sintetizzate e degradate in amminoacidi. La cinetica dei processi e le dinamiche dei network devono essere considerate per comprendere come la perturbazione delle proteine e delle loro interazioni conducano alla malattia. Infine, è importante considerare anche il contesto specifico (tessuto, stadio di malattia e di risposta) in cui avviene l'interazione proteica. E' chiaro poi, che esiste ancora un gap tra l'identificazione del network di proteine associate alla malattia e la comprensione completa del meccanismo della malattia. Questo gap, sfortunatamente, è riempito con più domande che risposte. Siamo ancora lontani dal comprendere l'eziologia della maggior parte delle malattie, ma ulteriori rilevanti progressi in ambito tecnologico e sperimentale,

come quelli riguardanti le interazioni proteiche, la struttura delle proteine, l'espressione genica, insieme agli strumenti di calcolo per organizzare, visualizzare e integrare questi dati forniranno un passo avanti in quella direzione.

Idealmente, dal momento che gli approcci strutturali e del network sono complementari, la combinazione di questi due ha il potenziale di essere un eccellente quadro per lo studio dei meccanismi delle malattie.

1.9 Scopo del progetto di tesi

Le interazioni proteina-proteina sono basilari nella regolazione della trasduzione dei segnali e l'identificazione in vivo di tali proteine è utile per la descrizione a livello molecolare dei meccanismi cellulari in cui sono coinvolte.

Il lavoro svolto durante il progetto di tesi è stato, dunque, incentrato sullo sviluppo di metodologie innovative in proteomica funzionale ed articolato come segue:

- ✓ Studio del “traffico” molecolare di SUMF1.

SUMF1 è una proteina del reticolo endoplasmatico (RE) capace di attivare tutte le solfatasi mediante la modifica post- traduzionale di una cisteina presente nel loro dominio catalitico in formilglucina. Questa unica modifica è essenziale per l'attività delle solfatasi. Infatti, mutazioni nel gene di SUMF1 rendono inattive tutte le solfatasi causando una grave malattia, la Multiple Sulfatase Deficiency (MSD), che colpisce diversi organi e tessuti, soprattutto quello nervoso. Studi precedenti hanno dimostrato che SUMF1 può essere secreta nel mezzo di coltura cellulare e poi reinternalizzata da altre cellule che la localizzano nuovamente nel RE, dove attiva le solfatasi endogene.

Obiettivo del progetto di ricerca è stato l'identificazione delle proteine che interagiscono con SUMF1 al fine di delucidare i pathways molecolari attraverso cui essa è secreta dalle cellule.

- ✓ Studio dell'interattoma del fattore trascrizionale CBX7.

CBX7 (Chromobox protein homologue 7) è una proteina nucleare appartenente al complesso PRC1 (Polycomb repressive complex 1), una famiglia di proteine impegnata nella repressione genica mediante l'inibizione dell'inizio della trascrizione, legando e comprimendo l'eterocromatina. E' noto ormai che CBX7 ha un importante ruolo nella regolazione dello sviluppo e della crescita cellulare, ma è letteratura recente l'ipotesi secondo cui questa proteina ha una funzione di oncosoppressore.

Nel secondo anno di Dottorato l'attività scientifica è stata indirizzata all'impiego delle metodologie di proteomica funzionale per l'identificazione dei partners molecolari di CBX7, al fine di delucidare il meccanismo attraverso cui la ridotta concentrazione del fattore CBX7 è coinvolta nella carcinogenesi.

- ✓ Messa a punto di strategie di pre-frazionamento di complessi proteici.

L'esperimento di proteomica mediante il quale si è cercato di delucidare il meccanismo attraverso cui il fattore Cbx7 controlla la proliferazione cellulare, ha fornito una lunga lista di putativi interattori del fattore trascrizionale, indicando chiaramente che Cbx7 è coinvolta in numerosi complessi probabilmente dotati di diversa funzione biologica

Il compito più impegnativo in questi esperimenti è l'interpretazione dei dati e in particolare il tentativo di assegnare i singoli componenti proteici allo specifico complesso cui essi appartengono. No-

nonostante la ricchezza delle informazioni fornite da banche dati di proteine e database di interazioni proteina-proteina e la disponibilità di strumenti software, la deduzione di un significato biologico dalla lista degli interattori rimane ancora un problema di difficile soluzione. Queste difficoltà potrebbero essere superate dall'impiego di tecniche di pre-frazionamento dell'estratto proteico di partenza in grado di preservare le interazioni native tra le proteine. In tal modo si tenderebbe ad isolare i singoli complessi funzionali ridimensionando la complessità dei dati di proteomica funzionale. A tal proposito durante il terzo anno di Dottorato, si è cercato di razionalizzare la lunga lista d'interattori ottenuta durante il secondo anno di dottorato, mettendo a punto tali strategie di pre-frazionamento.

Bibliografia Introduzione

1. Carmen L. de Hoog and Matthias Mann. PROTEOMICS. *Annu. Rev. Genomics Hum. Genet.* 2004. 5:267–93.
2. Hochstrasser DF, Sanchez JC, Appel RD. Proteomics and its trends facing nature's complexity. *Proteomics* 2002, 2, 807–812.
3. Anne-Claude Gingras, Matthias Gstaiger, Brian Raught and Ruedi Aebersold. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol.* 2007 Aug;8(8):645-54.
4. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003 422:198–207.
5. Dahlberg AE. Ribosome structure. The ribosome in action. *Science* 2001 292:868–69.
6. Kobayashi Y, Weigel D. Move on up, it's time for change-mobile signals controlling photoperiod-dependent flowering. *Genes Dev.* 2007, 21:2371-2384. *Rev proteomica.*
7. Qiu Z, Ghosh A. A calcium-dependent switch in a CREST-BRG1 complex regulates activity-dependent gene expression. *Neuron.* 2008, 60:775-787.
8. Michaud GA, Snyder M. Proteomic approaches for the global analysis of proteins. *Biotechniques.* 2002, 33:1308–16.
9. Souchelnytskyi S. Proteomics in studies of signal transduction in epithelial cells. *J Mammary Gland Biol.* 2002, 7:359 – 371.
10. de la Fuente van Bentem S, Mentzen WI, de la Fuente A, Hirt H. Towards functional phosphoproteomics by mapping differential phosphorylation events in signalling networks. *Proteomics.* 2008, 8:4453-4465.
11. Lecaudey V, Cakan-Akdogan G, Norton WH, Gilmour D. Dynamic Fgf signaling couples morphogenesis and migration in the zebrafish lateral line primordium. *Development.* 2008, 135:2695-705
12. Pfeifer AC, Timmer J, Klingmüller U. Systems biology of JAK/STAT signalling. *Essays Biochem.* 2008;45:109-20.
13. Zak DE, Aderem A. Systems biology of innate immunity. *Immunol Rev.* 2009, 227:264-282.
14. Schlieker CD, Van der Veen AG, Damon JR, Spooner E, Ploegh HL. A functional proteomics approach links the ubiquitin-related modifier Urm1 to a tRNA modification pathway. *Proc Natl Acad Sci U S A.* 2008,105:18255-18260.
15. Souza TA, Chen X, Guo Y, Sava P, Zhang J, Hill JJ, Yaworsky PJ, Qiu Y. Mol. Proteomic identification and functional validation of activins and bone morphogenetic protein 11 as candidate novel muscle mass regulators. *Endocrinol.* 2008, 22:2689-2702.

16. Brown KA, Ham AJ, Clark CN, Meller N, Law BK, Chytil A, Cheng N, Pietenpol JA, Moses HL. Identification of novel Smad2 and Smad3 associated proteins in response to TGF-beta1. *J Cell Biochem.* 2008,105:596-611.
17. Wright CW, Duckett CS. The aryl hydrocarbon nuclear translocator alters CD30-mediated NF-kappaB-dependent transcription. *Science.* 2009, 323:251-255.
18. Markillie LM, Lin CT, et al. Simple protein complex purification and identification method for high-throughput mapping of protein interaction networks. *J Proteome Res.* 2005,4:268-274.
19. Wu F, Wang P, Young LC, Lai R, Li L. Proteome-Wide Identification of Novel Binding Partners to the Oncogenic Fusion Gene Protein, NPM-ALK, using Tandem Affinity Purification and Mass Spectrometry. *Am J Pathol.* 2009 Jan 8. [Epub ahead of print]
20. Zada AA, Pulikkan JA, Bararia D, Geletu M, Trivedi AK, Balkhi MY, Hiddemann WD, Tenen DG, Behre HM, Behre G. Proteomic discovery of Max as a novel interacting partner of C/EBPalpha: a Myc/Max/Mad link. *Leukemia.* 2006, 20:2137-2146.
21. K. Terpe. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol.* 2003, 60:523-533
22. Meier N, Krpic S, Rodriguez P, Strouboulis J, Monti M, Krijgsveld J, Gering M, Patient R, Hostert A, Grosveld F. Novel binding partners of Ldb1 are required for haematopoietic development. *Development.* 2006, 133:4913-4923.
23. Henkin JA, Jennings ME, Matthews DE, Vigoreaux JO. Mass processing--an improved technique for protein identification with mass spectrometry data. *J Biomol Tech.* 2004, 15: 230-237.
24. Dianzani I. Analysis of the ribosomal protein S19 interactome. *Mol Cell Proteomics.* 2007, 6:382-393.
25. Scholten A, van Veen TA, Vos MA, Heck AJ. Diversity of cAMP-dependent protein kinase isoforms and their anchoring proteins in mouse ventricular tissue. *J Proteome Res.* 2007, 6: 1705-1717.
26. Fraldi A, Zito E, Annunziata F et al.. Multistep, sequential control of the trafficking and function of the multiple sulfatase deficiency gene product, SUMF1 by PDI, ERGIC-53 and ERp44. *Hum Mol Genet.* 2008, 17: 2610-2621.
27. Pisa v, Cozzolino M, Gargiulo S et al. The molecular chaperone Hsp90 is a component of the cup-binding complex and interacts with the translational repressor Cup during *Drosophila* oogenesis. *Gene* 432, 67-74 (2009).

28. Bowen C, Stuart A, Ju JH, Tuan J, Blonder J, Conrads TP, Veenstra TD, Gelmann EP. NKX3.1 homeodomain protein binds to topoisomerase I and enhances its activity. *Cancer Res.* 2007, 67: 455-464.
29. Köcher T. and Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nature Methods.* 2007, 4: 807-815.
30. Lee PY, Bae KH, Kho CW, Kang S, Lee do H, Cho S, Kang S, Lee SC, Park BC, Park SG. Interactome analysis of yeast glutathione peroxidase 3. *J Microbiol Biotechnol.* 2008, 18: 1364-1367.
31. Gavin AC, Aloy P, Grandi P, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature.* 2006, 440: 631-636.
32. Krogan NJ, Cagney G, Yu H, et al Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature,* 2006. 440: 637-643.
33. Banks E, Nabieva E, Chazelle B, Singh M. Organization of physical interactomes as uncovered by network schemas. *PLoS Comput Biol.* 2008, 4:e1000203.
34. Zhang B, Park BH, Karpinets T, Samatova NF. From pull-down data to protein interaction networks and complexes with biological relevance. *Bioinformatics.* 2008, 24:979-986.
35. Klein T, P Geurink P, S Overkleeft H, K Kauffman H, Bischoff R. Functional Proteomics on Zinc-Dependent Metalloproteinases using Inhibitor Probes. *Chem. Med Chem.* 2008 Dec 12.
36. Rix U, Hantschel O, Dürnberger G et al. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood.* 2007, 110: 4055-4063.
37. Remsing Rix LL, Rix U, Colinge J, Hantschel O, Bennett KL, Stranzl T, Müller A, Baumgartner C, Valent P, Augustin M, Till JH, Superti-Furga G. Global target profile of the kinase inhibitor bosutinib in primary chronic myeloid leukemia cells. *Leukemia.* 2008 Nov 27.
38. Aye TT, Mohammed S, Toorn HW, Veen TA, Heyden MA, Scholten A, Heck AJ. Selectivity in enrichment of PKA RI and RII isoforms and their interactors using modified cAMP affinity resins. *Mol Cell Proteomics.* 2008 Dec 31
39. Russo A, Cirulli C, Amoresano A, Pucci P, Pietropaolo C, Russo G. cis-acting sequences and trans-acting factors in the localization of mRNA for mitochondrial ribosomal proteins. *Biochim Biophys Acta.* 2008, 1779: 820-829.

40. Lin JY, Li ML, Huang PN, Chien KY, Horng JT, Shih SR. Heterogeneous nuclear ribonuclear protein K interacts with the enterovirus 71 5' untranslated region and participates in virus replication. *J Gen Virol.* 2008, 89: 2540-2549.
41. Medugno L, Costanzo P, Lupo A, Monti M, Florio F, Pucci P, Izzo P. A novel zinc finger transcriptional repressor, ZNF224, interacts with the negative regulatory element (AldA-NRE) and inhibits gene expression. *FEBS Lett.* 2003, 534: 93-100.
42. Hall-Pogar T, Liang S, Hague LK, Lutz CS. Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *RNA.* 2007, 13: 1103-15.
43. Lambert JP, Mitchell L, Rudner A, Baetz K and Figeys D., 2008 .
44. Monti M, Orrù S, Pagnozzi D, Pucci P. Functional proteomics. *Clinica Chimica Acta* 2005, 357: 140–150.
45. Monti M, Orrù S, Pagnozzi D, Pucci P. Interaction proteomics. *Biosci Rep.* 2005, 25: 45-56.
46. Gavin AC, Bösch M, Krause R, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 2002, 415: 141–147.
47. Ho Y, Gruhler A, Heilbut A, et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature.* 2002, 415: 180–183.
48. Goll J, Uetz P. The elusive yeast interactome. *Genome Biol.* 2006, 7: 223.
49. Cottrell JS. Protein identification by peptide mass fingerprinting. *Pept Res.* 1994, 7(3):115-24.
50. Yates JR. Database searching using mass spectrometry data. *Electrophoresis* 1998, 19(6):893-900.
51. Godovac-Zimmermann J, Brown LR. Perspective for mass spectrometry and functional proteomics. *Mass Spectrom Rev.* 2001, 20: 1-57.
52. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem* 2001, 70: 437-73.
53. Aebersold R. A mass spectrometric journey into protein and proteome research. *J Am Soc Mass Spectrom* 2003, 14(7):685-95.
54. Lin D, Tabb DL, Yates JR. Large-scale protein identification using mass spectrometry. *Biochim Biophys Acta* 2003, 1646(1-2):1-10.
55. Vidal M. Interactome modeling. *FEBS Lett.* 2005 579:1834–1838
56. Berggård T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. *Proteomics.* 2007 Aug;7(16):2833-42.

57. Monti M, Cozzolino M, Cozzolino F, Vitiello G, Tedesco R, Flagiello A, Pucci P. Puzzle of protein complexes in vivo: a present and future challenge for functional proteomics. *Expert Rev Proteomics*. 2009 Apr;6(2):159-69.
58. Kann MG. Protein interactions and disease: computational approaches to uncover the etiology of diseases. *Brief Bioinform*. 2007 Sep;8(5):333-46. Epub 2007 Jul 16.

Capitolo II

Studio del “traffico” molecolare di SUMF1

2.1 Introduzione

2.1.1 Le Solfatasi

Le solfatasi sono una famiglia di enzimi che mostrano similarità sia strutturali che funzionali. Catalizzano l'idrolisi dei legami estere solfato da un'ampia varietà di substrati come i glicosamminoglicani, i solfolipidi e gli steroidi solfato. Questi enzimi possono essere divisi, almeno nei mammiferi, in due categorie principali basate sulla loro localizzazione subcellulare: quelli agenti a pH acido, localizzati nei lisosomi, e quelli agenti a pH neutri che si trovano nel reticolo endoplasmatico (RE), nell'apparato del Golgi, e sulla superficie cellulare. (Tabella 1) [1, 2, 3]. Le prime solfatasi di mammiferi sono state identificate parecchie decadi fa mediante lo studio d'individui affetti da malattie da accumulo lisosomiale come le mucopolissaccaridosi (MPS) e la leucodistrofia metacromatica (MLD). Queste malattie ereditarie sono errori innati del metabolismo in cui una mutazione di un gene codificante un enzima coinvolto nella degradazione di un dato componente conduce al blocco in un *pathway* catabolico con conseguente accumulo di questo componente nelle cellule e nei fluidi extracellulari. Le analisi dei tessuti, cellule e corpi fluidi derivanti da questi pazienti hanno portato a identificare materiale di accumulo, come glicosamminoglicani (GAG) o solfolipidi, che si pensò derivare dal deficit di una specifica solfatasi [4, 5].

Come menzionato sopra, le prime solfatasi descritte sono state identificate mediante lo studio di una malattia, ma è solo dopo la conoscenza della funzione e della sequenza di queste solfatasi che sono stati identificati i geni corrispondenti. Gli approcci utilizzati per identificare le solfatasi sono cambiati dopo il completamento della sequenza del genoma umano. Sei nuove solfatasi furono identificate studiando le sequenze genomiche disponibili pubblicamente (Tabella 1).

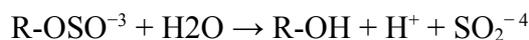
Sebbene questi sforzi abbiano notevolmente ampliato le possibilità di avere una visione completa della funzione di questa famiglia di enzimi nei mammiferi, ci sono ancora molte domande senza risposta sulla funzione e sul possibile coinvolgimento di queste nuove solfatasi in malattie umane.

TABELLA 1. Principali caratteristiche delle Solfatasi umane

Gene	Chromosomal Localization	Enzyme	Subcellular localization	Human Disorder
ARSA	22q13	Arylsulfatase A	Lysosomal	Metachromatic leukodystrophy (MLD)
ARSB	5q13	Arylsulfatase B	Lysosomal	Maroteaux-Lamy syndrome (MPSVI)
IDS	Xq27-28	Iduronate Sulfatase	Lysosomal	Hunter syndrome (MPSII)
SGSH	17q25.3	Sulfamidase	Lysosomal	Sanfilippo A syndrome (MPSIIIA)
G6S	12q14	Glucosamine-6-Sulfatase	Lysosomal	Sanfilippo D syndrome (MPSIIID)
GAL6S	16q24	Galactose-6-Sulfatase	Lysosomal	Morquio A syndrome (MPSIVA)
ARSC/STS	Xp22.3	Arylsulfatase C/ Steroid sulfatase	ER	X-linked Ichthyosis (XLI)
ARSD	Xp22.3	Arylsulfatase D	Golgi	NI
ARSE	Xp22.3	Arylsulfatase E	ER	Chondrodysplasia Punctata 1 (CDPX1)
ARSF	Xp22.3	Arylsulfatase F	ND	NI
ARSH	Xp22.3	Arylsulfatase H	Cell Surface	NI
HSULF1	8q13.2-13.3	Hsulf1	Cell Surface	NI
HSULF2	20q13.12	Hsulf2	ER	NI
ARSG	17q23-24	Arylsulfatase G	ER	NI
ARSJ	4q26	Arylsulfatase J	ER	NI
ARSI	5q32	Arylsulfatase K	ER	NI
TSULF	5q15	T. Sulfatase	ND	NI

2.1.2 Le Solfatasi umane

Le solfatasi catalizzano l'idrolisi degli esteri solfato come gli O-solfati e gli N-solfati.



Nei mammiferi le idrolisi degli esteri solfato richiedono l'azione di una famiglia di solfatasi, ognuna avente una differente specificità di substrato, coinvolta nella desolfatazione di vari substrati naturali che vanno da molecole complesse come GAG o componenti più piccoli come i sulfolipidi e gli idrosteroidi. Le solfatasi mostrano le seguenti caratteristiche: *a)* hanno dimensioni simili, che vanno dai 500 agli 800 aminoacidi; *b)* sono glicosilate estensivamente; *c)* mostrano un'omologia di sequenza significativa sull'intera lunghezza; *d)* hanno un sito attivo altamente simile, che subisce un solo tipo di modifica posttraduzionale. Analisi sulla sequenza del genoma umano ha rivelato la presenza di 17 diversi geni che codificano per proteine con le caratteristiche sopra menzionate e che potrebbero essere considerate *bona fide* membri della famiglia delle solfatasi (Tabella 1). La Fig.1 mostra l'allineamento delle sequenze delle 17 solfatasi umane in tre regioni, che rappresentano i domini proteici altamente conservati (domini a, b e c).

Più recentemente, gli scienziati hanno identificato due nuovi geni codificanti solfatasi, SULF1 e SULF2, che sono molto simili tra loro, hanno una dimensione leggermente più grande rispetto alle altre solfatasi e, soprattutto, si trovano sulla superficie cellulare. Studi funzionali su organismi modello hanno suggerito che queste solfatasi sono coinvolte nel *signaling* cellulare mediante la modulazione degli stati di solfatazione dell'eparansolfato sulla superficie cellulare dei proteoglicani. L'espressione dei geni codificanti queste solfatasi è regolata dallo sviluppo. Questi dati rivelano una nuova categoria funzionale entro la famiglia delle solfatasi che è coinvolta nel *signaling* cellulare e nello sviluppo embrionale [6, 7, 8].

a		**	b		* *
ARSC	PNIILVMADDLGIGDPG	43	ARSC	ASPLCTPSRAAFMTGRYPV	89
ARSD	PNILLIMADDLGTGDLG	57	ARSD	AAPLCTPSRAAFLTGRHSF	103
ARSH	PNIVLLMADDLGVGDLC	23	ARSH	AASMCTPSRAAFLTGRYPV	69
ARSE	PNILLMADDLGIGDIG	54	ARSE	AASLCTPSRAAFLTGRYPV	100
ARSF	PNIVLIMVDDLGIGDLG	47	ARSF	AASLCSPSRSAFLTGRYPV	93
ARSG	PNFVIIILADDMGWGDLG	52	ARSG	AASTCSPSRASLLTGRGLG	98
ARSA	PNIVLIFADDLGYGDLG	37	ARSA	PVSLCTPSRAALLTGRLPV	83
GALNS	PNILLLLMDDMGWDLG	47	GALNS	ANPLCSPSRASLLTGRLPV	93
ARSB	PHLVFLLADDLGWNDVG	61	ARSB	TQPLCTPSRSQLLTGRYQI	105
ARSJ	PHLIFILADDQGFQFDVG	92	ARSJ	VQPICTPSRSQFITGKYQI	136
ARSI	PHIIFILTDDQGYHDVG	63	ARSI	IQPICTPSRSQLLTGRYQI	107
SGSH	RNALLLLADDGGFESGA	39	SGSH	SVSSCSPSRASLLTGLPQH	84
GNS	PNVLLLLTDDQDEVLGG	63	GNS	PSALCCPSRASILTGYKYPH	105
SULF1	PNIILVPTDDQDVELGS	59	SULF1	TTPMCCPSRSSMLTGYKVH	101
SULF2	PNIILVLTDDQDVELGS	60	SULF2	TTPMCCPSRSSILTGYKVH	102
TSULF	PNVVLVVSDFDGRITF	48	TSULF	NSPICCPSRAAMWSGLFTH	94
IDS	LNVLIIIVDDLRLPSLGC	53	IDS	QQAVCAPSRVSVFLTGRRPD	98

c		* *
ARSC	KDQGYSTALIGKWHLGMS	140
ARSD	QQHGYATGLIGKWHQGVN	154
ARSH	QHRGYRTGLIGKWHLGLS	121
ARSE	KEKGYATGLIGKWHLGLN	151
ARSF	KKQGYSTGLIGKWHQGLN	144
ARSG	QQAGYVTGIIGKWHLGHH	143
ARSA	AARGYLTGMAGKWHLGVG	129
GALNS	KKAGYVSKIVGKWHLGHR	146
ARSB	KEAGYTTHMVGKWHLGMV	151
ARSJ	KEVGYSTHMVGKWHLGFY	182
ARSI	QEAGYSTHMVGKWHLGFY	153
SGSH	SQAGVRTGIIGKKHVGPE	129
GNS	SMCGYQTFPAGKYLNEYG	155
SULF1	NNTGYRTAFFGKYLNEYN	148
SULF2	NSTGYRTAFFGKYLNEYN	149
TSULF	ERHGYRTQKFGKLDYTSG	134
IDS	KENGYVTMSVGVFHPGI	141

Fig.2.1 Allineamento di sequenza dei domini conservati tra le solfatasi umane

Allineamento dei tre domini conservati in tutte le solfatasi umane. (a) e (c) rappresentano il dominio N-terminale delle proteine. (b) rappresenta il sito attivo. Gli aminoacidi ombreggiati sono conservati nel 90% di tutte le solfatasi. Gli asterischi evidenziano gli aminoacidi con una funzione catalitica nota.

2.1.3 Modifiche Post-traduzionali delle Solfatasi

Ancora grazie allo studio di una malattia è stato possibile definire un nuovo meccanismo biologico. In un raro disordine autosomico recessivo, noto come MSD, l'attività di tutte le solfatasi è drasticamente compromessa [2]. Studi portati avanti per far luce su questa questione hanno dimostrato che l'espressione dei cDNA delle solfatasi nei fibroblasti MSD produce solfatasi con attività enzimatica drasticamente ridotta [9]. Queste osservazioni hanno condotto all'ipotesi che i difetti del disordine MSD coinvolgevano modifiche delle solfatasi co- o posttraduzionali che avvengono prima che le solfatasi siano destinate ai differenti compartimenti cellulari.

Studi successivi hanno dimostrato che le solfatasi subiscono una modifica posttraduzionale di un residuo di cisteina localizzato nel loro sito attivo in acido 2-amino-3-ossopropanoico [α -formylglycina (FGly)] [10]. Le modifiche coinvolte convertono un tiolo in un gruppo aldeidico. La cisteina modificata è stata altamente conservata nel corso dell'evoluzione, ed è presente in tutte le solfatasi, con l'eccezione di poche solfatasi batteriche che hanno una serina, invece di una cisteina, nella stessa posizione. Questa serina subisce anch'essa la modifica in FGly, sebbene mediante un meccanismo differente [11].

La modifica posttraduzionale avviene nel RE su solfatasi unfoldate subito dopo la traduzione [12]. La sequenza contenente la cisteina che deve essere modificata a FGly è di 11 residui aminoacidici. Mutazioni e troncamenti di questa sequenza target hanno rivelato che un core con un motivo C(X/T)P(X/S)R, è la minima sequenza richiesta per ottenere modeste quantità di modificazioni [13, 14, 15].

E' stato dimostrato che le solfatasi che derivavano da cellule MSD recano una cisteina nel loro sito attivo che non è stata modificata in FGly [10]. Questo suggerì che il disordine MSD è causato dalla mutazione di un gene, o geni, implicato nel macchinario di conversione della cisteina a FGly. La natura di questo macchinario e le basi molecolari della MSD sono rimaste sconosciute per molti anni, nonostante l'intenso interesse medico e biologico. La ragione di questo è dovuta all'assenza di casi familiari di MSD che permettessero di mettere a punto studi di *mapping* genetico e la novità ed unicità del meccanismo di modificazione.

2.1.4 SUMF1

Il gene della MSD, chiamato *sulfatase modifying factor 1* (*sumf1*) e codificante la *formylglycine-generating enzyme* (FGE) è stato identificato solo molti anni dopo la scoperta della modifica post-traduzionale, mediante due approcci completamente indipendenti e complementari, uno biochimico e l'altro genetico [12, 16].

Il gene *sumf1* codifica una proteina di 374 aminoacidi localizzata nel lume del RE, sebbene essa non possieda alcun segnale di ritenzione per lo stesso RE.

La proteina ha un singolo potenziale segnale di N-glicosilazione ad Asn 141 e la sua forma intracellulare contiene oligosaccaridi di tipo *high mannose* [17]. Il gene è molto conservato tra i pro e gli eucarioti ed ha un omologo nei vertebrati chiamato *sumf2* [12,16]. SUMF2 colocalizza con SUMF1 nel RE e inibisce gli effetti di miglioramento di SUMF1 sulle solfatasi [19].

Analisi del gene *sumf1* in pazienti affetti da MSD rivelarono un largo spettro di mutazioni: dalle mutazioni nonsense, missenso fino a piccole delezioni, che portano alla perdita della funzione della proteina codificata [12, 16].

La struttura cristallografica di SUMF1 è stata risolta e si è notato che possiede un singolo dominio monometrico che adotta un unico fold, con una piccola struttura secondaria. E' stato possibile definire che l'enzima ha una funzione ossigenasica [18].

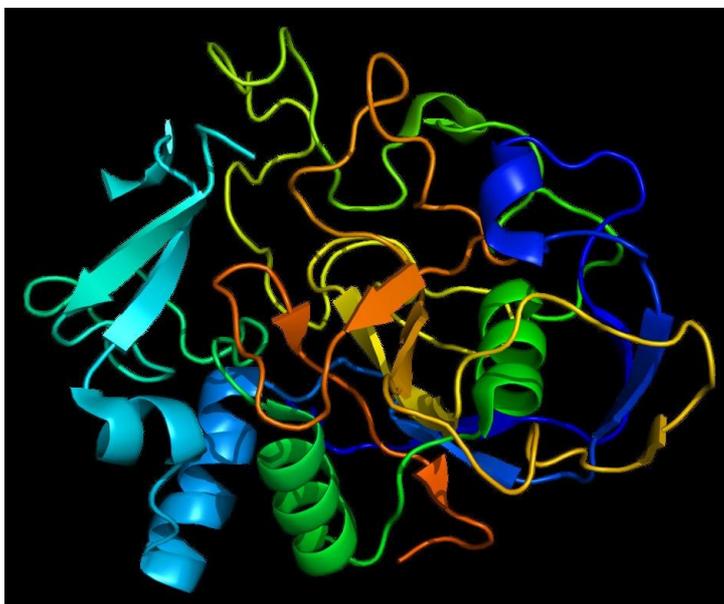


Fig.2.2 Struttura della proteina SUMF1.

2.2. Materiali e Metodi

2.2.1. Materiali

Il detergente usato per l'estrazione proteica è Nonidet P-40 della FLUKA.

I prodotti della SIGMA utilizzati sono:

la resina coniugata all'anticorpo ANTI-FLAG, ANTI-FLAG M2 Affinity Gel, così come l'Anti-Mouse Agarose Beads usate per l'immunoprecipitazione;

il ditioneitrilo (DTT);

la tripsina ;

l'acido α -ciano-4-idrossi-cinnamico;

l'anticorpo primario ANTI-FLAG M2;

Il peptide 3xFlag;

Le colonne di reazione usate per l'immunoprecipitazione sono state acquistate presso la USB.

Tutti i materiali utilizzati per le tecniche elettroforetiche, 30% Acrilammide/ Bis, N,N,N',N'-tetrametilendiammina (TEMED), sodio-dodecilsolfato (SDS), ammoniopersolfato (APS), la miscela di proteine a peso molecolare noto "total blue" usata come standard nella migrazione elettroforetica sono della BIO-RAD.

La iodoacetamide (IAM) e l'ammonio bicarbonato sono della FLUKA.

La colonna capillare del CapLC a fase inversa C18 (ID 75 μ m, L 20 cm, 300 Å) è della nanoseparations (The Netherlands).

L'anticorpo secondario Anti-Mouse è della AMERSHAM.

Gli inibitori di proteasi sono della ROCHE.

2.2.2 Metodi

2.2.2.1 Colture Cellulari

Le cellule utilizzate negli esperimenti eseguiti durante questo lavoro di tesi sono state fornite dal gruppo della dott.ssa Maria Pia Cosma, con la quale abbiamo collaborato per lo sviluppo del progetto.

Le cellule utilizzate sono cellule HeLa esprimenti stabilmente la proteina SUMF1 recante all'estremità C-terminale il peptide 3xFlag (clone HL3XFS1), necessario per purificare i complessi dei partners molecolari di SUMF1. Il clone di controllo, invece, è costituito da cellule HeLa nelle quali non è stato trasfettato il costrutto di espressione della proteina chimerica ma è costituito da cellule *wild type*. Quest'ultimo costituisce il controllo negativo dell'esperimento di "*fishing for partners*".

2.2.2.2 Estrazione delle proteine

Dopo aver rimosso il mezzo di coltura dalle piastre, le cellule sono state lavate con PBS, recuperate e la sospensione è stata quindi trasferita in un tubo per centrifuga sterile. Quest'ultimo è stato centrifugato per 5 min a 1000 rpm. Il precipitato di cellule così ottenuto è stato risospeso in tampone di lisis (RIPA buffer: 50 mM Tris/HCl, pH 8, 250 mM NaCl, 0.5% sodio deossicolato, 1% NP-40, 0.1% SDS, cocktail di inibitori di proteasi), vortexato, lasciato in ghiaccio per 10 minuti, ed aggiunto il detergente alla concentrazione desiderata nel caso specifico NP-40 1%. Infine sono state poste ad agitare per 30 minuti a 4° C. La sospensione è stata quindi centrifugata per 30 minuti a 4° C a 13200 rpm; il sopranatante, costituente l'estratto proteico totale, è stato filtrato con filtri da 0.45 µm e trasferito in un nuovo tubo. Di tale estratto è stata determinata la concentrazione proteica mediante saggio di Bradford, utilizzando il reattivo di "Bio Rad protein Assay", e la proteina di riferimento per costruire la retta di taratura è l'albumina di siero bovino.

2.2.2.3. Isolamento di complessi proteici mediante tecniche di immunoprecipitazione

Per poter isolare i complessi proteici che coinvolgono SUMF1 è stato eseguito un esperimento di immunoprecipitazione sfruttando la presenza, all'estremità C-terminale della proteina esca, del peptide 3XFLAG, per il quale sono disponibili commercialmente ottimi anticorpi che permettono di purificare i complessi con una buona resa.

- **Pre-cleaning**

È stato prelevato 1 mL di Anti-Mouse Agarose Beads (500 µL di resina secca) per ogni 40 mL di estratto proteico. La resina coniugata con gli anticorpi è stata lavata in modo da eliminare la soluzione di glicerolo. I lavaggi sono stati effettuati aggiungendo una quantità di tampone di lavaggio (Tris-Cl 50 mM, PH 8, NaCl 250 mM) pari a quella di resina per almeno cinque volte, lasciando la sospensione in ghiaccio per alcuni minuti ed infine centrifugando per 2 minuti a 3000 rpm. Gli estratti cellulari sono stati poi incubati con la resina precedentemente lavata, prima over night e poi per due ore.

- **Immunoprecipitazione**

L'estratto è stato poi centrifugato per 5 minuti a 2000 rpm ed è stato recuperato il sopranatante che è stato incubato con la resina anti-FLAG M2 Affinity Gel, lavata precedentemente per sei volte con il tampone di lavaggio. In particolare, la quantità di resina che è stata usata è 80 µL (40 µL di resina secca) per ogni mg di estratto proteico. La determinazione della concentrazione proteica, necessaria a stabilire la quantità di resina da usare, è stata effettuata mediante un saggio Bradford. L'incubazione è stata eseguita a 4° C su ruota over night.

La resina M2 anti-FLAG prima di essere lavata, è stata incubata con un eccesso di soluzione di latte al 5%, per un'intera notte. I lavaggi poi sono stati effettuati aggiungendo alla resina un eccesso di tampone di lavaggio ponendo su ruota a 4°C per 15 minuti ed infine centrifugando per 2 minuti a 3000 rpm.

Dopo l'incubazione con la resina M2 anti-FLAG, gli estratti sono stati centrifugati per 5 minuti a 2000 rpm, il sopranatante è stato scartato, mentre la resina è stata lavata sei volte con il tampone di lisi.

- **Eluizione**

L'eluizione dei complessi legati all'anticorpo presente sulla resina è avvenuta per competizione, grazie alla presenza nel tampone di eluizione del peptide 3XFLAG.

La resina è stata quindi trasferita nelle colonnine di reazione che sono state centrifugate fino a quando non è stato eliminato tutto il tampone di lavaggio. In ogni camera di reazione è stato aggiunto un volume di tampone di eluizione (50 mM Tris/HCl, pH 8, 250 mM NaCl, 0.5% sodio deossicolato, 1% NP-40, 0.1% SDS, cocktail di inibitori di proteasi, peptide 3XFLAG 200 µg ml⁻¹) uguale al volume di resina. E' stata effettuata una prima incubazione a 4° C su ruota per 4 h. Per recuperare tutto l'eluato contenente i complessi coinvolgenti SUMF1 coniugata con il peptide Flag si è quindi proceduto centrifugando a 3000 rpm per 2 minuti. Si è proceduto poi con una seconda eluizione, aggiungendo lo stesso volume di tampone aggiunto per la prima eluizione.

Gli eluati così ottenuti sono stati trattati con metanolo-cloroformio allo scopo di precipitare le proteine presenti.

2.2.2.4. Precipitazione in metanolo e cloroformio

Il protocollo di precipitazione ha previsto la suddivisione dell'eluato in aliquote da 150 µl, cui sono stati aggiunti 600 µl di metanolo per poi agitare su vortex. In seguito sono stati aggiunti, nell'ordine, 150 µl di cloroformio e 450 µl di H₂O. Dopo ciascuna aggiunta si è proseguito agitando su vortex. Si è centrifugato per 5 min a 12000 rpm a T ambiente per eliminare la fase acquosa evitando di allontanare anche il pellet proteico stratificato all'interfaccia con la fase organica. Sono stati aggiunti 450 µl di metanolo e si è di nuovo agitato su vortex prima di ripetere la centrifugazione per 5 minuti a 12000 rpm a T ambiente ed eliminare il surnatante. Il precipitato recuperato è stato portato a secco per 10 min in un sistema Speed-Vac, così da prepararlo per l'analisi elettroforetica.

2.2.2.5. SDS-PAGE

L'SDS-PAGE (Sodio Dodecyl Sulfate-Polyacrilammide Gel Electrophoresis) è una metodica che consente di separare le proteine in base alle loro dimensioni ed al loro peso molecolare relativo. Tale tecnica è eseguita in condizioni denaturanti mediante l'uso di un detergente anionico ossia l'SDS, sodio dodecil-solfato (CH₃(CH₂)₁₀-CH₂O-SO₃⁻). Questo denaturante proteico conferisce una carica netta negativa, praticamente uniforme, alle catene proteiche da separare, la cui migrazione nel gel dipenderà di conseguenza esclusivamente dal peso molecolare. La miscela da separare in SDS-PAGE viene dapprima bollita per 5 minuti in un tampone, loading buffer (100 mM Tris/Cl PH 6.8, 200 mM ditiotreitolo, 4% SDS, 0.2% blu di bromofenolo, 20% glicerolo), che contiene DTT capace di ridurre eventuali ponti di solfuro che tengono insieme la struttura terziaria della proteina e l'SDS che si lega fortemente ad essa e la denatura; in media una molecola di SDS si lega ogni due residui di amminoacido. La carica nativa della proteina viene quindi definitivamente eliminata dall'SDS. Tutti i complessi proteina-SDS si muoveranno verso l'anodo e, grazie alle proprietà di setaccio molecolare del gel, le proteine si separeranno. Facendo correre insieme al campione degli standard di peso molecolare noto è possibile avere un'indicazione del peso molecolare delle proteine della miscela proteica. La corsa elettroforetica è seguita grazie ad un colorante tracciante ionizzabile, di solito il blu di bromofenolo, e il glicerolo presente nel loading buffer rende la densità della soluzione campione tale da stratificare facilmente il campione attraverso il tampone elettroforetico sul fondo, quando viene iniettato nel pozzetto di caricamento.

2.2.2.6. Western blotting

Dopo l'esperimento d'immunoprecipitazione è possibile eseguire una verifica "qualitativa" su una piccola aliquota di campione, con un esperimento di Western blotting. Tale tecnica prevede un frazionamento elettroforetico del campione, seguita da un trasferimento su membrana delle proteine intrappolate nelle maglie del gel e riconoscimento dell'esca mediante anticorpo. Il gel è generalmente 10x10cm di dimensione e 1.5 mm di spessore; nel tampone di trasferimento (Tris-glicina), vengono equilibrati 2 fogli di carta di 3 mm (10x8) cm e 2 spugnette del sistema di trasferimento (Bio-Rad Protean) ed infine la membrana di nitrocellulosa Immobilon-P (Millipore). Si compongono le cassette di sostegno del gel contenenti due spugnette, due fogli di carta di 3 mm, il gel e la membrana di trasferimento, tenendo conto, nella disposizione delle parti, che le proteine sottoposte all'azione del campo elettrico migrano verso l'anodo. Il trasferimento è stato effettuato in costanza di voltaggio a 25 V per tutta la notte a 4° C. Dopo il trasferimento i filtri sono stati bloccati in una soluzione di latte no fat al 5% in PBS 1X per circa 2 ore a temperatura ambiente. Trascorso tale tempo la membrana è stata incubata con l'anticorpo primario diretto contro l'epitopo in una soluzione di latte 1% in PBS 1X più Tween 0.05% per 1 h 30 min.; l'incubazione è seguita da tre lavaggi di 15, 5, 5 minuti rispettivamente ognuno con PBS 1 X più tween 0.05%. Ai lavaggi segue l'incubazione con l'anticorpo secondario, che riconosce l'anticorpo primario, per 45 min., terminata la quale si eseguono nuovamente tre lavaggi ognuno in PBS più tween 0.05%. La rivelazione dell'avvenuto riconoscimento antigene-anticorpo è stata effettuata facendo avvenire sul filtro una reazione di chemiluminescenza (ECL) utilizzando reagenti Amersham Biosciences; i filtri sono poi stati posti ad impressionare una lastra fotografica autoradiografica (Amersham Biosciences).

2.2.2.7. SDS-PAGE preparativa

Una volta preparati e risospesi in loading buffer 2x (100 mM Tris/Cl PH 6.8, 200 mM ditiotreitolo, 4% SDS, 0.2% blu di bromofenolo, 20% glicerolo) i campioni sono stati denaturati a 98° C per 5 min e rapidamente caricati su un gel al 12%, 10x10 cm. La corsa elettroforetica è stata eseguita a voltaggio costante.

2.2.2.8. Colorazione

Le proteine sono state visualizzate mediante una colorazione con Coomassie Colloidale blue (PIERCE).

La visualizzazione delle proteine avviene attraverso l'immersione del gel in una soluzione di Coomassie per un'intera notte; dopodichè il colorante in eccesso è stato rimosso mediante estensivi lavaggi con acqua deionizzata.

2.2.2.9 Deglicosilazione in situ

2.2.2.9.1. Rimozione delle bande proteiche dal gel e decolorazione

Dopo la separazione elettroforetica e la colorazione, le bande proteiche sono state tagliate dal gel con un bisturi, sminuzzate e poste in eppendorf. Sono state quindi decolorate trattandole con 50 μ l di acetonitrile; dopo 15 minuti il solvente è stato rimosso e sono stati aggiunti 50 μ l di una soluzione di NH_4HCO_3 50 mM, pH 8.0. Dopo 5 minuti le particelle di gel sono state centrifugate, il solvente rimosso e quindi eliminato.

2.2.2.9.2. Riduzione ed alchilazione delle proteine in situ

I campioni proteici presenti nelle bande di gel tagliate sono stati sottoposti ad un trattamento con agenti riducenti ed, infine, le cisteine presenti sono state alchilate. A tale scopo le bande di gel sono state trattate con 50 μ l di una soluzione di ditione 10 mM in NH_4HCO_3 50 mM ed incubate per 45 minuti a 56° C. Al termine della riduzione, l'eccesso di soluzione riducente è stato rimosso e i pezzetti di gel sono stati disidratati con acetonitrile; successivamente essi sono reidratati con una soluzione NH_4HCO_3 50 mM, contenente iodoacetammide 55 mM; la reazione è stata condotta al buio e a temperatura ambiente per 30 minuti, al fine di alchilare i gruppi tiolici presenti. Passati 30 minuti la iodoacetammide in eccesso è stata rimossa e le particelle sono state nuovamente trattate con acetonitrile per disidratarle. Esse sono state infine sottoposte a lavaggi NH_4HCO_3 50 mM/ acetonitrile come precedentemente descritto.

2.2.2.9.3 Deglicosilazione delle proteine presenti nel gel

Le bande di gel disidratate sono state trattate con 3 μ l di una soluzione di PNGaseF in AMBIC 50 mM pH 8 e poste a 37°C per 30 minuti. A questo punto, è stato aggiunto NH_4HCO_3 50 mM, in un volume minimo ma tale da coprire i pezzettini di gel; le eppendorf sono quindi state poste a 37° C over night.

2.2.2.9.4 Lavaggi delle bande dopo deglicosilazione

La miscela così ottenuta è stata centrifugata alla massima velocità per pochi secondi ed il surnatante contenente la miscela glicosidica è stata raccolta e posta in eppendorff.

Le bande rimanenti sono state lavate per tre volte con 50 μ l di acqua milliQ e sonicate in un bagnetto apposito per 10 minuti. Tra un lavaggio e l'altro le bande sono state disidratate con acetonitrile.

2.2.2.10. Idrolisi in situ

2.2.2.10.1. Digestione in situ delle proteine presenti nel gel

Le bande di gel disidratate sono state trattate con 10 µl di una soluzione di tripsina (10 ng/µl) in AMBIC 50 mM pH 8 e poste a 4°C per 1 ora. A questo punto, è stato aggiunto NH₄HCO₃ 50 mM, in un volume minimo ma tale da coprire i pezzettini di gel; le eppendorf sono quindi state poste a 37° C per 12 ore.

2.2.2.10.2 Estrazione dei peptidi

La miscela di idrolisi è stata centrifugata alla massima velocità per pochi secondi ed il surnatante contenente la miscela peptidica, raccolto, posto in eppendorf ed acidificato con 2-5 µl di TFA 20% (1 µl di TFA 20% per ogni µl di surnatante).

Ai rimanenti pezzettini di gel è stata aggiunta una piccola quantità di acetonitrile, lasciando il tutto a temperatura ambiente per circa 15 minuti al fine di estrarre completamente eventuali peptidi ancora trattenuti nelle maglie del gel. Alla fine dell'incubazione si è centrifugato e recuperato l'estratto, che riunito al primo surnatante è stato quindi liofilizzato.

2.2.2.11 Identificazione

2.2.2.11.1 Analisi mediante Matrix Assisted Laser Desorption Ionizzazione (MALDI/MS)

Le miscele peptidiche liofilizzate sono state risospese in 10 µl di TFA 2% e caricate negli opportuni pozzetti della piastra portacampioni dello spettrometro, mescolate ad 1 µl di matrice (acido α-ciano-4-idrossicinnamico 10 mg/mL sciolto in una miscela di 70% acetonitrile e 30% Acido citrico 50 mM).

La miscela è stata fatta essiccare all'aria per consentire la formazione dei cristalli di matrice in cui il campione proteico resta intrappolato. Infine, la piastra portacampioni è stata introdotta nella camera di ionizzazione dove in condizioni di alto vuoto il campione è stato bombardato mediante un raggio laser pulsato tra 1 e 10 ns. L'intervallo di m/z analizzato è stato calibrato usando standard interni aventi un opportuno valore di massa.

Gli spettri sono stati acquisiti utilizzando uno spettrometro di massa MALDI Voyager DE-PRO (Applied Biosystems) a sorgente di ionizzazione laser equipaggiato con analizzatore a tempo di volo in grado di separare gli ioni in modalità sia linear che reflectron.

2.2.2.11.2. Analisi mediante LCMS/MS

L'analisi LCMS/MS è stata condotta mediante un sistema LC/MSD Trappola XCT Ultra (Agilent Technologies), equipaggiato con un sistema HPLC 1100 ed un chip cube (Agilent Technologies).

Dopo il caricamento, la miscela peptidica (8 µl in 0.2% acido formico) è stata prima concentrata e poi desalificata a 4 µl /min su una colonna di arricchimento inversa (chip, Agilent Technologies), con 0.1% acido formico (CHOOH) come eluente. Il campione è stato poi frazionato su una colonna capillare C18 a fase inversa (75 µm, 43 mm) ed un flusso di 200nl/min, con un gradiente lineare di B in A dal 5 al 60% in 50 min.

L'eluato è stato introdotto nello spettrometro di massa programmato nella modalità MS/MS data-dependent, dove ogni spettro di massa (intervallo 400-2000 Da/e) è stato seguito da uno o più spettri di massa tandem ottenuti a partire dagli ioni parenti più intensi eluiti in ogni frazione cromatografica.

Una rampa di collisioni energetiche tra 30 e 60 eV è applicata in base alla massa e alla carica del precursore ionico. Gli spettri di frammentazione così ottenuti sono stati analizzati con il programma Analysis List che generano delle peak list che vengono inserite in banche dati disponibili in rete conducendo all'identificazione della proteina.

2.2.2.11.3. Identificazione mediante Mascot

I valori di massa dei peptidi ottenuti in seguito all'analisi mediante MALDI/MS sono stati inseriti nella banca dati di Mascot utilizzando l'opzione Peptide Mass Fingerprint.

Le peak list ottenute in seguito all'analisi mediante LCMSMS mediante Q-TOF sono state inserite, come nel primo caso, nella banca dati di Mascot utilizzando in questo caso l'opzione MS/MS Ion Search.

2.3. Risultati

2.3.1 SUMF1 è secreta e reinternalizza nel Reticolo Endoplasmatico (RE)

SUMF1 è localizzata nel RE, dove attiva tutte le solfatasi neo sintetizzate durante o subito dopo l'importo cotraduzionale [12, 16], ma può essere anche secreta.

A tal proposito, nel laboratorio della dott.ssa Cosma del Tigem con cui questo progetto è stato condotto in collaborazione, è stato prodotto un clone cellulare HeLa che esprime stabilmente SUMF1-3XFlag, chiamato per semplicità clone HL3XFS1. Nell'estratto cellulare derivante da queste cellule, usando un anticorpo anti-Flag, sono state identificate tre bande di circa 42, 39 e 33 kDa che corrispondono alle tre forme della proteina SUMF1 (Figura 2.3A).

Queste differenti bande di SUMF1 corrispondono alle sue forme differenzialmente glicosilate e/o proteoliticamente processate, com'è stato dimostrato precedentemente [18, 19] e come mostrato da analisi di spettrometria di massa degli estratti proteici del clone HL3XFS1. Le solfatasi sono proteine secrete che possono essere reinternalizzate nelle cellule mediante il recettore MPR ubiquitariamente espresso. Iniziando da questa osservazione, ci si è chiesti se la stessa SUMF1 può essere reinternalizzata nelle cellule dal mezzo.

Così cellule Cos7 ed HeLa, sono state incubate nel mezzo di coltura condizionato, raccolto dal clone a confluenza HL3XFS1 dopo 12 h di coltura. Gli estratti proteici derivanti dalle cellule recipienti sono stati analizzati mediante Western Blotting con un anticorpo anti-Flag. Si è visto che SUMF1 è stata reinternalizzata in tutti e due tipi cellulari (Figura 2.3B, lanes 1–4). I controlli includono il mezzo condizionato e non-condizionato (Figura 2.3B, lanes 5 and 6) e un campione del mezzo condizionato non applicato alle cellule recipienti (Figura 2.3B, lane7). Sono stati effettuati dei controlli anche usando un anticorpo anti-tubulina (Figura 2.3B).

Tuttavia, nelle cellule che hanno internalizzato SUMF1, è presente una banda a MW 42 kDa. Questa osservazione ha suggerito l'ipotesi che SUMF1 sia reinternalizzata solo se glicosilata.

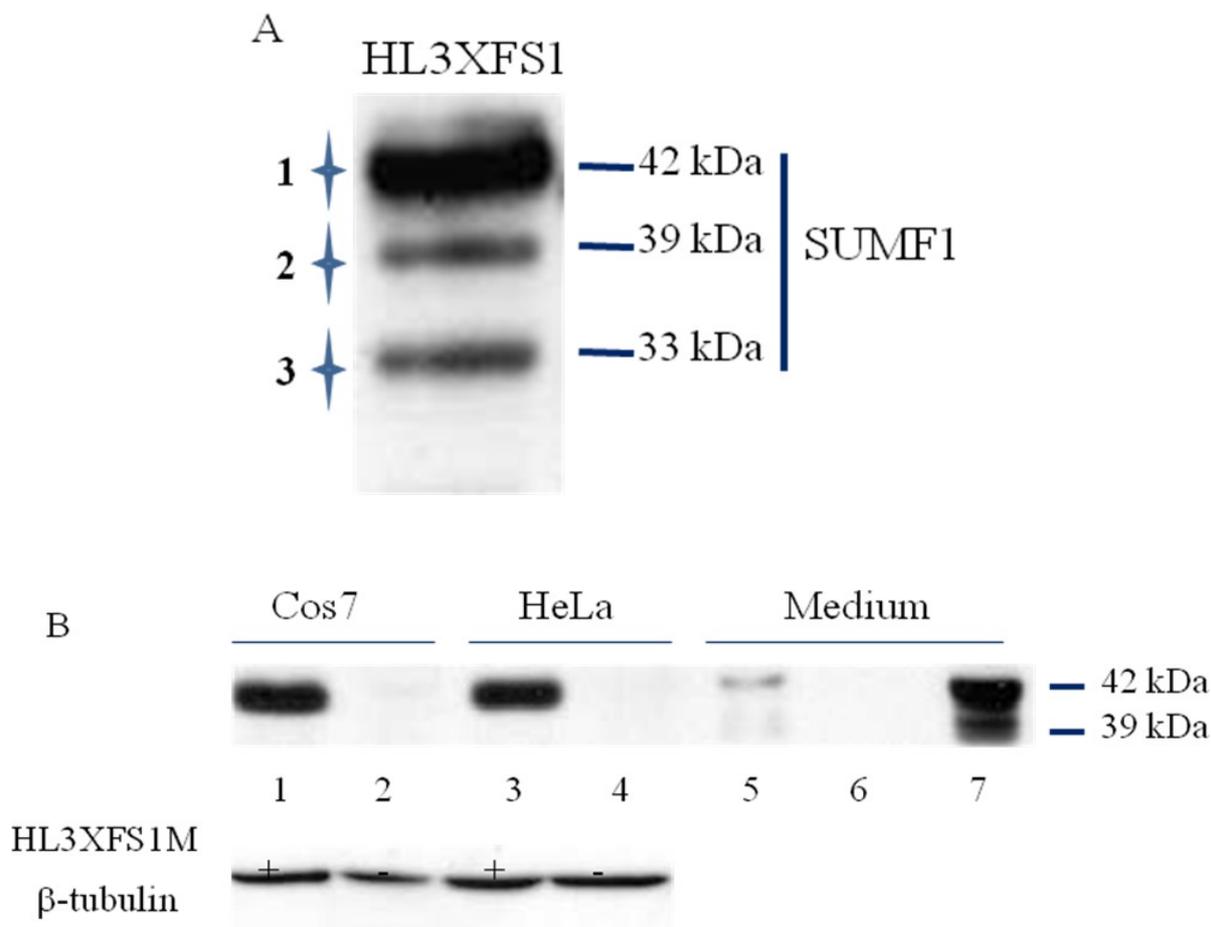


Fig.2.3 Struttura della proteina SUMF1.

Secrezione e reinternalizzazione di SUMF1. (A) Espressione di SUMF1 nel clone HL3xFS1. Sono state rivelate tre bande di circa 42, 39 and 33 kDa mediante Western blotting con un anticorpo anti-Flag. (B) SUMF1 è secreta dal clone HL3xFS1 e reinternalizzata nelle cellule Cos7 and HeLa. Lanes 1–4: estratti cellulari di cellule Cos7 and HeLa incubate con mezzo di coltura condizionato (HL3xFS1M) raccolto dal clone HL3xFS1 o con il mezzo di controllo di cellule HeLa, come analizzato mediante Western blotting. Lane 5 e 6: 5% di SUMF1-Flag condizionato e il mezzo di controllo usato per mettere in cultura le cellule recipienti. Lane 7: 10% del mezzo totale condizionato con SUMF1-Flag. I Western blotting furono condotti con anticorpi anti-Flag o anti-tubulina.

Per confermare questi risultati, è stata eseguita un'analisi di spettrometria di massa di SUMF1-Flag seguendo la sua reinternalizzazione in cellule HeLa recipienti incubate nel mezzo condizionato, recuperato dal clone HL3XFS1. Gli estratti proteici delle cellule recipienti sono stati immunoprecipitati con un anticorpo anti-Flag (come descritto nel paragrafo 2.3.2) e sono state condotte analisi MALDI ed LCMS/MS delle bande escisse, deglicosilate e digerite con tripsina.

Dopo immunoprecipitazione dell'estratto del clone produttore HL3XFS1 con anticorpo anti-Flag, SUMF1 è stato chiaramente identificato in tre bande (Fig 2.4A), corrispondenti alle tre forme di SUMF1, e in particolare alle bande B, D e F nell'estratto cellulare HL3xFS1, mentre nelle cellule dopo uptake sono presenti solo le bande BeD. La banda B corrisponde alla proteina intera, come dimostrato dalla presenza del peptide 1-12 ad m/z 1215.6 (Fig 2.4B). La banda D manca dei primi 33 aminoacidi, come dimostrato dall'esclusiva presenza del peptide N-terminale 34-69 a m/z 3235.7. La banda F corrisponde alla forma tronca all'N-terminale di SUMF1 che inizia a Tyr70. Le bande B e D inoltre sono glicosilate ad Asn141, come dimostrato dalla presenza del segnale ad m/z 1459.7, 1 Da in più rispetto valore di massa teorico del peptide in seguito alla conversione di Asn 141 in Asp dopo trattamento con PNGase F. La banda F invece mostra solo una piccola percentuale di Asn141 glicosilata, come indicato dalla presenza del segnale del peptide non glicosilato 139-151 (m/z 1458.7).

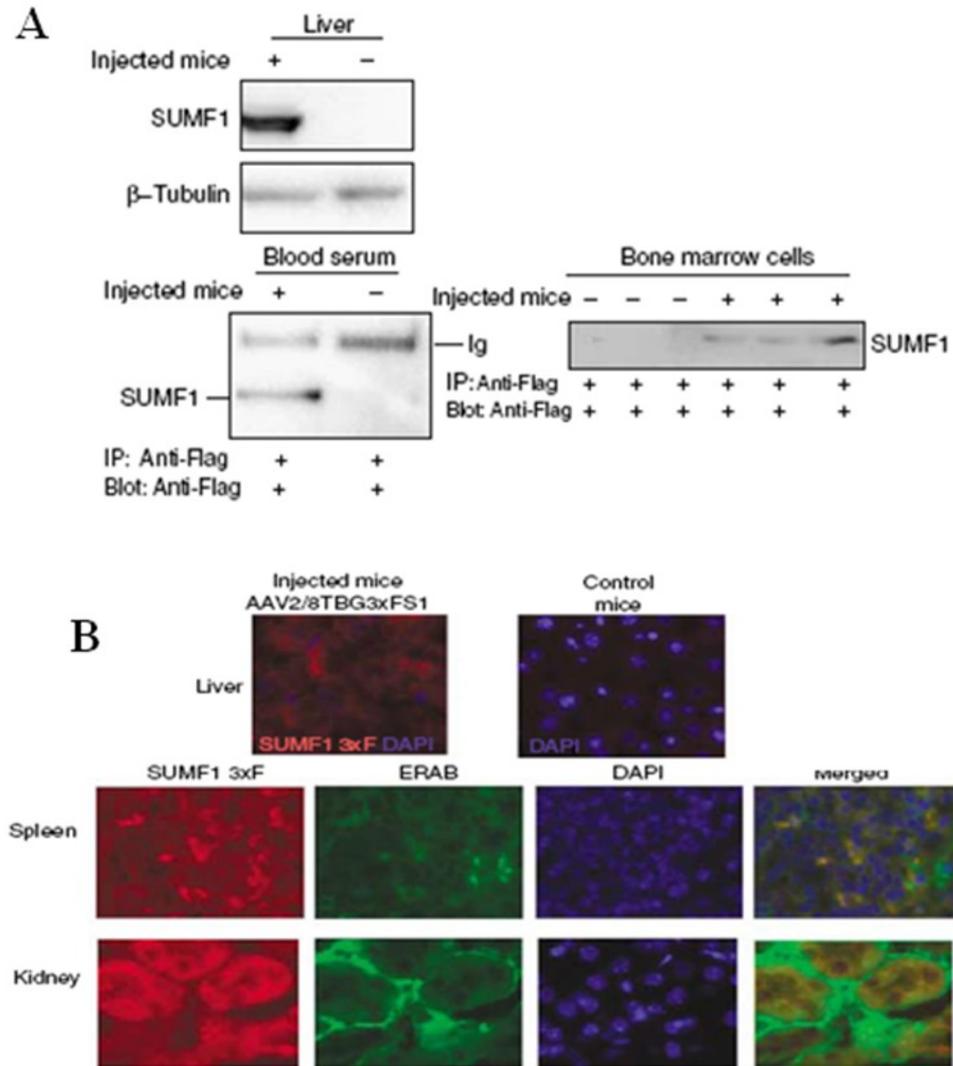


Fig.2.5 Secrezione di SUMF1 in cultura cellulare e in modello animale

(A) Secrezione di SUMF1 dal clone HL3xFS1 e sua reinternalizzazione in cellule Cos7 and HeLa. Analisi western blot di estratti di cellule Cos7 and HeLa dopo la loro incubazione con il mezzo di coltura recuperato da cellule HL3xFS1 overesprimenti Flag-SUMF1; (B) Uptake e secrezione di SUMF1 in tessuti di topi *wild type*; (C) analisi dell'espressione di SUMF1-Flag in fegato. L'uptake e la localizzazione subcellulare di SUMF1-Flag sono state analizzate in milza, rene e polmone mediante fluorescenza con anticorpi policlonali Anti-Flag e anticorpi monoclonali anti-ERAB.

2.3.2 Identificazione delle proteine che controllano la funzione ed il traffico di SUMF1

SUMF1 modifica le solfatasi appena esse entrano nel RE [20] ed è noto che non possiede nessuna sequenza segnale nota di localizzazione nel RE. Inoltre una frazione di SUMF1 può essere secreta e reinternalizzata dalle cellule vicinali, dove rilocalizza nel RE ed attiva le solfatasi [21]. Allo scopo di identificare gli interattori che controllano la sua localizzazione subcellulare e la sua secrezione, 3 mg di estratto proteico derivante da un clone stabile di cellule HeLa esprimenti SUMF1-3xFlag (clone HL3xFS1) [21] e 3 mg derivanti da cellule HeLa non trasfettate, usate come controllo, sono stati immunoprecipitati con anticorpi anti-Flag immobilizzati su una resina di agarosio. In seguito a competizione con il peptide Flag, l'eluato dapprima è stato frazionato mediante elettroforesi monodimensionale e poi analizzato mediante nanoLCMS/MS.

Il primo step d'immunoprecipitazione ha previsto l'incubazione dell'estratto con una resina Anti-Mouse per la fase di pre-cleaning. In questa fase gli estratti sono stati incubati con resina derivatizzata con anticorpi aspecifici al fine di impoverire l'estratto proteico di quelle proteine che, avendo affinità per l'agarosio, lo legano generando poi, nell'esperimento d'immunoprecipitazione, dei "falsi positivi". Gli estratti sono stati incubati con una resina cui sono legati anticorpi di ratto di natura totalmente aspecifica. Per l'esperimento preparativo d'immunoprecipitazione, sono stati utilizzati per ogni campione 25 µl di resina secca Anti-Mouse per il pre-cleaning, sia per l'incubazione overnight su ruota a 4 °C, che per la seconda incubazione di 2 ore.

Trascorso tale tempo il surnatante recuperato è stato incubato con la resina anti-Flag, precedentemente saturata con latte No Fat. Il rapporto ottimale tra anticorpo ed estratto proteico è di 40 µl di resina secca per mg di estratto proteico. Pertanto, l'estratto derivante sia dal clone HL3xFS1 che dalle cellule HeLa non trasfettate è stato incubato con 120 µl di resina secca Anti-Flag. L'incubazione è stata eseguita overnight a 4°C su ruota. Rimosso il surnatante, la resina che ha trattenuto la proteina esca con i suoi complessi è stata sottoposta a una serie di lavaggi con tampone di lisi. Si è infine proceduto con la fase di eluizione. Per limitare l'eluizione di proteine aspecifiche è stata condotta un'eluizione per competizione con il peptide Flag. A tal proposito la resina Anti-Flag è stata incubata per due volte con il tampone di eluizione contenente il peptide Flag a una concentrazione pari a 200 µg/ml, la prima volta per 4 ore mentre la seconda volta overnight. In tale eccesso il peptide Flag compete con il peptide analogo presente all'estremità carbossiterminale della proteina esca nel legame all'anticorpo consentendo pertanto l'eluizione della proteina esca e dei partners ad essa legati.

L'eluato così ottenuto è stato precipitato con metanolo-cloroformio (secondo la procedura descritta nei Metodi), allo scopo di concentrare le proteine presenti, e il precipitato così ottenuto è stato pre-

parato per l'analisi elettroforetica monodimensionale, risospingendolo direttamente nel tampone di caricamento. Le proteine sono state frazionate su un gel monodimensionale al 12%. Il gel colorato con Comassie Colloidale è mostrato in figura 2.6.

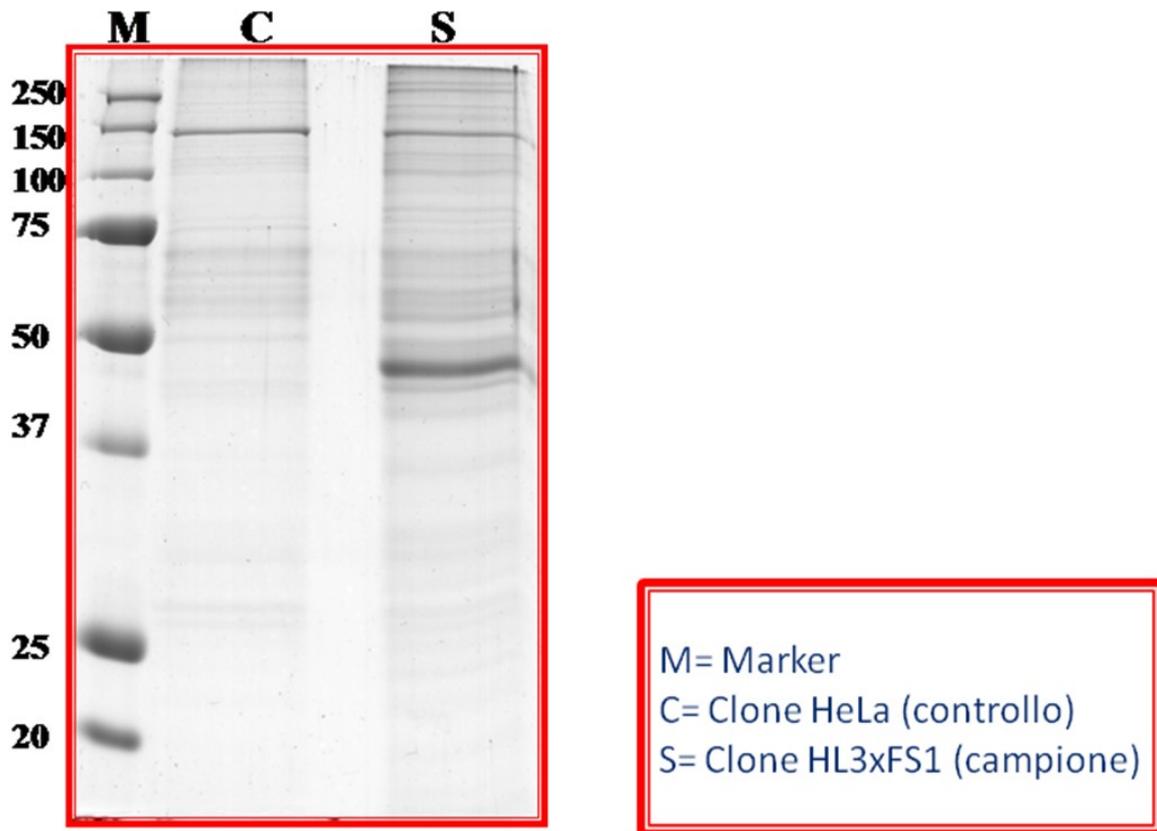


Fig. 2.6

Prima di procedere all'esperimento preparativo è stata verificata l'efficienza dell'immunoprecipitazione utilizzando la resina su cui è legato covalentemente l'anticorpo anti-Flag e rivelando il risultato mediante analisi Western Blot, utilizzando un anticorpo anti-Flag (Figura 2.7).

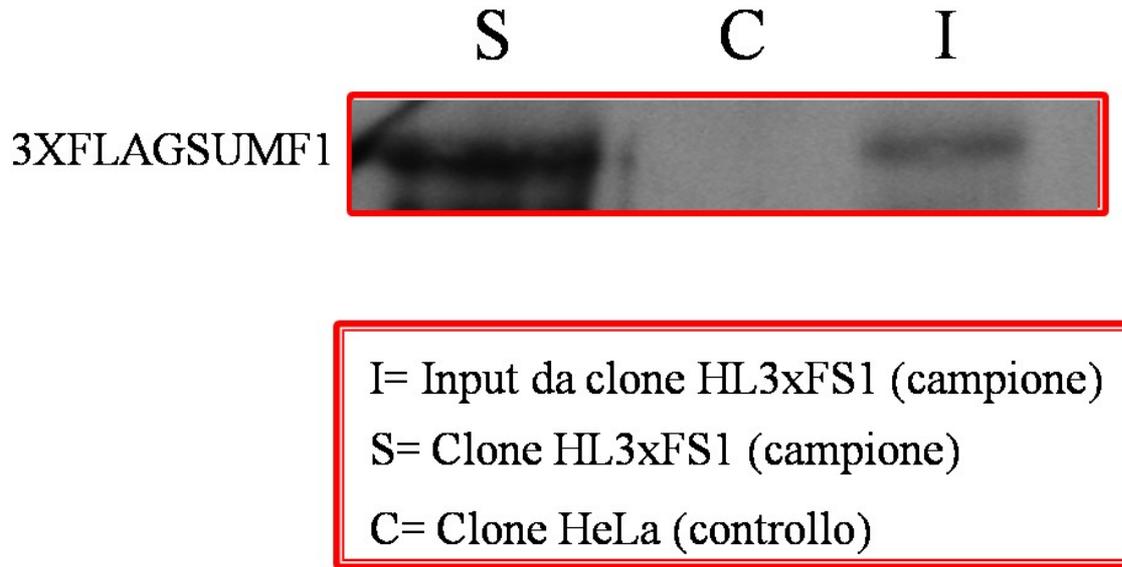


Fig. 2.7 Western Blotting dell'esperimento preparativo.

Dal confronto tra le lane del controllo (C) e quelle del campione (S) si osservano delle bande discriminanti in entrambi i gel; tuttavia un gel monodimensionale non consente di risolvere proteine con un peso molecolare o una migrazione elettroforetica simile. Non è stato quindi possibile affidarsi all'analisi d'immagine per l'identificazione delle bande discriminanti tra le due lane, ma è stato necessario analizzare interamente le lanes di controllo e campione. Tutte le bande sia del campione sia del controllo, sono state escisse, sottoposte ad idrolisi in situ con tripsina e le miscele peptidiche ottenute sono state analizzate mediante la metodologie di Spettrometria di Massa LCMSMS. Tale metodologia abbina procedure di cromatografia liquida capillare a fase inversa con analisi mediante spettrometria di massa tandem, riuscendo non solo ad ottenere i valori di massa accurati dei singoli peptidi ma anche, man mano che la cromatografia procede, di isolare gli ioni corrispondenti ai peptidi e di frammentarli. Dagli spettri di frammentazione si ricavano informazioni concernenti le sequenze dei peptidi. In quest'ultimo caso quindi l'identificazione delle proteine in banca dati è condotta utilizzando due tipi di informazioni e risulta più stringente perché basata non solo sull'impiego dei pesi molecolari accurati, ma anche sulle sequenze delle specie peptidiche analizzate. Le liste dei picchi (*peak list*) così ottenute sono state introdotte nel programma Mascot permettendo di eseguire delle ricerche tra le proteine umane depositate nelle banche dati. Infine, le proteine comuni presenti nel campione e nel controllo sono state scartate e quelle identificate solo ed esclusivamente nel campione sono state considerate putativi interattori di SUMF1 e sono listate in Tabella 2.

Tabella 2. Putativi interattori di SUMF1

Protein name	Swiss Prot code
Filamin A (Actin BP280, FLN-A)	P21333
Filamin B (Actin BP278, FLN-B)	O75369
Myosin heavy chain 9 Non-muscle	P35579
CAD Protein (PYR-1)	P27708
Myosin heavy chain 10 Non-muscle	P35580
Ras GTPase-activating-like protein IQGAP1 (p195)	P46940
UDP-glucose: glucosyltransferase 1 (HUGT1)	Q9NYU2
Hypoxia up-regulated protein 1 (Orp150)	Q9Y4L1
ATP-dependent RNA helicase A (DEAH box P9)	Q08211
ATP-citrate synthase (ACLY)	P53396
Heterogeneous nuclear ribonucleoprotein U (HNRP-U)	Q00839
Endoplasmic (HSP90 member, GRP94. gp96 homolog)	P14625
Neutral alpha-glucosidase AB (Glucosidase II)	Q14697
Alpha-actinin-4 (F-actin cross-linking protein)	O43707
Programmed cell death 6-int.P (ALG-2 intP)	Q8WUM4
Prolyl 3-hydroxylase 1 (LEPRE 1)	Q32P28
Lysyl hydroxylase 2 (PLOD2)	O00469
Mannosyl-oligosaccharide glucosidase	Q13724
Prolyl 3-hydroxylase 2 (LEPREL 1)	Q8IVL5
ATP-dependent RNA helicase DDX1 (DEAD box P1)	Q92499
DNA replication licensing factor MCM5	P33992
Mitochondrial inner membrane protein (Mitofilin)	Q16891
DNA replication licensing factor MCM7	P33993
Semaphorin-3B (Sema V)	Q13214
Lysyl hydroxylase 3 (PLOD3)	O60568
Lysyl hydroxylase 1 (PLOD1)	Q02809
Galectin-3-binding protein (Mac 2 BP)	Q08380
Glucosamine—F6P aminotransferase (G-FAT1)	Q06210
FK506-BP 10 (PPIase, Rotamase)	Q96AY3
Glycosyltransferase 25 family member 1	Q8NBJ5

Protein disulfide-isomerase A4 (Erp72)	P13667
Solute carrier family 25 member 13 (Alarar)	Q9UJS0
Arginyl-tRNA synthetase	P54136
Ribophorin I (RPN1)	P04843
Calnexin	P27824
Proteasome subunit p58	O43242
IGF2 mRNA-binding protein 1	Q9NZI8
Histone-arginine methyltransferase (CARM1)	Q86X55
T-complex protein 1 subunit alpha (CCT1)	P17987
T-complex protein 1 subunit zeta (CCT zeta 1)	P40227
Clusterin (Apo J)	P10909
Heterogeneous nuclear ribonucleoprotein K (HNRP-K)	P61978
T-complex protein 1 subunit theta (CCT8)	P50990
Protein disulfide-isomerase A3 (Erp60)	P30101
Protein disulfide-isomerase (PDI)	P07237
Inosine-5'-monophosphate DH 2 (IMPDH2)	P12268
ERGIC-53 (Lectin mannose-binding 1)	P49257
Nicotinamide phosphoribosyltransferase (Visfatin)	P43490
Aspartyl aminopeptidase	Q9ULA0
Endoplasmic reticulum P46 (Erp46)	Q8NBS9
Protein disulfide-isomerase A6	Q15084
Endoplasmic reticulum resident P44 (Erp44)	Q9BS26
Ataxin-10 (Brain protein E46 homolog)	Q9UBB4
Sulfatase-modifying factor 1 (SUMF1)	Q8NBK3
Reticulocalbin-2 (Calcium-binding protein ERC-55)	Q14257
Reticulocalbin-1 (RCN1)	Q15293
Lactotransferrin (Lactoferrin)	P02788
Endoplasmic reticulum lipid raft-associated P1 (Erlin 1)	O75477
Transmembrane protein 43	Q9BTV4
Serum paraoxonase/arylesterase 2 (PON2)	Q15165
Heterogeneous nuclear ribonucleoproteins C1/C2	P07910
Hornerin (HRNR)	Q86YZ3

ELAV-like protein 1 (HuR)	Q15717
Inorganic pyrophosphatase (PPase)	Q15181
60S Ribosomal Protein L7a	P62424
Voltage-dependent anion-selective channel protein 2	P45880
GTP-BP subunit beta-2-like 1 (RACK1)	P63244
Cell division control protein 2 homolog (CDK1, CDC2)	P06493
Uncharacterized protein KIAA0152	Q14165
40S Ribosomal Protein S2	P15880
40S Ribosomal Protein S3	P23396
60S Ribosomal Protein L13	Q567Q8
Peroxiredoxin-4 (Thioredoxin peroxidase AO372)	Q13162
Heat shock protein beta-1 (HSP 27)	P04792
BAG family molecular chaperone regulator 2 (BAG2)	O95816
40S Ribosomal Protein S7	P62081
60S Ribosomal Protein L10	P27635
Flavin reductase (Biliverdin reductase B)	P30043
Ras-related protein Rab-9A (Rab-9A)	P51151
60S Ribosomal Protein L15	P61313
60S Ribosomal Protein L13a	P40429
60S Ribosomal Protein L21	P46778
Peptidyl-prolyl cis-trans isomerase B (Cyclophilin B)	P23284
60S Ribosomal Protein L18a	Q02543
Programmed cell death protein 6 (ALG-2)	O75340
60S Ribosomal Protein L31	P62899
Myosin regulatory light chain 2 (MRLC2)	P19105
Cofilin-1 (p18 Phosphoprotein)	P23528
ADP-ribosylation factor 1 (ARF1)	P84077

Tra le proteine identificate quelle che hanno destato maggiore interesse sono state PDI, ERp44, ER-GIC-53 ed ERp57. Si è proceduto quindi ad analizzare l'interazione di queste proteine con SUMF1, per stabilire se potessero in qualche modo controllare il suo *trafficking* e la sua localizzazione.

2.3.3 IL legame di PDI con SUMF1 è mediato sia da interazioni covalenti che non covalenti.

PDI è una proteina multifunzionale residente nel RE che catalizza l'ossidazione, la riduzione e l'isomerizzazione dei ponti disolfuri e agisce come chaperone [22]. Possiede quattro domini *thioredoxin-like*, abb'a'. I domini a ed a' posseggono i motivi redox-attivi CxxC e sono richiesti per l'attività ossidasica. I domini b e b' hanno funzioni strutturali, partecipano al legame con il substrato e sono essenziali per l'isomerizzazione dei disolfuri [23-25].

Per confermare e caratterizzare l'associazione di SUMF1 con PDI, cellule HeLa sono state trasfettate con SUMF1-Flag, PDI o con la proteina mutata PDiasas che non possiede le quattro cisteine attive nei domini a ed a' della proteina [26]. Effettuando un esperimento di Western Blotting con un anticorpo anti-Flag ed utilizzando un gel non riducente, nelle cellule over-esprimenti SUMF1-Flag, è stato possibile individuare solo due bande principali, una a 40 e l'altra a 80 kDa che corrispondono al monomero e all'omodimero della proteina [19]. La co-espressione di PDI wild-type (wt) fa decrescere l'intensità della banda corrispondente all'omodimero di SUMF1 (Fig. 2.8, lane 2) Si notano anche due bande a pesi molecolari inferiori (Fig. 2.8, lane 2, asterischi), che corrispondono agli eterodimeri covalenti PDI-SUMF1. Queste bande scompaiono nelle cellule in cui SUMF1-Flag è stata co-trasfettata con il mutante PDiasas (Fig. 2.8, lane 3). Questi esperimenti dimostrano che PDI lega covalentemente SUMF1 attraverso un legame S-S intermolecolare facendo diminuire fortemente la formazione degli omodimeri di SUMF1.

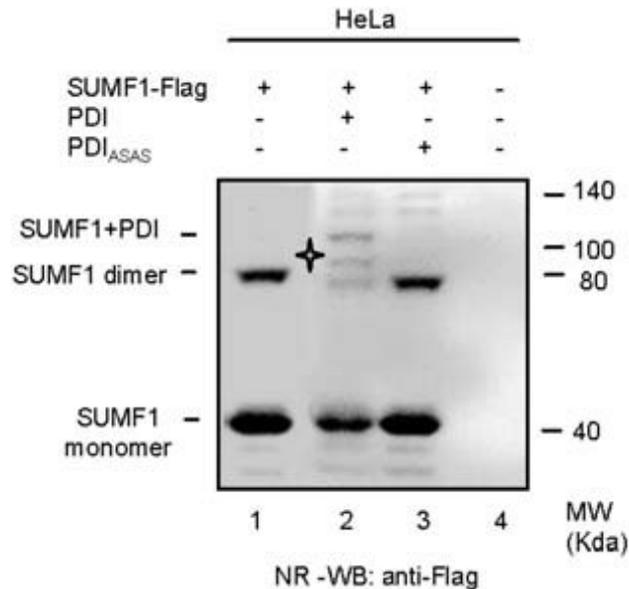


Fig.2.8 PDI interagisce con SUMF1.

Cellule HeLa sono state co-trasfettate con SUMF1-Flag and PDI o PDI_{ASAS}. Le cellule sono state trattate con N-ethylmaleimide per evitare l'interscambio dei disolfuri. Aliquote dei lisati sono stati resolti mediante un gel SDS-PAGE gradientato non-riducente. I filtri sono stati rivelati con anticorpi anti-Flag. Gli asterischi indicano le bande corrispondenti ai disolfuri misti SUMF1-PDI.

Per stabilire se PDI e SUMF1 potessero associare anche mediante interazioni non covalenti, cellule HL3xFS1 trasfettate sia con PDI sia con PDiasas sono state immunoprecipitate con anticorpo anti-Flag. Uguali quantità di immunocomplessi furono caricati su un gel riducente e poi analizzati mediante Western Blot con anticorpo anti- PDI. Chiaramente, un po' di PDI endogeno immunoprecipita con SUMF1-Flag (Fig 2.9, lane 1). La banda è più intensa nelle cellule overesprimenti PDI che nelle cellule overesprimenti PDiasas che non può interagire covalentemente con il suo substrato (Fig. 2.9, lanes 2 e 3). Questi risultati dimostrano che PDI associa covalentemente con SUMF1 mediante le cisteine dei siti attivi a ed a'. Prendono luogo anche delle interazioni non covalenti, come confermato dall'interazione di PDiasas con SUMF1.

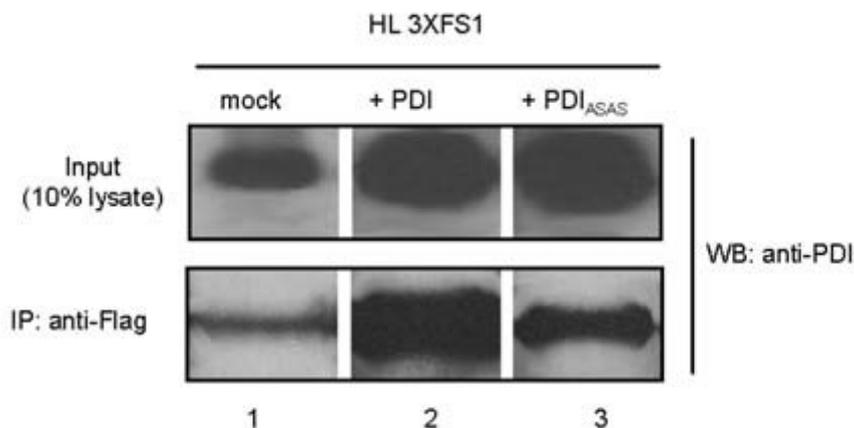


Fig.2.9 PDI associa con SUMF1 anche mediante interazioni non-covalenti.

Cellule HL3xFS1 esprimenti SUMF1-Flag sono state trasfettate con PDI o PDI_{ASAS}. Gli estratti proteici sono stati immunoprecipitati con anticorpi anti-Flag, gli immunocomplessi sono stati risolti in condizioni denaturanti ed infine e' stato effettuato un esperimento di Western Blot con anticorpi anti-PDI.

2.3.4 Il legame di ERGIC-53 con SUMF1 è mediato da N-glicani.

ERGIC-53 è una lectina che si lega alla membrana ed è capace di muoversi tra il RE e il Golgi. Ha la funzione di recettore cargo [27], catturando certe glicoproteine (es. Fattore V ed VIII) nel RE e rilasciandole nel Golgi in maniera dipendente dal pH [28]. Inoltre in associazione con ERp44, Ergic-53 può assistere alla polimerizzazione delle IgM [29].

Per validare l'interazione di ERGIC-53 con SUMF1 sono state analizzate cellule HeLa non trasfettate. Dopo aver immunoprecipitato con anticorpo anti-ERGIC-53 gli estratti proteici derivanti da queste cellule, gli immunocomplessi sono stati risolti in condizioni denaturanti su gel 1D e poi rivelati con anticorpi anti-SUMF1. Sono state evidenziate bande corrispondenti a SUMF1 sia nell'immunoprecipitato sia nell'input, ma non nel controllo (Fig.2.10, lane 1, 2 e 3).

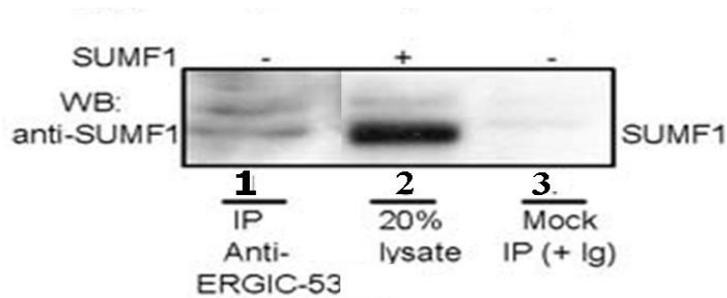


Figura 2.10. Interazione tra SUMF1 ed ERGIC-53 endogeni

Gli estratti proteici di cellule HeLa non trasfettate sono stati immunoprecipitati con anticorpi anti-ERGIC-53, gli eluati frazionati su gel riducente e i blot rivelati con anticorpi anti-SUMF1. Nella lane 3 è stata caricata la proteina SUMF1 overespressa per facilitare l'identificazione della banda rilevante.

ERGIC-53 è una lectina che lega glicoproteine mediante un dominio di riconoscimento dei carboidrati (CRD), in modo Ca^{2+} dipendente (ref 19 HMG). Per investigare se anche l'associazione tra SUMF1 ed ERGIC-53 è mediato dall'interazione lectina-carboidrato, sono state trasfettate cellule KKFF con SUMF1N141A-Flag, un mutante di SUMF1 non glicosilato [21]. In accordo con l'ipotesi che l'interazione potrebbe avvenire via carboidrati, il mutante non glicosilato di SUMF1 non associa con ERGIC-53 (Fig. 2.11, lane 3). Questi dati dimostrano che il legame di SUMF1 ed ERGIC-53 è mediato da interazione N-glicani-lectina.

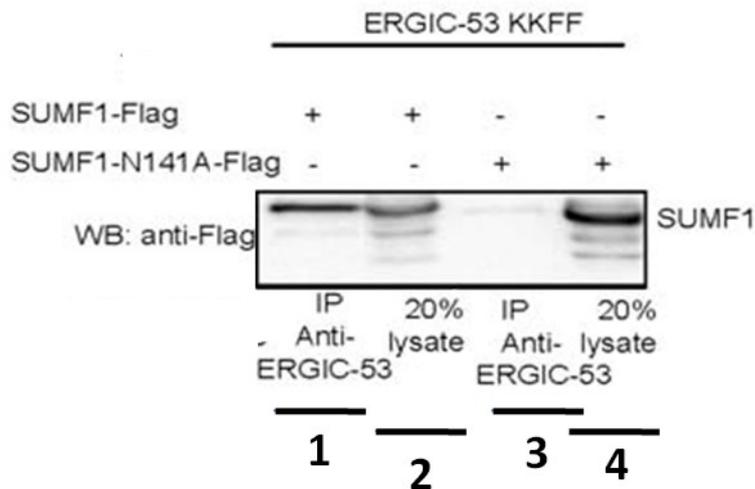


Figura 2.11. Il legame tra SUMF1 ed ERGIC-53 dipende da N-glicani.

Cellule ERGIC-53KKFF sono state transfettate con SUMF1-Flag o SUMF1N141A-Flag non glicosilata e incubate per 48 h+ con tetraciclina per modulare l'espressione di ERGIC-53. L'immunoprecipitazione e il Western Blotting sono stati condotti con gli anticorpi indicati.

2.3.5 L'interazione di ERp44 con SUMF1 è mediata da tioli.

ERp44 è una proteina solubile appartenente alla famiglia delle tioredossine (TRX), codificata con un segnale di localizzazione per il RE, RDEL [32], che si accumula nell'ERGIC (compartimento intermedio tra il RE ed il Golgi) ed il cis-Golgi [29-31]. E' responsabile della ritenzione mediata da tioli di molte proteine substrato, includendo Ero1a, adiponectina ed IgM [29, 31, 33]. ERp44 è implicata anche nell'omeostasi del calce del RE, mediante l'interazione e la regolazione di IP3R1 [34]. L'interazione tra SUMF1-Flag ed ERp44 endogena è stata confermata sia in cellule HeLa sia in cellule 293T (Fig. 2.12).

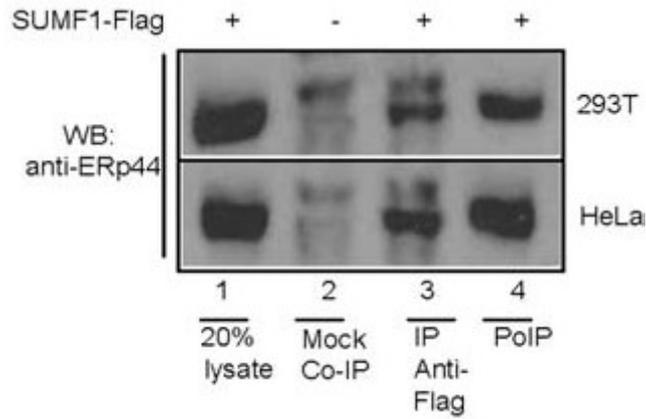


Fig. 2.12. ERp44 endogena associa con SUMF1 overespressa.

Cellule HeLa and 293T sono state trasfettate con il plasmide esprimente SUMF1-Flag. Gli estratti proteici sono stati immunoprecipitati con anticorpi anti-Flag e i western blot rivelati con anticorpi anti-ERp44.

Diversamente da ERp44 wt, il mutante ERp44C29S che al posto della Cys in posizione 29 presenta una Ser, non coprecipita con SUMF1 (Fig.2.13, lane 5,8 pannello in basso), confermando che il legame dipende dalla Cys29 di ERp44.

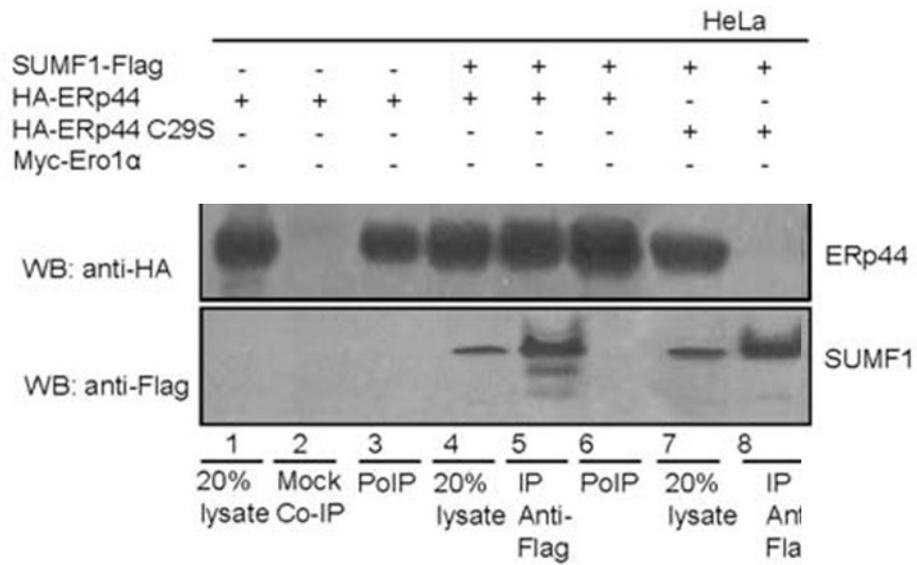


Fig. 2.13. SUMF1 lega ERp44 via interazioni tiolo-dipendenti.

Cellule HeLa sono state trasfettate con SUMF1-Flag, HA-ERp44, HA-ERp44C29S, come indicato. Gli estratti cellulari sono stati immunoprecipitati con anticorpi anti-Flag e blottati con anti-HA o anti-Flag.

2.3.6 PDI, ERGIC-53 e ERp44 modulano co-operativamente la ritenzione e la secrezione di SUMF1 e la sua attività enzimatica.

Avendo stabilito che SUMF1 interagisce specificamente con PDI, ERp44 ed ERGIC-53, si è testato il loro ruolo nella localizzazione subcellulare SUMF1.

Inizialmente si è determinato se l'over-espressione di PDI, PDI_{ASAS} o PDI Δ KDEL, un mutante che manca della sequenza C-terminale KDEL, ha effetto sulla secrezione di SUMF1. Come mostrato in figura (Fig.2.14a), PDI previene la secrezione di SUMF1 e incrementa il suo accumulo intracellulare (Fig.2.14a, lanes 1 e 2), indicando che PDI può ritenere SUMF1. PDI_{ASAS} ha invece un minore effetto sulla ritenzione/secrezione di SUMF1, riflettendo la sua capacità di stabilire solo interazioni non covalenti (Fig. 2.14a, lane 3). Come atteso, il mutante di secrezione PDI Δ KDEL ha un effetto solo marginalmente sulla localizzazione di SUMF1. È interessante notare che ERp57, un'altra thiol disulfide ossidoreduttasi identificata mediante analisi MS come interattore di SUMF1, inibisce anch'essa la secrezione di SUMF1 (Fig. 2.14a, lane 5).

Per confermare che PDI controlla la ritenzione e l'attività di SUMF1, questa proteina è stata silenziata nelle cellule HeLa. Nelle cellule interferite si nota un significativo decremento di SUMF1 nel pellet cellulare con un incremento della secrezione (Fig. 2.14 b). Questo risultato indica che PDI controlla la ritenzione di SUMF1 e ne modula la secrezione.

Analogamente per chiarire la funzione di ERGIC-53 ed ERp44 nel traffico di SUMF1, le cellule HeLa sono state depletate di una o entrambe le proteine mediante *silencing*. Silenziando ERGIC-53 l'ammontare di SUMF1 detectabile negli estratti cellulari diminuisce sensibilmente, ma la frazione di SUMF1 secreta nel mezzo non aumenta significativamente (Fig.2.14c). Anche silenziando ERp44 si causa una simile diminuzione nell'estratto cellulare di SUMF1, ma a differenza di quanto avviene con ERGIC-53, riducendo i livelli di ERp44 aumentano i livelli di secrezione di SUMF1 (Fig. 2.14 d). Presi insieme, questi dati suggeriscono che sia ERp44 che ERGIC-53 modulano il trafficking di SUMF1; comunque esse hanno differenti funzioni. ERp44 gioca un ruolo importante nella localizzazione di SUMF1 nel RE.

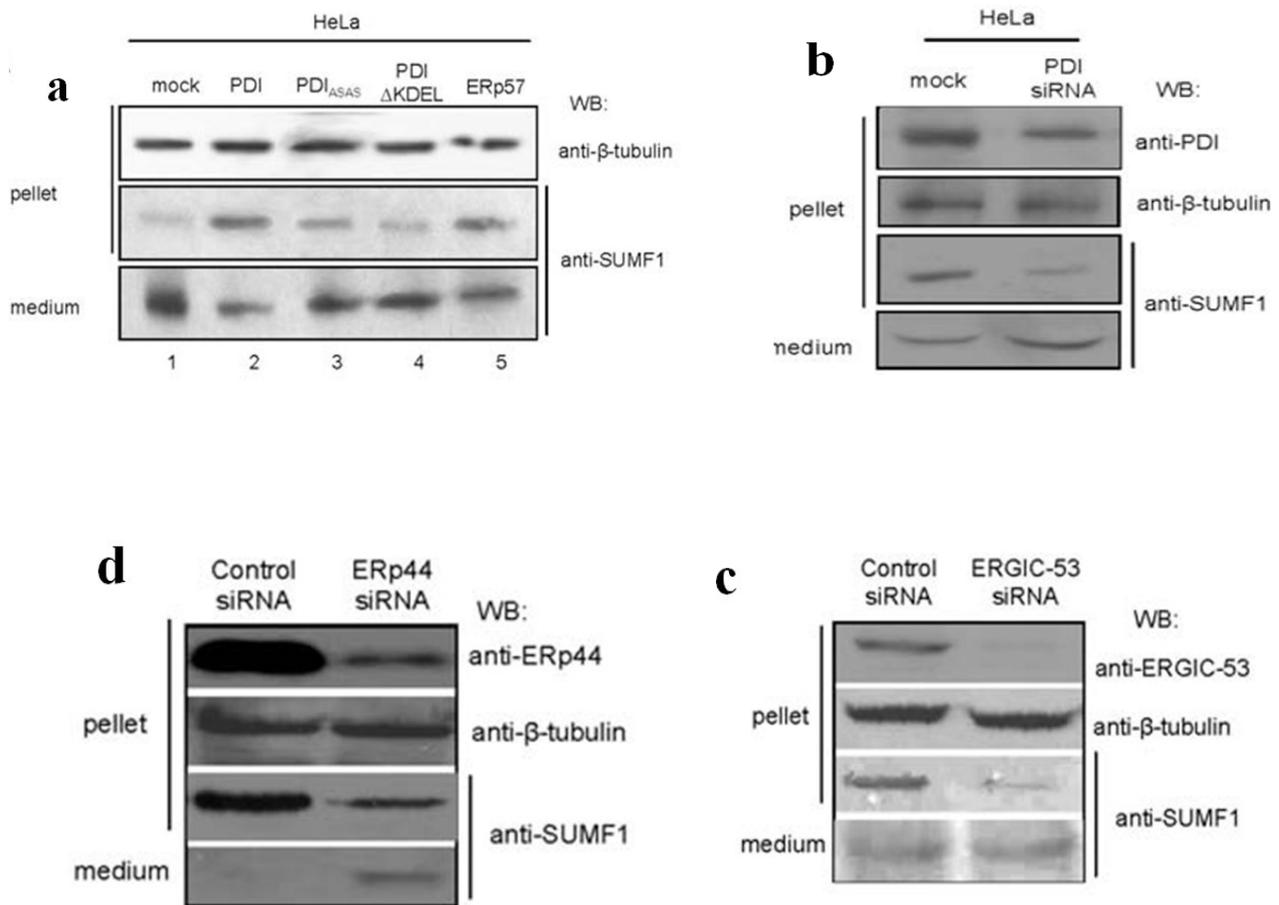


Fig. 2.14. PDI, ERGIC-53 ed ERp44 controllano il traffico di SUMF1.

a. Cellule HeLa sono state trasfettate con i plasmidi di espressione di PDI, PDI_{ASAS}, PDI_{ΔKDEL} o ERp57. Due giorni dopo la trasfezione, gli estratti proteici derivanti da 2×10^6 cellule ed il mezzo di coltura corrispondente (condizionato per 16 ore) furono testati con l'anticorpo anti-SUMF1. L'anticorpo anti-tubulina è stato usato come controllo; **b.** Cellule HeLa sono state trasfettate con il mock o con il plasmide esprimente il siRNA diretto verso PDI. Ventiquattro ore dopo la trasfezione i pellet cellulari e il corrispondente mezzo di coltura condizionato sono stati analizzati mediante Western Blotting con gli anticorpi specifici indicati; **c.** Il silenziamento di ERGIC-53 causa la perdita di SUMF1 intracellulare senza incrementare la sua secrezione. Le cellule HeLa sono state trasfettate con il siRNA fluorescente diretto contro ERGIC-53. Tre giorni dopo la trasfezione i pellet cellulari ed il mezzo di coltura (condizionato 16 h) sono stati testati con gli anticorpi indicati; **d.** Il silenziamento di ERp44 aumenta la secrezione di SUMF1. Le cellule HeLa sono state trasfettate con il siRNA fluorescente di ERp44. Tre giorni dopo la trasfezione i pellet cellulari ed il mezzo di coltura (condizionato 16 h) sono stati testati con gli anticorpi indicati.

2.3.7 Distribuzione subcellulare di SUMF1

Infine è stata determinata la co-distribuzione di SUMF1 con PDI, ERp44 ed ERGIC-53 mediante colorazione ad immunofluorescenza (Fig 2.15A). Come atteso, è stato trovato che la maggior parte di SUMF1 co-localizza con PDI nel RE, come mostrato in precedenza [12], così come con ERp44. La co-localizzazione di SUMF1 con ERGIC-53 è stata trovata principalmente nel RE, ed in misura minore nell'ERGIC probabilmente perché le interazioni in questo compartimento sono molto transienti. Questo suggerisce che SUMF1 endogeno è efficacemente e rapidamente recuperato nel RE. La distribuzione di SUMF1 in compartimenti differenti rispetto all'apparato di secrezione iniziale è stata confermata dagli esperimenti del gradiente di densità. L'immunoreattività di SUMF1 è abbondante nelle frazioni contenenti la maggior parte di PDI e parte di ERp44. Comunque, una minore quantità di SUMF1 si co-distribuisce con ERp44 ed ERGIC-53 entro le frazioni meno dense (Fig. 2.15B).

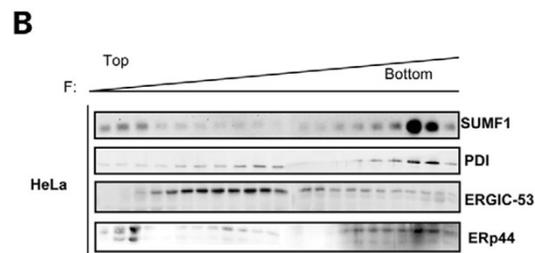
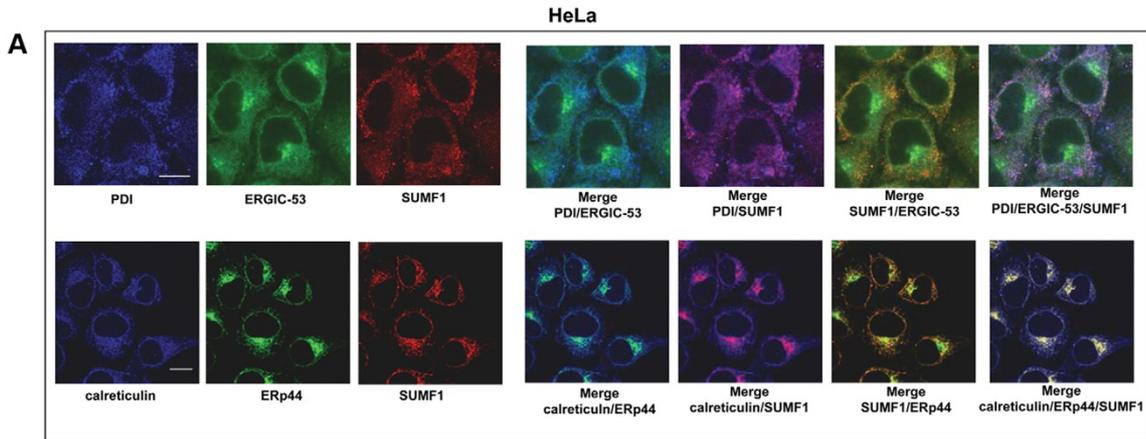


Fig. 2.15. Didistribuzione subcellulare di SUMF1.

(A) Localizzazione per immunofluorescenza di SUMF1 con PDI, ERp44 and ERGIC-53. Cellule HeLa sono state co-colorate con anticorp specifici fluorescenti per SUMF1, ERp44, PDI, Calreticulin and ERGIC-53. (B) Co-frazionamento di SUMF1, ERp44, ERGIC-53 and PDI. Le membrane cellulari postnucleari di cellule HeLa sono state frazionate mediante gradiente continuo Optiprep (8–24%). Sono state collezionate venti differenti frazioni e risolte su un gel riducente. Le proteine sono state rivelate con anticorpi anti- SUMF1, ERp44, ERGIC-53 o PDI, come indicato.

2.4. Conclusioni

E' stato possibile dimostrare che SUMF1 è secreta e poi reinternalizzata dal mezzo di coltura di diverse linee cellulari e dal plasma nei tessuti di topi transgenici. Le sole proteine che sono state caratterizzate come capaci di muoversi dall'ambiente esterno verso le cellule mediante la membrana plasmatica per raggiungere il RE sono solo alcune tossine [35]. In questo caso invece siamo di fronte ad una proteina di mammifero. In questo modo sorge una questione: come fa SUMF1 ad essere secreta dal RE di alcune cellule e raggiungere il RE di altre cellule? Per quanto riguarda la nostra ricerca, abbiamo solo iniziato a determinare come SUMF1 lascia il RE per raggiungere l'ambiente extracellulare.

SUMF1 non ha un segnale di ritenzione per il RE, sebbene risieda nel lume del RE. Date le condizioni di estrazione con cui abbiamo recuperato SUMF1 dalle cellule, si potrebbe pensare che SUMF1 è associato alle membrane. Quindi, la nostra ipotesi è che dal momento che SUMF1 non ha un segnale per il RE, la proteina deve essere associata alla membrana del RE mediante un interattore sconosciuto che agisce come un recettore per ritenerla nel RE. E' stato suggerito che il trasporto delle glicoproteine fuori dal RE è regolato dalle lectine, che si legano specificamente alle glicoproteine e funzionano come recettori. Il ruolo di queste lectine "recettoriali" è di ritenere le glicoproteine piuttosto che trasportarle attivamente. Così, il trafficking di molte proteine potrebbe essere regolato da recettori, che armonizzano la loro ritenzione e la loro uscita. In base a questo, si può pensare che la glicosilazione è importante per SUMF1 per essere ritenuta nel RE mediante un recettore ancora sconosciuto che interagisce con essa, sebbene una frazione di SUMF1 può essere secreta. Interessatamente, SUMF1 non glicosilata è ancora capace di attivare le solfatasi. Così, si potrebbe pensare che SUMF1 non glicosilata folda correttamente e la glicosilazione ha un ruolo esclusivo nella ritenzione della proteina nel RE, piuttosto che per la sua attività. Quindi sebbene la maggior parte delle glicoproteine ha bisogno di oligosaccaridi *N-linked* durante il folding, il grado di dipendenza è variabile. Alcune proteine mostrano totale o solo parziale misfolding in assenza di zuccheri, ma ci sono numerose proteine che foldano correttamente senza i loro zuccheri *N-linked* [36].

Una volta che SUMF1 è secreta, può essere reinternalizzata da altre cellule, includendo linee cellulari immortalizzate, così com'è stato confermato mediante analisi di spettrometria di massa. Nelle nostre condizioni cellulari, la perdita degli zuccheri, cioè del contenuto di mannosio, è cruciale per abrogare la reinternalizzazione. Comunque, non possiamo escludere che in differenti linee cellulari, meccanismi di compensazione alternativa possono permettere la parziale reinternalizzazione di SUMF1. Dalla struttura cristallina e dalla spettrometria di massa, è chiaro che SUMF1 non contiene mannosio-6-fosfato [37].

D'altra parte, la questione di quanto SUMF1 viaggia dalla membrana plasmatica al RE rimane completamente aperto. Le tossine sono state le prime molecole per le quali è stato possibile dimostrare la capacità di entrare nella cellula mediante gli endosomi ed essere trasportate al Golgi, e da quel compartimento al RE [35]. Il trasporto retrogrado dal Golgi al RE potrebbe essere sia COP-I dipendente che indipendente. Il pathway COP-I-indipendente è regolato dalla GTPase Rab6, ed è usato per la glicosilazione degli enzimi, la tossina Shiga (ST) e ricina [38]. Al contrario, la lectina ERGIC-53, che è un recettore cargo per le glicoproteine, e la tossina del colera usano un pathway COP-I-dipendente per raggiungere il RE [35, 38].

A tal proposito, è stato interessante identificare quale pathway usa SUMF1 per viaggiare attraverso il RE e il Golgi. Infatti, il passo successivo è stato proprio di identificare quelle proteine che aiutano SUMF1 in questo suo viaggio verso l'esterno della cellula.

SUMF1 attiva tutte le solfatasi nel RE ossidando una specifica cisteina del loro dominio catalitico [20]. Comunque, SUMF1 perde il segnale di localizzazione del RE e parte di essa è secreta nel mezzo di coltura. Quanto sono armonizzate e regolate queste caratteristiche? Sono state identificate tre proteine, PDI, ERGIC-53 ed ERp44 che orchestrando il bilanciamento tra la ritenzione di SUMF1 nel RE e la sua secrezione, modulando la sua attività. Dopo il suo traslocamento nel RE, SUMF1 deve foldare, con l'assistenza di chaperones, lectine ed enzimi, includendo due molecole, Calnexin e ERp57, che sono state identificate nella nostra analisi MS. Una volta foldata, un'ampia frazione di SUMF1 deve rimanere nel RE, in modo da esercitare la propria funzione sulle solfatasi in arrivo. L'incremento e il decremento dei livelli di PDI mostra che l'interazione con questa proteina è importante per il mantenimento di SUMF1 nel RE. PDI è uno chaperone redox-dipendente che media la formazione, l'isomerizzazione e la riduzione dei ponti disolfuri di numerosi substrati, mediante la formazione di disolfidi misti transienti. In generale è difficile intercettare questi intermedi nelle cellule viventi, a meno che la seconda cisteina del motivo CxxC, che risolve gli intermedi, sia soppressa. SUMF1 è chiaramente uno speciale substrato di PDI e gli eterodimeri formati da queste due proteine sono facilmente rivelati nelle cellule. Comunque, il ciclo di vita di SUMF1 non è limitato al RE: parte di essa è secreta e reinternalizzata da cellule distanti. I nostri risultati identificano SUMF1 come un nuovo substrato di ERGIC-53. Quest'ultima media il trasporto di glicoproteine selezionate [27] e interagisce con SUMF1 via dominio CRD, e potrebbe promuovere il suo esporto dal RE mediante le vescicole rivestite da COPII. Una volta nel cis-Golgi, il pH più basso potrebbe favorire la dissociazione di SUMF1 da ERGIC-53 (Fig. 2.16), come descritto per altri substrati [28]. Sorprendentemente, il destino di SUMF1 in assenza di ERGIC-53 implica la degradazione proteosomiale, piuttosto che l'accumulo nel RE. Quest'osservazione suggerisce che la degradazione e l'esporto sono in competizione per una frazione di molecole di SUMF1 [39]. Una volta nel cis-Golgi, SUM-

F1 può procedere verso la membrana plasmatica ed essere secreta o legarsi a ERp44 ed essere riportata nel RE (Fig. 2.16). Questa nozione è supportata da molte evidenze: primo, l'over-espressione di ERp44 attivo inibisce la secrezione. Secondo, i disolfuri misti SUMF1-ERp44 sono facilmente osservabili in diversi tipi cellulari. Terzo, il danneggiamento della funzione di ERp44 aumenta la secrezione dell'enzima. Simili risultati sono stati recentemente ottenuti indipendentemente [40]. Una significativa differenza emersa rispetto agli studi da noi condotti è che i mutanti ERp44 Cys29Ala mantengono la capacità di legame di SUMF1 [40]. I fenotipi differenti osservati con i due mutanti di ERp44 potrebbero essere spiegati con l'inserzione di un residuo più idrofilico in posizione 29, e quindi entro il sito attivo di legame, che potrebbe parzialmente inibire le interazioni non covalenti che ERp44 stabilisce con le proteine assistite. In tutte le linee cellulari analizzate finora, ERp44 localizza anche a valle del RE, e una frazione considerevole si accumula nell'ERGIC e nel cis-Golgi. In questi compartimenti, si pensa che ERp44 catturi gli intermedi ossidativi di folding (es. IgM, adiponectina) e li riporta al RE mediante meccanismi dipendenti da RDEL [29,31]. I nostri dati suggeriscono che ERp44 riporta SUMF1 rapidamente indietro nel RE. Pertanto, essendo ERp44 essenziale per la funzione di SUMF1, le cellule depletate di ERp44 secernono SUMF1 e le solfatasi nascenti sono parzialmente inattivate. Il silenziamento sia di ERGIC-53 sia di ERp44 quindi causa una rilevante diminuzione di SUMF1 intracellulare. Presi insieme, questi risultati suggeriscono un meccanismo di localizzazione altamente dinamico. In conclusione, SUMF1 può essere ritenuto nel RE o secreto nella matrice extracellulare, secondo l'azione concertata di PDI, ERp44 ed ERGIC-53, e questo controllo multistep modula a sua volta la sua attività sulle solfatasi. I nostri dati rivelano un sottile meccanismo di localizzazione che coinvolge il riciclaggio di molecole enzimaticamente attive, competenti nel trasporto. Questa duplice localizzazione fornisce un pool di molecole SUMF1 facilmente mobilitate in secrezione. Il controllo multistep delle interazioni individuato sintonizza la quantità di SUMF1 circolante. L'aumento della secrezione di SUMF1 in seguito a certi stimoli, come l'iperossidazione del RE, potrebbe permettere così un addizionale livello di regolazione dell'attività delle solfatasi in alcune cellule e tessuti.

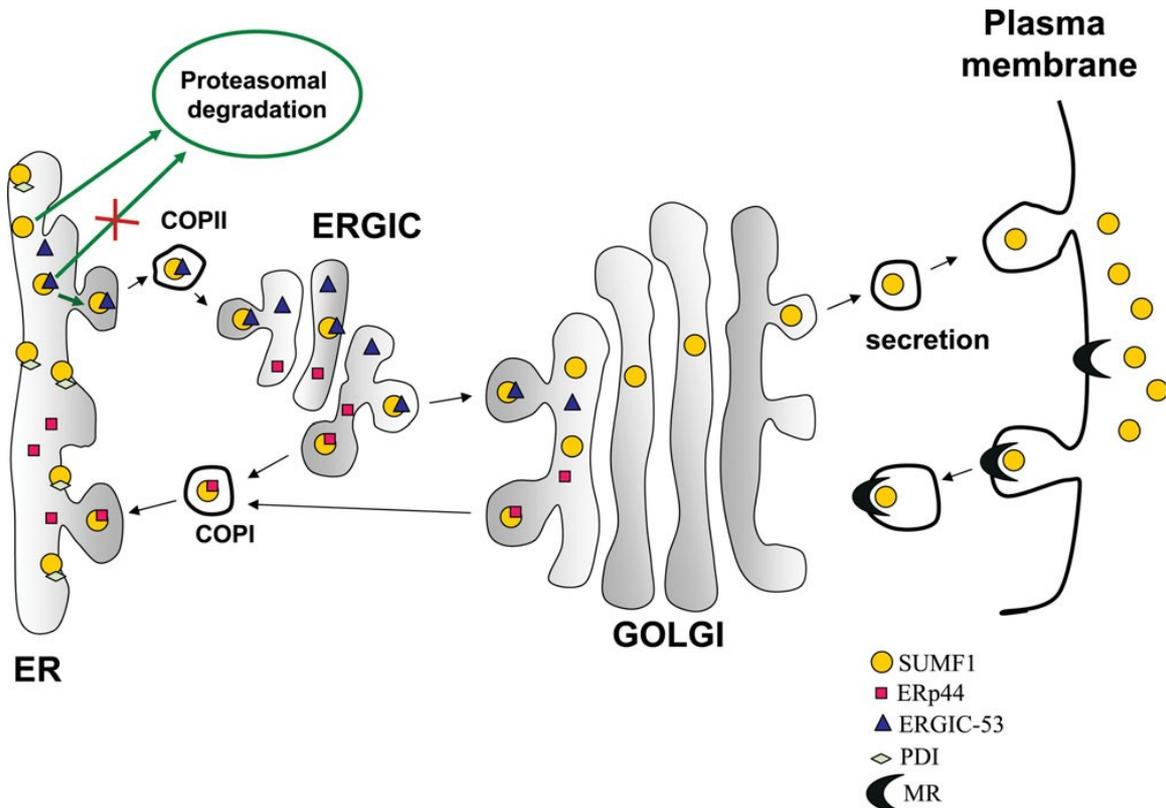


Fig. 2.16. Il traffico di SUMF1: un modello.

All'interno del RE, SUMF1 folda e interagisce con PDI. L'attività redox di PDI è importante per l'attivazione delle solfatasi dipendenti da SUMF1. Quest'ultima lega anche ERGIC-53 mediante interazioni lectinoglicane. Quest'interazione protegge SUMF1 dalla degradazione proteosomiale e la consegna al cis-Golgi. Qui, parte di SUMF1 è recuperata da ERp44 mediante meccanismi RDEL-dipendenti e parte di essa può muoversi in avanti e raggiungere il mezzo di coltura.

Bibliografia SUMF1

1. Hanson SR, Best MD, Wong CH. 2004. Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed. Engl.* 43:5736–63.
2. Hopwood JJ, Ballabio A. 2001. Multiple sulfatase deficiency and the nature of the sulfatase family. In *The Metabolic and Molecular Basis of Inherited Disease*, ed. CR Scriver, AL Beaudet, WS Sly, pp.3725–32. New York: McGraw-Hill.
3. Parenti G, Meroni G, Ballabio A. 1997. The sulfatase gene family. *Curr. Opin. Genet. Dev.* 7:386–91.
4. Neufeld EF, Muenzer J. 2001. The mucopolysaccharidoses. In *The Metabolic and Molecular Basis of Inherited Disease*, ed. CR Scriver, AL Beaudet, WS Sly, D Valle, pp. 3421–52. New York: McGraw-Hill.
5. Von Figura K, Gieselmann V, Jaeken J. 2001. Metachromatic leukodystrophy. In *The Metabolic and Molecular Basis of Inherited Disease*, ed. CR Scriver, AL Beaudet, WS Sly, D Valle, pp. 3695–724. New York: McGraw-Hill.
6. Dhoot GK, Gustafsson MK, Ai X, Sun W, Standiford DM, Emerson CP Jr. 2001. Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase. *Science* 293:1663–66.
7. Morimoto-Tomita M, Uchimura K, Werb Z, Hemmerich S, Rosen SD. 2002. Cloning and characterization of two extracellular heparin-degrading endosulfatase in mice and humans. *J. Biol. Chem.* 277:49175–85.
8. Ohto T, Uchida H, Yamazaki H, Keino-Masu K, Matsui A, Masu M. 2002. Identification of a novel nonlysosomal sulphatase expressed in the floor plate, choroid plexus and cartilage. *Genes Cells* 7:173–85.
9. Rommerskirch W, von Figura K. 1992. Multiple sulfatase deficiency: catalytically inactive sulfatases are expressed from retrovirally introduced sulfatase cDNAs. *Proc. Natl. Acad. Sci. USA* 89: 2561–65.
10. Schmidt B, Selmer T, Ingendoh A, von Figura K. 1995. A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* 82:271–78.
11. Miech C, Dierks T, Selmer T, von Figura K, Schmidt B. 1998. Arylsulfatase from *Klebsiella pneumoniae* carries a formylglycine generated from a serine. *J. Biol. Chem.* 273:4835–37
12. Dierks T, Schmidt B, Borissenko LV, Peng J, Preusser A, et al. 2003. Multiple sulfatase deficiency is caused by mutations in the gene encoding the human c(alpha)-formylglycine generating enzyme. *Cell* 113:435–44.

13. Dierks T, Lecca MR, Schlotterhose P, Schmidt B, von Figura K. 1999. Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. *EMBO J.* 18:2084–91.
14. Recksiek M, Selmer T, Dierks T, Schmidt B, von Figura K. 1998. Sulfatases, trapping of the sulfated enzyme intermediate by substituting the active site formylglycine. *J. Biol. Chem.* 273:6096–103.
15. Waldow A, Schmidt B, Dierks T, von Bulow R, von Figura K. 1999. Amino acid residues forming the active site of arylsulfatase A. Role in catalytic activity and substrate binding. *J. Biol. Chem.* 274:12284–88.
16. Cosma MP, Pepe S, Annunziata I, Newbold RF, Grompe M, et al. 2003. The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell* 113:445–56.
17. Landgrebe J, Dierks T, Schmidt B, von Figura K. 2003. The human SUMF1 gene required for posttranslational sulfatase modification, defines a new gene family which is conserved from pro- to eukaryotes. *Gene* 316:47–5.
18. Preusser-Kunze A, Mariappan M, Schmidt B, Gande SL, Mutenda K, Wenzel D, von Figura K, Dierks T (2005) Molecular characterization of the human Calpha-formylglycine-generating enzyme. *J Biol Chem* 280: 14900–14910.
19. Zito E, Fraldi A, Pepe S, Annunziata I, Kobinger G, Di Natale P, Ballabio A, Cosma MP (2005) Sulphatase activities are regulated by the interaction of sulphatase-modifying factor 1 with SUMF2. *EMBO Rep* 6: 655–660.
20. Dierks, T., Schmidt, B. and von Figura, K. (1997) Conversion of cysteine to formylglycine: a protein modification in the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA*, 94, 11963–11968.
21. Zito, E., Buono, M., Pepe, S., Settembre, C., Annunziata, I., Surace, E.M., Dierks, T., Monti, M., Cozzolino, M., Pucci, P. et al. (2007) Sulfatase modifying factor 1 trafficking through the cells: from endoplasmic reticulum to the endoplasmic reticulum. *Embo. J.*, 26, 2443–2453.
22. Tsai, B., Rodighiero, C., Lencer, W.I. and Rapoport, T.A. (2001) Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin. *Cell*, 104, 937–948.
23. Tian, G., Xiang, S., Noiva, R., Lennarz, W.J. and Schindelin, H. (2006). The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell*, 124, 61–73.

24. Ellgaard, L. and Ruddock, L.W. (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep.*, 6, 28–32.
25. Ellgaard, L. and Helenius, A. (2003) Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell. Biol.*, 4, 181–191.
26. Otsu, M., Bertoli, G., Fagioli, C., Guerini-Rocco, E., Nerini-Molteni, S., Ruffato, E. and Sitia, R. (2006) Dynamic retention of Ero1alpha and Ero1beta in the endoplasmic reticulum by interactions with PDI and ERp44. *Antioxid. Redox. Signal.*, 8, 274–282.
27. Appenzeller, C., Andersson, H., Kappeler, F. and Hauri, H.P. (1999) The lectin ERGIC-53 is a cargo transport receptor for glycoproteins. *Nat. Cell. Biol.*, 1, 330–334.
28. Appenzeller-Herzog, C., Roche, A.C., Nufer, O. and Hauri, H.P. (2004) pH-induced conversion of the transport lectin ERGIC-53 triggers glycoprotein release. *J. Biol. Chem.*, 279, 12943–12950.
29. Anelli, T., Ceppi, S., Bergamelli, L., Cortini, M., Masciarelli, S., Valetti, C. and Sitia, R. (2007) Sequential steps and checkpoints in the early exocytic compartment during secretory IgM biogenesis. *Embo. J.*, 26, 4177–4188.
30. Gilchrist, A., Au, C.E., Hiding, J., Bell, A.W., Fernandez-Rodriguez, J., Lesimple, S., Nagaya, H., Roy, L., Gosline, S.J., Hallett, M. et al. (2006). Quantitative proteomics analysis of the secretory pathway. *Cell*, 127, 1265–1281.
31. Wang, Z.V., Schraw, T.D., Kim, J.Y., Khan, T., Rajala, M.W., Follenzi, A. and Scherer, P.E. (2007) Secretion of the adipocyte-specific secretory protein adiponectin critically depends on thiol-mediated protein retention. *Mol. Cell. Biol.*, 27, 3716–3731.
32. Anelli, T., Alessio, M., Mezghrani, A., Simmen, T., Talamo, F., Bachi, A. and Sitia, R. (2002) ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. *Embo. J.*, 21, 835–844.
33. Anelli, T., Alessio, M., Bachi, A., Bergamelli, L., Bertoli, G., Camerini, S., Mezghrani, A., Ruffato, E., Simmen, T. and Sitia, R. (2003) Thiol-mediated protein retention in the endoplasmic reticulum: the role of ERp44. *Embo. J.*, 22, 5015–5022.
34. Higo, T., Hattori, M., Nakamura, T., Natsume, T., Michikawa, T. and Mikoshiba, K. (2005) Subtype-specific and ER lumenal environment-dependent regulation of inositol 1,4,5-trisphosphate receptor type 1 by ERp44. *Cell*, 120, 85–98.
35. Sandvig K, van Deurs B (2000) Entry of ricin and Shiga toxin into cells: molecular mechanisms and medical perspectives. *EMBO J* 19: 5943–5950.
36. Helenius A (1994) How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol Biol Cell* 5: 253–265.

37. Dierks T, Dickmanns A, Preusser-Kunze A, Schmidt B, Mariappan M, von Figura K, Ficner R, Rudolph MG (2005) Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme. *Cell* 121: 541–552.
38. Girod A, Storrie B, Simpson JC, Johannes L, Goud B, Roberts LM, Lord JM, Nilsson T, Pepperkok R (1999) Evidence for a COP-I-independent transport route from the Golgi complex to the endoplasmic reticulum. *Nat Cell Biol* 1: 423–430.
39. Wiseman, R.L., Powers, E.T., Buxbaum, J.N., Kelly, J.W. and Balch, W.E. (2007). An adaptable standard for protein export from the endoplasmic reticulum. *Cell*, 131, 809–821.
40. Mariappan, M., Radhakrishnan, K., Dierks, T., Schmidt, B. and von Figura, K. (2008). ER-p44 mediates a thiol independent retention of formylglycine generating enzyme in the endoplasmic reticulum. *J. Biol. Chem.*, 283, 6375–6383.
41. Dierks T, Dickmanns A, Preusser-Kunze A, Schmidt B, Mariappan M, von Figura K, Ficner R, Rudolph MG (2005) Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme. *Cell* 121: 541–552.

Capitolo III

Studio dell'interattoma del fattore trascrizionale CBX7

3.1 Introduzione

3.1.1 CBX7 e la carcinogenesi tiroidea

La tiroide è la più grande ghiandola endocrina umana [1]. E' situata nella regione del collo, sulla superficie anteriore della trachea, ed è costituita da due differenti tipi di cellule, le follicolari (TFC) e le parafollicolari o cellule C [2]. Più del 95% dei carcinomi tiroidei deriva dalle cellule follicolari [1] e sono le più comuni neoplasie endocrine, con circa 2500 casi stimati ogni anno negli Stati Uniti. Di contro, solo il 3% dei tumori tiroidei, definiti carcinomi tiroidei midollari, è di origine parafollicolare. I tumori tiroidei derivano dall'accumulo di diverse modificazioni nei geni critici coinvolti nel controllo della proliferazione cellulare. Le neoplasie tiroidee rappresentano un buon modello per lo studio degli eventi multistep coinvolti nella carcinogenesi, poiché esse comprendono un ampio spettro di lesioni con differenti gradi di malignità, che sono diagnosticate sulla base di parametri istologici e clinici.

I tumori tiroidei derivanti da cellule follicolari includono: 1) adenomi benigni, che non sono invasivi e molto bene differenziati; 2) carcinomi, che sono divisi in ben differenziati, poco differenziati (PDTC) e non differenziati (ATC). I carcinomi tiroidei ben differenziati sono papillari (PTC) che rappresentano più del 70% dei tumori tiroidei maligni e follicolari (FTC) che rappresentano il 10% dei carcinomi tiroidei. Questi tumori, essendo differenziati, hanno una buona prognosi [1,3]. PDTC e ATC sembrano derivare dalla progressione di carcinomi differenziati [4]. Sebbene gli ATC rappresentano il 2-5% dei tumori tiroidei maligni, essi sono tra le neoplasie umane più letali che crescono rapidamente, sono molto aggressivi e sempre fatali. I PDTC invece rappresentano il 7% dei tumori tiroidei, sono forme intermedie tra i ben differenziati e i non differenziati, sia morfologicamente che dal punto di vista comportamentale. Sebbene sono seguiti vari approcci terapeutici nella pratica clinica, la maggior parte di essi sono fatali. Quindi, la scoperta di nuovi approcci per la diagnosi del cancro a stadi precoci e per stabilire terapie migliori è un problema critico e urgente. Per arrivare a questo scopo, l'identificazione e la caratterizzazione di molecole chiave che partecipano nella carcinogenesi sono steps essenziali. Studi di microarray sono ampiamente usati per definire situazioni diagnostiche e prognostiche nel cancro e hanno condotto all'identificazione di una lunga lista di geni regolati nei carcinomi [5]. Anche per identificare i geni coinvolti nel processo di carcinogenesi tiroideo è stato usato questo approccio, che ha condotto all'identificazione di geni la cui espressione è up- o down- regolata in linee cellulari di carcinomi paragonata con le culture cellulari pri-

marie di origine tiroidea normale [6]. Tra i geni down-regolati è stato identificato CBX7. Questa proteina appartiene alla famiglia chromobox, codifica per una nuova proteina polycomb e la sua espressione mostra una riduzione progressiva, ben relazionata con il grado di malignità delle neoplasie tiroidee. Inoltre, i livelli della proteina diminuiscono in una percentuale crescente di casi andando dagli adenomi benigni ai PTC, FTC e gli ATC. La correlazione tra la bassa espressione di CBX7 e gli istotipi più aggressivi sembra poter essere applicata anche ai carcinomi al seno, ovaie e intestino e sembra ridurre la sopravvivenza nel carcinoma al colon [7].

3.1.2 CBX7 e i geni Polycomb

Il gene *cbx7* (Chromobox homolog 7) è down-regolato nei carcinomi tiroidei e codifica per la proteina CBX7 che fa parte del complesso Polycomb [8].

I geni del gruppo Polycomb (PcG) collegano le modifiche degli istoni con la repressione trascrizionale [9,10]. Le proteine PcG sono parte di due complessi ad alto peso molecolare chiamati Polycomb repressive complexes 1 e 2, PRC1 e PRC2 (Figura 3.1).

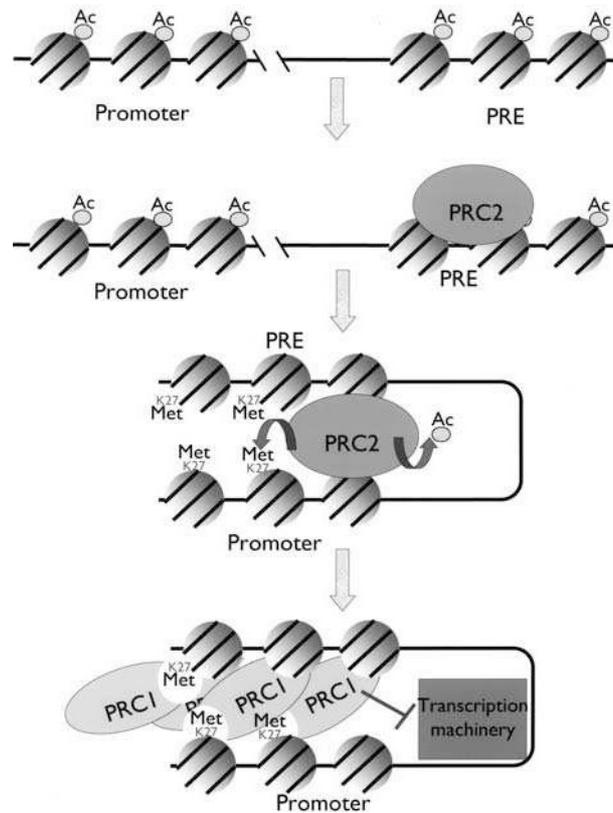


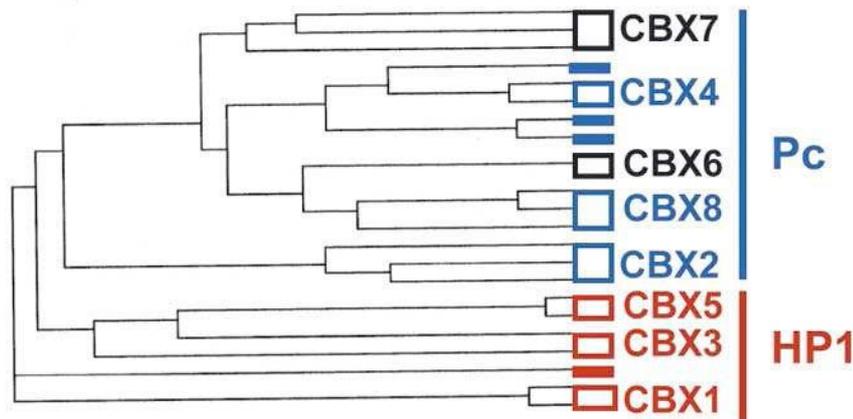
Figura 3.1. Meccanismo di repressione trascrizionale operato dalle proteine del gruppo Polycomb.

Il legame del Polycomb repressive complex 2 (PRC2) al Polycomb responsive elements (PRE) attiva la deacetilazione dell'istone e la metilazione sulla Lisina 27 dell'istone H3, entrambi nella regione PRE e del promotore dei geni silenziati. Questa modifica costituisce uno dei segnali per il reclutamento del Polycomb repressive complex 1 (PRC1), o complesso di mantenimento, che reprime l'espressione genica mediante l'inibizione dell'inizio della trascrizione.

Il complesso PRC1 riconosce e lega la lisina trimetilata in posizione 27 dell'istone H3 mediante le proteine Polycomb e contribuisce alla repressione genica inibendo l'inizio della trascrizione [11]. I membri del *core* di questo complesso nei mammiferi sono gli omologhi di *Drosophila* delle proteine Polycomb (Pc), Posterior sex combs (Psc), Sex combs extra (Sce) e Polyhomeiotic (Ph). Il complesso PRC2, invece, è associato alle attività istone deacetilasi e istone metiltransferasi (specifica per la lisina 27 dell'istone H3) [12,13]. Tra i membri di questo secondo complesso ci sono Enhancer of Zeste (EZH2), Early embryonic deficient (EED), Suppressor of Zeste (SUZ12) e altre proteine associate. E' ben noto che la deregolazione delle proteine PcG è coinvolta nella tumorigenesi. Molti studi infatti dimostrano che, ad esempio, Bmi-1 risulta amplificato nei tumori del sangue e nei medulloblastomi [14-17]; EZH2 è coinvolto nei tumori del sangue [14-18], e la sua espressione risulta incrementata nelle metastasi dei tumori alla prostata ed al seno [19,20]; alterazioni di Rae28/Ph1 sono osservate nella leucemia linfoblastica acuta [21] e SUZ12 risulta up-regolata nei

tumori al colon [22]. Evidenze recenti mostrano che i geni PcG sono una nuova classe di oncogeni e antioncogeni, che potrebbero in futuro diventare centrali per lo sviluppo di nuove terapie del cancro basate su silenziamento genico [23,24]. I geni PcG come Bmi-1 ed EZH2 sono capaci di trasformazione cellulare, ma sono anche vitali per la sopravvivenza cellulare. Ad oggi, l'espressione anomala dei geni PcG è stata descritta nella maggior parte dei tumori umani. Inoltre, la correlazione tra l'espressione dei geni PcG e il comportamento biologico di sottotipi clinicamente definiti di cancro suggerisce che questi geni svolgono un ruolo centrale nell'oncogenesi, e possono essere presi in considerazione per la scoperta di nuovi marcatori diagnostici [25].

CBX7 e le altre proteine PcG contengono un chromodominio all'estremità N-terminale (CBX7 tra gli aminoacidi 10 e 46). Questo dominio fu originariamente identificato in *Drosophila melanogaster* come una regione di 37 aminoacidi omologa ad heterochromatin protein 1 (HP1) and Polycomb (Pc) proteins [26]. Analisi filogenetiche e di sequenza del chromodominio (Figura 3.2) hanno rivelato che CBX7 mostra una grande similitudine ad altre proteine note o putative del gruppo Pc come CBX2 (Pc1), CBX4 (Pc2), CBX6 and CBX8 (Pc3). Questa similitudine è meno pronunciata rispetto alle proteine HP1, come CBX1 (HP1 β), CBX3 (HP1 γ), CBX5 (HP1 α) e, inoltre, CBX7 non contiene il “*chromo-shadow domain*” che è una caratteristica delle proteine HP1. CBX7 contiene anche altri specifici residui molto conservati che permettono di distinguere i chromodomini Pc e HP1: questi residui specifici delle proteine Pc sono necessari sia per la dimerizzazione delle proteine Pc che per il riconoscimento della Lys 27 trimetilata sull'istone H3 [27]. Oltre ad avere un'alta omologia entro il chromodominio, CBX7 è anche omologa alle proteine Pc nella regione carbossi-terminale (Figura 4C) definito Pc-box o C-box [28]. E' noto che il Pc box (aminoacidi 231-243) e il chromodominio sono entrambi necessari per la funzione di CBX7. E' stato visto che la proteina di fusione tra CBX7 e il DNA-binding domain di Gal4 è capace di reprimere, in modo dose-dipendente, la trascrizione del gene reporter 4xGal4-tk-luc, suggerendo che CBX7 funziona come repressore della trascrizione [8]. Così, CBX7 sembra agire come altre proteine Pc esercitando il suo effetto mediante la repressione trascrizionale.

A**B**

hcbx7	9	QVFAVESIRKKRVRKGGKVEYLVKWKGWPPKYSTWEPEEHI	48
hcbx4	10	-VFAVESIEKKRIRKGRVEYLVKWRGWSPKYNTWEPEENI	48
hcbx6	9	RVFAAESIIKRRIRKGRIEYLVKWKGWAIKYSTWEPEENI	48
hcbx8	9	RVFAAEALLKRRIRKGRMEYLVKWKGWSQKYSTWEPEENI	48
hcbx2	10	QVFAAECILSKRLRKGKLEYLVKWRGWSSKHNSWEPEENI	49

C

hCBX8	364	TDVTSNFLTVTIKE	377
hCBX6	371	TDVTSNLLTVTIKE	384
hCBX2	269	TDVTANLITVTVKE	292
hCBX4	541	TDVTANCLTVTFKE	554
hCBX7	231	TDITANSITVTFRE	244

Figura 3.2 CBX7 e chromodominio

Identificazione di CBX7 come proteina appartenente alla famiglia Polycomb. (A) Analisi filogenetica delle proteine Chromobox (CBX). Le linee raggruppate mediante rettangoli aperti denotano ortologhi della stessa proteina CBX (nome a destra). I rettangoli pieni si riferiscono a proteine che non possono essere considerate puramente ortologhi di una qualsiasi delle proteine CBX. Le proteine Pc sono mostrate in blu, mentre le proteine HP1 in nero. (B) Allineamento del chromodominio di CBX7 con quello di altre Polycomb umane note e putative. Residui identici sono mostrati in giallo, mentre residui simili in azzurro. (C) Allineamento della porzione Polycomb di CBX7 con quello di altre Polycomb umane note e putative. I colori sono descritti in (B).

CBX7 è anche capace di interagire con vari membri del gruppo Pc. Infatti, come altre proteine Pc, CBX7 è capace di interagire con se stessa e con la proteina ring-finger Ring1 [8], come riportato da esperimenti Glutathione S-transferase (GST) pull-down [29,30]. Inoltre, nel nucleo di diverse linee cellulari, CBX7 localizza con Ring1 in strutture distinte chiamate Pc-bodies [31], come dimostrato

da studi di immunofluorescenza. In topo è stato dimostrato che il chromodominio di CBX7 è essenziale per il legame a H3K9me3 e H3K27me3, in alcuni casi associato con lo stato repressivo. E' stato anche trovato che, *in vitro*, il cromodominio di CBX7 può legare RNA e che, *in vivo*, l'interazione di CBX7 con la cromatina, in particolare con il cromosoma X inattivo, parzialmente dipende dalla sua associazione con l'RNA [32]. CBX7 è altamente espresso in parecchi tessuti, quali il cervello, cuore, muscoli scheletrici e reni [8]. Uno studio recente ha dimostrato che il gene *cbx7* è drasticamente ridotto in sei linee cellulari di carcinoma tiroideo rispetto a normali cellule tiroidee. Conseguentemente, l'analisi dell'espressione di *cbx7* in un grande numero di campioni di carcinomi tiroidei rivela una progressiva riduzione dei livelli di CBX7 che risultavano ben relazionati con la malignità della neoplasia tiroidea [7]. Inoltre l'espressione di CBX7 decresce in una crescente percentuale di casi andando dagli adenomi benigni ai carcinomi tiroidei papillari, follicolari e anaplastici [33]. Risultati più recenti confermano una correlazione tra la bassa espressione di CBX7 e una ridotta sopravvivenza nei pazienti con carcinoma al colon. Inoltre, l'associazione tra la perdita dell'espressione di CBX7 e un istotipo più aggressivo è stata anche mostrata nel carcinoma al seno, ovaie e prostata. Il ripristino dell'espressione di CBX7 nelle cellule cancerose tiroidee riduce la loro velocità di crescita, indicando che CBX7 gioca un ruolo critico nella regolazione della proliferazione delle cellule tiroidee trasformate [7].

3.1.3 Pre-frazionamento di complessi proteici

Un singolo esperimento di proteomica funzionale in genere fornisce una lunga lista di putativi interattori, il che indica chiaramente la presenza della proteina esca in diversi complessi. Come in altre applicazioni di proteomica, il compito più impegnativo e che richiede maggior tempo in questi esperimenti è l'interpretazione dei dati, e in particolare il tentativo di assegnare tutti i componenti proteici allo specifico complesso cui appartengono. Nonostante la ricchezza delle informazioni fornite da banche dati di proteine e database di interazioni proteina-proteina e la disponibilità di strumenti software, la deduzione di un significato biologico dalla lista di interattori rimane ancora un problema sfuggente. Una possibile soluzione per superare queste difficoltà potrebbe essere l'introduzione di modifiche alle metodologie classiche di purificazione dei complessi proteici. Il pre-frazionamento di un estratto totale di proteine con metodi in grado di preservare le interazioni native tra le proteine potrebbe essere utile per isolare i singoli complessi funzionali riducendo così notevolmente la complessità dei dati di proteomica funzionale. La pre-purificazione dei singoli complessi in condizioni native potrebbe essere realizzata sulla base delle loro caratteristiche chimico-fisiche, per esempio le dimensioni e la grandezza (gel filtrazione) [34-36], la densità (gradiente ultracentrifugazione) [37-39], la mobilità elettroforetica (Blue Native) [40-42]. Questi metodi non possono includere la cromatografia a fase inversa o altre procedure condotte in condizioni di denatura-

zione per evitare la dissociazione dei complessi. A seguito della pre-purificazione del campione, i vari complessi contenenti l'esca sono separati in diverse frazioni e la presenza dell'esca all'interno di queste frazioni può essere semplicemente rilevata da analisi con western blot. Le frazioni positive all'immunorivelazione sono selezionate, mentre le altre vengono scartate, ed i complessi sono poi singolarmente isolati utilizzando diverse tecniche, quali l'immunoprecipitazione o l'affinity [34]. Diminuire la complessità di un estratto cellulare con procedure di pre-frazionamento comporta una notevole diminuzione delle identificazioni di putativi interattori per ogni frazione. Inoltre, tutti i componenti trovati nella stessa frazione verosimilmente appartengono allo stesso complesso funzionale, rendendo l'interpretazione dei dati molto più semplice. Infine, la rimozione di una quantità enorme di proteine non correlate con l'esca influisce positivamente sul range dinamico del metodo portando ad una forte diminuzione del background ed in generale ad una maggiore fiducia nelle identificazioni. La presenza di falsi positivi è un problema così critico negli esperimenti di proteomica funzionale che pochi anni fa Matthias Mann e collaboratori hanno sviluppato una apposita strategia per discriminare interattori specifici e non, isolati durante un esperimento di proteomica funzionale. Questa strategia, definita QUICK (quantitative IP combined with knockdown) [43] combina SILAC (stable-isotope labelling amino acids) con IP e RNA interference (RNAi).

3.1.3.1 Gradiente di ultracentrifugazione

Il gradiente di ultracentrifugazione con saccarosio, glicerolo, CsCl, etc ha trovato la sua principale applicazione nella separazione di organelli sub-cellulari [44, 45]. Vogelmann et al. separarono compartimenti della membrana usando un gradiente di densità lineare iodixanolo 10-20-30 % [38]. Sono stati segnalati finora solo alcuni esempi di applicazione di questi esperimenti per la separazione complessi proteici [37, 38]. Nella procedura standard l'estratto proteico è caricato su un gradiente di adeguata densità e sottoposto a ultracentrifugazione per un tempo fisso. Le frazioni recuperate sono poi analizzate per individuare la presenza della proteina esca di solito mediante western blot con un anticorpo specifico per la proteina di interesse. Tuttavia, è stato anche riportato l'uso di attività enzimatica o specifici saggi biologici per l'identificazione della proteina esca [38]. Va rilevato che interattori più deboli potrebbero dissociare dal complesso durante lo step di pre-frazionamento; d'altra parte, questi interattori dissociano anche durante le condizioni stringenti di lavaggio necessarie a eliminare gli interattori non specifici in assenza di prefrazionamento.

3.1.3.2. Cromatografia ad esclusione molecolare

Recentemente, Holaska e Wilson hanno segnalato l'isolamento di diversi complessi coinvolgenti la proteina emerin dai nuclei di cellule HeLa in tre successive fasi cromatografiche [46]. I complessi contenenti emerin sono stati prima arricchiti con immunoprecipitazione utilizzando anticorpi anti-emerin e poi frazionati mediante cromatografia a scambio ionico ed esclusione molecolare. Oltre un certo numero di complessi contenenti emerin, quest'ultima analisi ha suggerito la presenza di almeno altri due complessi: uno ad alto peso molecolare che comprendeva emerin, NMI, actina e RII-spettrina, e un complesso di piccole dimensioni contenenti emerin, NMI, e actina [47]. In quest'approccio l'immunoprecipitazione è stata effettuata a monte del frazionamento cromatografico dei complessi. Ciò è stato possibile perché i complessi contenenti emerin sono stati eluiti dalla resina di immunoaffinità per competizione con la proteina ricombinante. Questa procedura di eluizione ha permesso di recuperare i complessi proteici in condizioni quasi-native, mantenendo stabile l'interazione proteina-proteina. L'identificazione delle componenti proteiche all'interno dei complessi di emerin ha suggerito l'esistenza di un nucleo centrale del complesso, comune a più complessi ai quali sono state associate proteine discriminanti.

In un approccio leggermente diverso, la cromatografia ad esclusione molecolare può essere utilizzata come procedura di purificazione preliminare per frazionare gli estratti proteici totali prima della fase di immunoprecipitazione [34, 35]. Le frazioni contenenti la proteina di interesse possono essere ulteriormente sottoposte a immunoprecipitazione per isolare specifici complessi contenenti l'esca [34]. L'isolamento di complessi proteici con la procedura cromatografica ha molti vantaggi in confronto con l'ultracentrifugazione. Il caricamento dei campioni, il frazionamento dei complessi e il recupero delle frazioni sono abbastanza facili e richiede meno tempo rispetto alla procedura di ultracentrifugazione, garantendo una più elevata conservazione delle interazioni native proteina-proteina. Inoltre, la gel filtrazione ha mostrato una più alta risoluzione nella separazione di complessi nativi e ciò consente l'individuazione di complessi a basso peso molecolare, non osservabili nell'esperimento di ultracentrifugazione molto probabilmente perché vengono esclusi dal gradiente di glicero.

3.2. Materiali e Metodi

3.2.1. Materiali

I prodotti abcam utilizzati sono:

la resina coniugata all'anticorpo Anti-v5 e il peptide v5 così come l'anticorpo primario Anti-v5 usato per il Western Blot.

Tutti gli altri prodotti della SIGMA, USB, Roche utilizzati sono quelli riportati nella sezione 2.2.

3.2.2 Metodi

3.2.2.1 Colture Cellulari

Le cellule utilizzate negli esperimenti eseguiti durante questo lavoro di tesi sono state fornite dal gruppo del prof. Alfredo Fusco, con il quale si è collaborato per lo sviluppo del progetto.

Le cellule utilizzate sono cellule HEK 293 trasfettate in maniera transiente col costrutto di espressione del gene CBX7 recante all'estremità amminotermale il peptide V5 (necessario per purificare i complessi dei partners molecolari di CBX7).

3.2.2.2 Estrazione delle proteine nucleari

Per ottenere l'estratto nucleare (NE), le cellule furono lavate due volte con un tampone fosfato e risospese in tre volumi di una soluzione (A) contenente 10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl₂, 0.1 mM EGTA, 0.5 mM DTT (soluzione di omogeneizzazione). Le cellule furono distrutte mediante il passaggio attraverso un ago. I nuclei sono stati raccolti mediante centrifugazione a 1500 rpm e risospesi in 1,2 volumi di soluzione di estrazione (B) contenente 10 mM HEPES pH 7,9, 0,4 M di NaCl, 1,5 mM MgCl₂, 0.1 mM EGTA, 0,5 mM DTT, 5% glicerolo, consentendo l'eluizione delle proteine nucleari mediante delicata agitazione a 4 ° C. I nuclei sono stati nuovamente pellettati per centrifugazione a 12 000 rpm e il surnatante è stato recuperato e filtrato per essere . Gli inibitori di proteasi leupeptina (5 mM), aprotinina (1,5 mM), phenylethylsulfonylfluoride (2mM), pepstatin A (3 mM), e benzamidine (1mM) sono stati aggiunti sia durante l'omogeneizzazione che nelle soluzioni di estrazione. Di tale estratto è stata determinata la concentrazione proteica mediante saggio di Bradford, utilizzando il reattivo "Bio Rad Protein Assay". Come proteina di riferimento per costruire la retta di taratura è stata utilizzata l'Albumina di siero bovino.

3.2.2.3 Isolamento di complessi proteici mediante tecniche d'immunoprecipitazione

Per isolare i complessi proteici che coinvolgono CBX7 è stato eseguito un esperimento d'immunoprecipitazione, sfruttando la presenza, all'estremità amminoterminale della proteina esca, del peptide v5, per il quale sono disponibili commercialmente anticorpi che permettono di purificare i complessi.

- ***Pre-cleaning***

È stato prelevato 1 mL di *Anti-Mouse Agarose Beads* (500 µL di resina secca) per ogni 40 mL di estratto proteico. La resina coniugata con gli anticorpi è stata lavata in modo da eliminare la soluzione di glicerolo. I lavaggi sono stati effettuati aggiungendo una quantità di tampone di lavaggio B (10 mM HEPES pH 7,9, 0,4 M di NaCl, 1,5 mM MgCl₂, 0.1 mM EGTA, 5% glicerolo,) pari a quella di resina per almeno cinque volte, lasciando la sospensione in ghiaccio per alcuni minuti ed infine centrifugando per 2 minuti a 3000 rpm. 1 mg di estratto nucleare è stato poi incubato con la resina precedentemente lavata. L'incubazione è stata realizzata a 4°C su ruota per due ore.

- ***Lavaggi beads Anti-v5***

Si sono eseguiti 6 lavaggi delle beads Anti-v5 con il buffer di lisi B per allontanare la soluzione in cui sono di norma conservate, centrifugando ogni volta per 2 minuti a 3000 rpm.

- ***Immunoprecipitazione***

L'estratto è stato poi centrifugato per 5 minuti a 2000 rpm ed è stato recuperato il sopranatante che è stato incubato con la resina anti-v5, lavata precedentemente per sei volte con il tampone di lavaggio. In particolare, la quantità di resina che è stata usata è 50 µL (25 µL di resina secca) per ogni mg di estratto proteico. La determinazione della concentrazione proteica, necessaria per stabilire la quantità di resina da usare, è stata eseguita mediante un saggio Bradford. L'incubazione è stata eseguita a 4° C su ruota over night. Dopo l'incubazione con la resina M2 anti-FLAG, gli estratti sono stati centrifugati per 5 minuti a 2000 rpm, il sopranatante è stato scartato, mentre la resina è stata lavata quattro volte con il tampone B.

- ***Eluizione***

L'eluizione dei complessi legati all'anticorpo presente sulla resina è avvenuta per competizione, grazie alla presenza nel tampone di eluizione del peptide v5. La resina è stata quindi trasferita nelle colonnine di reazione che sono state centrifugate fino a quando non è stato eliminato tutto il tampo-

ne di lavaggio. In ogni camera di reazione è stato aggiunto un volume di tampone di eluizione (10 mM HEPES pH 7,9, 0,4 M di NaCl, 1,5 mM MgCl₂, 0.1 mM EGTA, 5% glicerolo, 5 mM leupeptina, 1,5 mM aprotinina, 2mM phenylethylsulfonilfluoride, 3 mM peptastatin A, 1mM benzamidine, peptide v5 200 µg ml⁻¹) uguale al volume di resina. L'incubazione è stata effettuata a 4° C su ruota per tutta la notte. Per recuperare tutto l'eluato contenente i complessi coinvolgenti CBX7 coniugata con il peptide v5 si è quindi proceduto centrifugando a 3000 rpm per 2 minuti. L'eluato così ottenuto è stato trattato con metanolo-cloroformio allo scopo di precipitare le proteine presenti come descritto in 2.2.2.4. Il campione è stato analizzato seguendo esattamente le procedure descritte nella sezione di materiali e metodi precedente.

3.2.2.4 Prefrazione dei complessi proteici

3.2.2.4.1 Prepurificazione dei complessi mediante Gel filtration

Al fine di eseguire una separazione dei complessi proteici contenenti la proteina di interesse, è stata effettuata una cromatografia di esclusione molecolare.

L'apparato strumentale utilizzato è un FPLC AKTA della GE Healthcare su cui è stata installata una colonna per cromatografia ad esclusione molecolare. La FPLC è un sistema cromatografico a media pressione che consente di purificare proteine e peptidi più rapidamente e con maggiore efficienza rispetto ai sistemi cromatografici tradizionali.

Il tampone utilizzato durante la separazione è: 0,05M NaH₂PO₄ + 0,15M NaCl ad un pH = 7.2.

La prima fase di equilibratura della colonna è stata effettuata in tampone da corsa ad un flusso di 0,3 mL/min, finché l'assorbimento a 220 nm non è arrivato ad un valore costante. In questa fase bisogna controllare che il pH resti uguale a 7 e bisogna tenere sotto controllo la pressione, che non deve superare i 70mPa. L'intera separazione cromatografica è stata eseguita in camera fredda a 4°C per preservare l'integrità del campione. Una volta equilibrata la colonna sono stati iniettati 3,6 mg di campione proteico e l'eluizione è stata eseguita a 0,5mL/min.

Sono raccolte mediante un collettore automatico 26 frazioni da 0,5mL.

3.2.2.4.2 Prepurificazione dei complessi mediante Centrifugazione in gradiente di densità

Un gradiente discontinuo di glicerolo 5-45% è stato preparato stratificando in un tubo apposito, dal basso verso l'alto le soluzioni a densità decrescente di glicerolo. I tubi utilizzati per creare il gradiente sono di poliallometere (Beckman), e le soluzioni a densità differente sono state preparate separatamente, aggiungendo prima la quantità di acqua e buffer di lisi contenete inibitori di proteasi, e infine glicerolo, secondo le proporzioni riportate in Tabella 1:

Tabella 1

Frazione	Acqua	Buffer Lisi	Glicerolo
5%	120 μ L	660 μ L	420 μ L
10%	120 μ L	600 μ L	480 μ L
15%	120 μ L	540 μ L	540 μ L
20%	120 μ L	480 μ L	600 μ L
25%	120 μ L	420 μ L	660 μ L
30%	120 μ L	360 μ L	720 μ L
35%	120 μ L	300 μ L	780 μ L
40%	120 μ L	240 μ L	840 μ L
45%	120 μ L	180 μ L	900 μ L

In maniera decrescente, quindi dalla soluzione a % maggiore a quella a % minore, 1mL di ciascuna soluzione preparata è stata stratificata nel tubo apposito, alloggiato in ghiaccio in modo da non miscelare i vari. Il tutto è stato lasciato per 15 minuti in ghiaccio per renderlo continuo ed in fine è stato caricato il campione sulla sommità del gradiente. Il tubo poi è stato adagiato nell'alloggio specifico del rotore dell'ultracentrifuga. Allo stesso modo è stato preparato un gradiente anche per il controllo. Siccome il rotore dell'ultracentrifuga deve essere bilanciato, i due tubi sono stati bilanciati per pesata con tampone di lisi.

L'ultracentrifuga impiegata è stata Opima™ L-80XP della Beckman ed i parametri impostati per l'esperimento sono stati: tempo = 16 h; velocità = 39000rpm ; temperatura = 4°C.

Terminate le 16 h, le frazioni sono state prelevate dai tubi procedendo dall'alto verso il basso (gradiente crescente), tenendo conto che la prima frazione è pari al volume del campione caricato, le altre sono di 1 mL, mentre l'ultima che è sul fondo del tubo, rappresenta tutto ciò che "esce fuori gradiente", ovvero non sedimenta nella frazione a più alta % di glicerolo.

3.2.2.5 Identificazione

3.2.2.5.1. Analisi mediante LCMS/MS

Le miscele peptidiche liofilizzate sono state risospese in 10 µl di acido formico (HCOOH) allo 0.2%.

L'analisi LCMS/MS è stata condotta mediante un sistema LC/MSD Trappola XCT Ultra (Agilent Technologies), equipaggiato con un sistema HPLC 1100 ed un chip cube (Agilent Technologies). Dopo il caricamento, la miscela peptidica (8 µl in 0.2% acido formico) è stata prima concentrata e poi desalificata a 4 µl/min su una colonna di arricchimento inversa (chip, Agilent Technologies), con 0.1% acido formico (HCOOH) come eluente. Il campione è stato poi frazionato su una colonna capillare C18 a fase inversa (75 µm, 43 mm) ed un flusso di 200nl/min, con un gradiente lineare di B in A dal 5 al 60% in 50 min.

Il campione è stato trasferito on-line su una colonna capillare C18 a fase inversa (75 µm, 20 cm) e un flusso di 4 µl/min, dove è stato frazionato usando un gradiente lineare di eluente B (0.2% acido formico in 95% acetonitrile) in A (0.2% acido formico, 5% acetonitrile in acqua) dal 7 al 50% in 64 minuti.

L'eluato è stato introdotto nello spettrometro di massa programmato nella modalità MS/MS data-dependent, dove ogni spettro di massa (intervallo 400-2000 Da/e) è stato seguito da uno o più spettri di massa tandem ottenuti a partire dagli ioni parenti più intensi eluiti in ogni frazione cromatografica.

Una rampa di collisioni energetiche tra 30 e 60 eV è applicata in base alla massa e alla carica del precursore ionico. Gli spettri di frammentazione così ottenuti, così ottenuti dalla Trappola, sono stati analizzati con il programma Analysis List. Tale programma genera delle peak list che vengono inserite in banche dati disponibili in rete conducendo all'identificazione della proteina.

3.2.2.5.2. Identificazione mediante Mascot

Le peak list ottenute in seguito all'analisi mediante LCMSMS sono state inserite nella banca dati di Mascot utilizzando l'opzione MS/MS Ion Search.

3.3. Risultati

3.3.1 Identificazione delle proteine che interagiscono con CBX7

Per investigare i meccanismi mediante cui la perdita di espressione di CBX7 correla con i fenotipi tumorali più maligni, sono stati identificati nuovi partners molecolari della proteina CBX7. A tale scopo cellule umane HEK293 sono state trasfettate in maniera transiente sia con un vettore di espressione codificante per la proteina CBX7 coniugata all'epitopo V5, sia con un vettore vuoto come controllo. Per eseguire l'esperimento è stato utilizzato circa 1 mg di estratto nucleare (NE). In particolare, le proteine associate con CBX7 sono state purificate per immunoprecipitazione (IP) usando un anticorpo anti-V5 coniugato covalentemente a una resina di agarosio. Il primo step d'immunoprecipitazione ha previsto l'incubazione dell'estratto con la resina Anti-Mouse per la fase di pre-cleaning. In questa fase gli estratti sono stati incubati con resina derivatizzata con anticorpi aspecifici al fine di impoverire l'estratto proteico di quelle proteine che, avendo affinità per l'agarosio lo legano generando poi nell'esperimento d'immunoprecipitazione, dei "falsi positivi". Gli estratti sono stati incubati con una resina cui sono legati anticorpi di ratto di natura totalmente aspecifica. Per l'esperimento preparativo, la resina Anti-Mouse usata per il pre-cleaning è stata utilizzata in rapporto di 500 µl di resina secca per ogni 40 ml di estratto totale, effettuando un'incubazione di 2 ore su ruota a 4 °C. Trascorso tale tempo il surnatante recuperato è stato incubato con la resina Anti-V5, precedentemente saturate con latte No Fat. Il rapporto ottimale tra anticorpo ed estratto proteico è di 25 µl di resina secca per mg di estratto proteico. Pertanto, l'estratto derivante dalle cellule HEK293 è stato incubato con 25 µl di resina Anti-V5. L'incubazione è stata eseguita over night a 4°C su ruota. Rimosso il surnatante, la resina che ha trattenuto la proteina esca con i suoi complessi è stata sottoposta a una serie di lavaggi in tampone. Si è infine proceduto con la fase di eluizione. Per limitare l'eluizione di proteine aspecifiche, è stata condotta a un'eluizione per competizione con il peptide V5. A tal proposito la resina Anti-V5 è stata incubata con il tampone di eluizione contenente il peptide v5 a una concentrazione pari a 200 ug/ml over night. In tale eccesso il peptide V5 compete con il peptide analogo presente all'estremità amminoterminale della proteina esca nel legame all'anticorpo consentendo pertanto l'eluizione della proteina esca e dei partners ad essa legati. L'eluato così ottenuto è stato precipitato con metanolo-cloroformio (secondo la procedura descritta nei metodi nel cap.2), allo scopo di concentrare le proteine presenti, e il precipitato così ottenuto è stato preparato per l'analisi elettroforetica monodimensionale, risospingendolo direttamente nel tampone di caricamento. Le proteine sono state frazionate su un gel monodimensionale al 12%. Il gel colorato con Comassie Colloidale è mostrato in figura 3.3. Prima di procedere al-

l'esperimento preparativo è stato opportuno confermare l'espressione del costrutto CBX7-v5 nelle cellule HEK 293 e l'efficienza dell'immunoprecipitazione utilizzando la resina su cui è legato covalentemente l'anticorpo anti-v5 e rivelando il risultato mediante analisi Western Blot, utilizzando un anticorpo anti-V5 (Figura 3.3).

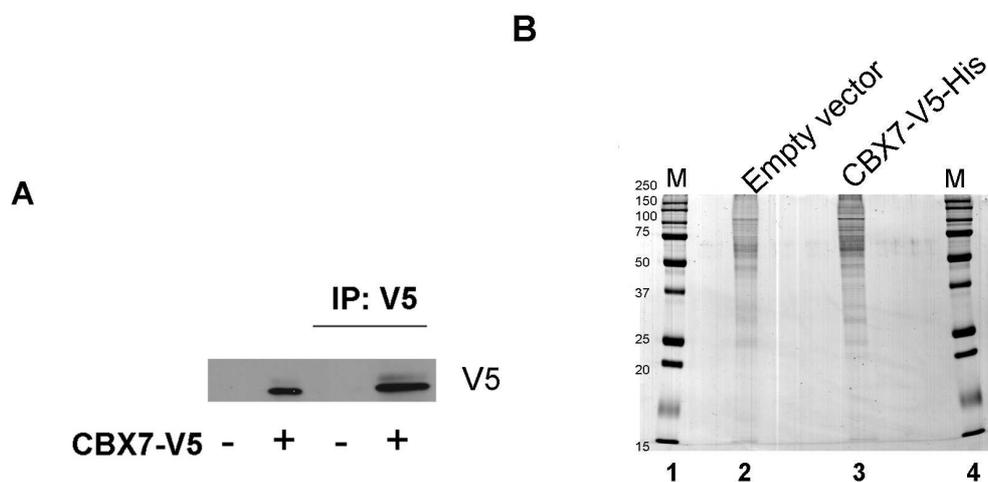


Figura 3.3 Analisi Proteomica. A) Una parte dell'immunoprecipitato del vettore vuoto e del vettore CBX7-V5 è stato usato per verificare l'espressione di CBX7 nelle cellule trasfettate e l'efficienza dell'immunoprecipitazione. B) Gel preparativo colorato con comassire colloidale (M: marker).

Dal confronto tra le lane del controllo (C) e quelle del campione (S) si osservano delle bande discriminanti. Tuttavia un gel monodimensionale non consente di risolvere proteine con un peso molecolare o una migrazione elettroforetica simile. Non è stato quindi possibile affidarsi all'analisi d'immagine per l'identificazione delle bande discriminanti tra le due lane, ma è stato necessario analizzare interamente le lanes di controllo e campione. Tutte le bande sia del campione sia del controllo, sono state escisse, sottoposte ad idrolisi in situ con tripsina e le miscele peptidiche ottenute sono state analizzate mediante LCMSMS. Tale metodologia abbina procedure di cromatografia liquida capillare a fase inversa con analisi mediante spettrometria di massa tandem, riuscendo non solo ad ottenere i valori di massa accurati dei singoli peptidi ma anche, man mano che la cromatografia procede, di isolare gli ioni corrispondenti ai peptidi e di frammentarli. Dagli spettri di frammentazione si ricavano informazioni concernenti le sequenze dei peptidi. Quindi l'identificazione delle proteine in banca dati è condotta utilizzando due tipi di informazioni e risulta più stringente perché basata non solo sull'impiego dei pesi molecolari accurati, ma anche sulle sequenze delle specie peptidiche analizzate. Le liste dei picchi (*peak list*) così ottenute sono state introdotte nel programma Mascot permettendo di effettuare delle ricerche tra le proteine umane depositate nelle banche dati.

Infine, le proteine comuni identificate sia nel campione che nel controllo sono state scartate mentre quelle presenti esclusivamente nel campione sono state considerate putativi interattori di CBX7 e sono listate in Tabella 2.

Tabella 2. Putativi interattori di CBX7

Protein name	Swiss Prot Code
Heat shock protein beta-1 (HSP-27)	P04792
Heterogeneous nuclear ribonucleoprotein A1	P09651
60S acidic ribosomal protein P0	P05388
Nucleolysin TIAR (TIA-1-related protein)	Q01085
Poly(rC)-binding protein 2 (PCBP2)	Q15366
Heterogeneous nuclear ribonucleoprotein D0	Q14103
Eukaryotic translation initiation factor 3 sub F	O00303
Interleukin enhancer-binding factor 2 (ILF2)	Q12905
Polyadenylate-binding protein 2 (PABP2)	Q86U42
Eukaryotic initiation factor 4A-I (eIF-4A-1)	P60842
Heterogeneous nuclear ribonucleoprotein L (hnRNP-L)	P14866
Ras GTPase-activating protein-binding P 1	Q13283
Cleavage stimulation factor 64 kDa subunit (CstF-64)	P33240
Polyadenylate-binding protein 1 (PABP 1)	P11940
Nuclear RNA export factor 1 (Tip-associating protein)	Q9UBU9
Interleukin enhancer-binding factor 3 (ILF3)	Q12906
Eukaryotic translation initiation factor 3 sub B (eIF-3b)	P55884
5'-3' exoribonuclease 2	Q9H0D6
Eukaryotic translation initiation factor 3 sub C (eIF-3c)	Q99613
Regulator of nonsense transcripts 1	Q92900
Splicing factor 3B subunit 1 (SAP155)	O75533
Splicing factor 3B subunit 2 (SAP145)	Q13435
Valyl-tRNA synthetase	P26640
Elongation factor 1-gamma	P26641
Elongation factor 1-alpha 1 (eEF-1A-1)	P68104
Elongation factor 1-delta	P29692
60S ribosomal protein L18	Q07020
GTP-binding nuclear protein Ran	P62826
Exosome complex exonuclease RRP42	Q15024
Ran GTPase-activating protein 1 (RANGAP1)	P46060
Nuclear pore complex protein Nup214 (Nup214)	P35658

Matrin-3	P43243
Chromobox protein homolog 7 (ESCA)	O95931
General transcription factor IIIH subunit 3 (BTF2-p34)	Q13889
Protein DERPc	Q8NDH8
Histone deacetylase 2 (HDAC2)	Q92769
Poly(U)-binding-splic fact PUF60	Q9UHX1
Alpha-globin transcription factor CP2	Q12800
Paraspeckle component 1	Q8WXF1
Transcriptional repressor p66 beta	Q8WXI9
TFIIH basal TF complex helicase XPB sub (ERCC3)	P19447
Mediator of RNA polymerase II transcription sub 23	Q9ULK4
Transcription elongation factor SPT5 (hSPT5)	O00267
Death-inducer obliterator 1 (hDIDO1)	Q9BTC0
Interferon regulatory factor 2-BP 1	Q8IU81
DNA-binding protein SATB1	Q01826
DNA replication licensing factor MCM5	P33992
Septin-2 (Protein NEDD5)	Q15019
Katanin p60 ATPase-containing sub A1	O75449
T-complex protein 1 subunit theta (CCT-theta)	P50990
T-complex protein 1 subunit eta (CCT-eta)	Q99832
T-complex protein 1 subunit gamma (CCT-gamma)	P49368
Actin (alpha Actin)	P62736
γ-tubulin complex component 2 (hSpC97)	Q9BSJ2
Src substrate cortactin (Amplaxin)	Q14247
Mitogen-activated P kinase kinase kinase 10	Q02779
GTP-binding protein 1	O00178
CDKN2A-interacting protein (Collaborator of ARF)	Q9NXV6
Casein kinase II subunit alpha (CK 2)	P68400
Annexin A1 (Lipocortina 1)	P04083
Annexin A2 (Lipocortina 2)	P07355
P-glutamine gamma-glutamyltransferase	Q08188

Zymogen granule P16 homolog B (ZG16B)	Q96DA0
14-3-3 protein gamma	P61981
Protein S100-A14	Q9HCY8
Small proline-rich protein 3 (Esophagin)	Q9UBC9
Glyceraldehyde-3-phosphate dehydrogenase	P04406
14-3-3 protein sigma	P31947
Short palate, lung and nasal epithelium carcinoma-AP 2	Q96DR5
Protein SEC13 homolog	P55735
Protein arginine N-methyltransferase 1 (PRMT1)	Q99873
N-acylneuraminate cytidyltransferase	Q8NFW8
WD40 repeat-containing protein SMU1 (SMU1)	Q2TAY7
UPF0027 protein C22orf28	Q9Y3I0
Ubiquitin-associated protein 2-like (Protein NICE-4)	Q14157
Periodic tryptophan protein 2 homolog	Q15269
C-1-tetrahydrofolate synthase	P11586
Structural maint.chromos flexible hinge domain-cont. P1	A6NHR9

E' stato possibile classificare funzionalmente i putativi interattori di CBX7 sulla base delle informazioni presenti in Swiss Prot o in letteratura. Da tale classificazione si osserva la presenza di proteine coinvolte nella sintesi e nella degradazione del RNA (RNA destiny), nella trasduzione del segnale, nell'organizzazione del citoscheletro e nella trascrizione (Fig. 3.4).

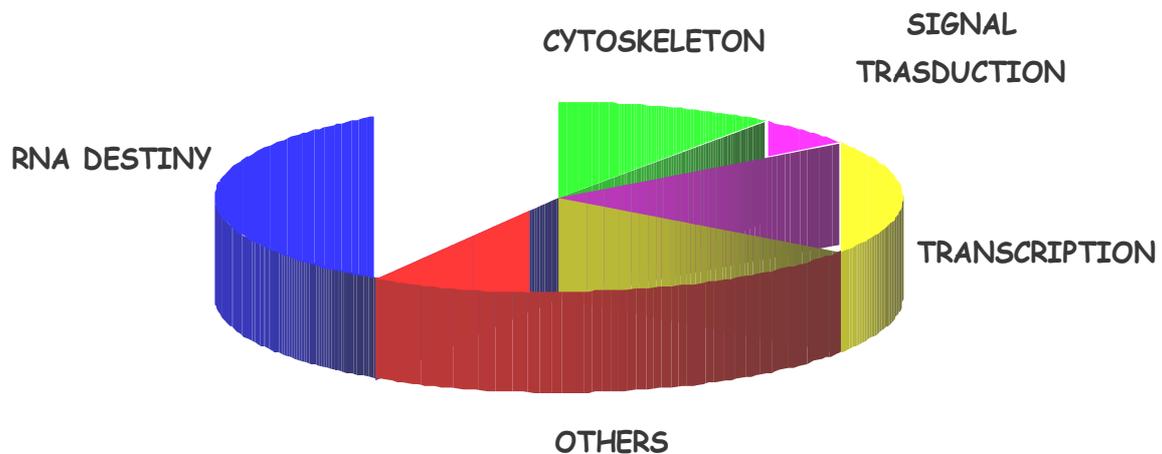


Figura 3.4 Classificazione funzionale degli interattori di CBX7 identificati mediante spettrometria di massa.

3.3.1.1 Caratterizzazione dei complessi molecolari coinvolgenti CBX7 in vivo

I putativi partners di CBX7 identificati con l'approccio di proteomica prima descritto sono stati confermati mediante esperimenti indipendenti, quali ad esempio la coimmunoprecipitazione.

Tra le proteine interagenti con CBX7, è stata studiata l'interazione con HDAC2 a causa della rilevanza di quest'ultima nella biologia tumorale [48]. Le proteine HDAC sono enzimi che catalizzano la rimozione del gruppo acetile dal core istonico e, a causa della loro capacità di indurre una locale condensazione della cromatina, sono considerati repressori della trascrizione.

Per verificare l'interazione *in vivo* di CBX7/HDAC2, cellule HEK293 sono state trasfettate transientemente con il vettore di espressione V5-CBX7. Le proteine lisate sono poi state immunoprecipitate con un anticorpo anti-V5 e rivelate mediante saggio Western Blot con gli anticorpi anti-V5 e anti-HDAC2. Come mostrato nella figura 3.5, è stato possibile stabilire l'associazione tra CBX7 e la proteina HDAC2 endogena, dimostrando che formano un complesso *in vivo*. L'esperimento reciproco è stato eseguito immunoprecipitando con l'anticorpo anti-HDAC2 e rivelando con l'anticorpo anti-V5: è stata confermata l'interazione tra HDAC2 e CBX7.

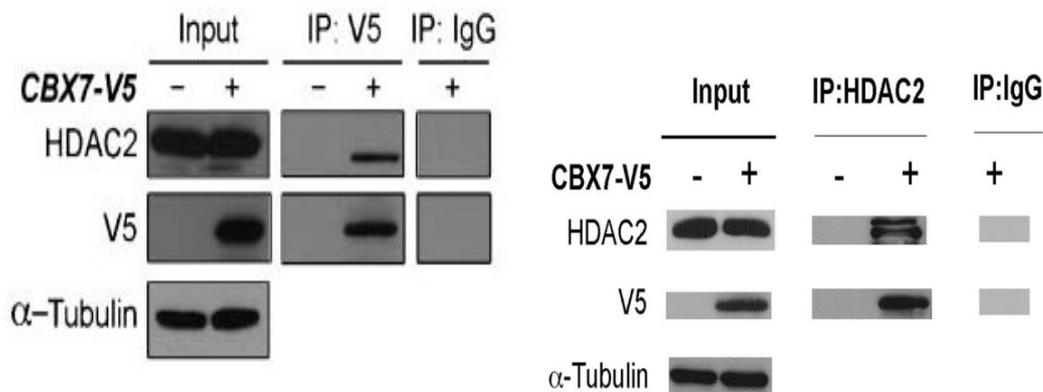


Figura 3.5. Interazione di HDAC2/CBX7 in vivo mediante co-immunoprecipitazione

Da cellule HEK 293 trasfettate con il vettore di espressione V5-CBX7 è stato preparato un estratto totale cellulare e immunoprecipitato con un anticorpo anti-V5. Gli immunocomplessi sono stati analizzati mediante Western Blot usando un anticorpo anti-HDAC2. L'input è rappresentato da un'aliquota dell'estratto totale di partenza. IgG indica il controllo negativo dell'immunoprecipitazione usando un anticorpo aspecifico.

3.3.1.2. *CBX7 inibisce l'attività di HDAC2*

Per verificare quale sia l'effetto dell'interazione CBX7/HDAC2 sull'attività di HDAC2, sono stati preparati estratti nucleari derivanti sia da cellule HeLa che NPA in modo da testare la capacità di CBX7 di modulare l'attività di HDAC2, incrementando di volta in volta la quantità della proteina trasfettata nelle cellule. L'espressione di CBX7 inibisce significativamente l'attività di HDAC in maniera dose-dipendente (la percentuale di inibizione è stata del 61.15% e del 42.55% nelle cellule HeLa ed NPA rispettivamente, dopo la trasfezione di 10 µg di vettore di espressione di CBX7; Fig.3.6 A e B). Come controllo positivo per l'inibizione di HDAC, è stato usato il sodio butirrato, un forte inibitore di HDAC, a una concentrazione di 250 mmol/L [48,49]. Gli estratti nucleari derivanti da cellule NPA trasfettate con CBX7 sono stati immunoprecipitati con un anticorpo anti-HDAC2. Il materiale immunoprecipitato è stato testato usando il saggio di attività per HDAC. Come mostrato nella Fig. 3.6C, CBX7 inibisce specificamente l'attività di HDAC2.

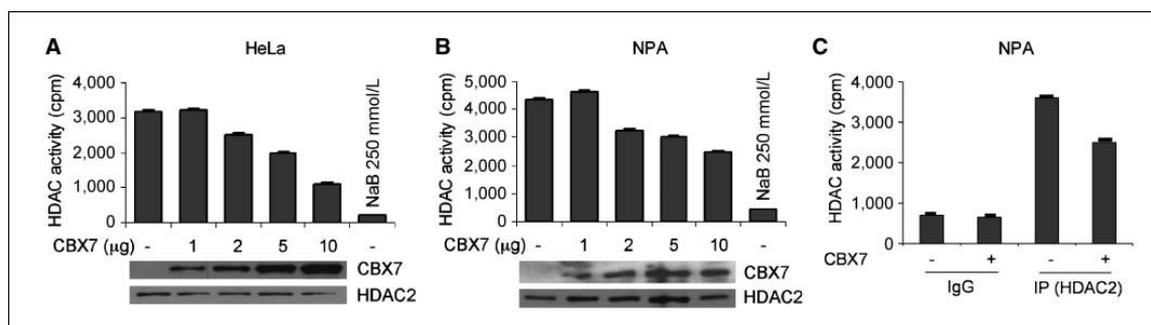


Fig.3.6. CBX7 inibisce l'attività di HDAC.

Cellule HeLa e NPA (A e B, rispettivamente), trasfettate transientemente con quantità crescenti del vettore di espressione di CBX7, sono state testate per l'attività di HDAC. I campioni trattati con sodio butirrato (NaB) sono stati usati come controllo positivo dell'inibizione dell'attività di HDAC. Aliquote degli stessi lisati sono state immunoblottate con gli anticorpi indicati (sotto). C, cellule NPA sono state trasfettate transientemente con il vettore di espressione di CBX7 o con il vettore vuoto. Gli estratti nucleari sono stati immunoprecipitati usando anticorpi anti-HDAC2 e saggiati per l'attività di HDAC.

3.3.1.3. CBX7 lega il promotore del gene dell'E-cadherin

E' stato recentemente mostrato che HDAC2 è coinvolto nel complesso di repressione che silenzia l'espressione del gene dell'E-cadherin durante la progressione tumorale [50]. L'E-cadherin è uno dei custodi del fenotipo epiteliale in quanto la perdita della sua espressione è stato dimostrato avere un ruolo critico nella transizione epitelio-mesenchimale (EMT) [51,52], cioè un meccanismo indispensabile durante la morfogenesi che permette la formazione di organi e tessuti. Durante questo processo, attraverso una serie di eventi, le cellule epiteliali perdono molte delle loro caratteristiche e acquisiscono proprietà tipiche delle cellule mesenchimali, con complessi cambiamenti nell'architettura cellulare. Il preciso spettro di cambiamenti che avviene durante l'EMT è probabilmente determinato dall'integrazione dei segnali extracellulari che le cellule ricevono, sebbene quest'aspetto non sia ancora chiaro. Questo programma di sviluppo ha un ruolo anche nello sviluppo tumorale. L'EMT provvede a dare nuove basi per comprendere la progressione tumorale verso stadi dedifferenziati e più maligni.

Quindi per verificare se anche CBX7 è capace di legare il promotore dell'E-cadherin è stata effettuata una *chromatin immunoprecipitation* (ChIP). Cellule HEK 293 and HeLa sono state transientemente trasfettate con il vettore di espressione HA-CBX7. Dopo aver verificato mediante Western blotting l'espressione proteica di HA-CBX7 (Fig. 3.7A), la proteina è stata crosslinkata al DNA ed immunoprecipitata con anticorpi anti-HA o IgG. L'immunoprecipitazione della cromatina è stata poi analizzata mediante una PCR semiquantitativa usando primers che mimano la regione del promotore dell'E-cadherin (da 300 bp a monte del promotore a +40 bp a valle). Gli anticorpi anti-HA precipitano questa regione del promotore dell'E-cadherin sia nelle cellule HEK 293 che nelle cellule HeLa trasfettate con la proteina CBX7 taggata con HA. Nessuna immunoprecipitazione è stata osservata con anticorpi aspecifici IgG o quando sono stati usati i primers per il promotore di GAPDH o P16 (Fig. 3.7A), indicando che il legame di CBX7 è specifico per il promotore dell'E-cadherin.

Successivamente si è investigato se l'interazione tra CBX7 ed HDAC2 avviene fisicamente sul promotore dell'E-cadherin. E' stato eseguito un'esperimento di *re-chromatin immunoprecipitation* (Re-ChIP). Cellule HEK 293 sono state trasfettate transientemente con HA-CBX7, sono state crosslinkate e immunoprecipitate con anticorpi anti-HDAC2. I complessi così isolati sono stati rilasciati, re-immunoprecipitati con anticorpi anti-HA e poi analizzati mediante PCR. I risultati mostrati in Fig. 3.7B rivelano che l'anticorpo contro HA precipita il promotore dell'E-cadherin dopo il rilascio dall'anticorpo anti-HDAC2. L'esperimento reciproco permette di comparare i risultati. (Fig. 3.7C).

Presi insieme questi risultati indicano che CBX7 ed HDAC2 formano un complesso che in vivo lega il promotore umano dell' E-cadherin.

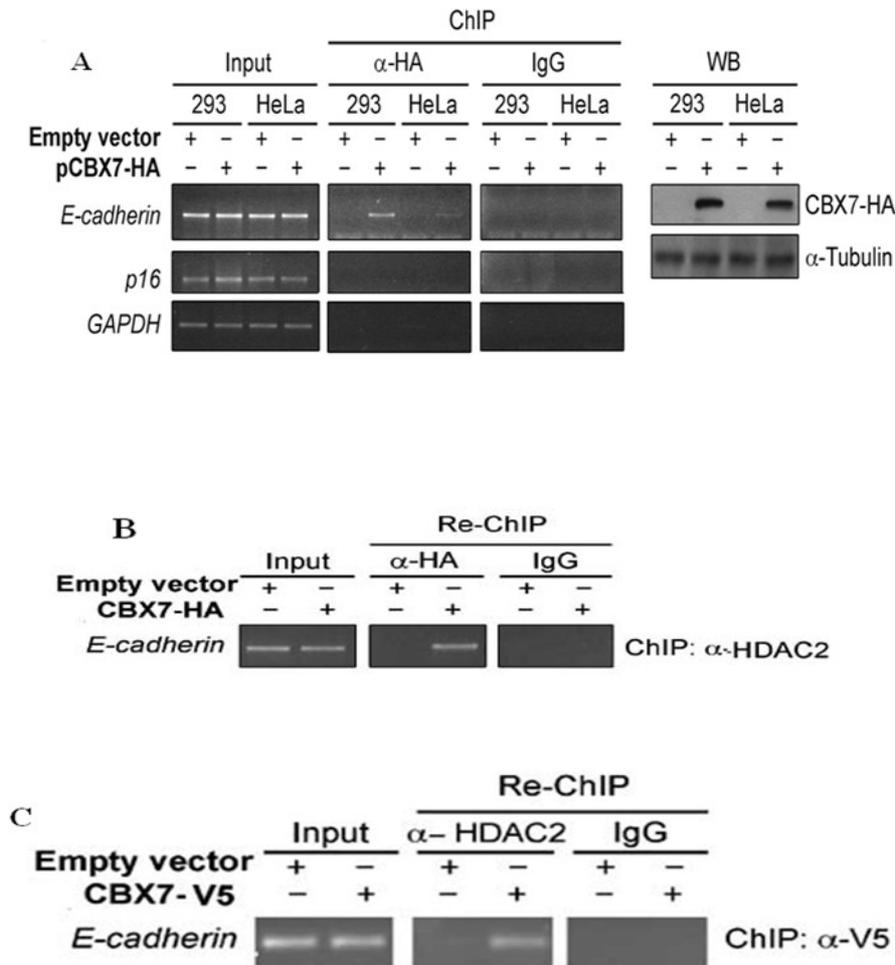


Fig. 3.7 CBX7 lega il promotore del gene dell'E-cadherin.

A, esperimento di *chromatin immunoprecipitation* (ChIP) usando anticorpi anti-HA in cellule HEK 293 e HeLa trasfettate transientemente con HA-CBX7 o il vettore vuoto. Il DNA associato è stato amplificato mediante PCR usando primers specifici per i promotori del gene umano (sinistra). Per il promotore dell'E-cadherin, è stata utilizzata una regione dal nucleotide 300 al +40 del gene. Anticorpi aspecifici IgG furono usati come immunoprecipitazione di controllo. **B**, esperimenti di *re-chromatin immunoprecipitation* (Re-ChIP) in cui la cromatina solubile immunoprecipitata con anti-HDAC2 è stata re-immunoprecipitata con anti-HA. Il controllo con IgG si riferisce alla Re-ChIP con anti-HA. **C**, esperimenti di Re-ChIP in cui la cromatina solubile immunoprecipitata con anti-V5 è stata re-immunoprecipitata con anti-HDAC2. Il controllo con IgG si riferisce alla Re-ChIP con anti-HDAC2.

3.3.1.4. CBX7 regola positivamente il promotore di E-cadherin

Per verificare l'effetto dell'espressione di CBX7 sulla trascrizione di E-cadherin, cellule HEK 293 sono state transientemente cotrasfettate con un vettore di espressione codificante CBX7 e con un vettore reporter che contiene il gene della luciferasi sotto il controllo del promotore dell'E-cadherin. Come mostrato in figura 3.8A, CBX7 aumenta l'attività trascrizionale del promotore dell'E-cadherin in modo dose-dipendente. Il trattamento delle cellule con tricostatina A, un potente inibitore dell'attività di HDAC [53], coopera con CBX7 a indurre la trascrizione del gene di E-cadherin (Fig 3.8B). Questi risultati hanno fortemente suggerito che la proteina CBX7 è coinvolta nella trascrizione del gene E-cadherin contrastando probabilmente gli effetti inibitori di HDAC2 già conosciuti sul promotore di questo gene.

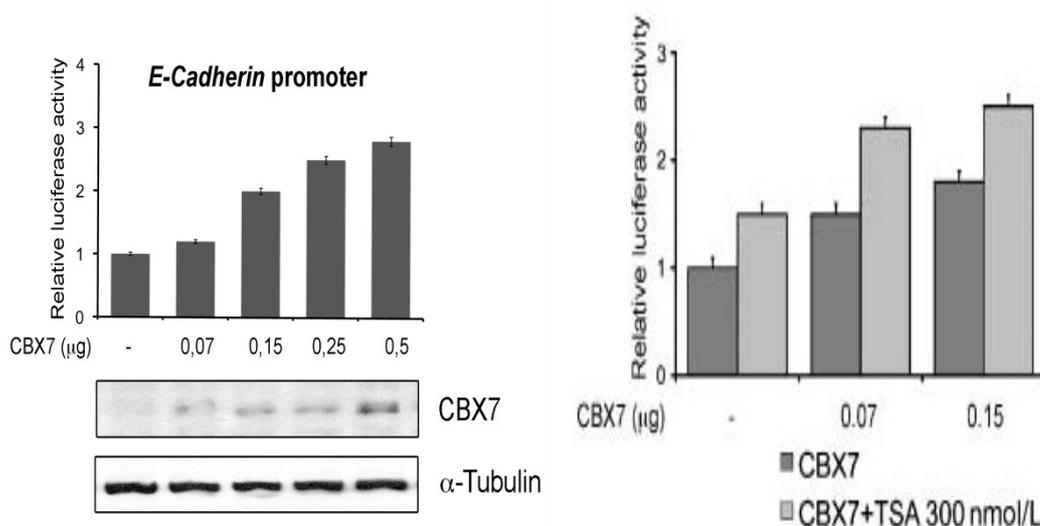


Fig. 3.8 CBX7 aumenta l'attività del promotore dell'E-cadherin.

A, in alto, analisi della risposta avuta incrementando la quantità di CBX7 sul vettore reporter dell'E-cadherin sotto il controllo della luciferase transientemente trasfettato in cellule HEK 293. L'analisi Western blot conferma l'incremento dell'espressione di CBX7. L'espressione seguita con α -Tubulin è servita come controllo per la quantità di proteina caricata. **B**, analisi della risposta avuta incrementando la quantità di CBX7 sul vettore reporter dell'E-cadherin sotto il controllo della luciferase transientemente trasfettata in cellule HEK 293 trattate o non con 300 nmol/L dell'inibitore di HDAC tricostatina A (TSA).

3.3.1.5. I livelli di espressione di CBX7 ed E-cadherin sono correlati nei carcinomi tiroidei umani.

Allo scopo di verificare se il legame di CBX7 al promotore di E-cadherin possa giocare un ruolo nella carcinogenesi tiroidea, è stata analizzata l'espressione di E-cadherin in carcinomi tiroidei umani di differenti istotipi. Come mostrato in Fig. 3.9 è stata riscontrata una correlazione positiva tra l'espressione di CBX7 ed E-cadherin. I livelli dell'mRNA di E-cadherin infatti sono drasticamente ridotti nei campioni ATC, mentre è stato osservato un leggero decremento in quelli PTC. L'espressione di CBX7 mostra lo stesso andamento dell'E-cadherin. Questi risultati sono d'accordo con altri dati che mostrano una down-regolazione del gene di E-cadherin in carcinomi umani di diversa origine anatomica caratterizzati da un fenotipo molto aggressivo [54].

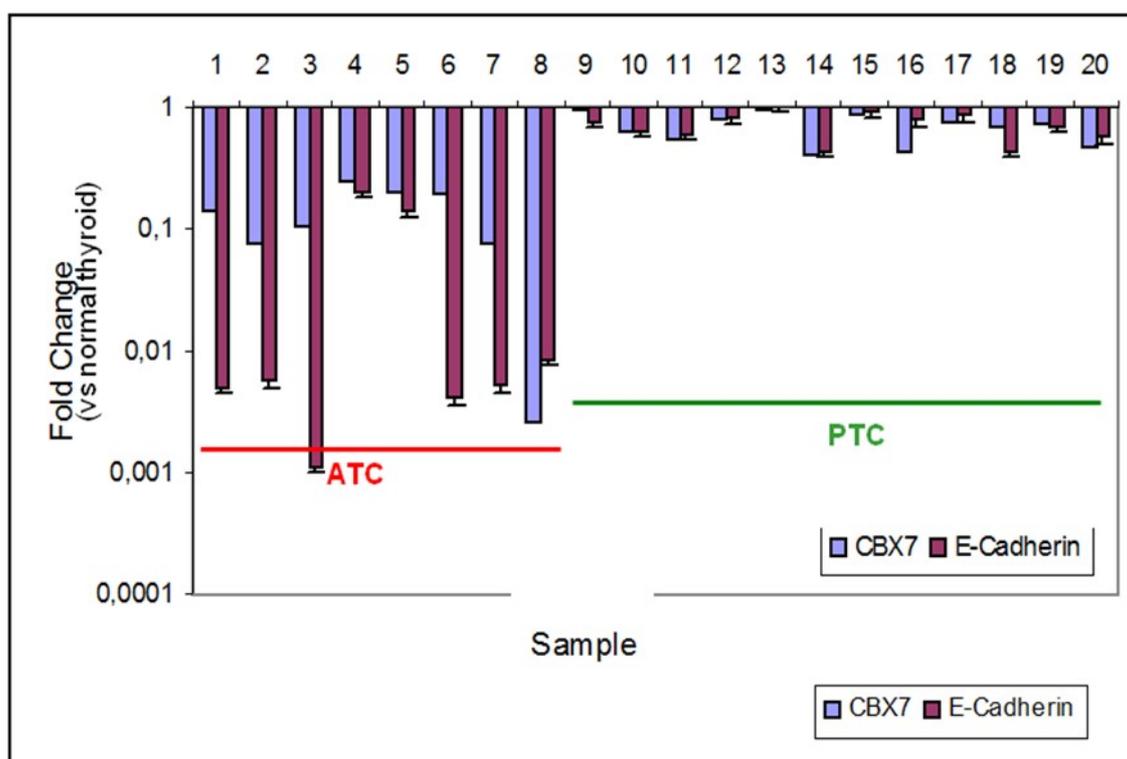


Fig. 3.9 CBX7 aumenta l'attività del promotore dell'E-cadherin.

Il Fold Change indica il relativo cambio dei livelli di espressione tra i campioni tumorali e i campioni normali, assumendo che il valore di ogni campione normale sia uguale ad 1. PTC, carcinoma papillare tiroideo; ATC, carcinoma tiroideo anaplastico. I numeri indicano i pazienti.

3.3.2 Separazione dei complessi di CBX7 mediante tecniche di Frazionamento.

Dall'esperimento di immunoprecipitazione sopra descritto, oltre ad HDAC2, sono stati identificati un numero elevato di putativi interattori. Per razionalizzare la lista di questi interattori, che hanno suggerito un eventuale coinvolgimento di CBX7 in diversi complessi funzionali, si è pensato di utilizzare delle tecniche di pre-purificazione dei diversi complessi senza destabilizzare le interazioni native. Le metodologie utilizzate sono state la centrifugazione in gradiente di densità e la Gel Filtration.

3.3.2.1. Separazione mediante centrifugazione in gradiente di densità

Tra le tecniche di frazionamento per separare i complessi è stata utilizzata in primo luogo l'ultracentrifugazione zonale in gradiente di densità, che ha permesso di separare i complessi proteici formati in vivo in base alla loro velocità di sedimentazione e quindi alla loro densità. Il gradiente di densità 5-45% è stato preparato utilizzando soluzioni di glicerolo a diversa concentrazione come menzionato in Materiali e Metodi. Un campione di 2,5 mg di estratto proteico nucleare, è stato caricato sulla sommità del gradiente di glicerolo, e l'ultracentrifugazione condotta a 39000 rpm. Dopo 16 h di ultracentrifugazione, tempo necessario affinché avvenga la stratificazione del materiale proteico alla giusta densità, sono state recuperate in maniera discendente 11 frazioni da 1mL ciascuna. Un'aliquota di 200 μ L di ognuna è stata precipitata con metanolo e cloroformio ed i precipitati sono stati risospesi in loading buffer e frazionati mediante tecniche di elettroforesi monodimensionale. Le bande proteiche sono state immunorivelate mediante l'utilizzo di un anticorpo anti-V5 che riconosce il peptide fuso all'N-terminale della proteina CBX7. In questo modo è stato possibile seguire la distribuzione di CBX7, e indirettamente dei complessi in cui essa è coinvolta lungo tutto il gradiente di glicerolo (Fig. 3.10).



Figura 3.10. Western Blot delle frazioni recuperate da Gradiente di glicerolo sviluppato con anticorpo anti-v5.

Dal risultato dell'esperimento riportato in Fig.3.10 è stato dedotto che CBX7 è presente in tutte le frazioni, anche se le bande più intense si osservano in corrispondenza delle frazioni da 1 a 6 con un picco di massima intensità nelle frazioni 3 e 4. Alla luce di questo andamento si è scelto di riunire le frazioni 1-6 e di sottoporle ad immunoprecipitazione con anticorpo anti-v5 al fine di isolare i soli complessi contenenti CBX7. Le frazioni riunite sono state incubate ciascuno con 30 μ l di resina anti-mouse come precleaning. Il surnatante è stato rimosso e trattato con 50 μ l di resina anti-v5. Infine è stata eseguita l'eluizione per competizione con il tampone di eluizione in cui è presente un'elevata concentrazione di peptide V5. Gli eluati sono stati analizzati mediante elettroforesi monodimensionale su gel di poliacrilammide al 12% in condizioni denaturanti (SDS-PAGE). Il gel è stato colorato al Comassie. I risultati sono mostrati in Fig.3.11, dove è riportato il gel riguardante le frazioni riunite.

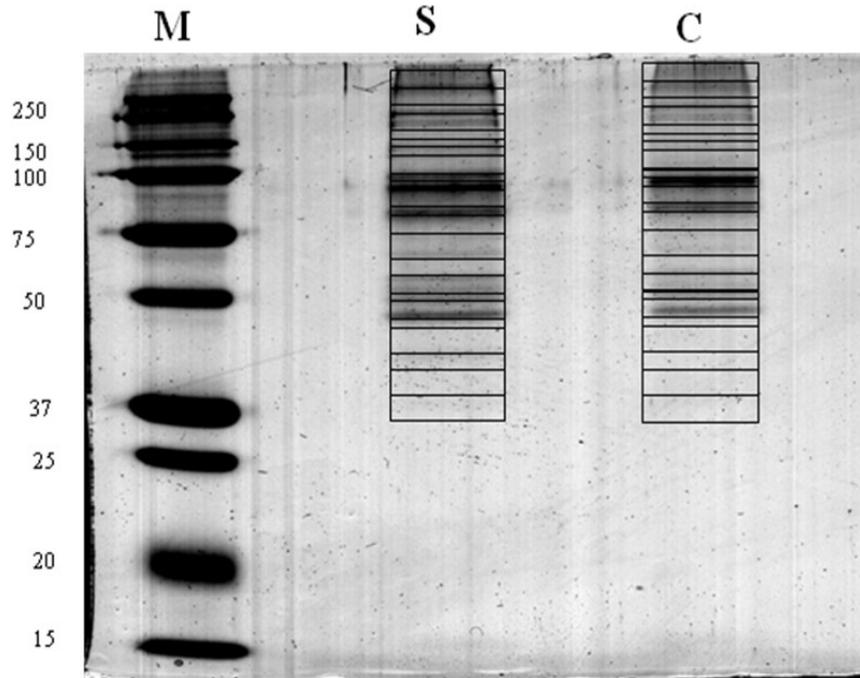


Fig.3.11. SDS-PAGE delle frazioni riunite e immunoprecipitate da gradiente di glicerolo. S= eluato da cellule trasfettate con v5-CBX7; C= eluato da cellule trasfettate con vettore vuoto.

Le bande segnalate in figura 3.11 sono state escisse da gel, sottoposte ad idrolisi in situ con tripsina e le miscele peptidiche ottenute sono state analizzate mediante LC-MS/MS.

Nella Tabella 3 sono riportate le identificazioni delle bande analizzate.

Tabella 3. Proteine identificate dall'esperimento di gradiente di glicerolo.

Protein name	Swiss prot code
RNA-binding protein 6 (Protein G16)	P78332
Polyhomeotic-like protein 1 (hPH1)	P78364
Ubiquitin-associated protein 2	Q5T6F2
Ribonucleoprotein PTB-binding 1 (RAVER1)	Q8IY67
Polyhomeotic-like protein 3 (hPH3)	Q8NDX5
Heterogeneous nuclear ribonucleoprotein K (hnRNP K)	P61978
Heterogeneous nuclear ribonucleoprotein L (hnRNP L)	P14866
Non-POU domain-contain. octamer-BP (p54nrb, NMT55)	Q15233
Polycomb group RING finger P 6 (RING finger P 134, MBLR)	Q9BYE7
Elongation factor 1-alpha 1 (EF-Tu)	P68104
Polycomb complex protein BMI-1 (RING finger P 51)	P35226
Polyhomeotic-like protein 2 (hPH2)	Q8IXK0
E3 ubiquitin-protein ligase RING2 (RING finger P 2)	Q99496
Chromobox protein homolog 7 (CBX7)	O95931
Polycomb group RING finger P 1 (RING finger P 68)	Q9BSM1

Queste proteine sono state classificate funzionalmente grazie alle informazioni reperibili in letteratura e su Swiss Prot (Fig.3.12). Da questa classificazione si evince che oltre a proteine coinvolte nello splicing, sono state identificate proteine specifiche del complesso PRC1.

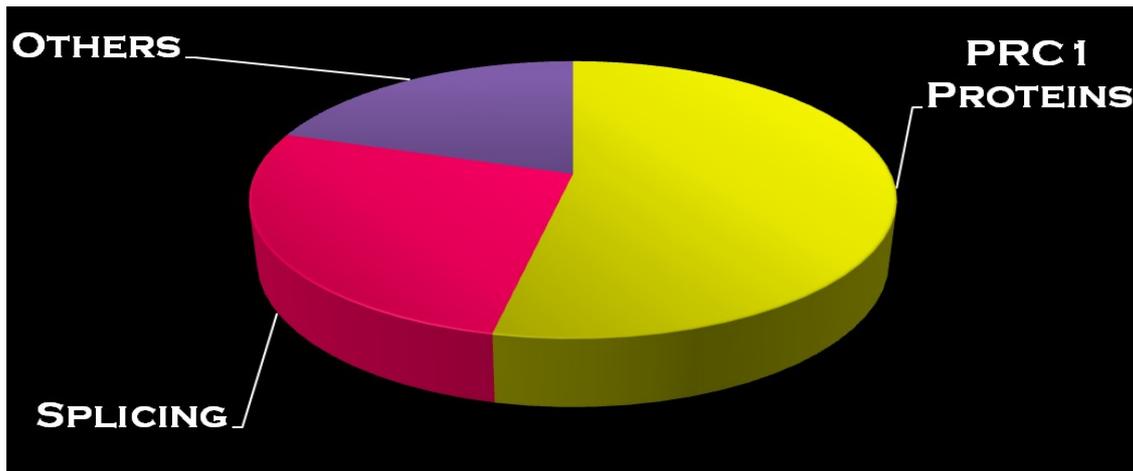


Fig.3.12. Classificazione funzionale delle proteine identificate dall'esperimento di gradiente di glicerolo.

3.3.2.2. Separazione mediante Gel Filtration

Una seconda metodologia di pre-frazionamento degli estratti utilizzata per separare i complessi in cui è coinvolta CBX7 è stata la cromatografia ad esclusione molecolare (Gel Filtration). Tale tecnica ha permesso di separare i complessi operando in condizioni native in base alle loro dimensioni molecolari e alla loro forma.

Circa 3,6 mg di estratto proteico nucleare e sospeso in 900 μ l di tampone fosfato è stato caricato, in tre tempi diversi, su una colonna Superose 6 10/300 GL eluita in tampone fosfato pH = 7.4. Durante la corsa cromatografica sono state raccolte 26 frazioni ciascuna di 500 μ l. Un'aliquota di 200 μ l di ciascuna frazione è stata sottoposta a precipitazione metanolo-cloroformio; i precipitati sono stati risospesi in loading buffer e analizzati mediante western blot, al fine di localizzare CBX7 per immunorivelazione con anticorpi (Fig. 3.13)

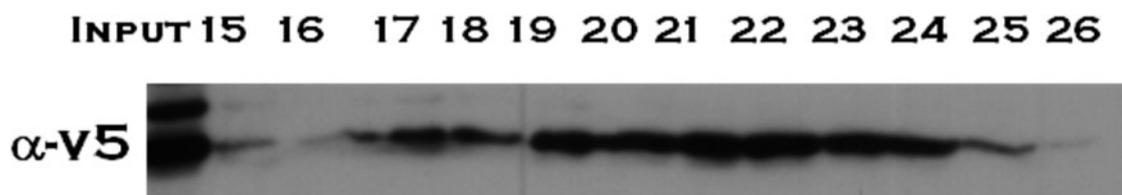


Figura 3.13: Western blot delle frazioni dopo gel filtration sviluppato con anticorpo anti-v5.

I risultati del Western blot indicano che è possibile individuare due picchi di massima intensità di CBX7 in corrispondenza delle frazioni 16,17,18 e 19, 20,21,22,23, chiamate rispettivamente pool A e pool B.

I pool A e B sono stati sottoposti separatamente ad immunoprecipitazione con anticorpo anti-V5 al fine di isolare i soli complessi contenenti CBX7 pre-frazionati. Le frazioni A e B sono state incubate ciascuno con 30 μ l di resina anti-mouse come precleaning. Il surnatante è stato rimosso e trattato con 50 μ l di resina anti-V5. Infine è stata eseguita l'eluizione per competizione con il tampone di eluizione in cui è presente un'elevata concentrazione di peptide V5. Gli eluati sono stati analizzati mediante elettroforesi monodimensionale su gel di poliacrilammide al 12% in condizioni denaturanti (SDS-PAGE). Il gel è stato colorato al Comassie. I risultati sono mostrati in Fig.3.14, dove è riportato il gel riguardante le frazioni riunite.

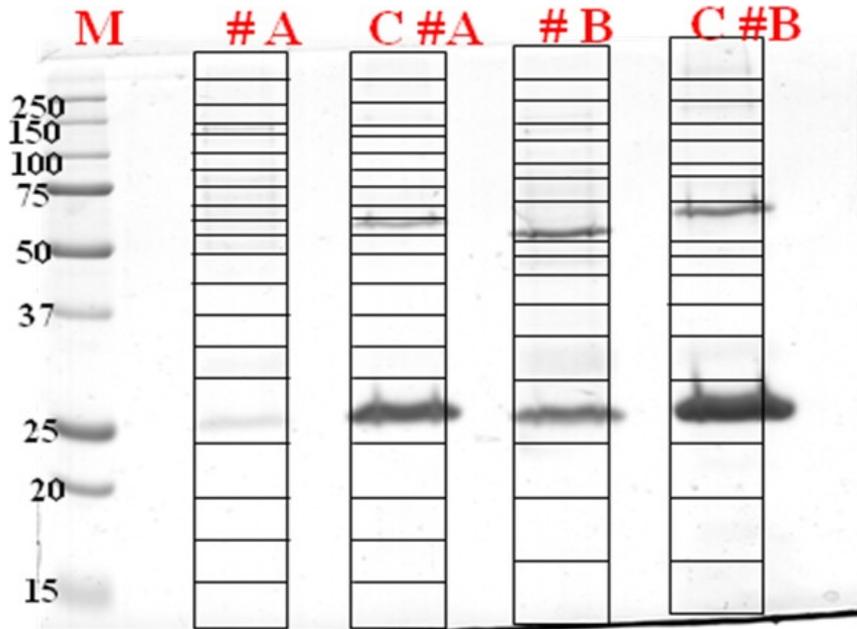


Fig.3.14. SDS-PAGE delle frazioni riunite ed immunoprecipitate da gel filtration.

#A= eluato dell'immunoprecipitazione della #A; C#A = eluato del controllo della #A; #B= eluato dell'immunoprecipitazione della #B; C#B = eluato del controllo della #B, M= marker.

Le bande segnalate in figura 3.14 sono state escisse da gel, sottoposte ad idrolisi in situ con tripsina e le miscele peptidiche ottenute sono state analizzate mediante LC-MS/MS.

Nelle Tabelle 4 e 5 sono riportate le identificazioni delle bande analizzate, rispettivamente delle frazioni A e B.

Tabella 4. Proteine identificate nella frazione A dell'esperimento di gel filtration

Protein name	Swiss Prot Code
poly(ADP-ribose) polymerase 1	P09874
Splicing factor Pro/Gln-rich (SFPQ)	P23246
Src substrate cortactin (Amplaxin)	Q14247
Far upstream element-bp 2	Q92945
Probable ATP-dep RNA helicase DDX17	Q92841
Zinc finger protein 143	P52747
heterogeneous nuclear ribonucleoprotein M	P52272
lamin A/C	P02545
KH domain containing, signal transduction associated 1 (sam68)	Q07666
Non-POU domain-containing octamer-bp	Q15233
60 kDa heat shock protein, mit. (HSP-60)	P10809
Nucleoporin-like protein RIP	P52594
heterogeneous nuclear ribonucleoprotein K	P61978
Probable Xaa-Pro aminopeptidase 3	Q9NQH7
RuvB-like 2	Q9Y230
Actin, cytoplasmic 1 (beta actin)	P60709
heterogeneous nuclear ribonucleoprotein D0	Q14103
Nuclear transcription factor Y subunit gamma	Q13952
heterogeneous nuclear ribonucleoprotein A/B	Q99729
peroxiredoxin 1	Q06830

Tabella 5. Proteine identificate nella frazione B dell'esperimento di gel filtration.

Protein name	SwissProt Code
poly(ADP-ribose) polymerase 1 (PARP-1)	P09874
Splicing factor Pro/Gln-rich	P23246
Elongation factor 2 (EF2)	P13639
Far upstream element-bp 2 (FUSEbp2)	Q92945
heterogeneous nuclear ribonucleoprotein M (hnRNP M)	P52272
Probable ATP-dep RNA helicase DDX5 (RNA helicase p68)	P17844
Far upstream element-bp 1 (FUSEbp1)	Q96AE4
Heat shock 70 kDa protein 1 (HSP70.1; HSP70-1/HSP70-2)	P08107
RNA-binding protein FUS (Oncogene TLS; POMp75)	P35637
RNA binding motif protein 14 (RBM14)	Q96PK6
Heat shock cognate 71 kDa protein (HSPA8)	P11142
Non-POU domain-containing octamer-bp (NONO protein)	Q15233
Polypyrimidine tract-bp 1 (PTB; hnRNP I)	P26599
Tyrosyl-tRNA synthetase, cyt (Tyrosyl-tRNA ligase)	P54577
Far upstream element-bp 1 (FUBP3)	Q96I24
heterogeneous nuclear ribonucleoprotein L (hnRNPL)	P14866
Probable Xaa-Pro aminopeptidase 3 (APP3)	Q9NQH7
heterogeneous nuclear ribonucleoprotein H (hnRNPH)	P31943
Regulator of chromosome condensation (RCC1)	P18754
Myc-binding protein-1 (alpha-enolase, MBP-1)	P06733
Beta-actin	P60709
DAZ-associated protein 1 (DAZAP1)	Q96EP5
heterogeneous nuclear ribonucleoprotein A/B (hnRNP A/B)	Q99729
heter nuclear ribonucleoprot. A1 (hnRNP core protein A1)	P09651
heter nuclear ribonucleoprotein A0 (hnRNP A0)	Q13151
Nascent polypeptide-associated complex sub alpha	Q13765
heter nuclear ribonucleoprot A2/B1	P22626
peroxiredoxin 1	Q06830

Anche in questo caso è stato possibile classificare funzionalmente le proteine identificate, sia per la frazione A che per la B (Fig.3.15). In entrambi i casi sono state identificate proteine impegnate nello splicing e nella regolazione della trascrizione.

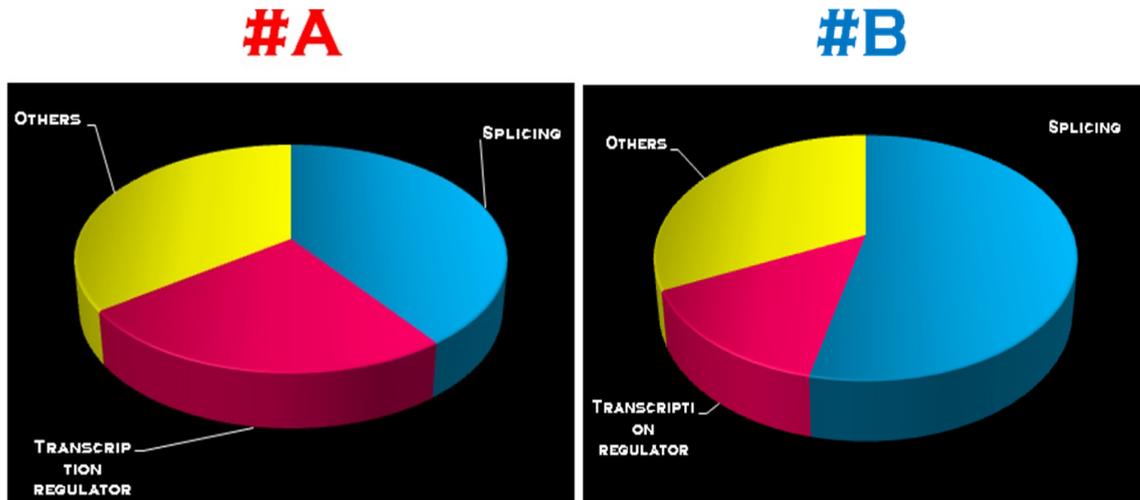


Fig.3.15. Classificazione funzionale delle proteine identificate dall'esperimento di gel filtration.

3.4. Conclusioni

I tumori tiroidei derivano dall'accumulo di varie modifiche nei geni critici coinvolti nel controllo della proliferazione cellulare. Sebbene nella pratica clinica siano seguiti vari approcci terapeutici per la cura di questi, la maggior parte di essi non raggiunge esiti positivi. Così, la scoperta di nuovi approcci per diagnosticare il cancro in stadi precoci e per stabilire terapie efficaci rappresenta una questione critica e urgente. Per raggiungere questo scopo, l'identificazione e la caratterizzazione di molecole chiave che partecipano alla carcinogenesi sono punti essenziali. A tal proposito ci si è soffermati sul ruolo di CBX7, che potrebbe essere coinvolto nel processo di carcinogenesi tiroidea. CBX7 è espressa abbondantemente nella ghiandola tiroidea normale, ma la sua espressione diminuisce con la malignità e lo stadio neoplastico. Infatti, l'espressione di CBX7 è comparabile al tessuto tiroideo normale negli adenomi follicolari benigni, leggermente ridotta nei carcinomi tiroidei papillari e drasticamente ridotta, e in molti casi assente, nei carcinomi tiroidei follicolari. Si potrebbe così pensare che i livelli di CBX7 potrebbero servire per differenziare le neoplasie tiroidee benigne e maligne.

Allo scopo di studiare la funzione di CBX7 nelle cellule normali e capire perché è drasticamente down-regolato nelle cellule tumorali, abbiamo deciso di investigare le proteine che interagiscono con CBX7, mediante un approccio di proteomica funzionale. Tali approcci sono molto validi perché l'associazione di CBX7 con dei partners appartenenti a uno specifico complesso proteico coinvolto in un particolare meccanismo può dare suggerimenti sulla funzione biologica della proteina [55]. Tra i risultati ottenuti è sembrata particolarmente interessante l'interazione tra CBX7 e HDAC2, un noto repressore della trascrizione di vari geni importanti tra cui E-cadherin.

Infatti, la perdita di espressione di CBX7 potrebbe essere correlata con l'acquisizione dell'invasività accompagnata dalla perdita delle caratteristiche epiteliali e il guadagno del fenotipo mesenchimale, un processo conosciuto come transizione epitelio-mesenchimale. L'E-cadherin è il principale elemento dei nodi di adesione cellula-cellula, indispensabile per la normale morfologia epiteliale della cellula, emergendo perciò come caratteristica del fenotipo epiteliale. Nella maggior parte dei tumori, la down-regolazione dell'E-cadherin durante la progressione neoplastica avviene mediante meccanismi epigenetici, che includono la repressione trascrizionale, in alcuni casi mediata dall'attività di HDAC [54], e l'ipermetilazione del promotore [56]. Solo in pochi casi sono state trovate mutazioni nel gene di E-cadherin che hanno portato all'assenza o all'espressione di una proteina non funzionale [52]. Per questo motivo è stata seguita l'espressione di E-cadherin in carcinomi tiroidei di differenti istotipi: è stata riscontrata un'evidente correlazione tra i livelli di espressione nei carcinomi tiroidei umani, tale per cui entrambi sono drasticamente down-regolate nei carcinomi tiroidei anaplastici paragonati al normale tessuto tiroideo. D'altra parte nei carcinomi tiroidei non è stata ri-

scontrata ipermetilazione del promotore dell'E-cadherin. In questo lavoro è stato dimostrato, mediante esperimenti di *chromatin immunoprecipitation* che in vivo CBX7 lega il promotore del gene dell'E-cadherin ed è capace di regolarne positivamente l'attività. Allo stesso tempo, mediante l'approccio di proteomica volto a delucidare il meccanismo attraverso cui la perdita di espressione di CBX7 è coinvolta nella progressione tumorale, è stata identificata HDAC2 tra le proteine interagenti con CBX7. Le proteine appartenenti alla famiglia delle HDAC regolano l'espressione e l'attività di numerose proteine coinvolte sia nell'inizio sia nella progressione del tumore, inducendo a una conformazione della cromatina non permissiva che previene la trascrizione dei geni codificanti proteine coinvolte nella tumorigenesi. Le proteine HDAC sono spesso overesprese in molti tumori [48]. In questo lavoro si è dimostrato che CBX7 interagisce con la proteina HDAC2 inibendo la sua attività. La *chromatin immunoprecipitation* dimostra inoltre che questa interazione avviene sul promotore dell'E-cadherin. E' stata anche dimostrata la capacità di CBX7 di regolare positivamente l'espressione di E-cadherin interagendo con HDAC2 e inibendo la sua attività sul promotore dell'E-cadherin.

In conclusione, nella prima parte di questo lavoro è stato proposto un nuovo *pathway* che regola la progressione della carcinogenesi in cui la proteina CBX7, la cui perdita di espressione correla con un fenotipo molto maligno, è una molecola chiave. Inoltre questi risultati indicano che la perdita di espressione di CBX7 contribuisce alla progressione tumorale mediante la down-regolazione dell'espressione di E-cadherin a causa della perdita del suo effetto inibitorio sull'attività di HDAC sul promotore dell'E-cadherin.

Nella seconda parte di questo progetto invece si è cercato di dare un significato alla rimanente lista di interattori identificati. Infatti, il cospicuo numero di interattori riportati in Tabella 2 e classificati funzionalmente in Figura 3.4 ha suggerito che CBX7 potrebbe essere coinvolta in diversi complessi funzionali, anche se è stato pressoché impossibile assegnare tutti i componenti proteici allo specifico complesso cui appartengono, nonostante la ricchezza delle informazioni fornite da banche dati di proteine e database di interazioni proteina-proteina.

Quindi poiché CBX7 prende parte a più complessi, rendendo difficile l'interpretazione dell'esperimento di proteomica, è nata la necessità di mettere a punto delle strategie che consentissero di separare fisicamente i complessi coinvolgenti la proteina esca, rendendo più semplice l'attribuzione delle funzioni biologiche ai singoli complessi. Sono state sviluppate quindi due metodologie di frazionamento dei complessi dall'estratto proteico nucleare di partenza, sfruttando le differenti caratteristiche chimico-fisiche dei complessi stessi.

In primo luogo i complessi contenenti CBX7 sono stati separati sulla base della loro velocità di sedimentazione utilizzando l'ultracentrifugazione zonale in gradiente di densità che consente la sepa-

razione dei complessi molecolari in base alla forma, dimensione e peso molecolare. Alternativamente i complessi proteici sono stati pre-frazionati in base alla loro grandezza molecolare e coefficiente idrodinamico mediante cromatografia ad esclusione molecolare. In entrambi i casi, la presenza della proteina esca nelle varie frazioni raccolte è stata verificata con un anticorpo specifico anti-V5.

In particolare, dall'esperimento di ultracentrifugazione zonale in gradiente di densità sono stati identificati una serie di componenti del complesso PRC1 quali BMI-1, hPH1, hPH2, hPH3, RING1B. In letteratura, il complesso PRC1 è spesso descritto come una singola entità, anche se le diverse purificazioni dei complessi PRC1 riportate hanno fornito diverse informazioni circa la sua composizione e stechiometria.

Il complesso PRC1 è formato da almeno un rappresentante di ciascuna delle quattro famiglie di proteine altamente conservate che ne costituiscono il core: Polycomb (CBX2/4/6/7/8), Polyhomeotic (HPH1/2/3), Posterior sex combs (PSC1-6) e Sex combs extra (RING 1A/1B) [57]. In completo accordo con quanto riportato in letteratura, in seguito a questo esperimento è stato co-purificato con CBX7 almeno un rappresentante di ognuna delle famiglie del core di PRC1, il che suggerisce la presenza dell'intero complesso PRC1 intatto. Inoltre, sono stati identificati tutti i membri della famiglia Polyhomeotic, suggerendo che nelle nostre condizioni sperimentali CBX7 potrebbe essere coinvolta in diversi complessi PRC1. Sebbene questi dati non sembrano escludere la possibilità di associazioni preferenziali *in vivo*, dovuti ai relativi livelli di espressione in differenti tipi cellulari, essi sembrerebbero suggerire che sono possibili diverse permutazioni dei complessi PRC1. Ci si potrebbe aspettare che questi diversi complessi abbiano funzioni simili o addirittura uguali tra loro e che possono essere usati in maniera intercambiabile senza una particolare logica. Comunque, una gran parte di letteratura che si occupa di quest'argomento ascrive funzioni specifiche a particolari proteine PRC1 o sostiene che queste siano implicate in meccanismi unici che la maggior parte delle volte non possono essere generalizzati. Un altro modo per razionalizzare la possibile molteplicità di complessi PRC1 in cellule eucariotiche sarebbe stabilire se essi hanno come bersagli diversi geni target.

Altre informazioni suggeriscono che le proteine PRC1 possono agire fuori del contesto classico del complesso PRC1, spesso associato con funzioni specifiche [58]. Un esempio è dato dal complesso BCOR in mammiferi, di cui abbiamo identificato il membro Ring finger protein 68, che è capace di ubiquitinare efficientemente l'istone H2A senza il pieno coinvolgimento di tutti i componenti del complesso PRC1 [59, 60].

Un'altra parte di proteine identificate in seguito a questo esperimento di ultracentrifugazione zonale in gradiente di densità è stata classificata come appartenenti allo splicing. Queste proteine sono state identificate anche in seguito all'esperimento di cromatografia ad esclusione molecolare.

In particolare, sono state identificate diverse proteine coinvolte nel processing dell'RNA, molte delle quali coinvolte nello splicing, suggerendo che CBX7 è associato con l'RNA. CBX7, infatti, così come altre proteine del complesso PRC1, associa con il cromosoma X inattivo mediante il suo chromodominio (CD) che lega la lisina 27 trimetilata dell'istone H3. Ma è stato dimostrato anche che CBX7 riesce a legare la cromatina mediante l'RNA, sempre utilizzando il suo CD. Quindi questo dominio può partecipare al legame di CBX7 con la cromatina in almeno due modi, mediante la metilazione dell'istone e mediante l'RNA. In particolare CBX7 lega l'RNA a singolo filamento con maggiore affinità rispetto all'RNA a doppio filamento e con minore affinità il DNA a doppio filamento [32]. Inoltre, sta diventando sempre più evidente che i processi di regolazione della cromatina, trascrizione e splicing sono fortemente integrati [61]. Letteratura recente suggerisce che lo splicing avviene direttamente a livello del trascritto nascente e perciò media la produzione dei trascritti varianti per tutti i geni di mammiferi [62, 63, 64]. Il coinvolgimento della cromatina e dei suoi modificatori in questo processo ha recentemente acquisito ulteriore sostegno in quanto si è riscontrato che i nucleosomi sono preferenzialmente localizzati in regioni esoniche e può giocare un ruolo importante nel riconoscimento dell'esone [65, 66]. Questo suggerisce che i processi coinvolti nella repressione mediata da PcG che modifica i nucleosomi o altera il loro compattamento potrebbero potenzialmente influenzare l'efficacia dello splicing.

Bibliografia CBX7

1. Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular cell neoplasia. *Nat Rev Cancer* 2006;6:292–306.
2. De Felice M, Di Lauro R. Thyroid Development and Its Disorders: Genetics and Molecular Mechanisms. *Endocrine Reviews* 2004; 25:722–46.
3. Saltman B, Singh B, Hedvat CV, Wreesmann VB, and Ghossein R. Patterns of expression of cell cycle/apoptosis genes along the spectrum of thyroid carcinoma progression. *Surgery* 2006;140:899-905.
4. Van der Laan BF et al. The association of well-differentiated thyroid carcinoma with insular or anaplastic thyroid carcinoma: evidence for dedifferentiation in tumor progression. *Endocr Pathol* 1993;4:215-221.
5. Delys L, Detours V, Franc B, et al. Gene expression and the biological phenotype of papillary thyroid carcinomas. *Oncogene* 2007;26:7894–903.
6. Pallante P, Berlingieri, MT, Troncone, G, Kruhoffer M, Orntoft TF, Viglietto G, Caleo A, Migliaccio I, Decaussin-Petrucci, M, Santoro M, et al. . UbcH10 overexpression may represent a marker of anaplastic thyroid carcinomas. . *Br J Cancer* 2005;93: 464-71.
7. Pallante P, Federico A, Berlingieri MT, Bianco M, Ferraro A, Forzati F, Iaccarino A, Russo M, Pierantoni GM, Leone V, Sacchetti S, Troncone G, Santoro M, Fusco A. Loss of the CBX7 gene expression correlates with a highly malignant phenotype in thyroid cancer. *Cancer Res* 2008;68(16):6770-8.
8. Gil J, Bernard D, Martínez D, Beach D. Polycomb CBX7 has a unifying role in cellular lifespan. *Nat Cell Biol* 2004;6:67-72.
9. Otte AP, Kwaks TH. Gene repression by Polycomb group protein complexes: a distinct complex for every occasion? *Curr Opin Genet Dev* 2003;13:448–54.
10. Lund AH, van Lohuizen M. Polycomb complexes and silencing mechanisms. *Curr Opin Cell Biol* 2004;16:239–46.
11. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. Role of histone H3 lysine 27 methylation in Polycombgroup silencing. *Science* 2002;298:1039–43.
12. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 2002; 16: 2893-905.

13. Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, Farnham PJ. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev* 2004; 18: 1592-605.
14. Raaphorst FM, van Kemenade FJ, Blokzijl T, Fieret E, Hamer KM, Satijn DP, Otte AP, Meijer CJ. Coexpression of BMI-1 and EZH2 polycomb group genes in Reed-Sternberg cells of Hodgkin's disease. *Am J Pathol* 2000; 157: 709-15.
15. Beà S, Tort F, Pinyol M, Puig X, Hernández L, Hernández S, Fernandez PL, van Lohuizen M, Colomer D, Campo E. BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. *Cancer Res* 2001; 61: 2409-12.
16. Van Kemenade FJ, Raaphorst FM, Blokzijl T, Fieret E, Hamer KM, Satijn DP, Otte AP, Meijer CJ. Coexpression of BMI-1 and EZH2 polycombgroup proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood* 2001;97:3896–901.
17. Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* 2003;423:255–60.
18. Visser HP, Gunster MJ, Kluin-Nelemans HC, Manders EM, Raaphorst FM, Meijer CJ, Willems R, Otte AP. The Polycomb group protein EZH2 is upregulated in proliferating, cultured human mantle cell lymphoma. *Br J Haematol* 2001;112:950–8.
19. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RGAB, Otte AP, Rubin MA, Chinnaiyan AM. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002;419:624–9.
20. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, Ghosh D, Sewalt RGAB, Otte AP, Hayes DF, Sabel MS, Livant D, Weiss SJ, Rubin MA, Chinnaiyan AM. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 2003;100:11606–11.
21. Tokimasa S, Ohta H, Sawada A, Matsuda Y, Kim JY, Nishiguchi S, Hara J, Takihara Y. Lack of the Polycomb-group gene rae28 causes maturation arrest at the early B-cell developmental stage. *Exp Hematol* 2001;29:93–103.
22. Kirmizis A, Bartley SM, Farnham PJ. Identification of the polycomb group protein SU(Z)12 as a potential molecular target for human cancer therapy. *Mol Cancer Ther* 2003;2:113–21.
23. Nakao M, Minami T, Ueda Y, Sakamoto Y, Ichimura T. Epigenetic system: a pathway to malignancies and a therapeutic target. *Int J Hematol* 2004;80:103–107.
24. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004;429:457–63.

25. Raaphorst FM, Vermeer M, Fieret E, Blokzijl T, Dukers D, Sewalt RG, Otte AP, Willemze R, Meijer CJ. Site-specific expression of Polycombgroup genes encoding the HPC-HPH/PRC1 complex in clinically defined primary nodal and cutaneous large B-cell lymphomas. *Am J Pathol* 2004;164:533–42.
26. Paro R, Hogness DS. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc Natl Acad Sci U S A* 1991;88:263–7.
27. Min J, Zhang Y, Xu R-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* 2003;17:1823–8.
28. Jones DO, Cowell IG, Singh PB. Mammalian chromodomain proteins: their role in genome organisation and expression. *Bioessays* 2000;22:124–37.
29. Satijn DP, Otte AP. Polycomb group protein complexes: do different complexes regulate distinct target genes? *Biochim Biophys Acta* 1999;1447:1–16.
30. Satijn DP, Gunster MJ, van der Vlag J, Hamer KM, Schul W, Alkema MJ, Saurin AJ, Freemont PS, van Driel R, Otte AP. RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. *Mol Cell Biol* 1997;17:4105–13.
31. Saurin AJ, Shiels C, Williamson J, Satijn DP, Otte AP, Sheer D, Freemont PS. The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol* 1998;142:887–98.
32. Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Mol Cell Biol* 2006;26:2560-9.
33. Hedinger C, Williams ED, Sobin LH. The WHO histological classification of thyroid tumours: a commentary on the second edition. *Cancer* 1989;63:908–11; Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat Rev Cancer* 2006; 6:292–306.
34. Parcellier A, Schmitt E, Gurbuxani S, Seigneurin-Berny D, Pance A, Chantome A, Plenchette S, Khochbin S, Solary E and Garrido C. HSP27 Is a Ubiquitin-Binding Protein Involved in I- κ B α Proteasomal Degradation. *Mol Cell Biol*. 2003, 23: 5790–5802.
35. Loyer P, Trembley JH, Grenet JA, Busson A, Corlu A, Zhao W, Kocak M, Kidd VJ and Lahti JM. Characterization of Cyclin L1 and L2 Interactions with CDK11 and Splicing Factors. *J Biol Chem*, 2008, 283: 7721–7732.
36. Liu X, Yang WC, Gao Q, Regnier F. Toward chromatographic analysis of interacting protein networks. *J Chromatogr A*. 2008, 1178:24-32.
37. Tanese N. Small-scale density gradient sedimentation to separate and analyze multiprotein complexes. *Methods*. 1997, 12: 224-234.

38. Vogelmann R, Nelson WJ. Fractionation of the Epithelial Apical Junctional Complex: Reassessment of Protein Distributions in Different Substructures. *Mol Biol Cell*. 2005, 16: 701–716.
39. Dörr J, Kartarius S, Götz C, Montenarh M. Contribution of the individual subunits of protein kinase CK2 and of hPrp3p to the splicing process. *Mol Cell Biochem*. 2008, 316: 187-193.
40. Schägger H, von Jagow G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem*. 1991, 199:223-231.
41. Camacho-Carvajal MM, Wollscheid B, Aebersold R, Steimle V, Schamel WW. Two-dimensional Blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach. *Mol Cell Proteomics*. 2004, 3: 176-182.
42. Krause F. Detection and analysis of protein-protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes. *Electrophoresis*. 2006, 27: 2759-2781.
43. Selbach M, Mann M, Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). *Nat. methods* 3, 981-983 (2006).
44. Castle JD. Purification of organelles from mammalian cells. *Curr Protoc Immunol*. 2003, Chapter 8:Unit 8.1B.
45. Chang J, Ruiz V, Vancura A. Purification of yeast membranes and organelles by sucrose density gradient centrifugation. *Methods Mol Biol*. 2008;457:141-9.
46. Holaska JM, Wilson KL. An Emerin “Proteome”: Purification of Distinct Emerin-Containing Complexes from HeLa Cells Suggests Molecular Basis for Diverse Roles Including Gene Regulation, mRNA Splicing, Signaling, Mechanosensing, and Nuclear Architecture. *Biochemistry*. 2007, 46: 8897-8908.
47. Gavin AC, Aloy P, Grandi P, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature*. 2006, 440: 631-636
48. Glozak MA, Seto E. Histone deacetylases and cancer. *Oncogene* 2007;26:5420–32.
49. Kruh J. Effects of sodium butyrate, a new pharmacological agent, on cells in culture. *Mol Cell Biochem* 1982;42:65–82.
49. Rada-Iglesias A, Enroth S, Ameer A, et al. Butyrate mediates decrease of histone acetylation centered on transcription start sites and down-regulation of associated genes. *Genome Res* 2007;17:708–19.
50. Peinado H, Ballestar E, Esteller M, Cano A. Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex. *Mol Cell Biol* 2004;24:306–19.

51. Thiery JP. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2002;2:442–54.
52. Thiery JP, Sleeman J. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol* 2006;7:131–42.
53. Ou JN, Torrisani J, Unterberger A, et al. Histone deacetylase inhibitor trichostatin A induces global and gene-specific DNA demethylation in human cancer cell lines. *Biochem Pharmacol* 2007; 73:1297–307.
54. Peinado H, Portillo F, Cano A. Transcriptional regulation of cadherins during development and carcinogenesis. *Int J Dev Biol* 2004; 48: 365-75.
55. Monti M, Orrù S, Pagnozzi D, Pucci P. Interaction proteomics. *Biosci Rep.* 2005; 25: 45-56.
56. Mick R, Ziober BL. c-Met expression in tall cell variant papillary carcinoma of the thyroid. *Cancer* 2003;98:1386–93.
57. Maertens, G.N., S. El Messaoudi-Aubert, T. Racek, J.K. Stock, J. Nicholls, M. Rodriguez-Niedenfuhr, J. Gil, and G. Peters. 2009. Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. *PLoS One.* 4:e6380.
58. Simon, J.A., and R.E. Kingston. 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol.*
59. Gearhart, M.D., C.M. Corcoran, J.A. Wamstad, and V.J. Bardwell. 2006. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol Cell Biol.* 26:6880-9.
60. Lagarou, A., A. Mohd-Sarip, Y.M. Moshkin, G.E. Chalkley, K. Bezstarosti, J.A. Demmers, and C.P. Verrijzer. 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.* 22:2799-810.
61. Allemand, E., E. Batsche, and C. Muchardt. 2008. Splicing, transcription, and chromatin: a menage a trois. *Curr Opin Genet Dev.* 18:145-51.
62. Listerman, I., A.K. Sapra, and K.M. Neugebauer. 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol.* 13:815-22.
63. Pandya-Jones, A., and D.L. Black. 2009. Co-transcriptional splicing of constitutive and alternative exons. *Rna.* 15:1896-908.
64. Wang, E.T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, and C.B. Burge. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 456:470-6.

- 65.** Tilgner, H., C. Nikolaou, S. Althammer, M. Sammeth, M. Beato, J. Valcarcel, and R. Guigo. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol.* 16:996-1001.
- 66.** Schwartz, S., E. Meshorer, and G. Ast. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol.* 16:990-5.