PROTEIN-PROTEIN INTERACTIONS: CHARACTERIZATION AND NEW STRATEGY FOR INDUSTRIAL APPLICATIONS

Gabriella Leo

Dottorato in Scienze Biotecnologiche – XIII ciclo
Indirizzo Biotecnologie Industriale e Molecolare
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XIII ciclo
Indirizzo Biotecnologie Industriale e Molecolare
Università di Napoli Federico II

PROTEIN-PROTEIN INTERACTIONS: CHARACTERIZATION AND NEW STRATEGY FOR INDUSTRIAL APPLICATIONS

Gabriella Leo

Dottoranda: Gabriella Leo

Relatore: Prof.ssa Leila Birolo

Coordinatore: Prof. Giovanni Sannia

Science can only ascertain what is,
but not what should be.

Albert Einstein

INDEX

Summary

Biological systems are made up of very large numbers of different components interacting at various scales. Most genes, proteins and other cell components carry out their functions within a complex network of interactions and a single component can affect a wide range of other components. Interactions involved in biological processes have been previously characterized individually, but this ‘‘reductionist’’ approach suffers from a lack of information about time, space, and context in which the interactions occur *in vivo*. A global, integrative, approach has been developed for several years, focusing on the building of protein-protein interaction maps (interactomes). These interaction networks are complex systems, where new properties arise. This is part of an emergent field, called systems biology which is ‘‘the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life’’ (http://www.systemsbiology.org/). This interdisciplinary approach, involving techniques from the mathematical, computational, physical and engineering sciences is required to understand complex networks. Detailed knowledge of protein interactions and protein complex structures is therefore fundamental to understand how individual proteins function within a complex and how the complex functions as a whole.

This PhD thesis targeted the characterization by structural and functional proteomics approaches of selected protein complexes involved in relevant molecular processes and can have an impact on their industrial applications. The following systems of biotechnological interest have been analysed in this study: **1) Phenol Hydroxylase (PH) from *Pseudomonas sp*. OX1.** It is a bacterium endowed with the ability to grow on a wide spectrum of hydroxylated and non-hydroxylated aromatic compounds for this reason it can be use in bioremediation of contaminated environments. In this context the research has been aimed at characterizing the endogenously expressed Phenol Hydroxylase (PH) complex from *Pseudomonas sp*. OX1, using biochemical approaches integrated to mass spectrometry analysis. Moreover, in order to define the protein-protein interaction networks of this complex a functional proteomic approach was carried out. **2) Characterization of the complex between pLG72 and human D-amino acid oxidase (hDAAO), linked to the onset of schizophrenia.** The molecular basis of schizophrenia is still elusive and current treatments focus on eliminating the symptoms of the disease. The clarification of the role of this specific complex and the identification of other physiological or conditional partners in this biochemical process will be useful in developing molecules that can be effective in the disease treatment.

Moreover, as a parallel goal, and strictly connected with the former issue, I worked at **3) the development of an innovative structural analysis approach based on the use of a femtosecond UV laser as a novel zero length protein-protein cross-linker.** Such photo-physical approach could obviate many of the problems associated with standard chemical cross-linking reagents, and put cross-linking in a proteome-wide position for the characterizations of protein-protein interactions *in vivo* in intact cells, the study of the transient interactions among proteins, mapping molecular interfaces in protein complexes, providing information on the dynamics of the contacts within a multi-protein complex.

Riassunto

I sistemi biologici sono costituiti da un elevato numero di diverse componenti che interagiscono tra loro su vari livelli e che svolgono la propria funzione all'interno di una complessa rete di interazioni. La disciplina volta a studiare gli organismi biologici nella loro interezza è la *systems biology* che, focalizzandosi sull'intero sistema di geni e prodotti genici, ha l'obiettivo di comprendere gli effetti delle interazioni tra le componenti di un dato sistema e come quest'ultimo interagisca con il suo ambiente [1]. Affinché ciò sia possibile è necessario conoscere i meccanismi molecolari che sono alla base di tali interazioni.
Tra le diverse macromolecole, presenti nella cellula, le proteine sono le più versatili e sono coinvolte nella maggior parte dei processi cellulari. Le proteine svolgono il loro ruolo biologico grazie alla particolare conformazione che assumono. La possibilità di subire cambiamenti conformazionali, modifiche post-traduzionali e di modulare le interazioni proteina-proteina, conferisce a queste macromolecole la capacità di adattarsi rapidamente ai bisogni cellulari. La capacità delle proteine di assemblarsi in diversi complessi proteici all'interno dell'ambiente cellulare, consente inoltre a queste macromolecole di svolgere ruoli funzionali diversi.
Negli ultimi anni è cresciuto l'interesse da parte della comunità scientifica per lo studio della funzione biologica delle proteine e la comprensione, a livello molecolare, dei meccanismi cellulari in cui sono coinvolte. Le proteine, infatti, non agiscono da sole ma si è stimato che oltre l'80% di esse è organizzato in complessi. Infatti, la maggior parte dei meccanismi molecolari richiede l'azione coordinata di un elevato numero di proteine che si assemblano in complessi multiproteici. Inoltre, la quasi totalità dei processi biologici è diretta e regolata da una cascata dinamica di segnali che comporta interazioni transienti proteina-proteina.
Siccome il ruolo di una proteina riflette la fitta rete di interazioni che è in grado di formare, è possibile inoltre ottenere informazioni sulla funzione di una molecola proteica anche attraverso la caratterizzazione dei suoi interattori molecolari.
Studiare l'eterogeneità delle proteine, sia da un punto di vista delle interazioni che esse instaurano all'interno della cellula, che della loro funzione, sta ricoprendo un ruolo sempre più importante. Ciò è fondamentale non solo per comprendere come sia evoluta la loro funzione nel tempo, o quale sia il loro riconoscimento cellulare, ma anche per lo sviluppo delle biotecnologie. In questo settore, infatti, la natura versatile delle proteine può essere utilizzata per la progettazione di nuovi farmaci (settore biomedico) e per ottimizzare le applicazioni dell'ingegneria proteica (settore industriale), al fine di utilizzare le proteine in modo innovativo [2].
L'identificazione e la caratterizzazione strutturale dei complessi multiproteici è, quindi, di fondamentale importanza per lo studio della loro funzione biologica e della conoscenza dei più importanti meccanismi molecolari cellulari [3]. Lo studio della rete di interazioni che presiedono al funzionamento dei complessi proteici può essere importante, infatti, non solo dal punto di vista della conoscenza stessa, ma anche dal punto di vista applicativo, immaginando di poter intervenire nei processi di interesse biologico, bloccando o immettendo specifici segnali che intervengano sulle interazioni proteina-proteina, o anche modulando la composizione di tali complessi [4]. Sono state sviluppate numerose metodologie, come ad esempio il sistema del doppio ibrido, o tecniche di coimmunoprecipitazione, per identificare i *partner* proteici all'interno di un complesso e determinare se due o più proteine interagiscono *in vivo*. Tuttavia l'organizzazione strutturale delle subunità che costituiscono un complesso non è sempre accessibile con questi approcci e, al tempo stesso, l'identificazione di

interazioni transienti o caratterizzate da costanti di dissociazione alte, risulta estremamente difficile.
E' necessario, quindi, sviluppare tecniche o strategie di analisi che permettano lo studio di interazioni proteina-proteina su scala proteomica, e che abbiano cinetiche che consentano la caratterizzazione di interazioni transienti in contesti cellulari. L'integrazione di metodologie quali la proteolisi complementare, il *cross-link*, già patrimonio della più classica chimica delle proteine, con le tecniche avanzate di spettrometria di massa può offrire un valido apporto in campi quali: i) la validazione di modelli di struttura, ii) l'analisi strutturale e delle variazioni conformazionali in sistemi dinamici, iii) lo studio delle interazioni proteina-proteina.
La finalità del presente progetto di dottorato è stata l'applicazione di metodiche di biochimica classica integrata a tecniche di spettrometria di massa e lo sviluppo di metodologie e strategie atte a caratterizzare e identificare complessi proteici *in vitro* e *in vivo*, in sistemi molecolari di interesse biotecnologico che non si prestavano ad analisi ad alta risoluzione. In particolare la mia attività di ricerca è stata incentrata su i) lo studio strutturale del complesso enzimatico della fenolo ossidrilasi da *Pseudomonas sp*. OX1, ii) la caratterizzazione del complesso hDAAO-pLG72 coinvolto nella patologia schizofrenica. iii) Parallelamente, la mia attività di ricerca è stata volta alla messa a punto di una metodologia innovativa che utilizza un laser pulsato a luce ultravioletta per indurre *cross-link* proteina-proteina.

Sistema della fenolo ossidrilasi da *Pseudomonas sp.* OX1.

Il complesso enzimatico della fenolo ossidrilasi da *Pseudomonas sp*. OX1 appartiene alla famiglia delle monoossigenasi batteriche ed è in grado di metabolizzare diversi composti aromatici ad elevato impatto ambientale; riveste pertanto un particolare interesse nel biorisanamento e nelle applicazioni di sintesi industriale [5,6]. Tale ceppo batterico è stato isolato da fanghi attivi di scarichi industriali ed è capace di crescere utilizzando come unica fonte di carbonio e di energia diverse molecole aromatiche tra cui benzene, fenolo, toluene, *o*-xilene, 2,3 e 3,4-dimetilfenolo e i cresoli. Nel genoma di *P. sp.* OX1 è presente una regione *ph* (*phenol hydroxylase*) che codifica per un'attività fenolo ossidrilasica, che catalizza reazioni di monossigenazione dell'anello aromatico. Nel *locus* genico *ph*, sono state inizialmente identificate cinque ORFs (*open reading frames*) codificanti per le cinque subunità costituenti il complesso, denominate PHL, PHM, PHN, PHO, PHP. La subunità PHP, un'ossidoreduttasi contenente FAD e un cofattore ferredossinico del tipo [2Fe-2S], rappresenta il primo accettore di elettroni nella catena di trasporto. Da tale subunità gli elettroni sono trasferiti al subcomplesso dimerico $PH(LNO)_2$ che rappresenta la componente ossidrilasica del sistema. La subunità PHM ha attività regolatoria e, nonostante non sia ancora completa una dettagliata caratterizzazione del suo ruolo, la sua presenza è indispensabile per l'attività catalitica dell'intero complesso PH. Recentemente, a monte del gene codificante per la subunità PHL, è stata identificata una sesta *orf*, denominata *phK* omologa alla proteina DmpK della fenolo ossidrilasi espressa dal batterio *Pseudomonas* CF600. E' importante sottolineare che la delezione del gene *phk* non consente ai ceppi di CF600 di crescere utilizzando il fenolo come unica fonte di carbonio e di energia, indicando il ruolo fondamentale che la subunità riveste nell'ambito del complesso, sebbene DmpK non sia indispensabile ai fini dell'attività catalitica [7]. È stata dunque avanzata l'ipotesi che DmpK, pur non partecipando direttamente agli eventi catalitici che portano alla ossidrilazione del substrato, rivesta un ruolo fondamentale durante l'assemblaggio del sito attivo della componente ossidrilasica.

Con il mio lavoro, in collaborazione con il gruppo del Prof. Di Donato del dipartimento di Biologia Strutturale e Funzionale dell'Università di Napoli "Federico II", ho contribuito a caratterizzare il ruolo funzionale di PHK.
Esprimendo il *cluster* genico della fenolo ossidrilasi in *E. coli* in presenza dell'*orf* codificante per PHK, si è osservato un aumento dell'attività ossidrilasica, suggerendo quindi un effetto positivo da parte della proteina accessoria sull'attività enzimatica della fenolo ossidrilasi.
Sono stati condotti diversi esperimenti incubando il complesso della fenolo ossidrilasi $PH(LNO)_2$ in presenza delle singole componenti purificate PHK e PHM, ed in presenza o in assenza di ferro, analizzando i prodotti mediante cromatografia ad esclusione molecolare integrata a tecniche di spettrometria di massa.
I dati ottenuti, parallelamente alla caratterizzazione enzimatica condotta presso il gruppo del Prof. Di Donato, hanno contribuito a formulare un modello caratterizzato da eventi molecolari in cui è evidenziato il possibile ruolo di PHK. Il modello prevede che PHK si lega inizialmente, in modo transiente, al complesso apo-$PH(LNO)_2$ favorendo l'inserimento di ferro intracellulare all'interno del sito attivo e portando alla formazione della specie holo-$PH(LNO)_2$. In seguito alla dissociazione di PHK si ha il legame della componente regolativa PHM che sembra proteggere il ferro all'interno del sito attivo, evitandone il rilascio, come è anche possibile osservare dalla struttura cristallografica. Una volta completato il ciclo di ossidrilazione, PHM si dissocia in modo da permettere alla reduttasi PHP di legarsi all'esamero e iniziare un nuovo un ciclo catalitico. La dissociazione di PHM permetterebbe il legame di PHK con la conseguente formazione del complesso inattivo PHK-PH(LNO), che si riorganizzerebbe a formare la specie apo-$PH(LNO)_2$.
A conferma che PHK possa formare *in vivo* un complesso con la fenolo ossidrilasi, è stato condotto un esperimento di *pull down* incubando un estratto cellulare di *Pseudomonas sp.* OX1 con PHK, precedentemente biotinilata, immobilizzata su resina di avidina. L'analisi degli interattori molecolari effettuata mediante spettrometria di massa ha mostrato che PHK interagisce in modo specifico con il complesso ossidrilasico $PH(LNO)_2$.
I dati ottenuti sono stati recemente pubblicati: Izzo V, et al (2010) Archives of Biochemistry and Biophysics.

Caratterizzazione del complesso hDAAO-pLG72 coinvolto nella patologia schizofrenica.

Altro oggetto di studio della mia tesi di dottorato è stata l'analisi strutturale e funzionale del complesso costituito dalle proteine hDAAO e pLG72, coinvolto nella patologia schizofrenica. Il ruolo fisiologico della DAAO nel cervello è di controllare i livelli di D-serina, un attivatore allosterico del recettore N-metil-D-aspartato (NMDA) [8]. La degradazione della D-serina, causata dalla DAAO, impedisce il suo legame al recettore NMDA, necessario per la trasmissione dell'impulso nervoso [9]. Si ipotizza che un'ipofunzionalità dei recettori NMDA potrebbe essere l'origine molecolare di alcuni sintomi della schizofrenia [10]. L'associazione tra alterazione del metabolismo della D-serina e la schizofrenia è stata dimostrata in studi che hanno evidenziato la diminuzione dei livelli di D-serina nel liquido cerebrospinale e nel siero di pazienti schizofrenici [11]. Studi recenti [12] hanno inoltre collegato l'insorgenza della patologia schizofrenica alla formazione di un complesso di DAAO con pLG72, proteina di cui non è nota la struttura e la funzione.
Recenti studi *in vitro* condotti dal gruppo del Prof. Pollegioni, dell'università dell'Insubria di Varese, con cui questo progetto è condotto in collaborazione, hanno

evidenziato la formazione di un complesso di circa 200 kDa costituito da due omodimeri di hDAAO (80kDa x 2) e due molecole di pLG72 (20kDa x 2), con una K_D di formazione del complesso di circa 8 x 10^{-6}M [13]. Inoltre esperimenti *in vivo* di co-immunoprecipitazione e immunofluorescenza hanno dimostrato che entrambe le proteine sono espresse nello stesso tipo cellulare, gli astrociti della corteccia cerebrale umana, e presentano una parziale colocalizzazione subcellulare.
La caratterizzazione di questo complesso, oltre a fornire dettagli molecolari sull'insorgenza della malattia, è fondamentale per la progettazione di farmaci mirati alla cura della patologia, piuttosto che al trattamento dei soli sintomi.
La caratterizzazione strutturale del complesso è stata, dapprima, condotta con esperimenti di proteolisi limitata e complementare. Gli esperimenti sono stati condotti seguendo nel tempo la cinetica di rilascio dei vari peptidi e sono stati impiegati enzimi proteolitici a diversa specificità di substrato.
Inoltre, l'indagine è stata condotta sulle proteine hDAAO e pLG72 isolate e sul complesso in modo da poter successivamente confrontare la topologia dei siti di idrolisi preferenziali e quindi definire le regioni delle proteine coinvolte in variazioni conformazionali indotte dalla formazione del complesso.
Dal confronto della topologia della proteina hDAAO in assenza e in presenza della proteina pLG72 è risultato evidente che la formazione del complesso provoca profonde modifiche conformazionali. La prima osservazione generale è che hDAAO quando è in complesso con pLG72 è molto più accessibile all'azione delle proteasi, infatti, sono state utilizzate condizioni d'idrolisi molto più stringenti che per hDAAO isolata. Quest'osservazione è peculiare del complesso in questione, in quanto, solitamente si osserva una riduzione della flessibilità ed accessibilità globale delle singole molecole proteiche quando si associano a formare un complesso. La seconda osservazione è la sostanziale diversificazione dei siti d'idrolisi nei due casi. La regione C-terminale della proteina, estremamente accessibile in hDAAO isolata, risulta decisamente meno esposta nel complesso hDAAO-pLG72. La formazione del complesso induce cambiamenti strutturali radicali in altre regioni della proteina: i siti L189, R191, L194 e i siti E249, L250 localizzati in due diversi *loop* della molecola non sono più accessibili in presenza di pLG72. Questa variazione è accompagnata dalla concomitante esposizione di diversi residui, Y34, Y55, F167, N180, L296 e R297, distribuiti lungo la sequenza e non accessibili in assenza di pLG72.
E' possibile inoltre ipotizzare che la formazione del complesso renda la struttura di pLG72 meno esposta all'azione delle proteasi, poiché da un'analisi qualitativa preliminare delle cinetiche d'idrolisi, si può osservare una minore accessibilità in presenza di hDAAO.
Esperimenti di *cross-linking* chimico con un reattivo omobifunzionale, integrati ad analisi di spettrometria di massa, hanno fornito informazioni complementari individuando zone di contatto in hDAAO in corrispondenza di quelle regioni che risultavano protette all'azione delle proteasi negli esperimenti di proteolisi limitata descritti in precedenza.
Sulla base della struttura tridimensionale della hDAAO si può dunque concludere che i) i siti idrolizzati negli esperimenti si trovano tutti in prossimità del sito di legame del cofattore flavinico, ii) i siti di idrolisi nella proteina isolata sono tutti localizzati spazialmente in una ristretta regione della struttura, iii) si può ipotizzare che il legame con pLG72 determini la protezione dei precedenti siti e l'esposizione di nuove regioni della macromolecola, comunque tutte localizzate nelle immediate vicinanze spaziali della zona protetta.

Al fine di delucidare i meccanismi molecolari alla base della patologia schizofrenica, sono stati condotti esperimenti di proteomica funzionale per individuare nuovi interattori delle proteine in questione. A tale scopo hDAAO è stata espressa in cellule umane come proteina di fusione con uno specifico peptide (c-Myc), in modo da poterla purificare mediante cromatografia di affinità insieme alle proteine a lei complessate.
È in fase di stesura un manoscritto che riporta i risultati ottenuti per la caratterizzazione strutturale del complesso hDAAO-pLG72.

Utilizzo di un laser ad impulsi ultrabrevi nella regione dell'UV per lo studio delle interazioni proteina-proteina.

La terza parte del mio progetto di dottorato è consistita nella messa a punto di una metodologia innovativa, che prevede l'utilizzo di un laser UV ultracorto per indurre *cross-link* proteina-proteina. L'uso di un laser pulsato con una durata di alcune decine di femtosecondi (fs), unitamente ad un dispositivo che modella l'impulso in modo specifico, può fornire numerosi vantaggi rispetto alle strategie convenzionali di *cross-link* delle proteine. Infatti, la scala temporale degli impulsi laser ultrabrevi, (intervalli nell'ordine dei nano- pico o anche femto- secondi) che è paragonabile e compatibile con il tempo di emivita delle interazioni transienti proteina-proteina, consente di investigare i contatti transienti, la dinamica molecolare delle proteine, la flessibilità conformazionale e la cinetica delle interazioni. Ulteriore e non minore vantaggio è la prospettiva di utilizzare questa metodologia per ottenere informazioni sulla dinamica dei *network* di complessi proteici *in vivo*, all'interno della cellula, senza bisogno di ingegnerizzazioni della cellula in esame. Questa parte del progetto è stata condotta in collaborazione con il gruppo del Prof. Velotta ed il Dr. Altucci del Dipartimento di Fisica, il gruppo della Prof.ssa Piccoli del Dipartimento di Biologia Strutturale e Funzionale e il gruppo della Prof.ssa Costello della Boston University.
In letteratura sono presenti studi di questo tipo rivolti soltanto al sistema acidi nucleici-proteine [14], in cui è evidenziato che questo tipo di meccanismo coinvolge gruppi in grado di assorbire la radiazione emessa dal laser. Inoltre la caratterizzazione del processo è incentrata esclusivamente sulle basi azotate e non sulle proteine.
Il punto di partenza di questo studio sono stati i peptidi, un sistema più semplice rispetto alle proteine. Sono stati condotti esperimenti utilizzando uno spettro di peptidi (xenopsina, angiotensina, interleuchina) che si differenziava per il contenuto in aminoacidi. Il sistema modello è stato irraggiato per tempi molto brevi (da 1 sec a 1 min) utilizzando un'energia del laser di 110 µJ, ad una frequenza di 2 kHz e ad una lunghezza d'onda di 267 nm. Mediante analisi di massa MALDI-TOF è stato possibile osservare in prima analisi che, a seguito dell'irraggiamento, solo quei peptidi contenenti aminoacidi aromatici davano come prodotto specie dimeriche costituite da due peptidi covalentemente legati, dimostrando dunque che è possibile indurre la formazione di *cross-link* per irraggiamento con laser UV ultracorto. Al fine di localizzare gli aminoacidi coinvolti nel legame, la specie d'interesse è stata sottoposta ad analisi di spettrometria di massa *tandem*, ed è stato possibile individuare quali aminoacidi coinvolti nel legame indotto dal laser, il Trp della xenopsina e la Tyr dell'angiotensina, per l'etero dimero xenopsina-angiotensina. Analogamente il Trp è anche risultato coinvolto nel *cross-link* dell'omodimero xenopsina-xenopsina.

Il passo successivo è stato studiare il meccanismo molecolare alla base di un processo di questo tipo. Partendo dall'ipotesi che il tipo di reazione comporta la formazione di specie radicaliche, abbiamo effettuato l'irraggiamento dei singoli peptidi in presenza di molecole quali 2-metile-2-nitrosopropano (MNP) e 5,5-dimetil-1-pirrolidin N-ossido (DMPO), dette *spin trap*, specie reattive verso i radicali [15]. Anche in questo caso per la xenopsina e per l'angiotensina era presente l'addotto con lo *spin trap*, assente invece per l'interleuchina (che non presenta aminoacidi aromatici), dimostrando dunque che a seguito dell'irraggiamento si generano radicali, verosimilmente sulla catena laterale di aminoacidi aromatici. I dati di spettrometria di massa *tandem* hanno confermato tali ipotesi, ovvero il legame dello *spin trap* è stato individuato sul Trp per la xenopsina e sulla Tyr per l'angiotensina.
Messo a punto la metodologia, abbiamo aumentato la complessità del sistema utilizzando come proteina modello l'alcool deidrogenasi (ADH) da *Saccaromyces cerevisiae,* proteina omo-tetramerica e contenente diversi aminoacidi aromatici. Variando differenti parametri, quali energia del laser, tempo di irraggiamento e concentrazione proteica, è stato possibile osservare, mediante elettroforesi denaturante, la formazione di aggregati ad alto peso molecolare assenti nella rifermento non irraggiato. Considerando la natura radicalica del meccanismo di reazione, abbiamo aggiunto alla nostra miscela di analisi un antiossidante, l'acido ascorbico, in concentrazione variabile, con lo scopo di ottenere bande discrete e una minore eterogeneità del campione. Le migliori condizioni utilizzate sono risultate, energia del laser 220 µJ e aggiunta di acido ascorbico ad una concentrazione finale di 5 mM dopo 1 min di esposizione. Mediante elettroforesi denaturante è stato, infatti, possibile osservare per i campioni irraggiati la comparsa di specie ad un peso molecolare corrispondente al dimero di ADH. La banda di interesse è stata escissa dal gel, idrolizzata *in situ* con tripsina ed analizzata mediante spettrometria di massa MALDI-TOF. Nello spettro di massa della miscela peptidica della specie irraggiata è stato possibile osservare dei segnali assenti nel riferimento ed imputabili a peptidi coinvolti in legami covalenti indotti dal laser.
Recentemente sono stati ottenuti risultati preliminari su cellule intatte. Abbiamo, infatti, esteso la metodologia a cellule umane esprimenti la proteina ricombinante ApoA-I, con l'obiettivo di isolare gli interattori della proteina di interesse. Le cellule risospese in un tampone a pH fisiologico sono state irraggiate e in seguito lisate. L'estratto cellulare è stato analizzato mediante *western blotting* utilizzando l'anticorpo specifico per ApoA-I ed è stato possibile rivelare, ottimizzate le condizioni di reazione, oltre alla proteina ricombinante, delle specie ad elevato peso molecolare, imputabili alla proteina complessata ai suoi interattori.
Obiettivo futuro sarà quello di purificare mediante cromatografia di immunoaffinità gli interattori della proteina di interesse, dall'estratto cellulare di cellule irraggiate. Sarà così possibile identificare quei *partner* proteici che *in vivo* interagiscono in modo transiente, avendo congelato mediante UV *cross-link* anche le interazioni più labili.
Gli spettri di frammentazione ad alta risoluzione sono stati ottenuti durante la permanenza presso il gruppo della Prof.ssa Costello della Boston University, School of Medicine, Mass Spectrometry Resource An NIH-Supported Resource Center.
Si ringrazia il Programma di scambi internazionali promosso dall'Università di Napoli "Federico II", perché fonte di finanziamento durante la permanenza presso il gruppo della Prof.ssa Costello.
È in corso di stesura un manoscritto che riporta i risultati ottenuti sui peptidi.

Bibliografia

[1] Villoslada P, Steinman L, Baranzini SE. Systems biology and its application to the understanding of neurological diseases. *Ann Neurol*. (2009);65(2):124-39. Review.

[2] Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for Biotechnology nature biotechnology. *Nat Biotechnol*. (2009); 27(2):157-67.

[3] Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol*. (2007);8(8):645-54. Review.

[4] Trakselis MA, Alley SC, Ishmael FT. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjug Chem*. (2005);16(4):741-50. Review.

[5] Sazinsky MH, Dunten PW, McCormick MS, Di Donato A, Lippard SJ. X-ray structure of a hydroxylase-regulatory protein complex from a hydrocarbon-oxidizing multicomponent monooxygenase, *Pseudomonas sp*. OX1 phenol hydroxylase. *Biochemistry*. (2006);45(51):15392-404.

[6] Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, Pucci P, Di Donato A. *Appl Environ Microbiol.* (2004);70(4): 2211-9.

[7] Powlowski J, Sealy J, Shingler V, CIDieux E. On the role of DmpK, an auxiliary protein associated with multicomponent phenol hydroxylase from Pseudomonas sp. strain CF600. *J Biol Chem*. (1997)10;272(2):945-51.

[8] Schell MJ, Molliver ME, Snyder SH. D-serine, an endogenous synaptic modulator: localization to astrocytes and glutamate-stimulated release. *Proc Natl ACID Sci U S A.* (1995);92(9):3948-52.

[9] Mothet JP, Parent AT, Wolosker H, Brady Jr RO, Linden DJ, Ferris CD, Rogawski MA, Snyder SH. D-serine is an endogenous ligand for the glycine site of the N-methyl-D-aspartate receptor. *Proc. Natl. ACID. Sci. U.S.A.* (2000);97:4926-31.

[10] Hashimoto K, Fukushima T, Shimizu E, Komatsu N, Watanabe H, Shinoda N, Nakazato M, Kumakiri C, Okada S, Hasegawa H, Imai K, Iyo M. Decreased serum levels of D-serine in patients with schizophrenia: evidence in support of the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia. *Arch Gen Psychiatry.* (2003);60(6):572-6.

[11] Tsai G, Yang P, Chung LC, Lange N, Coyle JT. d-Serine added to antipsychotics for the treatment of schizophrenia. *Biol Psychiatry*. (1998);44:1081-89.

[12] Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, Barry C, Tanaka H, La Rosa P, Puech A, Tahri N, Cohen-Akenine A, Delabrosse S, Lissarrague S, Picard FP, Maurice K, Essioux L, Millasseau P, Grel P, Debailleul V, Simon AM, Caterina D, Dufaure I, Malekzadeh K, Belova M, Luan JJ, Bouillot M, Sambucy JL, Primas G, Saumier M, Boubkiri N, Martin-Saumier S, Nasroune M, Peixoto H, Delaye A, Pinchot V, Bastucci M, Guillou S, Chevillon M, Sainz-Fuertes R, Meguenni S, Aurich-Costa J, Cherif D, Gimalac A, Van Duijn C, Gauvreau D, Ouellette G, Fortier I, Raelson J, Sherbatich T, Riazanskaia N, Rogaev E, Raeymaekers P, Aerssens J, Konings F, Luyten W, Macciardi F, Sham PC, Straub RE, Weinberger DR, Cohen N, Cohen D. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl ACID Sci U S A.* (2002);99(21):13675-80.

[13] Sacchi S, Bernasconi M, Martineau M, Mothet JP, Ruzzene M, Pilone MS, Pollegioni L, Molla G. pLG72 modulates intracellular D-serine levels through its interaction with D-amino acid oxidase: effect on schizophrenia susceptibility. *J Biol Chem* (2008);283:22244-56.

[14] Russmann C, Stollhof J, Weiss C, Beigang R, Beato M. Two wavelength femtosecond laser induced DNA-protein crosslinking. *Nucleic Acids Res.* (1998);26(17):3967-70.

[15] Davies MJ, Hawkins CL. EPR spin trapping of protein radicals. *Free Radic Biol Med.* (2004);36(9):1072-86. Review.

I. Introduction

I.1 Systems biotechnology for strain improvement

The indispensable role of biotechnology is increasing in nearly every industry, including the healthcare, pharmaceutical, chemical, food and agricultural industries, exponentially with the identification of organisms, isolation of novel compounds and their pathways, and the molecular characterization of cellular components. Production of large-volume low-value bioproducts requires the development of lower-cost and higher-yield processes. Towards this goal, improved microorganisms have traditionally been developed through random mutagenesis followed by screening processes [1]. Rational metabolic and cellular engineering approaches have also been successful in improving strain performance in several cases. However, such attempts were limited to the manipulation of only a handful of genes encoding enzymes and regulatory proteins selected using available information and research experience.

Recent advances in high-throughput experimental techniques supported by bioinformatics have resulted in rapid accumulation of a wide range of omics data at various levels, thus providing a foundation for in-depth understanding of biological processes. As DNA sequencing has become faster and cheaper the genome sequences of many microorganisms have been completed and many more are in progress. With the complete genome sequence, post-genomic research (the 'omics' fields) is increasing rapidly. New technologies have recently become available, which allow a thoroughly high-throughput assessment of changes at gene (genomics) and protein (proteomics) levels involved in determining productivity in different environmental conditions, and establish functional relationships between cellular organization and productivity. Considering that most cellular metabolic activities are directly or indirectly mediated by proteins, proteome profiling takes us one step further towards understanding cellular metabolic status and for deciphering the functions and characteristics of biological systems [2,3]. The central task of systems biotechnology is to comprehensively collect global cellular information, such as omics data, and to combine these data through metabolic, signalling and regulatory networks to generate predictive computational models of the biological systems.

I.2 Proteins as molecular machinery

The past decade has experienced a huge leap in the amount of data generated from different areas of life sciences and has seen the complete sequencing of several eukaryotic genomes. One of these advancements was the completion of the Human Genome project in 2003 [4,5]. The new challenge is to understand the function of all the genes identified by the human genome project shifting the focus from the DNA to proteins.

Biological systems are made up of very large numbers of different components interacting at various scales. Most genes, proteins and other cell components carry out their functions within a complex network of interactions and a single component can affect a wide range of other components. Vital cellular functions such as DNA replication, transcription and mRNA translation require the coordinated action of a large number of proteins that are assembled into an array of multiprotein complexes

of distinct composition and structure. Similarly, biological processes are orchestrated and regulated by dynamic signalling networks of interacting proteins that link chemical or physical stimuli to specific effector molecules.
Since proteins compose most of the functional units in the cell, the complete understanding of their role in the cell is very critical in understanding how the cell functions. Proteomics focuses on understanding the many aspects of proteins that can involve their structures, modifications, localization and their protein-protein interactions [6].
The analysis of protein complexes and protein-protein interaction networks - and the dynamic behaviour of these networks as a function of time and cell state - is therefore of central importance in biological research. Interactions involved in biological processes have been first characterized individually, but this ‘‘reductionist’’ approach suffers from a lack of information about time, space, and context in which the interactions occur *in vivo*.
A global, integrative, approach has been developed for several years, focusing on the building of protein-protein interaction maps (interactomes). These interaction networks are complex systems, where new properties arise. An interactome is the whole set of molecular physical interactions between biological entities in cells and organisms and it is essential to understand how gene functions and regulations are integrated at the level of an organism. Indeed, many proteins mediate their biological function through protein interactions, which are involved in supramolecular assemblies (collagens, elastic fibers, actin filaments), in the building of molecular machines (molecular motors, ribosomes, proteasome) and in major biological processes such as immunity (antigen-antibody interaction), metabolism (enzyme-substrate interaction), signalling (interaction of messenger molecules, hormones, neurotransmitters with their cognate receptors), and gene expression (DNA-protein interactions). This is part of an emergent field, called systems biology which is ‘‘the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life’’ (http://www.systemsbiology.org/) (Fig.I.1). This interdisciplinary approach, involving techniques from the mathematical, computational, physical and engineering sciences is required to understand complex networks [7].
Furthermore, the sequencing of the genome and advances in proteomics lead to the identification of proteins of unknown functions [8]. Interaction networks might give clues on the functions of these newly discovered proteins or on new functions of already identified proteins.
Proteins interact with each other to form distinct stable and transient assemblies. Such non-covalent interactions are mediated by van der Waals, electrostatic and hydrophobic forces and hydrogen bonding. Stable protein-protein interactions are strong and irreversible, whereas a complex qualifies as transient if it readily undergoes changes in the oligomeric state.Transient complexes can be subdivided into weak and strong. Weak transient complexes show a dynamic mixture of different oligomeric states *in vivo*, whereas strong transient complexes change their quaternary state only when triggered by, for example, ligand binding. Weak transient interactions are characterized by a dissociation constant (K_D) in the micromolar range and lifetimes of seconds. Strong transient interactions, stabilized by binding of an effector molecule, may last longer and have a lower K_D in the nanomolar range [9].
The connectivity between several different proteins may give the protein complex a new functional role. Proteins inside the cell do not interact randomly. Protein associations need precise regulation. The spatial and temporal regulation of enzyme

activities through extensive interactions bears remarkable functional relevance. In human, mutations or environmental factors that interfere with protein-protein interaction lead to pathology. This is the case in the Immunodeficiency, Centromeric instability, Facial anomalies (ICF) syndrome, caused by defects in DNMT3B, a DNA methyltransferase [10].
A wide variety of modular and specialized binding domains have been mapped that mediate protein-protein recognition [11]. Even though the distinction is not always very strict, interaction domains are thought to generally operate through two mechanisms. Domains sometimes bind other domains. Typically, this type of interactions involves large binding interfaces provided by the rigid globular domains. This is illustrated by the interaction taking place between Ras and its GTPase activating protein Ras-GAP, which play a crucial role in regulating cellular signal transduction processes [12]. Domain-domain interactions are thus generally characterized by relatively high stabilities and affinities in the low nM to pM range [13,14]. In other cases, interactions are mediated by short sequences, typically between 3 and 10 amino acids, present in disordered parts of proteins. These short linear motifs are critical to many biological processes. They often show low affinities (0.5-10 µM) [15,16]. They tend to be mediators in transient interactions, such as the ones seen in cell signalling [17].

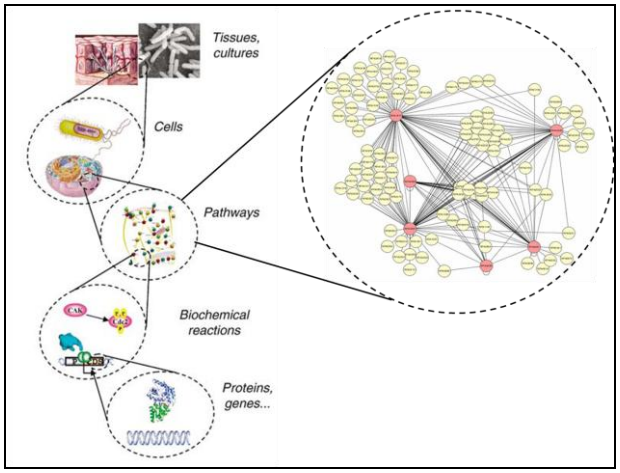


Fig.I.1. Biological systems hierarchy. Within the hierarchy, every constituent part is embedded in a more complex system that provides its context. Design of new behaviour occurs with the top of the hierarchy in mind but is implemented bottom-up. At the bottom of the hierarchy are DNA, RNA, proteins, and metabolites. The next layer, comprises biochemical reactions that regulate the flow of information and manipulate physical processes. At the module layer, the complex pathways that function like integrated circuits. The connection of these modules to each other and their integration into host cells allows the scientist to extend or modify the behaviour of cells in a programmatic fashion.

Various large-scale efforts have thus attempted to define protein interactomes in several organisms, including *Saccharomyces cerevisiae* [18], *Drosophila melanogaster* [19], *Caenorhabditis elegans* [20] and *Homo sapiens* [21,22].
Recent experimental evidence suggests that promiscuity, not only in interactions but also in the actual function of proteins, is not as rare as was previously thought. This has implications not only for our fundamental understanding of molecular recognition and how protein function has evolved over time but also in the realm of biotechnology (Fig.I.2). Understanding protein promiscuity is becoming increasingly important not only to optimize protein engineering applications in areas as diverse as synthetic biology and metagenomics but also to lower attrition rates in drug discovery

programs, identify drug interaction surfaces less susceptible to escape mutations and potentiate the power of polypharmacology [23].
Enhanced understanding of promiscuity can facilitate progress in protein engineering and drug design for industrial applications. The action of drugs relies on the promiscuous character of their protein targets, and their side effects relate to the promiscuity of unintended targets.

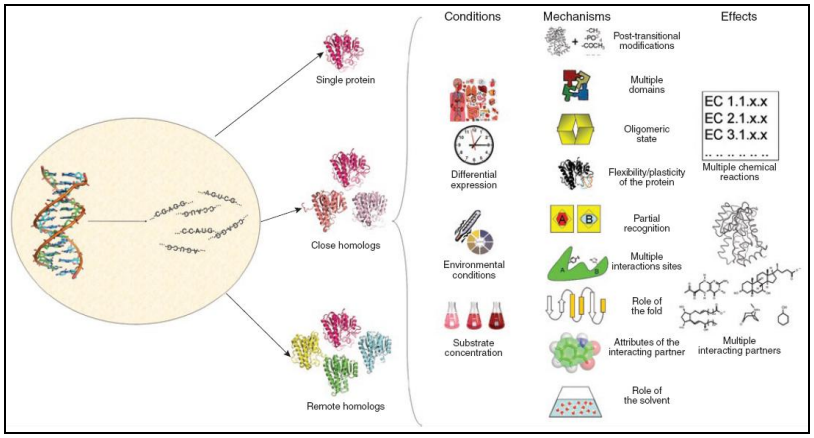


Fig.I.2. Schematic representation of protein promiscuity: levels, triggering conditions, molecular mechanisms and overall effects.

Thus, to design drugs less likely to have deleterious side effects and more likely to withstand resistance mutations, we need to understand how to exploit and manage a protein's tendency for binding promiscuity. In biotechnology, the promiscuous nature of proteins could also be exploited to evolve enzymes with different reaction/substrate specificities [24] or to use proteins in innovative ways (e.g., in the context of synthetic biology projects) [25], and the industrial exploitation of such and similar systems will greatly benefit from the knowledge and the molecular characterization of biotechnological relevant systems.

I.3 Biochemical approaches to map protein-protein interactions

The ultimate goal of functional proteomics is to decipher the molecular function of an entire cell by generating a construction master plan describing all molecular machines, their functions, their reactions to external stimuli and their interconnectivities. An interdisciplinary and community-wide effort will be required to realize this vision, even when limited to the characterization of a few cellular states.
Different approaches have been used to characterize protein complexes and protein-protein interaction networks.
The first interactome maps were obtained using the yeast two-hybrid system. More recently, a combination of affinity purification techniques and mass spectrometry (AP-MS) has been used to greatly advance our understanding of protein complex composition [26]. Alternative strategies are also emerging, like protein chips and luminescence-based mammalian interactome mapping (LUMIER). Fluorescence-based interaction assays such as fluorescence resonance energy transfer (FRET), hold great promise as they allow the monitoring of protein-protein interaction inside a living cell and computational prediction methods, some of which are based on known three-dimensional structures and binding motifs.

The yeast two-hybrid system is a genetic, *ex vivo* assay that allows the discovery of binary protein-protein interactions. Its principle relies on the modular nature of many transcription factors that contain both a site-specific DNA binding domain (DBD) and a transcriptional activation domain (AD) that recruits the transcriptional machinery to the promoters. The interaction between a 'bait' fusion protein (hybrid protein X-DBD) and a 'prey' fusion protein (hybrid protein Y-AD) reconstitutes a functional transcription factor which turns on the expression of a reporter gene or selectable marker [27].

The key advantage of the two-hybrid system, besides scalability and flexibility, is its capacity to detect transient interactions and those that are very weak. This includes interactions with dissociation constants down to 10^{-7} M, which correspond to the weakest interactions that occur within the cell. The system, however, has drawbacks due to its *ex vivo* nature and for this reason this does not necessarily reflect physiological events. Expressed fusion proteins are forced to the nucleus, which may not be where they are usually localised. Membrane proteins, for example, are usually not compatible with such a nuclear-based assay. The bait or prey proteins may not undergo appropriate post-translational modification. Similarly, interactions that involve co-operative, allosteric events or chaperone-assisted assembly may not occur correctly in the nucleus. Finally, transcription factors, as well as other proteins (about 5-10% of gene products), can auto activate the transcription of reporter genes. Such proteins are thus unsuitable to the approach.

In contrast to yeast two hybrid and related methods, AP-MS can be performed under near physiological conditions. AP-MS does not typically perturb relevant post-translational modifications, which are often crucial for the organization and/or activity of complexes (Fig.I.3).

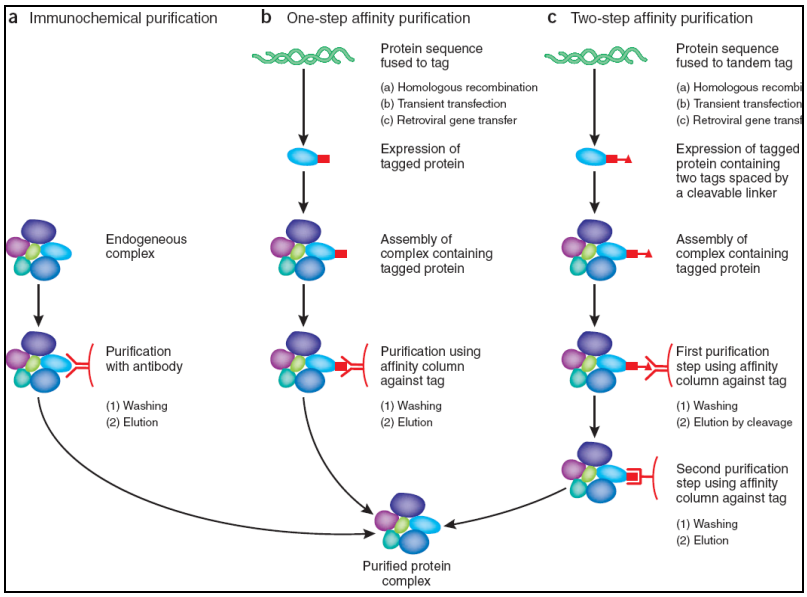


Fig.I.3. Main routes of protein-complex purification. (**a**) In immunochemical purification, the endogenous protein complex is precipitated using an antibody to the target protein, allowing protein-complex characterization without expression of a tagged protein. (**b**) In one-step affinity purification, the purified protein complex is obtained by expression of the tagged construct in the cell, followed by specific binding and elution from an affinity column. (**c**) In two-step affinity purification, two rounds of specific binding and specific elution assure a highly purified protein complex with little contaminating proteins at the cost of losing transient interactions.

Another advantage of AP-MS is that it can be used to probe dynamic changes in the composition of protein complexes, especially when used in combination with quantitative proteomics techniques [28,29].
In immunochemical methods, protein complexes are precipitated from a cell lysate by using an immobilized antibody to a known component of a complex. The protein complex is then purified by washing away nonspecific interactors. A fundamental advantage of this approach is that protein complexes can be isolated with a specific and efficient antibody from all types of biological sources, including tissue samples from patients, circumventing the need to express the target protein.
Affinity purification techniques exploit the biochemical properties of a tag attached to the bait protein to purify the other components of a protein complex. Using standard cloning techniques, target protein and peptide-tag coding sequences are fused, and the resulting construct is expressed in target cells. A variety of epitope tags have been developed in the past such as His tags, glutathione *S*-transferase (GST) tags, Flag tags, Myc tags, HA tags, the calmodulin-binding peptide, the streptavidin binding peptide or the *in vivo* biotinylation of the target tagged protein. Corresponding antibodies or affinity resins (e.g. anti-Myc, anti-HA, anti-Flag, anti-KT3, glutathione) are commercially available [30].
One of the most successful tags developed to date is the tandem affinity purification (TAP) tag [31,32], which uses two sequential enrichment steps. Developed for yeast, the original TAP tag is composed of a protein A tag, followed by a tobacco etch virus (TEV) protease cleavage site and a calmodulin binding peptide. In the first purification step, the protein complex is purified from the cell lysate on an immunoglobulin gamma (IgG) affinity resin. The target protein complex is cleaved from the protein A tag with TEV protease. The eluate is then selectively enriched by increased [Ca^{2+}] in a second affinity purification step on an immobilized calmodulin column. The TAP-tag significantly reduce background noise, but probably result in the loss of some of the more transient and weak binding partners during the purification procedure, as the second affinity step essentially causes infinite sample dilution. Indentification of the single protein components is then achieved by means of mass spectrometry.

I.4 Biochemical approach to map three-dimensional structure of proteins and protein complexes

A comprehensive study of the structural details of protein-protein interactions is an important step to understand how the information contained in a genome is put into action. Methods such as two-hybrid assays, functional genetics, FRET and, recently, large-scale chip-based proteomics have uncovered a vast number of protein-protein interactions; in yeast alone it is estimated that one protein molecule has on average about nine different interaction protein partners [33]. However, although a number of experimental techniques can report on the existence of a protein- protein interaction, very few can provide detailed structural information.
Since its inception in the 1950s of the 20th century, structural biology provided critical observations in biology, and it continues to contribute greatly to the understanding of many biological processes. Structures helped to decipher basic principles of protein assembly, mechanisms of biochemical reactions, and details of macromolecular interactions, and contributed to developing new pharmaceuticals.

I.4.1 High resolution structural characterization

Structural proteomics is the large scale structural analysis of proteins and complexes present in a given cell. It holds special significance since cellular mechanisms are functions of the structures and interactions between macromolecular complexes involved in cellular biological transactions. Important questions in structural proteomics involve elucidating the structure of these multicomponent assemblies, including their subcomponents and their assembly, and relating their structure to function. At present, the majority of proteins associated with many key cellular processes have structural representatives known. High-quality structural models are available for the majority of important protein families. Moreover, the structures of numerous complexes and multi-component assemblies have been also determined. At the same time it is worth noticing that some very large and important protein families are still poorly structurally studied.
Structural genomics researchers' intention was to fill this major gap by developing high-throughput methods, improving quality of structures and reducing the cost of structure determination. According to PDB records, the number of solved structures increased more than fourfold since 1993.
A detailed knowledge of three-dimensional protein structure and the elucidation of interacting amino acid sequences in protein complexes are of outstanding importance for the understanding of protein functions in a cell. A number of well-established analytical techniques are available to address the questions of spatial and topological organizations of proteins and protein complexes, currently primarily x-ray crystallography and NMR spectroscopy. Both techniques yield detailed information on the three dimensional structure of proteins at atomic resolution.
In principle x-ray crystallography is not limited to the size of the macromolecule, but the protein has to be crystallized and it is not easy to obtain well-structured crystals with good diffraction pattern for high-resolution reconstruction.
NMR is an appropriate technique to study weak and strong interactions of protein complexes in solution. It also allows studies of conformational changes in time courses. However most NMR structures can be solved of proteins with a molecular weight distribution in the range of 20 kDa and limited to 50 kDa and NMR spectroscopy requires large quantities of pure protein in a specific solvent.
Neither of these methods affords high throughput, which is an important consideration of the fast-evolving modern scientific regime. In this regard, MS is an attractive alternative to provide important low-resolution structural information.

I.4.2 Low resolution structural characterization

Mass spectrometry is an invaluable tool to provide low-resolution structural information on the interface of proteins and protein complexes. Although X-ray crystallography and NMR methods remain the gold standard in terms of resolution, MS possesses distinct advantages such as relative speed, sensitivity, and adaptability to different types of protein that account for its popularity to analyze protein complexes.
MS can monitor conformational alternations in flexible loop regions that are often not resolved by traditional methods. Another advantage of MS is that the targeted proteins can interact with their ligands in solution under native conditions, and can be monitored under a more physiological-like concentration.

Among the mass spectrometric techniques used for structural studies of proteins are hydrogen-deuterium exchange, limited proteolysis, gel filtration, analysis of the intact non-covalent complexes in the gas phase and chemical cross-linking followed by enzymatic digestion and mass spectrometric identification of the cross-linked peptides.

Limited proteolysis is based on the evidence that the three-dimensional structure of a protein constitutes a barrier against the enzymatic hydrolysis, which occurs only on exposed and flexible region of the backbone [34]. Controlled experimental conditions (pH, T, time, Ionic strength, E/S ratio), in order to ensure the stability of the target protein, will produce a single hydrolysis on the exposed/flexible region of the protein, and the mass of the resulting two complementary peptides is easily determined by MS, even if in trace amounts. By using a wide range of proteases is then possible to get a detailed picture of the exposed protein backbone. The use of identical proteolysis conditions for different states of the protein allows to compare different conformations occurring on formation of complexes with molecular effector, either of low or high molecular weight. It is applicable to analytes at low concentrations, it does not require specific solvent provided that the protease is active.

In recent years, chemical cross-linking coupled with mass spectrometry (MS) has become a powerful method for studying protein interactions [35,36]. Chemical cross-linking stabilizes protein interactions through the formation of covalent bonds and allows the detection of stable and weak protein-protein interactions in native cells or tissues. In addition to capturing protein interacting partners, many studies have shown that chemical cross-linking can yield low-resolution structural information about the constraints within a molecule or protein complex. Identification of cross-linked peptides provides distance constraints that aid in constructing the structural topology of proteins and/or protein complexes. Moreover, it is applicable to analytes at low concentrations, it does not require protein crystals or proteins in a specific solvent and it is a fast method provided that data analysis can be automated [37-39].

A cross-linker is simply a chemical reagent with two or more reactive groups connected by a spacer or linker region. The selection of the reactive groups depends on the target molecules to be cross-linked. For the purposes of labelling proteins, the reactive ends can be amine-, sulfhydryl-, or photo-reactive. The length of the spacer chain is often used as a ‘‘ruler’’ for estimating the distance of the two linked residues, providing approximate topological information on proteins or protein complexes. Many structural variants of the basic, conventional cross-linker (two reactive groups and a simple spacer chain) are commercially available.

However, protein-protein cross-linking techniques present multiple experimental and analytical challenges. The choice of cross-linker is crucial, as cross-linkers vary in cell-permeability, reactivity and spacer arm length. Cross-linker concentrations, reaction times, and buffer pH must be optimized to achieve a high yield of cross-linked product, while not disrupting the three-dimensional protein structures by introducing too many cross-links per molecule.

To conduct chemical cross-linking experiments of proteins, two alternative strategies exist in principle, which is commonly referred to as bottom-up and top-down approaches. In the bottom-up approach, the protein reaction mixture is enzymatically digested after the cross-linking reaction, and mass spectrometric identification of the cross-linked products is performed, based on the resulting proteolytic peptides.

One of the most direct techniques to analyze cross-linked products is the top-down approach, in which the cross-linked proteins are analyzed intact rather than being digested before the mass spectrometric analysis. Within the mass spectrometer, the

cross-linked protein is isolated and directly fragmented, without the introduction of the fractionation step before the mass spectrometric analysis. Because this approach starts from MS detection of the intact protein, it has the potential for full protein characterization.

I.5 The mass spectrometry role in the analysis of protein complexes

The trend toward mass spectrometry as the technique of choice for identifying and probing the covalent structure of proteins was accelerated by the genome project. Genomics demonstrated the power of high-throughput, comprehensive analyses of biological systems. Genomics also provides complete genomic sequences, which are a critical resource for identifying proteins quickly and robustly by the correlation of mass spectrometric measurements of peptides with sequence databases. The systematic analysis of all the proteins in a tissue or cell was popularized under the name proteomics, with mass spectrometry central to most proteomic strategies.

Mass spectrometry is currently the method of choice for peptide sequencing because it is sensitive and it routinely allows for the identification of peptides that are present at femtomole levels. MS is also rapid; sequencing of individual peptides can be achieved within hundreds of milliseconds, and thousands of peptides can therefore be identified in a single MS run. Last, MS is compatible with high-throughput strategies and is easily automated. It also allows for the characterization of peptide modifications (including naturally occurring post-translational modifications, such as phosphorylation, and exogenously added modifications, such as chemical cross-linkers). MS can also be adapted to quantitatively measure peptide abundance and does not require pre-existing knowledge of the proteins to be analysed [40,41]. Advances in sample processing and instrumentation have gone hand-in-hand with the development of software tools that automatically retrieve sequence information from acquired mass spectra and provide statistical validation of the accuracy of the determined sequences [42].

Mass spectrometry was restricted for a long time to small and thermostable compounds because of the lack of effective techniques to softly ionize and transfer the ionized molecules from the condensed phase into the gas phase without excessive fragmentation. The development in the late 1980s of two soft ionization techniques for the routine and general formation of molecular ions of intact biomolecules, electrospray ionization (ESI) [43] and matrix assisted laser desorption/ionization (MALDI) [44], dramatically changed this situation and made polypeptides accessible to mass spectrometric analysis [45].

Mass spectrometric measurements are carried out in the gas phase on ionized analytes. By definition, a mass spectrometer consists of an ion source, a mass analyser that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that registers the number of ions at each m/z value.

Matrix-assisted laser desorption ionization (MALDI) was developed by Karas & Hillenkamp in the late 1980s. To generate gas phase, protonated molecules, a large excess of matrix material is cocristallized with analyte molecules. The resulting solid is then irradiated by nanosecond laser pulses, usually from small nitrogen lasers with a wavelength of 337 nm. The matrix is typically a small organic molecule with absorbance at the wavelength of the laser employed and the matrices differ from each other in the amount of energy they impart to the biomolecules during desorption

and ionization and hence the degree of fragmentation (unimolecular decay) that they cause. Typically, only singly charged ions are observed in MALDI.
The precise nature of the ionization process in MALDI is still largely unknown and the signal intensities depend on incorporation of the peptides into crystals, their likelihood of capturing and/or retaining a proton during the desorption process, and a number of other factors including suppression effects in peptide mixtures [46].
Electrospray mass spectrometry (ESMS) has been developed for use in biological mass spectrometry by Fenn et al. Liquid containing the analyte is pumped at low microliter-per-minute flow rates through a fine needle at high voltage to electrostatically disperse, or electrospray, small, micrometer-sized droplets, which rapidly evaporate and which impart their charge onto the analyte molecules. ESI typically induces a range of charge states; because most mass spectrometers actually detect mass/charge or m/z, the resulting spectra may have many ions for each analysis. This includes proteins, oligonucleotides, sugars and polar lipids [47].
After ionization, the analytes enter in the mass analyser. It is, literally and figuratively, central to the technology. In the context of proteomics, its key parameters are sensitivity, resolution, mass accuracy and the ability to generate information-rich ion mass spectra from peptide fragments (tandem mass or MS/MS spectra) [48,49]. There are four basic types of mass analyser currently used in proteomics research (Fig.I.4). These are the quadrupole (Q), ion trap (IT), time-of-flight (TOF), and Fourier transform ion cyclotron resonance (FT-ICR) analysers [50]. They are very different in design and performance, each with its own strength and weakness. These analysers can be stand alone or, in some cases, put together in tandem to take advantage of the strengths of each.

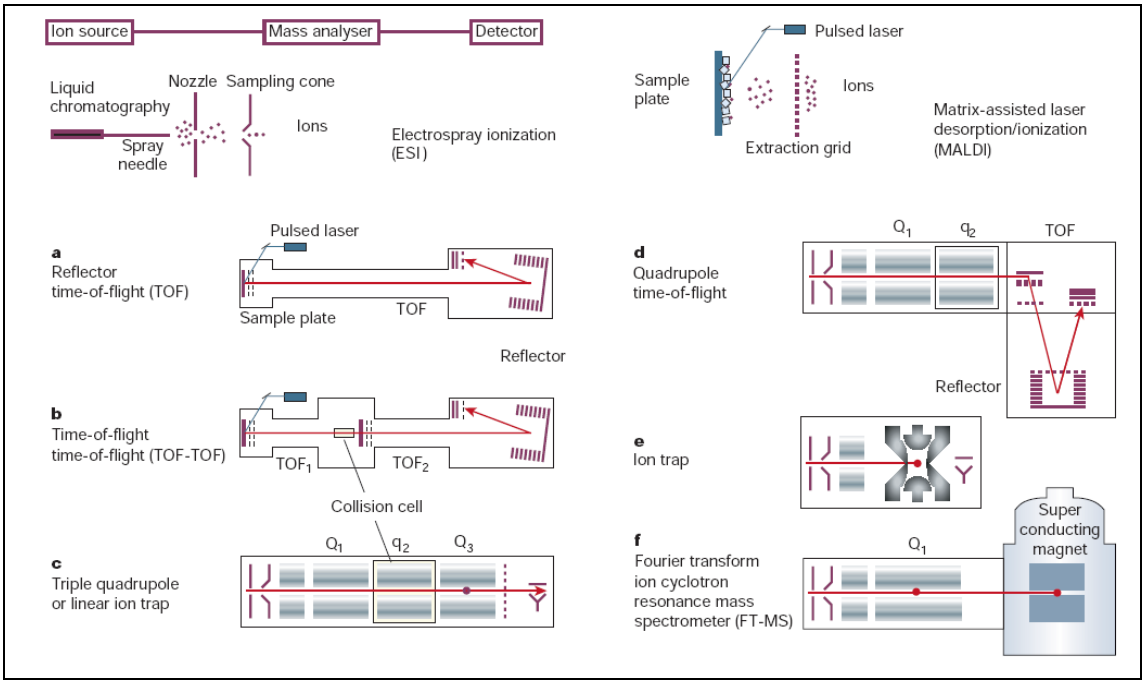


Fig.I.4. Mass spectrometers used in proteome research.

The quadrupole mass filter consists of a linear array of four symmetrically arranged rods to which radio frequency (RF) and DC voltages are supplied. Forces are exerted in a plane normal to the direction (z-direction) in which the ions drift. The RF potential gives rise to a field which alternatively reinforces and then dominates the DC field.

Ions oscillate in the x,y-plane with frequencies which depend on their m/z values. If the oscillations of an ion in this plane are stable, the ion will continue to drift down the rod assembly and reach the detector. Stable oscillations are only achieved by ions of given m/z value for a given rod assembly, oscillation frequency, RF voltages and DC voltages. Commercially available instruments usually have mass/charge limits ranging from 0 to 4000 m/z and at best are normally set to resolve the various ^{13}C isotope peaks for a singly charged ion, although the resolution may be intentionally degraded to improve sensitivity. In ESI, multiple charging enables quadrupole mass measurement of molecules >100,000 Da, if the molecule can be charged sufficiently.
The ion trap consists of three electrodes, the central ring electrode and two end-cap electrodes of hyperbolic cross-section. In this device too, ions are subjected to forces applied by an RF field but the forces occur in all three, instead of just two, dimensions. In ion-trap analysers, ions are resonantly activated and ejected by electronic manipulation of this field. Ion traps are robust, sensitive and relatively inexpensive, and so have produced much of the proteomics data reported in the literature. A disadvantage of ion traps is their relatively low mass accuracy, due in part to the limited number of ions that can be accumulated at their point-like centre before space-charging distorts their distribution and thus the accuracy of the mass measurement. The ‘linear’ or ‘two dimensional ion trap’ [51,52] is an exciting recent improvement in quadrupole ion traps, with higher scan ranges, larger electronic trap fields, increased sensitivity, higher resolution and mass accuracy.
In the Time-of flight (TOF) analyzers the ions are accelerated through a fixed potential into a drift tube. As all the ions with same charge obtain the same kinetic energy after acceleration, the lower m/z ions achieve higher velocities than higher m/z ions. Moreover, ion velocities are inversely related to the square root of the m/z. Thus, by measuring the time it takes to reach the detector, the m/z of the ion can be determined. Ion manipulations are used to increase resolution (delayed extraction of ions from the source, two stage sources with complex voltage gradients, and reflectron technology) and a commercial TOF instrument can typically achieve resolution of 10,000 or greater.
A Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometer (also referred to as a Fourier-transform mass spectrometer, or FTMS) uses a magnetic field to determine the m/z of an ion. In an FT-ICR ions have kinetic energies, at most, of a few tens of electron volts (eV). At low kinetic energies, ions are actually trapped under high vacuum in the magnetic field. For a constant magnetic field, ions oscillate around the magnetic field with a cyclotron frequency that is inversely related to the m/z. In a very simplified view of FT-ICR, the cyclotron frequencies of the ions trapped in the FT-ICR are measured and then converted into m/z. State-of-the-art electronic equipment is capable of measuring frequencies with extremely high precision. This translates to a very high mass resolution, which is the property FT-ICR is most widely known for. Mass resolutions in the hundreds of thousands are fairly easy to obtain on instruments with large magnetic field strengths (that is, > 7 Tesla), and resolutions in the millions have been demonstrated. Very high mass accuracies, down to the ppm level, can also be obtained (resolution >1,000,000) [53,54].
Very recently, a new type of mass analyzer called orbitrap has emerged. It was invented by A. Makarov [55-57]. As its name suggests, orbitrap is an ion trap. But it is not a conventional ion trap; there is neither RF nor a magnet to hold ions inside. Instead, moving ions are trapped in an electrostatic field. The electrostatic attraction towards the central electrode is compensated by a centrifugal force that arises from the initial tangential velocity of ions. The electrostatic field which ions experience

inside the orbitrap forces them to move in complex spiral patterns. The axial component of these oscillations is independent of initial energy, angles and positions, and can be detected as an image current on the two halves of an electrode encapsulating the orbitrap. A Fourier transform is employed to obtain oscillation frequencies for ions with different masses, resulting in an accurate reading of their m/z. Such measurements achieve very high resolution rivalling that of FT-ICR instruments, and surpassing, by an order of magnitude, the resolution presently obtainable with orthogonal time-of-flight analyzers. The instrument is capable of routinely achieving high sensitivity (attomole to femtomole range), high mass accuracy (low ppm), and isotope resolution of small proteins. This instrument presents characteristics similar to an FT-ICR spectrometer in terms of resolution and mass accuracy but without the burden of an expensive superconducting magnet.

I.5.1 Tandem mass spectrometry

The power of mass spectrometry can be dramatically increased by employing methods of tandem mass spectroscopy [58,59]. In the tandem MS (MS/MS) and MSn experiments, the first mass analyzer is used to selectively pass an ion into another reaction region where excitation and dissociation take place. The second mass analyzer is used to record the m/z values of the dissociation products. Low-energy collision-induced (activated) dissociation (CID or CAD) tandem mass spectrometry has been, by far, the most common method used to dissociate peptide ions for subsequent sequence analysis. Upon collisional activation with a nonreactive gas, such as argon or helium, the amide bond of the peptide backbone will fragment to produce, ideally, a homologous series of b and y-type fragment ions. The observed fragmentation pattern depends on various parameters including the amino acid composition and size of the peptide, excitation method, time scale of the instrument, the charge state of the ion, etc. Peptide precursor ions dissociated under the most usual low-energy collision conditions fragment along the backbone at the amide bonds forming structurally informative sequence ions and less useful non-sequence ions by losing small neutrals like water, ammonia, etc.

An alternative method for peptide dissociation (electron capture dissociation, ECD) was introduced by McLafferty and co-workers [60]. Low energy electrons were reacted with peptide cations in the magnetic field of a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS). The reaction resulted in the attachment of electrons to the protonated peptides producing peptide cations containing an additional electron. The odd-electron peptide then undergoes rearrangement with subsequent dissociation. Peptide backbone cleavage by ECD is relatively indifferent to either peptide sequence or length. Unlike CID, ECD does not cleave chemical modifications from the peptide, but rather induces random breakage of the peptide backbone. With ECD, labile modifications such as phosphorylation, N- and O-glycosylation, sulfonation, as well as other labile PTMs remain intact. However, ECD requires an FT-ICR mass spectrometer, instruments that are not readily available to many in the proteomics field.

Electron transfer dissociation (ETD) is another method to fragment peptides that utilizes ion/ion chemistry. ETD fragments peptides by transferring an electron from a radical anion to a protonated peptide. This induces fragmentation of the peptide backbone, causing cleavage of the Cα–N bond just as ECD does. This creates complementary c and z-type ions instead of the typical b and y-type ions observed in CID. ETD also preserves PTMs that are labile by CID and sequence information on

the peptide can be obtained. ETD uses a RF quadrupole ion trapping device instead of an FT-ICRMS for ion trapping and detection. RF ion trap mass spectrometers are low-cost, low-maintenance, and widely accessible as compared to the FT-ICR-MS [61].
A peptide contains three different bonds in its backbone: NH-CH, CH-CO and CO-NH. According to the nomenclature for peptide fragmentation each breakage at any of the bonds gives rise to a different ion depending on which fragment retains the charge. Peptide fragment ions are termed a-, b-, c-ions if the charge is retained on the N-terminus and x-, y- and z-ions if the charge is kept on the C-terminus (Fig.I.5). The peptide can be manually sequenced by systematically calculating the mass difference between two adjacent ions of one series that correspond to one amino acid at the position of cleavage. Nowadays computer programs (such as Mascot) facilitate this task.

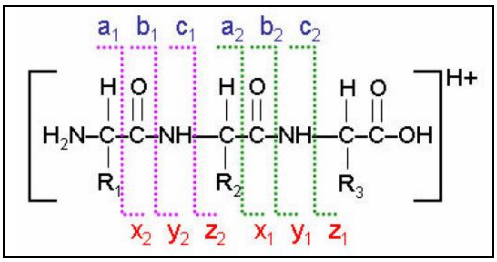


Fig.I.5. Roepstorff Nomenclature Scheme. Illustration of fragment ions formed from the backbone cleavage of protonated peptides.

I.5.2 Non-covalent protein complexes - Native mass spectrometry

Soon after the introduction of electrospray ionization (ESI), researchers realized that it could be used to monitor intact proteins and even protein complexes by MS [62], because ESI is an extremely soft ionization technique and non covalent interactions between molecules can be preserved [63]. Recent application of native MS have highlighted its ability to define stoichiometry, topology and dynamics of many protein complexes, ranging from recombinant expressed large oligomeric complexes, such as viruses and chaperones complexes, to endogenously expressed structurally less well characterized protein complexes. A major breakthrough was implemented with the development of nano-ESI that facilitates the ionization efficiency due to a decreased flow. The successful analysis of non covalent assemblies by MS requires the preservation of the native state of the complex in solution. In addition, to ensure efficient ES, it is crucial to use aqueous solutions containing only volatile buffers. Owing to their neutral pH and high volatility, the two most commonly used buffers are ammonium acetate and ammonium carbonate. Millimolar quantities (ranging from 5 mM to 1 M) of volatile buffer at physiological pH are typically employed so that complexes are introduced from solution conditions in which the native state is preserved.
Native mass spectrometry does not yield detailed molecular (and atomic) structure information, but it has some major advantages over traditional structural biology methods, such as its sensitivity, speed, selectivity and ability to simultaneously measure several species present in a mixture.

I.6 Aim of the PhD thesis

Proteins form stable and dynamic multi-subunit complexes under different physiological conditions to maintain cell viability and normal cell homeostasis. Detailed knowledge of protein interactions and protein complex structures is fundamental to understand how individual proteins function within a complex and how the complex functions as a whole.
This PhD thesis targets the characterization by structural and functional proteomics approach of selected protein complexes involved in relevant molecular processes and that can have an impact on their industrial applications.
Moreover, as a parallel goal, and strictly connected with the former issue, I worked at the development of an innovative structural analysis approach based on the use of a femtosecond UV laser in order to capture transient protein-protein interactions.

Chapter 2 describes the analysis of the endogenously expressed Phenol Hydroxylase (PH) complex from *Pseudomonas sp*. OX1, using biochemical approaches integrated to mass spectrometry analysis. Moreover, in order to define the protein-protein interaction networks of this complex a functional proteomic approach was carried out.
Pseudomonas sp. OX1 is a bacterium endowed with the ability to grow on a wide spectrum of hydroxylated and non-hydroxylated aromatic compounds. Aromatic hydrocarbons released in the biosphere by human activities are today a major threat not only to environment but also to human health due to their potential carcinogenicity [64]. By expressing various catabolic pathways, microorganisms can use a wide array of aromatic compounds as the sole carbon and energy source, thus providing a set of diversified tools that can be used in bioremediation of contaminated environments. Bacterial Multicomponent Monooxygenases (BMMs) are actually key enzymes that catalyze the hydroxylation of the aromatic ring at different positions and for this reason they can be use in bioremediation of contaminated environments.
This part of my thesis has been carried out in collaboration with the Prof. Di Donato's group of the Department of Structural Biology at the University of Naples "Federico II".

The work in **chapter 3** is aimed on the characterization of the complex between pLG72 and human D-amino acid oxidase (hDAAO), which has recently been linked to the onset of schizophrenia. Schizophrenia affects almost 1% of the world's population and accounts for ~2.5% of the healthcare costs of the European Community. It is a psychiatric disorder that develops as a result of neurobiological and genetic predispositions interacting with environmental factors. In the past years, yeast two-hybrid experiments with the pLG72 protein identified the enzyme D-amino acid oxidase (DAAO) as an interacting partner [65].
Because the causes of schizophrenia are still unknown, current treatments focus on eliminating the symptoms of the disease. The clarification of the key role of this protein-protein interaction and the structural effects of drugs on this interaction, as well as the identification of other physiological or conditional partners in this biochemical pathway, will be extremely useful in developing molecules effective in the disease treatment.
In this perspective the structural features of the pLG72-hDAAO complex was defined by low resolution strategies (limited proteolysis and chemical cross-linking) coupled

with mass spectrometric techniques. Moreover functional proteomics experiments using hDAAO as bait have been carried out with the aim of identifying other interactors of this protein.
This part of my thesis has been carried out in collaboration with the Prof. Pollegioni of the University of Insubria.

Chapter 4 introduces the development of an innovative methodology to induce cross-linking in proteins, based on the use of an UV ultra shorts pulsed laser with duration of few tens of femtoseconds as a novel zero length protein-protein cross-linker. This work has been in collaboration with Proff. R. Velotta and C. Altucci's group of the Physic Department at the University of Naples "Federico II".
Although the combination of chemical cross-linking and mass spectrometry has been successful, effective detection and accurate identification of cross-linked peptides as well as unambiguous assignment of cross-linked sites remain extremely challenging due to their low abundance and complicated fragmentation behaviour in MS analysis. Therefore, new reagents and methods are urgently needed to allow unambiguous identification of cross-linked products and to improve the speed and accuracy of data analysis to facilitate its application in structural elucidation of large protein complexes. It is expected that UV-laser cross-linking will offer a powerful tool to investigate protein-protein interactions, especially transient interactions and binding kinetics, because the number of photons required for covalent complex formation can be delivered very rapidly thus resulting in efficient cross-linking. Such photo-physical approach could obviate many of the problems associated with standard chemical cross-linking reagents, and put cross-linking in a proteome-wide position for the characterizations of protein-protein interactions *in vivo* in intact cells, since no chemical reagent has to enter the cell, and thus false-positive interactions such as those resulting from the loss of spatial organization during cell lyses are also avoided.

I.7 References

[1] Parekh S, Vinci VA, Strobel RJ. Improvement of microbial strains and fermentation processes. *Appl Microbiol Biotechnol*. (2000);54(3):287-301. Review.

[2] Selinger DW, Wright MA, Church GM. On the complete determination of biological systems. *Trends Biotechnol*. (2003);21(6):251-4.

[3] Stelling, J. Mathematical models in microbial systems biology. *Curr Opin Microbiol*. (2004);7(5):513-8. Review.

[4] Abu-Farha M, Elisma F, Figeys D. Identification of protein-protein interactions by mass spectrometry coupled techniques. *Adv Biochem Eng Biotechnol.* (2008);110:67-80. Review.

[5] Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* (2003);422(6934):835-47.

[6] de Hoog CL, Mann M. Proteomics. *Annu Rev Genomics Hum Genet*. (2004);5:267-93. Review.

[7] Villoslada P, Steinman L, Baranzini SE. Systems biology and its application to the understanding of neurological diseases. *Ann Neurol*. (2009);65(2):124-39. Review.

[8] Chautard E, Thierry-Mieg N, Ricard-Blum S. Interaction networks: from protein functions to drug discovery. A review. Pathol Biol (Paris). (2009);57(4):324-33. Review.
[9] Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein-protein interactions: structural, functional, and network properties. *Structure*. (2010);18(10):1233-43.
[10] Bader S, Kühner S, Gavin AC. Interaction networks for systems biology. *FEBS Lett*. (2008);582(8):1220-4.
[11] Pawson T, Raina M, Nash P. Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett*. (2002);513:2-10.
[12] Scheffzek K, Ahmadian MR, Kabsch W, Wiesmüller L, Lautwein A, Schmitz F, Wittinghofer A. The Ras-RasGAP complex: structural basis of GTPase activation and its loss in oncogenic Ras mutants. *Science*. (1997)*;*277(5324):333-8.
[13] Schreiber G. Kinetic studies of protein-protein interactions. Curr. Opin. Struct. Biol. (2002);12:41-47.
[14] Reichmann D, Rahat O, Cohen M, Neuvirth H, Schreiber G. The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol.* (2007);17:67-76.
[15] Ladbury JE, Lemmon MA, Zhou M, Green J, Botfield MC, Schlessinger J. Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. *Proc. Natl. ACID. Sci. USA* (1995);92:3199-203.
[16] Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB. Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol*. (2005); 3(12):e405.
[17] Koch CA, Anderson D, Moran MF, Ellis C, Pawson T. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* (1991);252:668-74.
[18] Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. Proteome survey reveals modularity of the yeast cell machinery. *Nature* (2006);440(7084):631-6.
[19] Formstecher E, Aresta S, Collura V, Hamburger A, Meil A, Trehin A, Reverdy C, Betin V, Maire S, Brun C, Jacq B, Arpin M, Bellaiche Y, Bellusci S, Benaroch P, Bornens M, Chanet R, Chavrier P, Delattre O, Doye V, Fehon R, Faye G, Galli T, Girault JA, Goud B, de Gunzburg J, Johannes L, Junier MP, Mirouse V, Mukherjee A, Papadopoulo D, Perez F, Plessis A, Rossé C, Saule S, Stoppa-Lyonnet D, Vincent A, White M, Legrain P, Wojcik J, Camonis J, Daviet L. Protein interaction mapping: a *Drosophila* case study. *Genome Res.* (2005);15:376-84.
[20] Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M*.* A map of the interactome network of the metazoan *C. elegans*. *Science* (2004);303:540-3.
[21] Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksöz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* (2005);122:957-68.
[22] Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE,

Roth FP, Vidal M. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* (2005);437:1173-78.
[23] Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for Biotechnology nature biotechnology. *Nat Biotechnol*. (2009); 27(2):157-67.
[24] Farinas ET, Bulter T, Arnold FH. Directed enzyme evolution. *Curr Opin Biotechnol*. (2001);12(6):545-51. Review.
[25] Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol. 2006;2:2006.0028.
[26] Gavin AC, Wilkins MR, Appel RD, Williams KL, Hochstrasser D.F. Protein-Protein Interactions. *Proteome Research: Concepts, Technology and Application* © Springer-Verlag Berlin Heidelberg 2007.
[27] Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature* (1989);340(6230):245-6.
[28] Kratchmarova I, Blagoev B, Haack-Sorensen M, Kassem M, Mann M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* (2005);308:1472-7.
[29] Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. *Nature Biotechnol.* (2003);21(3):315-8.
[30] Köcher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* (2007);4(10):807-15.
[31] Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Séraphin B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*. (2001);24(3):218-29. Review.
[32] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. *Nat Biotechnol.* (1999);17(10):1030-2.
[33] Russell RB, Alber F, Aloy P, Davis FP, Korkin D, Pichaud M, Topf M, Sali A. A structural perspective on protein-protein interactions. *Curr. Opin. Struct. Biol.* (2004);14:313-24.
[34] Zappacosta F, Pessi A, Bianchi E, Venturini S, Sollazzo M, Tramontano A, Marino G, Pucci P. Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody. *Protein Sci.* (1996);5(5):802-13.
[35] Kao A, Chiu CL, Vellucci D, Yang Y, Patel VR, Guan S, Randall A, Baldi P, Rychnovsky SD, Huang L. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics*. (2010). [Epub ahead of print]
[36] Zelter A, Hoopmann MR, Vernon R, Baker D, MacCoss MJ, Davis TN. Isotope signatures allow identification of chemically cross-linked peptides by mass spectrometry: a novel method to determine interresidue distances in protein structures through cross-linking. *J Proteome Res*. (2010);9(7):3583-9.
[37] Rappsilber, J., Siniossoglou, S., Hurt, E.C. & Mann, M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal. Chem.* (2000);72, 267-75.
[38] Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Rev.* (2006);25(4):663-82.
[39] Back JW, de Jong L, Muijsers AO, deKoster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol*. (2003);331(2):303-13. Review.
[40] Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol*. (2007);8(8):645-54. Review.
[41] Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov.* (2003);2(2):140-50.
[42] Nesvizhskii AI, Roos FF, Grossmann J, Vogelzang M, Eddes JS, Gruissem W, Baginsky S, Aebersold R. Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification

of post-translational modifications, sequence polymorphisms, and novel peptides. *Mol Cell Proteomics.* (2006);5(4):652-70.
[43] Fenn JB, Mann M, Meng CK, Wong S F, Whitehouse CM. Electrospray ionization for the mass spectrometry of large biomolecules. *Science* (1989);246:64-71.
[44] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular mass exceeding 10000 daltons. *Anal. Chem.* (1988) 60:2299-2301.
[45] Domon B, Aebersold R. Mass spectrometry and protein analysis *Science* (2006); 312(5771):212-7. Review.
[46] Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* (2001);70, 437-473.
[47] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* (2003);422(6928):198-207.
[48] Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* (2000);405:837-846.
[49] Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem. Rev.* (2001);101:269-295.
[50] Wysocki VH, Resing KA, Zhang Q, Cheng G. Mass spectrometry of peptides and proteins *Methods.* (2005);35(3):211-22.
[51] Hager JW. A new linear ion trap mass spectrometer. *Rapid Commun. Mass. Spectrom.* (2002);16:512-526.
[52] Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom.* (2002);13(6):659-69.
[53] Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev*. (1998);17(1):1-35.
[54] Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* (2000);72:4266-74.
[55] Makarov A. Electrostatic axially harmonic orbital trapping: a high performance technique of mass analysis. *Anal. Chem*. (2000);72(6):1156-62.
[56] Scigelova M, Makarov A. Orbitrap mass analyzer-overview and applications in proteomics. *Proteomics*. (2006);6 Suppl 2:16-21.
[57] Hu QZ, Noll RJ, Li HY, Makarov A, Hardman M, Cooks RG The Orbitrap: a new mass spectrometer. J. Mass Spectrom. (2005);40(4), 430-443.
[58] Paizs B, Suhai S. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev*. (2005);24(4):508-48. Review.
[59] McLafferty FW. Tandem mass spectrometry. *Science.* **(**1981);214(4518):280-7.
[60] Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. Anal Chem. 2000 Feb 1;72(3):563-73.
[61] Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta.* (2006);1764(12):1811-22.
[62] Heck AJ. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods*. (2008);5(11):927-33. Review.
[63] Sharon M, Robinson CV. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem*. (2007);76:167-93. Review.
[64] Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, Pucci P, Di Donato A. *Appl Environ Microbiol*. (2004);70(4): 2211-9.
[65] Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, Barry C, Tanaka H, La Rosa P, Puech A, Tahri N, Cohen-Akenine A, Delabrosse S, Lissarrague S, Picard FP, Maurice K, Van Duijn C, Gauvreau D, Riazanskaia N, Rogaev E, Raeymaekers P, Aerssens J, Konings F, Luyten W, Macciardi F, Sham PC, Straub RE, Weinberger DR, Cohen N, Cohen D. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl ACID Sci U S A.* (2002);99(21):13675-80.

II. Structural e functional characterization of Phenol Hydroxylase (PH) complex from Pseudomonas sp. OX1: key enzyme involved in bioremediation of contaminated environments

II.1 The bioremediation

Aromatic hydrocarbons released in the biosphere by human activities are today a major threat not only to environment but also to human health due to their potential carcinogenicity [1]. As an example, cyanides, phenols, polycyclic aromatic compounds or long chain aliphatics may be hazardous and, particularly if accidentally or deliberately released by the processing industry, they exhibit acute or chronic toxicity. Therefore, a major goal of environmental biotechnology is to establish highly efficient biological processes that use the naturally existing catabolic potential for the elimination and detoxification of these chemicals.
The advent of the post-genome era has fostered research on bacterial biodegradation at the level of the genome, enabling environmental microbiologists to better understand the process of biodegradation. This development in the study of biodegrading bacteria has resulted in the impetus to develop a suitable strategy for bioremediation. Therefore, the extensive and intensive characterization of the catabolic genes, proteins, and metabolites in microorganisms associated with pollutant degradation is absolutely necessary in order to devise the most effective strategy for bioremediation [2].
Microorganisms can use a wide array of aromatic compounds as the sole carbon and energy source, thus providing a set of diversified tools that can be used in bioremediation of contaminated environments [3,4]. In bacteria, aerobic catabolic pathways for aromatic hydrocarbon degradation can schematically be divided into two major biochemical steps. First, early reactions, the so-called *upper pathways* leads to the formation of partially oxidized aromatic intermediates. Then, in the *lower pathways*, dihydroxylated aromatic molecules that can undergo the cleavage of the ring are produced and further processed to give compounds that can enter the tricarboxylic acid cycle and eventually mineralized to carbon dioxide and water (Fig.II.1) [5].

II.2 Bacterial multicomponent monooxygenases (BMMs)

Bacterial multicomponent monooxygenases (BMMs) are key enzymes of the upper pathway that hydroxylate and epoxidize an array of hydrocarbon substrates, including alkanes, alkenes, and aromatics in a regio- and enantioselective fashion [6,7]. The substrate transformations performed by this family of enzymes are responsible, in part, for removing atmospheric methane, a greenhouse gas, and carcinogenic halogenated solvents like Trichloroethylene (TCE) from the environment.
Recently, it has been recognized that bacterial multicomponent monooxygenases (BMMs) constitute a family of enzymes which can be divided into six distinct groups, each with a characteristic subunit composition. BMMs are transcribed from single operons that code for four to six polypeptides.

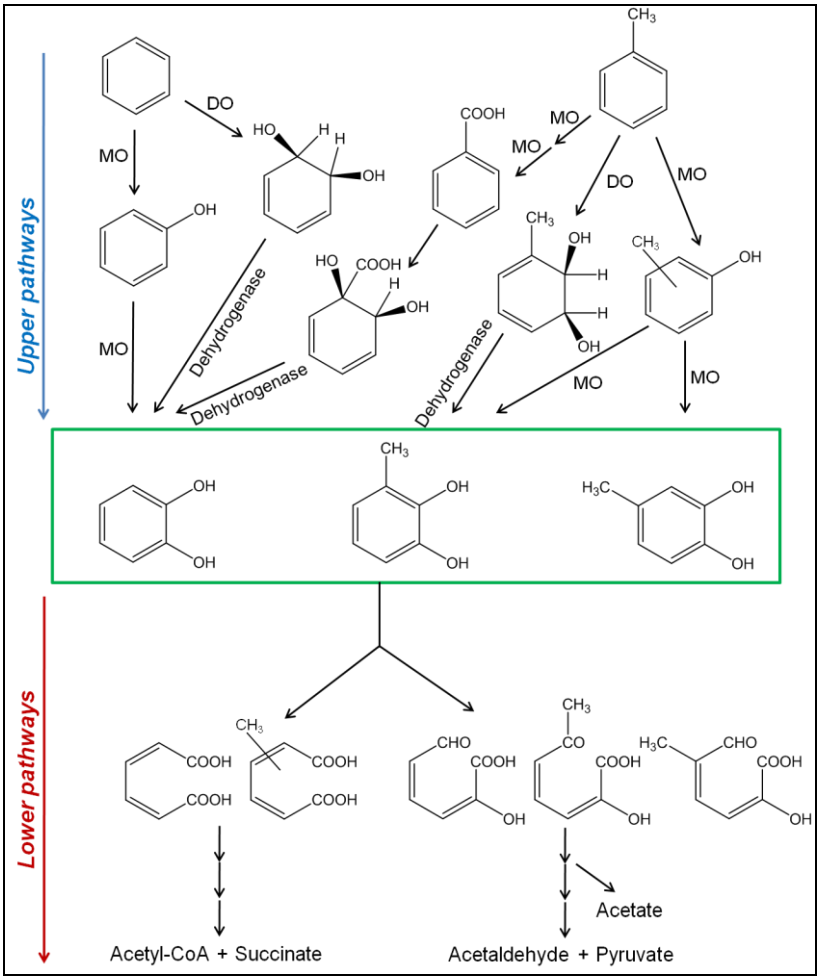


Fig.II.1. Catabolic pathways in microorganisms for benzene, phenol and toluene. In bacteria, aerobic catabolic pathways for aromatic hydrocarbon degradation can schematically be divided into two major biochemical steps, the *upper* and *lower* pathways. MO indicates monooxygenase, while DO dioxygenase.

Analysis of the sequences from nucleotide and protein databases indicates that most bacterial strains possess only one BMM, but a few cases (3 out of 31) of bacterial genomes coding for more than one monooxygenase have been found [8].
BMMs usually consist of a 200-250 kDa dimeric hydroxylase of the form $(\alpha\beta\gamma)_2$, a 10-16 kDa regulatory protein, devoid of any cofactor, that enhances catalytic turnover by 30-150-fold and a FAD- and [2Fe-2S]- containing 38-40 kDa reductase that mediates the electron transfer from NADH to the active site of the hydroxylase [7-9]. An additional Rieske protein may be present to assist the electron transfer between the reductase and hydroxylase components. The hydroxylation chemistry takes place at non-heme carboxylate-bridged diiron(III) center coordinated by four glutamate and two histidine ligands from a four-helix bundle. Solvent-derived water and hydroxide ions complete the octahedral coordination spheres [10].
Pseudomonas stutzeri OX1 is able to grow on a wide spectrum of aromatics, including phenol, cresols, and dimethylphenols, but also on nonhydroxylated molecules such as toluene, *o*-xylene and benzene. Two different monooxygenases have been found in the genome of *P. stutzeri* OX1, phenol hydroxylase (PH) [6] and toluene *o*-xylene monooxygenase (ToMO) [11,12].

II.3 Phenol Hydroxylase from *Pseudomonas stutzeri* OX1

Phenol Hydroxylase is a multicomponent monooxygenases enzyme composed by five polypeptides - PHL, PHM, PHN, PHO, PHP - organized as three components [6]. PHP is the NADH-oxidoreductase responsible for supplying electrons to the diiron cluster housed in the active site of the hydroxylase component. This latter component comprises three polypeptides (L, N, O) organized in a quaternary structure of the type $(LNO)_2$ (from here named PHH). Finally, PHM has been shown to be a regulatory protein, devoid of any cofactor or metal, which is essential for efficient catalysis of phenol hydroxylation (Fig.II.2).

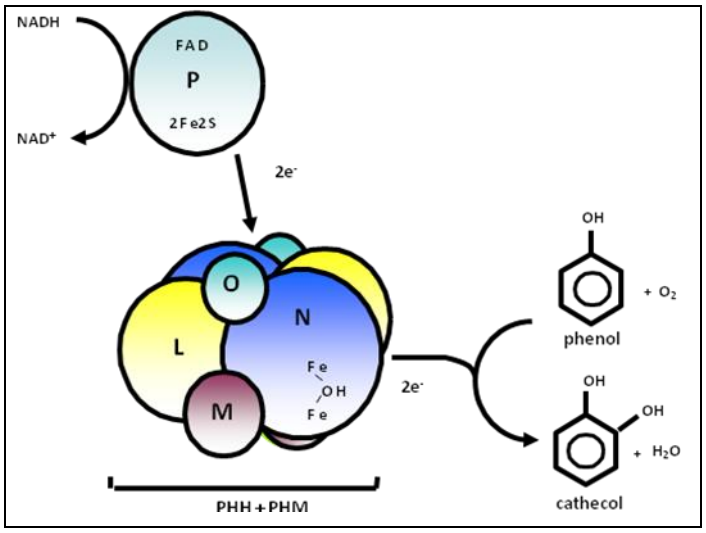


Fig.II.2. Mechanism of electron transfer from NADH to the PHH complex.

A comparative analysis of the polypeptide sequences of PH proteins showed a significant degree of identity with the components of the multicomponent phenol hydroxylase encoded by the dmp genes in *Pseudomonas sp.* CF600. Sequencing of 5' region of *ph* locus coding for the multicomponent phenol hydroxylase of *P. sp*.OX1 revealed the presence of a putative *orf* whose deduced amino acid sequence shared 62.6% identity (74% similarity) to DmpK [13-18]. In the *Pseudomonas sp*. strain CF600 PH, DmpK was suggested to play a role in the expression of active recombinant hydroxylase [18]. Auxiliary proteins that are required for correct metal center assembly are not a unique feature of BMMs since proteins with a similar role have been identified in a number of metalloenzymes including Fe/S proteins, nitrogenase, urease, CO-dehydrogenase, and hydrogenase [19-21]. These components share several common features: they are small proteins with a molecular mass of about 10 kDa, devoid of cofactors, expressed at low levels in their respective native organisms, and inhibiting enzyme activity *in vitro* when present in stoichiometric amounts.

Despite these common aspects the role of accessory proteins in BMMs has been elusive so far, although they have been hypothesized to be involved in the assembly of the active form of the hydroxylase component, as for example in the case of the protein MMOD, isolated from the methanotrophic bacterium *Methylococcus capsulatus* (Bath) [9,19]. Even when devoid of any significant indication of homology, these genes have been identified across all the different groups of BMMs [7,8,22-24],

suggesting that proteins someway involved in the assembly of the active diiron center might be more widespread than it is currently assumed.
Different molecular mechanisms employed to facilitate the insertion of metal sites into target proteins have been reported in literature [25]. However, none of the mechanisms described so far has been unambiguously assigned to proteins such as DmpK and MMOD.

II.4 Aim of the project

This part of the work was focused to gain further insight into the role of PHK, the accessory protein of phenol hydroxylase from *P. sp.* OX1, by a structural proteomics approach, including classical biochemical experiment (gel filtration) integrated with mass spectrometry analysis. Moreover, in order to gain further insight in this molecular process, to provide evidence for the specificity of the binding between PHK and the LNO subunits of the PH hydroxylase moiety, and to verify the possibility of other proteins interacting with PHK, we carried out an affinity purification strategy using an avidin pull-down assay on two different celluar extracts of *Pseudomonas sp.* OX1.
The start up of this project was the observation of an improved efficiency of the PHH complex when coexpressed with the accessory subunit PHK, that is not part of the functional complex, but can be found as a genetic element in most (if not all) of the phenol hydroxylase operons sequenced so far. General chaperone? Metal delivery factor? Metallochaperone? Several hypotheses can be outlined for this component, but none of these had been proved so far. The only evidences being, as stated above, that it is not essential to the enzymatic activity, but, when present during the recombinant production process, the efficiency of the active enzyme is improved.
The phenol hydroxylase gene cluster was cloned and expressed in *E. coli* in the absence or in the presence of the *orf* coding for PHK. In the latter case, recombinant cells showed an increased phenol hydroxylase activity compared to cells not expressing PHK, thus suggesting a positive effect of the accessory protein on the enzymatic activity of phenol hydroxylase. Interestingly, it has been isolated from the soluble extract of cells expressing the complete *ph* gene cluster, including *phk*, active $PH(LNO)_2$ and $PH(LNO)_2$/PHM complexes. It has been also isolated a stable complex formed by PHK bound to a PH(LNO) trimer, devoid of either iron or activity, which could not be reactivated, at least under our experimental conditions. To elucidate the role of PHK, this component was subcloned and overexpressed in *E. coli* and characterized its interaction with the active hydroxylase moiety, $PH(LNO)_2$, showing that the protein, like MMOD and DmpK, is not directly involved in the catalytic activity of the phenol hydroxylase complex. However, PHK does not appear to be required for the production of active recombinant hydroxylase as previously observed for DmpK [26]. PHK seems instead to be involved in increasing the apparent affinity of apo-$PH(LNO)_2$ for iron, thus possibly stimulating the incorporation of the transition metal into the active site of the protein.

II.5 Material and Methods

II.5.1 Analytical gel filtration

Analytical gel filtration experiments were carried out as follows: 200 µl of sample were loaded on a Superdex 200 HR 10/30 or on Superdex 75 HR 10/30 column previously equilibrated in 25 mM Tris-HCl, 5% glycerol, 0.2 M NaCl and 2 mM L-cysteine, pH 7.0 installed on an AKTA™fplc™ (GE Healthcare Life Science), and isocratically eluted at room temperature at a flow rate of 0.5 mL min^{-1}, and monitored at 280 nm. The column had been previously calibrated in the same buffer with the following proteins of known molecular mass: apotransferrin (400 kDa), alcohol dehydrogenase (150 kDa) bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome c (12 kDa). When necessary, the area of each peak was estimated by nonlinear curve-fitting of the elution profile using PeakFit software (Systat Software).

II.5.2 Preparation of iron-reconstituted PHH

The way of reconstituting phenol hydroxylase activity of apoproteins was carried out by adding a diluted solution of $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ directly in eppendorf along with the other purified components in the *in vitro* reconstitution of the PH complex and immediately after having added apo-PHH to the mixture.

II.5.3 Biotinylation of PHK

PHK (3.0 mg, in MOPS 25 mM pH 7.0, 5% glycerol, 0.2 M NaCl) was incubated with 1.37 mg of Sulfo-NHS-Biotin (sulfosuccinimidobiotin) (Pierce) (molar ratio 20:1), dissolved in the same buffer, on ice in the dark for 2 hours. The reaction was stopped with 0.1 M Tris pH 7.0 and the excess of non-reacted biotin reagent was removed by size exclusion chromatography on PD-10 columns (GE-Healthcare) in sodium phosphate 0.1 M pH 7.2, NaCl 0.15 M. , following absorbance at 220 nm and 280 nm to single out K-containing fractions. These fractions were pooled and added to the Avidin Agarose Resin (settled gel, Pierce) in order to immobilize the biotinylated protein.

II.5.4 Immobilization of biotinylated PHK on avidin and "Pull down" assay

450 µl of resin were equilibrated with five volumes of Binding Buffer (Sodium Phosphate 0.1 M pH 7.2, NaCl 0.15 M) and incubated with biotinylated PHK (3 mg biotinylated PHK per mL of settled avidin agarose resin) at 4°C for 1 hour. The resin was washed with 10 volumes of Binding Buffer.
Pseudomonas sp. OX1 cells grown on M9 minimal medium containing 5 mM phenol or 5 mM malat as unique carbon and energy source were disrupted by sonication in MOPS 25 mM pH 7.0, NaCl 0.15 M, 5% glycerol, 1 mM PMSF and protease inhibitors cocktail (Sigma Aldrich). The extracts were centrifuged at 12,000 rpm for 1 hour and filtered with 0.45 µm PVDF filter. Total protein cell extracts, quantified with the BioRad protein assay, were incubated with 125 µl of mouse anti-IgG agarose conjugated beads (Sigma) overnight at 4ºC (preclining step). Cell extracts were then incubated with 450 µl of biotinylated-PHK immobilized on avidin resin and incubated at 4ºC for 2 hours. The resin was washed with 10 volumes of Binding Buffer and the

protein samples were eluted with 70 µl of Laemmli sample buffer. Samples were analyzed by SDS-PAGE (12.5 %), and proteins were stained with Coomassie Brilliant Blue G-Colloidal (Pierce, Rockford, USA).

II.5.5 Mass spectrometric analysis

Identification of complex subunits was carried out on trypsin digested samples either in solution or by *in situ* digestion after fractionation on polyacrilamide gel elecrophoresis. Proteins were detected on the gel using colloidal Coomassie (Pierce, Rockford, USA). Excised bands were destained, reduced with 10 mM DTT, carbamidomethylated with 55 mM iodoacetamide in 0.1 M NH_4HCO_3 buffer, pH 7.5, and subjected to tryptic in-gel digestion for 16 h at 37°C, by adding 100 ng of trypsin. Peak top fractions from analytical gel filtration experiments (1 mL) were lyophilized, resuspended in 100 µl of H_2O, and digested at 37°C for 16 h with 100 ng of trypsin. The reaction was stopped by lowering pH to about 1 with formic acid and the peptide mixtures were concentrated and purified using a reverse phase Zip Tip pipette tips (Millipore). The peptides were eluted with 20 µl of a solution made of 50% acetonitrile, 0.1% formic acid in Milli-Q water. Peptide mixtures were analyzed either by MALDI-MS or capillary LC-MS/MS.
MALDI-MS experiments were performed on a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems, Framingham, MA) equipped with a nitrogen laser (337 nm). Typically, 1 µl of the total peptide mixture was mixed (1/1, v/v) with a 10 mg mL^{-1} solution of R-cyano-4-hydroxycinnamic acid in acetonitrile/50 mM citrate buffer (3/2, v/v). The experimental mass values obtained were compared with calculated masses from the predicted tryptic digestion of the different subunit sequences, confirming the identities of the corresponding protein bands.
The peptide mixtures were analysed using a CHIP MS 6520 QTOF equipped with a capillary 1200 HPLC system and a chip cube (Agilent Technologies, Palo Alto, Ca). After loading, the peptide mixture (8 µl in 0.1% formic acid) was first concentrated and washed at 4 µl/min in 40 nl enrichment column (Agilent Technologies chip), with 0.1% formic acid in 2% acetonitrile as eluent. The sample was then fractionated on a C_{18} reverse-phase capillary column (75 µm x 43 mm in the Agilent Technologies chip) at flow rate of 400 nl/min, with a linear gradient of eluent B (0.1% formic acid in 95% acetonitrile) in A (0.1% formic acid in 2% acetonitrile) from 7 to 60% in 50 min.

II.5.6 Protein identification

Peptide analysis was performed using data-dependent acquisition of one MS scan (mass range from 300 to 2000 m/z) followed by MS/MS scans of the five most abundant ions in each MS scan. Spectral data were analyzed using Mass Hunter software (Agilent Technologies) and raw data from nanoLC-MS/MS analyses were employed to query, using the licensed version of Mascot 2.1 (Matrix Science, Boston, USA), a non-redundant protein databases (NCBI, with taxonomy restriction to Bacteria), or an *ad hoc* created databases including only the sequences of PHK, PHL, PHM, PHN, PHO.
The Mascot search parameters were bacteria as taxonomy restriction, “trypsin” as enzyme allowing up to 2 missed cleavages, carbamidomethyl as fixed modification, oxidation of M, pyroGlu N-term Q, as variable modifications, 10 ppm peptide tolerance and 0.6 Da MSMS tolerance and auto protein entries. Spectra with a

MASCOT score < 25 having low quality were rejected. The score used to evaluate quality of matches for MSMS data was higher than 30.

II.6 Results

II.6.1 Purification and properties of PHK

The oligomeric state of PHK was assessed by analytical gel filtration on a Superdex 75 analytical column. The chromatographic profile resulted in a single peak, which eluted at an apparent molecular mass of about 25,500 ± 400 Da, suggesting the occurrence of a dimeric species (Fig.II.3).

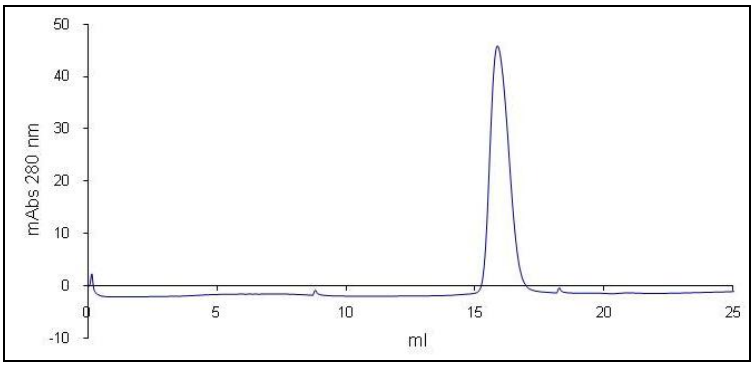


Fig.II.3. Characterization of PHK. Elution profile of analytical gel filtration chromatography of PHK.

The accurate subunit molecular weight of PHK was determined by ESI-MS analysis. The experimental value of 10,235.27 ± 0.19 Da was in agreement with the molecular weight predicted from the sequence of PHK (10,235.7 Da).

II.6.2 Inhibition of Phenol Hydroxylase activity by PHK

Studies carried out at the Prof. di Donato's group demonstrated that the addition of purified PHK to phenol hydroxylase complex resulted in the inhibition of the phenol hydroxylase activity in a dose-dependent way. Specifically, a 10-fold excess of PHK to $PH(LNO)_2$ led to retain only 16.8 ± 3.1% of the initial rate of catechol production. Interestingly, higher PHK:$PH(LNO)_2$ ratios did not lead to complete inactivation of phenol hydroxylase activity.
To gain further insight into this inhibition apparently mediated by the presence of the accessory protein PHK, we incubated, in a final volume of 500 µL of 25 mM MOPS pH 7.0, purified PHK 128 µM with the hydroxylase complex $PH(LNO)_2$ 2.6 µM at a final concentration of 57 µM in a molar ratio of 0.6:1 (PHK: $PH(LNO)_2$). The mixture was incubated at room temperature for 90 minutes, and 15 µL were analyzed on a native 15% polyacrilamide gel (Fig.II.4.A). Compared to samples containing purified $PH(LNO)_2$ and PHK (lanes 2 and 4), the sample containing both $PH(LNO)_2$ and PHK showed the presence of an additional species (lane 3) characterized by a migration pattern similar to that observed for the trimeric form PH(LNO) associated to PHK (lane 1) and previously purified from cell extracts of *E. coli*/JM109 expressing plasmid pGEM3Z/*phΔp*. Protein components of this additional species were identified by peptide mass fingerprinting. The protein band from lane 3 was excised, digested *in*

situ with trypsin, and analyzed by MALDI-TOF MS. Mass values were mapped onto the anticipated sequences of PHL, PHN, PHO, PHM, and PHK subunits revealing the simultaneous presence of PHL, PHN, PHO, and PHK (Table II.1).

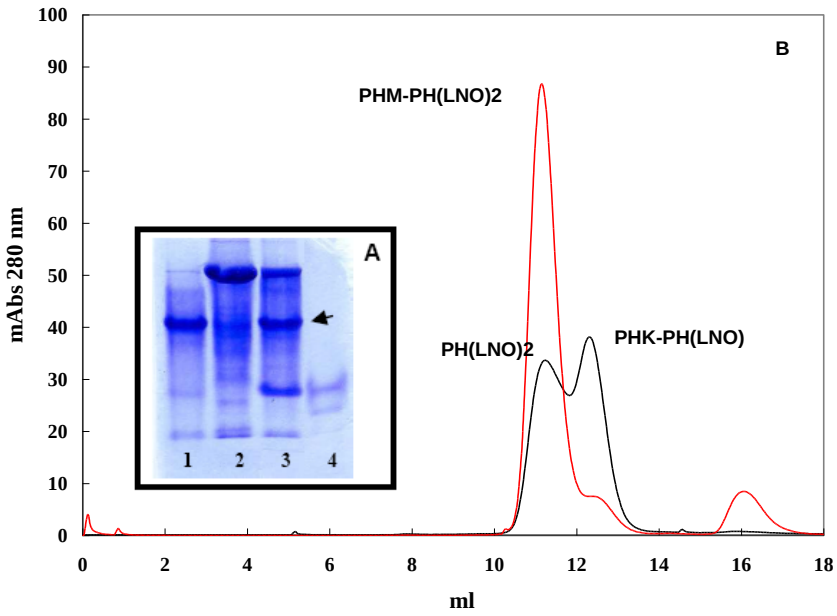


Fig.II.4. Characterization of the $PH(LNO)_2$ complex in the presence of PHK.
A) Native gel electrophoresis of a $PH(LNO)_2$/PHK mixture. *Lane 1*,10 µg of the PHK-PH(LNO) complex as purified from the anion exchange chromatography. *Lane 2*, 20 µg of $PH(LNO)_2$ active hexamer. *Lane 3*, 20 µg of $PH(LNO)_2$ incubated in a 1:0.6 molar ratio with respect to PHK for 90 minutes at room temperature. *Lane 4*, approximately 5 µg of purified recombinant PHK. The arrow indicates the additional band present in the incubated sample and absent in the starting protein samples. ***B***) Elution profiles of analytical gel filtration chromatography of a mixture of 0.88 nmoles of $PH(LNO)_2$ and 0.44 nmoles of PHK (black line), and of a mixture of 0.44 nmoles of PHM-$PH(LNO)_2$ and 0.88 nmoles of PHK (red line).

Phenol Hydroxylase Component	*Matched Peptide*	*Sequence Coverage (%)*
PHN	44	66
PHL	29	71
PHO	4	45
PHK	8	68
PHN	44	66

Table II.1. Identification of the subunit components in the phenol hydroxylase complex fractionated by native gel electrophoresis. The protein band indicated with an arrow in Fig. II.4.A was excised from the gel and *in situ* digested with trypsin. The peptide mixture was analyzed by LC-MS/MS and the experimental data used for searching the database with MASCOT software.

Another 200 µl aliquot of the incubation mixture containing both $PH(LNO)_2$ and PHK was loaded onto an analytical gel filtration carried out on a Superdex 200 column. The elution profile (Fig.II.4.B) reveals a protein peak with an apparent molecular weight of 220 kDa, corresponding to the $PH(LNO)_2$ hexameric complex of the active hydroxylase. Also evident is a new species with an apparent molecular weight of 124 kDa, likely corresponding to the trimeric PH(LNO) species complexed with PHK. Top peak fractions were analyzed by SDS-PAGE and mass spectrometry (data not shown), confirming the presence of PHK only in the peak eluting with the apparent

molecular weight of 124 kDa, whereas PHL, PHN, and PHO were observed in both peaks.
Thus, the molecular species produced by the interaction of PHK with the hexameric hydroxylase $PH(LNO)_2$ can be confidently identified as a PHK-PH(LNO) complex. This species was found to be devoid of both iron and catalytic activity on phenol, thus suggesting that the inhibition of the phenol hydroxylase activity, as observed in the *in vitro* assays and previously reported also in the case of accessory proteins DmpK and MMOD [26-27], depends on the formation of apo-trimers of the type PH(LNO), mediated by the intervention of PHK.
Interestingly, incubation of PHK with complex PHM-$PH(LNO)_2$, performed under the same experimental conditions described above, resulted in very little dissociation of the hexameric hydroxylase. In fact, little or no evidence of a PHK-PH(LNO) complex was found in the gel filtration profile (Fig.II.4.B), thus suggesting a stabilizing effect of PHM against hexamer dissociation. This protecting effect is also in agreement with the observation of a lower amount of PHK-PH(LNO) is formed when $PH(LNO)_2$ is incubated at the same time with M and K in comparison the incubation with the single PHM component (data not shown).

II.6.3 Role of PHK in iron uptake

Since bacterial multicomponent monooxygenases (BMMs) are members of a wide family of nonheme, diiron enzymes which contain a (2Fe-2S)- centre that mediates transfer of electrons from NADH to the active site of the hydroxylase, we performed experiment in order to investigate a possible role of PHK in modulating iron uptake by the PHH hydroxylase moiety, by evaluating the distribution and composition in subunits of the species obtained in the gel filtration runs of several samples when iron depleted $PH(LNO)_2$ (from now on referred to as apo-$PH(LNO)_2$) has been incubated under different conditions.
Apo-$PH(LNO)_2$ was further purified by gel filtration on a Superdex-200 gel filtration column. Two different species were obtained and identified by means of both native-PAGE analysis and analytical gel filtration which showed the presence of both hexameric apo-$PH(LNO)_2$ and trimeric apo-PH(LNO) complexes. Iron uptake by these complexes was attempted by adding increasing amounts of a freshly prepared 50 µM solution of $Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$ in deionized water to both species. Immediately after, the other components of the phenol hydroxylase complex, oxidoreductase PHP and regulatory protein PHM, were added at a ratio of 2:1 and 4:1, respectively, with respect to $PH(LNO)_2$; this latter was used in the assay at a final concentration of 0.75 µM.
In order to evaluate any potential role of iron in the dissociation of the hexameric complex upon incubation with PHK, we evaluated the distribution and composition in subunits of the species obtained in gel filtration experiments performed on iron depleted samples incubated under different conditions (Fig.II.5).
The identity of the subunits eluting in the different peaks of the gel filtration profile was assessed by a proteomic approach: 1 mL fractions corresponding to the peak tops were digested with trypsin and analyzed by MALDI-TOF or, when ambiguous results were obtained, by capillary LC-MS/MS. Proteins were then identified by comparing the experimental data to an in-house built database where the protein sequences of PHL, PHN, PHO, PHK, and PHM had been inserted, using MASCOT software.

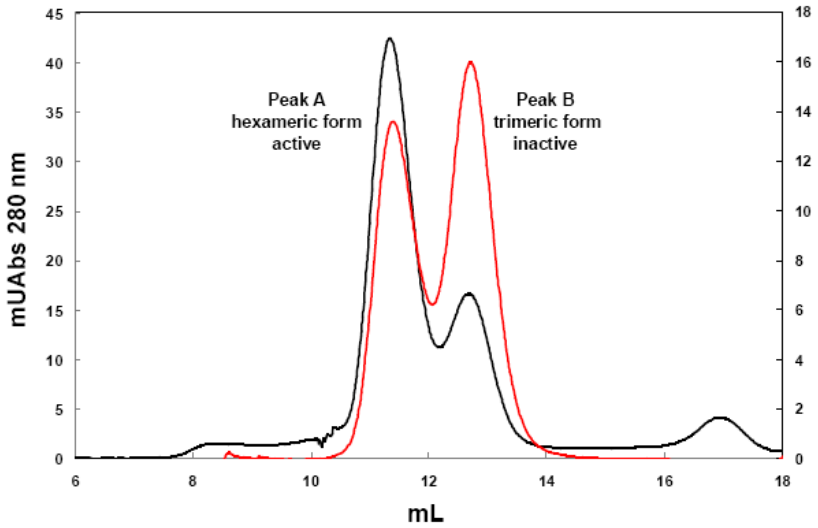


Fig.II.5. Characterization of the apo-PH(LNO)$_2$ complex in the presence of PHK, PHM and iron. Elution profiles of analytical gel filtration chromatography separation of a 200 µL mixture of 2.5 µM of apo-PH(LNO)$_2$ and 1.25 µM of PHK (red line), and of a mixture of 2.5 µM of apo-PH(LNO)$_2$, 1.25 µM of PHK, and 10 µM of PHM in the presence of 10 µM Fe^{2+} (black line). PeakFit software was used to deconvolute the peaks.

As observed with fully active holo-PH(LNO)$_2$, incubation of 2.5 µM of apo-PH(LNO)$_2$ with 1.25 µM of PHK at room temperature for 90 min, in a final volume of 200 µL, induced the dissociation of the hexameric species and the subsequent formation of PHK-PH(LNO) complexes as measured by gel-filtration experiments (Fig.II.5, red line, and Table II.2). Addition of either iron (10 µM) or exogenous regulatory protein PHM (10 µM) to the incubation mixture did not significantly alter the dissociation profile (data not shown).
Importantly, a significant increase of the relative abundance of active hexameric species was observed with respect to the inactive trimeric form when exogenous recombinant PHM (10 µM) and iron (10 µM) were added simultaneously to the reaction mixture containing apo-PH(LNO)$_2$ and PHK (Fig.II.5, black line). Moreover, in the later experiment a significant fraction of the regulatory protein PHM was bound to the active hexameric PH(LNO)$_2$ hydroxylase in a PHM-PH(LNO)$_2$ complex (Fig.II.5, black line, and Table II.2).

Sample	Peak A			Peak B		
	Subunits	N° of matched Peptides	Sequence Coverage (%)	Subunits	N° of matched Peptides	Sequence Coverage (%)
Apo PH(LNO)$_2$ incubated with PHK (FIG.II.6-red line)	PHN[a] PHL[a] PHO[a]	23 22 4	46 59 33	PHN[a] PHL[a] PHO[a] PHk[a]	24 23 5 7	46 61 38 43
Apo PH(LNO)$_2$ incubated with PHK, PHM, and Fe^{2+} (FIG.II.6-black line)	PHN[b] PHL[b] PHO[b] PHM[b]	22 24 5 3	47 75 46 55	PHN[a] PHL[a] PHO[a] PHk[a]	23 24 5 9	46 63 38 64

Table II.2. Identification of the subunits in the chromatographic peaks as separated in Fig.II.5. Proteins were identified in the peak top fractions after tryptic digestion and mass spectrometric analysis. [a] MALDI-TOF analysis. [b] Capillary LC-MSMS analysis.

Unexpectedly, PHK did not bind to pre-formed, trimeric apo-PH(LNO). When apo-PH(LNO) was incubated with PHK, up to a molar ratio of 1:8, we could not detect any

PHK co-eluting with the trimeric species, whereas all the three subunits PHL, PHN, and PHO could be easily identified by proteomics methods (data not shown).
This result strongly suggests that PHK binds to the hexameric form (either apo- or holo-) of the hydroxylase before inducing dissociation to the PHK-PH(LNO) complex.

II.6.4 Avidin pull-down assay using Pseudomonas sp. OX1 cell extract

In order to gain further insight in this molecular process, to provide evidence for the specificity of the binding between PHK and the PHL, PHN, PHO subunits of the PH hydroxylase moiety, and to verify the possibility of other proteins interacting with PHK, we carried out an affinity purification strategy using an avidin pull-down assay on two different extracts of cells of *Pseudomonas sp.* OX1. One grown on phenol as unique carbon and energy source and the other grown on malat (the control), in this last condition the bacteria does not express any PHH complex. This approach required that recombinant PHK was biotinylated in controlled reaction conditions (not more than two biotin molecules introduced in average per PHK molecule) in order to avoid extensive modification of the protein. We biotinylated 3 mg of PHK as described in the Experimental Procedures, section II.5.3, with Sulfo-NHS-Biotin (sulfosuccinimidobiotin); an aliquot was loaded on RP-HPLC and analysed by ES-MS, showing that, in the experimental conditions used PHK was efficiently biotinylated (up to 3 biotins per molecule).
The biotinylated PHK was immobilized on avidin beads that were subsequently incubated with total protein extracts from *Pseudomonas sp.* OX1 cells grown on phenol and on malat. These latter had been pre-cleaned on agarose beads to minimize non-specific binding on the cromatographic matrix during the pull-down procedure. After extensive washing, PHK recruited protein interaction partners were eluted in Laemmli buffer, separated on SDS-PAGE and stained with colloidal blue Coomassie (Fig.II.6). Stained gel displayed a number of discrete bands both in the control (Fig.II.6; lane 4) and sample lanes (Fig.II.6; lanes 3). By comparison of the electrophoretic patterns after elution (Fig.II.6; lanes 3 and 4), the sample profile showed enrichment of some specific bands. Three major proteins with molecular weight at approximately 12 kDa, 40 kDa and 60 kDa (Fig.II.6; lane 3) were pulled-down. Interestingly, the band pattern observed from *P.sp.* OX1 pull down appeared to be similar to purified reference proteins (Fig.II.6, lane 5). Thus the bands present only in the sample were excised from the gel. Each slice was destained reduced, alkylated and digested *in situ* with trypsin. To check for nonspecific pull down, the same procedure was applied to the control sample. The resulting peptide mixtures were analyzed by LC/MSMS analysis. The data obtained were then used for protein identification searching the bacteria subset of NCBI database with the MASCOT software.
All the corresponding protein bands, from the sample and control, were compared. However, all proteins identified in the control that were also detected in the sample, at the corresponding gel slices, were eliminated, in order to reduce the incidence of false positives. The three major proteins were confidently identified as PHL, PHO, PHN, beside the PHK bait. Thus, these results indicate a highly specific interaction of PHK with the LNO subunits of the hydroxylase moiety expressed in the native microorganism *P.sp.*OX1. Those proteins, identified by MASCOT search with at least 2 peptides, and present exclusively in the sample lane were selected, providing a full list of PHK putative interactors [27,28]. Eight putative PHK-interacting proteins were

identified by the above criteria and are listed in Table II.3, together with their corresponding genes and the number of peptides used for their identification.

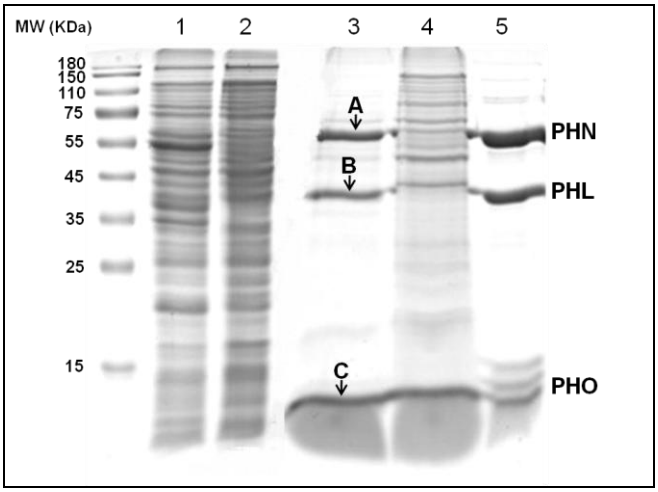


Fig.II.6. SDS-PAGE fractionation of *pull down* experiment with biotinylated PHK. Colloidal blue-stained gel for biotinylated PHK associated proteins from *Pseudomonas sp*. OX1 cells grown on phenol (lane 3) and malat (lane 4). Total *P.sp*.OX1 cell extract is loaded in lane 1 (grown on phenol) and 2 (grown on malat). An aliquot of PHM-PH$(LNO)_2$, was loaded (lane 5) as a control. Biotinylated PHK associated proteins were analysed by SDS-PAGE followed by colloidal blue staining. A, B, and C labels in lane 3 indicate the three major proteins detected among the proteins pulled-down by biotinylated PHK. Protein bands occurring in the sample (lane 3) and in the control (lane 4) were submitted to mass spectral analyses. Common proteins identified in both the sample and the control gel slices were eliminated. Molecular weight markers are also shown.

Gene Name	ID	Protein Name	Number of Peptides	MW (Da)	Sequence coverage (%)
fusA	Q889X4	Elongation factor G	7	77170	13
fadB	Q4KFC4	Fatty acid oxidation complex subunit alpha	7	77124	15
	Q88K15	Isoquinoline 1-oxidoreductase, beta subunit, putative	3	78675	43
htpG	Q4KFX8	Chaperone protein htpG	2	71349	6
phN	Q84AQ2	Phenol hydroxylase component phN	43	60141	77
touA	O87798	Toluene, o-xylene monooxygenase oxygenase subunit	12	57725	33
phO	Q84AQ1	Phenol hydroxylase component O	3	13235	26
touK	Q93JV2	Putative uncharacterized protein touK	3	10078	26

Table II.3. Full list of Phenol Hydroxylase component K (PHK) putative interactors identified by mass spectrometry.

II.7 Discussion

Microorganisms devote part of their genetic and biochemical resources to ensure the functional incorporation of metals into the active site of several enzymes whose catalytic chemistry requires metal ions. Proteins responsible for the tightly regulated homeostasis of metals are generally known as metallochaperones [20,29], proteins which, among others, have also a key role in confining toxic metals - such as iron and copper - to avoid non-specific interactions with cellular membranes, proteins, or DNA [30,31].

Based on data available for components similar to PHK from *Pseudomonas sp.* OX1, such as MMOD from *M. capsulatus* (Bath) and DmpK from *P. sp*.CF600 [18,19], several roles could be envisaged for this protein, including its role as a molecular chaperone, a metal delivery factor, or a metallochaperone. Interestingly, sequences homologous to that coding for PHK can be found in most of the phenol hydroxylase operons sequenced so far [8], thus clearly suggesting a critical role for a component with unclear functions in this specific subfamily of BMMs.
Study carried out at the Prof. di Donato's group indicated that whole cells of *E. coli* strain JM109 coexpressing recombinant multicomponent phenol hydroxylase from *Pseudomonas sp*. OX1 and the accessory subunit PHK were characterized by an improved catalytic efficiency in converting phenol to catechol when compared to the same strain expressing the PH operon deprived of the gene *phk*. This difference was confirmed when crude extracts of induced cells, instead of whole cells, were used in the enzymatic assays. Whereas total iron content was similar in cells expressing the two different *ph* gene clusters, a significant difference was found in their phenol hydroxylase activity, measured as rate of catechol formation. Surprisingly, in cell extracts lacking recombinant PHK, increase of phenol hydroxylase activity could be obtained by the sole addition of an excess of exogenous iron (II) to the sample tested. This result is different from what previously observed with the recombinant hydroxylase of multicomponent PH from *Pseudomonas sp.* CF600 [18]. In this case, in fact, the hydroxylating activity of crude extracts of cells not expressing the accessory component DmpK could be detected only after addition of DmpK together with exogenous iron (II), thus supporting the hypothesis of a role of DmpK not only in the delivery but mainly in the correct assembly of iron into the active site of the hydroxylase [18].
We provided this study with a structural approach to investigate the influence of PHK on phenol hydroxylase complex reconstituted from the purified components PHP (oxidoreductase), PHM (regulatory protein) and $PH(LNO)_2$ (hydroxylase moiety) [6]. PHK was found to be as a dimeric protein, devoid of any metal or organic cofactor. PHK did not seem to be essential for the enzymatic activity of the PH complex on phenol, instead study carried out at the Prof. di Donato's group showed an inhibitory effect on the rate of hydroxylation of phenol to catechol, even when present at low ratios with respect to the hydroxylase moiety $PH(LNO)_2$.
An insight into the molecular details of this inhibition was obtained by screening the molecular species formed after incubating the hydroxylase complex $PH(LNO)_2$ with a substoichiometric amount of purified PHK (Fig.II.4.A). Native gel electrophoresis, and gel filtration experiments followed by mass spectrometry identification of the species formed upon incubation, showed the presence of a novel molecular species, made up by PH(LNO) tightly associated with monomeric PHK. Biochemical assays revealed that this species is devoid of both iron and catalytic activity, providing an explanation for the inhibitory effect observed when exogenous PHK was added to reconstituted phenol hydroxylase in *in vitro* enzyme assays. The ability of PHK to promote dissociation of the hexameric complex $PH(LNO)_2$ was confirmed by the observation that inactive PHK-PH(LNO) could be isolated upon purification of cell extract of E. coli-strain JM109 expressing the complete ph gene cluster from plasmid pGEM3Z/*phΔk*. In the same gel-filtration experiment, the yield of the PHK-PH(LNO) complex is significantly lower when complex PHM-$PH(LNO)_2$ instead of $PH(LNO)_2$ was incubated with PHK (Fig.II.4.B). These results suggest a possible shielding effect of PHM towards the dissociation of the $PH(LNO)_2$ hexamer mediated by PHK. This "protective role" of the coupling protein PHM, observed in different experiments

presented in this work, is similar to what previously described, amog others, for the homologous regulatory component MMOB in the sMMO system from *Methylococcus capsulatus* [29].
Further insight on the possible role of PHK came from experiments of iron uptake by iron-deprived $PH(LNO)_2$ hydroxylase. Study carried out in collaboration with the Prof. di Donato's group showed that in the presence of the accessory protein PHK the recovery of the catalytic activity of apo-$PH(LNO)_2$ occurs at stoichiometric concentration of iron (II). This observation is physiologically relevant because iron uptake and recovery of catalytic activity measured in the absence of PHK occurs at a much higher iron (II) concentration.
Taking into account the data presented in this work and the information available in literature for the accessory components MMOD [19] and DmpK [18], a possible sequence of molecular events in which PHK might be involved can be hypothesized, which could be compatible with our data (Fig.II.7).

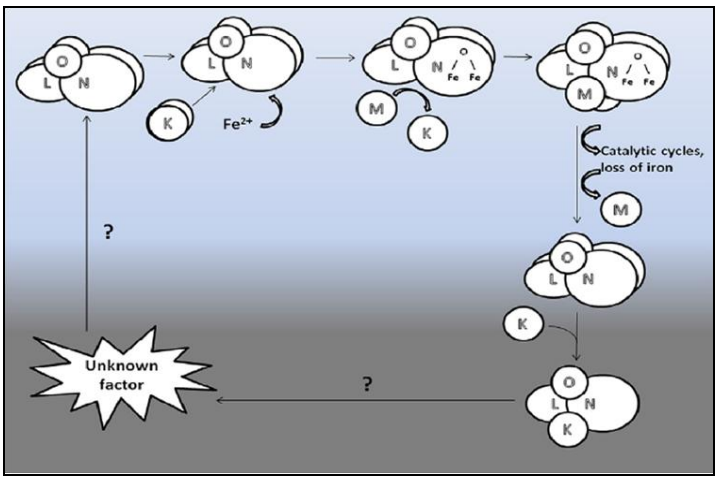


Fig.II.7. Model showing a possible role for PHK. In white and light grey background is the part of the model suggested by the experiments described in this work and the data presented in literature for DmpK and MMOD accessory proteins [18,19] . In dark grey is the sequence of events that is still under investigation. The presence of the oxidoreductase PHP in the PH catalytic cycle, has been omitted in the scheme for clarity.

Lacking any experimental evidence on early post-translational events, the model assumes apo-$PH(LNO)_2$ as the starting point of the sequence of events in which PHK transiently binds to the apo-hexamer influencing the uptake of intracellular iron inside the active site, thus facilitating the formation of holo-$PH(LNO)_2$ *i.e.*, the active hydroxylase moiety.
Based on the ability of PHM to "protect" the holo-$PH(LNO)_2$ from dissociation caused by PHK (Fig.II.4 and Fig.II.5), our model proposes that the regulatory component PHM would bind to the holo-$PH(LNO)_2$ complex, immediately after the PHK-facilitated insertion of iron (II). This event displaces PHK, which can then be immediately recycled to assist in iron insertion in newly assembled apo-$PH(LNO)_2$ molecules.
If inhibition of the phenol hydroxylase activity at PHK:$PH(LNO)_2$ ratios higher than 0.5 is of questionable physiological significance, the PHK-mediated inhibition of the catalytic activity of holo-$PH(LNO)_2$ exerted at substoichiometric ratios, and not detected in the gel filtration experiments, might be a consequence of the fact that, after completion of the hydroxylation reaction, the regulatory protein PHM probably dissociates from the hydroxylase so that the reductase PHP can bind and reinitiate

the catalytic cycle, as previously suggested in literature [10]. The dissociation of the regulatory protein would allow the binding of PHK and the consequent formation of the PHK-PH(LNO) inactive complex.
To support the presence of an interplay between the accessory and the regulatory proteins in BMMs, it should be taken also in account that *Sazinsky* and coworkers [25], using the purified components of the sMMO from *M. capsulatus* (Bath), highlighted, by means of ITC binding experiments, a higher affinity of the regulatory (MMOB) and the accessory (MMOD) proteins for the holo- and the apo- hydroxylase moiety MMOH, respectively. Moreover, *Powlowski* and coworkers [18] had previously advanced the hypothesis that the accessory component DmpK likely interferes with the interaction between the regulatory protein DmpM and the hydroxylase moiety $Dmp(LNO)_2$.
Once the regulatory component PHM is bound to the $PH(LNO)_2$ active complex, the iron is protected inside the active site from accidental loss as indicated by depletion and reconstitution experiments performed on the PHM-$PH(LNO)_2$ complex, and a new catalytic cycle on phenol can be initiated.
The protective role of the regulatory component in the phenol hydroxylase system of *P.sp*. OX1 is particularly evident in the crystal structure of the PHM-$PH(LNO)_2$ complex [10], whose arrangement shows that PHM, among others, functions to block the entrance of the solvent and substrate to the diiron active site through the four-helix bundle, offering some type of protection of the dioxygen-activated metal center as it traverses the reaction cycle [10].
Our current hypothesis is that PHK does not directly bind iron, as suggested by ITC experiments, but might be responsible for altering local conformation of apo-$PH(LNO)_2$, thus facilitating direct access of iron into the hydroxylase active site. It is actually under investigation whether this observation is related to the direct binding of iron to an apo-$PH(LNO)_2$-PHK transitory complex or it is a consequence of a conformational modification induced by PHK in the iron depleted hydroxylase moiety apo-$PH(LNO)_2$.
Among others, the question arises, also evidenced in the scheme of Fig. II.7, on whether the PHK-PH(LNO) complex is of physiological importance. Our current hypothesis is that during repeated catalytic cycles this complex might lose iron and could be coupled to PHK to induce the formation of the PHK-PH(LNO) species we have observed *in vitro* (lower part of the scheme of Fig. II.7). These complexes might reorganize, with the aid of further, yet undiscovered, auxiliary proteins, to form apo-$PH(LNO)_2$ complexes. Alternatively, the presence of PHK might be necessary to avoid the occurrence *in vivo* of holo-$PH(LNO)_2$ molecules in the absence of the regulatory protein PHM. In this case, in fact, the oxidoreductase PHP is still able to transfer electrons to the hydroxylase moiety [6], leading to an uncoupling, or a premature reduction of the oxygenated metal cluster, which would not only consume the reactive diiron species but also deplete the NADH supply of the cell in a wasteful manner [10].
At this stage we cannot exclude that formation of inactive PHK-PH(LNO) complexes could be just an artifact of our system due to the presence of high intracellular concentration of recombinant $PH(LNO)_2$ and PHK. This phenomenon would shift the physiological association equilibrium toward the formation of the PHK-PH(LNO) species. Nevertheless, the stable PHK-PH(LNO) complex might be a valuable tool in the future to shed light on the molecular determinants responsible for the interaction between the accessory protein and the hydroxylase subunits PHL, PHN and PHO.

In conclusion, novel details on the role of accessory proteins in bacterial multicomponent monoooxygenases have been highlighted by the experiments described in this work which add to what has been previously observed in the case of accessory proteins MMOD and DmpK [18,19].
The similarities and the differences among the different accessory proteins characterized so far lead to the conclusion that in the BMMs family similar protein components but diverse strategies are present aimed at modulating the correct assembly of the diiron cluster in the active site of the hydroxylase moiety.

II.8 References

[1] Dagley S. Biochemistry of aromatic hydrocarbon degradation in Pseudomonads, in: J. Sokatch, LN. Ornston, (Eds), The Bacteria, ACIDemic Press, Orlando (1986);10:527-55.

[2] Rieger PG, Meier HM, Gerle M, Vogt U, Groth T, Knackmuss HJ. Xenobiotics in the environment: present and future strategies to obviate the problem of biological persistence. *J Biotechnol.* (2002);94(1):101-23.

[3] Harayama S, Timmis KN. Aerobic degradation of aromatic hydrocarbons by bacteria, in: H.Sigel, A.Sigel (Eds), Metal ions in biological systems, Marcel Dekker, New York (1992);28:99-156.

[4] Díaz E. Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *Int. Microbiol*. (2004);(7):173-180.

[5] Arenghi FL, Berlanda D, Galli E, Sello G, Barbieri P. Organization and regulation of meta cleavage pathway genes for toluene and o-xylene derivative degradation in *Pseudomonas stutzer*i OX1. *Appl Environ Microbiol.* (2001);67(7):3304-8.

[6] Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, Pucci P, Di Donato A. Phenol hydroxylase and toluene/o-xylene monooxygenase from *Pseudomonas stutzeri* OX1: interplay between two enzymes. *Appl Environ Microbiol.* (2004);70(4):2211-9.

[7] Leahy JG, Batchelor PJ, Morcomb SM, Evolution of the soluble diiron monooxygenases, *FEMS Microbiol. Rev*. (2003);27:449-79.

[8] Notomista E, Lahm A, Di Donato A,Tramontano A, Evolution of bacterial and archaeal multicomponent monooxygenases, *J. Mol. Evol*. (2003);56:435-45.

[9] Merkx M, Kopp DA, Sazinsky MH, Blazyk JL, Müller J, Lippard SJ, Dioxygen activation and methane hydroxylation by soluble methane monooxygenase: A tale of two irons and three proteins, Angew. Chem. Int. Ed. (2001);40:2783-807.

[10] Sazinsky MH, Dunten PW, McCormick MS, Di Donato A, Lippard SJ, X-ray structure of a hydroxylase-regulatory protein complex from a hydrocarbon-oxidizing multicomponent monooxygenase, *Pseudomonas* sp. OX1 phenol hydroxylase, *Biochemistry* (2006);45:15392-404.

[11] Bertoni G, Martino M, Galli E, Barbieri P. Analysis of the gene cluster encoding toluene/*o*-xylene monooxygenase from *Pseudomonas stutzeri* OX1. *Appl. Environ. Microbiol.* (1998);64:3626-32.

[12] Bertoni G, Bolognesi F, Galli E, Barbieri P. Cloning of the genes for and characterization of the early stages of toluene catabolism in *Pseudomonas stutzeri* OX1. *Appl. Environ. Microbiol.* (1996);62:3704-11.

[13] Nordlund I, Powlowski J, Shingler V. Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600, J. Bacteriol. (1990);172:6826-33.

[14] Powlowski J, Shingler V. *In vitro* analysis of polypeptide requirements of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600, J. Bacteriol. (1990);172:6834-40.
[15] Powlowski J, Shingler V. Genetics and biochemistry of phenol degradation by *Pseudomonas* sp. CF600. *Biodegradation* (1994);5:219-36.
[16] CIDieux E, Vrajmasu V, Achim C, Powlowski J, Münck E. Biochemical, Mössbauer, and EPR studies of the diiron cluster of phenol hydroxylase from *Pseudomonas* sp. strain CF 600. *Biochemistry* (2002);41:10680-91.
[17] Newman LM, Wackett LP. Purification and characterization of toluene 2-monooxygenase from *Burkholderia cepacia* G4. *Biochemistry* (1995);34:14066-76.
[18] Powlowski J, Sealy J, Shingler V, CIDieux E. On the role of DmpK, an auxiliary protein associated with multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600. *J. Biol. Chem.* (1997);272:945-51.
[19] Merkx M, Lippard SJ. Why OrfY? Characterization of MMOD, a long overlooked component of the soluble methane monooxygenase from *Methylococcus capsulatus* (Bath). *J. Biol. Chem*. (2002);277:5858-65.
[20] Ribbe MW, Burgess BK. The chaperone GroEL is required for the final assembly of the molybdenum-iron protein of nitrogenase. *Proc. Natl. ACID. Sci. USA* (2001);98:5521-25.
[21] Kuchar J, Hausinger RP. Biosynthesis of metal sites. *Chem. Rev.* (2004);104:509-25.
[22] Cardy DL, Laidler V, Salmond GP, Murrell JC. Molecular analysis of the methane monooxygenase (MMO) gene cluster of *Methylosinus trichosporium* OB3b. *Mol. Microbiol.* (1991);5:335-42.
[23] Herrmann H, Müller C, Schmidt I, Mahnke J, Petruschka L, Hahnke K. Localization and organization of phenol degradation genes of *Pseudomonas putida* strain H. *Mol. Gen. Genet.* (1995);247:240-46.
[24] Santos PM, Sá-Correia I. Characterization of the unique organization and co-regulation of a gene cluster required for phenol and benzene catabolism in *Pseudomonas sp.* M1. *J. Biotechnol*. (2007);131:371-78.
[25] Sazinsky MH, Merkx M, CIDieux E, Tang S, Lippard SJ. Preparation and X-ray structures of metal-free, dicobalt and dimanganese forms of soluble methane monooxygenase hydroxylase from *Methylococcus capsulatus* (Bath). *Biochemistry* (2004);43;16263-76.
[26] Kuchar J, Hausinger RP. Biosynthesis of metal sites. *Chem. Rev*. (2004)104;509-25.
[27] Köcher T, Superti-Furga G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* (2007);4(10):807-15.
[28] Monti M, Orrù S, Pagnozzi D, Pucci P. Interaction proteomics. *Biosci Rep* (2005);25: 45-56.
[29] Kuchar J, Hausinger RP. Biosynthesis of metal sites. *Chem. Rev.* (2004);104:509-25.
[30] Daniel KG, Harbach RH, Guida WC, Dou QP. Copper storage diseases: Menkes, Wilson, and cancer. *Front. Biosci.* (2004);9:2652-62.
[31] Bartnikas TB, Gitlin JD. How to make a metalloprotein. *Nat. Struct. Biol.* (2001);8:733-734.

III. Molecular basis of schizophrenia: a multidisciplinary investigation to new diagnostic and therapeutic approaches

III.1 Molecular basis of schizophrenia

Schizophrenia is a chronic and severely debilitating psychiatric disorder affecting nearly 1% of the world's population and accounts for ~2.5% of the healthcare costs of the European Community. It is characterized by positive symptoms that include hallucinations, delusions, and thought disorder, by negative symptoms comprised of affective flattening and social isolation, and by profound cognitive deficits in attention, learning, memory, and behavioural flexibility [1,2]. Current antipsychotic treatments for schizophrenia show success in reducing the severity of positive symptoms, but have limited efficacy in ameliorating negative and cognitive deficits [1]. Furthermore, antipsychotic regimes are often poorly tolerated, leading to poor compliance and symptomatic relapse [2].

In order to develop effective therapies, much effort has been made to further understand the molecular alterations involved in the pathophysiology of schizophrenia. Efforts to identify the underlying disturbances in schizophrenia are currently focused on three general lines of inquiry: (i) examination of the mechanism of action of the drugs that alleviate the symptoms of schizophrenia, (ii) examination of neuroanatomical abnormalities in the brains of schizophrenia patients, and (iii) examination of candidate genes that confer susceptibility to schizophrenia.

Abnormalities in several neurotransmitter systems have been implicated in the pathophysiological processes underlying schizophrenia. The predominant theory has been the dopamine hypothesis, which postulates that schizophrenic symptoms arise from excessive dopaminergic transmission, particularly in the striatum, and the presence of dopaminergic deficits in prefrontal brain regions [3]. In addition to the dopaminergic abnormalities, *N*-methyl-D-aspartate of glutamate receptors (NMDAR) hypofunction has been proposed to be involved in schizophrenia [4-9].

This theory originated from studies demonstrating that non-competitive NMDAR antagonists like phencyclidine (PCP) reliably and immediately induce a syndrome similar to schizophrenia in healthy individuals and exacerbate symptoms in schizophrenia patients [4,10]. Since these initial observations, convergent evidence has supported a role for aberrant NMDAR mediated neurotransmission in schizophrenia pathogenesis [11,12] and the glutamatergic system in schizophrenia is considered to be part of a larger complex framework involving the interaction of multiple neurotransmitters and risk genes. More recently, studies have indicated that the NMDAR D-serine/glycine binding site and its modulatory enzymes may be crucially involved in the glutamatergic dysfunction thought to occur [11,12].

III.2 The NMDAR: structure and regulation

NMDARs play a key role in excitatory synaptic transmission and have been implicated in many physiological processes, including learning and memory.

NMDARs in the central nervous system (CNS) are heteromeric protein complexes composed of at least one NR1 subunit together with different combinations of NR2 and/or NR3 subunits [13]. Alternative splicing of the *Grin1* gene produces eight NR1 isoforms and by associating with different constellations of NR2 (NR2A-D) and NR3 subunits (NR3A and NR3B) form a multitude of different NMDAR receptors with

distinct biophysical properties [14] and specific patterns of expression during development and in the mature mammalian CNS [15]. NMDAR complexity is further enhanced through post-translational modifications, such as phosphorylation, glycosylation, and ubiquitination, affecting cellular localization and function of the receptor [16]. The NMDAR also contains a glutamate recognition site on the NR2 subunit and a glycine or D-serine modulatory site on the NR1 subunit (Fig.III.1) [17]. The D-serine/glycine site on the NMDAR must be occupied for glutamate to activate the receptor [18].

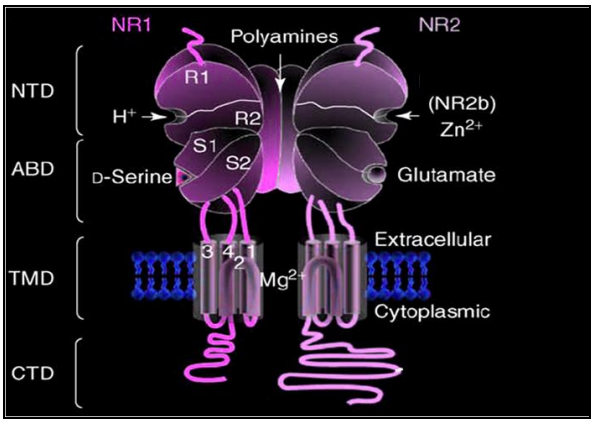


Fig.III.1. The NMDA receptor. D-Serine is synthesized in astrocytes and released near NMDA receptors (NMDARs), where it regulates glutamatergic neurotransmission. NMDAR requires ligand binding at two sites to open an ion channel. Thus, glutamate binds to NMDAR only if the NMDAR glycine site is occupied by the agonist.

Once activated, the NMDAR channel permits the influx of calcium, which stimulates intracellular signalling cascades that can subsequently affect synaptic plasticity and gene transcription [19]. NMDARs are present throughout the brain and are principally neuronal, though they can also be expressed on astrocytes. Beyond the glutamate and D-serine/glycine binding site, they contain several regulatory sites sensitive to polyamines, Zn^{2+}, protons, and glutathione [13,16]. The numerous influences that converge on the NMDAR highlight the importance of these receptors in diverse brain functions. In NMDAR complexes containing NR1 and NR2 subunits, D-serine and glycine both have excitatory effects, with D-serine being up to three times more potent than glycine [20]. Binding to the D-serine/glycine site allosterically influences the NMDAR to enhance the affinity and efficacy of glutamate [21], delays receptor desensitization to increase the duration and frequency of the open channel state [22], and promotes NMDAR turnover through priming of the receptor for internalization [23]. As a coagonist with glutamate, D-serine mediates NMDAR responses and long-term synaptic changes in the hippocampus, a region linked to learning and memory [24,25]. Interestingly, therapeutic trials with D-serine have been shown to significantly improve symptoms (positive, negative, and cognitive) in patients with schizophrenia, suggesting that D-serine may play an important role in the pathophysiology of schizophrenia [26-28].

The discovery of D-serine in the brain revolutionized the longstanding belief that only L-isomers of amino acids existed in mammalian tissues. It is now accepted that higher organisms produce D-amino acids (and in particular D-serine) by the enzyme serine racemase (SR), which directly converts L-serine to D-serine [29], in both astrocytes and neurons [30]. D-Serine was found to be a highly selective endogenous activator of the NMDAR D-serine/glycine site [24,31]. In the adult human

and rodent brain, D-serine and SR are predominantly localized to the forebrain, with high levels in the cerebral cortex and hippocampus, and minimal levels in the cerebellum and brainstem [32].
The low D-serine concentrations in caudal brain areas coincide with the emergence of D-amino acid oxidase (DAAO), the D-serine catabolic enzyme [33].

III.3 D-amino acid oxidase (DAAO) and pLG72. Two proteins involved in schizophrenia pathology

D-Amino acid oxidase (DAAO) is a FAD dependent flavoenzyme, localized in the peroxisome that at physiological pH catalyzes with a strict stereospecificity the oxidative deamination of D-amino acids to the corresponding imino acids, with concomitant reduction of FAD [34]. The reduced flavin is subsequently reoxidized by molecular oxygen generating H_2O_2 and the imino acid is released into solvent, where it spontaneously hydrolyzes to the corresponding α-keto acid and ammonia. (Fig.III.2). This enzymatic activity has been identified in most eukaryotic organisms, the only exception being plants.
The human flavoenzyme can be distinguished from the other known DAAOs because it is a stable 80 kDa homodimer of 347 amino acid, even in the apoprotein form, because the non-covalent binding of the FAD cofactor is the weakest among the known DAAOs (Kd value is $8x10^{-6}$ M) [35,36] (Fig.III.3).

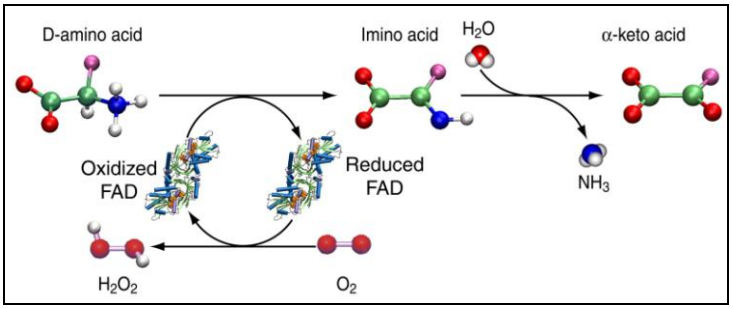


Fig.III.2. Scheme of the reaction catalyzed by D-amino acid oxidase (DAAO).

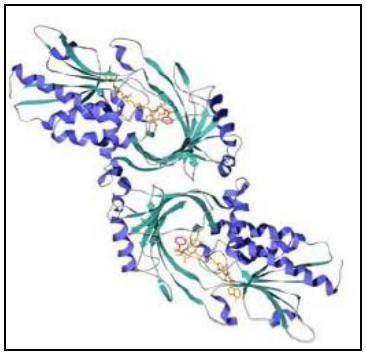


Fig.III.3. Human DAAO oloenzyme tridimensional model. (PDB code 2E4A).

The brain distribution of DAAO is inversely related to that of endogenous D-serine concentrations, with the highest levels of DAAO in astrocytes, Golgi-Bergmann glia, and tanycytes of the hindbrain and cerebellum [37]. Lower levels of DAAO have been detected in the neurons of the prefrontal cortex, hippocampus, and substantia nigra [38]. In the periphery, DAAO is most highly expressed in the kidneys and liver [39].

Modulating DAAO function is pLG72 (also known as LG72 or DAAO activator). pLG72 was initially identified by Chumakov et al. [40], who examined markers within a 5-Mb segment from chromosome 13q33, a region that had previously been linked to schizophrenia in earlier linkage analyses. Moreover, the genes encoding for human DAAO (hDAAO) and its putative binding partner pLG72 have been linked to schizophrenia susceptibility through genetic analysis of different human population [41].

The gene *G72* encodes for several splicing isoforms; pLG72 represents the longest open reading frame (153 amino acids), which is present only in primates and is mainly expressed in brain [41].

A yeast two-hybrid screening experiment identified D-amino acid oxidase as the interacting partner of pLG72 and *in vitro* assays confirmed a regulatory effect of pLG72 on DAAO activity [41,42].

The physiopathologic hypothesis proposed by Chumakov suggests pLG72 as an activator of DAAO (Fig.III.4 arrow 1), and therefore, they proposed that the onset of schizophrenia may result from the overexpression of pLG72, which induces hyperactivation of DAAO and, ultimately, decreases D-serine concentrations at the synapse resulting in hypofunction of NMDAR.

A recent study carried out by Sacchi et al. indicates that pLG72 may instead repress hDAAO activity (Fig.III.4 arrow 2) [42]. By using gel filtration analysis and surface plasmon resonance, they demonstrated that *in vitro* pLG72 specifically interacts with both the holo- and apoprotein forms of hDAAO, yielding an ~200-kDa complex constituted by 2 hDAAO homodimers (2x80 kDa) and 2 pLG72 molecules (2x20 kDa); *Kd* for the complex formation is ~$8x10^{-6}$ M [42-43].

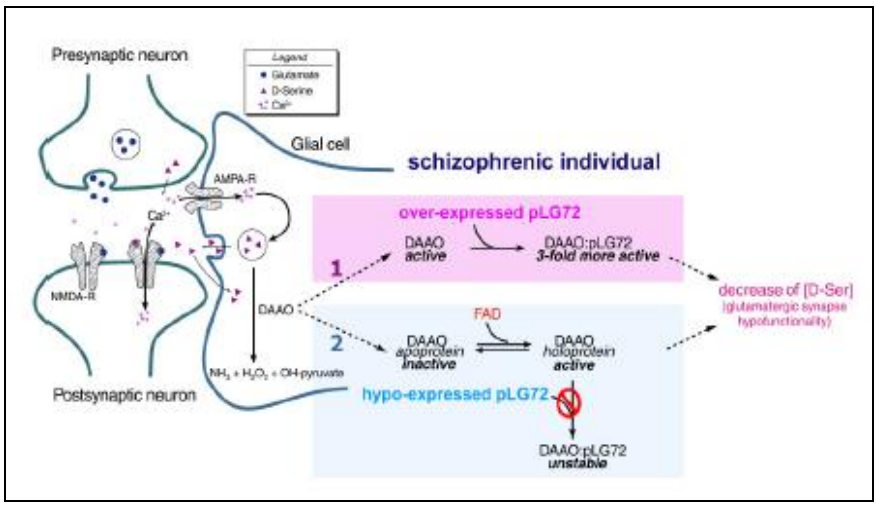


Fig.III.4. Hypothesis of the role of DAAO and pLG72 on the D-serine bioavailability at glutamatergic synapses under normal and pathological conditions. Glutamate is released by the presynaptic neuron into the synaptic space after the depolarization of nerve terminals, yielding the activation of non-NMDA receptors on the membrane of the postsynaptic neuron and, subsequently, an increase in cytoplasmic concentration of Ca^{2+}. D-Serine is then released by the glial cells and, together with glutamate, activates NMDA receptors on the membrane of the postsynaptic neuron, leading to the opening of ion channels. The Ca^{2+} entry induces several possible intracellular responses, *e.g.* it activates nitric-oxide synthase that produces NO, which can diffuse to the neighboring cells where it is proposed to inhibit serine racemase and (probably) activate DAAO, thus decreasing the local concentration of D-serine. (1) Hypothesis of involvement of DAAO and pLG72 in schizophrenia according to Chumakov *et al.* in which pLG72 is an activator of DAAO. (2) Sacchi's hypothesis, in which pLG72 modulates the amount of active hDAAO acting on the stability of the holoenzyme.

They demonstrated that pLG72 binding did not affect the catalytic efficiency and the kinetic parameters of the reaction catalyzed by the enzyme on D-serine, the affinity for coenzyme, or the rate constant of FAD binding to the apoprotein. The main effect observed was a faster time course of hDAAO inactivation when an excess of pLG72 was present, which they attributed to the decrease in the amount of the active form of holoenzyme is solution. In fact, the amount of cofactor bound to the enzyme decreased after the pLG72-hDAAO complex was formed. In their hypothesis pLG72 binding does not affect hDAAO functionality but, rather, modifies its tertiary structure and results in a time-dependent inactivation of the flavoenzyme. Furthermore, they confirmed *in vivo* the hDAAO-pLG72 interaction, by coimmunoprecipitation of the two proteins from human cortex extracts. By immunofluorescent assay they identified hDAAO and pLG72 in the same astrocytes and they demonstrated the partial subcellular colocalization of hDAAO and pLG72 in human glial cells. These finding support the possibility of interaction between these two proteins *in vivo*, suggesting that pLG72 might indirectly modulate NMDA receptor function (at least in cortex), controlling the levels of D-serine by acting on hDAAO.
Therefore, they proposed a molecular mechanism (Fig.III.5) by which hDAAO and pLG72 are involved in schizophrenia susceptibility: an anomalous increase in hDAAO activity (e.g., related to pLG72 hypoexpression) will result in a decrease in the synaptic concentration of D-serine, thus causing hypofunctionality of NMDA receptor mediated neurotransmission. Importantly, and in agreement with the aforementioned model, hDAAO is now considered the target for a new class of drugs (hDAAO inhibitors) to treat schizophrenia [44-45].

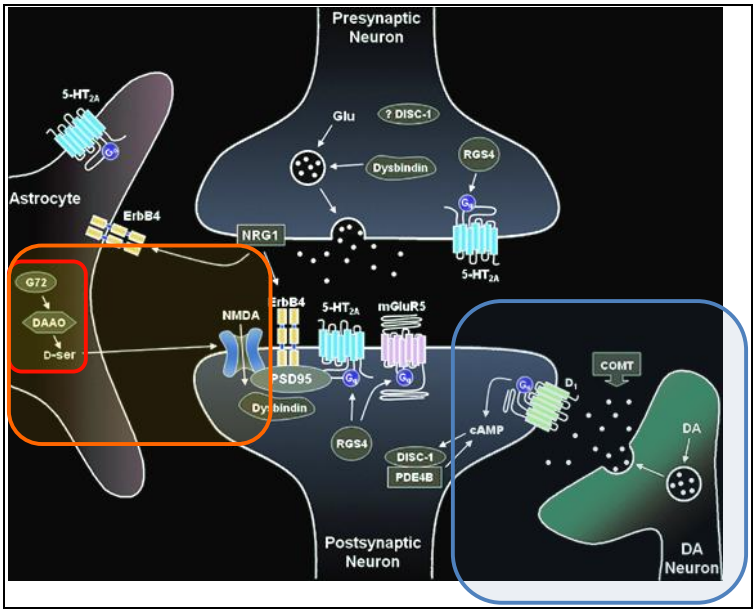


Fig.III.5. Diagram illustrating the regulation of NMDA receptors by astrocytic D-serine. In brain, hDAAO oxidises D-serine, a molecule that acts on glutamate signalling pathway through the binding to the glycine-site of N-methyl-D-aspartate (NMDA) receptor. hDAAO and pLG72 proteins are involved in the molecular mechanism underlying schizophrenia by modulating the (cellular) levels of the neuromodulator D-serine and altering glutamatergic neurotransmission through an hypofunction of NMDA-receptors.

III.4. Aim of the project

Schizophrenia is a multi-factorial disorder with both genetic and environmental contributions and up to now represented a puzzle for the researchers. In recent years, specific genes have been discovered that influence susceptibility to schizophrenia. The new challenge now is to identify the specific mechanisms by which the implicated genes (and protein products) alter the risk of schizophrenia and the molecular and cognitive processes that link these primary events to psychopathology. In recent years, the pLG72 protein and D-amino acid oxidase (DAAO) have been linked to schizophrenia [41].
This part of the work was aimed at providing structural details on the interaction between hDAAO and pLG72. These proteins control the concentration of important molecules (D-serine, L-proline and glutamate) involved in neurotransmission and thus this study is important for understanding the molecular basis of the onset of schizophrenia and, thereofore, for the development of therapeutic approaches as well as new specific drugs.
The molecular mechanism by which DAAO expression and acquisition of the catalytic activity is regulated is still unknown. Moreover i) structural informations on pLG72 are lacking, ii) pLG72 is highly unstable (it is soluble only in presence of mild denaturant) and iii) no homologous protein has been structurally characterized so far. These features make the characterizations of the complex extremely challenging and hardly feasible by high resolution techniques [42].
This part of the work was focused on the structural characterization of the hDAAO-pLG72 complex. Low resolution strategies, such as those based on the coupling of classical biochemistry approaches (limited proteolysis and chemical cross-linking) with mass spectrometric techniques, allowed for the *in vitro* analysis of the surface topology of pLG72-hDAAO complex, as well as the definition of the conformational changes of hDAAO associated to complex formation.
Since hDAAO is likely to be finely regulated in its activation/inhibition by protein partners, we have also performed a proteomic approach on extracts from U87 cells expressing hDAAO as c-Myc tagged protein. Our purpose was defining hDAAO role in the schizophrenic pathology through a comprehensive analysis of its interactome.

III.5 Material and Methods

Recombinant hDAAO and pLG72 were expressed in *E. coli* and purified as previously reported in [35] and [43]. The water-soluble, amino group-selective bifunctional cross-linking reagent Bis(sulfosuccinimidyl)suberate (BS^3) was from Pierce Biotechnology (Rockford, IL, USA). Trypsin TPCK-treated, trypsin proteomic grade, chymotrypsin, subtilisin and endoprotease Glu-C were purchased from Sigma. Acetonitrile Ultra gradient from Romil.

III.5.1. Complementary and limited proteolysis experiments

Enzymatic hydrolyses of hDAAO, pLG72, and hDAAO-pLG72 complex were performed at 37°C in 20 mM NaPPi pH 8.5, 150 mM NaCl, 5% glycerol, 0.06% NLS, 40 µM FAD and 1 mM DTT, with either trypsin, chymotrypsin, subtilisin or endoprotease Glu-C, enzyme-to-substrate ratios ranging from 1:5000 to 1:100 (w/w). The time-course of proteolysis was monitored by sampling the reaction mixture at

different time intervals (from 1 to 30 min). The reaction was stopped by adding 1% TFA or denaturing buffer (300mM Tris-HCl, pH 8.0, 6M guanidine, 10mM EDTA) and the incubation mixtures at different interval times were separated by RP-HPLC on a Vydac C4 column (250 x 4.6 mm, 300 Å pore size) with a 10 - 65% acetonitrile linear gradient in 0.1% TFA over 30 minutes, at a flow rate of 1 mL/min. Elution was monitored at 220 nm and individual fractions were collected and analysed by ESI-MS on a ZQ single quadrupole instrument (Waters-Micromass). Protein solution was injected into the ion source at a flow rate of 10 µL/min. Data were elaborated using the MassLynx program (Waters-Micromass). Mass calibration was performed by means of multiply charged ions from a separated injection of horse heart myoglobin (Sigma, average molecular mass: 16951.5 Da); all masses are reported as average values.

III.5.2. Chemical cross-linking experiments

hDAAO, pLG72, and hDAAO-pLG72 complex were equilibrated in 20 µL of 20 mM HEPES pH 8.0, 150 mM NaCl, 3 mM EDTA, 0.06% NLS and 1 mM DTT at 25°C for 15 minutes and then the cross-linking reaction was started by the addition of Bis(sulfosuccinimidyl)suberate (BS^3). The reactions were incubated at 25 °C for 30 min and terminated by quenching the excess of reagents with Laemmli sample buffer. Preliminary experiments were carried out on single hDAAO and pLG72 proteins, and on the hDAAO-pLG72 complex to determine the optimal protein concentrations and the cross-linking reagent to protein molar ratios.

III.5.3. In situ and in solution digestion

The protein samples were heated at 100°C for 5 min and separated by electrophoresis on a 15% SDS polyacrylamide gel. Proteins were stained with Coomassie Brilliant Blue G-Colloidal (Pierce, Rockford, USA), and selected stained protein bands positive to the western blotting analysis were excised from the gel.
Excised bands were destained by repetitive washes with 0.1 M NH_4HCO_3, pH 7.5, and acetonitrile. Samples were reduced with 10 mM DTT in 0.1 M NH_4HCO_3 buffer, pH 7.5, carboxyamidomethylated with 55 mM iodoacetamide in the same buffer. Enzymatic digestion was carried out with 100 ng of trypsin in 10 mM ammonium bicarbonate buffer, pH 7.8. Gel pieces were incubated at 4°C for 2 h. Trypsin solution was then removed and a new aliquot of the same solution was added. Samples were incubated for 18 h at 37°C. Peptides were then extracted by washing the gel particles with 10 mM ammonium bicarbonate and 1% formic acid in 50% acetonitrile at room temperature and then analyzed by capillary LC-MS/MS.
As a parallel approach the cross-linking reaction was stopped by adding TFA 1% and an in solution hydrolysis with trypsin was perfomed (E/S 1/50 w/w) for 16 h at 37°C. The mixture was analysed by LC-MS/MS.

III.5.4. Western blotting analysis

For immunodetection, after SDS-PAGE, the gel was run in Tris-Glycine buffer at 15 V for 1 h, and transferred to a PVDF membrane (Hybond-P, Amersham Biosciences). The membrane was saturated with 5% non-fat milk in TBS (10mM Tris-HCl pH 8.5, 0.5M NaCl), Tween® 0.1% for 1 h at room temperature and then incubated with rabbit polyclonal antibody anti-DAAO, or goat polyclonal antibody anti-

pLG72 at 4°C for two hours. An anti-rabbit or donkey anti-goat IgG were used as a secondary antibody and Dura system was used (Thermo) for detection.

III.5.5. Mass spectrometry analysis

LC-MS/MS analyses were carried out on a CHIP MS 6520 QTOF equipped with a capillary 1200 HPLC system and a chip cube (Agilent Technologies, Palo Alto, Ca). After loading, the peptide mixture (8 µl in 0.1% formic acid) was first concentrated and washed at 4 µL/min in 40 nl enrichment column (Agilent Technologies chip), with 0.1% formic acid in 2% acetonitrile as eluent. The sample was then fractionated on a C18 reverse-phase capillary column (75 µm x 43 mm in the Agilent Technologies chip) at flow rate of 400 nL/min, with a linear gradient of eluent B (0.1% formic acid in 95% acetonitrile) in A (0.1% formic acid in 2% acetonitrile) from 7 to 60% in 50 min. Peptide analysis was performed using data-dependent acquisition of one MS scan (mass range from 300 to 2000 m/z) followed by MS/MS scans of the five most abundant ions in each MS scan. Dynamic exclusion was used to acquire a more complete survey of the peptides by automatic recognition and temporary exclusion (30 sec) of ions from which definitive mass spectral data was previously acquired. Nitrogen at a flow rate of 3 L/min and heated to 325°C was used as the dry gas for spray desolvation. MS/MS spectra were measured automatically when the MS signal surpassed the threshold of 1000 counts. Double and triple charged ions were preferably isolated and fragmented over single charged ions. Spectral data were transformed in MzData format using Mass Hunter software (Agilent Technologies) and analysed using the licensed version of Mascot 2.1 (Matrix Science, Boston, USA), in a local database or using Protein Prospector software.
Putative cross-linked peptides were identified with the strategy developed by Chu et al. by searching for arbitrary mass modifications greater than 138 Da on Lys residues or at protein N-termini using Protein Prospector version 5.6. These searches were performed with parent mass tolerance of 10 ppm, and a fragment tolerance of 0.6 ppm. Carbamidomethyl cysteine was searched as a constant modification. Variable modifications included: methionine oxidation, and peptide N-terminal glutamine cyclization to pyroglutamate. Mass modifications of any integer value between 138 Da (corresponding to intramolecular cross-linking) and 5000 Da on lysine residues or protein N-termini were searched as variable modifications. Mass modification searches were performed against a restricted database consisting of only hDAAO and pLG72.

III.5.6. Immunoaffinity Purification of hDAAO-Associated Proteins

hDAAO complexes were affinity purified from 8 mg of total extracts prepared from human U87 glioblastoma cells transfected with expression vectors encoding for Myc-hDAAO fused proteins. U87 cells were disrupted by sonication in lysis buffer (50mM Tris-HCl, pH 8.0, 0.15 M NaCl, 0.1% Triton X-100, 1mM EDTA, 1mM DTT, 1mM PMSF) and protease inhibitors cocktail (Sigma Aldrich). The extracts were centrifuged at 12,000 rpm for 30 min at 4°C and filtered with 0.45 µm PVDF filter. Total protein cell extracts, quantified with the BioRad protein assay, were incubated with 160 µl of mouse anti-IgG agarose conjugated beads (Sigma) overnight at 4ºC (preclining step). Beads were pretreated with 5% non-fat milk. After clarification, extracts were centrifuged at 3000 rpm for 5 min at 4°C. Cell extracts were then incubated overnight with polyclonal anti-Myc IgG conjugated agarose beads (Santa

Cruz). The beads were washed extensively with lysis buffer. Control samples were prepared in parallel using untransfected cell extracts. Protein samples were eluted with Myc competitor peptide (200 µg/mL), precipitated in methanol/ clorophormium, resuspended in Laemmli sample buffer. Samples were analyzed by SDS–PAGE (12.5 % - 20 cm x 20 cm), and proteins were stained with Coomassie Brilliant Blue G-Colloidal (Pierce, Rockford, USA).

III.5.6.1. In situ digestion and Western Blotting Analysis

Identification of complex subunits was carried out on trypsin digested samples by *in situ* digestion after fractionation on polyacrilamide gel electrophoresis as previously described (sections III.5.3). Each step of the experimental procedure was paralleled by Western blotting step, as previously described III.5.4 section. After blotting the PVDF membrane was incubated with mouse monoclonal antibody anti-C-Myc (Santa cruz) at 4°C for two hours. An anti-mouse peroxidase-conjugated secondary antibodies (Sigma) was used as a secondary antibody and Dura system was used (Thermo) for detection.

II.5.6.2. Mass spectrometry analysis and protein identification

The peptide mixtures were analysed by LC-MS/MS analysis as previously described (section III.5.5). The Mascot search parameters were Human as taxonomy restriction, trypsin as enzyme allowing up to 2 missed cleavages, carbamidomethyl as fixed modification, oxidation of M, pyroGlu N-term Q, as variable modifications, 10 ppm peptide tolerance and 0.6 Da MS/MS tolerance and auto protein entries. Spectra with a MASCOT score < 25 having low quality were rejected. The score used to evaluate quality of matches for MSMS data was higher than 30.

III.6. Results

III.6.1. Complementary and limited proteolysis

Limited proteolysis experiments were designed to study the surface topology of hDAAO-plG72, on the basis that the portions of the protein that are involved in highly structured regions or are located at the inter-subunits interface are protected from proteases. In order to properly define the changes induced in the single protein components upon formation of the complex, the experiments were in parallel carried out on the isolated proteins.

III.6.1.2 Complementary and limited proteolysis of hDAAO

hDAAO, 1 nmol in 200 µL of 20mM sodium pyrophosphate pH 8.5, 150mM NaCl, 5% glycerol 0.06% NLS, 40 µM FAD and 1mM DTT, was individually incubated at 37°C with trypsin (enzyme-to-substrate ratio (E/S) 1/100 w/w), chymotrypsin (E/S 1/100 w/w), subtilisin (E/S 1/500 w/w), and endoprotease GluC (E/S 1/500 w/w).
The extent of the enzymatic hydrolysis was monitored on a time-course basis by sampling the incubation mixture at different interval times followed by RP-HPLC analysis. The fragments released from the hDAAO molecule were identified on the

basis of the accurate molecular mass determined by ES/MS, leading to the assignment of cleavage sites.
As an example, Fig.III.6, shows the RP-HPLC chromatograms of hDAAO aliquots withdrawn at 0, 5, and 15 min after endoprotease Glu-C addition. As expected, most of the protein remained undigested (fraction 1) and only two fragments were detected in the chromatograms. ES/MS analysis identified fractions 2 and 3 as peptides 1-249 and 250-347, respectively, generated by a single proteolytic cleavage at Glu249.
Fig.III.7 shows the absorbance profiles of the RP-HPLC analysis of the aliquots withdrawn at 0, 5, 15 min of digestion with trypsin. Peak 1 was identified as the mixture of undigested protein (1-347) and the peptide 1-341 (experimental mass value 38811.56 ± 1.70 Da, theoretical mass value 38811.16 Da). The hDAAO protein is immediately cleaved at Arg341 with the removal of C-terminal hexapeptide. The flexibility and accessibility of the C-terminal region deduced from these data is consistent with the three-dimensional structure of the protein determined by X-ray analysis, which could not be solved in the C-terminal portion due to the high flexibility of this region.
Because of the very fast proteolysis at the C-terminal end, and with the confidence of the structural data that represent the C-terminal arm as an exposed peptide not stably interacting with the hDAAO molecule, we confidently assumed that the loss of the 342-347 peptide does not significantly alter the native conformation of the protein. Therefore, in the experiments with trypsin, the truncated form 1-341 was considered as the basic molecule of hDAAO, on which subsequent primary hydrolysis occurred. Fractions 2 and 3 were identified as peptides 1-191 and 192-341, respectively, indicating the occurrence of another further cleavage site at Arg191.
Similar limited proteolysis experiments were then carried out using subtilisin and chymotripsin: these proteases identified cleavage sites at Gly 249 and at Leu 189, Leu 194 and Leu 230, respectively (see Table 1).
The overall results of the limited proteolysis experiments on hDAAO are summarized in Table III.1.

Protease (Protease/Substrate ratio (w/w))	**Identified Peptide**	**Experimental mass value (Da)**	**Theoretical mass value (Da)**	**Proteolytic site**
Trypsin (1/100)	1-341	38811.56±1.70	38811.16	R^{341}
	1-191 192-341	21763.63±1.46 17066.90±1.00	21762.60 17066.50	R^{191}
Chymotrypsin (1/100)	1-250 251-347	28392.60±1.75 11098.27±0.37	28393.20 11098.7	L^{250}
	1-194 195-347	22089.81±1.99 17404.48±1.75	22088.0 17404.0	L^{194}
	1-189 190-347	21476.70±1.53 18013.41±1.60	21478.3 18013.7	L^{189}
Glu-C (1/500)	1-249 250-347	28280.49±1.69 11211.94±1.03	28280.49 11211.90	E^{249}
Subtilisin (1/500)	1-189 190-347	21478.68±1.22 18041.14±0.59	21478.3 18013.7	L^{189}

Table III.1. Peptides generated in the digestions of hDAOO with proteases and identified by ES/MS analysis after fractionation by RP-HPLC.

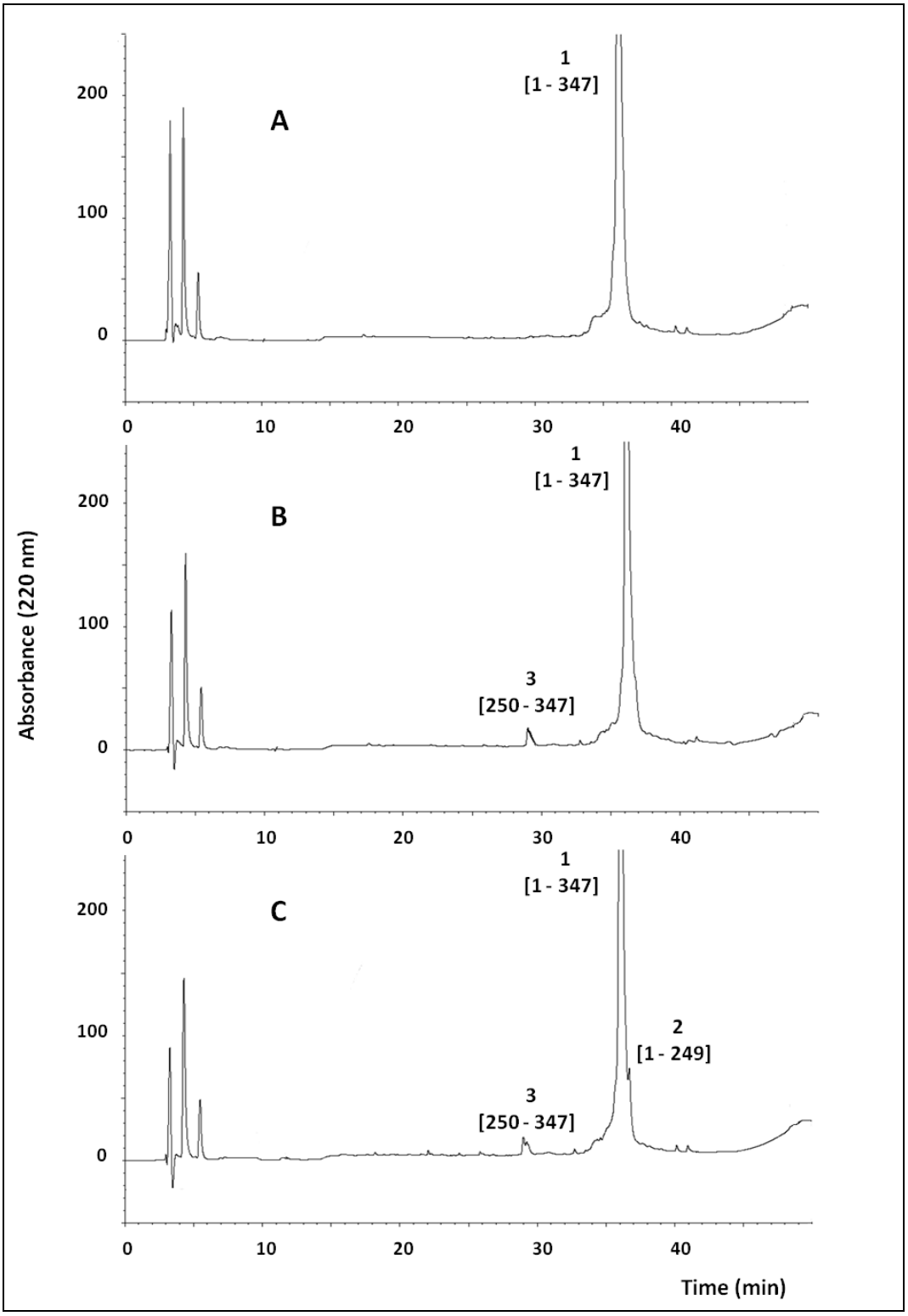


Fig.III.6. Time-course analysis of hDAAO digested with endoprotease Glu-C under controlled conditions using an enzyme/substrate ratio of 1:500 (w/w). HPLC chromatograms of the aliquots withdrawn from the incubation mixture at (A) 0 min, (B) 5 min, and (C) 15 min. Individual fractions were collected and analyzed by ES/MS.

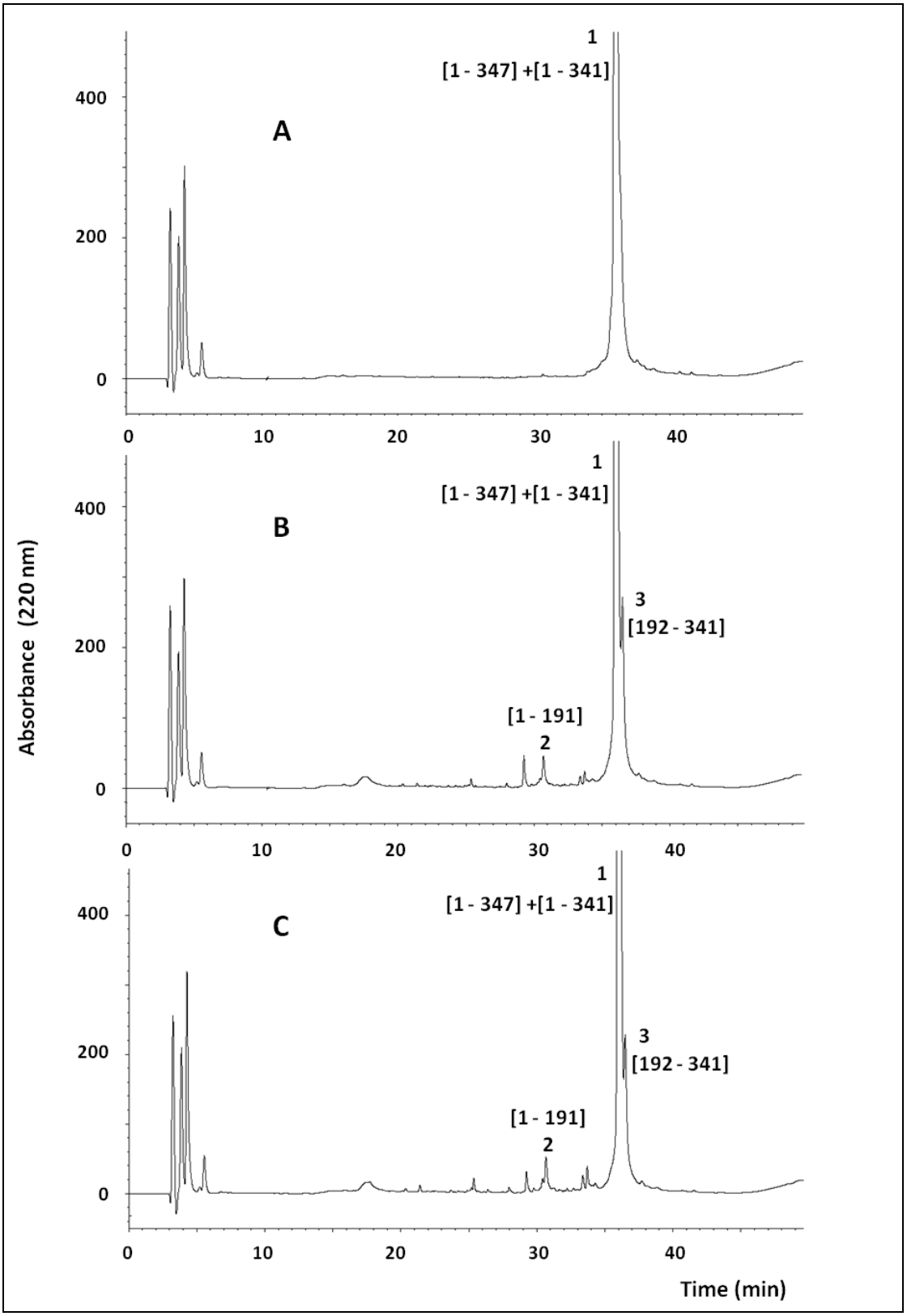


Fig.III.7. Time-course analysis of hDAAO digested with trypsin under controlled conditions using an enzyme/substrate ratio of 1:100 (w/w). HPLC chromatograms of the aliquots withdrawn from the incubation mixture at (A) 0 min, (B) 5 min, and (C) 15 min. Individual fractions were collected and analyzed by ES/MS.

III.6.1.3. Complementary and limited proteolysis of pLG72

The same experimental strategy described above was used to investigate the surface topology of the pLG72 protein. The optimal enzyme-to-substrate ratio was 1/2000 for trypsin, chymotrypsin and endoprotease Glu-C, and 1/5000 for subtilisin. In all cases, the time course of fragments formation was investigated by RP-HPLC followed by molecular mass determination by ES/MS analysis.
As an example, Fig.III.8 shows the absorbance profiles of the RP-HPLC analysis of the aliquots withdrawn following 0, 5, 15 min of trypsin digestion. Only two fragments were released from the intact protein after 5 min of incubation. On the basis of their accurate molecular mass, as determinate by ES/MS analysis, the fragments could be identified as peptides 1-128 and 129-153 which accumulated rapidly as the digestion time increases. These fragments were generated by the single proteolytic event at Arg128.
The overall results of the limited proteolysis experiments on pLG72 are summarized in Table III.2.
With the only exception of cleavage at site Trp 61 by chymotripsin, all the employed proteases acted on residues belonging to the C-terminal end of this small protein.

Protease (Protease/Substrate ratio (w/w))	**Identified Peptide**	**Experimental mass value (Da)**	**Theoretical mass value (Da)**	**Proteolytic site**
Trypsin (1/2000)	1-128 129-153	15110.15±1.49 2988.58±0.15	15109.22 2988.28	R^{128}
Chymotrypsin (1/2000)	1-61 62-153	7030.38±0.38 11066.37±0.73	7030.08 11067.41	W^{61}
	1-134 135-153	15908.85±1.12 2188.96±1.40	15908.20 2189.34	W^{130}
Glu-C (1/5000)	1-146	17347.57±1.22	17948.66	E^{146}
Subtilisin (1/2000)	1-150	17751.44±1.51	17751.11	T^{150}
	1-116 117-153	13643.55±0.99 4454.52±1.00	13642.59 4454.89	Y^{116}

Table III.2. Peptides generated in the digestions of pLG72 with proteases and identified by ESI-MS analysis after fractionation by RP-HPLC.

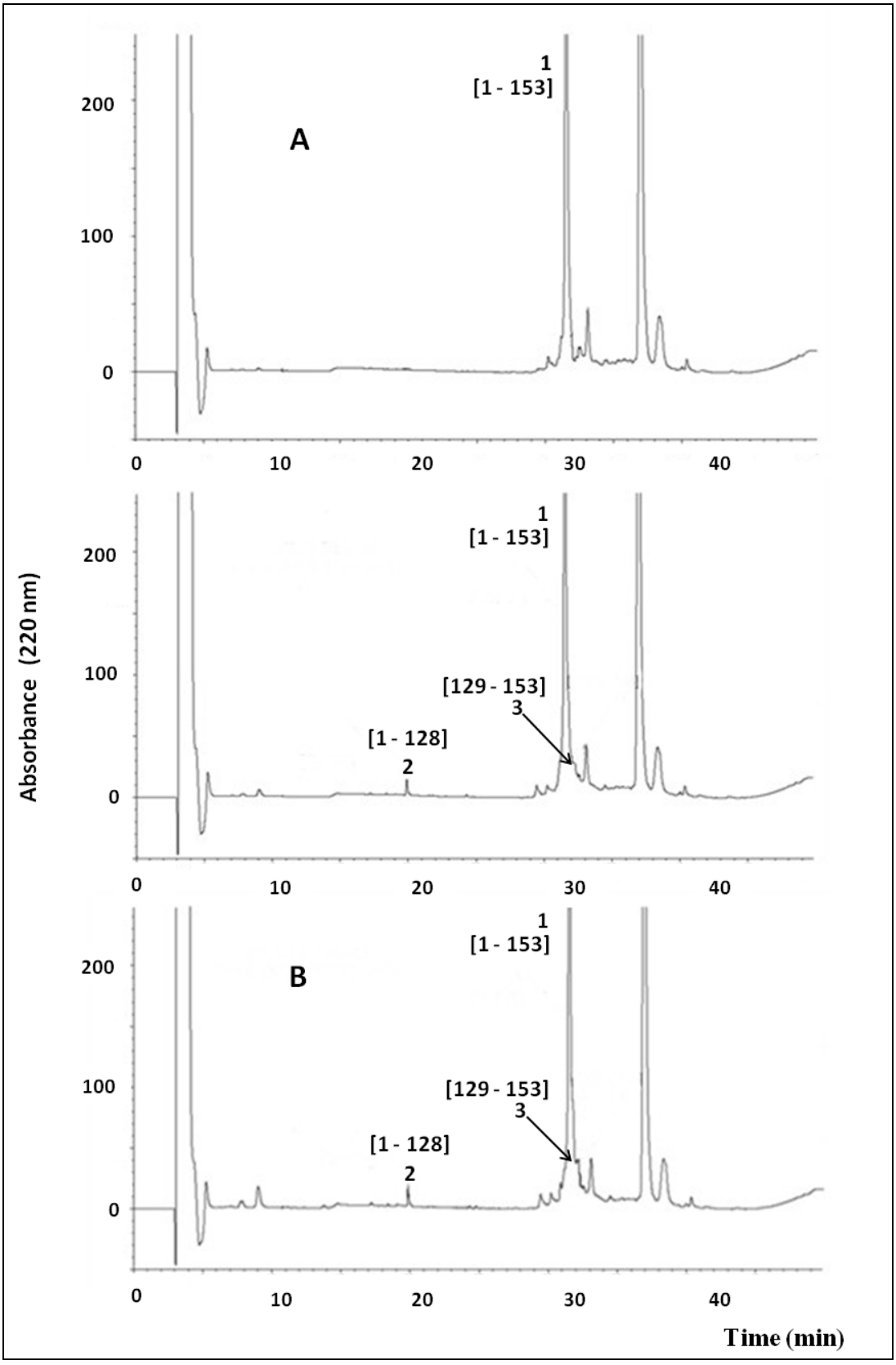


Fig.III.8. Time-course analysis of pLG72 complex digested with trypsin under controlled conditions using an enzyme/substrate ratio of 1/2000 (w/w). HPLC chromatograms of the aliquots withdrawn from the incubation mixture at (A) 0 min, (B) 5 min, and (C) 15 min. Individual fractions were collected and analyzed by ES/MS.

III.6.1.4. Complementary and limited proteolysis of DAAO-pLG72 complex

When the experiments were carried out on the hDAAO-pLG72 complex, in order to ensure that all of flavoenzyme was complexed with pLG72, we worked with a molar excess of pLG72 protein. Consequently, it is not possible to deduce any structural information about the pLG72 regions involved in the complex because most of the pLG72 molecules are not complexed with hDAAO, providing therefore results that have to be ascribed to free pLG72.
The hDAAO-pLG72 complex was formed by incubating hDAAO (4 nmol) with a 5:1 molar excess of pLG72 (20 nmol) protein at 0°C prior to proteolytic enzyme addition. Formation of the complex was confirmed by native gel analysis (data not shown). The optimal enzyme-to-substrate ratio was 1/2000 for trypsin, chymotrypsin and endoprotease Glu-C, 1/10000 for subtilisin.
Fig.III.9 shows the RP-HPLC analysis of the aliquots withdrawn at 0, 5, 15 min of digestion with subtilisin. ES/MS determination identified peaks 1 and 2 as hDAAO and pLG72 intact proteins, respectively. Fractions 3 and 4 were identified as fragments 1-167 and 1-180 of hDAAO protein, respectively, and fraction 5 contained the complementary peptides 168-347 and 181-347.
The overall results of the limited proteolysis experiments on the hDAAO-pLG72 complex are reported in Table III.3 and compared to those obtained on the isolated hDAAO in Figure III.12.

Protease (Protease/Substrate ratio (w/w))	**Identified Peptide**	**Experimental mass value (Da)**	**Theoretical mass value (Da)**	**Proteolytic site**
Trypsin (1/2000)	1-297 298-347	34004.15 ± 1.51 5488.24 ± 0.32	34003.70 17066.50	R^{297}
Chymotrypsin (1/2000)	1-35 36-347	3851.66±0.13 35640.87±1.0	3851.62 35640.35	Y^{35}
	1-55 56-347	6056.99±0.5 3343.86±1.48	6057.07 33434.90	Y^{55}
	1-296 297-347	33846.89±1.39 5644.50±0.43	33847.51 5644.47	L^{296}
Glu-C (1/2000)	1-336	38138.09±1.10	38198.39	E^{336}
Subtilisin (1/10000)	1-167 168-347	19237.07±0.82 20255.54.±1.81	20619.31 18872.69	F^{167}
	1-180 181-347	20619.58±0.84 18872.35±0.64	20619.31 18872.69	N^{180}

Table III.3. Peptides generated in the digestions of hDAAO when in complex with pLG72, with proteases and identified by ESI-MS analysis after fractionation by RP-HPLC.

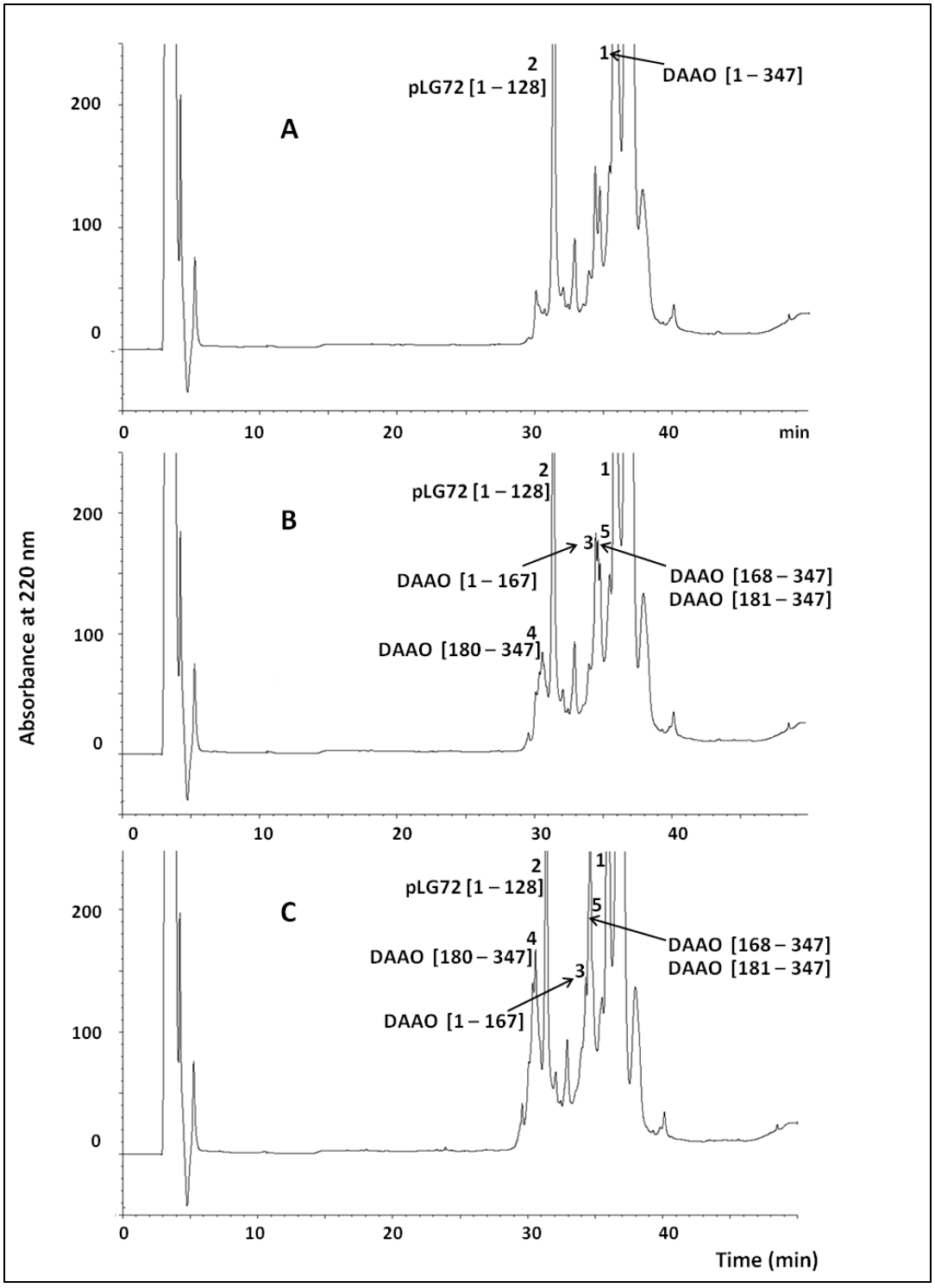


Fig.III.9. Time-course analysis of hDAAO/pLG72 complex digested with subtilisin under controlled conditions using an enzyme/substrate ratio of 1/10000 (w/w). HPLC chromatograms of the aliquots withdrawn from the incubation mixture at (A) 0 min, (B) 5 min, and (C) 15 min. Individual fractions were collected and analyzed by ES/MS.

III.6.2. Chemical crosslinking of hDAAO-pLG72 complex

A general strategy of studying the contact regions in protein complexes is by chemical crosslinking.
In this study, the interacting regions in the hDAAO-pLG72 complex were investigated by chemical cross-linking with the homobifunctional reagent Bis(sulfosuccinimidyl)suberate (BS^3), that contains an amine-reactive N-hydroxysulfosuccinimide (NHS) ester at each end of an 8-carbon spacer arm.
hDAAO and pLG72 (1 nmol and 5 nmol, respectively, in 20 µL) were incubated in ice. After 10 min, BS^3 was added to the cross-linking reaction and incubated at 25°C for 30 min. For quenching the reactions 20 mM NH_4HCO_3 was added to each aliquot.
Several concentrations of BS^3 and different reaction times were tested and the final experimental conditions were selected as those resulting in the formation of an hDAAO-pLG72 heterodimeric complex, avoiding formation of high molecular mass oligomers and preventing excessive cross-linking.
As an example, Fig.III.10, panel A, shows the electrophoretic analysis of the cross-linking reaction on hDAOO-plG72 complex, obtained by mixing hDAOO and pLG72 in a 1:5 molar ratio (monomer of hDAOO:plG72) treated with 5, and 50 molar excess of BS^3 in respect to hDAOO monomer. The reaction was carried out for 30 min at 25°C. However, upon cross-linking treatment we could only detect (lane 6 and 8 of panel A, Fig.III.10), the appearance of a specie migrating at an apparent molecular weight of about 37 kDa, an apparent molecular weight that is even lower than monomeric hDAAO (migrating at about 40 kDa).

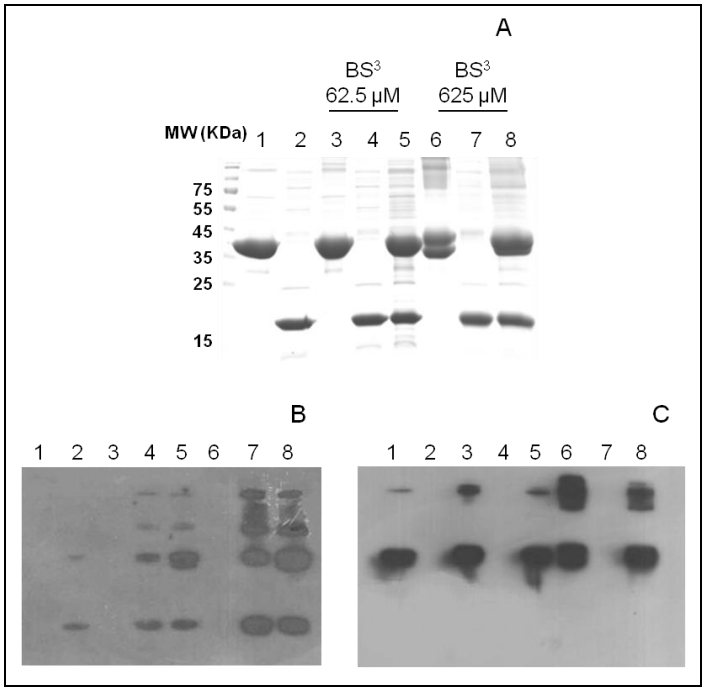


Fig.III.10. hDAAO-pLG72 complex cross-linking with BS^3. **A**) SDS-PAGE of reaction mixtures of hDAAO (12.5 µM) and pLG72 (62.5 µM) with BS^3. The samples are labelled as follows: 1)hDAAO 2) pLG72 3) hDAAO with BS^3 62.5 µM 4) pLG72 with BS^3 62.5 µM, 5) hDAAO-pLG72 complex with BS^3 62.5 µM 6) hDAAO with BS^3 625 µM 7) pLG72 with BS^3 625 µM 8) hDAAO-pLG72 complex with BS^3 625 µM. All the reactions were carried out at 25°C for 30 min in 20 mM Hepes pH 8.0 and stopped by the addition of Laemmli sample buffer. Western blotting analyses with an antibody directed against pLG72 **B**) and against hDAAO **C**).

As a further analysis, the different mixtures of products generated by the cross-linking reaction were analyzed by Western blotting with anti-pLG72 and anti-hDAAO, as shown in Fig.III.10, panel B and C, respectively. It can be observed that the band at about 37 kDa is positive to anti-DAAO staining, while immunodetection with anti-pLG72 antibodies indicates that, upon treatment with BS^3 pLG72 is also present in a band at about 40 kDa, comigrating with hDAAO, probably due to two molecules of pLG72 cross-linked.
The identity of the gel band was definitively assessed by identification of the protein components by *in situ* digestion followed by mass spectrometry analysis. The protein bands were excised from the gel and subjected to enzymatic *in situ* digestion with trypsin. The resulting peptide mixture was directly analyzed by LC-MS/MS and the corresponding mass values were mapped onto the anticipated sequences of hDAAO and pLG72 revealing the simultaneous presence of both proteins.
In order to locate the crosslinks in the protein molecules, the general strategy is the *in situ* digestion of the gel band where both proteins are present and that is generated upon treatment with the cross-linking agent, followed by LC-MS/MS analysis of the petide mixture and careful inspection of the MS/MS spectra. Usually, the advantage of *in situ* digestion is that the cross-linked sample can be isolated from not cross-linked materials, thus reducing the complexity of the mixture.
A possible disadvantage is that it is difficult to extract from the gel band some peptides, manly the cross-linked ones, because of their dimensions. We decided, therefore, to perform, as an alternative approach, an in-solution digestion of the cross-linked hDAAO-pLG72 complex, in order to improve sample recovery. The mixture was analysed by LC-MS/MS in presence of an exclusion list created with the theoretical tryptic hDAAO and pLG72 peptides. The exclusion list would favour the instrument to select low abundant species, such as the cross-linked ones, in respect to the highly abundant ions generated by the expected and canonical digestion of the proteins.
This strategy, coupled with the use of the open mass modification option in the Prospector software, allowed to identify several putative ion species that could account for cross-linked peptides, either intermolecular or intramolecular, whose spectra manual inspection is currently in progress. As an example, Fig.III.11 shows the MS/MS spectrum of the triply charged ion $[M+3H]^{3+}$ at m/z 546.29 that can be interpreted as an intramolecular cross-link between the peptides $K^{163}VESFEEVAR^{172}$ and $M^1R^2$of the hDAAO. The fragment ions corresponding to these peptides are therefore with either the α or β subscript, respectively, to indicate the peptide of origin. The only amines in these two sequences that would have been available for cross-linking in the intact protein are Lys-163 and the amino-terminus of M-1. The MS/MS spectrum revealed a complete y-ion series for the 163-172 peptide chains, indicated with $y_{1\alpha\text{-}9\alpha}$. Moreover there are three ions originate from cleavage reactions involving both peptide chains, i.e., m/z 671.39 ($b_{2\alpha}b_{2\beta}$; KV-MR), m/z 800.44 ($b_{3\alpha}b_{2\beta}$; KVE-MR) and 887.46 ($b_{4\alpha}b_{2\beta}$; KVES-MR). It is important to notice the specie at m/z 444.23 corresponding to the MR peptide plus 138 Da, which is the mass of the spacer arm of BS^3 cross-linker.
The strategy described above produced promising data, that are suggestive of some distinct putative cross-links between hDAAO and pLG72, but the careful manual inspection of the spectra collected is still in progress.

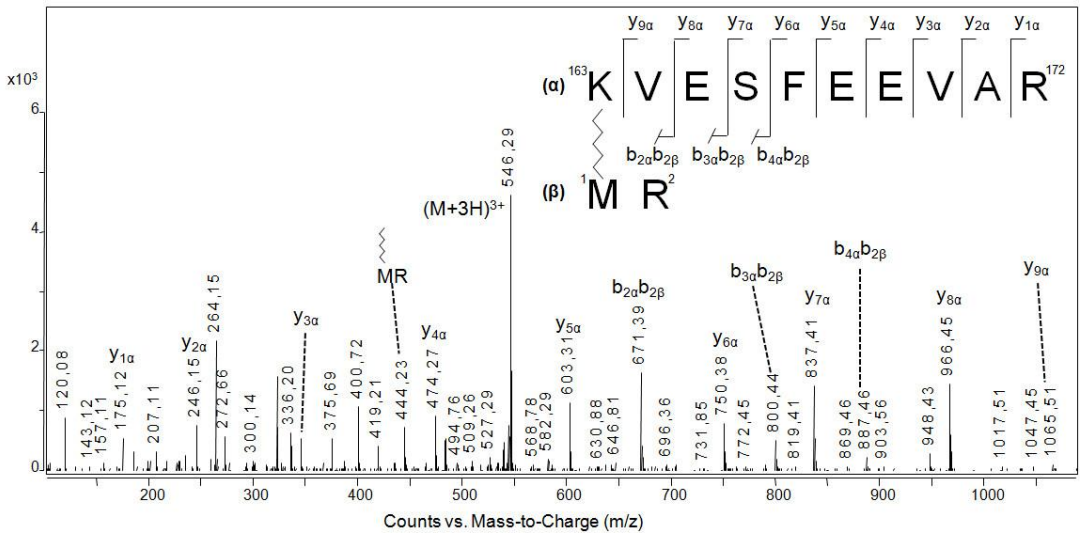


Fig.III.11. ESI-MS/MS spectra of cross-linked peptides KVESFEEVAR and MR of hDAAO with cross-link between N-term of M-1 and Lys-163. Tandem mass spectra of the precursor ion $[M+3H]^{3+}$ at m/z 546.29. Fragments from peptide KVESFEEVAR are labelled with an α subscript, those from peptide MR with a β subscript. The cross-linked sites on the sequence are indicates with a crossed line.

III.6.3. Identification of novel human DAAO-interacting proteins in neuronal cells

Proteins specifically interacting with hDAAO were identified in neuronal cells (U87, human glioblastoma-astrocytoma, epithelial-like cell line) by a functional proteomics approach, using transfected Myc-tagged hDAAO protein as bait. Total protein extract (8 mg) from Myc-tagged hDAAO transfected U87 cells was precleared with underivatized agarose beads to minimize non-specific binding on the cromatographic matrix during the immunoaffinity procedure. The unbound material was immunoprecipitated with polyclonal anti-Myc IgG conjugated agarose beads. After extensive washing, proteins specifically retained were extensively washed and eluted with a solution of Myc peptide. Control experiments were carried out, in parallel, using extracts from untransfected U87 cells (Figure III.12.B, lane 1 and lane 2). Following each step of the procedure, Western blotting was performed on a sample aliquot using anti-Myc antibodies as a control (Figure III.12.A). The protein mixture from the Myc elution was subjected to SDS-PAGE and visualized with Coomassie Blue staining (Figure III.12.B). Stained gel displayed a number of discrete bands both in the control and sample lanes (lanes 1 and 2, respectively). A band occurring at approximately 55 kDa, absent in the pool of non-specifically bound proteins and in the control, was positive to the western blotting staining with anti-Myc antibodies, as well as anti-hDAAO antibodies. (Fig.III.12, panel A).
Because of the complexity of the gel patterns and the low resolution of 1D electrophoresis, several proteins can occur in the same gel band. Therefore, protein bands specifically present in the sample lane and absent in the control lane cannot be identified by simply comparing the two gel profiles. Thus, the entire lanes from the sample and the control were cut in 38 slices each and the slices were destained, reduced, alkylated and digested with trypsin. The resulting peptide mixtures were analyzed by LC-MS/MS analysis generating sequence information on individual peptides. The obtained data were then used for protein identification searching the human subset of NCBI database with the MASCOT software.

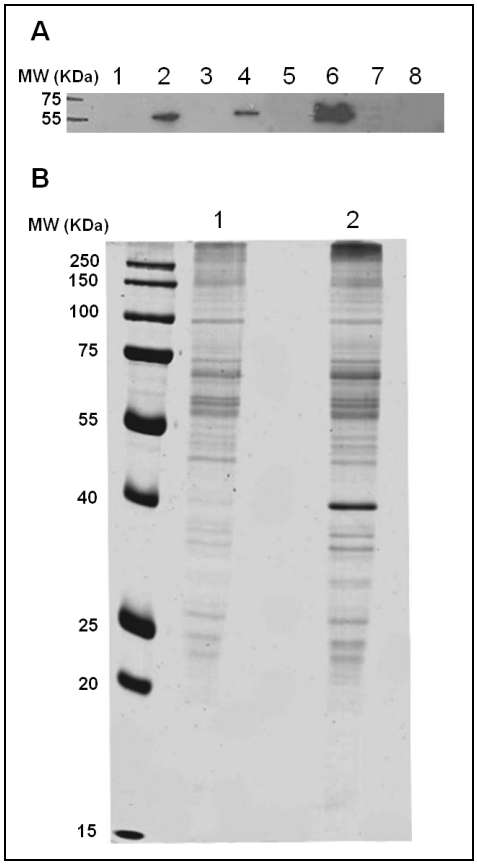


Fig.III.12. Western blotting and SDS-PAGE fractionation of immunoprecipitation experiment with Myc-hDAAO. (A) Western blotting of extracts from Myc-hDAAO transfected human U87 cells (sample). Untransfected human U87 cell extracts were used as control. Aliquots (20 μg) corresponding to each step of the experimental procedure were subjected to Western blotting and revealed with anti-Myc and peroxidise conjugated secondary antibodies. Whole cell extract control (lane 1) and sample (lane 2). Unbound control (lane 7) and sample (lane 8). Elutions with Myc-peptide competitor, control (lanes 3 and 5), sample (lanes 4 and 6). (B) SDS-PAGE fractionation of hDAAO protein complexes. hDAAO protein complexes were isolated by immunoprecipitation experiment, sample (lane 2) and control (lane 1). Molecular weight markers are also shown.

All the gel bands from the sample and control lanes were then compared on the basis of the proteins effectively identified in each gel slice. Common proteins identified both in the control and the sample slices were discarded and only those proteins solely identified in the sample slices and absent in the control were considered. As further selection criteria, only the proteins, identified by MASCOT search with at least 2 peptides, providing a full list of putative hDAAO interactors. This strategy let us to obtain more confident identification of truly hDAAO binding candidates minimizing the number of false positives.

A total of 48 proteins are listed in Table III.4 together with their corresponding genes and the number of peptides used for their identification. Protein identification procedures confirmed that the band at about 55 kDa corresponds to the bait hDAAO.

As a preliminary screening of the results obtained, we extensively searched the literature and protein interaction databases for known interactors. Two pricipal distinct functional clusters were represented: one (cell morphology, cellular assembly and organization) which covers the 67% of the total proteins and the other composed by amino acid metabolism, post-traslational modification, small molecule biochemistry, which is the 29% of the total proteins (Fig.III.13). The most important

finding is that the statistical analysis revealed that 27 proteins are involved in neurological disorder, as assessed by experiments of genome wide expressions, SNPS, differential proteomics on post-mortem tissues from brains of patient suffering for schizophrenia or other neurological disorders.

Gene Name	ID	Protein Name	Location	Number of Peptides	MW (Da)
NNT	Q13423	NAD(P) transhydrogenase, mitochondrial	Mitochondrion inner membrane	4	113896
FLNC	Q14315	Filamin C	Cytoplasm cytoskeleton	16	291022
CAD	P27708	CAD protein	Cytoplasm	3	242984
MYH9	P35579	Myosin 9	Cytoplasm	9	226532
CLTC	Q00610	Clathrin heavy chain 1	Plasma Membrane	3	191615
IMMT	Q16891	Mitofilin - Mitochondrial inner membrane protein	Mitochondrion inner membrane	7	83678
CLPTM1	O96005	Cleft lip and palate transmembrane protein 1	Plasma Membrane	9	76097
LMNA	P02545	Prelamin-A/C	Nucleus	9	74139
POR	P16435	NADPH-cytochrome P450 reductase	Endoplasmic reticulum membrane; Peripheral membrane protein	17	76690
HSPA5	P11021	78 kDa glucose-regulated protein	Endoplasmic reticulum	12	72333
EWSR1	Q01844	RNA-binding protein EWS	Nucleus	2	68478
NDUFS1	P28331	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	Mitochondrion inner membrane	7	79468
ALB	P02768	Serum albumin	Extracellular Space	5	69367
ATAD3A	Q9NVI7	ATPase family, AAA domain containing 3A	Nucleus	6	71369
PIGT	Q969N2	GPI transamidase component PIG-T (Phosphatidylinositol-glycan biosynthesis class T protein)	Endoplasmic reticulum membrane	4	65700
VPS45	Q9NRW7	Vacuolar protein sorting-associated protein 45	Golgi apparatus membrane; Peripheral membrane protein	13	65388
GALNT2	Q10471	Polypeptide N-acetylgalactosaminyltransferase 2	Golgi apparatus	12	64733
TOR1AIP1	Q5JTV8	Torsin-1A-interacting protein 1	Nucleus	3	66248
FAF2	Q96CS3	Fas associated factor 2	Cytoplasm	3	52623
B4GALT1	P15291	Beta-1,4-galactosyltransferase 1	Golgi apparatus	2	43920
FKBP8	Q14318	Peptidyl-prolyl cis-trans isomerase FKBP8	Mitochondrion membrane	3	44562
SAMM50	Q9Y512	Sorting and assembly machinery component 50 homolog	Mitochondrion outer membrane	6	52270
SPTLC1	O15269	Serine palmitoyltransferase 1	Endoplasmic reticulum membrane	3	52744
UFSP2	Q9NUQ7	UFM1-specific protease 2		3	53261
HNRNPF	P52597	Heterogeneous nuclear ribonucleoprotein F	Nucleus	2	45672
DAO	P14920	D-amino-acid oxidase	Peroxisome	7	39474
PREB	Q9HCU5	Prolactin regulatory element-binding protein	Nucleus	4	45468
NDUFS2	O75306	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	Mitochondrion inner membrane	4	52546
HNRNPC	P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	Nucleus	7	33670
ANXA2	B4DNH8	Annexin A2	Plasma Membrane	13	21670
SCD	D3DR68	Stearoyl-CoA desaturase (Delta-9-desaturase), isoform CRA_c	Membrane	3	41505
RDH14	Q9HBH5	Retinol dehydrogenase 14	Cytoplasm	3	36865
GNB1	P62874	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit	Plasma Membrane	3	37377

		beta-1			
NAPA	P54920	Alpha-soluble NSF attachment protein (SNAP-alpha)	Membrane	13	33233
MLEC	Q14165	Malectin	Endoplasmic reticulum membrane	2	32234
GAPDH	P04406	Glyceraldehyde-3-phosphate dehydrogenase	Cytoplasm	2	36053
CYC1	P08574	Cytochrome c1, heme protein, mitochondrial	Mitochondrion inner membrane	3	35422
RALA	P11233	Ras-related protein Ral-A	Cytoplasm	3	23567
FAM3C	Q92520	Protein FAM3C	Cytoplasmic vesicle	2	24680
RAB5C	P51148	Ras-related protein Rab-5C	Cytoplasm	2	23483
RAB11B	Q15907	Ras-related protein Rab-11B	Cytoplasm	3	24489
RAB18	Q9NP72	Ras-related protein Rab-18	Cytoplasm	3	22977
RAC1	P63000	Ras-related C3 botulinum toxin substrate 1	Cytoplasm	4	21450
SPCS2	Q15005	Signal peptidase complex subunit 2	Cytoplasm	2	25003
CYB5B	O43169	Cytochrome b5 type B	Mitochondrion outer membrane	2	16332
LCN1	P31025	Lipocalin	Extracellular Space	2	19250
LYZ	P61626	Lysozyme	Extracellular Space	2	16537
CISD2	Q8N5K1	CDGSH iron-sulfur domain-containing protein 2	Endolasmic reticulum or mitochondrion outer membrane	2	15278

Table III.4. Full list of hDAAO putative interactors identified by mass spectrometry. The proteins in green are involved in neurological disorder.

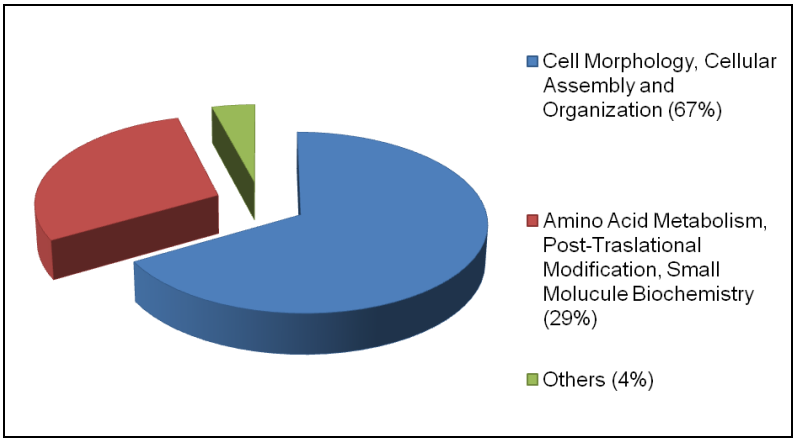


Fig.III.13. Functional clusters of hDAAO interacting partners. Schematic representation of the interactors identified in the proteomic screening. The interactors identified by proteomics are grouped in different functional categories on the basis of their reported biological activities.

III.7. Discussion

III.7.1. Structural characterization

Since the causes of schizophrenia are still unknown, current treatments focus on eliminating the symptoms of the disease, rather than curing the disease. Antipsychotic medications have been available since the mid-1950s, and they effectively alleviate the positive symptoms of schizophrenia. The older antipsychotic medications include chlorpromazine, haloperidol, perphenazine, and fluphenzine, and can cause extrapyramidal side effects, such as rigidity, persistent muscle spasms, tremors, and restlessness. In the 1990s, new drugs, called atypical

antipsychotics (such as clozapine), that rarely produced these side effects, were developed. Clozapine treats psychotic symptoms effectively but it can produce a serious problem called agranulocytosis, a loss of the white blood cells that fight infection. Medical evidence supporting the role of NMDA receptor hypofunction in schizophrenia has prompted clinical trials of agents that enhance NMDA receptor functionality [46,47]. For example, schizophrenia patients receiving D-serine (an allosteric activator of NMDA-type glutamate receptor) together with antipsychotics showed significant improvement in their positive, negative, and cognitive symptoms [48]. In human brain hDAAO is the main perpetrator of oxidation of the gliotransmitter D-serine. Modulating DAAO function is pLG72 which was initially identified by Chumakov et al. [40], who by a yeast two-hybrid screening experiment identified D-amino acid oxidase (DAAO) as the interacting partner of pLG72. Structural characterization of both proteins isolated and in complex is a challenging task for the understanding of the molecular basis of schizophrenia.

Integrated strategies such as limited proteolysis and chemical cross-linking coupled to Mass Spectrometry (MS) were used for the *in vitro* analysis of the surface topology of pLG72-hDAAO complex, as well as for the definition of the conformational changes of hDAAO associated to complex formation.

The optimal conditions for the *in vitro* complex formation were obtained incubating in ice for 10 min a 5 molar excess of pLG72 respect hDAAO.

As first approach proteolysis experiments on the two separated protein have been carried out, in order to obtain low resolution structural characterization of the two proteins and to define the portions of the proteins involved in conformational changes after the complex formation.

Several considerations can be drawn: i) hDAAO is more accessible to proteases when in complex with pLG72 than as single protein, as assed by the lower enzyme/substrate ratio used than the single hDAAO protein; ii) a significative conformational change is induced in the hDAAO molecule by the formation of the complex with pLG72, as can be deduced by the completely different patterns of preferential proteolytic sites observed for hDAAO in the presence and in absence of pLG72. Specifically, the peptide bonds following L189, R191, L194 E249 and L250 that were proteolitically hydrolyzed in the isolated hDAAO, were not recognized in the presence of pLG72: on the contrary, the Y34, Y55, F167, N180, L296 and R297 sites were hydrolyzed only in the hDAAO-pLG72 complex (Fig.III.14).

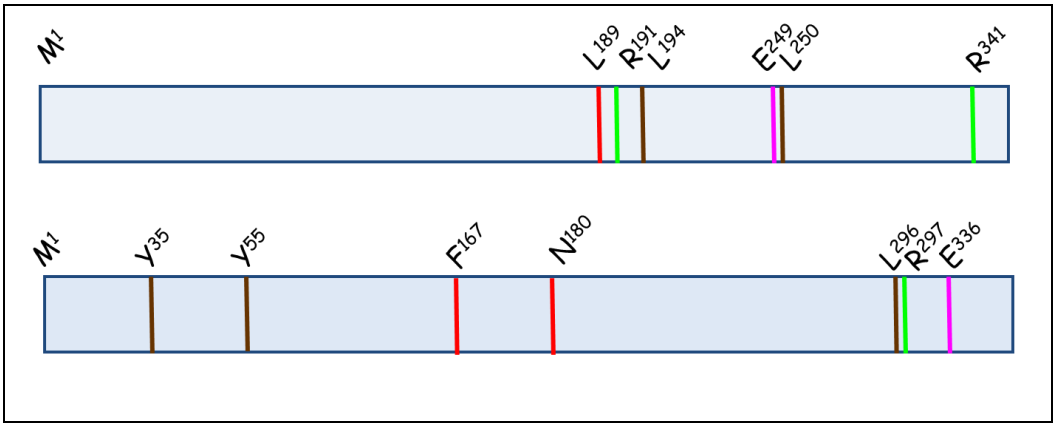


Fig.III.14. Schematic representation of complementary proteolysis results of isolate hDAAO (upper panel) and in complex with pLG72 (lower panel). The results of the trypsin experiments are in green, the ones of the chimotripsin in brown, the ones of the Glu-C in pink and subtilisin in red.

Because of the lower protease/substrate ratio adopted with respect to the experiments carried out on the isolated hDAAO, and the molar excess of pLG72 present in the reaction mixture, in the experiments carried out on the hDAAO-pLG72 complex we confidently refer the proteolytic sites observed to hDAAO molecules actually complexed with pLG72.
The increased accessibility of hDAAO when in complex with pLG72 is rather peculiar as, generally, proteins in complexes show reduced flexibility and accessibility to proteases.
Overall, in the complex, the C-terminal portion of hDAAO is folded in a compact structure not accessible to proteolytic enzymes, differently from the isolated hDAAO. Cleavage most consistently occurred at sites located in different portions of the protein.
As far as pLG72 component is concerned, we remind that a crystallographic structure of pLG72 is not available because of its highly instability (it is soluble only in presence of mild denaturant) and no homologous protein has been structurally characterized so far. However our results are in agreement with circular dicroism data [35], as the proteolytic sites are localised on region predicted to be at the extremes of secondary structure or in loops (data not shown). In Fig.III.13 are indicated the preferential hydrolysis sites of pLg72 carried out on the isolated protein.

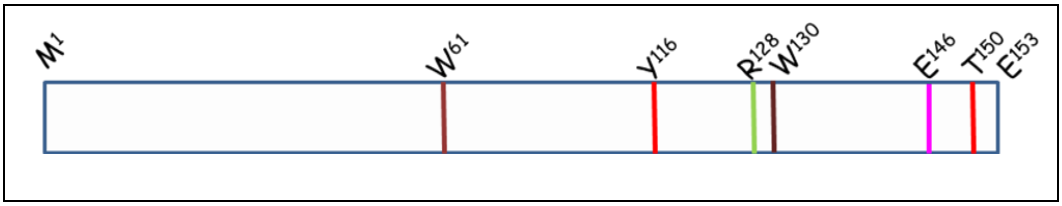


Fig.III.13. Schematic representation of complementary proteolysis results of isolate pLG72. The results of the trypsin experiments are in green, the ones of the chimotripsin in brown, the ones of the Glu-C in pink and subtilisin in red.

Unfortunately, we cannot draw any consideration on conformational changes occurring on pLG72 when in complex with hDAAO, since the experiments were carried out in presence of a molar excess of pLG72 that was therefore present in solution also in the non complexed form, although, in the presence of hDAAO, the general impression was a reduced accessibility of pLG72 to proteases.
To give more structural details, when the hydrolysis sites were mapped on the 3D-structure of a monomer of hDAAO, (Fig.III.14, in green the sites identified on isolated hDAAO, in red the sites identified on the hDAAO molecule when in complex with pLG72) it is possible to notice that i) the preferential hydrolysis residues are all localised in proximity of the binding site of the flavinic cofactor and in a defined portion of the three-dimensional structure, ii) it can be assumed that the formation of the complex with pLG72 determines the protection of the specific region of the molecule while exposing sites that are however in the immediate surroundings of that region. The exact definition of the contact regions between the two molecules by chemical cross-linking is still in progress.

Fig.III.14. Three-dimensional model of the hDAAO monomer. The flavinic cofactor is reported in yellow, the preferential hydrolysis sites of the isolated protein are in green and the ones of the protein involved in the complex in red.

III.7.2 Identification of novel protein partners of human DAAO

hDAAO is likely to be finely regulated in its activation/inhibition by protein partners. A number of recent studies focused on the development of new and effective hDAAO inhibitors as a treatment strategy for schizophrenia by modulating D-serine concentrations in the brain [44,45,49,50].
However, the mechanistic aspects of how hDAAO accomplishes its function are still enigmatic and controversial. To meet this aim, we have performed a comprehensive analysis of the interactome of hDAAO by a functional proteomic approach. Human U87 glioblastoma cells has been transfected with an expression vector encoding for Myc-hDAAO fused protein. To identify protein specifically interacting with hDAAO, we performed immunoprecipitation and mass-spectrometry analysis. We identified a total of 48 proteins, in particular, 27 proteins of which had been previously reported to be involved in neurological disorder (Table III.4).
Among them, we focus out attention on the members of the Rab family of small GTPases are key mediators of intracellular membrane sorting in eucariotic cell and they have been implicated in the pathology of multiple human deseases. Rabs, in conjunction with their effectors and regulators, participate in central nervous system development (Rab23), polarized neurite growth (Rab8 and Rab11), endocytosis and axonal retrograde transport (Rab5 and Rab7). Rab-mediated membrane traffic processes in neurons and glia may be sensitive to certain pathological factors [51].
Through a GDP-GTP exchange cycle, Rabs function as molecular switches to mediate vesicular transport along the cytoskeleton by engaging specific motor proteins, ensuring efficient tethering and docking of vesicles to target membranes, and timing vesicle fusion with the engagement of soluble N-ethylmaleimide sensitive factor (NSF) attachment receptor (SNARE) proteins.
We have found two member of the family, Rab 5 and Rab 11, which are correlated with the synaptic plasticity. This process, which is widely thought to be the cellular correlate for cognitive functions, such as learning and memory, consist of a continuously remodelling of the synaptic connections between neurons in some region of the brain in response to neuronal activity. Not surprisingly, the dysfunction of synapses has emerged as a major etiologic factor in many brain disease including widespread neurodegenerative, neurodevelopmental and psychiatric conditions

(schizophrenia depression). In particular, Rab 5, which is localized both to synaptic vesicles and endosomal intermediates involved in synaptic vesicle recycling, has been implicated in protein transport from plasma membrane to early endosomes during clathrin-dependent endocytosis in a variety of cellular systems.
Rab5 is activated downstream by Ca^{2+} binding after NMDAr opening. Its activation drives the specific internalization of synaptic AMPA receptors. Neuronal activity-dependent insertion or removal of AMPA receptors from the postsynaptic plasma membrane underlies the phenomenon of long-term depression (LTD), which is experimental manifestation of synaptic plasticity.
Although the details remained elusive, the emerging rough picture is that synaptic activity activates Rab5, which in turn drives the removal of AMPA receptors from synapses into endosomes.
Moreover, Rab5 has some effects on the activity of Rac1, another hDAAO interacting partner which we have found, and which is essential for actin remodelling. Actin is the principal component of the synaptic dendritic spines and directly determines the morphology of the spines and allows them to be highly dynamic in shape and size. Synaptic spines are dynamic structures that regulate neuronal responsiveness and plasticity. The cytoskeleton of dendritic spines is particularly important in their synaptic plasticity; without a dynamic cytoskeleton, spines would be unable to rapidly change their volumes or shapes in responses to stimuli. Disturbances in neuronal connectivity, particularly in glutamatergic synaptic connections, underlie schizophrenia and associated disorders. Neuropathological studies with autopsied brains from individuals with schizophrenia have reported reduced numbers of dendritic spines. Morphological changes of the dendritic spine are directly correlated with its functional deficits [52]. Rac1 is one of main upstream regulators of JNK. c-Jun N-terminal kinase (JNK) regulates the morphological changes of migrating neurons through the promotion of microtubule dynamics and actin reorganization. It serves as important mediators of spine enlargement and constitutive Rac1 activation decreases spine size [53]. This mechanism likely underlies disturbances in glutamatergic neurotransmission that have been frequently reported in schizophrenia that can lead to alteration of dendritic spines with consequential major pathological changes in brain function.
Rab11 appears to have a specific role in neuronal and neuroendocrine cells as a GTP-dependent switch that could shunt certain cargoes between the regulated and constitutive secretory pathways. Rab11 endosomes are key trafficking intermediates that control vesicle exocytosis and membrane growth during the process of cellularization. It is therefore presumably important for developmental processes in particular for synaptic morphogenesis.
Among the other putative interactors that we have identified, there is mitofilin, which has a role in neurological disorder, also called mitochondrial inner membrane protein, (IMMT). First identified and characterized in the 1990s [54,55], mitofilin is essential to normal mitochondrial function and morphology.
Mitochondria play critical roles in fundamental cellular processes, encompassing energy production, programmed cell death, Ca^{2+} homeostasis, and monoamine metabolism. Intriguingly, mitochondrial dysfunction has also been found to be associated with various psychiatric disorders, including schizophrenia.
For example, a reduction in the number of mitochondria, an abnormal morphology of mitochondria, and likely, as a consequence, defects in the process of oxidative phosphorylation have been reported in postmortem brain studies of schizophrenia patients [56,57]. More recently, gene expression studies using the postmortem brains

of schizophrenia patients also suggest a role of mitochondrial dysfunction in the development of schizophrenia [58,59]. Intriguingly, it has recently been shown that Mitofilin is associated with the general protein complex involved in the import of mitochondrial proteins such as metaxin1, metaxin2 and sorting and assembly machinery component 50 homolog (SAM50) [60], a protein which we have identified as an interactor of DAAO.
A last example of interesting putative interactors, there is CAD, a trifunctional protein which is associated with the enzymatic activity of the first three enzymes in the pyrimidine biosynthesis: carbamoyl phosphate synthetase, aspartate transcarbamoylase, and dihydroorotase. Uracil, one of the three pyrimidines, is converted to uridine, which has been reported to have an antidepressant effect. A large-scale clinical trial of uridine (RG2417) for bipolar depression is underway. It has been reported that biosynthesis of pyrimidine is inhibited in the depressive state. Reduced uridine biosynthesis is involved in the biochemical basis of depressive state in relevance to the clinical effect of uridine [61].
Some proteins we have identified are enzymes involved the regulation of energy metabolism. Their altered levels can lead to an overall disturbance, and ultimately contribute to the establishment of pathological states.
Enzymes with functions in glycolysis such as GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) were found to be very likely to affect the glucose metabolism in schizophrenic patients. Alterations in the glucose metabolism have been extensively reported as a central component of schizophreina and are unlikely due to antipsychotic effects. A highly altered energy metabolism profile within the schizophrenia brain was observed, particularly implying an increased demand for glucose, resulting in increased glycogen catabolism.
Abnormal serum glucose profiles in first-onset schizophrenia patients have been reported and the prevalence of diabetes type II is significantly increased in schizophrenia patients where there is substantial evidence for an increased glucose demand and/or cellular hypoxia within the schizophrenia prefrontal cortex [62].
The panel of putative interactors is extremely promising and demands for further studies that might elucidate the functional links among the different component, and most importantly, to understand the specific role of hDAAO in such mechanism. These are preliminary results and the next step will be to define how these proteins interact each other in the cell, in order to define a pathway who can elucidate the molecular basis of schizophrenia.

III.5. References

[1] Ross CA, Margolis RL, Reading SA. Pletnikov M, Coyle JT. Neurobiology of schizophrenia. *Neuron* (2006);52:139-53.
[2] Lewis DA, Gonzalez-Burgos G. Pathophysiologically based treatment interventions in schizophrenia. *Nat. Med.* (2006);12:1016-22.
[3] Davis KL, Kahn RS, Ko G, Davidson M. Dopamine in schizophrenia: a review and reconceptualization. *Am. J. Psychiatry* (1991);148:1474-86.
[4] Javitt DC, Zukin SR. Recent advances in the phencyclidine model of schizophrenia. *Am J Psychiatry*. (1991);148:1301-08.
[5] Olney JW, Farber NB. Glutamate receptor dysfunction and schizophrenia. *Arch Gen Psychiatry.* (1995);52:998-1007.

[6] Coyle JT. The glutamatergic dysfunction hypothesis for schizophrenia. *Harv Rev Psychiatry*. (1996);3:241-53.
[7] Tamminga CA. Schizophrenia and glutamatergic transmission. *Crit Rev Neurobiol.* (1998);12:21-36.
[8] Krystal JH, D'Souza DC, Petrakis IL, Belger A, Berman R, Charney DS, Abi-Saab W, Madonick S. NMDA agonists and antagonists as probes of glutamatergic dysfunction and pharmacotherapies for neuropsychiatric disorders. *Harv Rev Psychiatry*. (1999);7:125-33.
[9] Hashimoto K, Iyo M. Glutamate hypothesis of schizophrenia and targets for new antipsychotic drugs [in Japanese]. *Nihon Shinkei Seishin Yakurigaku Zasshi*. (2002);22:3-13.
[10] Krystal JH, Karper LP, Seibyl JP, Freeman GK, Delaney R, Bremner JD, Heninger GR, Bowers MBJ, Charney DS. Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Arch. Gen. Psychiatry* (1994);51:199-214.
[11] Coyle JT. Glutamate and schizophrenia: beyond the dopamine hypothesis. *Cell. Mol. Neurobiol.* (2006);26:365-84.
[12] Millan MJ. N-Methyl-D-aspartate receptors as a target for improved antipsychotic agents: novel insights and clinical perspectives. *Psychopharmacology* (2005);179:30-53.
[13] Cull-Candy S, Brickely S, Farrant M. NMDA receptor subunits: diversity, development and disease. *Curr. Opin. Neurobiol.* (2001);11:327-35.
[14] Monyer H, Burnashev N, Laurie DJ, Sakmann B, Seeburg PH. Developmental and regional expression in the rat brain and functional properties of four NMDA receptors. *Neuron* (1994);12:529-40.
[15] Dumas TC. Developmental regulation of cognitive abilities: modified composition of a molecular switch turns on associative learning. *Prog. Neurobiol.* (2005);76:189-211.
[16] Dingledine R, Borges K, Bowie D, Traynelis SF. The glutamate receptor ion channels. *Pharmacol. Rev.* (1999);51:7-61.
[17] Johnson JW, Ascher P. Glycine potentiates the NMDA response in cultured mouse brain neurons. *Nature* (1987);325:529-31.
[18] Clements JD, Westbrook GL. Activation kinetics reveal the number of glutamate and glycine binding sites on the N-methyl-D-aspartate receptor. *Neuron* (1991);7:605-13.
[19] Bading H, Ginty DD, Greenberg ME. Regulation of gene expression in hippocampal neurons by distinct calcium signaling pathways. *Science* (1993);260:181-6.
[20] Matsui T, Sekiguchi M, Hashimoto A, Tomita U, Nishikawa T, Wada KJ. Functional comparison of D-serine and glycine in rodents: the effects on cloned NMDA receptors and the extracellular concentration. *J. Neurochem.* (1995);65:454-58.
[21] Fadda E, Danysz W, Wroblewski JT, Costa E. Glycine and D-serine increase the affinity of N-methyl-D-aspartate sensitive glutamate binding sites in rat brain synaptic membranes. *Neuropharmacology* (1988);27:1183-5.
[22] Vyklický Jr L, Benveniste M, Mayer ML. Modulation of N-methyl-D aspartic acid receptor desensitization by glycine in mouse cultured hippocampal neurones. *J. Physiol.* (1990);428:313-31.
[23] Nong Y, Huang YQ, Ju W, Kalia LV, Ahmadian G, Wang YT, Salter MW. Glycine binding primes NMDA receptor internalization. *Nature* (2003);422:302-7.
[24] Mothet JP, Parent AT, Wolosker H, Brady Jr RO, Linden DJ, Ferris CD, Rogawski MA, Snyder SH. D-serine is an endogenous ligand for the glycine site of the N-methyl-D-aspartate receptor. *Proc. Natl. ACID. Sci. U.S.A.* (2000);97:4926-31.
[25] Yang Y, Ge W, Chen Y, Zhang Z, Shen W, Wu C, Poo M, Duan S. Contribution of astrocytes to hippocampal long-term potentiation through release of D-serine. *Proc. Natl. ACID. Sci. U.S.A.* (2003);100:15194-99.
[26] Hashimoto K, Fukushima T, Shimizu E, Komatsu N, Watanabe H, Shinoda N, Nakazato M, Kumakiri C, Okada S, Hasegawa H, Imai K, Iyo M. Decreased serum levels of D-serine in patients with schizophrenia: evidence in support of the N-methyl-

D-aspartate receptor hypofunction hypothesis of schizophrenia. *Arch Gen Psychiatry.* (2003);60(6):572-6.
[27] Tsai G, Yang P, Chung LC, Lange N, Coyle JT. d-Serine added to antipsychotics for the treatment of schizophrenia. *Biol Psychiatry*. (1998);44:1081-89.
[28] Goff DC, Coyle JT. The emerging role of glutamate in the pathophysiology and treatment of schizophrenia. *Am J Psychiatry*. (2001);158:1367-77.
[29] Wolosker H, Sheth KN, Takahashi M, Mothet J, Brady ROJ, Ferris CD, Snyder SH. Purification of serine racemase: Biosynthesis of the neuromodulator D-serine. *Proc. Natl. ACID. Sci. U.S.A.* (1999);96:721-5.
[30] Pollegioni L, Sacchi S. Metabolism of the neuromodulator D-serine. *Cell Mol Life Sci.* (2010);67(14):2387-404. Review.
[31] Shleper M, Kartvelishvily E, Wolosker, H. D-serine is the dominant endogenous coagonist for NMDA receptor neurotoxicity in organotypic hippocampal slices. *J. Neurosci.* (2005);25:9413-17.
[32] Xia M, Liu Y, Figueroa DJ, Chiu CS, Wei N, Lawlor AM, Lu P, Sur C, Koblan KS, Connolly TM. Characterization and localization of a human serine racemase. Brain Res. Mol. Brain Res. (2004);125:96-104.
[33] Wang LZ, Zhu XZ. Spatiotemporal relationships among D-serine, serine racemase, and D-amino acid oxidase during mouse postnatal development. *Acta Pharmacol. Sin.* (2003);24:965-74.
[34] Pollegioni L, Piubelli L, Sacchi S, Pilone MS, Molla G. Physiological functions of D-amino acid oxidases: from yeast to humans. *Cell Mol Life Sci* (2007);64:1373-94.
[35] Molla G, Sacchi S, Bernasconi M, Pilone MS, Fukui K, Pollegioni L. Characterization of human D-amino acid oxidase. *FEBS Lett* (2006);580:2358-64.
[36] Campaner S, Pollegioni L, Ross BD, Pilone MS. Limited proteolysis and site-directed mutagenesis reveal the origin of microheterogeneity in Rhodotorula gracilis D-amino acid oxidase. *Biochem J.* (1998);330(2):615-21.
[37] Moreno S, Nardacci R, Cimini A, Ceru` MP. Immunocytochemical localization of D-amino acid oxidase in rat brain. *J. Neurocytol*. (1999);28:169-85.
[38] Verrall L, Walker M, Rawlings N, Benzel I, Kew JN, Harrison PJ, Burnet PW. D-Amino acid oxidase and serine racemase in human brain: normal distribution and altered expression in schizophrenia. *Eur. J. Neurosci.* (2007);26:1657-69.
[39] Katagiri M, Tojo H, Horiike K, Yamano T. Immunochemical relationship of D-amino acid oxidases in various tissues and animals. *Comp. Biochem. Physiol. B* (1991);99:345-50.
[40] Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, Barry C, Tanaka H, La Rosa P, Puech A, Tahri N, Cohen-Akenine A, Delabrosse S, Lissarrague S, Picard FP, Maurice K, Essioux L, Millasseau P, Grel P, Debailleul V, Simon AM, Caterina D, Dufaure I, Malekzadeh K, Belova M, Luan JJ, Bouillot M, Sambucy JL, Primas G, Saumier M, Boubkiri N, Martin-Saumier S, Nasroune M, Peixoto H, Delaye A, Pinchot V, Bastucci M, Guillou S, Chevillon M, Sainz-Fuertes R, Meguenni S, Aurich-Costa J, Cherif D, Gimalac A, Van Duijn C, Gauvreau D, Ouellette G, Fortier I, Raelson J, Sherbatich T, Riazanskaia N, Rogaev E, Raeymaekers P, Aerssens J, Konings F, Luyten W, Macciardi F, Sham PC, Straub RE, Weinberger DR, Cohen N, Cohen D. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl ACID Sci U S A.* (2002);99(21):13675-80.
[41] Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, Barry C, Tanaka H, La Rosa P, Puech A, Tahri N, Cohen-Akenine A, Delabrosse S, Lissarrague S, Picard FP, Maurice K, Essioux L, Millasseau P, Grel P, Debailleul V, Simon AM, Caterina D, Dufaure I, Malekzadeh K, Belova M, Luan JJ, Bouillot M, Sambucy JL, Primas G, Saumier M, Boubkiri N, Martin-Saumier S, Nasroune M, Peixoto H, Delaye A, Pinchot V, Bastucci M, Guillou S, Chevillon M, Sainz-Fuertes R, Meguenni S, Aurich-Costa J, Cherif D, Gimalac A, Van Duijn C, Gauvreau D, Ouellette G, Fortier I, Raelson J, Sherbatich T, Riazanskaia N, Rogaev E, Raeymaekers P, Aerssens J, Konings F, Luyten W, Macciardi F, Sham PC, Straub RE,

Weinberger DR, Cohen N, Cohen D. *Proc. Natl. ACID. Sci. U. S. A.* (2002);99:13675-80.

[42] Sacchi S, Bernasconi M, Martineau M, Mothet JP, Ruzzene M, Pilone MS, Pollegioni L, Molla G. pLG72 modulates intracellular D-serine levels through its interaction with D-amino acid oxidase: effect on schizophrenia susceptibility. *J Biol Chem* (2008);283:22244-56.

[43] Molla G, Bernasconi M, Sacchi S, Pilone MS, Pollegioni L. Expression in Escherichia coli and in vitro refolding of the human protein pLG72. *Protein Expr Purif* (2006);46:150-5.

[44] Adage T, Trillat AC, Quattropani A, Perrin D, Cavarec L, Shaw J, Guerassimenko O, Giachetti C, Gre´co B, Chumakov I, Halazy S, Roach A, Zaratin P. *In vitro* and *in vivo* pharmacological profile of AS057278, a selective D-amino acid oxidase inhibitor with potential anti-psychotic properties. *Eur Neuropsychopharmacol* (2008);18:200-14.

[45] Sparey T, Abeywickrema P, Almond S, Brandon N, Byrne N, Campbell A, Hutson PH, Jacobson M, Jones B, Munshi S, Pascarella D, Pike A, Prasad GS, Sachs N, Sakatis M, Sardana V, Venkatraman S, Young MB. The discovery of fused pyrrole carboxylic acids as novel, potent D-amino acid oxidase (DAO) inhibitors. *Bioorg Med Chem Lett* (2008) 18:3386-91.

[46] Harrison PJ, Weinberger DR. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry* (2005);10:40-68.

[47] Lang UE, Puls I, Muller DJ, Strutz-Seebohm N, Gallinat J. Molecular mechanisms of schizophrenia. *Cell Physiol Biochem* (2007);20:687-702.

[48] Tsai G, Van Kammen DP, Chen S, Kelley ME, Grier A, Coyle JT. Glutamatergic neurotransmission involves structural and clinical deficits of schizophrenia. *Biol Psychiatry* (1998);44:667-74.

[49] Ferraris D, Duvall B, Ko YS, Thomas AG, Rojas C, Majer P, Hashimoto K, Tsukamoto T. Synthesis and biological evaluation of D-amino acid oxidase inhibitors. *J Med Chem* (2008);51:3357-59.

[50] Duplantier AJ, Becker SL, Bohanon MJ, Borzilleri KA, Chrunyk BA, Downs JT, Hu LY, El-Kattan A, James LC, Liu S, Lu J, Maklad N, Mansour MN, Mente S, Piotrowski MA, Sakya SM, Sheehan S, Steyn SJ, Strick CA, Williams VA, Zhang L. Discovery, SAR, and pharmacokinetics of a novel 3-hydroxyquinolin- 2(1H)-one series of potent D-amino acid oxidase (DAAO) inhibitors. *J Med Chem* (2009);52:3576-85.

[51] Ng EL, Tang BL. Rab GTPases and their roles in brain neurons and glia. *Brain Res Rev.* (2008);58(1):236-46. Review.

[52] Hayashi-Takagi A, Takaki M, Graziane N, Seshadri S, Murdoch H, Dunlop AJ, Makino Y, Seshadri AJ, Ishizuka K, Srivastava DP, Xie Z, Baraban JM, Houslay MD, Tomoda T, Brandon NJ, Kamiya A, Yan Z, Penzes P, Sawa A. Disrupted-in-Schizophrenia 1 (DISC1) regulates spines of the glutamate synapse via Rac1. *Nat Neurosci*. (2010);13(3):327-32.

[53] Kawauchi T, Sekine K, Shikanai M, Chihama K, Tomita K, Kubo K, Nakajima K, Nabeshima Y, Hoshino M. Rab GTPases-dependent endocytic pathways regulate neuronal migration and maturation through N-CIDherin trafficking. *Neuron.* (2010);67(4):588-602.

[54] Odgren PR, Toukatly G, Bangs PL, Gilmore R, Fey EG. Molecular characterization of mitofilin (HMP), a mitochondria-associated protein with predicted coiled coil and intermembrane space targeting domains. *J Cell Sci* (1996); 109:2253-64.

[55] Gieffers C, Korioth F, Heimann P, Ungermann C, Frey J. Mitofilin is a transmembrane protein of the inner mitochondrial membrane expressed as two isoforms. *Exp Cell Res* (1997);232(2):395-9.

[56] Maurer I, Zierz S, Möller H. Evidence for a mitochondrial oxidative phosphorylation defect in brains from patients with schizophrenia. *Schizophr Res* (2001);48:125-36.

[57] Kung L, Roberts RC. Mitochondrial pathology in human schizophrenic striatum: A postmortem ultrastructural study. *Synapse* (1999);31:67-75.

[58] Prabakaran S, Swatton JE, Ryan MM, Huffaker SJ, Huang JT, Griffin JL, Wayland M, Freeman T, Dudbridge F, Lilley KS, Karp NA, Hester S, Tkachev D, Mimmack ML, Yolken RH, Webster MJ, Torrey EF, Bahn S. Mitochondrial dysfunction in schizophrenia: Evidence for compromised brain metabolism and oxidative stress. *Mol Psychiatry* (2004);9:684-697.

[59] Park YU, Jeong J, Lee H, Mun JY, Kim JH, Lee JS, Nguyen MD, Han SS, Suh PG, Park SK. Disrupted-in-schizophrenia 1 (DISC1) plays essential roles in mitochondria in collaboration with Mitofilin. *Proc Natl ACID Sci U S A.* (2010);107(41):17785-90.

[60] Xie J, Marusich MF, Souda P, Whitelegge J, Capaldi RA. The mitochondrial inner membrane protein mitofilin exists as a complex with SAM50, metaxins 1 and 2, coiledcoil- helix coiled-coil-helix domain-containing protein 3 and 6 and DnaJC11. *FEBS Lett* (2007);581:3545-9.

[61] Tochigi M, Iwamoto K, Bundo M, Sasaki T, Kato N, Kato T. Gene expression profiling of major depression and suicide in the prefrontal cortex of postmortem brains. *Neurosci Res.* (2008);60(2):184-91.

[62] Prabakaran S, Swatton JE, Ryan MM, Huffaker SJ, Huang JT, Griffin JL, Wayland M, Freeman T, Dudbridge F, Lilley KS, Karp NA, Hester S, Tkachev D, Mimmack ML, Yolken RH, Webster MJ, Torrey EF, Bahn S. Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Mol Psychiatry.* (2004);9(7):684-97.

IV. UV Laser Cross-linking. Characterization and new strategy for industrial applications

IV.1. Shining a light on UV Laser Cross-linking

The detection and analysis of protein-protein interactions is one of the central tasks of proteomics in the post-genomic era.

Protein-protein interactions are among the essentials that make up life. Most biological processes are controlled by dynamic molecular network of enormous complexity rather than by individual proteins. Mapping of protein interactions on a proteome-wide scale is therefore essential for fully understanding what's going on in living cells. Characterization of protein interactions is indeed an extremely challenging task since many protein interactions are not stable, most interactions are transient, and multiprotein complexes possess no common factors or physical properties that can be used as an analytical handle.

Chemical cross-linking strategies combined with mass spectrometric (MS) analysis have been pursued for many years with the goal to fulfil two primary needs in proteomics research: (i) identification of protein interaction network and (ii) low-resolution mapping of protein and protein complex structures [2,3]. Both are important and both represent unique opportunities to provide critical information on biological systems. Cross-linking has been widely applied in protein-protein interaction studies due to the inherent potential these approaches hold for stabilizing and freezing labile interactions with covalent bonds. Chemical cross-linking also has significant advantages for structural characterization of individual purified proteins and protein complexes where the identified cross-linked residues/peptides provide distance constraints that help define 3-D structural models. However, identification of cross-linked peptides is not trivial even for purified protein complexes which may be available in large quantity. In this regard, the challenges associated with cross-linking approaches arise primarily from three aspects: (i) enormous complexity inherent in the cross-linked samples which contain predominantly unmodified peptides, inter-cross-linked peptides, intra-cross-linked peptides, dead-end labelled peptides, multiple-labelled peptides, and non-specific labelled peptides; (ii) low abundance of the cross-linked species; and (iii) data complexity of MS/MS spectra. Significant effort has been dedicated to designing novel cross-linkers that allow improved capability for cross-linked peptide identification via a signature pattern in the data or reduction of sample complexity by enrichment. Alternative efforts have also been applied to informatics software development with the use of the conventional cross-linkers [4].

As an alternative approach to classical chemical cross-linking is the UV laser cross-linking. It is already well documented that UV irradiation with a high-intensity femtosecond pulsed laser results in an efficient protein-nucleic acid cross-linking [5-7].

The mechanism leading to UV laser cross-linking induces the excitement of the electronic state of the biomolecules which culminates with the creation of a stable bond between them. Several investigations have concluded that the drawbacks of UV irradiation could be overcome by using high-intensity UV lasers [8-11]. These studies found that the total yield of cross-linked material is one to two orders of magnitude greater with pulsed UV lasers (pulse durations ~20 ps to 10 ns) than with usual continuous wave (CW) sources. These data indicated that the cross-linking reaction yield is enhanced by a two-step process [12] in a simplified energy level

diagram (Fig.IV.1) [6,7]. A first UV photon excites the DNA base from the singlet S0 ground state to the first excited S1 manifold. From the excited S1 level the molecule can absorb a second photon which promotes the DNA base to high lying singlet SN levels. There is, however, a finite probability for an intersystem crossing to the lowest triplet state T1, which has a longer lifetime compared to the S1 state. Absorption of a second photon then leads to a transition to high lying TN levels above the ionization limit. From both the SN and TN levels cross-linking can be induced [13]. The ratio between excitation via the singlet and triplet paths is a critical function of the laser parameters and cross-linking efficiency can be maximized by optimizing the relevant laser parameters. By changing the pulse length, the path leading to cross-linking can be chosen.

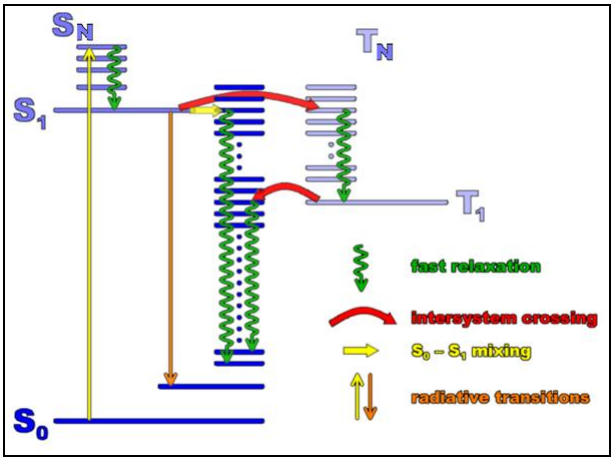


Fig.IV.1. Simplified energy level diagram. The pathways leading to cross-linking via the singlet S1 and the triplet T1 states are indicated. After excitation of the singlet S1 state the bases are promoted either directly with a second photon to high lying SN states above the ionization threshold or, after intersystem crossing to the T1 level, to states TN.

The practical use of UV lasers, however, depends on the efficiency of the cross-linking reaction as well as on the availability of methods adequate to determine DNA-protein contact points [14] and the identity of cross-linked proteins along the DNA sequences [15,16]. Although several techniques [16] have been used in the past, a major breakthrough, which will make the method generally attractive, has not yet been achieved. Considerable effort has been devoted in the last few years to studying these questions in various biological models using different laser systems, including nanosecond Nd:YAG lasers [15,17], picoseconds mode-locked Nd:YAG lasers [18,19] and excimer lasers.
All these studies put much emphasis on the nucleic acid side, neglecting the possibility that also proteins could be susceptible of photo-excitation by exposure UV. UV-laser cross-linking could therefore offer, as a novel zero length protein-protein cross-linker, an alternative, potentially powerful tool to investigate such protein-protein interactions. UV-laser cross-linking could obviate many of the problems associated with standard chemical cross-linking reagents, and put cross-linking in a proteome-wide position from the characterizations of protein-protein interactions, especially transient interactions, because the number of photons required for covalent complex formation can be delivered very rapidly, in nano-, pico- or even femtosecond intervals, and the high energy of the pulses should result in efficient cross-linking. The rationale of the approach is that light irradiation of the target

protein(s) should mediate the rapid production of cross-linked species. Schematically, photo-excitation abstracts an electron from a protein residue to produce a protein radical. The radical should form preferentially at the amino acids side chains that offer stabilization through aromatic group, mainly tyrosine, phenylalanine and tryptophane residues. Once the protein radical is created, it can react with neighbouring molecules to create either intermolecular or intramolecular cross-links with the extremely rapid kinetics of reaction typical of the chemistry of radicals.

IV.2. Laser set-up

Light Amplification by Stimulated Emission of Radiation (LASER or laser) is a mechanism for emitting electromagnetic radiation, often visible light, via the process of stimulated emission. The emitted laser light is (usually) a spatially coherent, narrow low-divergence beam, which can be manipulated with lenses. In laser technology, "coherent light" denotes a light source that produces (emits) light of in-step waves of identical frequency, phase, and polarization. The laser's beam of coherent light differentiates it from light sources that emit *incoherent* light beams, of random phase varying with time and position. Laser light is generally a narrow-wavelength electromagnetic spectrum monochromatic light.
A laser consists of three principal parts: i) an energy source (usually referred to as the pump or *pump source*), ii) a gain medium or laser medium, iii) two or more mirrors that form an optical resonator or optical cavity (Fig.IV.2).

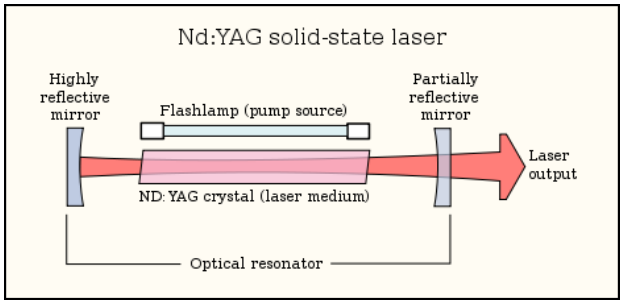


Fig.IV.2. Schematic rapresentation of a typical laser, showing the three major components.

The process of supplying the energy required for the amplification is called pumping. The energy is typically supplied as an electrical current or as light at a different wavelength. Most practical lasers contain additional elements that affect properties such as the wavelength of the emitted light and the shape of the beam.
The gain medium is a material with properties that allow it to amplify light by stimulated emission. It is a material of controlled purity, size, concentration, and shape, which amplifies the beam by the process of stimulated emission. It can be of any state: gas, liquid, solid or plasma. The gain medium absorbs pump energy, which raises some electrons into higher-energy (excited) quantum states. When the number of particles in one excited state exceeds the number of particles in some lower-energy state, population inversion is achieved and the amount of stimulated emission due to light that passes through is larger than the amount of absorption. Hence, the light of a specific wavelength is amplified (increases exponentially in power). The gain medium will amplify any photons passing through it, regardless of

direction; but only the photons aligned with the cavity manage to pass more than once through the medium and so have significant amplification. Each stimulated emission event returns a particle from its excited state to the ground state, reducing the capacity of the gain medium for further amplification. When this effect becomes strong, the gain is said to be *saturated*. The balance of pump power against gain saturation and cavity losses produces an equilibrium value of the laser power inside the cavity; this equilibrium determines the operating point of the laser.
The *optical resonator*, or optical cavity, in its simplest form contains a coherent beam of light between two parallel mirrors placed around the gain medium, so that the light passes through the gain medium more than once before it is emitted from the output aperture or lost to diffraction or absorption. Typically one will be a high reflector, and the other will be a partial reflector. The latter is called the output coupler, because it allows some of the light to leave the cavity to produce the laser's output beam.
The beam in the cavity and the output beam of the laser, if they occur in free space rather than waveguides (as in an optical fiber laser), are, at best, low order Gaussian beams.
The output of a laser may be a continuous constant-amplitude output (known as cw or continuous wave); or pulsed, by using the techniques of Q-switching, modelocking, or gain-switching. In pulsed operation, much higher peak powers can be achieved. Some types of lasers, such as *dye lasers* and *vibronic solid-state lasers* can produce light over a broad range of wavelengths; this property makes them suitable for generating extremely short pulses of light, on the order of a few femtoseconds (10^{-15} s).

IV.2.1. Nd:glass laser

The laser system used is a Nd:glass, which produces pulses with energy up to 4.3 mJ at wavelength of 1055 nm (pulse duration 1 ps). The Nd:Glass is a solid state laser whose active medium is constituted by Nd^{3+} ions installed as impurities in a bar of glass. It constitutes a variation of the Nd:YAG laser where the ions of Neodimium replace the ions Y^{3+} in some of the sites of the $Y_3Al_2O_{15}$ crystal. In such systems the laser oscillation happens on many longitudinal ways that can be locked in phase for producing ultrashorts impulses. The infrared pulse is doubled in frequency by a II-type KDP crystal in which also a temporal compression down to 200 fs takes place (energy up to 1 mJ and wavelength of 527 nm). The green pulse frequency is subsequently doubled by a non-linear II HG crystal getting a wavelength of 267 nm with variable intensity (between 350 µJ and 30 µJ).
In detail, this laser utilizes the chirped pulse amplification technique (CPA) to deliver ultra-short, ultra-intense pulses with no damage to the employed optics. Hence, it is constituted of four parts: i) laser oscillator, ii) pulse stretcher, iii) regenerative amplifier and iv) pulse compressor (Fig.IV.3).
The oscillator is composed by two active medium rods and it’s optically pumped by Xe flash-lamps. After the pumping, it delivers a pulse train having time spacing of some microseconds, few µJ being the pulse energy and 1 ps the pulse duration. This pulse then goes to the stretcher. It is an optical device composed by two diffraction gratings.
The first grating implies spatial dispersion whereas the second one has the role of spatially re-collimating the previously dispersed spectral component of the pulses.
The stretcher’s role is to extend the temporally duration of pulse and so to reduce the power of the peaks. Peaks with low power can be amplified in the regenerative

amplifier with no risk of damages for the active medium and the optics of the regenerative amplifier.

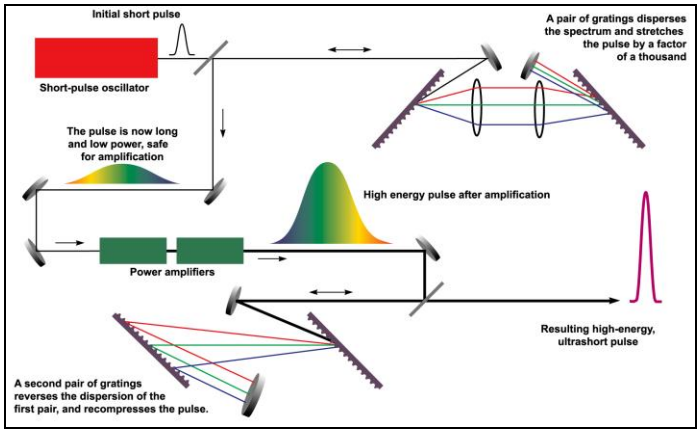


Fig.IV.3. Diagramatic scheme of chirped pulse amplification.

The pulse leaves the stretcher, having a length of 300 ps and goes to the regenerative amplifier, which is a laser cavity itself. The pulse coming from the stretcher gets strongly amplified by passing back and forth into the amplifier laser rod. The pulse oscillates within the amplifier cavity until it becomes amplified by a factor of about 10^4. It is, then, ejected from the regenerative amplifier cavity by means of an electro-optic ultra-fast switch. The amplified pulse is, then, sent to the compressor. It works in the opposite way with respect to the stretcher, so it compensates for the dispersion introduced by the stretcher through two diffraction gratings. This time, those colors which had traveled a longer distance into the stretcher and were, therefore, delayed take a shorter path and are re-phased in time with colors that arrive first.
The compressor is made of all-reflection optics so it can generate a pulse with a peak great power with no risk of damages.
The recompressed pulse has duration of 1 ps, energy of about 4 mJ and a repetition frequency of 33 Hz.

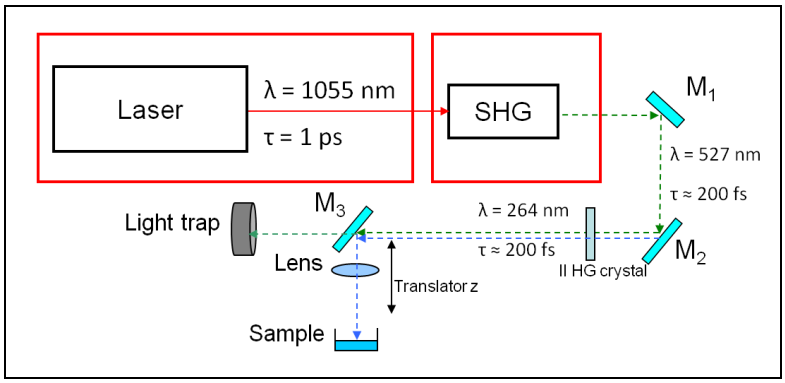


Fig.IV.4. Schematic diagram of the experimental set-up.

Such laser source is followed by a second harmonic generator (SHG), where the infrared pulse is doubled in frequency and a temporal compression down to 200 fs takes place (energy up to 1 mJ and wavelength of 527 nm) (Fig.IV.4). Second harmonic generation (SHG; also called frequency doubling) is a nonlinear optical

process, in which photons interacting with a nonlinear material (in this case two non-linear crystals of KDP, potassium dihydrogen phosphate crystal) are effectively "combined" to form new photons with twice the energy, and therefore twice the frequency and half the wavelength of the initial photons. It is able to generate ultras-short visible laser pulses starting with an infrared longer laser source.
The SHG produces pulses with duration of 250 fs, central wavelength of 527 nm and with a repetition frequency of 33 Hz.
The green pulse frequency is subsequently doubled by a non-linear II HG crystal getting a wavelength of 267 nm with variable intensity (between 350 µJ and 30 µJ).

IV.3. Reaction mechanism

In the hypothesis that the most likely reaction mechanism for protein-protein UV induced cross-linking is a radicalic one, we took advantages of the established strategies that have been developed to investigate the molecular mechanisms of radicalic reactions. Protein radical species have been generally detected using electron paramagnetic resonance spectroscopy (EPR). The technique of spin trapping was developed in the late 1960s to facilitate the detection of reactive free radicals by EPR spectroscopy [20]. Because most radicals are very reactive, and thus, short-lived (i.e., µsec to secs), the spin-trapping approach has been more widely used. With the spin-trapping approach, spin trap molecules can react with radical sites in proteins *in situ* and in real time. During this reaction, the spin trap forms a covalent bond at the radical site in the protein primary sequence forming a protein-spin trap radical adduct. In doing so, the radical adduct is, essentially, footprinting where the radical was initially formed and consequently trapped in the protein. The protein radical adduct is more stable and, consequently, more long lived (i.e. secs to minutes) than a protein radical, but decays to form an ESR-silent spin trap protein nitrone adduct in which the spin trap is covalently-bound to the protein.
Though spin trapping was developed initially to examine reactive radicals formed from low-molecular-mass compounds [20], developments over the last three decades have shown that this methodology can also be used to study radical formation on proteins and other macromolecules in both isolated and complex biological systems.
Nitroso spin traps, of which 2-methyl-2-nitrosopropane (MNP) (Fig.IV.5), have the advantage that the reactive radical attaches directly to the nitroso nitrogen atom, and is therefore in close proximity to the unpaired electron which is located primarily on the nitroxide function.

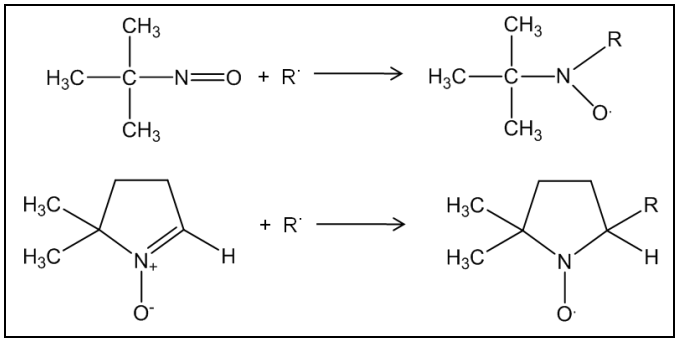
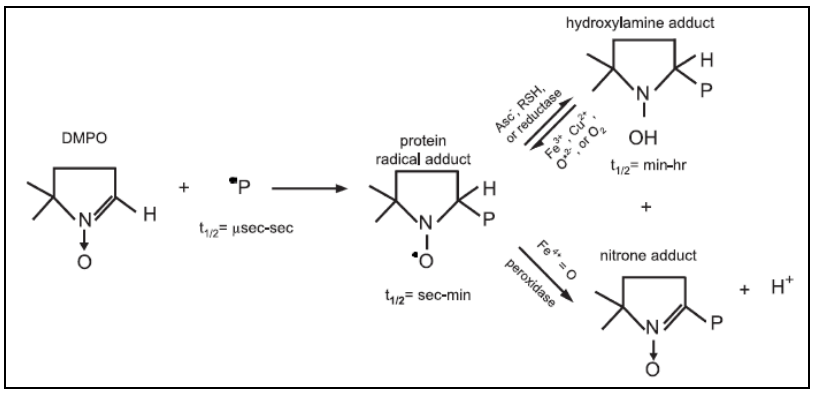


Fig.IV.5. Structures of 2-methyl-2-nitrosopropane (MNP) and 5,5-dimethyl-1-pyrroline N-oxide (DMPO) and their reaction with transient radicals to give nitroxide adducts.

The most popular of these spin traps is 5,5-dimethyl-1-pyrroline N-oxide (DMPO) (Fig.IV.5), which often form long lived adducts with a wider range of radical types (e.g., carbon-, oxygen-, sulfur-, and nitrogen-centered species). With nitrone traps such as DMPO, the reactive protein radical adds to a carbon atom adjacent to the nitrogen atom of the spin trap (Scheme 1). When DMPO traps a protein radical, a nitroxide radical adduct is formed with a covalent bond from DMPO to the protein (Scheme 1). The atom at which this bond is formed is a site of high electron spin density (or localization), which is sterically accessible to DMPO. Nitrone-derived adducts exist in three redox forms (Scheme 1): (1) the nitroxide radical adduct, (2) the corresponding hydroxylamine formed by one-electron reduction of the radical adduct, and (3) the corresponding nitrone formed by one-electron oxidation of the radical adduct. Of these, only the nitrone species is a stable product. At high enough concentrations, two nitroxide radical adducts will disproportionate to the hydroxylamine and the nitrone adducts [20,21].



Scheme 1

To take advantage of the more stable, trapped adducts, the Mason's research group reported the development of an antibody with specificity for the nitrone adduct, which is the only stable product of trapping radicals with DMPO [22]. Using this DMPO-specific antibody (immuno-spin trapping), they have successfully detected DMPO adducts in several heme-containing proteins [23-25]. In proteins, however, structural assignment of the specific amino acid residue or residues that form radical adducts with the spin trap molecule is difficult with immuno-spin trapping or ESR; thereby, requiring the use of additional biochemical techniques, such as site-directed mutagenesis and amino acid derivatization. These methods, however, suffer from potential drawbacks, such as perturbation of the protein's stability and conformation by mutations or derivatization. Hence, the formation of the amino acid radical(s) and/or transfer of the radical may be affected.

Recently, mass spectrometry has been utilized to study protein free radicals [26-29]. The spin trap molecules form a covalent bond with the protein resulting in a new nitroxide radical, which may be stable under electrospray ionization mass spectrometry conditions. By using a combination of peptide mapping by mass spectrometry it is possible to investigate the formation and the structures of radical adducts generated on proteins. To identify the precise amino acid residues (i.e. radical site) trapped by spin trap, mass spectrometry-based sequencing can be used.

IV.4. Aims of the project

The main goal of this work is to shape UV ultrashort laser pulses to induce efficient protein-protein cross-linking through the quantum control of molecular excitation. This would have a strong impact on the possibility study transient interactions among proteins, to map molecular interfaces in protein complexes, to gain dynamic information regarding how the contacts within a multi-protein complex might change over the time, because the number of photons required for covalent complex formation can be delivered very rapidly. Such photo-physical approach could obviate many of the problems associated with standard chemical cross-linking reagents, and put cross-linking in a proteome-wide position for the characterizations of protein-protein interactions *in vivo* in intact cells, since no chemical reagent has to enter the cell.

Essential to this aim is the multidisciplinary nature of this new bio-laser science, which benefits from the huge progresses in ultrashort laser physics. Due to the universality of computer-controlled learning optimization of ultrashort laser pulses, a broad field of applications opens up in all areas which use femtosecond laser technology. Among them, applications in complex systems, such as many biological systems, will greatly progress as there will be no requirement of a pre-constituted, theoretical model of the complex system under exam, unfeasible in most cases due to high degree of complexity.

In the first part of the work we have investigated the role played by different parameters characterizing the ultrashort laser source in peptide-petide photo-cross-link. We have varied laser pulse energy, the duration of exposition to the laser, laser beam focussing geometry and sample concentration in order to i) study the physical process of UV excitation of petide (proteins) and ii) maximize the peptide-peptide cross-linking absolute yield.

Moreover, since the software for the analysis of the cross-linked proteins require the mass shift introduced by the chemical modification, in order to put this strategy on a proteomics perspective, we have investigated the molecular details of the reactions. For this purpose we have studied the effect of the UV laser on peptides in absence and in presence of spin trap. The product of reaction was analysed by high resolution mass spectrometry in order to localize the amino acidic residues involved in the covalent bond formed after the irradiation.

The final aim of protein cross-linking is to introduce covalent bonds where non covalent interactions already exist, without affecting the three-dimensional structure by excessive introduction of “non physiological” bonds. The final goal of the “cross-linking project” will be to use the cross-linking reaction as a tool to “freeze” such “physiological” and transient interactions in protein complexes *in vivo*.

With these aims in mind, we have choose an oligomeric protein as the model protein in our tests: the stabilization of intersubunit interactions by cross-links would be “physiologically correct”, and the cross-link formation detected by SDS-PAGE analysis as the formation of species at molecular weight higher than the reference sample.

Finally we have applied this approach also to intact cell, in order to optimize the experimental parameter in a more complex contest and to define, in future, the molecular partners of a selected protein of interest.

The results herein presented were obtained in a collaborative and interdisciplinary project with the research group of Prof. Raffaele Velotta from the Dipartimento di Scienze Fisiche, Università di Napoli Federico II, the research group of Prof. Renata

Piccoli from the Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli Federico II, and the research group of Prof. Catherine Costello from the Departments of Biochemistry, Biophysics and Chemistry, Boston University School of Medicine, Boston, USA.
In particular, I acknowledge the “Programma di scambi internazionali” of Università di Napoli Federico II”, for funding my permanence in the lab of Prof. Costello.
A manuscript about the peptides section is in preparation.

IV.5. Material and Methods

Xenopsin, angiotensin, interleukin, Alcohol dehydrogenase (ADH) from *Saccaromyces cerevisiae*, 5,5-dimethyl-1-pyrroline N-oxides (DMPO) and the 2-methyl-2-nitrosopropane (MNP), dithiothreitol (DTT), L-ascorbic acid, trypsin, MALDI matrix α-cyano-4-hydroxycinnamic acid, iodoacetamide (IAM) and ammonium bicarbonate (AMBIC) were purchased from Sigma. Trifluoroacetic acid (TFA) is from Carlo Erba, Formic acid from Baker and Acetonitrile (ACN) Ultra gradient from Romil; Comassie Brilliant Blue was from Pierce.
The standards of molecular weight for the calibration of the Voyager-DE STR system has been purchased from Applied Biosystems.
The monobasic sodium phosphate (NaH_2PO_4) was purchased from Sigma.

IV.5.1. UV laser peptide-peptide cross-linking

Standard peptides (10 nmol of angiotensin, xenopsin, interleukin) individually or mixed together were dissolved in 6 µl of ammonium bicarbonate buffer (10 mM pH 7.0) and irradiated with a laser energy of 110 µJ, at a frequency of 2 kHz and at a λ of 267 nm, at room temperature for different time intervals from 1 sec to 3 min. The reaction was stopped by adding 4 µl of ascorbic acid at final concentration of 20 mM dissolved in the same buffer immediately before use.
The same conditions were used to irradiate each single peptide in presence of 4 µl DMPO (final concentration 100 mM and 10 mM) or MNP (final concentration 10 mM). Sample concentration and desalting were performed using a reverse phase C18 Zip Tip pipette tips (Millipore). The peptides were eluted with 20 µl of a solution made of 50% Acetonitrile, 0.5% Formic acid in Milli-Q water at a final concentration of 25 µM.

IV.5.2. UV laser protein-protein cross-linking

Alcohol deidrogenase (2 µg/µl) was dissolved in ammonium bicarbonate buffer (10 mM pH 7.0) in presence of ascorbic acid 1mM and fractions of 18 µl (1 nmol of monomer) were exposed to the laser in the same previously conditions for time intervals of 7 sec and 15 sec. After the irradiation ascorbic acid at final concentration of 20 mM was added to stop the radical reaction. The same experiments were performed on 500 µl of a diluite ADH solution (72 ng/µl) in the same buffer.
The protein samples were concentrated and purified using Amicon® Ultra-0.5 mL Centrifugal Filter Devices (Millipore) with 30 kDa. The proteins were resuspended with 20 µl of a solution made of 50% Acetonitrile, 0.5% Formic acid in Milli-Q water, at a final concentration of 10 µM or resuspended in Laemmli sample buffer and analysed by SDS-PAGE.

IV.5.3. UV laser in vivo

Recombinant Human ApoA-I, cloned in the expression vector pGEX-4T-3, was expressed in Hep75 human cells. Each pellet containing 200.000 cells was resuspended in 500 µL of PBS buffer (13.7 mM NaCl, 0.27 mM KCl, 10 mM Na_2HPO_4, 0.2 mM KH_2PO_4) and irradiated at this different laser energy, frequency and time conditions (110 µJ, 0,2 kHz for 2 minutes; 110 µJ, 2 kHz for 12 seconds; 55 µJ, 2 kHz for 12 seconds; 55 µJ, 0,2 kHz for 2 minutes; 110 µJ, 2 kHz for 1 minute). After irradiation the cells were centrifugated at 13,000 rpm for 5 minutes. The pellet was resuspended in 100 µL of PBS buffer, NP40 1% and protease inhibitors cocktail (Sigma Aldrich), incubated for 30 min at 4°C and lysed by sonication. The extracts were centrifuged at 12,000 rpm for 30 min at 4°C and the pellet was resuspendend in Laemmli sample buffer.

IV.5.4. SDS-PAGE analysis

Proteins were separated on SDS-PAGE on 12.5% polyacrylamide gel (Laemli, 1970). Proteins were detected by Coomassie Brilliant Blue G-Colloidal (Pierce, Rockford, USA). Bands of interest were excised and subjected to *in situ* digestion with trypsin (100 ng in 50 µL of 10 mM ammonium bicarbonate pH 7.0) for 16 h at 37°C.

IV.5.5. Hep75 human cells western blotting analysis

For immunodetection, after SDS-PAGE, proteins were electroblotted on a Hybond-ECL membrane (GE Healthcare) for 1 h at 15 V. The membrane was saturated with 5% non-fat dry milk for 1 h at room temperature and then incubated with anti-human ApoA-I polyclonal antibodies (Dako) at 4°C for two hours. An anti-rabbit was used as a secondary antibody and a chemiluminescence detection system (West Pico, Pierce) was used to visualise immunopositive proteins. The amount of recombinant ApoA-I was estimated densitometrically on blots by means of ImageJ 1.40 g software. Signals were compared to those obtained with serial dilutions of commercial ApoA-I.

IV.5.6. MALDI-TOF analysis

MALDI-TOF mass spectra were recorded using an Applied Biosystem Voyager STR instrument equipped with a nitrogen laser (337 nm). 1 µL of the analyte mixture was mixed (1/1, v/v) with a 10 mg/mL solution of R-cyano-hydroxycinnamic acid in acetonitrile/50 mM citrate buffer (2/3, v/v) and was applied to the metallic sample plate and dried down at room temperature. Acceleration and reflector voltage were set up as follows: target voltage at 20 kV, grid at 66% of target voltage, delayed extraction at 150 ns to obtain the best signal to-noise ratios and the best possible isotopic resolution. Mass calibration was performed using external peptide standards purchased from Applied Biosystems. Raw data were analyzed using the computer software provided by the manufacturer as monoisotopic masses.

IV.5.7. MS/MS analysis

The cross-linked product was analyzed in MS and MS/MS mode by high resolution mass spectrometry using hybrid Qh/FTICR (SolariX equipped with a 12 T actively shielded magnet, Bruker Daltonics, Billerica, MA, USA). This instrument was equipped with a nano-spray source operated in positive mode. The high voltage used for ionization was between 1,000 and 1,500 V and nitrogen was used as a counter current drying gas with a temperature maintained at 180°C. To record the spectrum the ECD a current of 1.6 A, an ECD bias of 1.5 V and an electron pulse length of 0.07s was employed. For the CID spectrum the collision voltage was set between 8 and 15 V. The collision gas was argon at a pressure reading of $6x10^{-6}$ mbar. In the ETD mode, radical negative ions of fluoranthene, produced in the CI source, are used as the reagent ions. During the ETD experiments, the crucible is heated to about 60°C so as to sublime the fluoranthene. The fluoranthene gas then passes into the CI chamber where it is ionized via chemical ionization with methane. Negatively charged fluoranthene ions are then extracted out of the CI source via a set of lenses. The methane was connected to the CI source with stainless steel tubing. The filament was set to emit a 3 µA. The acceleration time of reagent was 50-100 ms and the reaction time was 20 ms.
Mass spectra were acquired in the positive-ion mode over the m/z range 200-2000 at a resolution of 60,000. Mass accuracy was within 1 ppm. Compass Data analysis software (Bruker Daltonics) was used for data analysis; peptide sequencing and cross-linking site assignments were conducted manually employing a 1 ppm error on the fragment ions.

IV.6. Results

IV.6.1. Laser UV-induced peptide-peptide cross-linking

The first priority of this study was to verify whether an ultrashort UV pulsed laser is able to induce an efficient cross-link in simple model such as peptide and to localize the amino acid involved in the formed bond. We assumed that upon absorption of the ultraviolet light the chemical moieties of amino acids (above all tyrosines, phenylalanines and tryptophans) can produce radicals that are then able to react and eventually introduce covalent bonds between peptides.
To induce the cross-link we used as source of powerful UV radiation the Nd:glass one, which produces pulses with energy up to 4.3 mJ at wavelength of 1055 nm (pulse duration 1 ps). The infrared pulse is doubled in frequency by a II-type KDP crystal in which also a temporal compression down to 200 fs takes place (energy up to 1 mJ and wavelength of 527 nm). The green pulse is subsequently doubled by a non-linear II-type KDP crystal getting a wavelength of 267 nm with variable intensity (between 350 µJ and 30 µJ). The repetition rate of the whole laser source is 33 Hz.
Samples of different peptides were exposed to the UV laser at a laser energy of 110 µJ and different irradiation time (1 sec-3 min), purified by a reverse phase C18 Zip Tip and subsequently analysed by MALDI-TOF mass spectrometry, in order to detect the formation of dimeric species due to the introduction of intermolecular cross-links.

IV.6.1.1 MS spectra

As an example, the MALDI-TOF spectra of a mixture of xenopsin, interleukin and angiotensin that have been exposed to the UV-laser light for 10 sec are reported in comparison to the spectra obtained for the same mixture before the exposure to the laser (Fig.IV.6, panel B) Two signals are generated upon UV-laser exposure: (Fig.IV.6, panel A) the peak at m/z 1958.13 could account for two molecules of xenopsin cross-linked -2 Da, and the signal at m/z 2274.24 that correspond to one molecule of xenopsin cross-linked to one molecule of angiotensin, -2 Da. The other three signals present in the spectrum correspond to the $[M+H]^+$ of the standard peptides (m/z 980.57 for xenopsin, m/z 1005.44 for interleukin and m/z 1296.69 for angiotensin). It should be noted that, apparently, interleukin (that has no aromatic residue) is not involved in any of the species generated upon UV exposure, as also confirmed when the same experiment was carried out with interleukin as unique peptide model, where no species but monomeric interleukin could be detected. On the contrary, the same experiment performed in presence of isolated angiotensin shows the presence of the homodimer of the angiotensin (data not shown).
In order to verify the specificity of the observed homodimer and heterodimer, a mixture of xenopsin, interleukin and angiotensin without performing UV cross-linked irradiation was analysed by MALDI-MS in the similar conditions than those of the previous samples. No homodimer and heterodimer was evidenced confirming the specificity of cross-linking species observed after UV irradiation.

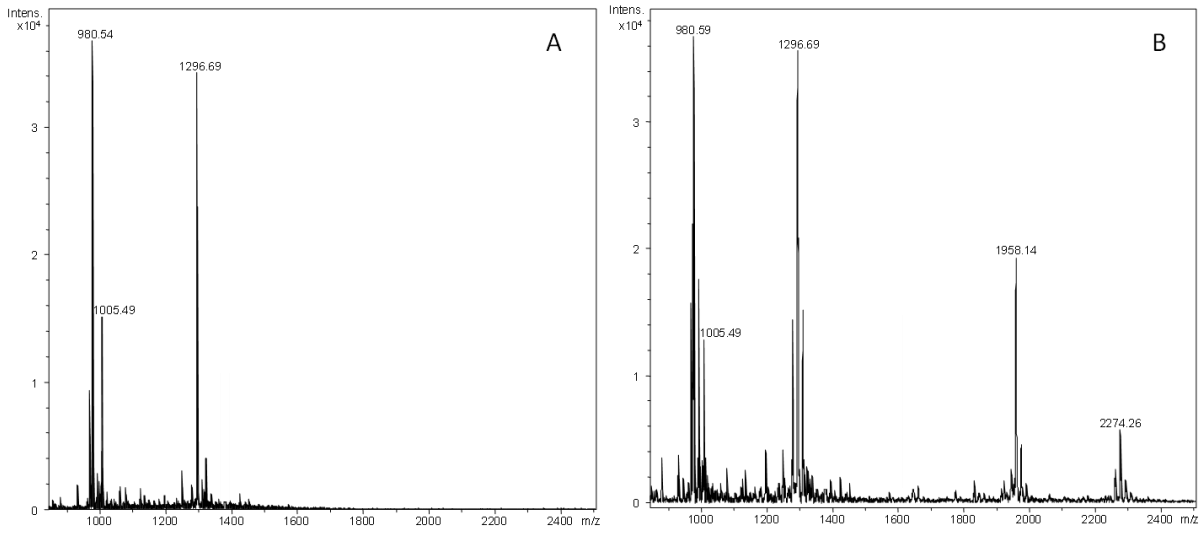


Fig.IV.6. MALDI-TOF m/z spectrum of mixture of xenopsin (980.54), interleukin (1005.49) and angiotensin (1296.69) not irradiated (panel A) and irradiated for 10 sec (panel B). The peaks at m/z 1958.13 corresponding to two molecules of xenopsin cross-linked and the peak at m/z 2274.24 at one molecule of xenopsin cross-linked with one molecule of angiotensin.

Only aromatic side chain containing peptides seemed to be affected by the exposure to UV-laser light suggesting a new concept: the reaction mechanism requires the aromatic amino acid.
In order to verify this view, the amino acids covalently linked were identified by tandem mass spectrometry experiments.

IV.6.1.2 Tandem mass spectrometry

Tandem mass spectrometry experiments were carried out on the Solarix (Bruker Daltonics, Billerica, MA, USA) using a 25 µM solution of the peptide mixture exposed to the UV-laser as described above, dissolved in 50% acetonitrile, 0.5% formic acid was directly infused in the nano-spray source. After isolation of interested ions in the Quadripole/hexapole, they were fragmented by CID, ECD and ETD mode. Displayed in Fig IV.7 and IV.8 are the CID (panel A) and ECD (panel B) MS/MS spectra generated from $(M+3H)^{3+}$ ion (m/z 653.38) of the homodimer of xenopsin and $(M+4H)^{+4}$ ion (m/z 569.31) of the heterodimer of angiotensin-xenopsin.
Single cleavage products are present in these spectrum that can be assigned as originating from both of the two peptide chains, primarily type a-, b-, and y-ions. The fragment ions corresponding to these two peptides are, therefore, designated with either the α or the β subscript to indicate the peptide of origin.
Fragmentation observed in the low-energy CID spectrum of the homodimer (Fig.IV.7, panel A) is dominated by ions corresponding to a, b and y ions.
In the ECD spectrum of this study, the signal corresponding to the charge reduced molecular ions $[M+nH]^{(n-1)+\cdot}$ was observed. As expected, these spectra were dominated by c and z· ion series andthe ECD spectrum (Fig.IV.7, panel B) is more informative than the low-energy CID spectrum.
In both spectra in addition to short z- and a,b,c-ion series, it is possible to observe several $y_\alpha y_\beta$, $z_\alpha z_\beta$, $b_\alpha b_\beta$, $c_\alpha c_\beta$-ions detected in their single and doubly charged states.
For example in the ECD spectrum of the cross-linked xenopsin EGKRPWIL, two c-ions are present at m/z 314.18 (c_3) and 567.33 (c_5). Moreover, fragment ions originating from cleavages involving both peptide chains were observed at m/z 1730.94 ($a_{8\alpha}a_{6\beta}$), 865.98 $(c_{8\alpha}c_{6\beta})^{2+}$, 1884.02 ($a_{8\alpha}a_{7\beta}$), 922.52 $(c_{8\alpha}c_{7\beta})^{2+}$, 1392.78 ($z_{3\alpha}z_{8\beta}$), 1645.93 ($z_{5\alpha}z_{8\beta}$), 823.47, 887.52, 916.03 $(z_{5\text{-}7\alpha}z_{8\beta})^{2+}$.
The $z_{3\alpha}z_{8\beta}$, $(c_{8\alpha}c_{6\beta})^{2+}$ and $a_{8\alpha}a_{6\beta}$ ions at m/z 1392.78, 865.98 and 843.47 constituted the key signals to univocally assess that the Trp is the amino acid involved in the cross-link induced upon laser exposure.
The tandem mass spectra of the angiotensin (α-chain) cross-linked to xenopsin (β-chain), shown in Fig.IV.8, generated a pattern similar to the example described above, and consisted of dominant b,y-series ions for the CID spectrum and c,z-series ions for the ECD spectrum. The only putative aromatic amino acids in these two sequences are Tyr-4 and Phe-8 on the angiotensin and the Trp-6 on the xenopsin. The MS/MS spectra of the quadruply charged ion at m/z 569.31 revealed several ions for both peptide chains; $y_{2\alpha}$ at m/z 269.16, $z_{3\alpha}$ at m/z 400.21, $y_{4\alpha}$ at m/z 513.27 and $z_{5\alpha}$ at m/z 634.32 (α-chain) and $c_{3\beta}$ at m/z 314.18, $b_{4\beta}$ at m/z 453.25 and $b_{5\beta}$ at m/z 567.33 (β-chain). However a nearly complete a,b and z,y-ion series originate from cleavage reactions involving both peptide chains are also present, i.e., m/z 944.50 and 994.03 $(z_{7\text{-}8\alpha}z_{8\beta})^{2+}$, m/z 938.51 and 1012.05 $(c_{7\text{-}8\alpha}z_{8\beta})^{2+}$, 764.92 and 821.46 $(c_{4\text{-}5\alpha}z_{8\beta})^{2+}$, 981.53, 697.38 and 716.39 $(z_{10\alpha}z_{5\text{-}7\beta})^{2+}$.
The $c_{10\alpha}c_{6\beta}{}^{3+}$ at m/z 697.38, $(z_{10\alpha}z_{3\beta})^{2+}$ at m/z 854.95, $(z_{7\alpha}z_{8\beta})^{2+}$ at m/z 938.51, and $(c_{4\alpha}z_{8\beta)}{}^{2+}$ at m/z 764.91 ions constituted the key signals to univocally assess that the Tyr-4 and the Trp-6 are the amino acids involved in the cross-link induced upon laser exposure.

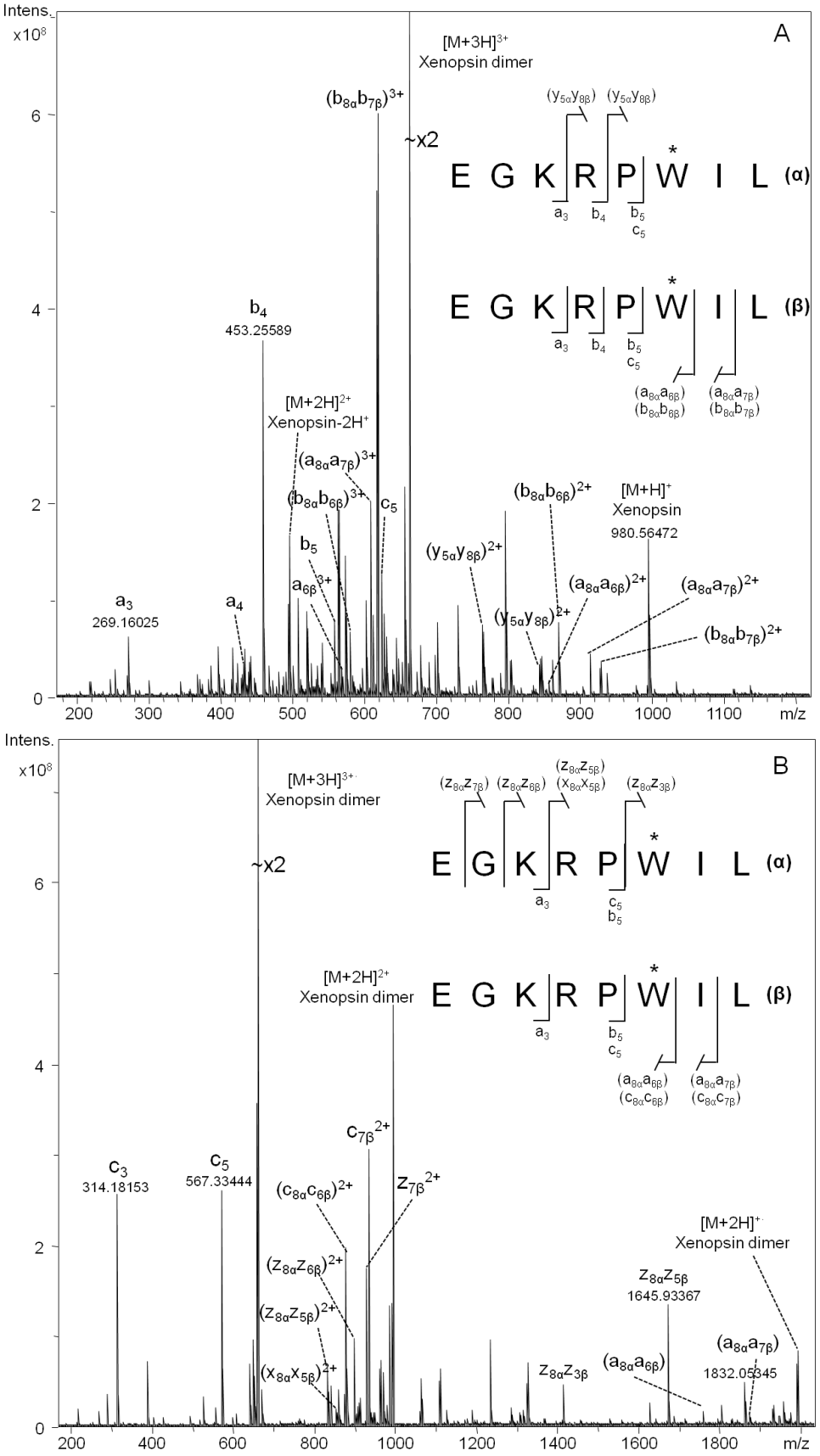


Fig.IV.7. ESI-CID (panel A) and ESI-ECD (panel B) m/z spectra of xenopsin homodimer. The $(M + 3H)^{3+}$ peak at m/z 653.38 was selected as the precursor ion.The asterisk indicates the residues imply cross-linking sites. The cross-linked sites on the sequence are indicates with a crossed line.

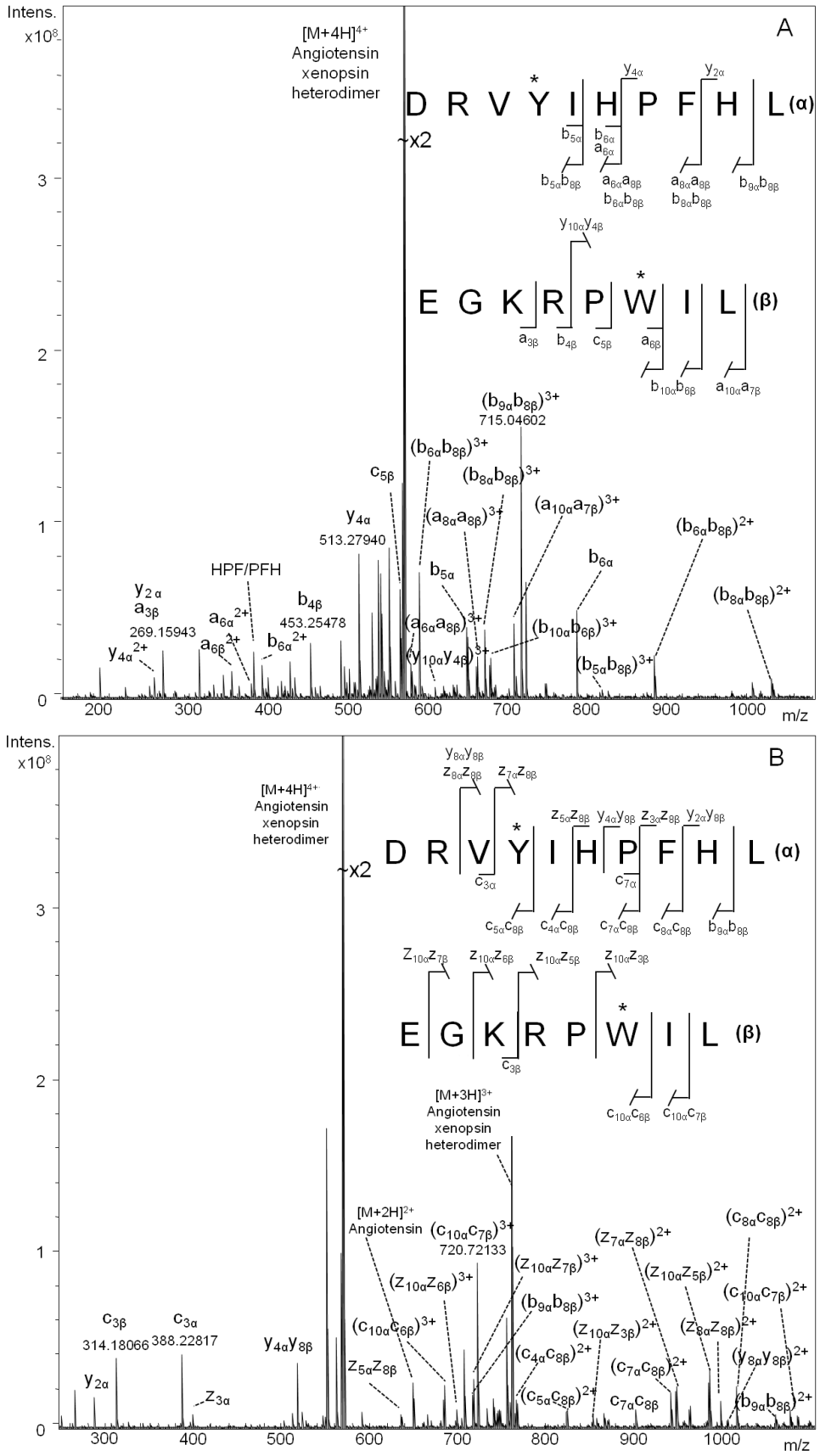


Fig.IV.8. ESI-CID (panel A) and ESI-ECD (panel B) m/z spectra of angiotensin-xenopsin heterodimer. The $(M + 4H)^{4+}$ peak at m/z 569.31 was selected as the precursor ion. The asterisk indicates the residues imply cross-linking sites. The cross-linked sites on the sequence are indicates with a crossed line.

IV.6.2. Characterization of the reaction mechanism

The most likely reaction mechanism for UV induced cross-linking is a radicalic one. Protein radical species can be spin trapped to give an adduct more stable than the primary radical. Nitrone spin traps, of which 5,5-dimethyl-1-pyrroline-N-oxide (DMPO) is the most commonly employed, and nitroso spin traps, such as 2-methyl-2-nitrosopropane (MNP), often form long-lived adducts that can be easily detected by EPR. Moreover, the spin trap molecules form a covalent bond with the protein resulting in a new nitroxide radical.
Therefore, in order to investigate the molecular nature of the intermediates involved in the process, UV-laser induced peptides cross-linking experiments were carried out as described above for either 10 sec or 1 min but in the presence of radical spin trap molecules, either 5,5-dimethyl-1-pyrroline N-oxide (DMPO), or 2-methyl-2-nitrosopropane (MNP). Similar experiments were also carried out on aliquots where the spin traps were added only after the exposure to the UV-laser beam for some extent (30 - 60 sec).

IV.6.2.1 MS spectra

MALDI-MS analysis of the xenopsin irradiated in presence of 10 mM DMPO (Fig.IV.9, panel B) showed, compared to the same sample not exposed to laser (Fig.IV.9, panel A), the presence of an additional signal at a 1091.61 m/z value, which is shifted of +111.07 Da in respect to the theoretical mass of the unmodified peptide. These data suggest that a single DMPO molecule is trapped on the xenopsin peptide. The DMPO nitrone adduct is a specific marker for where the radical is formed.

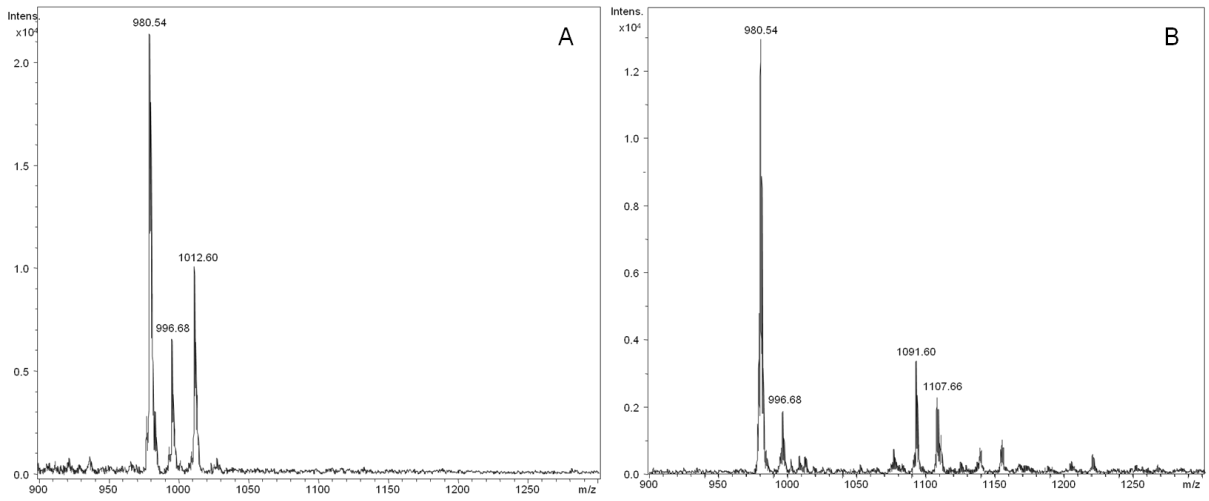


Fig.IV.9. MALDI-TOF m/z spectrum of mixture of xenopsin (980.54) not irradiated (panel A) and irradiated in presence of DMPO for 10 sec (panel B). The signals at 1091.60 m/z value corresponding to one molecule of xenopsin cross-linked with one molecule of DMPO.

IV.6.2.2 Tandem mass spectrometry

To validate the specificity of DMPO binding and to determine which amino acids contained a trapped DMPO molecule, we perfomed MS/MS analyses of the same mixture on the Solarix (Bruker Daltonics, Billerica, MA, USA) using CID, ECD

and ETD activation modes as described above. As an example the ECD spectrum of the $(M + 2H)^{2+}$ ion (m/z 546.32) is visualised in figure IV.10.
Some b and c (b_4, c_5, c_6) series ions were observed that accounts for cleavages of amino acids from the N-terminus of the peptide backbone. In addition, a peak corresponding to the charge reduced molecular ion $[M+2H]^{+\cdot}$ at m/z 1092.61 and fragment ion of m/z 1075.60 (labelled as $-H_2O$), corresponding to the loss of water molecule from the peptide, were observed. Several ions are observed which correspond to a fragment ion plus a DMPO molecule. For example, the ion of m/z 639.38 (labelled as y_{4DMPO}) corresponds in mass to a y_4 ion plus DMPO and the ion of m/z 960.54 (labelled as b_{7DMPO}) corresponds in mass to a b_7 ion containing DMPO. Moreover, it is possible to notice the presence of different internal fragments of the peptide plus one DMPO molecule (i.e. KRPW+DMPO at m/z value 679.38).
The concomitant presence of the series ions y_{4DMPO}, $b_6{}^{2+}{}_{DMPO}$ and c_5 indicates the DMPO adduct on Trp-6 in the xenopsin peptide. It should be noted that fragmentations didn't occurred only on the peptide backbone because few ions (y and c type) resulting from the cleavage of DMPO binding site were observed.

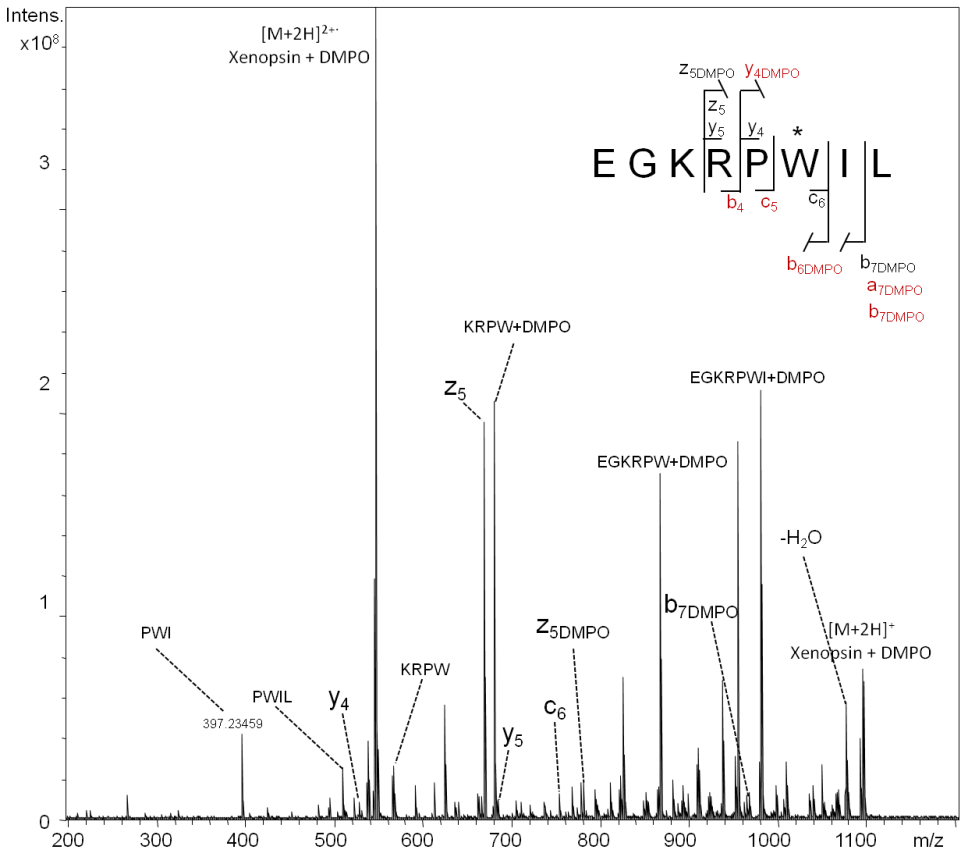


Fig.IV.10. ESI-ECD m/z spectra of xenopsin cross-linked with DMPO. The $(M + 2H)^{2+\cdot}$ peak at m/z 546.32 was selected as the precursor ion. The asterisk indicates the residues imply cross-linking sites. In red are indicated the ion series found in the CID spectrum.

Similar MS/MS analyses were performed for the same peptide irradiated in presence of 10 mM MNP (Fig. IV.11).
Again, the fragment ions preserving the MNP allowed to assess univocally that the Trp-6 is the amino acid involved in the bond with MNP.

Displayed in Fig.IV.12 and IV.13 are ECD spectra generated from $(M + 3H)^{3+}$ ion (m/z 469.92 and 461.92) of angiotensin peptide linked to DMPO and MNP respectively. We can assume analogous consideration for the angiotensin peptide linked to the spin trap molecule. A closely examination of both spectra shows the presence of fragment ions originating from cleavages involving the peptide chains plus the spin trap molecule. The interpretation of these data was more difficult because of competition of the backbone peptide cleavage with the DMPO binding cleavage. Nevertheless, the product ions preserving the DMPO allowed the assignment of the DMPO and MNP molecule to the Tyr-4 residue of the angiotensin.

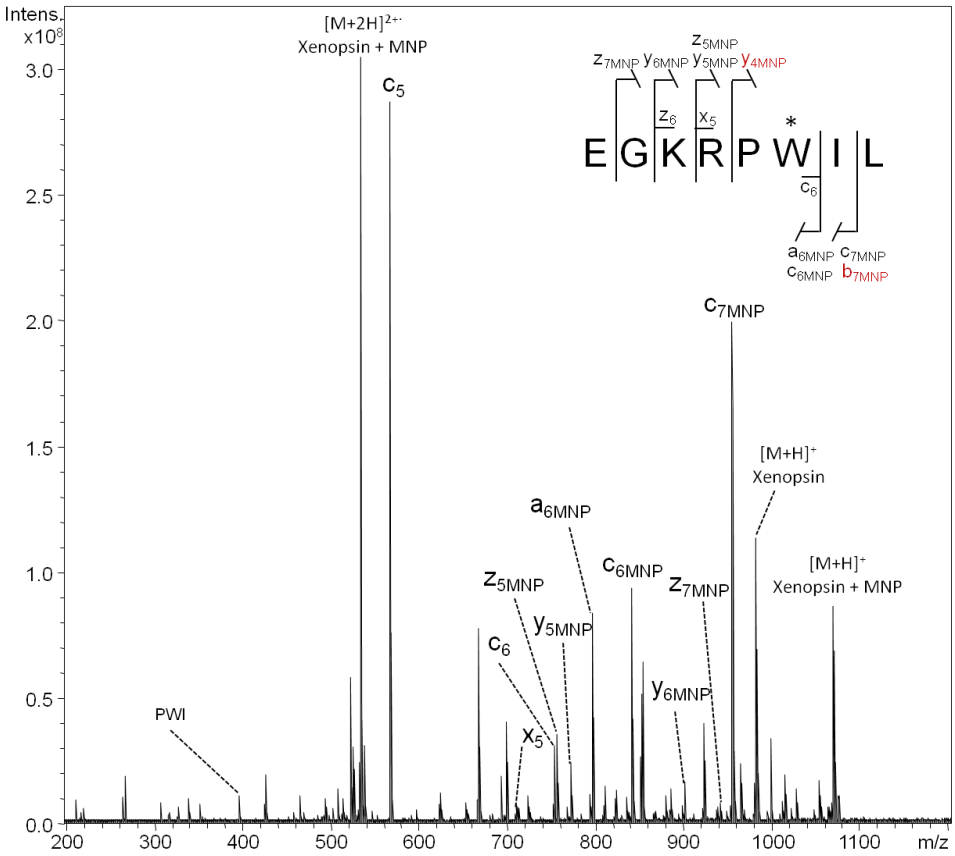


Fig.IV.11. ESI-ECD m/z spectra of xenopsin cross-linked with MNP. The $(M + 2H)^{2+\cdot}$ peak at m/z 534.32 was selected as the precursor ion. The asterisk indicates the residues imply cross-linking sites. In red are indicated the ion series found in the CID spectrum.

It is worth mentioning that addition of the spin trap immediately after UV-laser exposure did not result in any spin trap-peptide adduct. This suggests that radical species that have formed during UV-laser exposure, extinguished within the exposure time, and no long-lasting radical is formed that survives the exposure time.

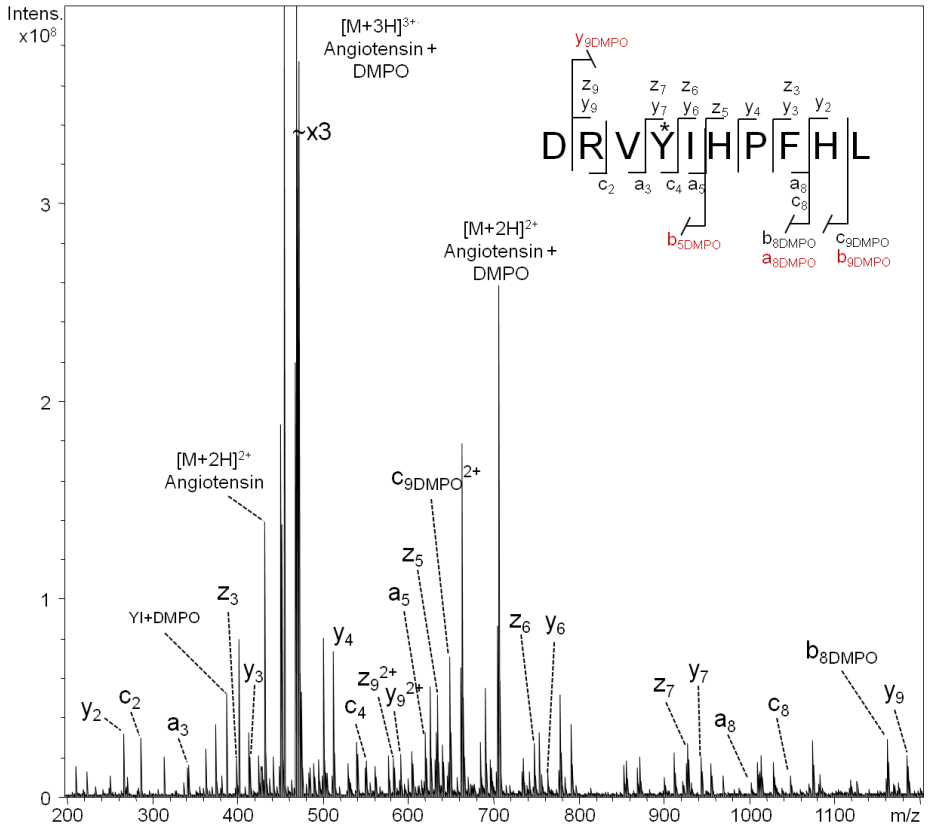


Fig.IV.12. ESI-ECD m/z spectra of angiotensin cross-linked with DMPO. The $(M + 3H)^{3+\cdot}$ peak at *m/z* 469.92 was selected as the precursor ion. The asterisk indicates the residues imply cross-linking sites. In red are indicated the ion series found in the CID spectrum.

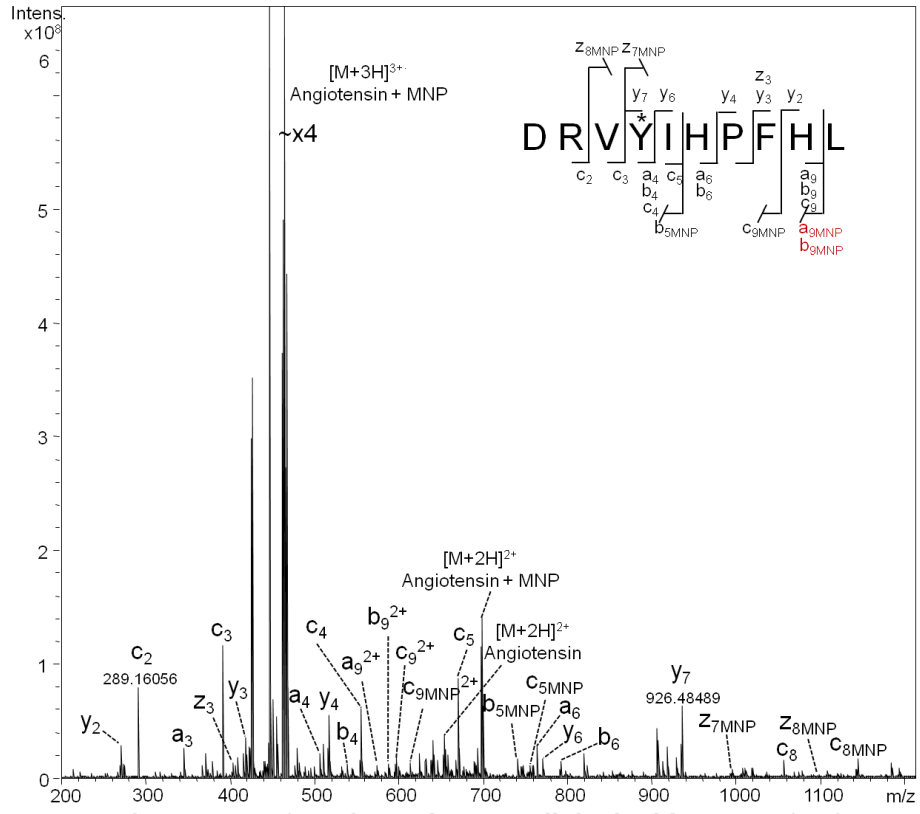
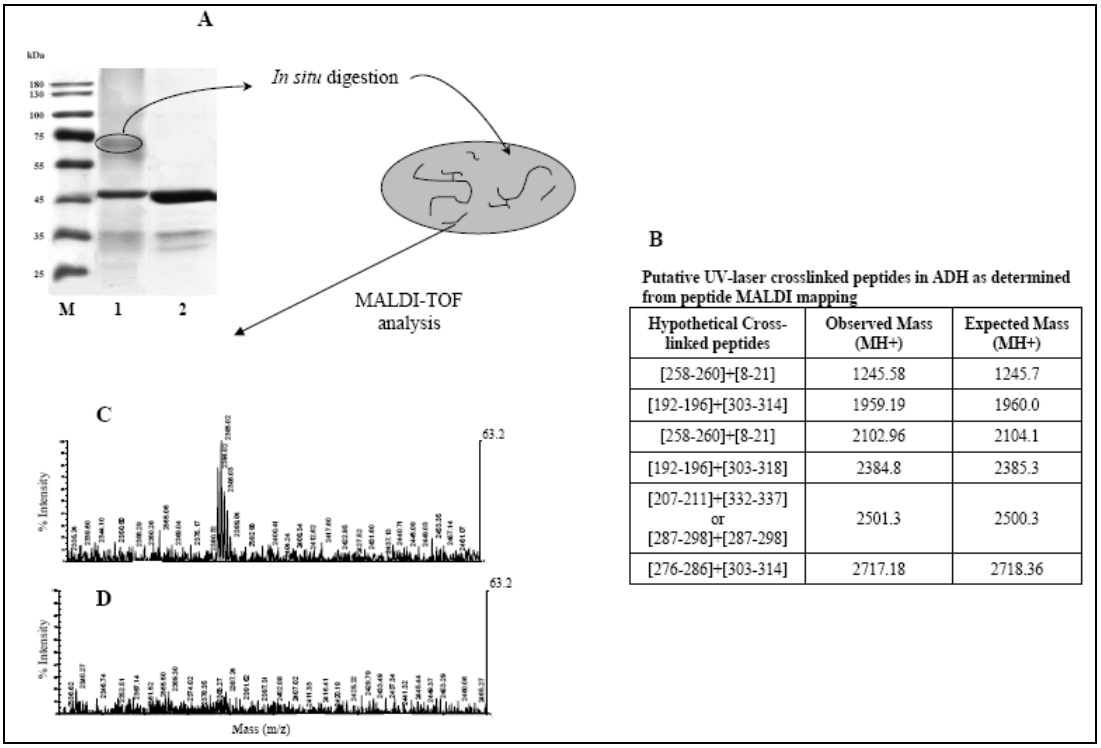


Fig.IV.13. ESI-ECD m/z spectra of angiotensin cross-linked with MNP. The $(M + 3H)^{3+\cdot}$ peak at m/z 461.92 was selected as the precursor ion. The asterisk indicates the residues imply cross-linking sites. In red are indicated the ion series found in the CID spectrum.

IV.6.3. Laser UV-induced protein-protein cross-linking

In our experiments, the alcohol dehydrogenase from *Saccaromyces cerevisiae* (ADH) was used as protein model: It is a homo-tetrameric, not glycosylated protein containing several aromatic amino acids, whose three-dimensional structure has been solved at high resolution (PDB code 2HCY). The stabilization of intersubunit interactions by cross-links would be “physiologically correct” and, none the less, the intersubunit cross-link formation can be easy monitored by SDS-PAGE analysis as the formation of species at molecular weight higher than the reference sample. The availability of a three-dimensional model will be helpful in the analysis of the “physiological significance” of the introduced cross-links.
Samples of ADH were exposed to the UV laser with different combinations of laser energy (30-350 µJ) and irradiation time (0.5-5 min), the SDS-PAGE analysis of selected samples of preliminary experiments carried out with 0.2 nmols of protein in 12 µl sodium-phosphate buffer 10 mM PH 7.0 (data not shown).
Depletion of the species with the apparent MW corresponding to ADH subunit was observed in all the conditions tested. Moreover, in all the samples is evident the formation of aggregates with high molecular weight, that did not penetrate inside the meshes of the gel.
Parameters such as protein concentration, exposure time and laser energy were therefore further varied in order to optimize cross-linking conditions. Moreover, considering the radicalic nature of the mechanism for the cross-linking reaction, we tested whether we could modulate the extent of this reaction by addition of a common antioxidant, such as the ascorbic acid, to the mixture.
Several tests were performed varying the laser energy (30-350 µJ), the concentration of ascorbic acid (0,1-20 mM), exposure time (0.5-30 min), and protein concentration (0.5-30 mM) (data not shown), in order to modulate and optimize the cross-linking reaction.
The best condition tested was exposure of ADH to the laser beam at an energy of 220 µJ with the addition of ascorbic acid to a final concentration 5 mM after 1 min exposure. After laser exposure the reaction mixture was separated by SDS-PAGE followed by Comassie blue staining (Fig.IV.14, panel A). Distinctive bands above the ADH subunit are clearly observed, located at an apparent molecular weight of about 70 kDa (compatible with a dimeric form of ADH). Dimeric species were considered our target for further analysis of the cross-linked protein, since they correspond to the simplest product of covalent oligomerization and therefore, the chance of cross-links to perturb the conformation of the protein is reduced to a minimum.
These bands were excised and subjected to enzymatic in-gel digestion with trypsin over night. For comparative purposes, the gel bands from lane corresponding to the monomer in the exposed sample as well as from the reference sample were also excised and *in situ* trypsin digested. The tryptic digests were then analysed by MALDI-MS and the experimentally obtained mass values were mapped onto the anticipated sequence of ADH, thus confirming the identity of the protein bands.
Fig.IV.14, panel B reports the signals, distinctive of the dimeric bands that could not be assigned to any linear peptide within the amino acid sequence of ADH and that could tentatively be ascribed to cross-linking products. As an example the signal at 2384,8 (Fig.IV.14, panel C), that is absent in the digest of the corresponding monomeric species (Fig.IV.14, panel D) as well as in the reference sample, can be interpreted as peptide 192-196 (m/z 597,3) cross-linked to peptide 303-318 (m/z 1791), with the loss of two atoms of hydrogen.

However, identification of the amino acids actually cross-linked, could not be possible simply on MALDI-MS data, and further analyses by MSMS are currently in progress.



Hypothetical Cross-linked peptides	Observed Mass (MH+)	Expected Mass (MH+)
[258-260]+[8-21]	1245.58	1245.7
[192-196]+[303-314]	1959.19	1960.0
[258-260]+[8-21]	2102.96	2104.1
[192-196]+[303-318]	2384.8	2385.3
[207-211]+[332-337] or [287-298]+[287-298]	2501.3	2500.3
[276-286]+[303-314]	2717.18	2718.36

Fig.IV.14. Molecular characterization of the cross-linked ADH dimer. (A) Electrophoretic separation of the UV-cross-linked dimer. UV-laser irradiated ADH (0.2 nmol in Na-phosphate 10mM, pH 7.1 quenched with 5 mM ascorbic acid after exposure for 1 min to the laser beam (E=220 µJ)) (lane 1). 0.1 nmol of ADH as reference sample, not irradiated (lane 2). Molecular weight standards (lane M). The gel was stained with Coomassie Blue. (B) Putative cross-linked peptides as determined from MALDI mass mapping. MALDI-TOF-MS spectrum from m/z 2333 to 2466 of the tryptic digest of the UV-laser cross-linked ADH (C), and of the reference sample, not irradiated.

IV.6.4. Laser UV-induced in vivo cross-linking

What learnt so far on peptides and protein models in *in vitro* studies, was useful for setting up experiments *in vivo*. We have applied this approach also to intact cells, in order to exploit this technique as the final, most attractive future perspective, and tremendously powerful tool for the definition of the molecular partners of a selected protein of interest.

We have performed the *in vivo* experiment on Hep75 human cells expressing the apolipoprotein A-I (ApoA-I), our bait, which is known to be responsible for hereditary, systemic amyloidoses, when specific mutations occur in the ApoA-I gene. The 13 amyloidogenic variants of ApoA-I described so far are directly associated with *in vivo* formation of amyloid deposits [30] that accumulate in tissues and organs, such as heart, kidney or liver, with severe consequences [31].

Samples of ApoA-I recombinant Hep75 cells in PBS buffer were exposed to the UV laser with different combination of laser energy (55 and 110 µJ), frequency (0.2 and 2 KHz) and irradiation time (12 sec to 2 min). Fig.IV.15 shows the SDS-PAGE and western blotting analysis of total cellular extract after exposure to the UV laser.

In each lane of the western blotting, it is possible to observe the Apo-A1 protein band at the molecular weight of 25 kDa. Moreover, in the lanes 2 and 3 the presence of aggregates at high molecular weight is evident. These species would correspond to covalent adducts of proteins with the bait ApoA-1.
The strongly encouraging results obtained demonstrate: 1) the exposure of intact cells to UV-laser induces the formation of covalent adduct inside the cell; 2) the phenomena can be modulated since the responses are function of the time of the exposure, the intensity of the laser beam, the frequency of the impulse.
These positive preliminary results will lead to further development of the project to be carried out as follows: 1) Human cells expressing the apolipoprotein A-I (ApoA-I) will be irradiated by laser beam; 2) the cells are lysed; 3) proteins covalently cross-linked to the bait protein will be co-purified by immunoaffinity chromatography using an anti-human ApoA-I polyclonal antibody linked to a resin; 4) the proteins eluted from the resin will be then identified by ordinary proteomics strategies.

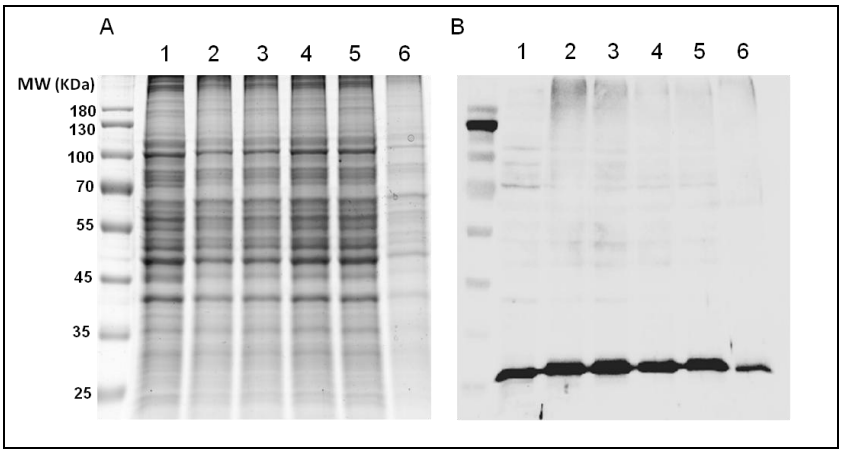


Fig.IV.15. SDS-PAGE fractionation and Western blotting analysis of Hep 75 cells irradiated under different conditions. (A) Electrophoretic separation of the UV-cross-linked Hep75 cells transfected with Apo-A1 protein. Not irradiated human Hep75 cell extracts were used as control (lane 1). UV-laser irradiated Hep75 cells (buffer) after exposure for 2 min to the laser beam (E=110 μJ, v=0.2 KHz) (lane 2); exposure for 12 sec (E=110 μJ, v=2 KHz) (lane 3); exposure for 12 sec (E=55 μJ, v=2 KHz) (lane 4); exposure for 2 min (E=55 μJ, v=0.2 KHz) (lane 5); exposure for 1 min (E=110 μJ, v=2 KHz) (lane 6). (B) Western blotting. Aliquots were subjected to SDS-PAGE and revealed with anti-Apo and peroxidise conjugated secondary antibodies. Molecular weight markers are also shown.

IV.7. Discussion

The development of novel strategies and methodologies that could help in depicting the complex traffic map of the rapidly changing sets of protein interactions, what has been defined as the “the molecular sociology of the cell” [32], is still of extraordinary importance for understanding cellular processes.
Chemical cross-linking is potentially a powerful alternative to genetic analyses of protein-protein interactions, such as the two hybrid system, and complement pull-down experiments. Despite the apparent straightforwardness of the cross-linking approach, cross-linking has always been a trial and error process for a particular protein or protein complex and identification of cross-linking products can be hampered by the complexity of the cross-linking reaction mixture. This is due largely

to the propensity of common bivalent cross-linking reagents to provide false positives and negatives. Clearly, false positives cause serious complications in building a model for a complex of unknown architecture from cross-linking data. On the other hand, the relatively modest reactivity of commonly used electrophilic moieties in cross-linkers (a necessity because the chemistry is carried out in aqueous solution) means that many interactions are not detected by common cross-linkers.
Moreover, the use of cross-linking to study multi-protein complexes or to search for binding partners has been limited by the relative slow and often inefficient chemistry employed by most classical cross-linkers.
The purpose of this study was to investigate the feasibility of exploiting UV-laser induced cross-linking as a tool for the analysis of protein-protein interactions on an extremely rapid time scale and with a good efficiency. If cross-linking was very rapid, the strategy might be useful for probing the dynamics of transient protein-protein interactions in a particular complex. Moreover, a light triggered reaction would be useful for probing protein-protein interactions in living cells or in crude extract, circumventing the problems of false positive detection due to unspecific artefacts. If successful, UV-induced laser cross-linking would result in a key strategy that will allow to capture and identify weakly interacting proteins *in vivo*.
Photochemical cross-linking of proteins to the nucleic acid has been applied to study a variety of nucleic acid binding proteins, but this methodology has not been ever used so far to analyse protein-protein interaction. Irradiation with UV light of wavelength near 267 nm can produce a “zero-length” covalent bond between contact points inside protein complexes and it is believed to produce less perturbation to structure than chemical cross-links. Moreover protein integrity is crucial for analysis of cross-linked products. Irradiation with a nanosecond laser was shown to degrade up to 25% of proteins *in vitro* [15]; however, a pico-second laser did not degrade proteins in vitro at three to four times higher doses [9]. Lejnine and co-workers [33] demonstrated that almost no degradation of proteins has been found after DNA-protein irradiation.
Moreover, it was not considered in any of the studies so far carried out investigating the laser UV induced DNA-protein cross-linking, the hypothesis that the high molecular aggregation could originate from reactions “born” on the protein side rather than on nucleic acid one. In all the papers investigating DNA-protein UV cross-linking, excitation upon absorption of UV light is always ascribed to nucleic acid bases [34,35].
The start up of this project was to test if 2 kHz, 267 nm pulsed laser was able to induce protein-protein cross-link.
We firstly demonstrated that it is possible to induce a cross-link when peptides are exposed to UV-laser. We observed that a covalent bond is formed only in the presence of aromatic amino acids in the peptide sequence. In order to verify which amino acids are involved we have performed high resolution tandem spectra of the cross-linked products. The obtained results asses univocally that the amino acids involved in this reaction are only the aromatic ones. The next step was to define which reaction mechanism is activated by ultra short UV laser. Since the most likely reaction mechanism is a radicalic one, we have performed femtosecond laser irradiation in the presence of spin trap molecules. After irradiation we observed the formation of covalent adducts with spin trap molecules, demonstrating the radical nature of the reaction mechanism. The specific location of the adducts on the model peptides was determined by high resolution mass spectrometry, demonstrating that radicalic adducts occurs only on aromatic side chain.

Once assessed the possibility to induce cross-link upon exposure to laser UV in peptides, the simplest model available, we have applied the same methodology to a more complicate model, an oligomeric protein, the alcohol dehydrogenase from *Saccaromyces cerevisiae*.
In preliminary experiments we have observed, upon exposure to laser UV, the formation of high molecular aggregates that do not enter the gel or stay at the stacking interface of a discontinuous gel, as has also been observed in DNA-protein UV-laser cross-linking experiments [35].
Therefore, in order to optimize cross-linking conditions, we have varied different parameters such as protein concentration, exposure time, laser energy and, because of the radicalic nature of the cross-linking reaction, we have added ascorbic acid.
In several conditions, bands that could be assigned to dimeric form (apparent molecular weight of 70 kDa) of ADH induced by the formation of one or more events of cross-link was observed (Fig.IV.14 panel A). The portions of the protein involved in the cross-link were defined by a classical proteomics approach. By means of it a discrete number of couple of peptides has been considered involved in the covalent bond induced by the laser. However, the validation of these hypotheses needs to MS/MS information.
In this perspective, preliminary results were obtained with intact Hep75 human cells, recombinant for apolipoprotein A-I (ApoA-I), that is involved in the systemic amyloidoses. It is worth observing that exposure time as short as just 12 sec to a femtosecond laser irradiation are enough to “freeze” some ApoA-I complexes. These times are not yet compatible with complex formation kinetics or with transient interactions, but certainly photo cross-linking with UV laser delivers high-power monochromatic light in a rapid and well controlled manner. Moreover the exposure times are reduced compared to classical chemical cross-linking, which require incubation times not less than 30 min. The laser version of the UV light is not popular yet, but its obvious potential as a protein-protein cross-linking agent will hopefully stimulate many scientists to develop the use of this technique for various purposes.

IV.8. References

[1] Sinz A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Rev.* (2006);25(4):663-82.
[2] Singh P, Panchaud A, Goodlett DR. Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique. *Anal Chem*. (2010);82(7):2636-42.
[3] Panchaud A, Singh P, Shaffer SA, Goodlett DR. xComb: a cross-linked peptide database approach to protein-protein interaction analysis. *J Proteome Res*. (2010);9(5):2508-15.
[4] Tang X, Bruce JE. A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. *Mol Biosyst*. (2010);6(6):939-47.
[5] Granneman S, Kudla G, Petfalski E, Tollervey D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A*. (2009);106(24):9613-8.
[6] Russmann C, Fix M, Naumer C, Herrmann T, Schmitt J, Stollhof J, Beigang R, Beato M. Crosslinking of progesterone receptor to DNA using tuneable nanosecond,

picosecond and femtosecond UV laser pulses. *Nucleic Acids Research*. (1997);25(12):2478-84.
[7] Russmann C, Stollhof J, Weiss C, Beigang R, Beato M. Two wavelength femtosecond laser induced DNA-protein crosslinking. *Nucleic Acids Res.* (1998);26(17):3967-70.
[8] Angelov D, Charra M, Muller CW, CIDet J, Dimitrov S. Solution study of the NF-kB p50-DNA complex by UV laser protein-DNA cross-linking. *J. Photochem. Photobiol.* (2003);77:592-6.
[9] Angelov D, Stefanovsky VY, Dimitrov SI, Russanova VR, Keskinova E, Pashev IG. Protein-DNA cross-linking in reconstituted nucleohistone, nuclei and whole cells by picoseconds UV laser irradiation. *Nucleic Acids Res.* (1988);16:4525-38.
[10] Budowsky EI, Axentyeva MS, Abdurashidova GG, Simukova NA, Rubin LB. Induction of polynucleotideprotein cross-linkages by ultraviolet irradiation. *Eur. J. Biochem.* (1986);159:95-101.
[11] Dobrov EN, Arbieva ZK, Timofeeva EK, Esenaliev RO, Oraevsky AA and Nikogosyan DN. UV laser induced RNA-protein crosslinks and RNA chain breaks in tobacco mosaic virus RNA *in situ*. *J. Photochem. Photobiol.* (1989);49:595-8.
[12] Pashev IG, Dimitrov SI, Angelov D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci*. (1991);16:323-6.
[13] Nikogosyan DN. Two-quantum UV photochemistry of nucleic acids: comparison with conventional low-intensity UV photochemistry and radiation chemistry. *Int J Radiat Biol.* (1990);57(2):233-99. Review.
[14] Buckle M, Geiselmann J, Kolb A, Buc H. Protein-DNA cross-linking at the lac promoter. *Nucleic Acids Res.* (1991);19(4):833-40.
[15] Ho DT, Sauvé DM, Roberge M. Detection and isolation of DNA-binding proteins using single-pulse ultraviolet laser crosslinking. *Anal Biochem.* (1994);218(2):248-54.
[16] Dimitrov SI, Moss T. UV laser-induced protein-DNA crosslinking. *Methods Mol Biol.* (1994);30:227-36.
[17] Hockensmith JW, Kubasek WL, Vorachek WR, Evertsz EM, von Hippel PH. Laser cross-linking of protein-nucleic acid complexes. *Methods Enzymol.* (1991);208:211-36.
[18] Dobrov EN, Arbieva ZKh, Timofeeva EK, Esenaliev RO, Oraevsky AA, Nikogosyan DN. UV laser induced RNA-protein crosslinks and RNA chain breaks in tobacco mosaic virus RNA in situ. *Photochem Photobiol.* (1989);49(5):595-8.
[19] Stefanovsky VYu, Dimitrov SI, Russanova VR, Angelov D, Pashev IG. Laser-induced crosslinking of histones to DNA in chromatin and core particles: implications in studying histone-DNA interactions. *Nucleic Acids Res*. (1989);17(23):10069-81.
[20] Davies MJ. Recent developments in EPR spin trapping. In: Gilbert, B. C.; Davies, M. J.; Murphy, D. M., eds. Electron spin resonance. Cambridge: Royal Society of Chemistry (2002);18:47-73.
[21] Mason RP. Using anti-5,5-dimethyl-1-pyrroline N-oxide (anti-DMPO) to detect protein radicals in time and space with immuno-spin trapping. *Free Radic Biol Med*. (2004);36(10):1214-23.
[22] Detweiler CD, Deterding LJ, Tomer KB, Chignell CF, Germolec D, Mason RP. Immunological identification of the heart myoglobin radical formed by hydrogen peroxide. *Free Radic Biol Med.* (2002);33(3):364-9.
[23] Deterding LJ, Ramirez DC, Dubin JR, Mason RP, Tomer KB. Identification of free radicals on hemoglobin from its self-peroxidation using mass spectrometry and immuno-spin trapping: observation of a histidinyl radical. *J Biol Chem.* (2004);279(12):11600-7.
[24] Detweiler CD, Lardinois OM, Deterding LJ, de Montellano PR, Tomer KB, Mason RP. Identification of the myoglobin tyrosyl radical by immuno-spin trapping and its dimerization. *Free Radic Biol Med.* (2005)1;38(7):969-76.
[25] Ehrenshaft M, Mason RP. Protein radical formation on thyroid peroxidase during turnover as detected by immuno-spin trapping. *Free Radic Biol Med.* (2006);41(3):422-30.

[26] Zhang H, He S, Mauk AG. Radical formation at Tyr39 and Tyr153 following reaction of yeast cytochrome c peroxidase with hydrogen peroxide. *Biochemistry.* (2002);41(46):13507-13.
[27] Wright PJ, English AM. Scavenging with TEMPO* to identify peptide- and protein-based radicals by mass spectrometry: advantages of spin scavenging over spin trapping. *J Am Chem Soc.* (2003);125(28):8655-65.
[28] Deterding LJ, Bhattacharjee S, Ramirez DC, Mason RP, Tomer KB. Top-down and bottom-up mass spectrometric characterization of human myoglobin-centered free radicals induced by oxidative damage. *Anal Chem.* (2007);79(16):6236-48.
[29] Gomez-Mejiba SE, Zhai Z, Akram H, Deterding LJ, Hensley K, Smith N, Towner RA, Tomer KB, Mason RP, Ramirez DC. Immuno-spin trapping of protein and DNA radicals: "tagging" free radicals to locate and understand the redox process. *Free Radic Biol Med*. (2009);46(7):853-65.
[30] Obici L, Franceschini G, Calabresi L, Giorgetti S, Stoppini M, Merlini G, Bellotti V. Structure, function and amyloidogenic propensity of apolipoprotein A-I. *Amyloid* (2006);13:191-205.
[31] Joya T, Wanga J, Hahn A, Hegele RA. ApoAI related amyloidosis: a case report and literature review. *Clin Biochem* (2003);36:641-5.
[32] Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. *Nature.* (2007);450(7172):973-82. Review.
[33] Li L, Lejnine S, Makarov V, Langmore JP. In vitro and in vivo reconstitution and stability of vertebrate chromosome ends. *Nucleic Acids Res.* (1998);26(12):2908-16.
[34] Russmann C, Truss M, Fix A, Naumer C, Herrmann T, Schmitt J, Stollhof J, Beigang R, Beato M. Crosslinking of progesterone receptor to DNA using tuneable nanosecond, picosecond and femtosecond UV laser pulses. *Nucleic Acids Res.* (1997);25:2478-84.
[35] Fecko CJ, Munson KM, Saunders A, Sun G, Begley TP, Lis JT, Webb WW. Comparison of Femtosecond Laser and Continuous Wave UV Sources for Protein-Nucleic Acid Crosslinking. *Photochem Photobiol*. (2007);83(6):1394-404.

Table of abbreviations

ACN	acetonitrile
ADH	alcohol dehydrogenases
AMBIC	ammonium hydrogen carbonate
α-ciano	alpha-cyano-4-hydroxycinnamic acid
BS^3	bis(Sulfosuccinimidyl) suberate
CID	collision induced dissociation
DC	direct current
DTT	dithiothreitol
ECD	electron collision dissociation
ESI	electrospray Ionisation
ETD	electron transfer dissociation
FTICR	fourier transform ion cyclotron resonance
hDAAO	human d-amino acid oxidase
IAM	iodoacetamide
IT	ion trap
LC	liquid chromatography
LIT	linear ion trap
m/z	mass-to-charge ratio
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MS	mass spectrometry
NLS	N-lauryl sarcosine
NMR	Nuclear Magnetic Resonance
PTMs	post translational modifications
Q	quadrupole
QqQ	triple quadrupole
RF	radio frequency
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
SDS	Sodium Dodecyl Sulphate
TFA	Trifluoroacetic acid
TOF	Time Of Flight

Pubblications

1) Leo G, Bonaduce I, Andreotti A, Marino G, Pucci P, Colombini MP, Birolo L.
   Deamidation at Asparagine and Glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings.
   Submitted to Anal Chem.

2) Del Vecchio C, Lettera V, Leo G, Pezzella C, Piscitelli A, Birolo L, Sannia G.
   Pleurotus ostreatus classical breeding: an useful tool to improve and increase laccase production in dikaryotic strains.
   Work in preparation.

3) Lettera V, Piscitelli A, Leo G, Birolo L, Pezzella C, Sannia G.
   Identification of a new member of Pleurotus ostreatus laccase family from mature fruiting body.
   Fungal Biol. 2010 Sep;114(9):724-30. Epub 2010 Jun 17.

4) Izzo V, Leo G, Scognamiglio R, Troncone L, Birolo L, Di Donato A.
   PHK from phenol hydroxylase of Pseudomonas sp. OX1. Insight into the role of an accessory protein in bacterial multicomponent monooxygenases.
   Arch Biochem Biophys. 2010 Oct 21. [Epub ahead of print]

5) Pagnozzi D, Birolo L, Leo G, Contessi S, Lippe G, Pucci P, Mavelli I.
   Stoichiometry and topology of the complex of the endogenous ATP synthase inhibitor protein IF(1) with calmodulin.
   Biochemistry. 2010 Sep 7;49(35):7542-52.

6) Amoresano A, Di Costanzo A, Leo G, Di Cunto F, La Mantia G, Guerrini L, Calabrò V.
   Identification of DeltaNp63alpha protein interactions by mass spectrometry.
   J Proteome Res. 2010 Apr 5;9(4):2042-8.

7) Leo G, Cartechini L, Pucci P, Sgamellotti A, Marino G, Birolo L.
   Proteomic strategies for the identification of proteinaceous binders in paintings.
   Anal Bioanal Chem. 2009 Dec;395(7):2269-80. Epub 2009 Oct 6.

Leo G, Marino G, Birolo L. “Spettrometria di massa” chapter of “La diagnostica nei beni culturali. Moderni metodi di indagine” by Paolillo L, Giudicianni I. Ed. Loghia.

Congress comunications

1) *Analysis of the human D-amino acid oxidase (DAAO) - pLG72 complex involved in the onset of schizophrenia.*
   Leo G., Sacchi S., Pollegioni L., Birolo L., Pucci P.
   Proteine 2010, Parma, Italy

2) *Analytical integrated platform for wine characterisation.*
   Galano E., Leo G., Amoresano A.
   SCI 5-10 July 2009, Sorrento, Italy

3) *Proteomics for the study of casein degradation in the wall paitings of monumental cemetary in Pisa.*
   Leo G., Andreotti A., Colombini MP., Marino G., Pucci P., Birolo L.
   SCI 5-10 July 2009, Sorrento, Italy

4) *Proteomics for the study of casein degradation in the wall paitings of monumental cemetary in Pisa.*
   Leo G., Andreotti A., Colombini MP., Marino G., Pucci P., Birolo L.
   Itpa 22-26 June 2009, Università degli Studi Bicocca - Milano, Italy 4

5) *Proteomics in cultural heritage.*
   Leo G., Cartechini L., Sgamellotti A., Marino G., Birolo L.
   9th International Conference on Ancient DNA and Associated Biomolecules
   Pompei, 19-22 October 2008

6) *Proteomics in cultural heritage.*
   Leo G., Cartechini L., Sgamellotti A., Marino G., Birolo L.
   Itpa 2008, Fasano, Italy

7) *UV-LASER induced protein crosslinking as a novel tool for studying protein-protein interactions.*
   Leo G., Altucci C., Donnarumma D., Velotta R., Marino G., Birolo L.
   Itpa 2008, Fasano, Italy

8) *Highly efficient protein-protein crosslinking induced by fs-UV laser pulses.*
   Altucci C., Velotta R., Esposito R., Birolo L., Leo G. and Marino G.
   Giornate polo delle scienze 20-21 September 2007, Naples, Italy

Visiting Appointment

Agosto-Ottobre 2010

Prof. Costello's laboratory. Mass Spectrometry Resource, NIH-Supported Resource Center, School of Medicine, Boston University. Boston, MA.

Acknowledgements

I wish to express my gratitude to my supervisor Leila Birolo, for her critical presence and constant support and for her passion in the research.
I would like to thanks all the people that have contributed to the realization of my PhD work. So I thank Prof. Di Donato and Dr. Viviana Izzo for the study concerning the Phenolo Hydroxylase complex, Prof. Polleggioni and Dr. Silvia Sacchi to introduce me in the fascinating world of Schizophrenia, Prof. Raffaele Velotta and Dr. Carlo Altucci with whom I have met a new field, the UV laser cross-linking and Prof. Piccoli for her contribution in this work.
I want also to acknowledge Dr. Angela Amoresano with whom I shared a side of my scientific work and (in alphabetic order!) Prof. Marino, Prof. Pucci, Prof. Sannia for their suggestions and scientific conversations.
A special thank goes to Prof. Cathy Costello for giving me the possibility to spend part of my research activity in her lab in Boston and to Sandrine for her patience and her precious help.
Last but not least, a special thank goes to all the people with whom I shared these last three years and more in the lab, colleagues, students, but, above all, friends. I am really thankful to Annalisa Leone who was always ready to help me, reminding all my duties.
Finally, I must specially thank all the people who have been present along these years.

All the best and take care!

Deamidation at Asparagine and Glutamine as a major modification upon deterioration/aging of proteinaceous binders in mural paintings.

Gabriella Leo[1], Ilaria Bonaduce[2], Alessia Andreotti[2], Gennaro Marino[1], Pietro Pucci[1], Maria Perla Colombini[2], Leila Birolo[1,*]

[1]Dipartimento di Chimica Organica e Biochimica, Università di Napoli "Federico II", Napoli, Italy.

[2]Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Italy

*To whom correspondence should be addressed. E-mail: birolo@unina.it

Abstract

Proteomic strategies are herein proved to be a complementary approach to the well established amino acid composition analysis for the characterization of the aging and deterioration phenomena occurring to proteinaceous materials in works-of-art.

Amino acid analyses on several samples demonstrated that proteins in the frescoes from the Camposanto Monumentale in Pisa are deteriorated as revealed by the decrease in Met, Lys and Tyr content and by the presence in all the samples of amino malonic acid as the results of Ser, Phe and Cys oxidation. Proteomic analysis identified deamidation at Asn and Gln as a further major event occurred.

This work paves the way to the exploitation of proteomic strategy for the investigation of the molecular effects of aging and deterioration in historical objects. Results show that proteomic searches for deamidation by LC-MSMS could constitute a routine analysis for paintings or any artistic and historic objects where proteins are presents. Two peptides, namely peptide 106-YLGYLEQLLR-115 from alpha S1 casein (P02662) and peptide 189-FALPQYLK-196 from alpha-S2 casein (P02663), were identified after a careful inspection of several candidates, to be used as molecular markers when casein is present.

Introduction

Safeguarding our Cultural Heritage is an important duty to ensure that future generations will have the opportunity to appreciate art masterpieces. Conservation must therefore be seen as one of the most exciting new fields and science can provide tools for the chemical, physical, and structural characterisation of the materials used to create a work-of-art, and on the basis of this knowledge, ongoing degradation processes can be highlighted.

Among the materials composing the art object, organic materials used as binding media or protective coatings are the most subject to degradation phenomena or ageing. Traditionally, a number of naturally occurring substances have been used in artworks given their ability to form homogeneous and flexible thin films when mixed with pigments and colouring materials on the appropriate supports[1]. Inevitably, these materials degrade with age, with the result that some features of the artist's work are more difficult to appreciate. Identification of the organic binders is of primary importance, and characterising degradation products of the organic material and identifying specific markers which describe the degradation and aging processes undergone the work-of-art is fundamental to better understand the history of a painting and to design appropriate restoration processes. The knowledge of original materials present in a work of art can in fact influence its cleaning, treatment and storage, and may also assist in the attribution of a painting to a particular workshop or artist.

The physico-chemical analysis of paintings is particularly challenging due to the vast range of inorganic and organic materials which often constitute a work-of-art: canvas fibres, wooden supports and inorganic fillers, mixtures of binding media and pigments, metal decorations, varnishes, modern polymer treatments, and protective layers. The most common and often reliable analytical approaches for the analysis of organic painting materials are based on GC-MS and Py-GC-MS[2-4] analyses. However, although the paint media can be often reliably identified, changes in the composition, degradation phenomena, formation of minor components, and molecular

modifications that have occurred cannot be completely understood by using only these approaches. The application of proteomics strategy to the characterization of proteinaceous binders in paintings has been recently reported in the literature[5-8]. So far, these studies were dedicated to the identification of proteinaceous paint media, rather than understanding modifications and degradation phenomena of the binders themselves.

Very little is known about ageing and modifications induced in the proteinaceous binders. Proteinaceous materials loose water as soon as they are applied as a thin film in a paint layer. The most likely modification event of proteinaceous material is then oxidation. Aminomalonic aldehyde and acid are considered oxidation products of Ser, Phe, and Cys[9]. The oxidation of other amino acids was observed by fluorescence and Raman spectroscopy, suggesting that photooxidation of paint films can produce several fluorescent compounds, such as dityrosine, N formil kyneurine (NFK) and 3,4-dihydroxy-phenyl alanine (DOPA), also depending on the simultaneous occurrence of pigments[10]. Moreover, it has been shown that the solubility of proteinaceous materials decreases with time, and that some cations can give rise to analytical interferences, suggesting the formation of strong complexes with proteins[4].

This paper presents some significant results obtained from the systematic characterization of proteinaceous binders aimed at identifying molecular markers to be used in detecting aged and degraded proteinaceous binders by LC-MSMS. LC-MSMS analysis has the potential to reveal chemical modifications undergone by proteinaceous materials in a painting at a molecular level, which cannot be achieved with any other analytical technique.

In this respect deamidation of asparagine and glutamine was detected as a major modification undergone by milk and animal glue in a sample collected from the Camposanto Monumentale in Pisa. Deamidation is a spontaneous nonenzymatic post-synthetic modification of proteins. It plays an important role in protein degradation and is postulated to function as a timer in aging[11-14]. Deamidation introduces negative charges and results in a +0.984-Da mass shift, which can be

detected by MS techniques[15-18], and MSMS methods can also provide very specific information on deamidation sites.

In the contest of the samples herein characterized, deamidation could be related to the reactions taking place between the proteinaceous binder and the other materials used in the past restoration as well as to the ageing of the proteinaceous binders.

EXPERIMENTAL SECTION

Materials. Ammonium hydrogen carbonate (AMBIC), Ethylenediaminetetraacetic acid (EDTA) and iodoacetamide were purchased from Fluka; Tri(hydroxymethyl)aminomethane (TRIS), urea, dithiothreitol (DTT), TPCK-treated trypsin were from Sigma; Formic acid from Baker and Acetonitrile (ACN) were purchased from Romil. Deionised water was obtained with a Millipore cartridge equipment.

Painting samples. Replicas were prepared using whole milk and fresh whole egg, and rabbit skin glue. Animal glue was dissolved in warm water; whole egg was slightly beaten to obtain a fluid homogeneous medium; whole milk was used without any treatment. The pigments (Prussian blue $Fe_4[Fe(CN)_6]_3$, verdigris $Cu(CH_3COO)_2 2Cu(OH)_2$, minium Pb_3O_4, cadmium yellow CdS+ZnO] were mixed with the fluid binder in ratios ranging approximately between 1 part pigment and 2-3 parts fluid binder. The paint was then applied with a brush on tiles covered by a layer of plaster, which were prepared in the summer 2003. The paint replicas were prepared in spring 2008, and stored at ambient temperature in the darkness since then.

All samples of painting layers were available as few milligrams fragments (1-3 mg).

Sample treatment. Samples and blank reference sample were digested in heterogeneous phase incubating for 4 or 16 h, approximately 1-3 mg of solid sample, in ammonium bicarbonate 10 mM pH 7.5, containing trypsin 0.16 µM. The supernatant was then recovered by centrifugation, dried in vacuum, resuspended in 10 µl in 0.1% formic acid, and the peptide mixture analysed by LC-MSMS.

LC-MSMS analysis. The peptide mixtures were analysed using a CHIP MS 6520 QTOF equipped with a capillary 1200 HPLC system and a chip cube (Agilent Technologies, Palo Alto, Ca) as previously reported[7]. Raw data from nano LC–MSMS analyses were employed to query non-

redundant protein databases (UniprotSprot with taxonomy restriction to *Chordata*) using in house MASCOT software (Matrix Science, Boston, USA).

RESULTS

The Camposanto Monumentale, a 13th century cemetery located alongside the Leaning Tower in Pisa (Italy), was decorated with important frescoes painted by Benozzo Gozzoli, Taddeo Gaddi, Spinello Aretino and Buonamico Buffalmacco, amongst others. During the Second World War an incendiary bomb exploded in the cemetery, burning the wood and melting the lead of the roof. During the restoration following the Second World War, the paintings (i.e., paint layer and plaster) were detached from the wall using animal glue, and was subsequently glued onto another canvas using non water soluble glues. At this stage, in most cases casein was used, but its quality was very poor and, with the passing of time it has slowly lost its adhesive properties[19]. In addition to prevent it from rotting quickly during the restorations, formaldehyde was often added: unfortunately in some cases this gave rise to cross-linking reactions, transforming all the paint layers into a unique insoluble and waterproof block[20]. Slaked lime was generally mixed with the casein, in order to increase the strength of the glue after carbonation of the calcium hydroxide, but it has since been proven that the carbonation was still not at all complete even 20-30 years later.

GC-MS analyses of the paint samples. In several samples from the restoration campaign, the proteinaceous content was analysed by GC-MS procedure based on the acidic hydrolysis of the samples assisted by microwave, followed by derivatisation with a silylating agent[3]. Purification from inorganic materials was also performed when necessary[3]. This procedure permits to evaluate the quantitative amino acidic content of 14 amino acids. Animal glue was used to detach the paintings and casein was used as glue to the new support. As a result, one of these materials, or both of them, are always found in the samples collected from these paintings. Animal glue can be easily identified due to the presence of Hyp. Moreover it is characterised by high contents of Gly. Casein is characterised by a high content of Glu and a relatively high content of Pro. To identify a

proteinaceous binder by GC-MS, the relative amino acid percentage content of a sample is compared to a dataset of reference samples by means of Principal Component Analysis (PCA).

Table 1 reports an example of the aminoacidic compositions observed in some samples of the mural paintings compared to the average profiles of reference samples of animal glue, egg and casein obtained with the same analytical procedure.

Samples GUsp1, GUsp4 and Ab show the occurrence of animal glue, as indicated by the presence of Hyp. By processing the samples with PCA (Score Plot in Figure 1) it is possible to evaluate that all samples contain casein as well, as indicated by their position in the score plot. Finally it is important to stress that some samples do not show any Met (GUsp1, GUsp4 and Ab), some any Tyr (GUsp1, CSG16, A6S1a and CSG-C bis), and some others any Lys (CSG16, A6S1a and CSG-C bis). Ser and Glu are also extremely low in CSG-C bis. Finally in all samples amino malonic acid, which is considered to be the degradation product of Ser, Phe, and Cys[21], was identified

The situation shown here with few examples is typical of the samples collected from the mural paintings of the Camposanto Monumentale in the course of the years. Although in most cases identification of the proteinaceous binder is still possible, most part of the samples analysed show modified profiles. But there is no information about how and where these modifications occur in the protein sequence and which can thus be the consequences from point of view of the stability of the materials and the conservation of the work of art itself.

Proteins Identification in painting samples by proteomic strategies. The identification of the proteinaceous material in the painting sample from the Camposanto Monumentale of Pisa, was carried out following the minimally invasive proteomic analytical procedure previously described[7] and reported in the experimental section.

A preliminary search in the UniprotSprot database for protein identification was carried out in the MASCOT software without the insertion of any variable chemical modification but oxidation of methionine and the possible formation of pyroglutammic acid from glutamine residues at the N-

terminal position of peptides. The identification of 6 proteins form bovine milk and two bovine collagens confirmed the use of casein and bovine animal glue as natural adhesives[20, 22].

Because of the presence of collagens, hydroxylation on proline and lysine residues, a post–translational modification extensively found in collagen chains, were subsequently inserted as variable modifications in the database search. This resulted in a much higher confidence in the identification of proteins from the animal glue component. As an example, addition of hydroxylations as variable modifications increased the identification score from 204 to 780 for collagen alpha-1 (I) chain (P02453, entry in the UniprotSprot Database,) and the number of identified peptides increased from 5 to 30. In the Supplementary materials, Table S1 reports the results on the sample from the Camposanto Monumentale of Pisa, introducing oxidation of methionine, hydroxylation of proline and lysine, pyro-Glu formation from Gln at the N-terminus of peptides as variable modifications.

Similar analyses were carried out on reference materials (i.e. dried skimmed milk, model samples prepared with caseins and animal glue as binders), and database searches were performed in the same conditions as above. Proteins identified in the different samples are quite always the same. In all the samples that contain milk, bovine beta, alpha S1, alpha S2, and kappa caseins were identified as well as beta lactoglobulin with high confidence. In all the samples containing bovine glue the presence of collagen alpha 1 (I) chain and collagen alpha 2 (I) chain was confidently assessed. These results are fully in agreement with the screening and statistical analysis carried out by Fremout and colleagues on the most common identified proteins in paint replicas[8].

Identification of Modification Events in Proteins in Painting Samples. Database searches were successively carried out, on the sample form the Camposanto Monumentale in comparison to reference materials, with the insertion of several additional variable chemical modifications, one or two at a time, because of search algorithm optimization. Kynurenine, formyl-kynurenine, hydroxykynurenine, oxidation of His, Trp, Arg, Lys, nitration of Tyr, beta elimination of Ser and

Thr, oxidation of Cys to cysteic acid, and formation of amino malonic acid from Ser, Phe and Cys were all tested as possible modifications of the proteins present in the samples Best search results were obtained by trial and error. As expected, and in agreement with the results obtained by amino acid analysis, formation of amino malonic acid from Ser (predominantly), Phe, and Cys was observed in several peptides (data not shown). On the contrary, oxidation of sulphur containing aminoacids, or modifications of tryptophan was not significantly detected.

When deamidation was considered as variable modification, several of the Gln and Asn containing peptides that had been identified in the sample from the Camposanto Monumentale, were also found as deamidated peptides (Table 2). As an example, 6 of the 10 peptides belonging to the bovine beta casein were also detected in the deamidated form. Similarly, 9 of the 15 peptides mapping the Alpha-S1-casein were also found as deamidated peptides. We were therefore prompted to investigate the occurrence of deamidation as a marker for deterioration and aging of the materials. It is worth mentioning that deamidation escapes detection in amino acids analysis because the extreme acidic conditions in the hydrolysis of the proteins systematically transform all the Gln and Asn into Glu and Asp, respectively.

Systematic database searches were therefore eventually repeated including deamidation on glutamine and asparagine as variable modification and Table 2 reports the number of deamidated peptides observed in the single proteins identified in the sample from the Camposanto Monumentale, in a sample of skimmed milk and in the test models.

Assessment of deamidation. Deamidation introduces a small (+0.984 Da) mass shift that overlaps with isotopic pattern. Moreover, non-deamidated and deamidated peptides behave quite similarly during reverse-phase chromatography, with deamidated peptides eluting at moderately higher concentration of acetonitrile than non-deamidated ones.

Peptide fragmentation does facilitate detection of deamidation and provide very specific information on deamidation sites. We manually inspected several MS and MSMS spectra originated by deamidated peptides as indicated by Mascot software, in order to confirm software assignments.

However, as has been shown by a large number of authors (e.g., ref. 23), many post-synthetic modifications including deamidation, can occur *in vitro* during proteomics sample handling.

Moreover, as Li and coworkers demonstrated[30], artificial deamidation occurring during the tryptic digestion can be reduced with faster trypsin digestion: digestion carried out for up to 4 h generally would not introduce additional detectable deamidations, even for rapidly deamidating peptides.

To reduce the incidence of artifactual deamidation, we therefore subjected the samples to a short trypsin digestion (~4 h) as a systematic strategy.

Despite the shorter hydrolysis time, proteins could be confidently identified in all the samples (Table 2).

Identification of peptides to be considered as aging/deterioration signatures. Deamidation was found widespread over most of the identified proteins, both from the old sample and the reference materials, (Table 2). A systematic analysis was carried out to individuate peptides that could be sorted out as aging/deterioration signatures for further investigation.

We designed specific criteria for the identification of candidate peptides as follows: 1) Peptides should have been identified in all the samples analysed with good scores; 2) fragmentation spectra should be of good quality in all the samples tested; 3) when deamidation was detected, fragmentation spectra of the deamidated peptide should be of good quality as well.

The limited variability of the proteins identified in the different samples (Table 2) (i.e., as highlighted above, the same proteins were identified in the sample from the Camposanto Monumentale of Pisa as well as in the reference samples) was instrumental in our search for biomarkers of the material deterioration. We could screen a large panel of peptides common to all

samples, with a wide range of physico-chemical characteristics, looking for candidates that do satisfy the prerequisites.

Table 3 lists all the glutamine or asparagine containing peptides that had been detected in the sample from the Camposanto Monumentale of Pisa, and, for each of them, the occurrence of deamidation is indicated and compared to the other samples. Table 3A and Table 3B list the peptides of the milk derived proteins and in the animal glue component, respectively.

Several deamidated peptides detected in the sample from the Camposanto Monumentale of Pisa were not deamidated either in the model samples or in the skimmed milk sample, (Table 3A and Table 3B). Moreover, peptides with multiple deamidations were detected in the sample from the Camposanto Monumentale of Pisa, whereas only singly or not deamidated in model samples or in the skimmed milk sample.

A preliminary analysis of the peptides identified in the different samples clearly indicates that peptides from caseins are better candidates than collagen fragments. The co-presence of a variable extent of hydroxylations on almost all of the collagen peptides enormously increases the heterogeneity of these fragments decreasing the relative abundance of the single species and thus making the confident assignment of deamidation of these peptides more challenging.

Our survey produced at least two possible candidates, whose mass and fragmentation spectra for both the normal and deamidated peptides are reported in Fig. 2 and 3, respectively.

Peptide 106-YLGYLEQLLR-115 from alpha S1 casein (P02662) was considered a good candidate for a number of considerations. This peptide is present in all the casein containing samples, it contains only one deamidation site, it was found to be deamidated only in the sample from the Camposanto Monumentale of Pisa, and deamidation could be easily visualised in the corresponding fragmentation spectra.

Figure 2 (panel A) shows the MSMS spectra of the doubly charged ion at 634.36 m/z of the YLGYLEQLLR peptide from alpha S1 casein (P02662), in the sample from Camposanto Monumentale of Pisa. Panel B shows the MSMS of the doubly charged ion at 634.85 m/z

corresponding to the deamidated form of the same peptide. Deamidation at Gln113 can be readily assessed since all the signals from the *y* series containing the Gln residue are shifted by one unit in the fragmentation spectra of the deamidated form, with deviation of the experimental data from the expected values within the 1-21 ppm range (Table S2, Supplementary materials).

The second candidate peptide was defined as the 189-FALPQYLK-196 peptide from alpha-S2 casein (P02663) on the basis of the same considerations as discussed above. Deamidation could be readily assessed since all the signals from the *y* series containing the Gln residue are shifted by one unit in the fragmentation spectra (Figure S1, supplementary materials), with deviation of the experimental data from the expected values within the 1-30 ppm range (Table S2).

DISCUSSION

Only in recent years proteomic approaches have been proposed as valuable tool for the identification of the proteinaceous constituents of artistic and historic objects. Herein, these procedures proved to be extremely promising also for the characterization of the natural and artificial aging and deterioration products in the proteins in masterpieces.

We worked on the 14th century frescoes from the Camposanto Monumentale in Pisa, that were detached in 1945 from the wall and relocated on an asbestos cement supports, using a glue based on a mixture of casein, animal glue and calcium hydroxide. Amino acid analyses on several samples demonstrated that proteins in the frescoes were deteriorated as revealed by the decrease in Met, Lys and Tyr and by the presence of amino malonic acid as the results of oxidation of Ser, Phe and Cys in all the samples.

Proteomic analyses confirmed the presence of amino malonic acid, and identified deamidation at Asn and Gln of proteinaceous binders as a major deterioration event. Deamidation is a spontaneous nonenzymatic post-synthetic modification of proteins and cannot be assessed by aminoacidic analysis because of the extreme conditions of the acidic hydrolysis that convert all the amide moieties into the corresponding acids. It plays an important role in protein degradation and is postulated to function as a timer in aging[11-14].

Deamidation at Asn residues takes place much more rapidly than at Gln (up to ten times faster), and the reaction mechanisms of nonenzymatic deamidation have been extensively studied. Asn deamidation proceeds through the formation of a succinimide ring, which in turn opens up to give a mixture of aspartic and iso-aspartic residue[11]. Gln deamidation, similarly, can proceed through formation of a glutarimide moiety, but, since this reaction is relatively slow, direct hydrolysis also contribute significantly[11].

The systematic analysis of the samples herein reported indicate extensive deamidation occurring on most of the peptides identified, suggesting that the miniature molecular clock, as Robinson and

Robinson defined any amide residue present in peptides or proteins[11], might well be used as molecular marker in artworks.

However, many factors can influence deamidation, such as protein sequence[14, 24, 25], secondary structure[26], local three-dimensional structure[27], pH, temperature, ionic strength, buffer ions, turnover of the protein, and other solution properties[28, 29]. Moreover, in this respect, it must be kept in mind that the sample taken into consideration was collected from a work of art. Milk is an emulsion, and is already in the liquid form. To use casein, which is sold as a powder, as an art material, a fluid must be obtained, which can be achieved by using an alkaline solution as a solvent. In the case of the Camposanto Monumentale, it is known that slaked lime was added to casein, to solubilise it and to obtain a stronger glue, after the carbonatation of the lime took place. Slaked lime is extremely alkaline, and it must be expected that this has reacted with casein during years, although a reduced reaction rate must be expected when the glue was dried and water thus evaporated.

It is therefore clear that the modifications detected here have to be considered as the results of the work of art preparation, the treatment itself (i.e. the presence of slaked lime) and the actual aging, without any possibility, at this stage, of distinguishing the different phenomena.

Our data indicate that, whenever caseins are present, several peptides can constitute the molecular signatures of degradation, and we identified, after a careful inspection of several candidates, two peptides that can be used as markers, namely peptide 106-YLGYLEQLLR-115 from alpha S1 casein (P02662) and peptide 189-FALPQYLK-196 from alpha-S2 casein (P02663).

Interestingly, both the selected peptides present Gln as the deamidation site. We suggest that Gln deamidation, intrinsically slower than deamidation at Asn site, might reveal to be more useful in the quantitative perspective on historical object: Asn deamidation might occur too fast to be used for the actual evaluation of the aging of the object itself.

Awareness and measurement of deamidation could constitute a routine analysis for paintings or any artistic and historic objects where proteins are presents. However, a contribution in the level of

deamidation due to artwork preparation itself, and deamidation artifacts produced during the analysis must be carefully guarded against and considered. Future works will be devoted to the quantitative analysis of deamidation as function of aging and to deeply investigate and possibly separate the contributions of actual aging and/or the preparation and treatment of the work-of art.

In conclusion, our data identified deamidation as an extensive deterioration process occurring in pictorial objects, and pave the way to the identification of molecular signatures that can be useful in characterizing of artworks that contain proteins and, suggest to further exploit proteomic strategy for the investigation of aging molecular effects.

Supplementary materials

The full list of the peptides and the corresponding proteins identified in the sample of the Camposanto Monumentale of Pisa is available free of charge via the Internet at http://XXXXXX.

REFERENCES

(1) Mills, J. S.; White, R. *The Organic Chemistry of the Museum Objects*, 2nd ed.; Butterworth Heinemann Ltd: Oxford, **1994**.

(2) Colombini, M. P.; Andreotti, A.; Bonaduce, I.; Modugno, F.; Ribechini, E. *Acc. Chem. Res.* **2010**, *395*, 715-727

(3) Bonaduce, I.; Andreotti, A.; In *Organic Mass Spectrometry in Art and Archaeology*; Colombini, M. P.; Modugno, F., Eds.; John Wiley & Sons Ltd, Chichester, UK, **2009**; pp 303-326

(4) Andreotti, A.; Bonaduce, I.; Colombini, M.P.; Modugno, F.; Ribechini, E. In *New Trends in Analytical, Environmental and Cultural Heritage Chemistry*; Tassi, L.; Colombini, M.P., Eds.; Transworld Research Network: Kerala, **2008**, pp 389-423.

(5) Tokarski, C.; Martin, E.; Rolando, C.; Cren-Olive, C. *Anal. Chem.*, **2006**, *78*, 1494-1502.

(6) Domènech-Carbò, M. T. *Anal. Chim. Acta*, **2008**, *621*, 109–139.

(7) Leo, G.; Cartechini, L.; Pucci, P.; Sgamellotti, A.; Marino, G.; Birolo, L. *Anal. Bioanal. Chem.* **2009**, *395*, 2269-2280.

(8) Fremout, W.; Dhaenens, M.; Saverwyns, S.; Sanyvova, J.; Vandenabeele, P.; Deforce, D.; Moens, L. *Anal. Chim. Acta*, **2010**, *658*, 156-162.

(9) Boon, J. J.; Peulvè S. L.; van der Brink, O. F.; Duursma, M. C.; Rainford, D. In *Proceedings of Early italian paintings techniques and analysis symposium*; Bakkenist, T.; Hoppenbrouwers, R.; Dubois, H. Eds. Limburg Conservation Institute, Maastricht **1996**, pp 35-56.

(10) Nevin, A.; Anglos, D.; Cather, S.; Burnstock A. *Appl. Phys. A: Mater. Sci. Process.* **2008**, *92*, 69-72

(11) Robinson, N. E.; Robinson, A. B. *Molecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins*; Althouse Press: Cave Junction, OR, **2004**.

(12) Robinson, N. E. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5283–5288.

(13) Robinson, N. E.; Robinson, A. B. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 944–949.

(14) Bischoff, R.; Kolbe, H. V. J. *J. Chromatogr. B Biomed. Appl.* **1994**, *662*, 261–278.

(15) Robinson, N. E.; Zabrouskov, V.; Zhang, J.; Lampi, K. J.; Robinson, A. B. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3535–3541.

(16) Robinson, N. E.; Lampi, K. J.; Speir, J. P.; Kruppa, G.; Easterling, M.; Robinson, A. B. *Mol. Vis.* **2006**, *12*, 704–711.

(17) Robinson, N. E.; Lampi, K. J.; McIver, R. T.; Williams, R. H.; Muster, W. C.; Kruppa, G.; Robinson, A. B. *Mol. Vis.* **2005**, *11*, 1211–1219.

(18) Schmid, D. G.; von der Mulbe, F.; Fleckenstein, B.; Weinschenk, T.; Jung, G. *Anal. Chem.* **2001**, *73*, 6008 – 6013.

(19) Colombini, M. P.; Fuoco, R.; Giacomelli, A.; Moscatello, B.; Baracchini, C.; Caponi, G. In *Scienza e Beni Culturali: "La pulitura delle superfici dell'architettura*", Padova: Libreria Progetto Ed., **1995**, pp 227-236.

(20) Bonaduce, I.; Colombini, M. P. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2523-2527.

(21) van den Brink, O. F.; Boon, J. J.; O'Connor, P. B.; Duursma, M. C.; Heeren, R. M. *J. Mass Spectrom.* **2001,** *36*, 479-492.

(22) Colombini, M. P.; Modugno, F.; Giacomelli, A. *J. Chromatogr. A* **1999,** *846*, 101-111.

(23) Xing, G.; Zhang, J.; Chen, Y.; Zhao, Y. *J. Proteome Res.* **2008,** *7*, 4603-8.

(24) Capasso, S. J. *Pept. Res.* **2000**, *55*, 224–229.

(25) Geiger, T.; Clarke, S. *J. Biol. Chem.* **1987**, *262*, 785–794.

(26) Xie, M. L.; Schowen, R. L. *J. Pharm. Sci.* **1999**, *88*, 8–13.

(27) Robinson, N. E.; Robinson, A. B. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4367–4372.

(28) Tylercross, R.; Schirch, V. *J. Biol. Chem.* **1991**, *266*, 22549–22556.

(29) Wright, H. T. *Crit. Rev. Biochem. Mol. Biol.* **1991**, *26*, 1–52.

(30) Li, X.; Cournoyer, J. J.; Lin, C.; O'Connor, P. B.; *J Am Soc Mass Spectrom.* **2008,** *19*, 855-864.

ACKNOWLEDGMENT.

This work was supported by Italian MIUR grants, PRIN 2008 (2008XXAMZT).

LEGENDS TO FIGURES

Figure 1: Score plot obtained by comparing the reference samples of animal glue, casein and egg with the samples from the Camposanto Monumentale.

Figure 2: MSMS spectra of the doubly charged ions at m/z 634.36 (A) and 634.85 (B) of the peptide 106-YLGYYLEQLLR-115 from bovine alpha S1 casein (P02662) and its deamidated form, respectively, identified in the analysis of the sample from Camposanto Monumentale of Pisa. The product ions are indicated with the observed masses.

Table 1. Aminoacidic composition in samples from the Monumental Cemetery.

Aminoacid / Sample	Ala	Gly	Val	Leu	Ile	Met	Ser	Pro	Phe	Asp	Glu	Lys	Hyp	Tyr
GUsp1	6.8	17.8	6.0	8.4	4.8	0.0	3.6	10.6	4.1	8.1	17.3	2.0	10.3	0.0
GUsp4	7.4	12.4	11.1	14.8	8.4	0.0	4.4	16.4	6.5	4.3	10.6	0.7	2.6	0.3
Ab	3.7	6.3	7.1	10.0	4.9	0.0	5.1	14.3	2.1	10.3	27.3	0.2	8.3	0.4
CSG16	6.6	14.5	6.8	8.9	5.1	2.4	2.9	12.6	2.4	9.0	28.8	0.0	0.0	0.0
A6S1a	5.2	6.0	8.1	13.8	7.9	4.8	3.6	6.5	6.4	9.1	28.5	0.0	0.0	0.0
CSG-C bis	8.2	5.1	15.5	18.3	12.7	7.5	1.7	14.0	10.5	3.8	1.8	0.0	0.0	0.0
Glue	14.3	23.1	5.0	6.0	2.4	1.3	5.1	12.6	3.1	5.3	7.8	3.7	9.2	1.2
Casein	3.7	1.9	6.6	11.5	7.0	3.5	7.1	11.7	6.9	8.8	25.7	2.0	0.0	3.6
Egg	8.6	4.7	10.1	14.7	8.6	6.0	7.6	4.4	8.7	11.8	10.3	3.7	0.0	0.8

Table 2. Identification of the proteins in painting samples by trypsin digestion in heterogeneous phase. Proteins were identified searching UniprotSprot database with MSMS Ion search Mascot software (Matrix Science). *Chordata* was used as taxonomy restriction, deamidation on Gln and Asn, oxidation on Met, hydroxylation on Lys and Pro as variable modifications. Only proteins identified with at least two peptides were considered as significative.

		Overnight trypsin digestion		**4 h trypsin digestion**	
SAMPLE	**Identified protein** (Accession number)	**Number of identified peptides**[b]	**Number of deamidated peptides**[c]	**Number of identified peptides**	**Number of deamidated peptides**
Sample from Cimitero Monumentale of Pisa [a]	Beta-casein1 (P02666)	10	6	8	6
	Alpha-S1-casein (P02662)	15	9	11	7
	Collagen alpha-2 (I) chain (P02465)	26	7	33	11
	Collagen alpha-1 (I) chain (P02453)	35	9	40	16
	Kappa-casein (P02668)	4	2	5	2
	Alpha-S2-casein (P02663)	11	7	9	6
	Beta-lactoglobulin (P02754)	7	0	6	0
	Lactadherin (Q95114)	3	0	-	-
Model on intonaco with casein	Beta-casein (P02666)	6	3	7	3
	Alpha-S1-casein (P02662)	8	4	9	3
	Beta-lactoglobulin (P02754)	4	1	6	1
	Alpha-S2-casein (P02663)	5	2	6	3
	Kappa-casein (P02668)	2	0	2	0
	Serum albumin (P02769)	4	0	2	0
	Polymeric immunoglobulin receptor (P81265)	2	0	-	-
Model on intonaco with animal glue	Collagen alpha-2 (I) chain (P02465)	26	12	38	13
	Collagen alpha-1 (I) chain (P02453)	34	15	38	17
	Collagen alpha-1(III) chain (P04258)	10	6	13	6
Skimmed milk	Alpha-S1-casein (P02662)	9	1	8	2
	Serum albumin (P02769)	13	0	6	0
	Beta-lactoglobulin (P02754)	7	2	9	-
	Kappa-casein (P02668)	3	1	4	0
	Beta-casein (P02666)	4	1	-	-
	Lactotransferrin (P24627)	4	0	3	0
	Alpha-S2-casein (P02663)	4	2	6	1
	Butyrophilin subfamily 1 member A1 (P18892)	3	0	1	-
	Glycosylation-dependent cell adhesion molecule I (P80195)	3	0	3	0
Model on glass with casein	Alpha-S1-casein (P02662)	16	7	9	2
	Beta-casein (P02666)	12	6	5	1
	Beta-lactoglobulin (P02754)	7	1	-	-
	Alpha-S2-casein (P02663)	14	4	7	2
	Kappa-casein precursor (P02668)	3	2	3	2
Model on glass with animal glue	Collagen alpha-2 (I) chainI (P02465)	53	27	36	12
	Collagen alpha-1 (III) chain (P04258)	27	14	11	5
	Collagen alpha-1 (I) chain (P02453)	49	27	48	18

[a] An extended version of the table, with the sequence of the identified peptides, is reported in the Supplementary material, S1

[b] The number of peptides was calculated regardless the presence of modifications, i.e., for this purpose, if a peptide was observed both as modified and not modified, herein it is considered as a single peptide.

[c] The number of deamidated peptides was herein calculated regardless the number of deamidations observed on the peptide, i.e., for this purpose, if a peptide was observed a singly or multiply deamidated, it is herein considered as a single modification.

Table 3. Peptides identified as deamidated in the sample from the Camposanto Monumentale of Pisa. Analysis of the peptides observed as deamidated in the proteins identified in the sample from the Camposanto Monumentale of Pisa digested with trypsin in heterogenous phase for 4 h (see Table 2 for the details of the protein identification). UniprotSprot Database search was carried out as with MSMS Ion search Mascot software (Matrix Science) and *Chordata* was used as taxonomy restriction. Detection of peptides with single, double and >2 deamidations are indicated with X. Table 3A reports the analysis for the casein containing samples and Table 3B for the animal glue containing samples.

A

Peptide (n° deamidation sites)	Sample from Cimitero Monumentale				Skimmed milk				Model on intonaco with casein				Model on glass with casein			
	No deam	+1 deam	+2 deam	>2 deam	No deam	+1 deam	+2 deam	>2 deam	No deam	+1 deam	+2 deam	>2 deam	No deam	+1 deam	+2 deam	>2 deam
Casein alpha-S1 (P02662)																
EKV**N**ELSK (**1**)	X	X			-	-			-	-			X	-		
VP**Q**LEIVP**N**SAEER (**2**)	X	X	-		X	-	-		X	-	-		-	-	-	
YLGYLE**Q**LLR (**1**)	X	X			X	-			X	-			X	-		
H**Q**GLP**Q**EVL**N**E**N**LLR (**4**)	X	X	X	-	X	X	-	-	X	X	-	-	X	-	-	-
HI**Q**KEDVPSER (**1**)	X	-			-	-			-	-			-	-		
YKVP**Q**LEIVP**N**SAEER (**2**)	X	X	-		X	-	-		-	-	-		-	-	-	
HPIKH**Q**GLP**Q**EVL**N**E**N**LLR (**4**)	X	X	-	-	X	X	-	-	-	-	-	-	-	-	-	-
Kappa-casein precursor (P02668)																
YIPI**Q**YVLSR (**1**)	X	X			X	-	-		X	-			X	X		
SPA**Q**IL**Q**W**Q**VLS**N**TVPAK (**4**)	X	X	X	X	X	-	-	-	X	-	-	-	X	X	-	-
Casein alpha-S2 (P02663)																
L**N**FLKK (**1**)	X	-			-	-			-	-			-	-		
FALP**Q**YLK (**1**)	X	X			X	-			X	-			X	-		
AMKPWI**Q**PK (**1**)	X	X			X	-			-	-			X	-		
NAVPITPTL**N**R (**2**)	X	X	X		X	X	-		X	X	-		X	X	-	
R**N**AVPITPTL**N**R (**2**)	X	X	X		X	-	-		X	X	-		X	X	-	
AL**N**EI**NQ**FY**Q**K (**4**)	X	X	-	-	X	-	-	-	X	-	-	-	-	-	-	-
Beta-casein precursor (P02666)																
VLPVP**Q**K (**1**)	X	X			-	-			X	-			X	-		
AVPYP**Q**R (**1**)	X	X			-	-			X	-			X	-		
VLPVP**Q**KAVPYP**Q**R (**2**)	X	X	-		-	-	-		X	-	-		X	-	-	
DMPI**Q**AFLLY**Q**EPVLGPVR (**2**)	X	X	X		-	-	-		X	X	-		X	X	-	
AVPYP**Q**RDMPI**Q**AFLLY**Q**EPVLGPVR (**3**)	-	X	X	-					-	-	-	-	-	-	-	-
F**Q**SEE**QQQ**TEDEL**Q**DK (**5**)	-	X	-	-					-	-	-	-	-	-	-	-
Beta-lactoglobulin (P02754)																
IDAL**N**E**N**K (**2**)	X	-	-		-	-	-		X	-	-		-	-	-	
VYVEELKPTPEGDLEILL**Q**K (**1**)	X	-			X	X			X	-			X	-		

B

Peptide (n° deamidation sites)	Sample from Cimitero Monumentale				Model on intonaco with animal glue				Model on glass with animal glue			
	No deam	+1 deam	+2 deam	>2 deam	No deam	+1 deam	+2 deam	>2 deam	No deam	+1 deam	+2 deam	>2 deam
Collagen alpha-1 (I) chain (P02453)												
GVVGLPG**Q**R (**1**)	X	X			X	X			X	X		
GV**Q**GPPGPAGPR (**1**)	X	X			X	X			X	X		
G**Q**AGVMGFPGPK (**1**)	X	-			X	X			X	X		
GPSGP**Q**GPSGPPGPK (**1**)	X	-			X	-			X	X		
DGL**N**GLPGPIGPPGPR (**1**)	-	X			-	X			-	X		
GA**N**GAPGIAGAPGFPGAR (**1**)	-	X			-	X			-	X		
VGPPGPSG**N**AGPPGPPGPAGK (**1**)	X	X			X	X			-	X		
GEPGPTGI**Q**GPPGPAGEEGKR (**1**)	X	X			X	X			X	X		
TGPPGPAG**Q**DGRPGPPGPPGAR (**1**)	X	-			X	X			X	X		
GDAGPPGPAGPAGPPGPIG**N**VGAPGPK (**1**)	X	-			-	-			-	-		
GPPGSAGSPGKDGL**N**GLPGPIGPPGPR (**1**)	-	X			-	X			-	X		
GAPGADGPAGAPGTPGP**Q**GIAG**Q**R (**2**)	X	X	-		X	X	X		X	X	X	
GFSGL**Q**GPPGPPGSPGE**Q**GPSGASGPAGPR (**2**)	X	X	-		X	-	-		-	-	-	
G**N**DGATGAAGPPGPTGPAGPPGFPGAVGAK (**1**)	-	X			-	-			-	-		
GAPGDRGEPGPPGPAGFAGPPGADG**Q**PGAK (**1**)	X	X			X	X			X	X		
GEPGPPGPAGFAGPPGADG**Q**PGAK (**1**)	-	X			-	X			-	-		
GVPGPPGAVGPAGKDGEAGA**Q**GPPGPAGPAGER (**1**)	X	X			X	X			X	X		
GLPGPPGAPGP**Q**GF**Q**GPPGEPGEPGASGPMGPR (**2**)	X	X			-	-	-		-	-	-	
GE**Q**GPAGSPGF**Q**GLPGPAGPPGEAGKPGE**Q**GVPGDLGA PGPSGAR (**3**)	X	X	-	-	-	X	-	-	X	X	X	-
Collagen alpha-2(I) chain (P02465)												
GD**Q**GPVGR (**1**)	X	-			-	-			X	X		
GVVGP**Q**GAR (**1**)	X	X			X	X			X	X		
IG**Q**PGAVGPAGIR (**1**)	X	-			X	X			X	X		
GFPGSPG**N**IGPAGK (**1**)	X	X			X	X			X	X		
GEPG**N**IGFPGPK (**1**)	X	X			X	X			X	X		
GELGPVG**N**PGPAGPAGPR (**1**)	X				X	-			-	-		
GP**N**GDSGRPGEPGLMGPR (**1**)	-	X			-	X			-	-		
HG**N**RGEPGPAGAVGPAGAVGPR (**1**)	-	X			-	-			X	X		
GEVGPAGP**N**GFAGPAGAAG**Q**PGAK (**2**)	-	-	X		-	-	-		-	-	X	
GYPG**N**AGPVGAAGAPGP**Q**GPVGPVGK (**2**)	X	X	-		-	-	-		X	X	X	
GEVGLPGLSGPVGPPG**N**PGA**N**GLPGAK (**2**)	-	X	X		-	-	X		X	-	-	
GPSGPPGPDG**N**KGEPGVVGAPGTAGPSGPSGLPGER (**1**)	X	X			-	-			-	-		
GAPGAIGAPGPAGA**N**GDRGEAGPAGPAGPAGPR (**1**)	-	X			-	X			-	X		
GPPGASGAPGP**Q**GF**Q**GPPGEPGEPG**Q**TGPAGAR (**3**)	X	X	X	-	-	X	-	-	-	-	-	-

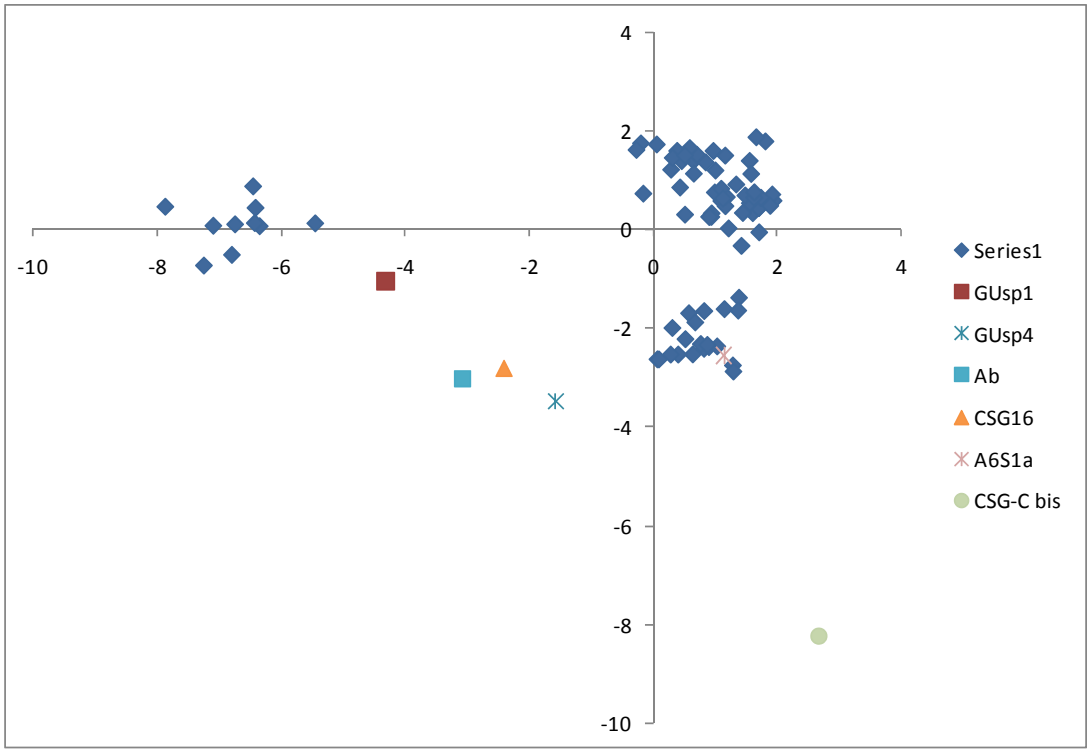
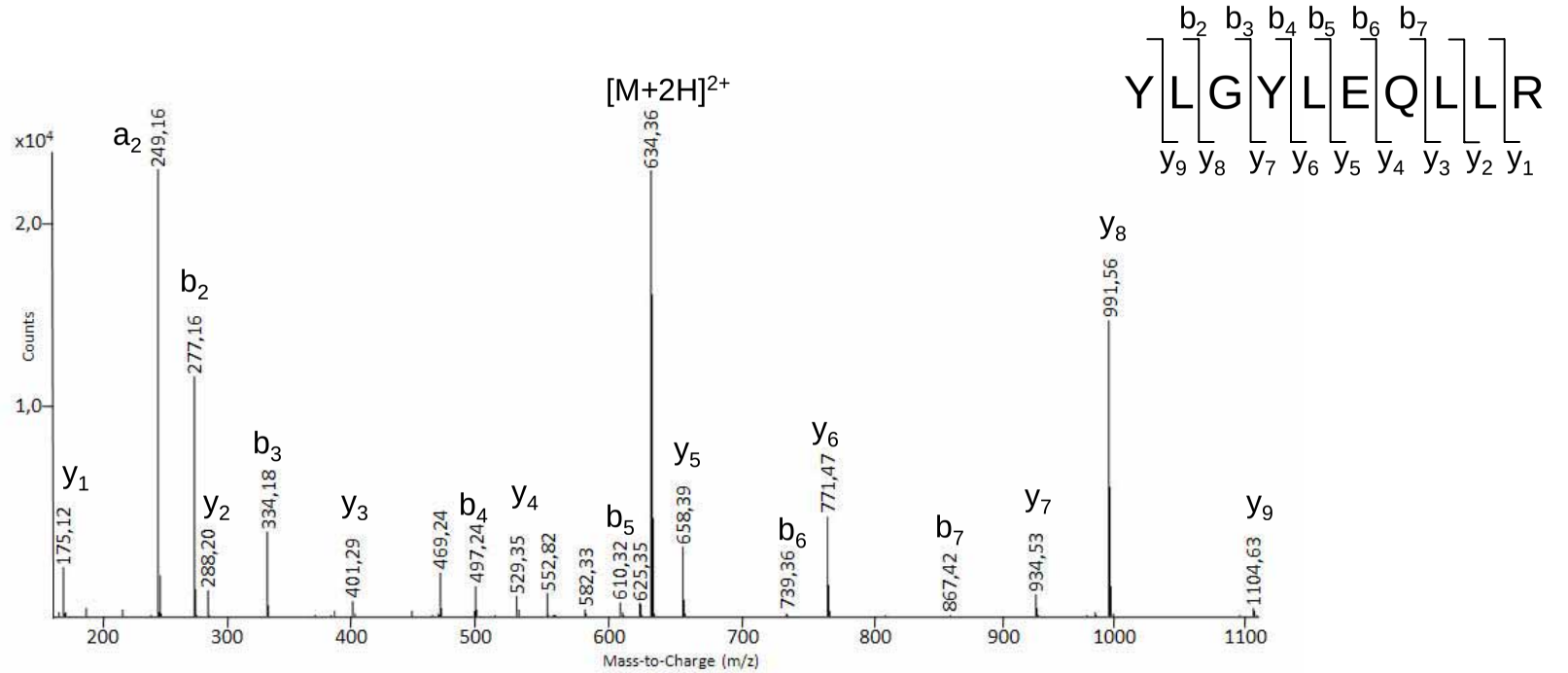

4
2
0
-2
-4
-6
-8
-10
-10
-8
-6
-4
-2
0
2
4
Series1
GUsp1
GUsp4
Ab
CSG16
A6S1a
CSG-C bis

Figure 1

A


b2 b3 b4 b5 b6 b7
Y L G Y L E Q L L R
y9 y8 y7 y6 y5 y4 y3 y2 y1
[M+2H]2+
x10⁴
Counts
Mass-to-Charge (m/z)
a2 249,16
b2 277,16
b3 334,18
y1 175,12
y2 288,20
y3 401,29
b4 497,24
469,24
y4 529,35
552,82
582,33
b5 610,32
625,35
y5 658,39
634,36
b6 739,36
y6 771,47
b7 867,42
y7 934,53
y8 991,56
y9 1104,63

B

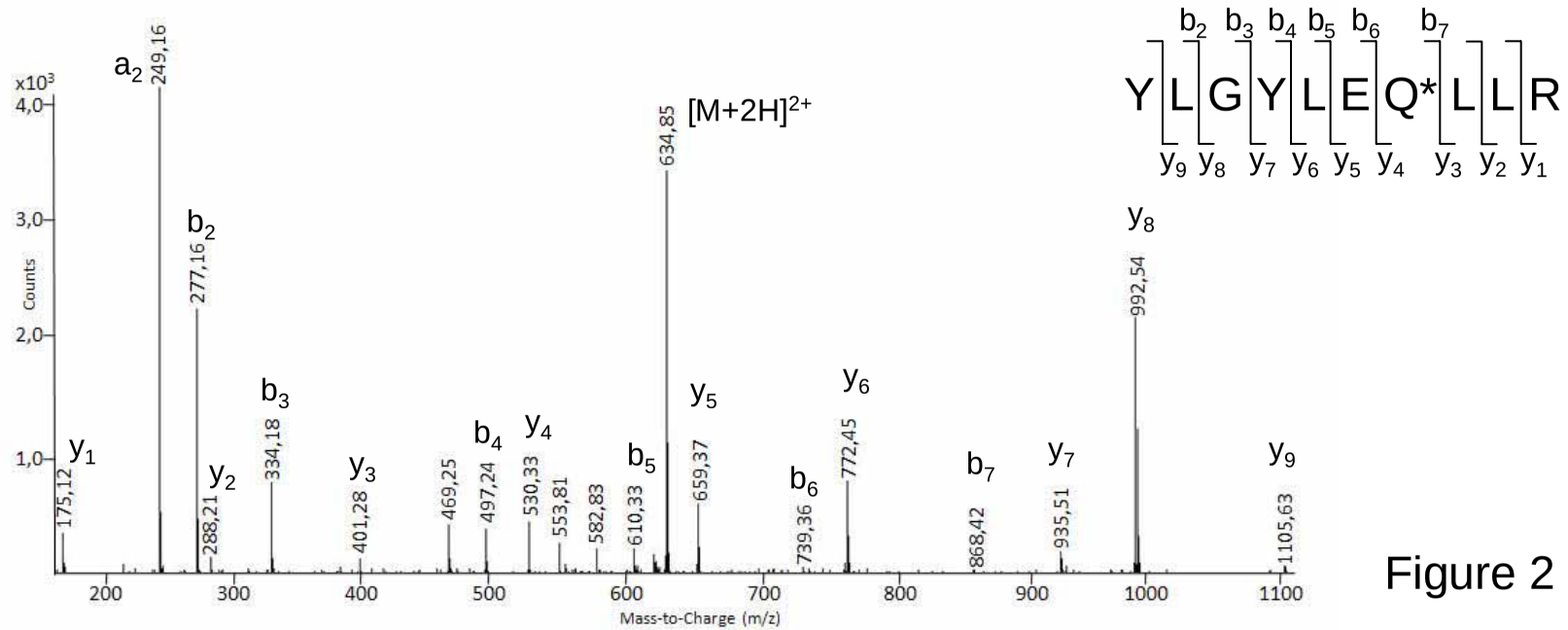

b2 b3 b4 b5 b6 b7
Y L G Y L E Q* L L R
y9 y8 y7 y6 y5 y4 y3 y2 y1
[M+2H]2+
x10³
Counts
Mass-to-Charge (m/z)
a2 249,16
b2 277,16
b3 334,18
y1 175,12
y2 288,21
y3 401,28
b4 497,24
469,25
y4 530,33
553,81
582,83
b5 610,33
y5 659,37
634,85
b6 739,36
y6 772,45
b7 868,42
y7 935,51
y8 992,54
y9 1105,63

Figure 2

Table S1. Identification of the proteins in the sample from the Cimitero Monumentale of Pisa by overnight trypsin digestion in heterogeneous phase and LCMSMS analysis[a]. Proteins were identified searching UniprotSprot database, with *Chordata* as taxonomy restriction, with MSMS Ion search Mascot software (Matrix Science). Only identification of proteins with at least two peptides were considered as significative.

Identified protein (Accession number)	**Total score[a]**	**Sequence coverage (%)**	**Matched sequence** (Only oxidation of methionine, hydroxylation of proline and lysine, pyro-Glu formation at Gln at the N-terminus of peptides were inserted as variable modifications in the MSMS Ion Search Program).
Beta-casein precursor (P02666)	1689	75	EMPFPK VLPVPQK AVPYPQR VLPVPQKAVPYPQR FQSEEQQQTEDELQDK DMPIQAFLLYQEPVLGPVR DMPIQAFLLYQEPVLGPVRGPFPIIV AVPYPQRDMPIQAFLLYQEPVLGPVR IHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK YPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK
Casein alpha-S1 (P02662)	1464	50	TTMPLW EDVPSER EKVNELSK YLGYLEQLLR HIQKEDVPSER FFVAPFPEVFGK VPQLEIVPNSAEER FFVAPFPEVFGKEK HQGLPQEVLNENLLR YKVPQLEIVPNSAEER EDVPSERYLGYLEQLLR HPIKHQGLPQEVLNENLLR EPMIGVNQELAYFYPELFR EGIHAQQKEPMIGVNQELAYFYPELFR
Collagen alpha-1(I) chain (P02453)	780	35	GAAGLPGPK GFSGLDGAK GSEGPQGVR GVVGLPGQR GFPGADGVAGPK GVQGPPGPAGPR GVPGPPGAVGPAGK GQAGVMGFPGPK GLTGSPGSPGPDGK GPSGPQGPSGPPGPK GFPGLPGPSGEPGK GEPGPAGLPGPPGER GSAGPPGATGFPGAAGR GETGPAGPAGPIGPVGAR GLTGPIGPPGPAGAPGDK GSPGEAGRPGEAGLPGAK EGAPGAEGSPGRDGSPGAK GPPGPMGPPGLAGPPGESGR VGPPGPSGNAGPPGPPGPAGK GEPGPTGIQGPPGPAGEEGK SGDRGETGPAGPAGPIGPVGAR GEPGPTGIQGPPGPAGEEGKR TGPPGPAGQDGRPGPPGPPGAR GAPGADGPAGAPGTPGPQGIAGQR GEPGPPGPAGFAGPPGADGQPGAK GETGPAGRPGEVGPPGPPGPAGEK GAPGDRGEPGPPGPAGFAGPPGADGQPGAK GFSGLQGPPGPPGSPGEQGPSGASGPAGPR GLTGPIGPPGPAGAPGDKGEAGPSGPAGPTGAR GVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGER GLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPR
Kappa-casein precursor (P02668)	646	26	YIPIQYVLSR HPHPHLSFMAIPPK HPHPHLSFMAIPPKK

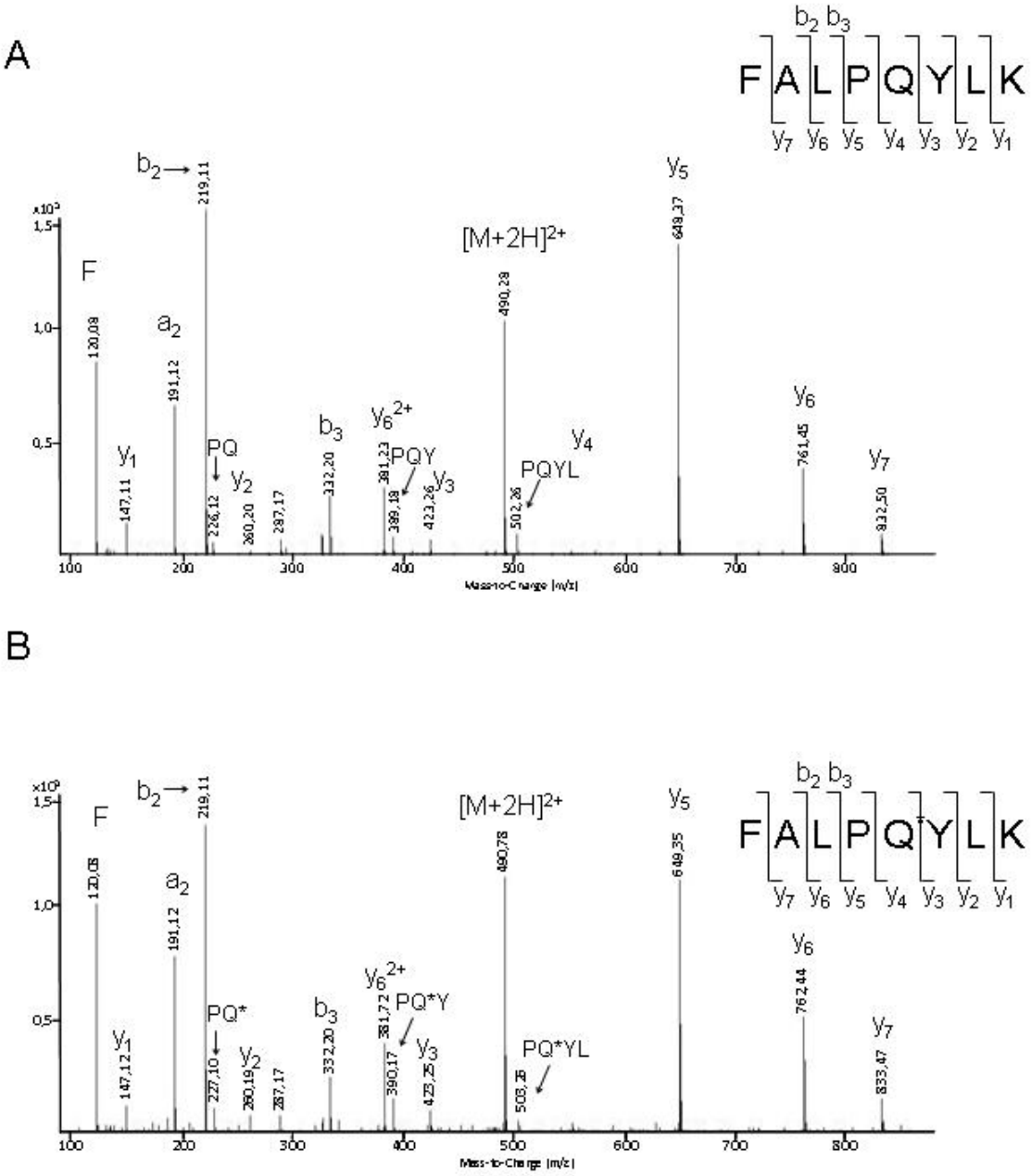
			SPAQILQWQVLSNTVPAK
Collagen alpha-2(I) chain (P02465)	592	30	GVVGPQGAR GPSGPQGIR VGAPGPAGAR GLVGEPGPAGSK GEPGNIGFPGPK IGQPGAVGPAGIR GPAGPSGPAGKDGR GIPGPVGAAGATGAR GFPGSPGNIGPAGK GIPGEFGLPGPAGAR GDGGPPGATGFPGAAGR SGETGASGPPGFVGEK GEPGPAGAVGPAGAVGPR GAAGLPGVAGAPGLPGPR GPPGESGAAGPTGPIGSR GELGPVGNPGPAGPAGPR GSTGEIGPAGPPGPPGLR RGSTGEIGPAGPPGPPGLR EGPVGLPGIDGRPGPIGPAGAR GLPGVAGSVGEPGPLGIAGPPGAR GPPGASGAPGPQGFQGPPGEPGEPGQTGPAGAR
Casein alpha-S2 (P02663)	275	46	VIPYVR LNFLKK NMAINPSK FALPQYLK VIPYVRYL AMKPWIQPK NAVPITPTLNR RNAVPITPTLNR ALNEINQFYQK TVDMESTEVFTK FALPQYLKTVYQHQK FPQYLQYLYQGPIVLNPWDQVK
Beta-lactoglobulin (P02754)	129	33	ALPMHIR IDALNENK VLVLDTDYK VLVLDTDYKK TPEVDDEALEK TPEVDDEALEKFDK VYVEELKPTPEGDLEILLQK
Lactadherin (Q95114)	52	9	VTGIITQGAR QFQFIQVAGR DFGHIQYVAAYR

[a] Ions score is -10*Log(P), where P is the probability that the observed match is a random event. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. (http://www.matrixscience.com/help/interpretation_help.html)

Table S2: Main series matches of the fragmentation spectra reported in Figure 1 and Figure 2. Those ions containing the deamidation site are highlighted reported in bold.

	YLGYLEQLLR	**YLGYLEQ*LLR**	**FALPQYLK**	**FALPQ*YLK**
Daughter ion	m/z	m/z	m/z	m/z
y_1	175.1213	175.1188	147.1142	147.1164
y_2	288.2069	288.2066	260.2044	260.1990
y_3	401.2924	401.2823	423.2609	423.2469
y_4	529.3491	**530.3313**	551.3244	**552.2882**
y_5	658.3893	**659.3660**	648.3701	**649.3549**
y_6	771.4714	**772.4542**	761.4525	**762.4367**
y_7	934.5386	**935.5100**	832.4970	**833.4685**
y_8	991.5664	**992.5379**		
y_9	1104.6354	**1105.6353**		
b_2	277.1605	277.1568	219.1142	219.1140
b_3	334.1784	334.1803	332.2001	332.1984
b_4	497.2403	497.2425	429.2483	
b_5	610.3256	610.3282		
b_6	739.3653		720.3621	
b_7	867.4202			
b_8	980.4983			
b_9	1093.6009	**1094.5944**		

Figure S1: MSMS spectra of the doubly charged ions at m/z 490.28 (A) and 490.78 (B) of the peptide 189-FALPQYLK-196 from bovine alpha-S2 casein (P02663) and its deamidated form, respectively, identified in the analysis of the sample from Camposanto Monumentale of Pisa. The product ions are indicated with the observed masses.



WORK IN PREPARATION

PLEUROTUS OSTREATUS CLASSICAL BREEDING: AN USEFUL TOOL TO IMPROVE AND INCREASE LACCASE PRODUCTION IN DIKARYOTIC STRAINS.

Del Vecchio C., Lettera V., Leo G., Pezzella C., Piscitelli A., Birolo L. and Sannia G.

Department of Organic Chemistry and Biochemistry, University of Naples Federico II
Complesso Universitario Monte S. Angelo - Via Cinthia 4, IT-80126 Napoli

Abstract

White-rot basidiomycetes, the most common wood-rotting organisms, are characterized by their ability to produce extracellular oxidative enzymes, which are mainly involved in the degradation of lignin. A non specific reaction mechanism characterizes lignin-modifying enzymes allowing them to be also able to degrade a wide range of environmental pollutants, such as polycyclic aromatic hydrocarbons, chlorophenols, and aromatic dyes. Due to their broad substrate range, these oxidative enzymes are being increasingly evaluated for a variety of biotechnological applications. In particular, in the last decades the physiological role of laccases and their use in several industrial fields has been deeply investigated. However, large amounts of enzyme are required both for fundamental studies and for their practical application at industrial scale. Nevertheless, laccases secreted from wild-type fungal organisms may not be suitable for commercial purposes mainly because of low yields of production and high costs of preparation procedures.

In the present work the white rot fungus *Pleurotus ostreatus* has been used as organism for classical breeding experiments. The genome of this basidiomycete has been recently sequenced and several genes coding for laccase isoenzymes have been annotated. Besides recombinant heterologous expression, classical breeding is an useful alternative method to OGM either to improve ligninolytic enzymes production and to increase isoenzymatic variability. In particular, starting from two different *P. ostreatus* variants, three laccase higher-producing dikaryotic strains have been obtained through crossing compatible characterized monokaryons. The three selected strains reached expression levels of 100,000 U/L, increasing the trite of parental strains up to four folds. Furthermore, this work allowed production of a new isoenzyme, the laccase POX1, whose cDNA was previously isolated, but the corresponding native protein was not identified yet.

Introduction:

White-rot fungi (WRF) have been widely recognized as the most efficient lignin degraders [1]. This ability results from the production of various isoforms of extracellular ligninolytic enzymes including manganese peroxidases, lignin peroxidases and laccases [2, 3]. The latter are involved in a variety of physiological roles including morphogenesis, fungal plant pathogen/host interactions, stress defence and lignin degradation [3-5]. Furthermore, due to their broad substrate range, laccase enzymes can play a role not only in the degradation of lignin in their natural lignocellulosic substrates but also in the degradation of various xenobiotic compounds [6]. Thus, WRF and their enzymes are widely considered to have potential for industrial purposes. As a fact, laccases offer industrially useful

advantages of great interest for biotechnological applications such as biodegradation of environmental pollutants (e.g. like textile dye or explosives), bioconversion of lignin and detoxification of agricultural by-products, including olive mill wastes or coffee pulp [7].

Laccases secreted from native sources are usually not suitable for large scale applications, mainly due to low production yields and high cost of preparation and purification procedures. To successfully utilize laccases for biotechnological purposes, production of large quantities of enzymes at low cost is required. Only few examples of industrial uses of laccases currently exist and most of them utilize genetically modified tailored laccase, produced by recombinant expression [8]. Therefore, any attempt to increase the production of enzymes from microbial sources could be of considerable industrial interest.

The constantly growing popularity of the edible fungus *P. ostreatus*, as well as the significantly increased knowledge of its ligninolytic enzymatic system, render worthwhile spending efforts for obtaining strains, derived from this mushroom, improved for phenoloxidases production at large-scale.

So far, eight members of the *P. ostreatus* laccase multigene family have been isolated and sequenced [9, 10] and up to 12 members of this family have been identified in the released *P. ostreatus* genome databank. However, the redundancy of laccase genes raises the question about their respective functions *in vivo*, and this question is even more pertinent since cDNAs or proteins for some of these genes have not been found yet [9].

Several methods have been carried out for strain improvement in *Pleurotus* including selection, hybridization and gene transformation [11-14]. However, based on current legislation (European Directive 2001/18/CE), genetic transformation and mutagenic treatments produce strains not suitable for “natural or safe processes”. Therefore, the construction of genetically modified organisms can not be chosen to improve the addressed quality of the fungus, and breeding should be based on classical genetic approaches. This technique is based on the mating of two monokaryotic compatible strains, whose hyphae are able to fuse and give rise to a dikaryotic mycelium in which the two parental nuclei remain independent [13, 14]. Production of tmonokaryotic strain, germinating from uninucleate basidiospores, is achieved when the fungus enters into the reproductive phase triggering basidiocarp formation: during basidia formation, karyogamy takes place immediately before the onset of the meiosis, giving rise to four uninucleate basidiospores. At this stage, genetic recombination occur, although some reports have also suggested the occurrence of parasexual somatic recombination in higher basidiomycetes [15]. As a fact, dykariotic strains derived from single spore germinating monokaryons could result in improved specific multifactorial traits like mycelium growth, colony fitness and protein expression or secretion.

Chaudhary *et al.* [16] developed single spore isolates from the white-rot fungi *Pleurotus djamor, P. ostreatus var. florida, Pleurotus citrinopileatus* and *Hypsizygus ulmarius*. The hybrids showed improved mycelial growth rate compared to that of the parental strains. In another work, Sawashe and Sawant [17] developed hybrid cultures which required a significantly shorter period for spawn run as compared to the parent species. In the frame of improvement of quantity and variability of enzymatic expression, selection of new hybrid strains for laccase production could be viewed as a solution to make the entire process cost effective. Till now only few data in literature have been reported to improve laccase production by classical breeding in *P. ostreatus* [18]. To the best of our knowledge, increasing of extracellular laccase

activities have ever been not correlated to the production of phenoloxidase isoenzymes, previously unexpressed.
In the present work, improvement of different strains of *P. ostreatus* obtained by classical breeding and screening for their laccase productivity has been performed. The obtained dikaryotic strains, in according to European legislation, are considered generally recognized as safe (GRAS) and therefore they could be good candidates for their use in bioproduction as natural strains. We have performed dikaryotization through crossing of characterized basidiospore-derived monokaryons of two *P. ostreatus* variants in order to increase the trite and the expressed pattern of laccase isoenzymes.

Materials and methods

Substrates and chemicals

Unless otherwise specified, all substrates and chemicals were purchased from Sigma-Aldrich.

Organism

All *P. ostreatus* monokariotyc and dikaryotic strains were maintained through periodic serial transfers and kept at 4°C on agar plates in the presence of 2.4% potato dextrose and 0.5% yeast extract (PDY) (Difco).

Culture conditions

Shacked submerged cultivation was carried out in 100 ml Erlenmeyer flasks (125 rpm) containing 30 ml of PDY supplemented with copper sulphate (final concentration 150 µM) and ferulic acid 2 mM (dissolved in ethanol and added at the second day of growth). The flasks were inoculated with four agar plugs (8 mm diameter) cut from the actively growing part of the colony on a Petri dish and incubated for at least 15 days at 28° C in the dark.

Fructification and basidiospores isolation

P. ostretus mushrooms were cultivated in 500 mL flasks containing 400 g of wheat-straw (65% water content), which were twice autoclaved for 1 h at 121 °C separated by a period of 24 h at room temperature. Each flask was inoculated with four agar plug (13 mm diameter), and left to grow at 28 °C for 30 days in the dark. Fructification was promoted by opening the jars, and placing them in presence of daylight in a chamber at 15±5 °C and 90% relative humidity. Primordia appeared after a further 15 days of growth, and basidiocarps were harvested 7 days later and weighed [10].
P. ostreatus basidiospores were collected by sporal print on a glass Petri dish, previously sterilised in autoclave (1 h at 121°C). A fresh healthy fruit body of *P. ostreatus* was attached to the cap by eukit® resin under aseptic conditions in such a way that the gills of the fruit body were facing underneath. After 24 h, the lid was removed from the top of the Petri dish holding the spore print.
A spore suspension was prepared in 1 mL of sterile physiological salt solution (0.9% NaCl). Sporal concentration was estimated by counting them in a Thoma chamber on optical microscopy. The basidiospores suspension was plated on PDY solid medium in Petri dishes after appropriate dilution to obtain distinct monosporic colonies and incubated at 28°C. Colonies formed were transferred onto agar slants and tested for production of ligninolytic enzymes.

Mating test
Small pieces of mycelium of two monokaryons (agar plugs 5 mm diameter) were inoculated close to each other on 2% malt extract (Difco) agar plates; after 5–7 days growth the presence of clamp connections, in the interaction zone was checked microscopically.

Enzyme assays
Phenol-oxidase activity was assayed at 25°C using 2,2' –azinobis-(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) as substrate [19]. The assay mixture contained 2 mM ABTS and 0.1 M sodium-citrate buffer, pH 3.0. Oxidation of ABTS was followed by absorbance increase at 420 nm (ε = 36,000 M^{-1} cm^{-1}) for 1 minute. Enzyme activity was expressed in IU.
Enzyme activity was measured in the filtrates from four parallel flasks after removing the mycelia. All measurements were repeated at least in triplicate.

Native polyacrylamide gel electrophoresis
Polyacrylamide gel electrophoresis (PAGE) was carried out at alkaline pH under non denaturing conditions. The resolving and stacking gels contained 9% and 4% acrylamide, respectively. The buffer solution used for the resolving gel contained 50 mM Tris-HCl (pH 9.5), and the buffer solution used for the stacking gel contained 18 mM Tris-HCl (pH 7.5). The electrode reservoir solution contained 25 mM Tris and 190 mM glycine (pH 8.4). Gels were stained to visualize laccase activity by using ABTS as substrate, in 0.1M sodium citrate buffer pH3.0. Samples containing 0.015 laccase unities were loaded on each lane.

Laccase isoenzyme fractionation
Ultra-filtrated protein samples were analyzed against 50mM Tris–HCl and were loaded on a Resource Q (GE Healthcare 1mL) column equilibrated with the same buffer. The column was washed at a flow rate of 1 ml/min with 3 column volume (cv) of buffer, and a 0–0.5 M NaCl linear gradient (20 cv) was applied. Fractions containing laccase activity were pooled and concentrated.

Protein identification by mass spectrometry
Protein band corresponding to the laccase of interest was excised from the native gel. Slices of interest were destained by washes with 0.1 M NH_4HCO_3 pH 7.5 and acetonitrile, reduced for 45 minutes with 100 µl of 10 mM dithiothreitol in 0.1 M NH_4HCO_3 buffer pH 7.5 and carboxyamidomethylated for 30 minutes in the dark with 100 µl of 55 mM iodoacetamide in the same buffer. Tryptic digestion was performed by adding for each slice 100 ng of enzyme in 10 µl of 10 mM NH_4HCO_3 pH 7.5 for 2 hours at 4 °C. The buffer solution was then removed and were added 50 µl of 10 mM NH_4HCO_3 pH 7.5 for 18 hours at 37 °C. Peptides were extracted by washing the gel slices with 10 mM NH4HCO3 and 1% formic acid in 50% acetonitrile at room temperature.
The peptide mixtures were filtered by using 0.22 µm PVDF membrane (Millipore) and analysed using a 6520 Accurate-Mass Q-TOF LC/MS System (Agilent Technologies, Palo Alto, CA) equipped with a 1200 HPLC system and a chip cube (Agilent Technologies). After loading, the peptide mixture was first concentrated and washed in 40 nL enrichment column (Agilent Technologies chip), with 0.1% formic acid in 2% acetonitrile as the eluent. The sample was then fractionated on a C18 reverse – phase capillary column (Agilent Technologies chip) at flow rate of 400nl/min, with a

linear gradient of eluent B (0.1% formic acid in 95% acetonitrile) in A (0.1% formic acid in 2% acetonitrile) from 7 to 80% in 50 min. Peptide analysis was performed using data - dependent acquisition of one MS scan (mass range from 300 to 180 0 m/z) followed by MS/MS scans of the five most abundant ions in each MS scan. Dynamic exclusion was used to acquire a more complete survey of the peptides by automatic recognition and temporary exclusion (0.5 min) of ions from which definitive mass spectral data had previously acquired. Nitrogen at a flow rate of 3 L/min and heated to 325°C was used as the dry gas for spray desolvation. MS/MS spectra were measured automatically when the MS signal surpassed the threshold of 50000 counts. Double and triple charged ions were preferably isolated and fragmented over single charged ions. The acquired MS/MS spectra were transformed in mz.data format and used for proteins identification with a licensed version of MASCOT 2.1 (Matrix Science, Boston, USA). Raw data from nanoLC–MS/MS analyses were used to query the *P. ostreatus* database and the Mascot search parameters were: trypsin as enzyme, allowed number of missed cleavage 3, carbamidomethyl as fixed modification, oxidation of methionine, pyro-Glu N-term Q, as variable modifications, 10 ppm MS tolerance and 0.6 Da MS/MS tolerance, peptide charge, from +2 to +3. Individual ions scores >32 indicate identify or extensive homology ($p<0.005$). Protein scores are derived automatically by MASCOT Software from ions scores as non-probabilistic for ranking protein hits.

Analysis of protein sequence

The predicted protein sequence of newly identified laccase POX1 was analyzed by comparison with those of the previously defined *Pleurotus ostreatus* laccases. Alignments of amino-acidic sequences were generated with ClustalW2 (http://www.ebi ac.uk/Tools/clustalaw2/index.html). Signal peptide was predicted with SignalP V2.0 (http://www.cbs.dtu.dk/services/). The amino-acid conservation scoring among the laccase signature sequences was performed by PRALINE (http://www.ibi.vu.nl/programs/pralinewww/).

Results

Production of *P. ostreatus* monokaryotic strains

Two varieties of the fungus *P. ostreatus, var. florida* (ATCC MYA-2306) and *var. ostreatus* (D1208), have been used as parental dikaryotic strains. The two varieties show different morphological and physiological features like size, temperature tolerance, colour, etc., and have been previously extensively characterized for their ability to produce oxidative enzymes [9, 20, 21], hydrophobins [22, 23], and natural compounds [24]. In order to produce new dikaryotic hybrids with increased production capabilities, basidiospores-derived monokaryons, obtained from each variety, have been isolated and analysed. Collected spores from the two different basidiocarps were successfully germinated in solid medium. Microcolonies progeny was microscopically analyzed and monokaryotic state was confirmed by the absence of mycelial clamp connections, as reported by Eichlerov´a I. and Homolka L. [18].

Isolated monokaryons (marked “A” and “D”, for strains derived from the parental dikaryons ATCC MYA2306 and D1208, respectively) were inoculated in PDY submerged culture in order to detect laccase activity during fungal fermentation. Time courses analysis of extracellular laccase activities showed high variability among the strains and a maximum of production between the 9^{th} and the 10^{th} day of growth. Among twenty-eight randomly chosen germinating spores, three monokaryons (5A,

6A, 18A) from the parental strain *P. florida* and three monokaryons (3D, 9D, 11D) from the parental strain *P. ostreatus* exhibit higher or comparable production levels than the two parental strains. Corresponding data are reported in table.1 (a) and (b), derived from three independent experiments and the mean values were taken. The standard deviation for the experiments was less than ±10%. As also reported by Eichlerov´a and Homolka [18], the isolates differ also in morphology and growth rate when they grew on solid media. However, these colony appearances were transient and, in general, not connected with laccase production levels.

Isoenzymatic pattern analyses were performed on the samples derived from extracellular medium of monokaryotic fungal cultures. Native PAGE of laccase isoenzymes stained for laccase activity were carried out and the identity of the laccase isoenzymes was ascertained on the basis of the different electophoretic mobility. Analysis of samples withdrawn from the media at different growth times showed that the activity is associated mainly to the production of three isoenzymes POXA1b, POXA3, and POXC, [9] as reported in figure 1.

Moreover, the monokaryon 6A secretes a new isoform, previously never identified (figure 1). When the protein band, corresponding to the new laccase was excised and analyzed by tandem mass spectrometry, 5 peptide (sequence coverage, 20%; protein summary score, 286) could unambiguously be identified. *P. ostreatus* database search with the raw MS/MS data identified the protein encoded by the predicted laccase transcript POX1. The cDNA of this 'new' native laccase was previously isolated [25], but the corresponding produced protein in an active form was not identified yet.

Dikaryotization of *P. ostreatus* generated by classical crossing

Because laccase production is regulated by multifactorial and mutiallelic expression systems which are dependent on extra- and intra-cellular regulations [13, 14, 26, 27], all crossings among the six monokaryotic strains were performed in order to obtain new high laccase-producers dikaryotic strains (table 2). Anastomosis induction followed by formation of clamp connections (figure 2) was confirmed microscopically at ×45 magnification, indicating that compatibility among two strains and subsequent formation of the corresponding dykarion.

New dikaryotic hybrids derived from the two parental strains were obtained through crossing of selected strains, excepting 5AX9D, that results incompatible. Time course analysis on laccase activity produced by the new dikaryotic strains grown in submerged cultures were perfomed. Six out of the eight dikaryons produced a higher amount of laccase activity. Moreover, three of them (5AxD3, 5AxD11 and 6AxD11) reached a production level of up to 70,000 U/L (figure 3), and were chosen for further investigations.

Effect of ferulic acid inducer [28] on the selected dikaryotic hybrids 5AxD3, 5AxD11 and 6AxD11 was also analyzed. Addition of the inducer to the culture broth positively affects extracellular laccase activity of the analyzed strains. Nevertheless, results indicate that hybrids differentially responded to the inducer presence (figure 4). As a fact, all the selected hybrids were more sensitive to the presence of ferulic acid than the parental strains. Moreover, the 5AXD3 strain resulted the best laccase producer also in inducing conditions, reaching a laccase expression of up to 110,000 U/L.

Laccase fractionation

Time course of the isoenzymatic pattern of the selected dikaryons was performed analyzing the extracellular medium of fungal culture by native PAGE stained for

laccase activity. Analysis of samples withdrawn from the media at different growth times confirmed the presence of the three isoenzymes POXA1b, POXA3, and POXC, as reported above for the corresponding monokaryotic parents, either in the presence or in the absence of inducer.
Furthermore, the strain 6AxD11, in both conditions, secretes the new isoform POX1, previously detected in the monokaryotic parental strain 6A. Relative quantification of laccase isoenzymes, in basal conditions, were also carried out in correspondence of laccase activity peak (9^{th} day). Fungal cultures were harvested and fractionated by anionic exchange chromatography ResourceQ (table 3). The fractions were pooled, concentrated, assayed and analyzed by electrophoretic separation on native PAGE gels stained for laccase activity. The extracellular enzyme mixtures of the strains 5AxD3 and 5AxD11 were showed to contain POXA3 and POXC isoenzymes, the latter being the most abundant (> 90%). Performing the same analysis on the isoenzymatic pattern of 6AxD11, POXA1b laccase and a significant production of POX1 were also detected (35% relative abundance).

Discussion

Laccase enzymes exhibit an extraordinary range of natural substrates (phenols, anilines, inorganic/organic metal compounds, etc) which is the major reason of their attractiveness for several biotechnological applications [29]. More than one laccase isoenzyme, either constitutive or inducible, has been detected in most white-rot fungi [3, 30]. Nevertheless, constitutive extracellular phenoloxidase from basidiomycetes are produced only in small amounts. In according to Galhaup *et al.* [31], one of the major limitations for the large-scale industrial applications of fungal laccases is the low production yields by both wild type and recombinant fungal strains and, as consequence, the high costs for enzyme production [32, 33]. To successfully utilize laccases in industrial applications, the production of large quantities of enzyme at low cost is required, and therefore, any attempt to increase their production from microbial sources could be of considerable interest.
In our study improvement of *P. ostreatus* laccase production was achieved by classical breeding. In particular, basidiospores-derived monokaryons, obtained from *P. ostreatus var. florida* and *var. ostreatus,* were isolated and analyzed. Crossing were performed in order to produce hybrids with favourable cultural characteristics, applicable in large-scale production.
Comparison of the morphological characteristics and laccase levels production of the randomly isolated monokaryons indicated a significantly high genetic variation. The surprisingly high variability between monokaryotic strains for the laccase production may indicate an heterozygosity in genes responsible for the synthesis and secretion of the analyzed enzymes [18].
Moreover, isoenzymatic variety was deeply investigated and a novel laccase protein was identified in one of the selected monokaryotic strains, undetectable in parental dykarion *P. florida*. The new laccase protein was unambiguously identified as POX1 through MS/MS analysis and querying of the *P. ostreatus* genome data bank. Discovery of POX1 proves the production of a functional protein by the related encoding transcript, previously identified by Giardina *et al*. [25]. The new isoenzyme POX1 was also expressed by one of the selected dikaryotic strains, the 6AxD11, either in basal and inductive analyzed conditions. Not surprisingly, the 6A phenotype related to the laccase expression resulted co-dominant in the corresponding dikayriotic strain 6AxD11. As a fact, during *P. ostreatus* dikaryotization process and

dikaryotic mitotic division, the two parental nuclei remain independent and probably silencing on the homologous loci coding for laccase enzymes does not occur [34].
All compatible crossings were performed and allowed the identification and characterization of three new high laccase producer *P. ostreatus* strains. Laccase production levels of the dikaryotic hybrids increased up to 4 folds respect to the maximum level of dikaryotic parents. The extraordinary enhanced laccase activity in dikaryotic strains (hybrid of high-yielding monokaryons), as well as certain physiological differences, may be due to genetic recombination effects, occurred during basidia formation [15].
As well as reported by Chaudhary *et al.* [16] and Sawashe and Sawant [17], results indicate that classical breeding is an useful method for obtaining isolates with interesting combination of enzyme and growth characteristics without inducing physics or chemical mutagenization. The classical natural method of hybrid preparation represents a considerable advantage in comparison with the other common methods (e.g. mutagenization), because undesirable side-effects (e.g. pleiotropic) can be avoided. In addition, new generated dikaryotic strains, in according with current legislation (European Directive 2001/18/CE) are still considered as GRAS and thus could be good candidates for their use in bioprocesses as natural strains.
This study also extends recent reports [9, 10] that demonstrated the production of multiple laccase isoforms in the basidiomycete white rot fungus *P. ostreatus*, enlarging the assortment of these industrially useful enzymes. Further investigations will reveal the peculiar molecular aspects and the biochemical properties of POX1 laccase.

References

[1] Ten Have R., Teunissen P.J. Oxidative mechanisms involved in lignin degradation by white-rot fungi. *Chem Rev*. (2001); 101(11):3397-413.
[2] Cullen D. and Kersten P. Fungal enzymes for lignocellulose degradation. *In: Kinghorn JR and Turner G (eds) Applied molecular genetics of filamentous fungi.* New York: Chapman & Hall (1992) Chapter 4.
[3] Giardina P., Faraco V., Pezzella C., Piscitelli A., Vanhulle S., Sannia G. Laccases: a never-ending story. *Cell Mol Life Sci*. (2010); 67(3):369-85.
[4] Thurston C.F. The structure and function of fungal laccases. *Microbiology*. (1994); 140:19-26.
[5] Gianfreda L., Xu F., Bollag J.M. Laccases: a useful group of oxidoreductive enzymes. *Bioremediat J*. (1999); 3:1-26.
[6] Hatakka A. Lignin-modifying enzymes from selected white-rot fungi: production and role from in lignin degradation. *FEMS Microbiology Reviews.* (1994); 13:125-135.
[7] Arora D.S. and Sharma R.K. Ligninolytic fungal laccases and their biotechnological applications. *Appl Biochem Biotechnol* (2010); 160(6):1760-88.
[8] Piscitelli A., Pezzella C., Giardina P., Faraco V. and Sannia G. Heterologous laccase production and its role in industrial applications. *Bioengineered Bugs*. (2010); 1:4, 1-11.
[9] Pezzella C., Autore F., Giardina P., Piscitelli A., Sannia G. and Faraco V. The *Pleurotus ostreatus* laccase multi-gene family: isolation and heterologous expression of new family members. *Current Genetics.* (2009); 55:45-57.

[10] Lettera V., Piscitelli A., Leo G., Birolo L., Pezzella C., Sannia G. Identification of a new member of *Pleurotus ostreatus* laccase family from mature fruiting body. *Fungal Biology.* (2010); doi:10.1016/j.funbio.2010.06.004.
[11] Irie T., Honda Y., Watanabe T. and Kuwahara M. Efficient transformation of filamentous fungus *Pleurotus ostreatus* using single -strand carrier DNA. *Appl Microbiol Biotechnol* (2001); 55:563-565.
[12] Irie T., Honda Y., Hirano T., Sato T., Enei H., Watanabe T. and Kuwahara M. Stable transformation of *Pleurotus ostreatus* to hygromycin B resistance using Lentinus edodes GPD expression signals. *Appl Microbiol Biotechnol* (2001); 56(5-6):707-709.
[13] Kothe E. Mating-type genes for basidiomycete strain improvement in mushroom farming. *Appl Microbiol Biotechnol* (2001); 56(5 -6):602-12.
[14] Kaur J., Sodhi H.S. and Kapoor S. Breeding of *Pleurotus florida* (oyster mushroom) for phenotypic pigmentation and high yield potential. *Journal of the Science of Food and Agriculture* (2008); 88:2676-2681.
[15] Xu J., Horgen P.A., Anderson J.B. Somatic recombination in the cultivated mushroom *Agaricus bisporus. Mycol Res* (1996); 100:188–192.
[16] Chaudhary A., Dwivedi R.R. and Singh R.P. Evolution and evaluation of hybrids of oyster mushroom, in Proceedings of an International Conference on Mushroom Biology and Biotechnology, NRCM, Solan, India. Mushroom Society of India Publication, Solan, H.P., India, (2007):51.
[17] Sawashe S.G. and Sawant D.M. Studies on evaluation of new strains of Pleurotus spp. obtained through hybridization. In Proceedings of an International Conference on Mushroom Biology and Biotechnology, NRCM, Solan, India,. Mushroom Society of India Publication, Solan, H.P., India, (2007):52.
[18] Eichlerov´a I., Homolka L. Preparation and crossing of basidiospore-derived monokaryons useful tool for obtaining laccase and other ligninolytic enzyme higher-producing dikaryotic strains of *Pleurotus ostreatus*. *Antonie van Leeuwenhoek.* (1999); 75:321-327.
[19] Palmieri G., Giardina P., Bianco C., Scaloni A., Capasso A., Sannia G. A novel white laccase from *Pleurotus ostreatus*. *The Journal of Biological Chemistry.* (1997); 272:31301-31307.
[20] Das N., Sengupta S., Mukherjee M. Importance of laccase in vegetative growth of *Pleurotus florida*, *Appl Environ Microbiol.* (1997); 63:4120-4122.
[21] Giardina P., Aurilia V., Cannio R., Marzullo L., Amoresano A., Siciliano R., Pucci P., Sannia G. The gene, protein and glycan structures of laccase from *Pleurotus ostreatus*. *Eur J Biochem.* (1996); 235:508-515.
[22] Armenante A., Longobardi S., Rea I., De Stefano L., Giocondo M., Silipo A., Molinaro A., Giardina P. The *Pleurotus ostreatus* hydrophobin Vmh2 and its interaction with glucans, *Glycobiology.* (2010); 20:594-602.
[23] Peñas M.M., Asgeirsdóttir S.A., Lasa I., Culiañez-Macià F.A., Pisabarro A.G., Wessels J.G., Ramírez L. Identification, characterization, and in situ detection of a fruit-body-specific hydrophobin of *Pleurotus ostreatus*. *Appl Environ Microbiol.* (1998); 64:4028-4034.
[24] Venkateshwarlu G., Chandravadana M.V., Meera P., Tewari R.P., Selvaraj Y. Volatile flavour compounds from oyster mushroom (*Pleurotus florida*) in submerged culture. *Flavour and fragrance journa.l* (2001); 15:320-322.
[25] Giardina P., Cannio R., Martirani L., Marzullo L., Palmieri G. and Sannia G. Cloning and sequencing of a laccase gene from the lignin-degrading basidiomycete *Pleurotus ostreatus. Applied and Environmental Microbiol.* (1995); 61: 2408–2413.

[26] Soden D.M. and Dobson A.D. Differential regulation of laccase gene expression in *Pleurotus sajor-caju. Microbiology* (2001); 147:1755-1763.
[27] Elisashvili V., Kachlishvili E. Physiological regulation of laccase and manganese peroxidase production by white-rot basidiomycetes. *J Biotechnol.* (2009); 144:37-42.
[28] Palmieri G., Giardina P., Sannia G. Laccase-mediated Remazol Brillant Blue R decolourization in a fixed-bed bioreactor. *Biotechnol Prog* (2005); 21(5):1436-41.
[29] Kunamneni A., Camarero S., Garcı´a-Burgos C.' Plou F.J., Ballesteros A., Alcalde M. Engineering and applications of fungal laccases for organic synthesis. *Microb Cell Fact* . (2008); 7:32.
[30] Mayer A.M., Staples R.C. Laccase: new functions for an old enzyme. *Phytochemistry*. (2002); 60:551–565.
[31] Galhaup C. and Haltrich D. Enhanced formation of laccase activity by the white-rot fungus *Trametes pubescens* in the presence of copper. *Applied Microbiology and Biotechnology*. (2001); 56, 225–232.
[32] Hou H., Zhou J., Wang J., Du C., Yan B. Enhancement of laccase production by *Pleurotus ostreatus* and its use for the decolorization of anthraquinone dye. *Process Biochem*. (2004); 39:1415–1419.
[33] Minussi R., Pastore G., Dura´n N. Laccase induction in fungi and laccase/N-OH mediator systems applied in paper mill effluent. *Bioresour Technol*. (2007); 98:158–164.
[34] Eric U. Selker, Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? *Trends in Genetics*. (1997); 13(8): 296-301.

Table 1: Maximum laccase production of basidiospore-derived monokaryons: (a) monokaryons A from parental strain *P.ostreatus* MYA-2306; (b) monokaryons D from parental strain *P.ostreatus* D1208

(a)

Strain	Activity (U/mL)
	9th day
Parental strain MYA-2306	**17.3**
1A	1
2A	3.36
3A	2.6
4A	2.44
5A	**36.6**
6A	**17.10**
7A	17.11
8A	7.33
9A	3.18
11A	9.17
13A	2.66
15A	12.5
17A	2.08
18A	**26.3**

(b)

Strain	Activity (U/mL)
	9th day
Parental strain D1208	**15.8**
D1	0.65
D2	4.2
D3	**5.1**
D4	0.015
D5	0.03
D6	3.3
D7	1.3
D8	1.02
D9	**16.6**
D10	1.21
D11	**7.8**
D12	0.83
D13	1.49
D14	0.38
D15	2.62

Table 2: Compatibility tables of selected higher laccase producers monokaryons: A X D crossing

Strain	3D	9D	11D
5A	+	+	+
6A	+	+	+
18A	+	-	+

Table 3: L**accase isoenzymes produced by dikaryotic hybrids cultures. Fractionation was performed by anionic exchange chromatography ResourceQ. Relative quantification of laccase isoenzymes were carried out in correspondence of laccase activity peak (9th day), in basal condition.**

Strains	Activity U/L (day)	Isoenzymatic	
5AxD3	**47,000 ±18,000 (9)**	POXA3	<1%
		POXC	99%
6AxD11	**53,000 ± 13,000 (9)**	POXA1b	2%
		POXA3	15%
		POXC	48%
		POX1	35%

5AxD11	**45,000 ± 8,000 (9)**	POXA3	2,2%
		POXC	97,8%
Parentale	**17,000 ± 2,000 (9)**	POXA3	20%
		POXC	80%

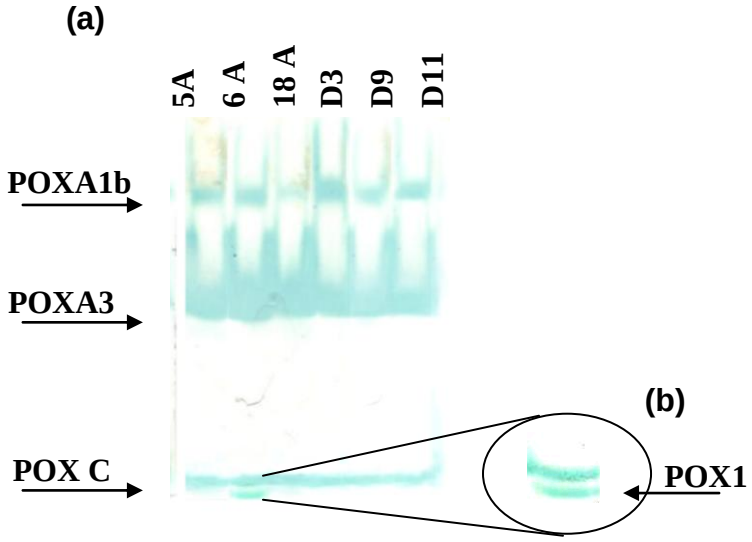


Figure 1: **Zymograms of laccase isoenzymes produced by the selected monokaryotic strains in PDY culture broth. Samples containing 0.015 U of laccase activity collected in the day of maximum production (9th days) were used. Known laccase isoforms POXA1b, POXA3 and POXC were loaded as standards. (b) Zoom of POX1 laccase identified in the monokaryotic strain 6A.**

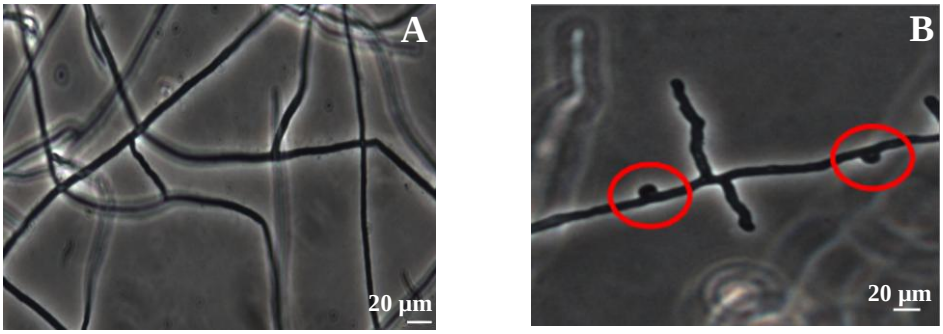


Figure 2: (A) *P. ostreatus* monokaryotic hyphae observed on optical microscope (×45 magnification); (B) *P. ostreatus* dikaryotic hyphae having clamp connections. Clamp structure is highlighted in red.

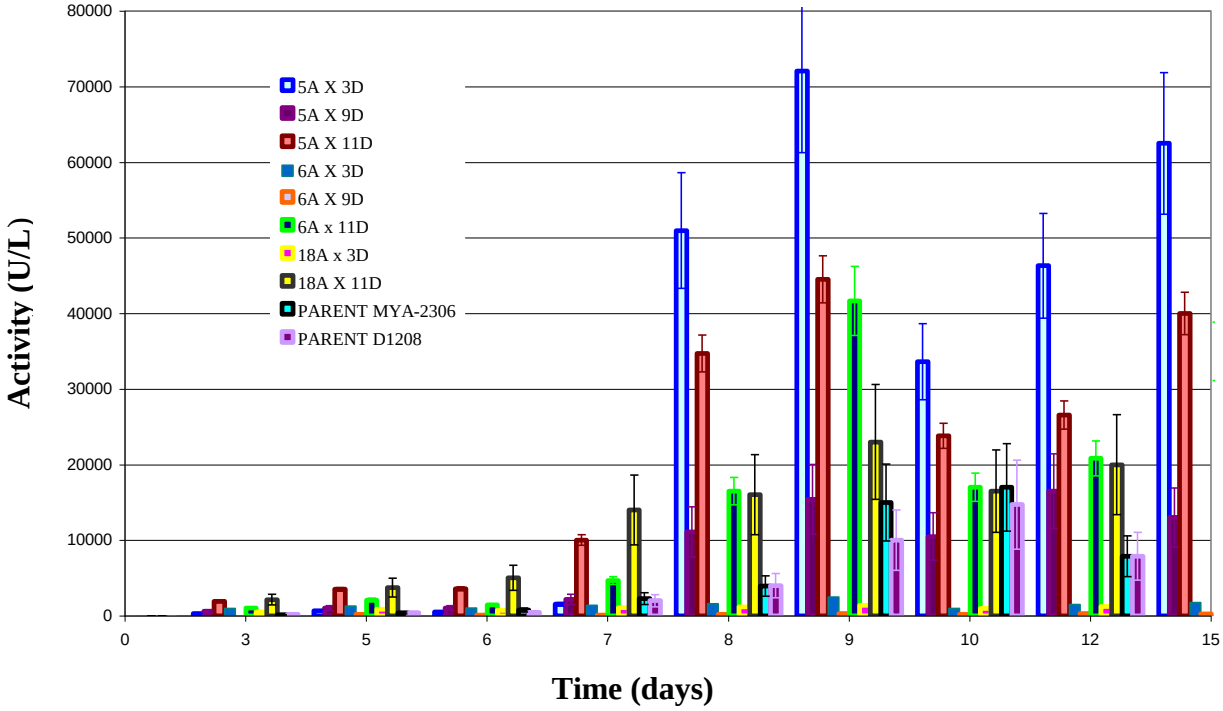


Figure 3: Time course analysis on extracellular laccase activity produced by dikaryotic strains, in basal condition (PDY+150 µMCuSO$_4$).

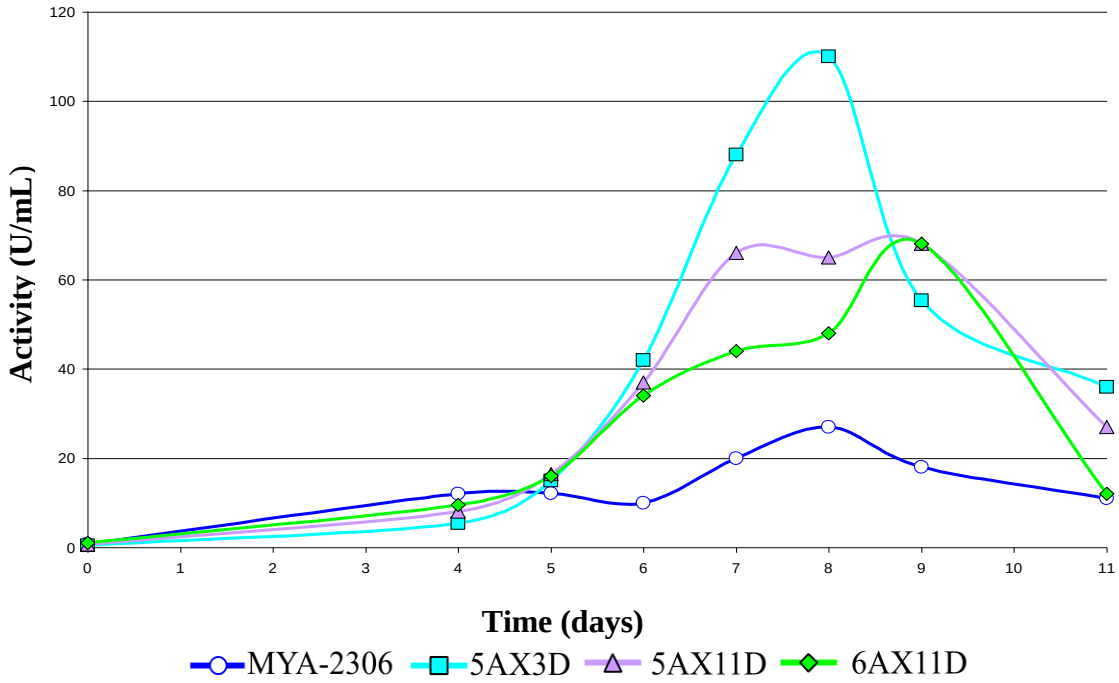


Figure 4: Time course analysis on extracellular laccase activity produced by dikaryotic strains, in addiction of ferulic acid (2mM final concentration).

British Mycological Society promoting fungal science

journal homepage: www.elsevier.com/locate/funbio

Identification of a new member of *Pleurotus ostreatus* laccase family from mature fruiting body

Vincenzo LETTERA, Alessandra PISCITELLI, Gabriella LEO, Leila BIROLO, Cinzia PEZZELLA, Giovanni SANNIA*

Department of Organic Chemistry and Biochemistry, University of Naples Federico II, Complesso Universitario Monte S. Angelo – Via Cinthia 4, IT-80126 Napoli, Italy

ARTICLE INFO

Article history:
Received 23 February 2010
Received in revised form
3 June 2010
Accepted 4 June 2010
Available online 17 June 2010
Corresponding Editor: Marc Stadler

Keywords:
Fruiting body
Fungi
Laccase family
Physiological role

ABSTRACT

Laccases (benzenediol:oxygen oxidoreductases, EC 1.10.3.2) are blue multicopper oxidases, catalyzing the oxidation of an array of aromatic substrates concomitantly with the reduction of molecular oxygen to water. Most of the known laccases have fungal or plant origins, although few laccases have been also identified in bacteria and insects. Most of the fungal laccases reported thus far are extra-cellular enzymes, whereas only few enzymes from fruiting bodies have been described so far. Multiple isoforms of laccases are usually secreted by each fungus depending on species and environmental conditions. As a fact, a laccase gene family has been demonstrated in the white-rot fungus *Pleurotus ostreatus*.

This work allowed identification and characterization of the first laccase isoenzyme from the fruiting body of *P. ostreatus*. Discovery through mass spectrometry of LACC12 proves the expression of a functional protein by the related deduced encoding transcript. The topology of phylogenetic tree of fungal laccases proves that LACC12 falls in cluster with the members of *P. ostreatus* LACC10 (=POXC) subfamily, although *lacc12* deduced intron–exon structure differs from that of the subfamily members and the related locus is located in a different chromosome. Results show that the evolutionary pattern of *lacc12* and that of the other laccase isozyme genes may have evolved independently, possibly through duplication–divergence events.

The reported data add a new piece to the knowledge about *P. ostreatus* laccase multigene family and shed light on the role(s) played by individual laccase isoforms in *P. ostreatus*.

Introduction

Laccase is one of the very few enzymes that have been studied since the end of 19th century. It was first identified in the exudates of *Rhus vernicifera*, the Japanese lacquer tree (Yoshida 1883). A few years later its presence was also demonstrated in fungi (Bertrand 1896). Laccases are thought to be nearly ubiquitous among fungi, and their presence has been documented in virtually every fungus examined thus far (Mayer & Staples 2002). More recently, proteins with features typical of laccases have been identified in insects (Dittmer *et al.* 2004) and prokaryotes (Claus 2003).

Laccases are multicopper-containing enzymes belonging to the group of blue oxidases, along with ascorbic oxidases and ceruloplasmins (Giardina *et al.* 2010). These enzymes are being increasingly evaluated for a variety of

* *Corresponding author*. Tel.: +39081674338; fax: +39081674313.
E-mail addresses: vincenzo.lettera@unina.it, apiscite@unina.it, gabriella.leo@unina.it, birolo@unina.it, cpezzella@unina.it, sannia@unina.it

1878-6146/$ – see front matter
doi:10.1016/j.funbio.2010.06.004

biotechnological applications due to their broad substrate range. The catalysis carried out by all members of this family is guaranteed by the presence of different copper centres. In particular, all blue multicopper oxidases (MCOs) are characterized by the presence of one type-1 (T1) copper, together with at least three additional copper ions: one type-2 (T2) and two type-3 (T3) copper ions, arranged in a trinuclear cluster.

Fungal laccases have been associated with delignification (Baldrian 2004), fruiting body structuring (Chen *et al.* 2004), pigment formation during asexual development (Tsai *et al.* 1999), and pathogenesis (Litvintseva & Henson 2002; Missal *et al.* 2005). The majority of fungal laccases are extra-cellular monomeric globular proteins of approximately 60–70 kDa with acidic isoelectric point (pI) around pH 4.0; they are generally glycosylated, with an extent of glycosylation ranging between 10 % and 25 % and only in a few cases higher than 30 % (Giardina *et al.* 2010). As a fact, to the best of our knowledge, among fungal laccases characterized thus far, only few of them have been reported as enzymes specifically purified from fruiting bodies. Laccases have been isolated from fruiting bodies of *Lentinula edodes* (Nagai *et al.* 2003), *Cantharellus cibarius* (Ng & Wang 2004), *Hericium erinaceum* (Wang & Ng 2004a), *Albatrella dispansum* (Wang & Ng 2004b), *Tricholoma giganteum* (Wang & Ng 2004c), *Ganoderma lucidum* (Wang & Ng 2006a), *Pleurotus eryngii* (Wang & Ng 2006b), *Pleurotus cornucopiae* (Wong *et al.* 2009), and *Ganoderma* sp. MK05 (Khammuang & Sarnthima 2009).

Many fungi produce several laccase isoenzymes endowed with different catalytic properties, being the physiological significance of this multiplicity still unknown. Likewise, fungi belonging to *Pleurotus* genera display this feature (Giardina *et al.* 2010). As a fact, four laccase gene members have already been isolated in *Pleurotus sajior-caju* (Soden & Dobson 2001), two in *P. eryngii* (Rodríguez *et al.* 2008), and seven in *Pleurotus ostreatus* (Pezzella *et al.* 2009). In this latter fungus, the existence of a "laccase subfamily" consisting of three members has been postulated, based on sequence similarity, and intron–exon structure. Furthermore, a careful investigation of the recently released genome of this fungus by the DOE – Joint Genome Institute (http://www.jgi.doe.gov/sequencing/why/50009.html, DOE – JGI) suggested an even more complex MCO family. New genes, putatively coding for previously uncharacterized laccases, have been identified, enriching the panel up to 12 members (unpublished data).

In this work the identification of a new laccase isoenzyme from the fruiting body of the white-rot fungus *P. ostreatus* has been achieved through the detection of six tryptic peptides by mass spectrometry. The identity of the functional laccase protein has been confirmed with the translated genomic sequence of *P. ostreatus* laccases.

Experimental/materials and methods

Organism and culture conditions

White-rot fungus *Pleurotus ostreatus* (type: *Florida* ATCC no. MYA-2306) was maintained through periodic transfer at 4 °C on potato dextrose agar plates (Difco) in the presence of 0.5 % yeast extract (Difco). Inoculum for liquid culture was prepared by pre-inoculating 300 mL of PDY broth in a shaken flask with *P. ostreatus* mycelia at 28 °C. Then, 50 mL of a 5-d-old culture was transferred to 1-L flasks containing 450 mL of PDY broth, supplemented with different inducer (150 μM $CuSO_4$, 1 mM vanillic acid dissolved in ethanol, and 2 mM ferulic acid dissolved in ethanol and added at the second day of growth). Cultures were incubated in the dark at 28 °C on a rotary shaker at 120 rpm.

Fruiting body production

Pleurotus ostreatus mushrooms were cultivated in 500 mL jars containing 400 g of wheat-straw (65 % water content), which were twice autoclaved for 1 h at 121 °C separated by a period of 24 h at room temperature. Each jar was inoculated with four agar plugs (13 mm diameter), and left to grow at 28 °C for 30 d in the dark. Fructification were promoted by opening the jars, and placing them in presence of daylight in a chamber at 15 ± 5 °C and 90 % relative humidity. Primordia appeared after a further 15 d of growth, and basidiocarps were harvested 7 d later and weighed.

Laccase extraction

Total extracts from the mycelium of *Pleurotus ostreatus* grown on wheat-straw were prepared as follows: lyophilized cells were ground in a mortar with a pestle. The ground material was resuspended in cold extraction buffer (200 mM Tris–HCl pH 8, 400 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 1 mM EDTA, 10 % glycerol, 1 mM phenylmethanesulphonylfluoride (PMSF)) followed by centrifugation at 4 °C for 1 h at 13 000×g.

Fungal proteins were extracted by homogenizing *P. ostreatus* fruiting bodies in a Waring blender in the presence of 50 mM sodium phosphate buffer (pH 6.5) containing 1 mM PMSF.

Laccase purification

Fungal proteins extracted by fruiting bodies were centrifuged at 13 000×g for 30 min at 4 °C, and the supernatant was filtered through cheesecloth. Proteins were precipitated by addition of $(NH_4)_2SO_4$ up to 80 % saturation at 4 °C and centrifuged at 10 000×g for 40 min. The ammonium sulphate precipitate was resuspended in 50 mM sodium phosphate (pH 6.5) and after extensive dialysis against the same buffer was loaded onto a Hi Trap DEAE Fast Flow column (GE Healthcare) in a fast protein liquid chromatography system (AKTA Explorer, Amersham Biosciences). Four fractions containing laccase activity were recovered with the equilibrating buffer and during the linear gradient from 0 to 0.4 M NaCl. They were separately pooled, concentrated and desalted on an Amicon PM-10 membrane. Fraction 3 was equilibrated in 50 mM Tris–HCl pH 8.0 and loaded onto an anion exchange Resource Q column (Amersham Biosciences) equilibrated with the same buffer. Three fractions containing laccase activity were recovered during the linear gradient from 0 to 0.4 M NaCl. They were separately pooled, concentrated and desalted on an Amicon PM-10 membrane.

Protein determination

Protein concentration was determined using the Bio-Rad protein assay with bovine serum albumin as standard.

Assay of laccase activity

Phenol-oxidase activity was assayed at 25 °C using 2,2-azinobis-(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS), 2,6-dimethoxyphenol (DMP) and syringaldazine as substrates, in the condition previously optimized (Palmieri *et al.* 1997) as below indicated.

The assay mixture contained 2 mM ABTS and 0.1 M sodium citrate buffer, pH 3.0. Oxidation of ABTS was followed by absorbance increase at 420 nm ($\varepsilon = 36\,000\ M^{-1}\,cm^{-1}$). The assay mixture contained 1 mM DMP and the McIlvaine's buffer adjusted to pH 5.0. Oxidation of DMP was followed by an absorbance increase at 477 nm ($\varepsilon = 14\,800\ M^{-1}\,cm^{-1}$). The assay mixture contained 0.5 mM syringaldazine (dissolved in methanol) and 50 mM phosphate buffer, pH 6.0 Oxidation of syringaldazine was followed by an absorbance increase at 525 nm ($\varepsilon = 65\,000\ M^{-1}\,cm^{-1}$). Enzyme activity was expressed in IU. K_M were calculated assaying laccase activity in these conditions by nonlinear regression fitting with the GraphPad Prism version 3.00 for Windows, GraphPad Software, San Diego, CA, USA (www.graphpad.com).

Native polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE) was performed at alkaline pH under non-denaturing conditions. The resolving and stacking gels were, respectively, at 9 and 4 % acrylamide, corresponding buffer solutions were 50 mM Tris–HCl (pH 9.5) and 18 mM Tris–HCl (pH 7.5) respectively. The electrode reservoir solution was 25 mM Tris, 190 mM glycine, pH 8.4. Gels were stained for laccase activity using ABTS as substrate. Samples containing 0.015 laccase unities were loaded on each lane.

Protein identification by mass spectrometry

Protein band corresponding to the laccase of interest was excised from the native gel. Slices were destained by washes with 0.1 M NH_4HCO_3 pH 7.5 and acetonitrile, reduced for 45 min with 100 μL of 10 mM dithiothreitol in 0.1 M NH_4HCO_3 buffer pH 7.5 and carboxyamidomethylated for 30 min in the dark with 100 μL of 55 mM iodoacetamide in the same buffer. Tryptic digestion was performed by adding for each slice 100 ng of enzyme in 10 μL of 10 mM NH_4HCO_3 pH 7.5 for 2 h at 4 °C. The buffer solution was then removed and 50 μL of 10 mM NH_4HCO_3 pH 7.5 were added for 18 h at 37 °C. Peptides were extracted by washing the gel slices with 10 mM NH_4HCO_3 and 1 % formic acid in 50 % acetonitrile at room temperature.

The peptide mixtures were filtered by using 0.22 μm PVDF membrane (Millipore) and analyzed using a 6520 Accurate-Mass Q-TOF LC/MS System (Agilent Technologies, Palo Alto, CA) equipped with a 1200 HPLC system and a chip cube (Agilent Technologies). After loading, the peptide mixture was first concentrated and washed in 40 nL enrichment column (Agilent Technologies chip), with 0.1 % formic acid in 2 % acetonitrile as the eluent. The sample was then fractionated on a C18 reverse-phase capillary column (Agilent Technologies chip) at flow rate of 400 nL/min, with a linear gradient of eluent B (0.1 % formic acid in 95 % acetonitrile) in A (0.1 % formic acid in 2 % acetonitrile) from 7 % to 80 % in 50 min. Peptide analysis was performed using data-dependent acquisition of one MS scan (mass range from 300 to 1800 m/z) followed by MS/MS scans of the five most abundant ions in each MS scan. Dynamic exclusion was used to acquire a more complete survey of the peptides by automatic recognition and temporary exclusion (0.5 min) of ions from which definitive mass spectral data had previously acquired. Nitrogen at a flow rate of 3 L/min and heated to 325 °C was used as the dry gas for spray desolvation. MS/MS spectra were measured automatically when the MS signal surpassed the threshold of 50 000 counts. Double and triple charged ions were preferably isolated and fragmented over single charged ions. The acquired MS/MS spectra were transformed in *mz.data* format and used for proteins identification with a licensed version of MASCOT 2.1 (Matrix Science, Boston, USA). Raw data from nanoLC–MS/MS analyses were used to query the *Pleurotus ostreatus* database (*Filtered Models ("best")* Proteins: PleosPC15_1_GeneModels_Filtered Models5_aa.fasta.gz – http://genome.jgi-psf.org/PleosPC15_1/PleosPC15_1.download.ftp.html) and the Mascot search parameters were: trypsin as enzyme, allowed number of missed cleavage 3, carbamidomethyl as fixed modification, oxidation of methionine, pyro-Glu N-term Q, as variable modifications, 10 ppm MS tolerance and 0.6 Da MS/MS tolerance, peptide charge, from +2 to +3. Individual ions scores >32 indicate identity or extensive homology ($p < 0.05$). Protein scores are derived automatically by MASCOT Software from ions scores as a non-probabilistic basis for ranking protein hits.

Analysis of protein sequence

The predicted protein sequence of newly identified laccase LACC12 was analyzed by comparison with those of the previously defined *Pleurotus ostreatus* laccases. Alignments of amino-acidic sequences were generated with ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Signal peptide was predicted with SignalP V2.0 (http://www.cbs.dtu.dk/services/). The amino-acid conservation scoring among the laccase signature sequences was performed by PRALINE (http://www.ibi.vu.nl/programs/pralinewww/).

Phylogenetic analysis

Phylogenetic analysis was carried out on *Pleurotus ostreatus* laccases using basidiomycete laccase sequences reported in Hoegger *et al.* (2006) and more recently released laccase sequences. The sequences were aligned with ClustalW using default settings for multiple sequence alignments. Based on this alignment, a Neighbour Joining tree was constructed using MEGA version 4 (Tamura *et al.* 2007). Bootstrapping was carried out with 1000 replications.

Results

Whole proteins extracted from fruiting bodies were analyzed by native electrophoresis gel and four laccase isoforms could

be identified. Besides the already studied laccases (LACC10, LACC6 (=POXA1b) and LACC2 (=POXA3)) (Palmieri *et al.* 2003), one new isoform, exclusively present in fruiting body extract, was also detected (Fig 1A, lane 8).

In order to characterize this new isoform, phenol-oxidases extracted from fruiting body were successfully separated into four distinct pools of protein fractions on a DEAE Sepharose column at pH 6.5. When analyzed by native PAGE, the pool of active fractions eluted at 0.2 M NaCl included the new isoenzyme (Fig 1B, lane 2). This new isoform was then successfully separated from LACC10 and LACC2 enzymes by anion exchange chromatography Resource Q column (Amersham Biosciences) (Fig 1B, lane 3).

When the protein band, corresponding to the new laccase, was excised and analyzed by tandem mass spectrometry, six peptides (sequence coverage, 18 %; protein summary score, 282) could unambiguously be identified. *Pleurotus ostreatus* database search with the raw MS/MS data unambiguously identified the protein encoded by the predicted laccase transcript *lacc12*.

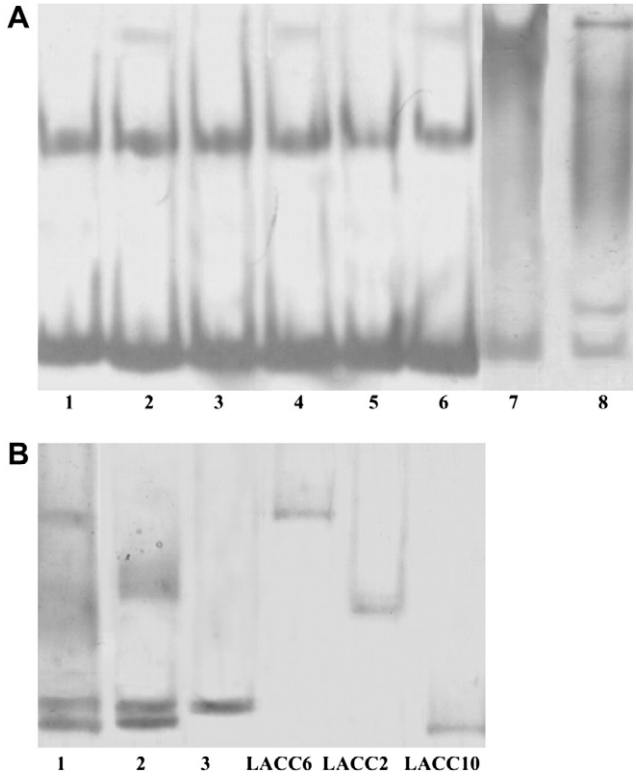


Fig 1 – A. Zymograms of *Pleurotus ostreatus* laccase isoenzymes. Laccases from mycelium grown in the presence of: no inducer (lane 1); $CuSO_4$ (lane 2); ferulic acid (lane 3); $CuSO_4$ and ferulic acid (lane 4); vanillic acid (lane 5); $CuSO_4$ and vanillic acid (lane 6); laccases extracted from the fungus overgrown on wheat-straw: active mycelium (lane 7); and fruiting body (lane 8) – B. Zymograms of fruiting body laccases. Ammonium sulphate precipitated (lane 1); fraction eluted at 0.2 M NaCl from DEAE Sepharose column (lane 2); laccase after Resource Q column (lane 3). Known laccase isoforms LACC6, LACC2, and LACC10 are loaded as standards.

LACC12 predicted protein sequence contains N-terminal secretion signal (23 amino-acids), conforms to the typical sequences of eukaryotic proteins (Nielsen *et al.* 1997), the conserved 10 histidine and 1 cysteine residues of the copper-binding centres of laccases, the fungal laccase signature sequences L1–L4 (Fig 2A), as defined by Kumar *et al.* (2003). Compared to the defined signature sequences, the only relevant substitutions observed in LACC12 are a phenylalanine to tyrosine change in L1, and a valine to isoleucine change in L4. Sequence of the substrate binding loops, previously described for other isoforms of the same family (Pezzella *et al.* 2009), are not conserved among *P. ostreatus* proteins (Fig 2B).

The molecular mass of LACC12, predicted from its whole deduced amino-acid sequence as 55 kDa, was confirmed by SDS-PAGE followed by Western-blotting analysis (data not shown).

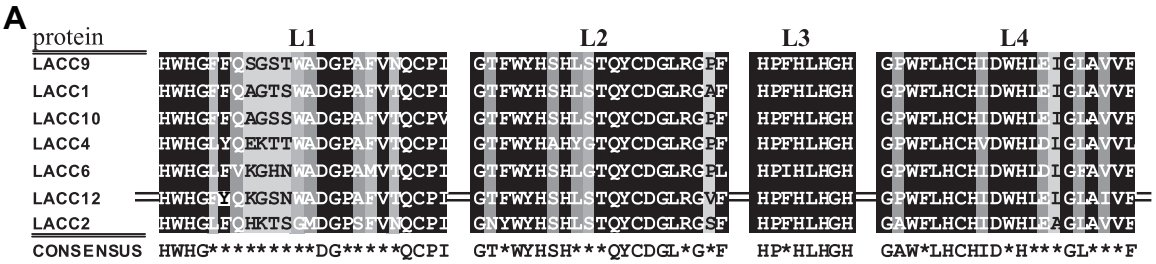
Kinetic proprieties of LACC12 with respect to three substrates (ABTS, DMP, and syringaldazine) were determined and compared with those of the other known members of the *P. ostreatus* laccase isoenzyme family (Table 1).

Discussion

Pleurotus ostreatus is an active lignin degrader that lives as a saprophyte on dead or decaying wood (Moncalvo *et al.* 2002). Within its ligninolytic enzymatic system, 7 members of the *P. ostreatus* laccase multigene family have been isolated and sequenced (Pezzella *et al.* 2009). Furthermore, up to 12 members of this family have been identified in the recently released *P. ostreatus* genome database (unpublished data). However, the redundancy of laccase genes raises the question about their respective functions *in vivo*, and this question is even more pertinent since cDNAs or proteins for some of these genes have not been found yet (Pezzella *et al.* 2009).

Laccases LACC10, LACC6, and LACC2 constitute the typical isoenzymatic pattern expressed by *P. ostreatus* in any condition analyzed thus far (*i.e.* solid/liquid state, presence of laccase inducers, *etc.*) (Palmieri *et al.* 2000; Giardina *et al.* 2002; Palmieri *et al.* 2005). In the present work a new laccase protein was isolated from *P. ostreatus* fruiting body and unambiguously identified as LACC12 through MS/MS analysis and querying of the *P. ostreatus* genome data bank. Discovery of LACC12 proves the expression of a functional protein by the related hypothetical encoding transcript. The new laccase isoform LACC12 seems to be exclusively expressed in the fruiting body stage as we have never detected it in the active mycelium, even when it is overgrown on the lignocellulosic substrate wheat-straw (Fig 1A, lanes 1–7). Thus, a specific role for this isoenzyme, possibly involved in the development of fruiting body, may be hypothesized. Indeed, a particular physiological role in gill browning for the fruiting body laccase LCC4 has been assumed in *Lentinula edodes* (Nagai *et al.* 2003). On the other hand, laccase genes in *Volvariella volvacea* (Chen *et al.* 2004) and *Laccaria bicolor* (Courty *et al.* 2009) are differentially expressed as a function of the developmental stage.

LACC12 affinity (K_M value) towards typical laccase substrates follows the rank order syringaldazine > DMP > ABTS (Table 1). It is almost of the same order of magnitude respect to those of the other known members of



B

	Substrate binding loop					
	Loop I		Loop II	Loop III	Loop IV	
LACC9	VVAP------QNAVLPT-	MSCDPN	ADPNL----GSTGF	MAFDVTNFELTI	IPALAVGGP	HIDWHLEIGL
LACC1	VVAP------QNGPIPT-	LSCDPN	AQPNL----GTVGY	LNFDFTTFEMTI	MPALAVGGP	HIDWHLELGL
LACC10	IVAP------QNAAIPT-	MSCDPN	ANPNL----GSTGF	MAFDFTTFELTI	MPALAVGGP	HIDWHLELGL
LACC4	PTAE----ELVARGGPPQ	IACEPN	GNPSA----GPTGF	FSFDPMTSRFAT	MPGGVNGGP	HVDWHLDLGL
LACC6	APSL------SLTGVPH-	TSCDSN	ANPNS----GDPGF	FAFDPATALFTA	MPALVFAGP	HIDWHLDLGF
LACC12	VPAP------SAGLVPV-	ISCDPN	ASPNL----GPQGF	IAFDEEKFKFSI	IPGLATGGP	HIDWHLELGL
LACC2	DLAPHAQNQFFQTGSVPI	ISCRPF	APLTGGNPAGNPNL	IAQPNPPF-FDI	IP---GAGA	HIDWHLEAGL

Fig 2 – A. Comparison of the laccase signature sequences from the *Pleurotus ostreatus* laccases LACC9, LACC10, LACC6, LACC2, LACC4, and LACC1 with LACC12. The fungal laccase signature sequences (L1–L4) are shown together with the fungal laccase consensus sequence. The amino-acid conservation scoring is performed by PRALINE (http://www.ibi.vu.nl/programs/pralinewww/). Residue shading has been applied from scoring 10 (black) to 7 (grey). The Y97 residue in L1 is boxed - B. Sequence alignment of the potential substrate binding loops of the *P. ostreatus* laccases. Loops L1 sequence at positions 182–192, B4–B5 at positions 230–235, B7–B8 at positions 288–297, C1–C2 at positions 357–368, C4–C5 at positions 413–421 and C7–C8 at positions 479–488.

the family, in agreement with what reported for other known fungal laccases (Baldrian 2006).

The protein sequence contains all the L1–L4 laccase consensus signatures and can be classified as a laccase in *sensu stricto*. LACC12 sequence displays the substitution phenylalanine to tyrosine (F97Y), already found in LACC4 (=POX3) (Pezzella *et al.* 2009) and in laccases belonging to *Coprinopsis cinerea* subfamily (Kilaru *et al.* 2006), being this a conserved position among many basidiomycete laccase sequences. The substitution of valine to isoleucine in loop 4, present in LACC12 in position 490, has also been found in the laccase sequences of *Coprinus cinereus* (Ducros *et al.* 2001), *Trametes versicolor* (Piontek *et al.* 2002) and *C. cinerea* (Kilaru *et al.* 2006). LACC12 shares the highest protein sequence identity (68 % identity) with a laccase from *Cyathus bulleri* (ABW75771) (Salony *et al.* 2008) and with laccase 2 from *C. cinerea* (AAR01243) (Kilaru *et al.* 2006). As far as *P. ostreatus* laccase family is concerned, the new laccase LACC12 shares the highest percents of identity with members of LACC10 subfamily (Pezzella *et al.* 2009), 68 % with LACC1 (=POX4), and 67 % with LACC9 (=POX1) and LACC10. As a fact, based on Neighbour Joining tree of basidiomycete laccases (Fig 3), LACC12 falls in cluster with the members of *P. ostreatus* LACC10 subfamily. On the contrary, this enzyme is more distantly related to the *L. edodes* fruiting body laccase LCC4 (Nagai *et al.* 2003).

Table 1 – Affinity constants of purified LACC12 towards ABTS, syringaldazine and DMP substrates compared to those of the other members of the isoenzyme family.

	LACC12	LACC10	LACC6	LACC2
K_M ABTS (mM)	15×10^{-2}	3.9×10^{-2}	47×10^{-2}	7.4×10^{-2}
K_M DMP (mM)	27×10^{-2}	0.76×10^{-2}	38×10^{-2}	880×10^{-2}
K_M syringaldazine (mM)	5.5×10^{-2}	2.0×10^{-2}	22×10^{-2}	7.9×10^{-2}

Concerning *P. ostreatus* laccase genes distribution in the genome, a cluster of 7 out of 12 laccase genes has been recently mapped by Pérez *et al.* (2009) in the sub-telomeric region of chromosome VI. The existence of MCO genes clustering has also been observed in the fungi *Phanerochaete chrysosporium* (Larrondo *et al.* 2004), *Agaricus bisporus* (Smith *et al.* 1998), *Rhizoctonia solani* (Wahleithner *et al.* 1996) and *C. cinerea* (Kilaru *et al.* 2006). Since sub-telomeric regions are genome areas where important rearrangements and gene regulation mechanisms occur, Pérez *et al.* (2009) suggested an evolutionary relationship among laccases located in this region, permitting the fungus to adapt rapidly to new lignocellulosic substrates. While all the members of the LACC10 subfamily lie in the gene cluster of the sub-telomeric region of chromosome VI, *lacc12* locus is located in a different chromosome (XI) (Antonio G. Pisabarro, pers. comm.). Moreover, deduced *lacc12* gene structure differs from that of the members of LACC10 subfamily. Thus, LACC12, specifically expressed in fruiting bodies, might have evolved independently from the subfamily, possibly through duplication–transposition events, probably to fulfil a key role in basidiocarp development.

This study shows for the first time the identification of a laccase from the fruiting body of *P. ostreatus*. The reported data add a new piece to the knowledge about *P. ostreatus* laccase multigene family. The enzymes – secreted and non-secreted forms – even though coming from the same source

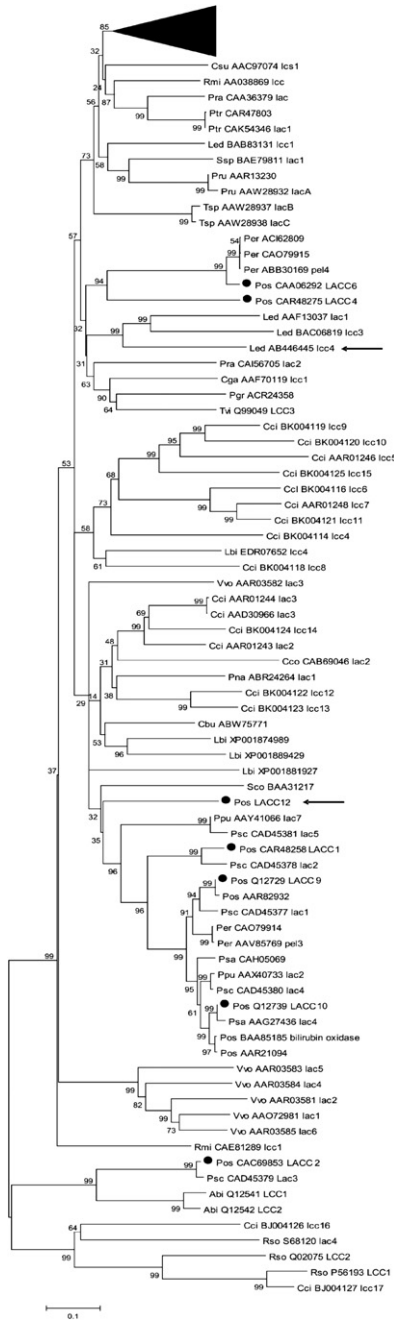


Fig 3 – Neighbour Joining tree of laccase amino-acid sequences. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino-acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). Phylogenetic analyses were conducted in MEGA4. *Pleurotus ostreatus* laccases are labelled with a black circle. LACC12 from *P. ostreatus* and LCC4 from *Lentinula edodes* are indicated with a black arrow.

might not share the same characteristics and/or may play different physiological roles. A further progress towards a more intelligible picture of the role(s) played by individual laccase isoforms in *P. ostreatus* has been attained with this work.

Acknowledgements

This work was supported by the European Commission, Sixth Framework Program (QUORUM, 032811), by grants from the Ministero dell'Università e della Ricerca Scientifica (Progetti di Rilevante Interesse Nazionale, PRIN), from Compagnia di San Paolo, Turin, Italy, project "Sviluppo di procedure di biorisanamento di reflui industriali (BIOFORM)", from COST Action FP0602 "Biotechnology for lignocellulose biorefineries" (BIOBIO), and from the Ministero Degli Affari degli Esteri di Intesa con il Ministero dell'Universita' e della Ricerca (Progetti di ricerca di base e tecnologica approvati nei protocolli di cooperazione scientifica e tecnologica bilaterale come previsto dal protocollo bilaterale tra Italia e Turchia).

REFERENCES

Baldrian P, 2004. Increase of laccase activity during interspecific interactions of white-rot fungi. *FEMS Microbiology Ecology* **50**: 245–253.

Baldrian P, 2006. Fungal laccases occurrence and properties. *FEMS Microbiology Reviews* **30**: 215–242.

Bertrand G, 1896. Sur la presence simultanee de la laccase et de la tyrosinase dans le suc de quelques champignons. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* **123**: 463–465.

Chen S, Ge W, Buswell JA, 2004. Molecular cloning of a new laccase from the edible straw mushroom *Volvariella volvacea*: possible involvement in fruit body development. *FEMS Microbiology Letters* **30**: 171–176.

Claus H, 2003. Laccases and their occurrence in prokaryotes. *Archives of Microbiology* **179**: 145–150.

Courty PE, Hoegger PJ, Kilaru S, Kohler A, Buée M, Garbaye J, Martin F, Kües U, 2009. Phylogenetic analysis, genomic organization, and expression analysis of multi-copper oxidases in the ectomycorrhizal basidiomycete *Laccaria bicolor*. *The New Phytologist* **182**: 736–750.

Dittmer NT, Suderman RJ, Jiang H, Zhu YC, Gorman MJ, Kramer KJ, Kanost MR, 2004. Characterization of cDNAs encoding putative laccase-like multicopper oxidases and developmental expression in the tobacco hornworm, *Manduca sexta*, and the malaria mosquito, *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology* **34**: 29–41.

Ducros V, Brzozowski AM, Wilson KS, Ostergaard P, Schneider P, Svendson A, Davies GJ, 2001. Structure of the laccase from *Coprinus cinereus* at 1.68-Å resolution: evidence for different 'type 2 Cu-depleted' isoforms. *Acta Crystallographica. Section D, Biological Crystallography* **57**: 333–336.

Giardina P, Cennamo G, Greco C, Palmieri G, Sannia G, 2002. Protease mediated processing of a Cu-induced laccase in *Pleurotus ostreatus*: a natural approach to improve protein stability. *Biotechnology in the Pulp and Paper Industry* **21**: 113–120.

Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S, Sannia G, 2010. Laccases: a never-ending story. *Cellular and Molecular Life Sciences: CMLS* **67**: 369–385.

Hoegger PJ, Kilaru S, James TY, Thacker JR, Kues U, 2006. Phylogenetic comparison and classification of laccase and related

multicopper oxidase protein sequences. *The FEBS Journal* **273**: 2308–2326.

Khammuang S, Sarnthima R, 2009. Laccase activities from fresh fruiting bodies of *Ganoderma sp*. MK05: purification and Remazol Brilliant Blue R decolorization. *The Journal of Biological Sciences* **9**: 83–87.

Kilaru S, Hoegger PJ, Kües U, 2006. The laccase multi-gene family in *Coprinopsis cinerea* has seventeen different members that divide into two distinct subfamilies. *Current Genetics* **50**: 45–60.

Kumar SVS, Phale PS, Durani S, Wangikar PP, 2003. Combined sequence and structure analysis of the fungal laccase family. *Biotechnology and Bioengineering* **83**: 386–394.

Larrondo LF, Gonzalez B, Cullen D, Vicuna R, 2004. Characterization of a multicopper oxidase gene cluster in *Phanerochaete chrysosporium* and evidence of altered splicing of the MCO transcripts. *Microbiology* **150**: 2775–2783.

Litvintseva AP, Henson JM, 2002. Cloning, characterization, and transcription of three laccase genes from *Gaeumannomyces graminis* var. *tritici*, the take-all fungus. *Applied and Environmental Microbiology* **68**: 1305–1311.

Mayer AM, Staples RC, 2002. Laccase: new functions for an old enzyme. *Phytochemistry* **60**: 551–565.

Missal TA, Moran JM, Corbett JA, Lodge JK, 2005. Distinct stress responses of two functional laccases in *Cryptococcus neoformans* are revealed in the absence of the thiol-specific antioxidant Tsa1. *Eukaryotic Cell* **4**: 202–208.

Moncalvo JM, Vilgalys R, Redhead SA, Johnson JE, James TY, Aime MC, Hofstetter V, Verduin SJ, Larsson E, Baroni TJ, Thorn RG, Jacobsson S, Clemencon H, Miller Jr OK, 2002. One hundred and seventeen clades of euagarics. *Molecular Phylogenetics and Evolution* **23**: 357–400.

Nagai M, Kawata M, Watanabe H, Ogawa M, Saito K, Takesawa T, Kanda K, Sato T, 2003. Important role of fungal intracellular laccase for melanin synthesis: purification and characterization of an intracellular laccase from *Lentinula edodes* fruit bodies. *Microbiology* **149**: 2455–2462.

Ng TB, Wang HX, 2004. A homodimeric laccase with unique characteristics from the yellow mushroom *Cantharellus cibarius*. *Biochemical and Biophysical Research Communications* **313**: 37–41.

Nielsen H, Engelbrecht J, Brunak S, Von Heijne G, 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *International Journal of Neural Systems* **8**: 581–599.

Palmieri G, Cennamo G, Faraco V, Amoresano A, Sannia G, Giardina P, 2003. Atypical laccase isoenzymes from copper supplemented *Pleurotus ostreatus* cultures. *Enzyme and Microbial Technology* **33**: 220–230.

Palmieri G, Giardina P, Bianco C, Fontanella B, Sannia G, 2000. Copper induction of laccase isoenzymes in the ligninolytic fungus *Pleurotus ostreatus*. *Applied Environmental Microbiology* **66**: 920–924.

Palmieri G, Giardina P, Bianco C, Scaloni A, Capasso A, Sannia G, 1997. A novel white laccase from *Pleurotus ostreatus*. *The Journal of Biological Chemistry* **272**: 31301–31307.

Palmieri G, Giardina P, Sannia G, 2005. Laccase-mediated Remazol Brilliant Blue R decolourization in a fixed-bed bioreactor. *Biotechnology Progress* **21**: 1436–1441.

Pérez G, Pangilinan J, Pisabarro AG, Ramírez1 L, 2009. Telomere organization in the ligninolytic basidiomycete *Pleurotus ostreatus*. *Applied and Environmental Microbiology* **75**: 1427–1436.

Pezzella C, Autore F, Giardina P, Piscitelli A, Sannia G, Faraco V, 2009. The *Pleurotus ostreatus* laccase multi-gene family: isolation and heterologous expression of new family members. *Current Genetics* **55**: 45–57.

Piontek K, Antorini M, Choinowski T, 2002. Crystal structure of a laccase from the fungus *Trametes versicolor* at 1.90-Å resolution containing a full complement of coppers. *The Journal of Biological Chemistry* **277**: 37663–37669.

Rodríguez E, Ruiz-Dueñas FJ, Kooistra R, Ram A, Martínez AT, Martínez MJ, 2008. Isolation of two laccase genes from the white-rot fungus *Pleurotus eryngii* and heterologous expression of the pel3 encoded protein. *Journal of Biotechnology* **134**: 9–19.

Salony, Garg N, Baranwal R, Chhabra M, Mishra S, Chaudhuri TK, Bisaria VS, 2008. Laccase of *Cyathus bulleri*: structural, catalytic characterization and expression in *Escherichia coli*. *Biochimica et Biophysica Acta* **1784**: 259–268.

Smith M, Shnyreva A, Wood DA, Thurston CF, 1998. Tandem organization and highly disparate expression of the two laccase genes lcc1 and lcc2 in the cultivated mushroom *Agaricus bisporus*. *Microbiology* **144**: 1063–1069.

Soden DM, Dobson AD, 2001. Differential regulation of laccase gene expression in *Pleurotus sajor-caju*. *Microbiology* **147**: 1755–1763.

Tamura K, Dudley J, Nei M, Kumar S, 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596–1599.

Tsai HF, Wheeler MH, Chang YC, Kwon-Chung KJ, 1999. A developmentally regulated gene cluster involved in conidial pigment biosynthesis in *Aspergillus fumigatus*. *Journal of Bacteriology* **181**: 6469–6477.

Wahleithner JA, Xu F, Brown KM, Brown SH, Golightly EJ, Halkier T, Kauppinen S, Pederson A, 1996. The identification and characterisation of four laccases from the plant pathogenic fungus *Rhizoctonia solani*. *Current Genetics* **29**: 395–403.

Wang HX, Ng TB, 2004a. A new laccase from dried fruiting bodies of the monkey head mushroom *Hericium erinaceum*. *Biochemical and Biophysical Research Communications* **322**: 17–21.

Wang HX, Ng TB, 2004b. A novel laccase with fair thermostability from the edible wild mushroom (*Albatrella dispansus*). *Biochemical and Biophysical Research Communications* **310**: 381–385.

Wang HX, Ng TB, 2004c. Purification of a novel low-molecular-mass laccase with HIV-1 reverse transcriptase inhibitory activity from the mushroom *Tricholoma giganteum*. *Biochemical and Biophysical Research Communications* **315**: 450–454.

Wang HX, Ng TB, 2006a. A laccase from the medicinal mushroom *Ganoderma lucidum*. *Applied Microbiology and Biotechnology* **72**: 508–513.

Wang HX, Ng TB, 2006b. Purification of a laccase from fruiting bodies of the mushroom *Pleurotus eryngii*. *Applied Microbiology and Biotechnology* **69**: 521–525.

Wong JH, Ng TB, Jiang Y, Liu F, Cho S, Sze W, Zhang KY, 2009. Purification and characterization of a laccase with inhibitory activity toward HIV-1 reverse transcriptase and tumor cells from an edible mushroom (*Pleurotus cornucopiae*). *Protein and Peptide Letters* **17**: 1040–1047.

Yoshida H, 1883. Chemistry of lacquer (urishi). Part 1. *Journal of the Chemical Society* **43**: 472–486.

Archives of Biochemistry and Biophysics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

PHK from phenol hydroxylase of *Pseudomonas* sp. OX1. Insight into the role of an accessory protein in bacterial multicomponent monooxygenases

Viviana Izzo [a,c,*], Gabriella Leo [b], Roberta Scognamiglio [a], Luca Troncone [a], Leila Birolo [b,c], Alberto Di Donato [a]

[a] *Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli Federico II, Via Cinthia, I-80126 Napoli, Italy*
[b] *Dipartimento di Chimica Organica e Biochimica, Università di Napoli Federico II, Via Cinthia, I-80126 Napoli, Italy*
[c] *School of Biotechnological Sciences, Università di Napoli Federico II, Napoli, Italy*

ARTICLE INFO

Article history:
Received 22 June 2010
and in revised form 6 September 2010
Available online xxxx

Keywords:
Diiron
Hydroxylation
Metallochaperone
Monooxygenase
Phenol hydroxylase

ABSTRACT

Bacterial multicomponent monooxygenases (BMMs) are members of a wide family of diiron enzymes that use molecular oxygen to hydroxylate a variety of aromatic compounds. The presence of genes encoding for accessory proteins not involved in catalysis and whose role is still elusive, is a common feature of the gene clusters of several BMMs, including phenol hydroxylases and several soluble methane monooxygenases. In this study we have expressed, purified, and partially characterized the accessory component PHK of the phenol hydroxylase from *Pseudomonas* sp. OX1, a bacterium able to degrade several aromatic compounds. The phenol hydroxylase (ph) gene cluster was expressed in *Escherichia coli*/JM109 cells in the absence and in the presence of the phk gene. The presence of the phk gene lead to an increase in the hydroxylase activity of whole recombinant cells with phenol. PHK was assessed for its ability to interact with the active hydroxylase complex. Our results show that PHK is neither involved in the catalytic activity of the phenol hydroxylase complex nor required for the assembly of apo-hydroxylase. Our results suggest instead that this component may be responsible for enhancing iron incorporation into the active site of the apo-hydroxylase.

Introduction

Aromatic hydrocarbons released in the biosphere by human activities are today a major threat not only to the environment but also to human health due to their potential carcinogenicity [1]. By utilizing various catabolic pathways, microorganisms can use a wide array of aromatic compounds as the sole carbon and energy source, thus providing a set of diversified tools that can be used in bioremediation of contaminated environments [2,3]. In aerobic microorganisms, processing of aromatic compounds is initiated by oxygenases, in the so-called *upper pathway* [1,2]. These enzymes are responsible for the insertion of one or more oxygen atoms in the aromatic rings that, once activated, are further transformed by ring-cleavage dioxygenases that are part of a metabolic route called *lower pathway* [1,2]. The resulting products are then converted into metabolites of the tricarboxylic acid cycle (TCA), and eventually mineralized to carbon dioxide and water [1–4].

Bacterial multicomponent monooxygenases (BMMs)[1] are key enzymes of the upper pathway and catalyze the hydroxylation of the aromatic ring at different positions [4,5]. They usually consist of a 200–250 kDa hydroxylase component organized in a $(\alpha\beta\gamma)_2$ quaternary arrangement, a 10–16 kDa regulatory protein devoid of any cofactor and enhancing catalytic turnover, and a FAD- and (2Fe-2S) containing reductase that mediates electron transfer from NADH to the active site of the hydroxylase [4–6]. An additional Rieske protein may be present to assist the electron transfer between the reductase and hydroxylase components [4,7,8]. The hydroxylation chemistry takes place at a non-heme carboxylate-bridged diiron center coordinated by four glutamate and two histidine residues from a four-helix bundle [4–6,10]. Although the four-helix bundle motif and the diiron center are part of a conserved framework probably present in all BMMs [4,9], these enzymes differ considerably in subunit composition and dimension, arrangement of the coding sequences, and genomic localization which can be either chromosomic, or plasmidic [4,11].

* Corresponding author at: Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli Federico II, Via Cinthia, I-80126 Napoli, Italy. Fax: +39 081 679313.
E-mail address: vizzo@unina.it (V. Izzo).

[1] *Abbreviations used:* BMM, bacterial multicomponent monooxygenase; ESI–MS, electro spray ionization mass spectrometry; ITC, isothermal titration calorimetry; LC-MSMS, liquid chromatography with tandem mass spectrometry detection; MALDI-TOF, matrix-assisted laser-desorption ionization-time of flight; MOPS, 3-(*N*-morpholino)propanesulfonic acid; PH, phenol hydroxylase; ORF, open reading frame; SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; sMMO, soluble methane monooxygenase; ToMO, toluene/*o*-xylene monooxygenase.

0003-9861/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.abb.2010.09.023

Two different multicomponent monooxygenases, toluene *o*-xylene monooxygenase (ToMO) and phenol hydroxylase (PH) [12–20], have been identified in the genome of *Pseudomonas* sp. OX1, a bacterium able to grow on a wide spectrum of hydroxylated and non-hydroxylated aromatic compounds [12]. PH is a BMM composed of five polypeptides: PHL, PHM, PHN, PHO, PHP. These five polypeptide chains are organized as three components [20]. PHP is the NADH-oxidoreductase responsible for supplying electrons to the diiron cluster housed in the active site of the hydroxylase component. The hydroxylase comprises three polypeptides (L, N, O) organized in a quaternary structure of the type $(LNO)_2$. Finally, PHM has been shown to be a regulatory protein, devoid of any cofactor or metal, that is essential for efficient catalysis [20]. A comparative analysis of the polypeptide sequences of PH polypeptides showed a significant degree of identity with the components of the multicomponent phenol hydroxylase encoded by the *dmp* genes in *Pseudomonas* sp. CF600 [21–24], and with the constituents of toluene-*o*-monooxygenase (TOM) from *Burkholderia cepacia* G4 [25].

When the nucleotide sequence of the *ph* locus coding for the multicomponent phenol hydroxylase of *Pseudomonas* sp. OX1 was first determined, its 5′ region revealed the presence of a putative open reading frame (*orf*) whose deduced amino acid sequence shared 62.6% identity (74% similarity) to DmpK [21,26], a protein translated from an *orf* located at the 5′ of the gene cluster coding for the multicomponent phenol hydroxylase from *Pseudomonas* sp. CF600. DmpK has been shown not to be essential for *in vitro* phenol hydroxylase activity, but several lines of biochemical evidence suggested a major role for this protein in the expression of active recombinant hydroxylase [26].

DmpK is not a unique case among multicomponent monooxygenases. In fact, genes expressing a group of proteins usually indicated as *accessory* or *auxiliary proteins*, are frequently clustered together with the *orfs* coding for the other components of the BMMs [4,5]. These components share several common features: they are small proteins with a molecular mass of about 10 kDa, do not harbor any inorganic or organic cofactors, are expressed at low levels in their respective native organisms, and inhibit enzyme activity *in vitro* when present in stoichiometric amounts [27]. Despite these common aspects, the roles of the accessory proteins in BMMs have been elusive so far, although they have been hypothesized to be involved in the assembly of the active form of the hydroxylase component. This function is exemplified by biochemical studies of the protein MMOD, which was isolated from the methanotrophic bacterium *Methylococcus capsulatus* (Bath) and is part of the soluble methane monooxygenase (sMMO) system [6,27]. Despite a lack of sequence homology, similar genes encoding small proteins of unknown function have been identified across many different groups of BMMs [4,5,28–30], suggesting that the involvement of additional protein factors in the assembly of the active hydroxylase diiron centers might be more widespread than currently believed.

Accessory proteins recruited for the assembly of the metal cofactor in the active site are not a unique feature of BMMs, because proteins with similar roles have been described as common molecular tools to allow correct or efficient insertion of metals into the active sites of other metalloenzymes including Fe/S proteins, nitrogenase, urease and hydrogenase [26,31,32]. Different molecular mechanisms employed to facilitate the insertion of metal sites into target proteins have been reported in literature [32]. However, none of the mechanisms described so far have been unambiguously assigned to individual proteins such as DmpK and MMOD.

In this paper we describe a set of experiments aimed at gaining insight into the role of PHK, the accessory protein of phenol hydroxylase from *Pseudomonas* sp. OX1. The phenol hydroxylase gene cluster was cloned and expressed in *Escherichia coli* in the absence or presence of the *orf* coding for PHK. In the latter case, recombinant cells showed an increased phenol hydroxylase activity compared to cells not expressing PHK, suggesting a positive effect of the accessory protein on the enzymatic activity of phenol hydroxylase. Active $PH(LNO)_2$ and $PH(LNO)_2$/PHM complexes were purified from the soluble extract of cells expressing the complete *ph* gene cluster including *phk*. Additionally, we isolated a stable complex formed by PHK bound to a PH(LNO) trimer that was devoid of iron and hydroxylase activity and could not be reactivated under our experimental conditions. To elucidate the role of PHK, we subcloned and overexpressed this component in *E. coli* and characterized its interaction with the active hydroxylase moiety, $PH(LNO)_2$, showing that the protein is not directly involved in the catalytic activity of the phenol hydroxylase complex, as for MMOD and DmpK. However, PHK does not appear to be required for the production of active recombinant hydroxylase as observed previously for DmpK [26]. PHK seems instead to be involved in increasing the apparent affinity of apo-$PH(LNO)_2$ for iron, thus possibly facilitating the incorporation of the crucial transition metal into the active site of the protein.

Experimental procedures

Materials and general procedures

Bacterial cultures, plasmid purifications and transformations were performed according to Sambrook et al. [33]. The pET22b(+) expression vector and *E. coli*-strain JM109 were purchased from Novagen, whereas *E. coli*-strain BL21(DE3) was from Invitrogen. Platinum *Pfx* DNA polymerase used for PCR amplification was from Life Technologies. Enzymes and other reagents for DNA manipulation were from New England Biolabs. The oligonucleotides were synthesized at the MWG-Biotech (http://www.mwgdna.com). Q-Sepharose Fast Flow, Sephacryl S300 High Resolution, Sephadex G75 Superfine, Superdex columns and disposable PD10 desalting columns were from Pharmacia. All other chemicals were from Sigma. Ammonium hydrogen carbonate (AMBIC), Ethylenediaminetetraacetic acid (EDTA) and iodoacetamide were purchased from Fluka; Tri(hydroxymethyl)aminomethane (Tris), urea, dithiothreitol (DTT), and sequencing-grade trypsin were from Sigma. Formic acid and Acetonitrile (ACN) were purchased from Romil.

Expression and purification of recombinant PHM, PHP and C2,3O from *Pseudomonas* sp. OX1 are described elsewhere [20,34]. DNA fragments were sequenced at the MWG-Biotech (http://www.mwgdna.com).

Construction of expression vectors

The plasmid pGEM3Z/*phΔk* containing the *ph-lmnop orfs* was kindly supplied by Dr. Valeria Cafaro (Dipartimento di Biologia Strutturale e Funzionale, Università Federico II, Napoli, Italy). The plasmid pGEM7Z/*k* containing the *ph-klm orfs* was a gift of Dr. Carla Caruso (Dipartimento di Agrobiologia e Agrochimica, Università della Tuscia, Viterbo, Italy).

To allow the expression of the complete *ph* gene cluster, the DNA sequence encoding the PHK component was subcloned from plasmid pGEM7Z/*k* into vector pGEM3Z/*phΔk* to create a construct containing the *phk, phl, phm, phn*, *pho* and *php* genes. A DNA fragment comprising the entire *phk* gene and part of the *phl* genes was amplified from the pGEM7Z/*k* vector using standard PCR procedures. A *Sal*I restriction site was inserted at the 5′ end of the *phk* gene. The 3′ primer was designed to amplify an internal region of the gene *phl*, 180 nucleotides downstream of the ATG start codon, incorporating a *Mlu*I restriction site. The resulting recombinant plasmid, named pGEM3Z/*ph*, was verified by DNA sequencing. To

construct plasmid pGEM3Z/*phΔp*, pGEM3Z/*ph* was digested with the restriction endonuclease *Eco*RV which cuts the *php* gene at nucleotides 196 and 335 causing the deletion of a 140 bp fragment.

In order to overexpress PHK, a third construct was prepared that contained only the *phk* gene. The gene *phk* was first amplified using plasmid pGEM7Z/*k* as template and a PCR method that inserted *Nde*I and *BamH*I restriction sites at the 5′ and 3′ ends of the gene, respectively. The amplified fragment was digested with endonucleases *Nde*I and *BamH*I, ligated with the pET22b(+) vector previously cut with the same enzymes. The resulting recombinant plasmid, named pET/*phk* was verified by DNA sequencing and used to transform BL21(DE3) competent cells.

Expression of PHK and PH(LNO)$_2$

The gene products of plasmids pGEM3Z/*phΔk* and pGEM3Z/*phΔp* were expressed in *E. coli*-JM109 cells. PHK was expressed from pET/*phk* in *E. coli* BL21(DE3) cells. All recombinant strains were routinely grown in Luria–Bertani (LB) medium [33] supplemented with 100 μg mL^{-1} of ampicillin. Expression of recombinant proteins was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 0.2 mM for pGEM3Z/*phΔk* and pGEM3Z/*phΔp*, and 0.1 mM for pET/*phk*. At the time of induction, $Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$ was added at a final concentration of 0.2 mM, except in the case of PHK from pET/*phk*. Growth was continued in all cases for 3 h at 37 °C. Cells were collected by centrifugation, washed in 25 mM of 3-(*N*-morpholino)propanesulfonic acid (MOPS), pH 7.0, containing 5% glycerol, and 2 mM L-cysteine (Buffer A). The cell paste was stored at −80 °C until needed.

Purification of the expression products of pGEM3Z/phΔp

The paste from a 2 L culture of induced cells was suspended in 40 mL of buffer A. Cells were disrupted by sonication (10 times for a 1-min cycle, on ice). Cell debris were removed by centrifugation at 18,000*g* for 60 min at 4 °C. The supernatant was immediately fractionated as described below. The soluble fraction from a 2 L culture of JM109 cells expressing the gene products of plasmid pGEM3Z/*phΔp* or pGEM3Z/*phΔk* was loaded onto a Q-Sepharose Fast Flow column (1 × 18 cm) equilibrated in buffer A containing 0.08 M NaCl at a flow rate of 10 mL h^{-1}. Elution was performed with a 500 mL linear salt gradient from 0.12 to 0.5 M NaCl in buffer A, at a flow rate of 18 mL h^{-1}. Fractions containing the proteins of interest were pooled, concentrated by ultrafiltration on an YM30 membrane, and loaded onto a Sephacryl S300 High Resolution column (2.5 × 50 cm) equilibrated in buffer A containing 0.2 M NaCl, at a flow rate of 8 mL h^{-1}. Fractions were pooled, concentrated by ultrafiltration on YM30 membranes, and stored under nitrogen at −80 °C. Unless otherwise stated, all chromatographic steps were performed at 4 °C. Buffers were purged by repeated cycles of flushing with nitrogen. Column operations were not strictly anoxic.

Purification of PHK

The soluble fraction from a 2 L culture of cells expressing the gene product of plasmid pET22b(+)/*phk* was loaded onto a Q-Sepharose Fast Flow column (1 × 18 cm) equilibrated in buffer A at a flow rate of 10 mL h^{-1}. Proteins were eluted with a 400 mL linear salt gradient from 0 to 0.45 M NaCl, at a flow rate of 10 mL h^{-1}. Relevant fractions were pooled, concentrated by ultrafiltration on YM10 membranes, and loaded onto a Sephadex G75 Superfine column (2.5 × 50 cm) equilibrated in buffer A containing 0.2 M NaCl at a flow rate of 8 mL h^{-1}. Fractions containing electrophoretically pure PHK were pooled, purged with N_2, concentrated by ultrafiltration on YM10 membranes, and stored at −80 °C.

Enzymatic assays

Enzymatic activities were measured as previously described [17] using phenol as a substrate and monitoring the production of catechol in a continuous coupled assay with recombinant catechol 2,3-dioxygenase from *Pseudomonas* sp. OX1 [34], which cleaves the catechol ring and produces 2-hydroxymuconic semialdehyde. Formation of this product, having an extinction coefficient of 12,000 M^{-1} cm^{-1}, can be monitored spectrophotometrically at 410 nm.

(i) *Assays of reconstituted PH components.* Conditions for this enzyme assay probing the activity of the reconstituted PH complex are reported elsewhere [17,20]. PH(LNO)$_2$ concentrations employed in the assay were typically 0.25–0.75 μM. Components PHP and PHM were added in a ratio of 2:4:1 compared to PH(LNO)$_2$ concentration. Semialdehyde yields were determined by measuring the absorbance at 410 nm. Specific activity (S.A.) was defined as nanomoles of substrate converted per minute per milligram (mU/mg) of PH(LNO)$_2$ hexamer at 25 °C.

(ii) *Whole-cells assays.* Whole-cells assays were performed as previously described [20] using *E. coli*-JM109 cells transformed with the plasmid of interest. One milliunit was defined as the amount of catalyst that oxidized 1 nmol of phenol per min at 25 °C.

(iii) *Assays on crude extracts.* Crude extracts of JM109 expressing either pGEM3Z/*ph* or pGEM3Z/*phΔk* were prepared by sonicating the cell suspension at an optical density at OD_{600} of about 7 in buffer A (3 times for a 1-min cycle, on ice). Cell debris was removed by centrifugation at 18,000*g* for 30 min at 4 °C. The supernatant was carefully decanted, and kept on ice. The reaction was carried out in 0.1 M Tris–HCl pH 7.5 in the presence of 0.25 mM NADH and 0.5 mM phenol. Semialdehyde yields were determined as noted.

Iron removal from PH(LNO)$_2$

Removal of iron from the hydroxylase component was performed as previously described [35] with the following modifications. A 1.2 mL solution of 10 μM PH(LNO)$_2$, 0.5 mM 1,10-phenanthroline in 0.05% HCl, and 0.8 mM methyl viologen in 25 mM MOPS (pH 7.0), was prepared in a glass vial fitted with a rubber septum and purged with nitrogen. A 20 μL aliquot of a 350 mM sodium dithionite solution, previously purged with a continuous flow of nitrogen, was added and the reaction mixture was incubated at room temperature for 3 h under a continuous flow of nitrogen. Apo-PH(LNO)$_2$ was separated from the reagents and from the $[Fe(o\text{-phen})_3]^{2+}$ complex by using a disposable PD10 desalting column (Pharmacia) equilibrated with a buffer containing 25 mM MOPS (pH 7.0), 200 mM NaCl, 5% glycerol and 2 mM L-cysteine (Buffer D). The mixture containing apo-PH(LNO)$_2$ (10 μM) was made anaerobic by purging the solution with nitrogen. Fractions of interest were pooled, concentrated and stored at −80 °C.

Isothermal titration calorimetry (ITC)

ITC experiments were performed on a Nano-ITC, CSC 5300 calorimeter from Calorimetry Science Corporation (Lindon, UT), having a cell volume of 1.3 mL. Calorimetric titrations were carried out at 25 °C, by injecting 10 μL aliquots of a 130 μM ligand solution into a 11 μM PHK solution at 200 s intervals for a total of 25 injections. The solution was stirred at 1000 rpm.

Analytical gel filtration

Analytical gel-filtration experiments were carried out as follows: 200 μL of a protein sample was loaded on a Superdex 200 HR 10/30 or on a Superdex 75 HR 10/30 column previously equilibrated in 25 mM Tris–HCl, 5% glycerol, 0.2 M NaCl and 2 mM L-cysteine, pH 7.0 installed on an AKTA™FPLC™ (GE Healthcare Life Science). The samples were eluted isocratically at room temperature at a flow rate of 0.5 mL min^{-1}. Protein separation was monitored at 280 nm. A molecular weight calibration was conducted in the same buffer with the following proteins of known molecular weight: apotransferrin (400 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66,000 Da), ovalbumin (45,000 Da), carbonic anhydrase (29,000 Da), trypsinogen (24,000 Da), cytochrome *c* (12,400 Da). When necessary, the area of each peak was estimated by nonlinear curve-fitting of the elution profile using PeakFit software (Systat Software).

Biotinylation of PHK

PHK (3.0 mg, in 25 mM MOPS pH 7.0, 5% glycerol, 0.2 M NaCl) was incubated with 1.37 mg of sulfosuccinimidobiotin (Sulfo-NHS-Biotin) (Pierce) dissolved in the same buffer at a molar ratio of 20:1, on ice in the dark for 2 h. The reaction was quenched with 0.1 M Tris–HCl pH 7.0 and the non-reacted biotin reagent was removed by size exclusion chromatography on a PD-10 column (GE-Healthcare) in 0.1 M Sodium Phosphate pH 7.2, 0.15 M NaCl. Absorbance at 220 and 280 nm was monitored to identify PHK-containing fractions. Appropriate fractions were pooled and added to Avidin Agarose Resin (settled gel, Pierce) in order to immobilize the biotinylated protein.

Immobilization of biotinylated PHK on avidin and "Pull down" assays

A 450 μL aliquot of resin was equilibrated with five volumes of Binding Buffer (0.1 M Sodium Phosphate pH 7.2, 0.15 M NaCl) and incubated with biotinylated PHK (3 mg biotinylated PHK per mL of settled avidin agarose resin) at 4 °C for 1 h. The resin was washed with 10 volumes of Binding Buffer.

Pseudomonas sp. OX1 cells grown on M9 minimal medium containing 5 mM phenol as a unique carbon and energy source were disrupted by sonication in 25 mM MOPS pH 7.0, 0.15 M NaCl, 5% glycerol, 1 mM PMSF containing a protease inhibitors cocktail (Sigma Aldrich). The extracts were centrifuged at 12,000 rpm for 1 h and filtered with a 0.45 μm polyvinylidene fluoride (PVDF) membrane. Total protein cell extracts, quantified using the BioRad protein assay, were incubated with 125 μl of mouse anti-IgG agarose conjugated beads (Sigma) overnight at 4 °C (preclining step). Cell extracts were then incubated with 450 μL of biotinylated-PHK immobilized on avidin resin and were incubated overnight at 4 °C. The resin was washed with 10 volumes of Binding Buffer and the protein samples were eluted with 70 μL of Laemmli sample buffer. Samples were analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) (12.5%, 20 cm × 20 cm), and proteins were stained with Coomassie Brilliant Blue G-Colloidal (Pierce, Rockford, USA).

Mass spectrometric analysis

Identification of complexed subunits was carried out on trypsin digested samples either by digestion in solution or *in situ* after separation by polyacrilamide gel electrophoresis. Proteins were detected on the gel using a colloidal Coomassie kit (Invitrogen Life Technologies). Excised bands were destained, reduced with 10 mM dithiothreitol (DTT), carbamidomethylated with 55 mM iodoacetamide in 0.1 M NH_4HCO_3 buffer, pH 7.5 and subjected to tryptic in-gel digestion for 16 h at 37 °C, by adding 100 ng of trypsin. Peak center fractions from analytical gel-filtration experiments (1 mL) were lyophilized, resuspended in 100 μl of H_2O, and digested at 37 °C for 16 h with 100 nmoles of trypsin. Reactions were quenched by lowering the pH to about 1 with formic acid, and the resulting peptide mixtures were concentrated and purified using a reverse phase Zip Tip pipette tips (Millipore). The peptides were eluted with 20 μL of a solution comprising 50% acetonitrile and 0.1% formic acid in deionized water. Peptide mixtures were analyzed either by matrix-assisted laser-desorption/ionization mass spectrometry (MALDI-MS) or capillary liquid chromatography with tandem mass spectrometry detection (LC-MSMS).

MALDI-MS experiments were performed on a Voyager DE-STR matrix-assisted laser-desorption ionization-time of flight (MALDI-TOF) mass spectrometer (Applied Biosystems, Framingham, MA) equipped with a nitrogen laser (337 nm). Typically, 1 μL of the total peptide mixture was mixed (1/1, v/v) with a 10 mg mL^{-1} solution of R-cyano-4-hydroxycinnamic acid in acetonitrile/50 mM citrate buffer (2/3, v/v). The experimental mass values obtained were compared with calculated masses from the predicted tryptic digestion of the different subunit sequences, confirming the identities of the corresponding protein bands.

The peptide mixtures were analyzed using a CHIP MS 6520 QTOF equipped with a capillary 1200 HPLC system and a chip cube (Agilent Technologies, Palo Alto, Ca). After loading, the peptide mixture (8 μl in 0.1% formic acid) was first concentrated and washed at 4 μL/min in a 40 nL enrichment column (Agilent Technologies chip), with 0.1% formic acid in 2% acetonitrile as the solvent. The sample was then fractionated on a C_{18} reverse-phase capillary column (75 μm × 43 mm in the Agilent Technologies chip) at flow rate of 400 nL/min with a linear gradient of eluent B (0.1% formic acid in 95% acetonitrile) in A (0.1% formic acid in 2% acetonitrile) from 7% to 60% over 50 min.

Peptide analysis was performed using data-dependent acquisition of one MS scan (mass range from 300 to 2000 m/z) followed by MS/MS scans of the three most abundant ions in each MS scan. Raw data from nanoLC-MSMS analyses were employed to query, using MASCOT software (Matrix Science, Boston, USA), a non-redundant protein databases (NCBI, with taxonomy restriction to Bacteria), or an *ad hoc* created databases including only the sequences of PHK, PHL, PHM, PHN, PHO.

Other methods

Polyacrylamide gel electrophoresis was carried out using standard techniques [33,36]. Tris–glycine gels (18% and 15%) were run under denaturing and native conditions respectively. SDS–PAGE "wide range" (200–6.5 kDa) molecular weight standard was from Sigma. Homology studies were performed by searching the public nucleotide and protein databases with BLAST. Total iron content was determined colorimetrically by complexation with Ferrozine [37]. Protein concentrations were determined colorimetrically with the Bradford Reagent [38] from Sigma, using 1–10 μg of bovine serum albumin (BSA) as a standard.

Results

Activity assays of whole cells of E. coli-JM109 expressing recombinant phenol hydroxylase in the presence or absence of PHK

To evaluate the influence of PHK on the phenol hydroxylase activity of PH, we initially investigated the catalytic activity of whole cells of *E. coli*-JM109 expressing the complete *ph l-m-n-o-p* gene cluster with or without *phk*, and compared the rate of conver-

sion of phenol to catechol by the induced cells in the presence or in the absence of the *phk* gene product.

To this purpose, the *orf* encoding PHK was amplified from plasmid pGEM7Z/*k* and subcloned into plasmid pGEM3Z/*phΔk* (see Construction of expression vectors). The primers used in the PCR reaction were designed to incorporate the putative PHK ribosome binding site (*rbs)* at the 5′ of the *phk* gene. This *rbs* was previously located at 8–10 bp upstream of the ATG start codon in the native 5′ UTR region of the *ph* gene cluster.

SDS–PAGE analysis of the crude extracts of induced cells expressing either pGEM3Z/*phΔk* or pGEM3Z/*ph* revealed that the hydroxylase components PHL, PHN and PHO were expressed at high levels from both plasmids. Additionally, no significant difference in the expression levels of PHM and PHP was observed. However, whole cells expressing the plasmid pGEM3Z/*ph*, showed a 2- to 4-fold increase in the rate of catechol production when compared to those lacking the expression product of gene *phk*. The latter had a specific activity of 0.8 ± 0.1 mU/OD_{600}; in the presence of gene *phk* the value of the activity with phenol was of 3.1 ± 0.5 mU/OD_{600}.

Similar results were obtained when crude extracts of *E. coli*-JM109 expressing either pGEM3Z/*ph* or pGEM3Z/*phΔk* were analyzed for catechol production. Recombinant *E. coli* cells harboring either plasmid were induced with 0.2 mM IPTG at 37 °C for 90 min and were collected by centrifugation. Crude extracts were prepared and used in continuous coupled assays as described in Enzymatic assays. A ~2–3-fold higher phenol hydroxylase activity was observed for crude extracts of *E. coli* expressing the *ph* gene cluster including gene *phk* than for those expressing the *ph* gene cluster in the absence of *phk*. Surprisingly, the activity of crude extract of *E. coli*-JM109 expressing pGEM3Z/*phΔk* could be restored to pGEM3Z/*ph* levels by adding 5 μM of exogenous iron (II). These results contrast those previously described for recombinant DmpK, where the addition of exogenous iron (II) was not sufficient to restore the enzymatic activity of the hydroxylase component in the absence of the auxiliary protein [26]. The total iron contents of crude extracts of cells expressing either pGEM3Z/*ph* or pGEM3Z/*phΔk* is similar, with 3.7 ± 0.7 nmol Fe/mg total protein and 4.2 ± 0.8 nmol Fe/mg total protein respectively, suggesting that PHK might be involved in directing intracellular iron to the hydroxylase moiety rather than influencing total amount of iron in the cell.

Purification and characterization of the expression products of pGEM3Z/phΔp

The results obtained on whole cells and on cells extracts of induced *E. coli* cells, described in the previous section, suggested a positive influence of the accessory protein PHK on the enzymatic activity of the phenol hydroxylase complex.

To gain additional insight into the role of PHK, we decided to purify PHK and all the other components of the phenol hydroxylase system from cells of *E. coli* expressing plasmid pGEM3Z/*ph*.

First a derivative of plasmid pGEM3Z/*ph*, having a partial deletion of the *php* gene, was constructed (see Construction of expression vectors). This construct was generated to avoid accumulation of indigo in the cell culture [39], which is the result of the hydroxylation of cellular aromatics by the active phenol hydroxylase system during induction of transformed *E. coli* cells.

Fractionation of the crude extracts of JM109 cells expressing pGEM3Z/*phΔp* was performed as described in the "Experimental procedures" section. Three distinct peaks eluted from the first anionic exchange column and were collected separately. As shown in Fig. 1A, all three peaks contained the PHL, PHN and PHO polypeptides of the hydroxylase protein, when analyzed by SDS–gel [20]. In peak 1, an additional protein band, with the molecular weight expected

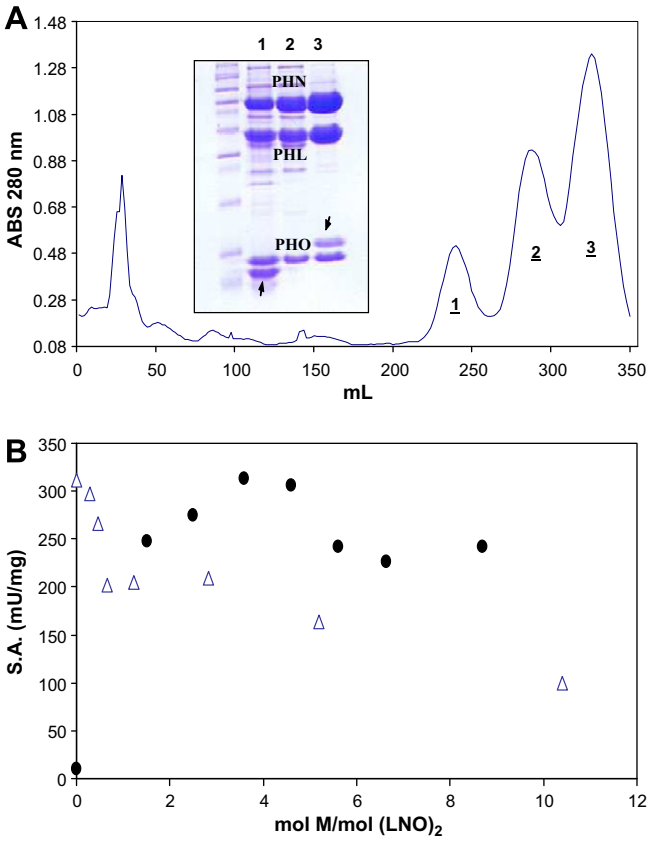


Fig. 1. Expression products of plasmid pGEM3Z/*phΔp*. (A) Q-Sepharose chromatogram and SDS–PAGE. MW, molecular weight standard ranging from 200 to 6.5 kDa (see Immobilization of biotinylated PHK on avidin and "Pull down" assays, for details). *Lane 1*, peak 1. *Lane 2*, peak 2. *Lane 3*, peak 3. Lane 2 contains the polypeptides PHN (60 kDa), PHL (38.3 kDa) and PHO (13.1 kDa). The arrows in lane 1 and 3 indicate the additional polypeptides, PHK and PHM respectively, that co-elute with the hydroxylase formed by subunits PHN, PHL and PHO. (B) Effect of different concentrations of regulatory protein PHM with respect to the hydroxylase $PH(LNO)_2$ on the rate of catechol production. The experiment was performed in the presence of either PHM–$PH(LNO)_2$ (empty triangles) complex isolated from peak 3, or $PH(LNO)_2$ (filled circles) isolated from peak 2. Optimized PHP/$PH(LNO)_2$ ratio of 4:1 was kept constant during the titration.

for PHK, was evident. In contrast the contents of peak 3 revealed a protein with the molecular weight expected for the regulatory protein PHM. The identities of the proteins of these three peaks were confirmed by peptide mass fingerprinting after *in situ* digestion of the bands excised from the gel and MALDI-TOF analyses.

Contents of the three peaks were further purified on a S-200 gel filtration column followed by anionic exchange chromatography on a Mono-Q column. Even after these purification steps, the composition and the ratio of the PH subunits of the three peaks remained as such, as evidenced by SDS–PAGE and mass spectrometry. These results strongly suggest the formation of stable complexes between the LNO subunits and PHK in peak 1, and between the LNO subunits and PHM in peak 3.

The protein complexes of all three peaks were individually characterized for iron content, molecular weight, and specific activity using phenol as a substrate. As shown in Table 1, the putative PHK-hydroxylase complex of peak 1 was devoid of both iron and enzymatic activity. The molecular weight of the putative complex measured by using a Superdex 200 gel filtration column accounted for a complex in which one molecule of PHK binds a trimeric LNO species. From now on we will refer to this complex as PHK-PH(LNO).

The same analyses carried out on peaks 2 and 3 (Table 1), revealed that peak 2 contains the hexameric hydroxylase with a qua-

Table 1

Pool	$Fe^{2+}/(LNO)_2$	Specific activity (mU/mg)	Calculated MW (kDa)	Expected MW (kDa)
1. PHK–PH(LNO)	ND[a]	ND[a]	136.30 ± 9.53	122.16
2. $PH(LNO)_2$	3.5 ± 0.2	254 ± 46	248.07 ± 2.31	223.58
3. PHM–$PH(LNO)_2$	4.2 ± 0.1	329 ± 17	259.30 ± 1.81	244.54

[a] ND, not detectable.

ternary structure of the type $PH(LNO)_2$, whereas peak 3 contains a complex in which two molecules of PHM binds the dimer $(LNO)_2$, from now on referred as PHM–$PH(LNO)_2$. The putative PHM–$PH(LNO)_2$ complex in peak 3 was fully active with phenol after the addition of an optimized ratio of the oxidoreductase component PHP (see Enzymatic assays and Fig. 1B). For peak 2, both the oxidoreductase and the recombinant component PHM, purified and characterized as described in Cafaro et al. [20], had to be added to obtain an optimally active catalytic system. In these latter experiments, 4 equiv. of PHM were added to $PH(LNO)_2$ for maximal activity (Fig. 1B). For both peak 2 and peak 3, adding higher concentrations of the regulatory component PHM inhibited the rate of phenol hydroxylation (Fig. 1B), [20].

Attempts to purify adequate amounts of PHK from cells expressing the whole *ph* gene cluster from pGEM3Z/*ph* were unsuccessful. Recombinant PHK was always found to be firmly bound to the trimeric complex PH(LNO) eluting in peak 1 and could not be dissociated unless strong denaturing conditions were used.

Purification and properties of PHK

To purify and characterize PHK, we subcloned the *orf* encoding this protein and expressed it in *E. coli*, strain BL21(DE3) (see Construction of expression vectors). Recombinant PHK was purified to homogeneity from *E. coli* harboring pET22b(+)/*phk* by an initial anionic exchange chromatography on a Q-Sepharose FF resin followed by a gel filtration step on a G-75 column (see Purification of PHK).

The oligomeric state of PHK was assessed by analytical gel filtration on a Superdex 75 analytical column. The chromatographic profile resulted in a single peak, which eluted at an apparent molecular weight of about 25,500 ± 400 Da, suggesting the occurrence of a dimeric species. The accurate subunit molecular weight of PHK was determined by electrospray ionization mass spectrometry (ESI–MS); the experimental value of 10,235.27 ± 0.19 Da is in agreement with the molecular weight predicted from the sequence of PHK (10,235.7 Da).

Purified PHK was found to be devoid of iron and did not contain any redox active prosthetic groups such as flavin, heme, or iron–sulfur centers. Incubation of PHK with an excess of ferrous iron and subsequent purification of the sample by using either a PD10 desalting column or dialysis did not lead the protein to retain any of the metal bound (data not shown).

Isothermal titration calorimetry (ITC) was then performed to assess whether PHK binds iron (II). 1 mL of purified 11 μM PHK in 25 mM MOPS pH 7.0 containing 5 mM L-ascorbic acid was titrated with the same buffer having 250 μL of 130 μM $Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$. L-ascorbic acid was added to both samples to avoid oxidation of iron (II) to iron (III) during the experiment. No binding isotherm was detected under the conditions reported in Materials and methods. To exclude the possibility that this result could arise from the oxidation of iron during the course of the experiment, the same experiment was repeated using a solution containing 130 μM $MnCl_2$ in 25 mM MOPS pH 7.0. Manganese (II) has already been used for replacing native iron (II) in several proteins ([35] and references therein). This metal is similar to iron (II) in charge and ionic radius and can exist in different oxidation states, but most notably its oxidation state (II) is much more stable at pH 7.0 when compared to iron (II). ITC experiments performed under conditions described above did not provide any evidence of binding between isolated PHK and manganese.

To provide evidence for the specificity of the binding event between PHK and the LNO subunits observed in the PHK-PH(LNO) complex isolated after the purification of the expression products of pGEM3Z/*phΔp* (see Purification and characterization of the expression products of pGEM3Z/*phΔp* and Fig. 1A), we carried out an affinity purification strategy using an avidin pull-down assay with extracts of cells of *Pseudomonas* sp. OX1 grown on phenol as a unique carbon and energy source (Fig. 2). This approach required recombinant biotinylated PHK. We biotinylated 3 mg of PHK as described in Biotinylation of PHK, with Sulfo-NHS-Biotin (sulfosuccinimidobiotin). To confirm that PHK was efficiently biotinylated under the experimental conditions, an aliquot was analyzed by RP-HPLC/ES-MS, which show that PHK contained up to 3 biotins per molecule. Biotinylated PHK was immobilized on avidin beads that were subsequently incubated with total protein extracts from *Pseudomonas* sp. OX1 cells grown on phenol; these latter had been pre-cleaned on agarose beads to minimize non-specific binding on the cromatographic matrix during the pull-down procedure. After extensive washing, PHK-recruited protein interaction partners were eluted in Laemmli buffer, separated on SDS–PAGE, and stained with colloidal blue Coomassie (Fig. 2). Three major proteins with molecular weights of ~12, ~40 and ~60 kDa (Fig. 2; lane 2) were identified. These proteins ran similarly to purified PHO, PHL and PHN standards on the gel (Fig. 2, lane 3). The three protein bands were excised from the gel, reduced, alkylated and digested *in situ* with trypsin. The resulting peptide mixtures were analyzed by LC-MSMS and the identities of the proteins were confirmed as PHO, PHL and PHN by comparing the MS patterns to those contained within the bacteria subset of

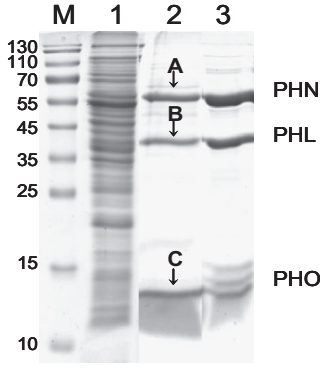


Fig. 2. Pull down experiment with biotinylated PHK. Colloidal blue-stained gel for biotinylated PHK associated proteins from *Pseudomonas* sp. OX1 cells grown on phenol (lane 2). Total *Pseudomonas* sp.OX1 cell extract is loaded in lane 1. An aliquot of PHM-$PH(LNO)_2$, was loaded (lane 3) as a control. Biotinylated PHK associated proteins were analyzed by SDS–PAGE followed by colloidal blue staining. A, B, and C labels in lane 3 indicate the three major proteins detected among the proteins pulled-down by biotinylated PHK.

the NCBI database using MASCOT software (Table S1, Supplementary Material). These results indicate a highly specific interaction of PHK with the LNO subunits of the hydroxylase moiety.

PHK-mediated inhibition of phenol hydroxylase activity

Once obtained the purified recombinant protein, the role of PHK was investigated by studying its influence on the phenol hydroxylase complex reconstituted from the purified components. Addition of purified PHK to the phenol hydroxylase complex reconstituted *in vitro* displayed a dose-dependent effect (Fig. 3). These assays contained PHP, PHM and $PH(LNO)_2$ at optimized ratios or PHP and the PHM–$PH(LNO)_2$ from peak 3 at optimized ratios with no exogenous recombinant PHM (see Enzymatic assays). Hydroxylation of phenol to catechol catalyzed by the *in vitro* reconstituted PH complex was measured at varying ratios of PHK to $PH(LNO)_2$. In all cases, addition of PHK resulted in the inhibition of the phenol hydroxylase activity (Fig. 3). Specifically, a 10-fold excess of PHK to $PH(LNO)_2$ resulted in only 16.8 ± 3.1% of the initial rate of catechol production. Interestingly, higher PHK:$PH(LNO)_2$ ratios did not lead to complete inactivation of phenol hydroxylase activity. A Kd value of 0.61 ± 0.15 μM over five different preparations of hydroxylase was obtained.

To gain further insight into the mechanism by which PHK inhibits its enzymatic activity, we incubated 57 μM of $PH(LNO)_2$ with 0.6 equiv. of purified PHK, in 500 μL of 25 mM MOPS pH 7.0. The mixture was let sit at room temperature for 90 min after which a 15 μL aliquot was analyzed on a native 15% polyacrilamide gel (Fig. 4A). Similar to a sample containing the partially purified PHK–$PH(LNO)_2$ complex from peak 1 (lane 1), the sample in which recombinant PHK was added to $PH(LNO)_2$ showed the presence of a high molecular weight species (lane 3) attributed to the trimeric form PH(LNO) associated to PHK (lane 1). Protein bands corresponding to the hydroxylase complex and uncomplexed PHK were also identified (lane 3) by comparison to a partially purified sample of $PH(LNO)_2$ (lane 2) and PHK (lane 4). The protein components of this putative identified PHK–PH(LNO) species were confirmed by peptide mass fingerprinting. The protein band from lane 3 was excised, digested *in situ* with trypsin, and analyzed by MALDI-TOF MS. Mass values were mapped onto the anticipated sequences of PHL, PHN, PHO, PHM, and PHK subunits confirming the presence of all four polypeptides (Table S2, Supplementary Material).

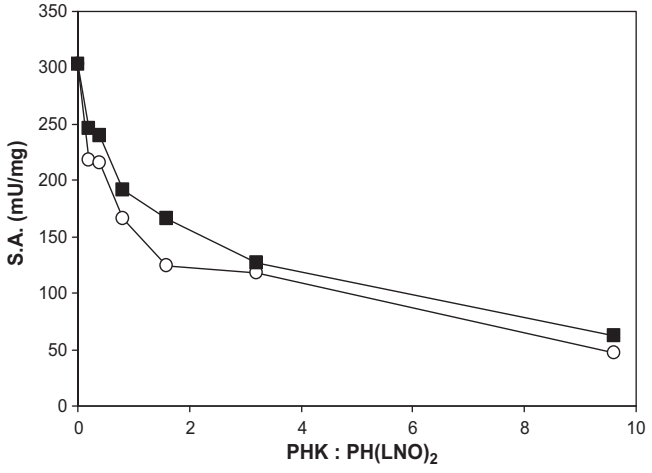


Fig. 3. Effect of increasing concentration of PHK on phenol hydroxylase activity of $PH(LNO)_2$. Activity assays were performed as described in Experimental Procedures. Filled squares represent experiments performed with the pre-formed PHM–$PH(LNO)_2$ complex. Empty circles are experiments with the $PH(LNO)_2$ complex employing exogenous recombinant PHM added at a 4:1 ratio with respect to the hydroxylase.

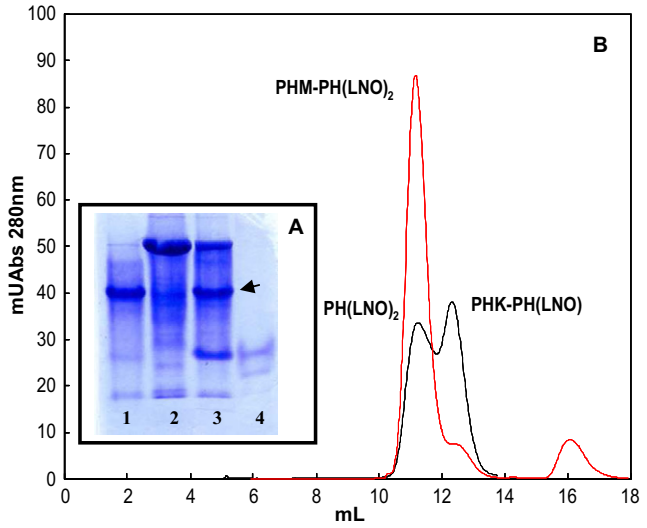


Fig. 4. Characterization of the $PH(LNO)_2$ complex in the presence of PHK. (A) Native gel electrophoresis of a $PH(LNO)_2$/PHK mixture. *Lane 1*, 10 μg of the PHK-PH(LNO) complex purified from peak 1 of the anion exchange chromatographic separation of Fig. 1. *Lane 2*, 20 μg of $PH(LNO)_2$ active hexamer. *Lane 3*, 20 μg of $PH(LNO)_2$ incubated with 0.6 equiv. of PHK for 90 min at room temperature. *Lane 4*, approximately 5 μg of purified recombinant PHK. The arrow indicates the additional band present in the incubated sample that is absent in the starting protein samples. (B) Elution profiles from analytical gel filtration chromatographic separation of a 200 μL mixture of 4.4 μM of $PH(LNO)_2$ and 2.2 μM of PHK (black line), and of a 200 μL mixture of 2.2 μM of PHM–$PH(LNO)_2$ and 4.4 μM of PHK (red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The contents of the reaction mixture containing both $PH(LNO)_2$ and PHK were also analyzed by analytical gel filtration chromatography. A 200 μL aliquot of the mixture was run onto a Superdex 200 column. The elution profile (Fig. 4B) revealed a protein peak with an apparent molecular weight of 220 kDa, corresponding to the $PH(LNO)_2$ hexameric complex of the active hydroxylase, as assessed by running an authentic standard of $PH(LNO)_2$. Also evident is a species with an apparent molecular weight of 124 kDa, likely corresponding to the trimeric PH(LNO) species complexed with PHK. Fractions located in the peak centers were analyzed by SDS–PAGE and mass spectrometry (data not shown), confirming the presence of PHK only in the peak eluting with the apparent molecular weight of 124 kDa, whereas PHL, PHN, and PHO were observed in both peaks.

The molecular species produced by the interaction of PHK with the hexameric hydroxylase $PH(LNO)_2$ can be confidently assigned as a PHK-PH(LNO) complex. The observation that this species is devoid of both iron and catalytic activity with phenol, suggests that the inhibition of the phenol hydroxylase activity, demonstrated in *in vitro* assays and previously reported in the case of accessory proteins DmpK and MMOD [26,27], depends on the formation of apo-trimers of the type PH(LNO) and is mediated by the intervention of PHK.

Interestingly, incubation of PHK with the PHM–$PH(LNO)_2$ complex (peak 3 in Fig. 1), performed under the same experimental conditions, resulted in very little dissociation of the hexameric hydroxylase (Fig. 4B).

The PHK-PH(LNO) complex, obtained either from fractionation of crude extract of *E. coli*/pGEM3ZΔ*p* (Fig. 1A) or from the incubation of purified component PHK with hexameric $PH(LNO)_2$ complex (Fig. 4), could not be reactivated by any of our experimental procedures. In one of these experiments a 1.2 mL solution

containing 10 μM of PHK–PH(LNO) complex and 0.8 mM methyl viologen in 25 mM MOPS pH 7.0, was prepared in a glass vial fitted with a rubber septum and purged with nitrogen. Next, 20 μL of a 350 mM solution of sodium dithionite in 25 mM MOPS pH 7.0, prepared as described in Iron removal from $PH(LNO)_2$, and 24 μL of a freshly prepared 5 mM solution of $Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$ in deionized water, were added. The reaction mixture was incubated at room temperature for 3 h under a continuous flow of nitrogen. Excess, unbound iron was removed by gel filtration on a PD10 desalting column (Pharmacia) equilibrated in 25 mM MOPS pH 7.0, having 200 mM NaCl, 5% glycerol and 2 mM L-cysteine. Enzymatic assays performed on fractions containing the purified PHK–PH(LNO) complex revealed that no catalytic activity could be recovered under these conditions (Fig. 5A). Moreover, colorimetric assays using ferrozine to detect iron (see Other methods) demonstrated that the metal was not present in the protein sample tested (Fig. 5B).

In contrast, addition of excess iron to iron-depleted $PH(LNO)_2$ or PHM-$PH(LNO)_2$ complexes resulted in the significant recovery of both enzymatic activity and iron content of the purified protein (Fig. 5). Recovery of the maximal specific activity was of 35.1 ± 3.5% for $PH(LNO)_2$ and 78.9 ± 2.5% for PHM–$PH(LNO)_2$. The reason for the different values for the $PH(LNO)_2$ and PHM–$PH(LNO)_2$ complexes can be explained by the fact that while the formation of apo-hydroxylase from the $PH(LNO)_2$ species resulted in the complete removal of iron from the sample, ~50% of the original iron was retained when the metal depletion procedure was conducted with PHM–$PH(LNO)_2$. Importantly, these results were not dependent of the ratio of PHK to $PH(LNO)_2$ employed in the iron reconstitution procedures.

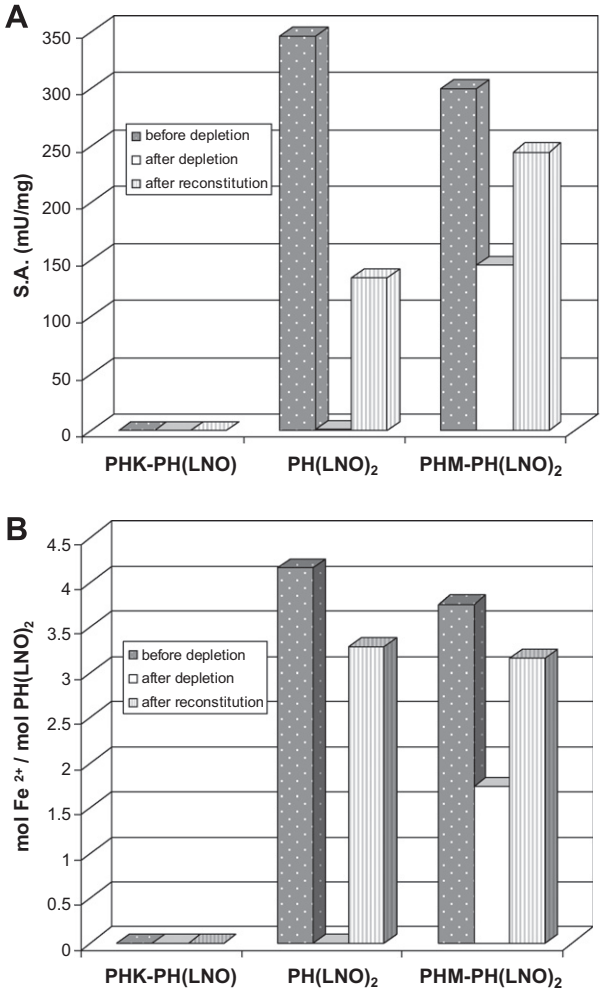


Fig. 5. Effect of iron depletion and reconstitution on the different purified hydroxylases expressed from plasmid pGEM3Z/*phΔp*. Iron content and phenol hydroxylase activity were measured as described in Materials and methods. Histogram A and B show the iron content and the residual phenol hydroxylase activities of the samples before and after depletion/reconstitution procedure, respectively.

Role of PHK in iron uptake of apo-PH(LNO)$_2$

Iron-depleted $PH(LNO)_2$, hereafter referred to as apo-$PH(LNO)_2$, was further purified by gel filtration on a S-200 gel filtration column. Two different species were obtained and identified by native-PAGE analysis and analytical gel filtration, which showed the presence of both hexameric apo-$PH(LNO)_2$ and trimeric apo-PH(LNO) complexes. Attempts to reconstitute the complexes with iron were performed by adding aliquots of a freshly prepared 50 μM solution of Fe(NH4)2(SO4)2·6H_2O in deionized water to 0.75 μM of either species in a 1 mL quartz cuvette. Immediately following iron addition, the other components of the phenol hydroxylase complex, the oxidoreductase PHP and the regulatory protein PHM, were added at a ratio of 2:1 and 4:1 with respect to $PH(LNO)_2$. Spectrophotometric continuous coupled assays were performed using phenol as substrate to monitor the rate of catechol formation.

Iron quantization of iron-reconstituted samples of apo-$PH(LNO)_2$ and apo-PH(LNO), in the absence or in the presence of PHK, always resulted in the measure of non-specific binding of iron to the hydroxylase. This is a common problem of *in vitro* iron reconstitution procedures. Since the enzymatic activity of the hydroxylase is related to the amount of iron correctly incorporated in the active site, we decided to indirectly measure iron binding by monitoring the increase in activity upon metal addition.

In the case of apo-PH(LNO) no enzymatic activity could be recovered after incubation with increasing concentration of iron (0–10 μM), neither in the presence of various amounts of recombinant PHK nor in its absence.

In the case of apo-$PH(LNO)_2$, different results were obtained in the absence and presence of PHK, which was added in a optimized ratio of 0.35:1 with respect to the iron-depleted hydroxylase protein (Fig. S1, Supplementary Material). As shown in Fig. 6, in the presence of PHK, addition of stoichiometric quantities of iron (II)

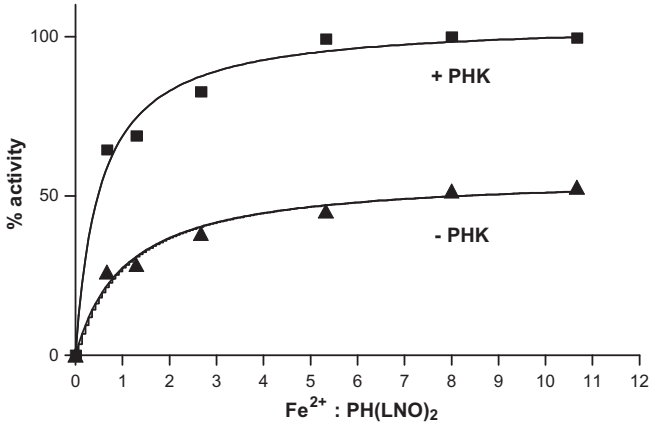


Fig. 6. Effect of increasing concentration of $Fe(NH_4)_2(SO_4)_2$ on the phenol hydroxylase activity of PH. Activity assays were performed as described in the Experimental procedures section in the presence (filled squares) or in the absence (filled triangles) of 0.35 equiv. of PHK with respect to apo-$PH(LNO)_2$. Values are normalized to the maximum specific activity obtained in the presence of accessory protein PHK.

to apo-PH(LNO)$_2$ resulted in the recovery of a maximal specific activity. Conversely, in the absence of PHK, only partial recovery (~50%) of apo-PH(LNO)$_2$ was possible when excess iron (II) was added to the reaction mixtures.

We cannot exclude the possibility that addition of a greater excess of iron (II) to apo-PH(LNO)$_2$ might allow for recovery of 100% of the activity obtained in the presence of PHK. Oxidation of iron (II) to iron (III) during the enzymatic assay hinders the use of high concentration of exogenous iron due to a background effect on the ABS410 monitored during the course of the reaction (see Enzymatic assays).

PHK-independent iron insertion into the hydroxylase at high iron (II) concentration was also demonstrated by the iron reconstitution experiments performed on whole cell extracts expressing either pGEM3Z/*ph* or pGEM3Z/*phΔk* and described in Activity assays of whole cells of *E. coli*-JM109 expressing recombinant phenol hydroxylase in the presence or absence of PHK. Using the same experiments performed on purified proteins and described in PHK-mediated inhibition of phenol hydroxylase activity (Fig. 5), apo-PH(LNO)$_2$ recovered specific activity with values comparable to those obtained in the presence or in the absence of accessory protein PHK when excess of exogenous iron (II) was added to the cell extracts. Together, these results suggest that PHK might facilitate iron uptake in the hexameric apo-PH(LNO)$_2$ complex, especially when the transition metal is present at low concentrations.

To evaluate the potential role of iron in the possible dissociation of the hexameric complex upon incubation with PHK, we monitored the distribution and subunit composition of the species obtained from gel-filtration experiments performed on iron-depleted samples incubated under different conditions (Fig. 7).

The identities of the subunits eluting in different peaks were assessed by a proteomic approach: 1 mL fractions taken from peak centers were digested with trypsin and analyzed by MALDI-TOF or, when ambiguous results were obtained, by capillary LC-MSMS. Proteins were then identified by comparing the experimental data to an in-house database using the MASCOT software. The protein sequences of PHL, PHN, PHO, PHK, and PHM were included in the search.

As observed with fully active holo-PH(LNO)$_2$, incubation of 2.5 μM of apo-PH(LNO)$_2$ with 1.25 μM of PHK at room temperature for 90 min, in a final volume of 200 μL, induced the dissociation of the hexameric species and facilitated the subsequent formation of PHK-PH(LNO) complexes, as measured gel-filtration experiments (Fig. 7, red line, and Table S3, Supplementary Material). Addition of either iron (10 μM) or exogenous regulatory protein PHM (10 μM) to the incubation mixture did not significantly alter the dissociation profile (data not shown).

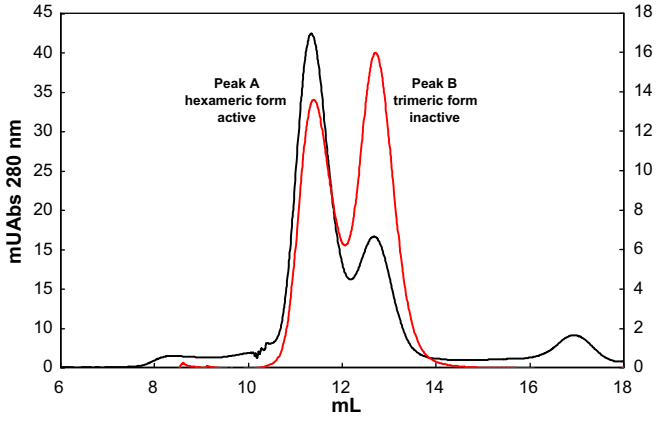


Fig. 7. Characterization of the apo-PH(LNO)$_2$ complex in the presence of PHK, PHM and iron. Elution profiles of an analytical gel filtration chromatographic separation of a 200 μL mixture of 2.5 μM of apo-PH(LNO)$_2$ and 1.25 μM of PHK (red line), and of a mixture of 2.5 μM of apo-PH(LNO)$_2$, 1.25 μM of PHK and 10 μM of PHM in the presence of 10 μM Fe^{2+} (black line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Importantly, a significant increase of the relative abundance of active hexameric species was observed with respect to the inactive trimeric form when exogenous recombinant PHM (10 μM) and iron (10 μM) were added simultaneously to the reaction mixture containing apo-PH(LNO)$_2$ and PHK (Fig. 7, black line). Moreover, in the later experiment a significant fraction of the regulatory protein PHM was bound to the active hexameric PH(LNO)$_2$ hydroxylase in a PHM-PH(LNO)$_2$ complex (Fig. 7, black line, and Table S3, Supplementary Material).

Unexpectedly, PHK did not bind to pre-formed, trimeric apo-PH(LNO). When apo-PH(LNO) was incubated with PHK, up to a molar ratio of 1:8, we could not detect any PHK co-eluting with the trimeric species. In these experiments the three hydroxylase subunits PHL, PHN, and PHO could be easily identified by proteomics methods (data not shown).

This result strongly suggests that PHK binds to the hexameric form (either apo- or holo-) of the hydroxylase before inducing dissociation to the PHK-PH(LNO) complex.

Discussion

Microorganisms devote part of their genetic and biochemical resources to ensure that cofactors are properly inserted into the active sites of metalloenzymes. Proteins responsible for the tightly regulated homeostasis of metals are generally known as metallochaperones [31,32]. These proteins also play a key role in preventing toxic metals – such as iron and copper – from engaging in non-specific interactions with cellular membranes, proteins, or DNA [32,40,41].

Based on biochemical studies of small BMM proteins similar to PHK from *Pseudomonas* sp. OX1, such as MMOD from *M. capsulatus* (Bath) and DmpK from *Pseudomonas* sp. CF600 [26,27], several roles could be envisaged for PHK. Possible functions include a molecular chaperone, a metal delivery factor, or a metallochaperone. Interestingly, sequences with high homology to that encoding PHK have been identified in most of the phenol hydroxylase operons sequenced so far [4,5], clearly suggesting a critical role for these components in this specific subfamily of BMMs.

We initially observed that *E. coli*-strain JM109 cells co-expressing recombinant multicomponent phenol hydroxylase from *Pseudomonas* sp. OX1 as well as the accessory subunit PHK were characterized by an improved catalytic ability to convert phenol to catechol when compared to the same strain expressing the PH operon devoid of the *phk* gene. This difference was confirmed when crude extracts of induced cells were used in the enzymatic assays instead of whole cells. Whereas total iron content was similar in cells expressing the two different *ph* gene clusters, a significant difference was observed in the ability of the cells to hydroxylate phenol. Surprisingly, cell extracts lacking recombinant PHK demonstrated an increase in hydroxylase activity when excess iron (II) was added. This result is different from what was previously observed with the recombinant hydroxylase of multicomponent PH from *Pseudomonas* sp. CF600 [26]. In this case the activity of crude extracts of cells not expressing the accessory component DmpK could be detected only after exogenous iron (II) and DmpK were both added, supporting the hypothesis that DmpK is involved not only in the delivery but also the proper insertion of iron into the active site of the hydroxylase [26].

The role of PHK was investigated here by initially studying its influence on the phenol hydroxylase complex reconstituted from the purified PHP (oxidoreductase), PHM (regulatory protein) and

$PH(LNO)_2$ (hydroxylase) components [20]. PHK was purified to homogeneity and found to be expressed as a dimer, devoid of any metal or organic cofactor. PHK was not essential for the enzymatic activity of the PH complex, when assayed with phenol, and instead inhibited the reaction even when present at low ratios with respect to the hydroxylase $PH(LNO)_2$ (Fig. 3).

An insight into the mechanism of phenol hydroxylase inhibition by PHK was obtained by characterizing the molecular species formed after incubating the hydroxylase complex $PH(LNO)_2$ with a sub-stoichiometric amount of purified PHK (Fig. 4). Native gel electrophoresis and gel-filtration experiments followed by mass spectrometric identification of the species formed upon incubation, showed that this is a novel molecular species comprised of PH(LNO) tightly associated with monomeric PHK. Biochemical assays revealed that this species is devoid of both iron and catalytic activity, providing an explanation for the inhibitory effect observed when exogenous PHK was added to reconstituted phenol hydroxylase in *in vitro* enzyme assays. The ability of PHK to promote dissociation of the hexameric complex $PH(LNO)_2$ was confirmed by the observation that inactive PHK–PH(LNO) could be isolated upon purification of cell extract of *E. coli*-strain JM109 expressing the complete *ph* gene cluster from plasmid pGEM3Z/*phΔp* (Fig. 1). In the same gel-filtration experiment the yield of the PHK–PH(LNO) complex was significantly lower when complex PHM–$PH(LNO)_2$ instead of $PH(LNO)_2$ was incubated with PHK (Fig. 4B). These results suggest a possible shielding effect of PHM towards the PHK-mediated dissociation of the $PH(LNO)_2$ hexamer, likely occurring in the absence of catalytic activity (Fig. 3).

Unexpectedly, experiments aimed at reactivating the PHK–PH(LNO) complex by the addition of exogenous iron (II) were not successful. Iron removal and restitution of $PH(LNO)_2$ and PHM-$PH(LNO)_2$ was possible when excess iron (II) was added, results that were somewhat independent of PHK (Fig. 5). Interestingly, iron removal from the hydroxylase was incomplete when PHM was present (Fig. 5). This "protective role" of the coupling protein PHM, observed in several experiments presented here, is similar to what previous reports of many BMMs described, as in the case of the sMMO system from *M. capsulatus* (Bath) [35].

Further insight on what appeared as contradictory results on the possible role of PHK came from attempts to incorporate iron into apo-$PH(LNO)_2$ hydroxylase. In the presence of the accessory protein PHK, complete recovery of the catalytic activity of apo-$PH(LNO)_2$ occurred when a stoichiometric concentration of iron (II) was added to the reaction mixture (Fig. 6). This observation is likely to be physiologically relevant. Notably, iron uptake and recovery of catalytic activity occurred at a much higher iron (II) concentration in the absence of PHK, a condition that is probably not compatible with the physiological concentration of this transition metal in the cell. This consideration also explains the observation that enzymatic activity was recovered when cell extracts of pGEM3Z/*phΔk* were incubated with iron (II), results that suggest the possibility that iron (II) can be inserted into apo-$PH(LNO)_2$ without the addition of exogenous PHK only when high concentrations of iron were present. In fact, many experiments probing PH were performed in the presence of an excess of $Fe(NH_4)_2(SO_4)_2$. Given the result presented here it is possible that these unnatural conditions might have prevented detection of the physiological role of PHK-like proteins in early experiments on BMMs. These experiments presented here show that at least in the PH from *Pseudomonas* sp. OX1, the accessory protein PHK is not necessary for proper diiron cluster assembly *in vitro*, a result that is different from what has been shown for the related component DmpK in *Pseudomonas* sp. CF600 [26].

Taking into account the data presented in this work and the information available in the literature for the accessory components MMOD [27] and DmpK [26], a possible mechanism (albeit incomplete) that does account for most of the observations and in which PHK might be involved can be hypothesized (Fig. 8).

Lacking experimental evidence on early post-translational events, the model assumes apo-$PH(LNO)_2$ as the starting point of the sequence of events. PHK transiently binds to the apo-hexamer, influencing the insertion of intracellular iron in the active site, and

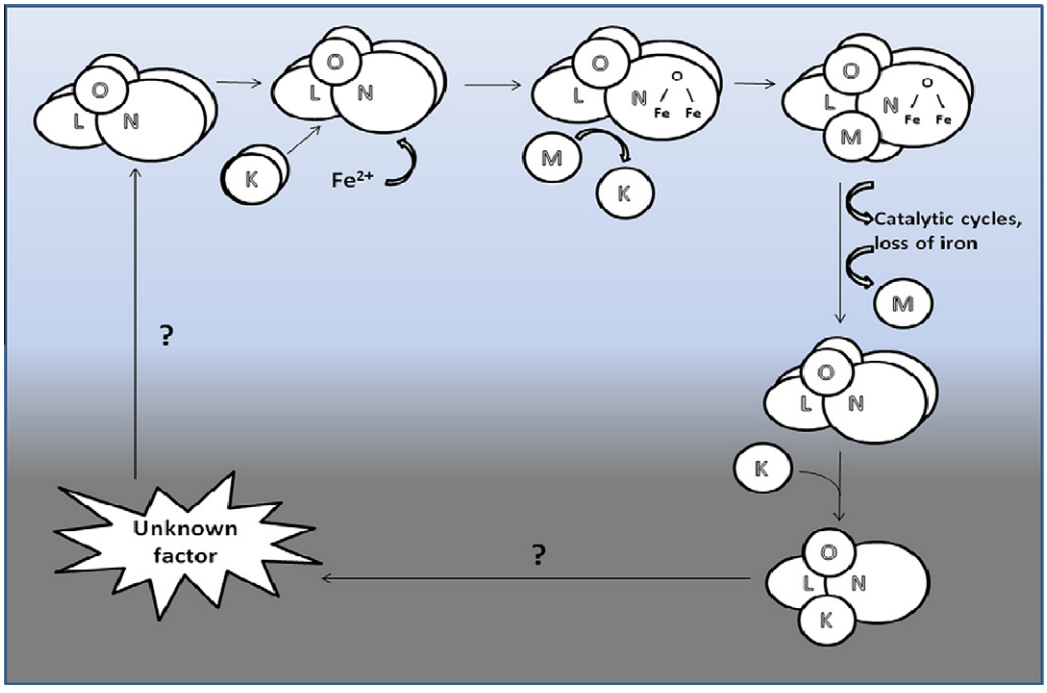


Fig. 8. Model showing a possible role for PHK. A possible model suggested by the experiments described in this paper and the data presented in literature for DmpK and MMOD accessory proteins [26,27]. In dark grey is the sequence of events that is still under investigation. The presence of the oxidoreductase PHP in the PH catalytic cycle has been omitted in the scheme for clarity.

thereby facilitating the formation of holo-PH(LNO)$_2$, the active hydroxylase complex.

Based on the ability of PHM to "protect" holo-PH(LNO)$_2$ from PHK-mediated dissociation (Fig. 4 and Fig. 7), our model proposes that the regulatory component PHM binds to the holo-PH(LNO)$_2$ complex immediately after the PHK-facilitated insertion of iron (II). This event displaces bound PHK, which can then be immediately recycled to assist in iron insertion in newly assembled apo-PH(LNO)$_2$ molecules.

In support of the existence of an interplay between the accessory and the regulatory proteins in BMMs, it should be underlined that Sazinsky and coworkers [35], using the purified components of the sMMO from *M. capsulatus* (Bath), highlighted, by means of ITC binding experiments, a preferential binding of the regulatory protein MMOB to the holo-hydroxylase complex MMOH and of the accessory protein MMOD to the apo-MMOH complex. Moreover, Powlowski and coworkers [26] had previously proposed that the accessory component DmpK likely interferes with the interaction between the regulatory protein DmpM and the hydroxylase moiety Dmp(LNO)$_2$.

Once the regulatory component PHM is bound to the PH(LNO)$_2$ active complex, the iron cofactor is protected within the active site from accidental loss, as indicated by our iron (II) removal and reconstitution experiments performed on the PHM–PH(LNO)$_2$ complex (Fig. 5). A new catalytic cycle can then be initiated (Fig. 8).

The protective role of the regulatory component in the phenol hydroxylase system of *Pseudomonas* sp. OX1 is particularly evident in the crystal structure of the PHM–PH(LNO)$_2$ complex [19]. This structure revealed that PHM blocks the entrance of solvent and substrate to the diiron active site, possibly offering protection of species formed at the dioxygen-activated metal center as it proceeds through its reaction cycle [19].

Our current hypothesis is that PHK does not directly bind iron, as suggested by ITC experiments, but instead might be responsible for altering the local conformation of apo-PH(LNO)$_2$, thus facilitating direct access of iron into the hydroxylase active site. We are currently investigating whether this observation is related to the direct binding of iron to an apo-PH(LNO)$_2$–PHK transitory complex or is a consequence of a conformational change induced by PHK in the iron-depleted hydroxylase moiety apo-PH(LNO)$_2$. The presence of iron in the active site of the PH(LNO)$_2$ hydroxylase could be made irreversible by rapid addition of O_2 once the two irons are in place, as already suggested in literature [42,43].

Among others, the question arises whether the PHK–PH(LNO) complex is of physiological importance. Our current hypothesis is that during repeated catalytic cycles the hydroxylase might lose iron which could induce the formation of the PHK–PH(LNO) species we observed *in vitro* (lower part of the scheme in Fig. 8). These complexes might reorganize, with the aid of further, yet undiscovered, auxiliary proteins, to form apo-PH(LNO)$_2$ species. Alternatively, the presence of PHK might be necessary to avoid the occurrence of holo-PH(LNO)$_2$ molecules *in vivo* in the absence of the regulatory protein PHM. In this case, the oxidoreductase PHP would still be able to transfer electrons to the hydroxylase moiety [20], leading to an uncoupling of NADH consumption and product formation, or a premature reduction of the oxygenated metal cluster, which would not only consume the reactive diiron species but also deplete the NADH supply of the cell in a wasteful manner [19].

At this stage we cannot exclude the possibility that formation of inactive PHK–PH(LNO) complexes could be just an artifact of our system due to the presence of high intracellular concentrations of recombinant PH(LNO)$_2$ and PHK. This phenomenon would shift the association equilibrium toward the formation of the PHK–PH(LNO) species. Nevertheless, the stable PHK–PH(LNO) complex might be a valuable tool to shed light on the molecular determinants responsible for the interaction between the accessory protein and the hydroxylase subunits PHL, PHN and PHO in future experiments.

In conclusion, novel details of the role of the accessory proteins in bacterial multicomponent monoooxygenases have been highlighted in the experiments presented here. These results add to what has been previously reported for accessory proteins MMOD and DmpK [26,27].

The similarities and differences between the different accessory proteins characterized to date lead to the conclusion that in the BMMs family seemingly similar protein components can have notably different ways of influencing the assembly of the diiron cluster in the active site of the hydroxylase moiety.

Acknowledgments

The authors are indebted to Prof. Giuseppe D'Alessio, Università di Napoli Federico II, for critically reading the manuscript, and to Dr. Luigi Martino and Prof. Tina Giancola, Università di Napoli Federico II, for ITC experiments.

This work was supported by a grant from the Ministry of University and Research (PRIN/2007).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.abb.2010.09.023.

References

[1] S. Dagley, in: J. Sokatch, L.N. Ornston (Eds.), The Bacteria, vol. 10, Academic Press, Orlando, 1986, pp. 527–555.
[2] S. Harayama, K.N. Timmis, in: H. Sigel, A. Sigel (Eds.), Metal Ions in Biological Systems, vol. 28, Marcel Dekker, New York, 1992, pp. 99–156.
[3] E. Díaz, Int. Microbiol. 7 (2004) 173–180.
[4] E. Notomista, A. Lahm, J. Mol. Evol. 56 (2003) 435–445.
[5] J.G. Leahy, P.J. Batchelor, S.M. Morcomb, FEMS Microbiol. Rev. 27 (2003) 449–479.
[6] M. Merkx, D.A. Kopp, M.H. Sazinsky, J.L. Blazyk, J. Müller, S.J. Lippard, Angew. Chem. Int. Ed. 40 (2001) 2783–2807.
[7] J.D. Pikus, J.M. Studts, C. Achim, K.E. Kauffmann, E. Münck, R.J. Steffan, K. McClay, B.G. Fox, Biochemistry 35 (1996) 9106–9119.
[8] N.L. Elsen, L.A. Moe, L.A. McMartin, B.G. Fox, Biochemistry 46 (2007) 976–986.
[9] M.H. Sazinsky, S.J. Lippard, Acc. Chem. Res. 39 (2006) 558–566.
[10] L.J. Murray, S.J. Lippard, Acc. Chem. Res. 40 (2007) 466–474.
[11] H. Saeki, M. Akira, K. Furuhashi, B. Averhoff, G. Gottschalk, Microbiology 145 (1999) 1721–1730.
[12] G. Baggi, P. Barbieri, E. Galli, S. Tollari, Appl. Environ. Microbiol. 53 (1987) 2129–2132.
[13] G. Bertoni, F. Bolognese, E. Galli, P. Barbieri, Appl. Environ. Microbiol. 62 (1996) 3704–3711.
[14] G. Bertoni, M. Martino, E. Galli, P. Barbieri, Appl. Environ. Microbiol. 64 (1998) 3626–3632.
[15] P. Barbieri, F.L. Arenghi, G. Bertoni, F. Bolognese, E. Galli, Antonie Van Leeuwenhoek 79 (2001) 135–140.
[16] F.L. Arenghi, D. Berlanda, E. Galli, G. Sello, P. Barbieri, Appl. Environ. Microbiol. 67 (2001) 3304–3308.
[17] V. Cafaro, R. Scognamiglio, A. Viggiani, V. Izzo, I. Passaro, E. Notomista, Eur. J. Biochem. 269 (2002) 5689–5699.
[18] M.H. Sazinsky, J. Bard, J. Biol. Chem. 279 (2004) 30600–30610.
[19] M.H. Sazinsky, P.W. Dunten, M.S. McCormick, A. Di Donato, S.J. Lippard, Biochemistry 45 (2006) 15392–15404.
[20] V. Cafaro, V. Izzo, R. Scognamiglio, E. Notomista, P. Capasso, A. Casbarra, P. Pucci, Appl. Environ. Microbiol. 70 (2004) 2211–2219.
[21] I. Nordlund, J. Powlowski, V. Shingler, J. Bacteriol. 172 (1990) 6826–6833.
[22] J. Powlowski, V. Shingler, J. Bacteriol. 172 (1990) 6834–6840.
[23] J. Powlowski, V. Shingler, Biodegradation 5 (1994) 219–236.
[24] E. Cadieux, V. Vrajmasu, C. Achim, J. Powlowski, E. Münck, Biochemistry 41 (2002) 10680–10691.
[25] L.M. Newman, L.P. Wackett, Biochemistry 34 (1995) 14066–14076.
[26] J. Powlowski, J. Sealy, V. Shingler, E. Cadieux, J. Biol. Chem. 272 (1997) 945–951.
[27] M. Merkx, S.J. Lippard, J. Biol. Chem. 277 (2002) 5858–5865.
[28] D.L. Cardy, V. Laidler, G.P. Salmond, J.C. Murrell, Mol. Microbiol. 5 (1991) 335–342.
[29] H. Herrmann, C. Müller, I. Schmidt, J. Mahnke, L. Petruschka, K. Hahnke, Mol. Gen. Genet. 247 (1995) 240–246.
[30] P.M. Santos, I. Sá-Correia, J. Biotechnol. 131 (2007) 371–378.

[31] M.W. Ribbe, B.K. Burgess, Proc. Natl. Acad. Sci. USA 98 (2001) 5521–5525.
[32] J. Kuchar, R.P. Hausinger, Chem. Rev. 104 (2004) 509–525.
[33] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
[34] A. Viggiani, L. Siani, E. Notomista, L. Birolo, P. Pucci, J. Biol. Chem. 279 (2004) 48630–48639.
[35] M.H. Sazinsky, M. Merkx, E. Cadieux, S. Tang, S.J. Lippard, Biochemistry 43 (2004) 16263–16276.
[36] U. Laemmli, Nature 227 (1970) 680–685.
[37] C.J. Batie, E. LaHaie, D.P. Ballou, J. Biol. Chem. 262 (1987) 1510–1518.
[38] M.M. Bradford, Anal. Biochem. 72 (1976) 248–254.
[39] S. Hart, K.R. Koch, D.R. Woods, J. Gen. Microbiol. 138 (1992) 211–216.
[40] K.G. Daniel, R.H. Harbach, W.C. Guida, Q.P. Dou, Front. Biosci. 9 (2004) 2652–2662.
[41] T.B. Bartnikas, J.D. Gitlin, Nat. Struct. Biol. 8 (2001) 733–734.
[42] H.H. Nguyen, J. Ge, D.L. Perlstein, J. Stubbe, Proc. Natl. Acad. Sci. USA 96 (1999) 12339–12344.
[43] J. Ge, D.L. Perlstein, H.H. Nguyen, G. Bar, R.G. Griffin, J. Stubbe, Proc. Natl. Acad. Sci. USA 98 (2001) 10067–10072.

Please cite this article in press as: V. Izzo et al., Arch. Biochem. Biophys. (2010), doi:10.1016/j.abb.2010.09.023

Table S1. ***Biotinylated PHK partners isolated by avidin pull-down and identified by LC-MSMS mass spectrometry.*** Proteomic identification of the major protein bands detected in the pull-down assay (Fig.2) using biotinylated PHK and cell extracts of *P*.sp.OX1 grown on phenol as unique carbon and energy source.

Protein band	***Identified protein***	***Accession No.***	***matched peptides***	***sequence coverage (%)***
A	**PHN**	gi|28848925 Q84AQ2	40	77
B	**PHL**	gi|28848923 Q84AQ4	27	82
C	**PHO**	gi|28848926 Q84AQ1	3	26
	PHK	gi|28848922 Q93JV2	4	26

Table S2. ***Identification of the subunit components in the phenol hydroxylase complex fractionated by native gel electrophoresis.*** The protein band indicated with an arrow in Fig. 4A was excised from the gel and digested *in situ* with trypsin. The peptide mixture was analyzed by LC-MS/MS and the experimental data used for searching the database with MASCOT software.

Phenol Hydroxylase component	***matched peptides***	***sequence coverage (%)***
PHN	44	66
PHL	29	71
PHO	4	45
PHK	8	68

Table S3. ***Identification of the subunits in the chromatographic peaks as separated in Fig. 7.*** Proteins were identified in the peak top fractions after tryptic digestion and mass spectrometric analysis.

Sample	***Peak A***			***Peak B***		
	Subunits	***N° of matched peptides***	***sequence coverage (%)***	***Subunits***	***N° of matched peptide***	***sequence coverage (%)***
Apo $PH(LNO)_2$ incubated with PHK (Fig.5A-red line)	PHN [a] PHL [a] PHO [a]	23 22 4	46 59 33	PHN [a] PHL [a] PHO [a] PHK [a]	24 23 5 7	46 61 38 43
Apo $PH(LNO)_2$ incubated with PHK, PHM and Fe^{2+} (Fig. 5A-black line)	PHN [b] PHL [b] PHO [b] PHM [b]	22 24 5 3	47 75 46 55	PHN [a] PHL [a] PHO [a] PHK [a]	23 24 5 9	46 63 38 64

[a] MALDI-TOF analysis.
[b] Capillary LC-MSMS analysis.

Figure S1

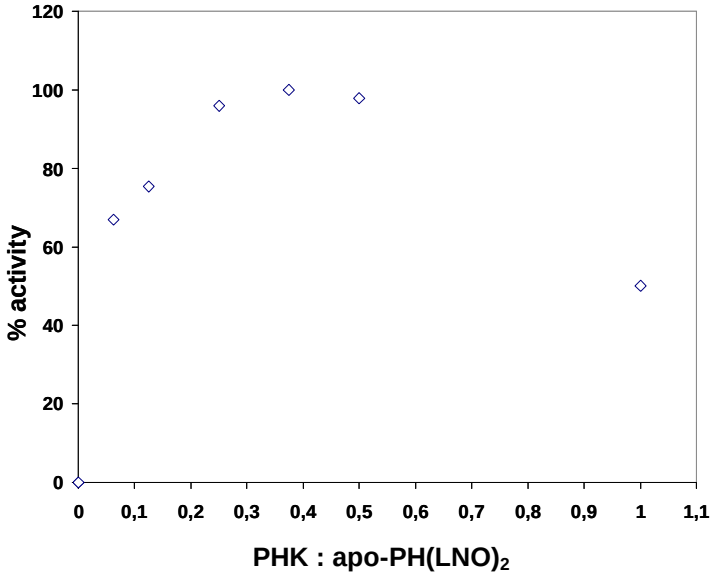


Fig. S1..***Effect of different concentrations of accessory protein PHK with respect to the apo-PH(LNO)$_2$ hydroxylase complex on the rate of catechol production.*** The experiment was performed in the presence of 5 μM of $Fe(NH_4)_2(SO_4)_2$ and in the experimental conditions described in section 3.5

7542 *Biochemistry* **2010,** *49,* 7542–7552
DOI: 10.1021/bi100447t

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Stoichiometry and Topology of the Complex of the Endogenous ATP Synthase Inhibitor Protein IF_1 with Calmodulin[†]

Daniela Pagnozzi,[‡,▽] Leila Birolo,*[,‡,§] Gabriella Leo,[‡] Stefania Contessi,[⊥] Giovanna Lippe,[#] Pietro Pucci,[‡,∥] and Irene Mavelli[⊥]

[‡]*Department of Organic Chemistry and Biochemistry, and* [§]*School of Biotechnological Sciences, University of Napoli "Federico II", via Cynthia 6, 80126 Napoli, Italy,* [∥]*CEINGE-Biotecnologie Avanzate, Napoli, Italy,* [⊥]*Department of Biomedical Sciences and Technologies, MATI Centre of Excellence, University of Udine, Piazzale Kolbe 4, I-33100 Udine, Italy, and* [#]*Department of Food Science, University of Udine, via Sondrio 2/A, 33100 Udine, Italy.* [▽]*Present address: Porto Conte Ricerche Srl, Loc. Tramariglio, SP 55 Porto Conte/Capo Caccia Km 8.400, 07041 Alghero (SS), Italy.*

Received March 25, 2010; Revised Manuscript Received July 29, 2010

ABSTRACT: IF_1, the natural inhibitor protein of F_OF_1ATP synthase able to regulate the ATP hydrolytic activity of both mitochondrial and cell surface enzyme, exists in two oligomeric states depending on pH: an inactive, highly helical, tetrameric form above pH 6.7 and an active, inhibitory, dimeric form below pH 6.7 [Cabezon, E., Butler, P. J., Runswick, M. J., and Walker, J. E. (2000) *J. Biol. Chem. 275*, 25460−25464]. IF_1 is known to interact *in vitro* with the archetypal EF-hand calcium sensor calmodulin (CaM), as well to colocalize with CaM on the plasma membrane of cultured cells. Low resolution structural data were herein obtained in order to get insights into the molecular interaction between IF_1 and CaM. A combined structural proteomic strategy was used which integrates limited proteolysis and chemical cross-linking with mass spectrometric analysis. Specifically, chemical cross-linking data clearly indicate that the C-terminal lobe of CaM molecule contacts IF_1 within the inhibitory, flexible N-terminal region that is not involved in the dimeric interface in IF_1. Nevertheless, native mass spectrometry analysis demonstrated that in the micromolar range the stoichiometry of the IF_1−CaM complex is 1:1, thereby indicating that binding to CaM promotes IF_1 dimer dissociation without directly interfering with the intersubunit contacts of the IF_1 dimer. The relevance of the finding that only the C-terminal lobe of CaM is involved in the interaction is two fold: (i) the IF_1−CaM complex can be included in the category of noncanonical structures of CaM complexes; (ii) it can be inferred that the N-terminal region of CaM might have the opportunity to bind to a second target.

Bovine IF_1, the natural inhibitor protein of F_OF_1ATP synthase, consists of an 84 residue long monomer which folds into a single cationic amphiphilic α-helix (*1*). In solution, two IF_1 molecules dimerize by forming an antiparallel α-helical coiled-coil structure in the C-terminal region, leading the N-terminal regions to be available for binding to F_OF_1 (*1*) and other putative partners (*2*). The dimer is the predominant form at acidic pH, while at pH 6.5, it is in equilibrium with a tetrameric form which becomes predominant at basic pH (*3*).

Pedersen and Carafoli groups had previously reported that IF_1 is a target for Calmodulin (CaM[1]) (*4*, *5*), the archetypal EF-hand calcium sensor (*6*−*9*). CaM interacts and regulates numerous target proteins that are structurally and functionally unrelated, including metabolic enzymes, structural proteins, transcription factors, ion channels, and pumps, and modulates a wide range of cellular processes in response to calcium (*10*−*12*). The list of known CaM-dependent proteins is extensive, exceeds 300 in number (*13*−*15*), with the diversity of targets being strictly dependent on the ability of CaM to interact with its targets in different ways (*11*, *15*−*17*).

The best known mode of interaction of CaM with the target is the classical wrap-around mode with the target peptide engulfed in a hydrophobic channel (*18*) and the two lobes of CaM coming close to one other (*18*−*21*). Although the wrap-around binding model was initially assumed to be the predominant structural mechanism for CaM binding to its targets, a growing body of evidence demonstrated that CaM can also bind to targets in an unusual compact or extended conformations in some instances with less than a full complement of bound calcium ions, as well as with novel stoichiometries (*10*, *20*). In the case of the plasma membrane Ca^{2+}-pump, $4Ca^{2+}$-CaM binding to its peptide C20W involves only the C-terminal lobe, as a consequence of a missing hydrophobic anchor residue with respect to the common target sequences (*22*). A radically novel target-binding mode has been revealed with the crystal structure of $2Ca^{2+}$-CaM in complex with a fragment of K^+-channel (*23*) and exotoxin from *Bacillus anthracis* (*23*). In all these structures, CaM maintains an extended, open conformation after binding to its target.

Our *in vitro* fluorimetric analyses clarified that both the bovine inhibitor protein of F_OF_1ATP synthase IF_1 and yeast inhibitors IF_1 and STF_1 interact with Ca^{2+}-saturated CaM with a 1:1 ratio

[†]This work was supported by Italian MIUR grants, PRIN 2007, and FIRB Italian Human ProteomeNet Project 2007. MRC Medical Research Council (Centre Cambridge) is acknowledged for having kindly provided bovine IF1 protein in accordance with the Material Transfer Agreements made among MRC, University of Udine, and University of Napoli "Federico II".

*To whom correspondence should be addressed: Department of Organic Chemistry and Biochemistry, University of Napoli "Federico II", via Cynthia 6, 80126 Napoli, Italy. Tel: +39-081-674114. Fax: +39-081-674113. E-mail: birolo@unina.it.

[1]Abbreviations: CaM, calmodulin; CNBr, cyanogen bromide; EDAC, 1-ethyl-3-(3-dimethylamino-propyl)-carbodiimide; TFA, trifluoroacetic acid.

 Published on Web 07/29/2010

over a wide pH range. Moreover, sequence visual inspection, together with binding studies of CaM with bovine IF_1 synthetic peptides, revealed the presence of a conserved putative CaM-binding motif within the N-terminal of IF_1 from bovine and yeast (*2*).

In addition, we have reported that IF_1 expressed on the cell surface colocalizes with CaM in human HepG2 cells (*25*), suggesting the hypothesis that the interaction may occur *in vivo* at this level. Such a hypothesis is reinforced by recent reports documenting that IF_1 located on the plasma membrane is able to regulate the activity of F_OF_1ATP synthase expressed on the cell surface (*25*−*27*).

As mentioned above, IF_1 exists in two oligomeric states depending on pH: an inactive, highly helical, tetrameric form above pH 6.7 and an active, inhibitory, dimeric form below pH 6.7 (*3*). In the IF_1 dimer, the two monomers associate through an extended antiparallel α-helical coiled-coil in the C-terminal region, while the inhibitory active portion is in the N-terminal region (*28*). Structural data are needed in order to understand whether the binding of CaM overlaps the inhibitory portion as well as the stoichiometry of the IF_1−CaM complex. The aim of the present work was to characterize the complex between IF_1 and Ca^{2+}-CaM, by defining the stoichiometry and topology of the protein complex, in the experimental conditions in which the IF_1 is in the dimeric, inhibitory form, i.e., pH 5.

We investigated the stoichiometry of the IF_1−CaM complex by native ESI-MS analyses. Complex topology was investigated by a combined strategy, which integrates limited proteolysis and cross-linking experiments with mass spectrometric analyses.

Our data indicated that IF_1 and CaM in the physiological micromolar range form a 1:1 complex, with four calcium ions bound, and clearly demonstrated that only the C-terminal lobe of CaM is involved in the formation of the complex and contacts IF_1 within the inhibitory, flexible N-terminal region.

MATERIALS AND METHODS

Materials. Recombinant IF_1 was provided by MRC Medical Research Council (Centre Cambridge). Recombinant bovine calmodulin, TPCK treated trypsin, endoprotease Glu-C, chymotrypsin, the cross-linking reagent 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDAC), α-cyano-4-hydroxycinnamic acid, and sinapinic acid were purchased from Sigma. Acetonitrile was HPLC-grade from Baker.

Biotinylation of IF_1. IF_1 (344 μg, in 50 mM MES, pH 5.0) was incubated with 8 μg of Sulfo-NHS-Biotin (sulfosuccinimidobiotin) (Pierce) (molar ratio 1:1), dissolved in the same buffer, on ice in the dark for 2 h. The reaction was stopped with 0.1 M Tris at pH 7.0, and the excess of nonreacted biotin reagent was removed by size exclusion chromatography on PD-10 columns (GE-Healthcare) in 0.1 M sodium phosphate at pH 7.2 and 0.15 M NaCl, following absorbance at 220 and 280 nm to single out IF_1-containing fractions. These fractions were pooled, and the extent of biotinylation was verified by ESI-MS on an aliquot corresponding to ~1 nmol of IF_1 protein previously purified by reverse-phase HPLC on a ZORBAX Eclipse XDB-C8 column (150 × 4.6 mm, 80 Å pore size) with a 25−65% acetonitrile linear gradient in 0.1% TFA over 10 min, at a flow rate of 1 mL/min. Elution was monitored at 220 nm, and individual fractions were collected and analyzed by electrospray mass spectrometry on a QUATTRO Micro LC-MS/MS system equipped with a Z-SPRAY source and a triple quadrupole analyzer (Waters-Micromass). Protein solution was injected into the ion source at a flow rate of 10 μL/min. Calibration of the instrument was achieved with the multicharge distribution of the equine myoglobin. Data were elaborated using the MassLynx program (Waters-Micromass).

Immobilization of Biotinylated PHK on Avidin and the Pull Down Assay. Two hundred microliters of avidin agarose resin (settled gel, Pierce) was equilibrated with five volumes of binding buffer (0.1 M sodium phosphate at pH 7.2 and 0.15 M NaCl) and incubated with biotinylated IF_1 (2 mg of biotinylated IF_1 per mL of settled avidin agarose resin) at 4 °C for 1 h. The resin was washed with 10 volumes of 50 mM MES at pH 5.0.

CaM (200 μL, 5 μM in 50 mM MES, pH 5.0) was incubated with 200 μL of biotinylated-IF_1 immobilized on avidin agarose resin and incubated overnight at 4 °C. The resin was washed with 10 volumes of 50 mM MES at pH 5.0, and the retained species were eluted with 30 μL of denaturing buffer (15.6mM Tris, 1.25% SDS, 2.5% Glycerol, 0.2M DTT). Samples were analyzed by SDS−PAGE (12.5%−10 cm × 10 cm), and proteins were stained with Coomassie Brilliant Blue G-Colloidal (Pierce, Rockford, USA).

Gel Filtration. Gel filtration experiments were carried out as follows: 50 μL of the 20 μM sample in 50 mM ammonium acetate at pH 5.0, 0.15 M NaCl, and 100 μM $CaCl_2$ were loaded on a Superdex 75 PC (3.2 × 300 mm) gel filtration column installed on Smart System (Pharmacia LKB) and isocratically eluted at 25 °C at a flow rate of 75 μL/min. The column had been previously calibrated in 50 mM sodium phosphate at pH 7.5 and 0.15 M NaCl with the following proteins of known molecular mass: bovine serum albumin (66000 Da), ovalbumin (45000 Da), carbonic anhydrase (29000 Da), trypsinogen (24000 Da), and cytochrome *c* (12400 Da).

Native Mass Spectrometry. IF_1 and the IF_1−calmodulin mixture (1/1, mol/mol) were analyzed on a Q-TOF hybrid mass spectrometer equipped with a nano Z-spray source (Waters, Manchester, UK). Protein solutions, 10 μM in 10 mM ammonium acetate at pH 5.0 and 100 μM $CaCl_2$, were injected into the ion source at a flow rate of 1 μL/min. Data were elaborated using the MassLynx program (Waters-Micromass). Calibration of the instrument was achieved with the multicharge distribution of the bovine pancreatic tripsinogen.

Complementary Proteolysis Experiments. Enzymatic hydrolyses were performed at 37 °C by incubating the 10 μM sample in 10 mM ammonium acetate at pH 5.0 and 100 μM $CaCl_2$ with either trypsin, chymotrypsin, or endoprotease Glu-C, enzyme-to-substrate ratios ranging from 1:4000 to 1:100 (w/w). The extent of proteolysis was monitored on a time-course basis by sampling the reaction mixture at different time intervals from 1 to 60 min and directly analyzing the released peptides by MALDI-MS. For each measurement, an aliquot of 1 μL of peptide mixture was applied to a sample slide and mixed with 4 μL of a 10 mg/mL sinapinic acid solution in acetonitrile/0.2% TFA (70:30, v/v) before air drying. The samples were then analyzed on a linear Voyager DE MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA). The mass range was internally calibrated with the $[M + H]^+$ and $[M + 2H]^{2+}$ ions from the entire IF_1 protein present in the proteolytic mixture.

Chemical Cross-Linking Experiments. Cross-linking reactions on the IF_1 dimer were carried out in a total volume of 20 μL of 50 mM MES at pH 5.0. IF_1 (0.5 nmol) was incubated at 37 °C for 15 min and then treated with 1-ethyl-3-(3-dimethylamino propyl)-carbodiimide (EDAC) cross-linking reagent for 30 min.

Cross-linking reactions on the IF_1−calmodulin complex were carried out in a total volume of 20 μL of 50 mM MES at pH5.0,

1 mM $CaCl_2$, and 0.15 M NaCl. Then, 0.5 nmol of IF_1 and 0.5 nmol of CaM were incubated at 37 °C for 15 min and then treated with 1-ethyl-3-(3-dimethylamino propyl)-carbodiimide (EDAC) cross-linking reagent for 30 min.

Preliminary experiments were carried out both for the IF_1 dimer and the IF_1−CaM complex to determine the optimal excess of EDAC. Cross-linking reactions were terminated by quenching the excess of reagents with gel loading buffer (15.6 mM Tris-HCl, 1.25% SDS, 2.5% glycerol at pH 6.8, and 0.2 M DTT). The protein samples were heated at 100 °C for 5 min and separated by electrophoresis on a 15% SDS−polyacrylamide gel. Proteins were detected by colloidal Coomassie, and selected positively stained protein bands were excised from the gel, *in situ* digested with trypsin, and analyzed by MALDI-MS for protein identification.

In IF_1 dimer cross-linking experiments on a preparative scale, 6 nmol in 120 μL of IF_1 were incubated at 37 °C for 15 min and then treated with a 10-fold molar excess of EDAC for 30 min at 37 °C. The reaction was stopped by acidification with 1% TFA. The cross-linked protein was purified by reverse-phase HPLC on a ZORBAX Eclipse XDB-C8 column (150 × 4.6 mm, 80 Å pore size) with a 15−65% acetonitrile linear gradient in 0.1% TFA over 30 min, at a flow rate of 1 mL/min. Elution was monitored at 220 nm, and individual fractions were collected and analyzed by electrospray mass spectrometry on a QUATTRO Micro LC-MS/MS system equipped with a Z-SPRAY source and a triple quadrupole analyzer (Waters-Micromass). Protein solution was injected into the ion source at a flow rate of 10 μL/min. Data were elaborated using the MassLynx program (Waters-Micromass). The cross-linked protein fraction was evaporated to dryness in a Speed Vac concentrator (Savant), resuspended in 10 mM ammonium bicarbonate at pH 8.0, digested with trypsin at 37 °C (E/S, 1:50 (w/w)) overnight, and subjected to MALDI-MS and LC-MS/MS analysis.

In the cross-linking experiment, the IF_1−CaM complex on a preparative scale, 15 nmol of both IF_1 and calmodulin was incubated in 400 μL of 50 mM MES at pH 5, 1 mM $CaCl_2$, and 0,15 M NaCl at 37 °C for 15 min and then treated with a 10-fold molar excess of EDAC for 30 min. The reaction was stopped by acidification with 1% TFA. The cross-linked protein was purified by reverse-phase HPLC on a Phenomenex Jupiter C4 column (250 × 4.6 mm, 300 Å pore size) with a linear gradient of 5−65% acetonitrile in 0.1% TFA over 40 min, at a flow rate of 1 mL/min. Elution was monitored at 220 nm, and individual fractions were collected and analyzed by electrospray mass spectrometry on a Quattro-Micro triple quadrupole mass spectrometer (Waters-Micromass). Protein solution was injected into the ion source at a flow rate of 10 μL/min. Data were elaborated using the MassLynx program (Waters-Micromass). The cross-linked protein fraction was digested with CNBr in 70% TFA overnight at room temperature under an inert atmosphere in the dark using a large weight excess of the reagent over protein. The sample was then diluted 10-fold with water and evaporated to dryness in a Speed Vac concentrator (Savant). The sample was then fractionated by reverse-phase HPLC on a Phenomenex Jupiter C4 column (250 × 2.00 mm, 300 Å pore size), and individual fractions were collected and analyzed by electrospray mass spectrometry. Selected fractions were digested with trypsin in 50 mM ammonium bicarbonate at pH 8 at 37 °C (E/S, 1:50 (w/w)) overnight and subjected to MALDI-MS and LC-MS/MS analysis.

MALDI mass spectra were recorded on an Applied Biosystems Voyager DE-PRO mass spectrometer equipped with a reflectron analyzer and used in delayed extraction mode. One microliter of the peptide sample was mixed with an equal volume of α-cyano-4-hydroxycynnamic acid as matrix (in acetonitrile/50 mM citric acid (70:30, v/v)), applied to the metallic sample plate, and air-dried. Mass calibration was performed by using the standard mixture provided by the manufacturer.

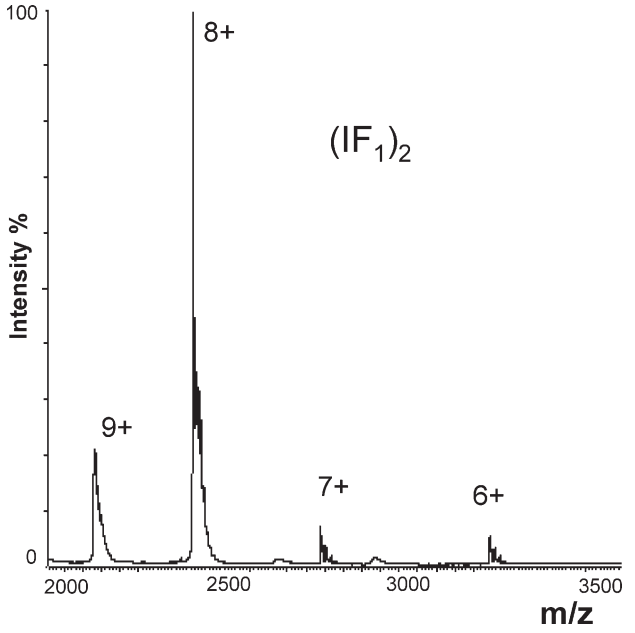


FIGURE 1: Native ESI-MS mass spectrum of the IF_1 dimer, showing the predominant dimeric ions.

LC-MS/MS analyses were performed on a Q-TOF hybrid mass spectrometer equipped with a Z-spray source and coupled online with a capillary chromatography system CapLC (Waters, Manchester, UK). After loading, the peptide mixture (4 μL) was first concentrated and washed at 10 μL/min onto a reverse-phase precolumn using 0.2% formic acid as eluent. The sample was then fractionated onto a C18 reverse-phase capillary column (75 μm × 20 mm) with 10−50% acetonitrile in 0.2% formic acid over 40 min, at a flow rate of 280 nL/min. The mass spectrometer was setup in a data-dependent MS/MS mode where a full scan spectrum (m/z acquisition range from 400 to 1400 Da/e) was followed by a tandem mass spectrum (m/z acquisition range from 100 to 2000 Da/e). Peptide ions were selected as the three most intense peaks of the previous scan. Suitable collision energy was applied depending on the mass and charge of the precursor ion.

RESULTS

Structural Characterization of IF_1. Oligomeric State of IF_1. The oligomeric state of IF_1 was assessed by direct ESI-MS measurement under native conditions. An aliquot of IF_1 (20 μM) was incubated in 10 mM ammonium acetate at pH 5.0 at 37° for 30 min and then directly injected into the electrospray source of a hybrid Q-TOF type mass spectrometer. Figure 1 shows the multiply charged ion spectrum of native IF_1, showing the presence of mass signal distribution encompassing $z = +9$ to $z = +6$ ions. The measured molecular mass calculated from the ESI-MS spectrum was 19166.3 ± 0.3 Da, in excellent agreement with the expected value for a dimeric IF_1 species ($(IF_1)_2$ = 19163 Da).

Limited Proteolysis. Limited proteolysis experiments were designed to study the surface topology of $(IF_1)_2$ on the grounds that the portions of the protein that are involved in highly

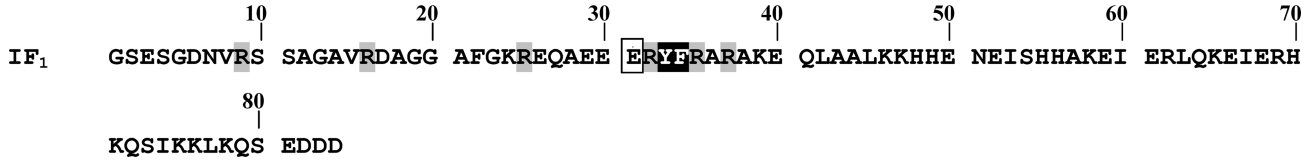


FIGURE 2: Preferential cleavage sites observed in the dimeric form of IF_1. The diagram shows the results obtained with trypsin (highlighting in gray), chymotrypsin (highlighting in black), and Glu-C (boxed residue). In all cases, the proteolytic cleavage took place between the indicated residue and the following one in the sequence.

Table 1: Peptides Generated in the Digestions of the Dimeric Form of IF_1 with Proteases and Identified by MALDI-TOF Analysis

protease (enzyme/IF1 ratio, w/w)	identified peptide	experimental $[MH^+]$ mass value (Da)	theoretical $[MH^+]$ mass value (Da)
Glu-C (1/400)	1−31	3125.69	3125.19
	32−84	6476.55	6476.28
trypsin (1/500)	1−9	919.70	920.90
	1−16	1548.81	1549.59
	1−25	2409.17	2409.52
	1−32	3281.35	3281.38
	1−35	3748.26	3747.92
	1−37	3975.7	3975.19
	38−84	5626.54	5626.29
	36−84	5853.75	5853.56
	33−84	6319.74	6320.10
	26−84	7191.11	7191.95
	17−84	8051.82	8051.89
	10−84	8680.66	8680.58
chymotrypsin (1/1000)	1−33	3444.10	3444.56
	1−34	3591.45	3591.73
	35−84	6009.01	6009.75
	34−84	6156.44	6156.92

structured regions or are located at the intersubunits interface are protected from proteases (*29*).

$(IF_1)_2$ (20 μM) was individually incubated with trypsin, chymotrypsin, or endoprotease Glu-C, as conformational probes. The extent of the enzymatic hydrolysis was monitored on a time-course basis by sampling 1 μL aliquots of the incubation mixture at different interval times followed by MALDI-MS analysis. Fragments released from the dimer were identified on the basis of their unique mass values and the specificity of the proteolytic enzyme. Identification of the two complementary peptides following a single proteolytic event led to the assignment of the preferential cleavage sites (*29*). In Supporting Information, an example of the MALDI-TOF analysis of the time course experiments with Glu-C endoprotease is reported.

The overall results of the limited proteolysis experiments are summarized in Table 1 and Figure 2, from which a number of considerations can be drawn. Preferential cleavage sites in $(IF_1)_2$ gathered into specific regions of the protein, the most exposed segment being the N-terminal portion spanning Glu31−Arg37. Interestingly, no cleavage sites were detected in the C-terminal region. These data are in perfect agreement with NMR data on the 44−84 C-terminal coiled-coil IF_1 domain (*30*), and with the crystallographic structure of the IF_1 mutant H49K (*1*) that could not be resolved in the N-terminal portion due to the high flexibility of this region. The high resolution data depict a central, highly structured, and, therefore, likely inaccessible, antiparallel coiled-coil structure formed by the opposite pairing of the two C-terminal regions with the two, very flexible and exposed N-terminal arms protruding at each end. This structural organization perfectly justifies our results.

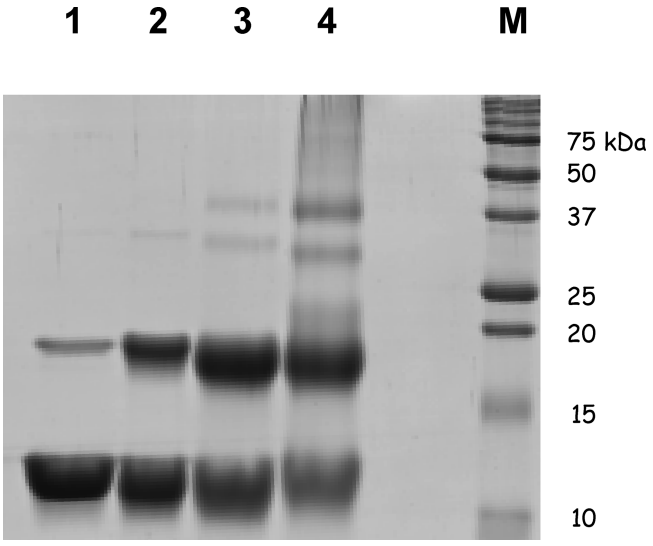


FIGURE 3: $(IF_1)_2$ cross-linking with EDAC. SDS−PAGE of reaction mixtures with different $(IF_1)_2$/EDAC ratios: (1) $(IF_1)_2$ without EDAC; (2) $(IF_1)_2$/EDAC 1/10; (3) $(IF_1)_2$/EDAC 1/100; (4) $(IF_1)_2$/EDAC 1/500; (M) protein molecular weight markers; molecular weights are reported in kDa on the side. All of the reactions were carried out at 37 °C for 30 min in 50 mM Na-MES at pH 5.0 and stopped by the addition of the SDS−PAGE sample buffer.

Chemical Cross-Linking of $(IF_1)_2$. The interacting contact regions in the IF_1 dimer complex were further investigated by chemical cross-linking. Cross-links were introduced with the heterobifunctional reagent EDAC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) that forms zero-length isopeptidic bonds between the carboxylic group of Glu residues and the ε-amino group of lysines or the free N-terminus of the protein.

$(IF_1)_2$ (20 μM) was incubated in 50 mM MES buffer at pH 5.0 and treated with EDAC at 37 °C. Several concentrations of EDAC and different reaction times were tested and the final experimental conditions were accurately selected to result solely in the formation of dimeric species, avoiding the formation of high molecular mass oligomers and distortion of the tertiary structure by excessive cross-linking. The cross-linking reaction mixtures were separated by monodimensional SDS−PAGE and the resulting gel stained with colloidal Coomassie Blue. Figure 3 shows the electrophoretic analysis of $(IF_1)_2$ treated with 10, 100, 500 molar excess of EDAC for 30 min at 37 °C. A protein band with an apparent molecular mass of 19 kDa was clearly detected in lanes 2, 3 and 4, which could tentatively be assigned to a covalent dimer of IF_1. IF_1 was also analyzed on the gel without treatment with EDAC as the control (Figure 3, lane 1). A thin band corresponding to a dimeric species could also be observed in these conditions, demonstrating the peculiar stability of the noncovalent $(IF_1)_2$ dimer, which can resist even to the highly denaturant conditions of the SDS−PAGE analysis. A 10-fold molar excess of EDAC was considered enough to introduce cross-links in a sufficient amount to freeze the dimeric IF_1 species

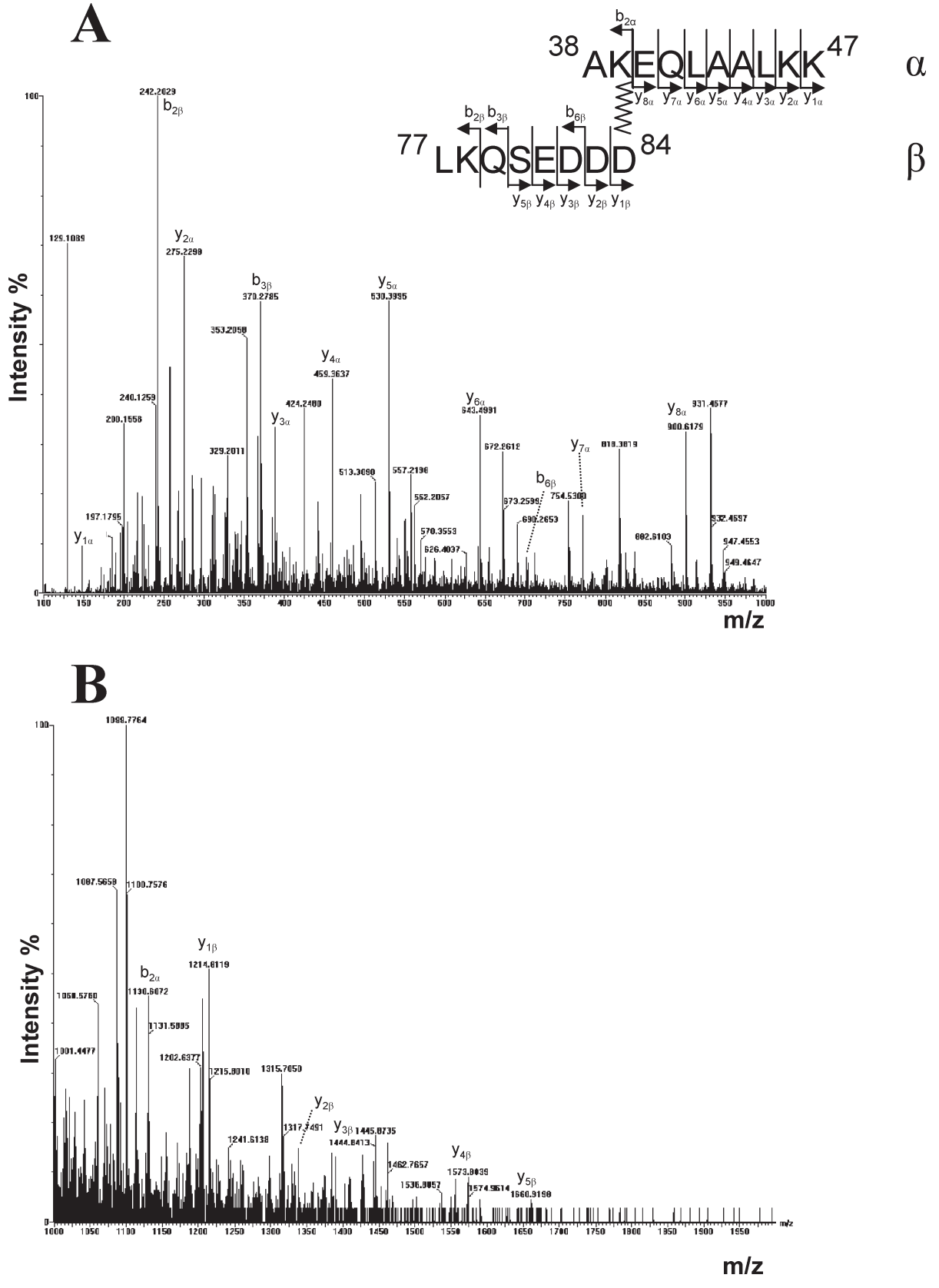


FIGURE 4: MS/MS spectra of the cross-linked peptides AKEQLAALKK and LKQSEDDD. Tandem mass spectra of the precursor ion $[M + 3H]^{3+} = 677.4$. Fragments from the peptide AKEQLAALKK are labeled with an α subscript and those from peptide LKQSEDDD with a *β* subscript. Most ion assignments are indicated on the peptide sequence in the figure. Panels A and B represent enlargements of different mass ranges of the same MS/MS spectrum.

for subsequent analysis, while higher excesses of EDAC generated higher molecular weight bands, thus suggesting aggregation of the protein caused by excessive cross-linking.

The dimeric protein band was excised from lane 2 and subjected to enzymatic in-gel digestion with trypsin. For comparative purposes, the monomeric IF_1 band from the reference sample

(lane 1) was also excised and treated in the same manner. The tryptic digests were then analyzed by MALDI-MS, and the experimentally obtained mass values were mapped onto the anticipated IF_1 sequence, confirming the identity of the protein bands and the absence of any contaminants.

The MALDI-MS fingerprint of dimeric IF_1 showed the occurrence of three signals that could be assigned to putative cross-linked products. The signal at m/z 2030.1 was assigned to the cross-linked peptides 38−47 and 77−84; the signal at m/z 2257.2 was interpreted as arising from either the peptide 36−47 linked to the peptide 77−84 or the peptide 36−46 joined to the peptide 76−84; the signal at m/z 3349.7 was assigned to the peptides 38−58 and 77−84. Although the IF_1 regions involved in the cross-link were clearly outlined by these data, discrimination on which amino acids were actually covalently linked could not be obtained by MALDI-MS data. A second aliquot of cross-linked IF_1 was then purified by HPLC and analyzed by ESI-MS providing the expected molecular mass of 19145.7 ± 0.7 Da. The sample was digested with trypsin and the resulting peptide mixture analyzed by nano LC-MS/MS. The multiply charged ions corresponding to the putative cross-linked species were isolated and fragmented by collision induced dissociation (CID). An example of the quality of the data is provided in Figure 4 where the fragmentation spectra of the triply charged ion at m/z 677.4 (corresponding to the MH^+ 2030.1) are shown.

The daughter ion spectra exhibited almost the entire y ion series for the two peptides 38−47 and 77−84. The $y_{1\beta}$ and $y_{8\alpha}$ ions constituted the key signals to assess the covalently linked residues. The former corresponds to the entire 38−47 peptide linked to Asp84, whereas the latter represents the residual 40−47 fragment following the release of Asp84 and the dipeptide Ala38-Lys39.

These data unequivocally indicated that the cross-linked species consisted of peptides 38−47 and 77−84 covalently linked by an isopeptide bond joining Asp84 to Lys39. Tandem mass spectra of ions corresponding to the other signals identified in the MALDI-MS analysis of the tryptic digests confirmed Asp84 and Lys 39 as the aminoacids involved in the cross-linked product. This isopeptide bond does not fit with the data of the IF_1 crystal structure. However, fragmentation spectra reported in Figure 4 can be unequivocally interpreted as such. The apparent discrepancy could be explained taking into account the conditions of our experiments, performed in solution at pH 5 and at low IF_1 concentration, rather different from the crystallization ones. Moreover, as stated by Walker's group (*3*), the region in wild-type IF_1 predicted to be in the coiled-coil extends from residues 37 to 84, consistent with our experimental observation, whereas for the IF1-H49K mutant, for which the three-dimensional structure has been solved at high resolution (*1*), the coiled-coil region is predicted to be shorter (involving only residues 47−84), in good agreement with the structural data. It should be noted that we worked with the wild-type IF_1 only, thus possibly explaining the discrepancy between the cross-linked residues observed and the high resolution data.

Structural Characterization of the IF_1−CaM Complex. IF_1−CaM Pull Down. The existence of a IF_1−CaM complex, as suggested by the Pedersen and Carafoli groups (*4*, *5*) and characterized by spectroscopic analysis (*2*) was further preliminarily assessed by a pull down assay exposing immobilized biotin labeled IF_1 to CaM.

Biotinylation of IF_1 was carried out with Sulfo-NHS-Biotin (sulfosuccinimidobiotin) and Biotin-IF_1 was immobilized on avidin, as described in the Materials and Methods. An aliquot

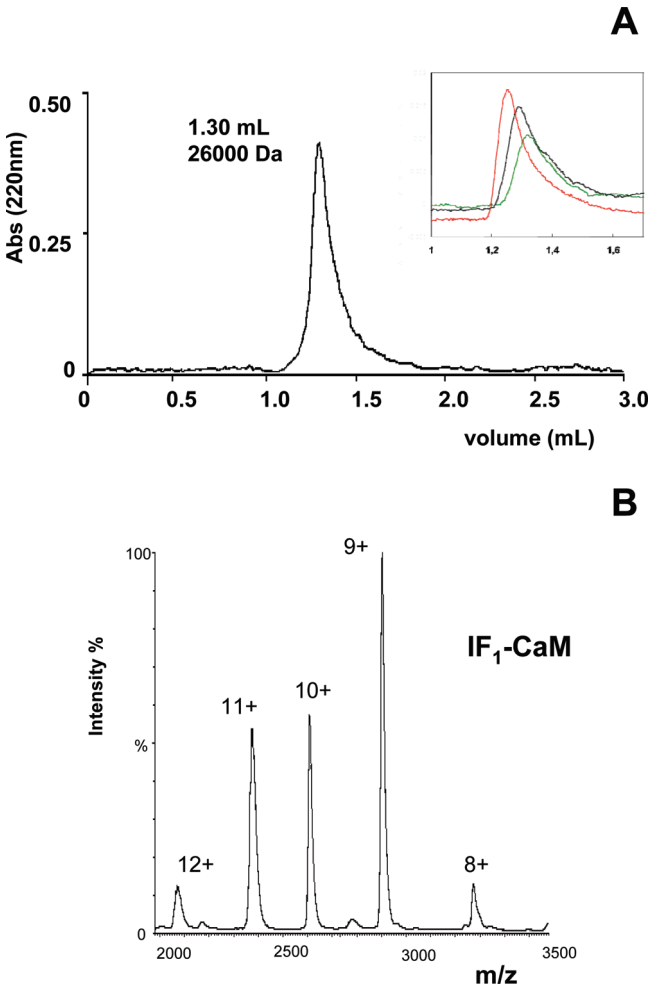


FIGURE 5: IF_1−CaM complex is a heterodimer. (A) Size exclusion chromatography on a Superdex-75 column of an equimolar mixture of IF_1 and CaM. In the inset, a zoom of the elution profile of IF_1−CaM (black line) is overlaid on those of isolated $(IF_1)_2$ (red line) and CaM (green line) obtained in the same chromatographic conditions (50 mM ammonium acetate at pH 5.0, 0.15 M NaCl, and 100 μM $CaCl_2$). (B) Native ESI-MS mass spectrum of the IF_1−CaM complex, showing the predominant heterodimeric ions.

of Biotin-IF_1 was purified by RF-HPLC an analized by ES-MS in order to assess the biotinilation yield showing that in the experimental conditions used IF_1 was efficiently biotinylated (up to 2 biotins per IF_1 monomer).

The biotinylated IF_1 was immobilized on avidin beads that were subsequently incubated with CaM in 50 mM MES buffer at pH 5.0. After extensive washing, the CaM that had been recruited by immobilized IF_1 was eluted in denaturing buffer, separated on SDS−PAGE and stained with colloidal blue Coomassie (Figure S2, Supporting Information). In a control experiment, an aliquot of CaM was loaded on avidin beads in the absence of immobilized IF_1, and treated as above: no CaM was unspecifically retained by avidin beads, thus unequivocally demonstrating the existence of a specific interaction between IF_1 and calmodulin.

Stoichiometry of the IF_1−CaM complex. The stoichiometry of the IF_1−calmodulin (CaM) complex was investigated by native electrospray mass spectrometry analyses. Preliminary size exclusion chromatography analysis was used to inspect the hydrodynamic behavior of the complex. Approximately equimolar amounts of IF_1 and CaM (20 μM of each protein) were incubated at 37° for 30 min at pH 5.0 and then loaded onto a Superdex 75 size exclusion column. The chromatography resulted

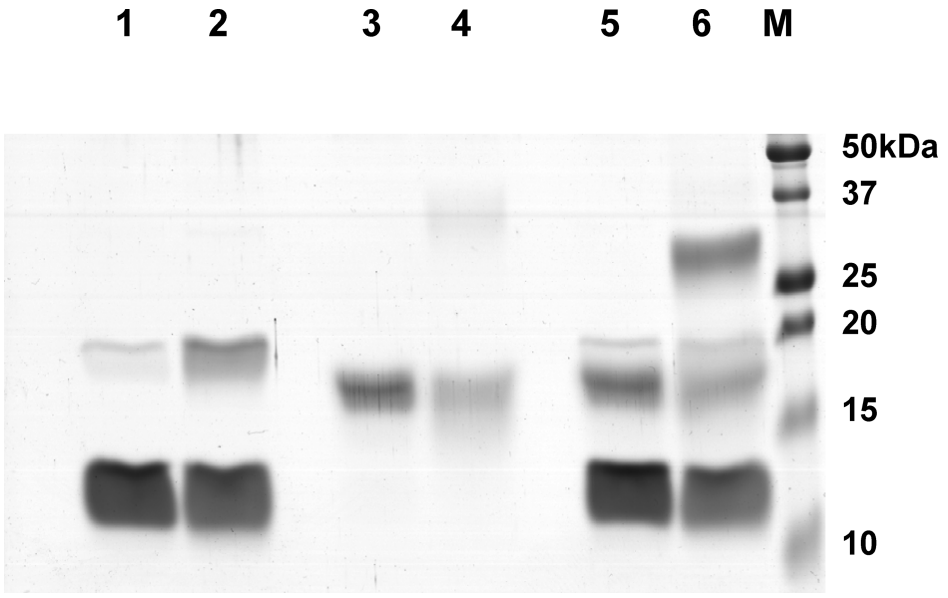


FIGURE 6: IF_1−CaM cross-linking with EDAC. SDS−PAGE of reaction mixtures of IF_1 (20 μM) and CaM (20 μM) with EDAC (200 μM). The samples are labeled as follows: (M) protein molecular weight markers, (1) CaM, (2) CaM with EDAC, (3) IF_1, (4) IF_1 with EDAC, (5) IF_1 and CaM mixture, and (6) IF_1 and CaM mixture with EDAC. All of the reactions were carried out at 37 °C for 30 min in 50 mM Na-MES at pH 5.0 and stopped by the addition of the SDS−PAGE sample buffer.

in a single peak, which eluted at an apparent molecular mass of about 26000 Da (Figure 5, panel A; elution volume, 1.30 mL; see the inset in the figure), suggesting a 1:1 stoichiometry of the complex. Peak top fractions were analyzed by MALDI-TOF, revealing the presence of both proteins, i.e., IF_1 and CaM (data not shown). Most importantly, no peak eluting at a volume corresponding to a higher apparent molecular weight was observed, thus suggesting that, in these experimental conditions, the complex is present in a single predominant homogeneous oligomeric state. However, a reliable definition of the stoichiometry of the complex from size exclusion analysis is impaired by the elongated shape of both components, which significantly alters the hydrodynamic behavior of the two isolated proteins (in the same experimental conditions, $(IF_1)_2$ elutes with an apparent molecular weight of 30000 Da (Figure 5, panel A; elution volume, 1.25 mL; see the inset in the figure) and CaM with an apparent molecular weight of 23000 Da (Figure 5, panel A; elution volume, 1.32 mL; see the inset in the figure), in fairly good agreement with the apparent molecular weight reported by other authors (*31*, *32*)).

Therefore, this issue was further addressed by native mass spectrometry, which is not affected by the elongated conformation of the singular components. An equimolar amount of the two proteins (20 μM of each protein) was mixed at 37° for 30 min and then directly injected into the electrospray source of a hybrid Q-TOF type mass spectrometer under native conditions. Panel B in Figure 5 shows the enlargement of the full spectrum exhibiting the presence of a clear distribution of mass signals encompassing $z = +12$ to $z = +8$ ions that was assigned to the IF_1−CaM complex containing 4 calcium ions. The measured molecular mass of this species was 26527.3 ± 0.7 Da, in excellent agreement with the expected value for one IF_1 molecule bound to one CaM molecule bearing four calcium ions (expected molecular mass = 26528.0 Da). This result clearly indicated a 1:1 stoichiometry for the IF_1−CaM complex.

Chemical Cross-Linking. The interacting regions in the IF_1−CaM complex were investigated by chemical cross-linking with the heterobifunctional reagent EDAC, using a procedure similar to the procedure described above for the IF_1 dimer.

IF_1 and CaM (20 μM of each protein) were incubated as reported in the Materials and Methods section, and then treated with a 10-fold molar excess of EDAC at 37 °C for a further 30 min. Isolated CaM and IF_1 were also individually incubated with EDAC under the same conditions as the control. Several concentrations of EDAC and different reaction times were tested, and the final experimental conditions were accurately selected to result in the formation of an IF_1−CaM heterodimeric complex, avoiding the formation of high molecular mass oligomers and preventing excessive cross-linking.

The different mixtures of products generated by the cross-linking reaction were analyzed by SDS−PAGE, as shown in Figure 6 for the reaction carried out with a 10-fold molar excess of EDAC. A protein band with an apparent molecular mass of 26 kDa was clearly detected in lane 6. This band was absent both when equimolar amounts of IF_1 and CaM were mixed without the cross-linking reagent (lane 5) and when CaM and IF_1 were individually treated with EDAC (lanes 2 and 4, respectively). Aggregation of proteins caused by excessive cross-linking was not observed as demonstrated by the absence of any band in the higher mass range.

The identity of the 26 kDa gel band was definitively assessed by identification of the protein components by mass spectral analysis. The protein band was excised from the gel and subjected to enzymatic in-gel digestion with trypsin. The resulting peptide mixture was directly analyzed by MALDI-MS and the corresponding mass values were mapped onto the anticipated sequences of IF_1 and CaM revealing the contemporaneous presence of both proteins. These results strongly supported the formation of a covalent IF_1−CaM complex following EDAC treatment and confirmed previous mass spectral data on the stoichiometry of the complex.

A second aliquot of the cross-linked IF_1−CaM complex was fractionated by reverse phase HPLC and the collected fractions analyzed by ESI-MS. The covalent IF_1−CaM complex eluted in two fractions exhibiting molecular masses of 26353.5 ± 1.0 Da and 26336.1 ± 1.6 Da, the latter being a minor component. These mass values correspond to the molecular masses of the two

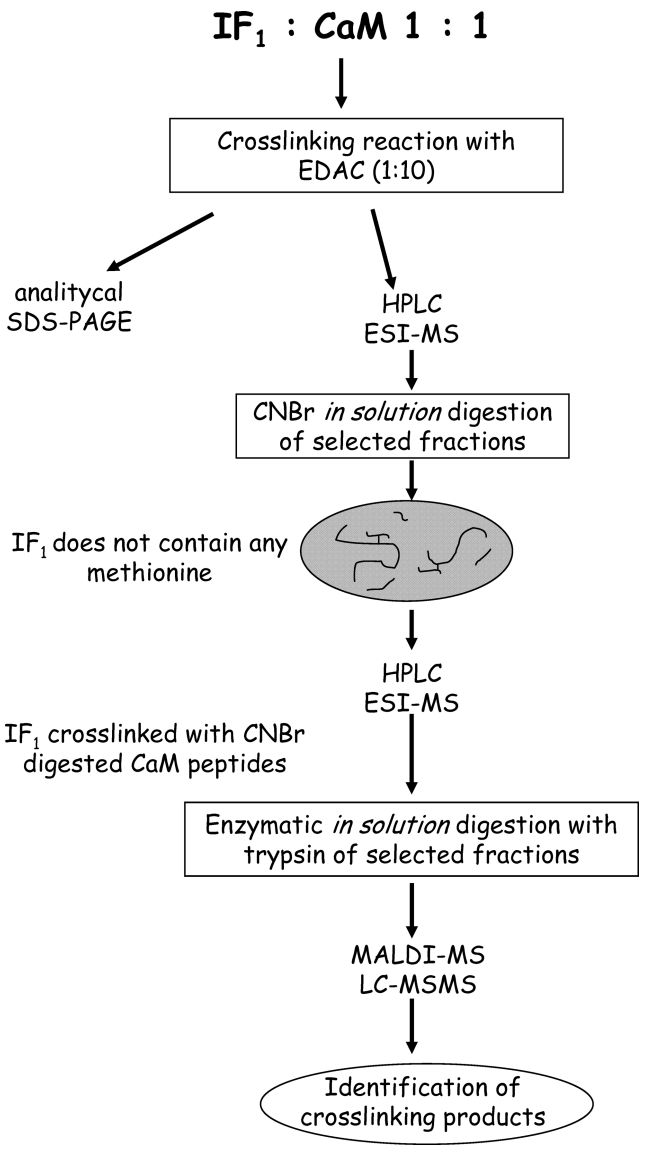


FIGURE 7: General analytical strategy for mapping the interaction between IF_1 and CaM by chemical cross-linking.

proteins minus 18.01 and 36.02 Da, indicating the presence of the 1:1 IF_1−CaM complex containing one and two cross-linking bonds.

Identification of the cross-linked residues in both IF_1 and CaM proteins was achieved following the strategy outlined in Figure 7 on the fraction containing the IF_1−CaM species where one covalent cross-link had been introduced. The rationale behind this strategy is based on the absence of methionine residues in the IF_1 sequence that is then unreactive with CNBr. Chemical hydrolysis of the complex would have therefore left the undigested IF_1 molecule covalently linked to CaM peptides, making these fragments immediately detectable by mass spectrometry.

The complex was then subjected to CNBr hydrolysis in 70% TFA at room temperature overnight, and the resulting peptide mixture was fractionated by HPLC. Individual fractions were manually collected and directly analyzed by ESI-MS. The fraction containing the species with a molecular mass 11280.2 ± 0.5 Da could be tentatively interpreted as IF_1 linked to peptide 110−124 of CaM (theoretical mass value 11280.3 Da). Two minor fractions containing the species with a molecular mass of 13560.3 ± 0.6 Da and 15063.0 ± 0.9 Da were also observed and interpreted as IF_1 linked to peptides 110−144 and 77−124, respectively (corresponding theoretical mass values of 13560.8 and 15061.5 Da, respectively).

The major component was further hydrolyzed with trypsin at 37 °C overnight, and the resulting peptide mixture directly analyzed by both MALDI-MS and nano LC-MS/MS. MALDI-MS spectra displayed the presence of three signals, at m/z 2594.1, 2636.1, and 3465.4 that could not be assigned to any linear peptide within the aminoacid sequence of either IF_1 or CaM. These species were then considered as putative cross-linked fragments and tentatively interpreted as the IF_1 peptides 17−25, 1−9, and 17−32 linked to the CaM fragment 110−124 (theoretical mass values 2594.2 Da, 2636.1 Da, and 3465.57 Da, respectively).

Validation of these hypotheses and identification of the IF_1 and CaM amino acids covalently linked were achieved by tandem mass spectrometry experiments. The peptide mixture was analyzed by nano LC MS/MS, and the multiply charged ions corresponding to the three putative cross-linked species were isolated and fragmented by collision induced dissociation (CID). Figure 8 shows the fragmentation spectra of the triply charged ion at m/z 879.3 (corresponding to the MH^+ 2636.1). The complete y-ion series attributed to the IF_1 peptide 1−9 and part of the b-ion series spanning the sequence 116−122 of the CaM fragment 110−124 could be detected, thus confirming the interpretation of the MALDI data. The specific fragmentation of the two peptides was certainly dictated by the presence of a trimethyl-lysine residue at position 115 of the CaM peptide and an Arg residue located at the C-terminus of the IF_1 fragment, both endowed with a high charge retention propensity.

The individual residues of the two proteins effectively involved in the covalent linkage could also be assessed by the MS/MS data. The two key signals in the spectrum were identified at m/z 1486.5 and 1265.5. The former belongs to the b-ion series and corresponds to the loss of the CaM Glu123-Met124 dipeptide together with the entire IF_1 1−9 fragment from the cross-linked species. On the contrary, the y-ion at 1265.5 accounts for the whole IF_1 fragment covalently bound to the CaM tripeptide Asp122-Glu123-Met124, with the positive charge retained on the Arg9 residue.

These data unequivocally indicated that the cross-linked species consisted of the 1−9 N-terminal peptide of IF_1 and the 110−124 fragment covalently linked by an isopeptide bond joining Glu123 in the CaM molecule to the IF_1 N-terminal amino group.

Similar analyses were carried out on the other two cross-linked species, demonstrating their identity as the IF_1 peptides 17−25 and 17−32 linked, respectively, to the CaM 110−124 fragment. Moreover, the cross-linked bond could be shown to involve the CaM Glu123 residue and the ε-amino group of IF_1 Lys 24.

Altogether, the cross-linking experiments pointed to an IF_1−CaM complex in which the N-terminal portion of IF_1 is in close contact with the calmodulin C-terminal lobe.

DISCUSSION

The IF_1−CaM complex was investigated in solution by a combined strategy which integrates mass spectrometric methodologies and cross-linking experiments. Crucial issues of the structural characterization were the stoichiometry of the IF_1−CaM complex and the understanding of the contact region between IF_1 and calmodulin.

Mass spectral analysis of the IF_1−CaM complex under native conditions highlighted the 1:1 stoichiometry, indicating that a single IF_1 moiety binds a single CaM molecule. Most interestingly,

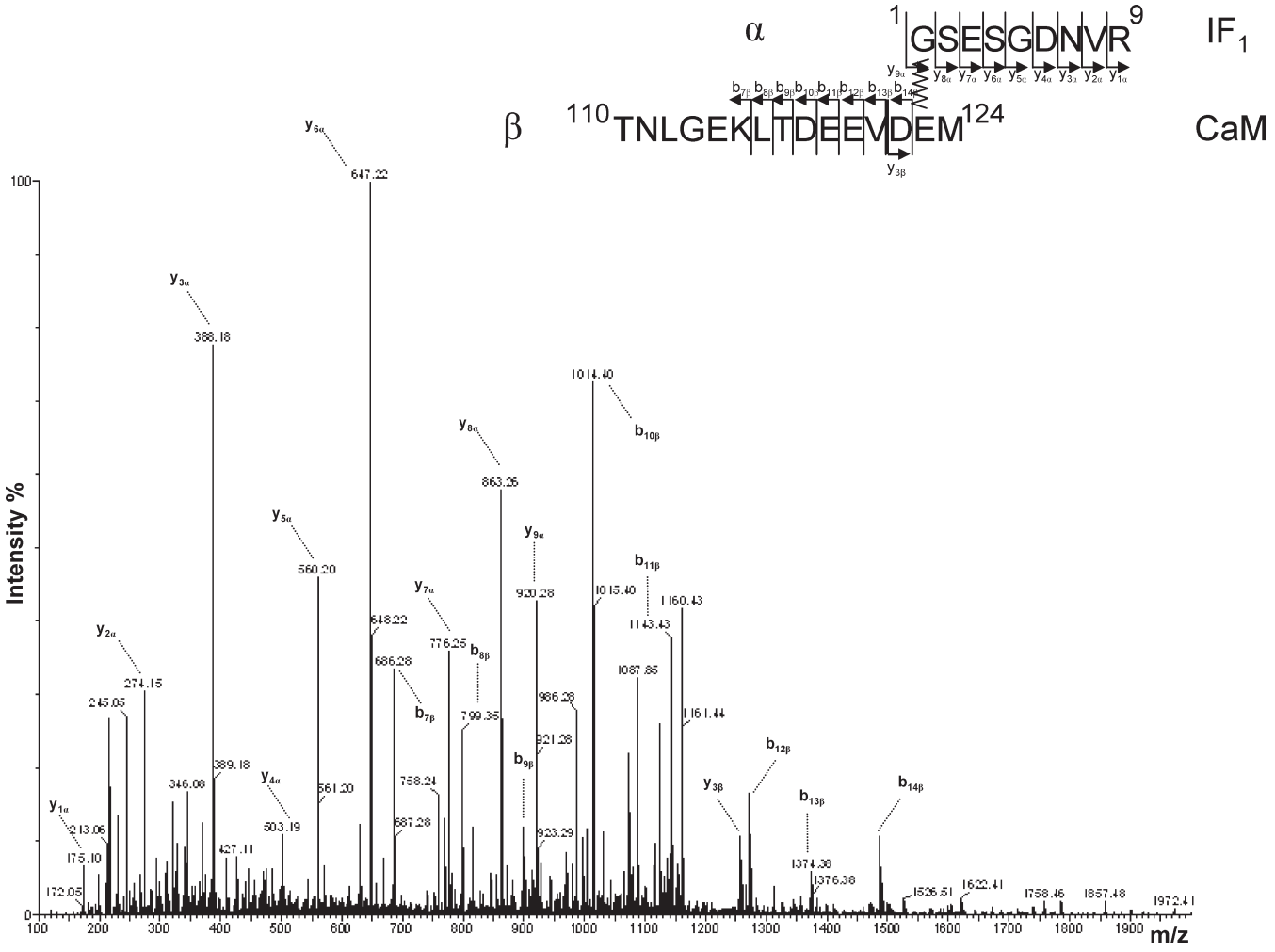


FIGURE 8: MS/MS spectra of the cross-linked peptides GSESGDNVR of IF_1 and TNLGEKLTDEEMDEM of CaM. Tandem mass spectra of the precursor ion $[M + 3H]^{3+} = 879.31$. Fragments from the peptide GSESGDNVR are labeled with an α subscript and those from peptide TNLGEKLTDEEMDEM with a β subscript. Most ion assignments are indicated on the peptide sequence in the figure.

the accurate mass value revealed that 4 Ca^{2+} ions remain bound to CaM, in accordance with the maintenance of a high Ca^{2+} affinity in the complex (*4*). The ESI-MS did not reveal the presence of higher order oligomers, as preliminarily suggested by gel filtration chromatography. The 1:1 stoichiometry implies, therefore, that CaM binds to monomeric IF_1, although IF_1 alone in solution at acidic pH exists and is functional as a dimer, as already suggested by several experimental data (*3*) and confirmed by mass spectral analysis under native conditions in the present study.

The present results, together with our previous data documenting the maintenance of a 1:1 molar ratio upon CaM fluorimetric titration with IF_1 up to 10 molar excess (*2*), are consistent with a 1:1 IF_1−CaM complex stoichiometry in the micromolar concentration range, which correlates with the physiological concentration of CaM present in living cells (8.8 ± 2.2 μM) (*33*).

More details on the IF_1−CaM interaction were needed to ascertain whether CaM might interfere with the intersubunit contacts of the IF_1 dimer or bind the inhibitory active N-terminal region of the molecule. Cross-linking experiments identified residues at the contact interface of the IF_1−CaM complex. Glu123 in the CaM molecule was shown to cross-link either the N-terminal amino group of IF_1 or the ε-amino group of IF_1 Lys24.

A merely qualitative examination of the cross-linking data suggested that a unique binding mode of IF_1 in the IF_1−CaM complex is incompatible with the experimental results since the same CaM residue (Glu123) was found cross-linked to two different IF_1 residues, suggesting the existence of at least two types of IF_1−CaM complexes, although the possibility of the spatial proximity of the IF_1 N-terminus and Lys 24 that can be alternatively cross-linked to Glu123 cannot be ruled out. In order to correctly interpret this result, it is also worth considering the ability of CaM to recognize its targets in multiple modes, as highlighted by similar experiments carried out on the well known CaM−melittin complex (*17*) where multiple modes of binding exist. Moreover, and mostly relevant in this case, the well documented (*1*) flexibility of the N-terminal region of IF_1 is an additional element that has to be taken into account to explain the multiple binding modes observed in the IF_1−CaM complex. This N-terminal segment, extending from the central coiled-coil of the molecule, appears to be essentially unstructured, as demonstrated by the high accessibility to proteases exhibited by the IF_1 dimer up to Arg 37, with the most exposed segment being the N-terminal portion spanning Glu31−Arg37. However, despite the multiple binding mode observed, it is clear that it is the N-terminal inhibitory region of IF_1 that is involved in binding to CaM, in agreement with the hypotheses that arose on the basis of sequence analysis and with the previous data obtained with synthetic peptides (*2*). The IF_1 sequence 33−42 was previously assessed as a putative target region for CaM on the basis of fluorimetric binding studies using dansyl-CaM and native bovine

IF_1 by comparing two synthetic peptides and two natural mutants, namely, yeast inhibitor proteins IF_1 and STF1 (*2*). We can consider the contact interface at the 33−42 region to be congruent with one of the multiple types of binding characterized by the cross-linking of CaM Glu123 to IF_1 Lys24.

Together, the findings that it is the N-terminal inhibitory region of IF_1 that is involved in the contact interface with CaM and that the complex stoichiometry is 1:1 suggest that binding to CaM promotes the dissociation of the active IF_1 dimer without directly interfering with the intersubunit contacts of the IF_1 dimer.

The overall cross-linking data produced in this article clearly indicate that only the C-terminal lobe of CaM is involved in binding IF_1. This definitively suggests that the IF_1−CaM complex deviates from the canonical or at least more frequent representation of CaM complexes in which the protein adopts a horseshoe-like structure with the N and C-terminal lobes, far apart in the dumbbell native protein (*34*), coming in close contact when the target peptide is bound, and originating an overall globular shaped complex.

Novel structures of CaM complexed with large target fragments, namely, the calcium-activated K^+ channel (*23*), and anthrax adenylate cyclase exotoxin (*24*) have been solved revealing unexpected binding modes, and our data clearly put the IF_1−CaM complex among the noncanonical CaM interaction mode. Moreover, a noncanonical binding mode might also explain the unchanged CaM affinity for Ca^{2+} previously observed in the complex (*2*). In the more common binding mode, the dissociation constants of Ca^{2+} from the EF hands of CaM in both C-terminal and the N-terminal lobes decrease significantly as a consequence of ternary complex formation (*35*−*37*).

Therefore, the IF_1−CaM complex can be included in the category of novel structures of CaM−target complexes. Interestingly, Vetter and Leclerc reported that in the case of a Ca^{2+} pump, the complex only involved the C-terminal lobe of CaM, similar to what occurs in the IF_1−CaM complex (*20*). In particular, NMR analysis of CaM complexed with the C20W peptide of the Ca^{2+} pump, showed that the protein remained in an extended conformation with a flexible central linker region between Arg74 and Glu84, thus explaining the changes in fluorescence emission spectra of Lys75-dansylated CaM upon binding to the C20W peptide (*20*). Similarly, the binding of Lys75-dansylated CaM to IF_1 showed changes in the fluorescence emission suggesting that the linker region of CaM responded with conformational changes to IF_1 binding (*2*).

The involvement of only the C-terminal lobe of CaM suggests that the N-terminal region of the protein might bind to a second target more likely on the plasma membrane where IF_1 and CaM colocalize and where CaM is anchored to lipid rafts (*25*). Moreover, we can safely state that CaM and F1 can compete for the binding of IF_1 within the physiologic pH range. In fact, *in vitro* at pH 6.7 CaM prevents IF_1 from inhibiting both F1 (*2*, *4*) and F_OF_1ATPase (*2*) and that both IF_1−F1 and IF_1−CaM complexes have K_d in the nanomolar range (*2*, *38*). Noteworthy, selective increase in IF_1 expression on the cell surface that may evoke some specific regulation has been observed to occur and modulate cell surface−F_OF_1 activity in response to different stimuli, i.e., TNFα in HUVEC cells and acute cholestasis in rat liver (*27*, *39*). Thus, we suggest that CaM on the cell surface may sequester free IF_1 even in the presence of F_OF_1 and may modulate the amount of the inhibitory IF_1 in response to a proper stimulus (*25*).

SUPPORTING INFORMATION AVAILABLE

Example of the limited proteolysis experiments, showing the MALDI-MS analysis of the aliquots withdrawn following 0, 5, and 20 min of digestion with Glu-C endoprotease; and the SDS−PAGE analysis of the pull-down experiment carried out with biotinylated IF_1. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Cabezon, E., Runswick, M. J., Leslie, A. G. W., and Walker, J. E. (2001) The structure of bovine IF(1), the regulatory subunit of mitochondrial F-ATPase. *EMBO J. 20*, 6990–6996.
2. Contessi, S., Haraux, F., Mavelli, I., and Lippe, G. (2005) Identification of a conserved Calmodulin-binding motif in the sequence of F_OF_1ATPsynthase inibhitor protein. *J. Bioenerg. Biomembr. 37*, 317–326.
3. Cabezon, E., Butler, P. J., Runswick, M. J., and Walker, J. E. (2000) Modulation of the oligomerization state of the bovine F1-ATPase inhibitor protein, IF1, by pH. *J. Biol. Chem. 275*, 25460–25464.
4. Pedersen, P. L., and Hullihen, J. (1984) Inhibitor peptide of mitochondrial proton adenosine triphosphatase. Neutralization of its inhibitory action by calmodulin. *J. Biol. Chem. 259*, 15148–15153.
5. Schwerzmann, K., Muller, M., and Carafoli, E. (1985) The inhibitor peptide of the mitochondrial F1.F0-ATPase interacts with calmodulin and stimulates the calmodulin-dependent Ca2+-ATPase of erythrocytes. *Biochim. Biophys. Acta 816*, 63–67.
6. Strynadka, N. C., and James, M. N. (1989) Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biochem. 58*, 951–998.
7. Persechini, A., Moncrief, N. D., and Kretsinger, R. H. (1989) The EF-hand family of calcium-modulated proteins. *Trends Neurosci. 12*, 462–467.
8. Meador, W. E., Means, A. R., and Quiocho, F. A. (1993) Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science 262*, 1718–1721.
9. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry 31*, 5269–5278.
10. Yamniuk, A. P., and Vogel, H. J. (2004) Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Mol. Biotechnol. 27*, 33–57.
11. Ikura, M., and Ames, J. B. (2006) Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: two ways to promote multifunctionality. *Proc. Natl. Acad. Sci. U.S.A. 103*, 1159–1164.
12. Izumi, Y., Watanabe, H, Watanabe, N., Aoyama, A., Jinbo, Y., and Hayashi, N. (2008) Solution X-ray scattering reveals a novel structure of calmodulin complexed with a binding domain peptide from the HIV-1 matrix protein p17. *Biochemistry 47*, 7158–7166.
13. Yap, K. L., Kim, J., Truong, K., Sherman, M., Yuan, T., and Ikura, M. (2000) Calmodulin target database. *J. Struct. Funct. Genomics 1*, 8–14.
14. Popescu, S. C., Popescu, G. V., Bachan, S., Zhang, Z., Seay, M., Gerstein, M., Snyder, M., and Dinesh-Kumar, S. P. (2007) Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays. *Proc. Natl. Acad. Sci. U.S.A. 104*, 4730–4735.
15. Bhattacharya, S., Bunick, C. G., and Chazin, W. J. (2004) Target selectivity in EF-hand calcium binding proteins. *Biochim. Biophys. Acta 1742*, 69–79.
16. Hoeflich, K. P., and Ikura, M. (2002) Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell 108*, 739–742.
17. Schulz, D. M., Ihling, C., Clore, G. M., and Sinz, A. (2004) Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin-melittin complex by chemical cross-linking and high-resolution FTICRMS: direct demonstration of multiple binding modes. *Biochemistry 43*, 4703–4715.
18. Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science 256*, 632–638.
19. O'Neil, K. T., and DeGrado, W. F. (1990) How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices. *Trends Biochem. Sci. 15*, 59–64.
20. Vetter, S. W., and Leclerc, E. (2003) Novel aspects of calmodulin target recognition and activation. *Eur. J. Biochem. 270*, 404–414.

21. Osawa, M., Tokumitsu, H., Swindells, M. B., Kurihara, H., Orita, M., Shibanuma, T., Furuya, T., and Ikura, M. (1999) A novel target recognition revealed by calmodulin in complex with Ca2+-calmodulin-dependent kinase. *Nat. Struct. Biol. 6*, 819–824.
22. Elshorst, B., Hennig, M., Försterling, H., Diener, A., Maurer, M., Schulte, P., Schwalbe, H., Griesinger, C., Krebs, J., Schmid, H., Vorherr, T., and Carafoli, E. (1999) NMR solution structure of a complex of calmodulin with a binding peptide of the Ca^{2+} pump. *Biochemistry 38*, 12320–12332.
23. Schumacher, M. A., Rivard, A. F., Bachinger, H. P., and Adelman, J. P. (2001) Structure of the gating domain of a Ca^{2+}-activated K^+ channel complexed with Ca^{2+}/calmodulin. *Nature 410*, 1120–1124.
24. Drum, C. L., Yan, S. Z., Bard, J., Shen, Y. Q., Lu, D., Soelaiman, S., Grabarek, Z., Bohm, A., and Tang, W. J. (2002) Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature 415*, 396–402.
25. Contessi, S., Comelli, M., Cmet, S., Lippe, G., and Mavelli, I. (2007) IF(1) distribution in HepG2 cells in relation to ecto-F(0)F (1)ATP-synthase and calmodulin. *J. Bioenerg. Biomembr. 39*, 291–300.
26. Burwick, N. R., Wahl, M. L., Fang, J., Zhong, Z., Moser, T. L., Li, B., Capaldi, R. A., Kenan, D. J., and Pizzo, S. V. (2005) An Inhibitor of the F1 subunit of ATP synthase (IF1) modulates the activity of Angiostatin on the endothelial cell surface. *J. Biol. Chem. 280*, 1740–1745.
27. Giorgio, V., Bisetto, E., Franca, R., Harris, D. A., Passamonti, S., and Lippe, G. (2010) The ectopic F(O)F(1) ATP synthase of rat liver is modulated in acute cholestasis by the inhibitor protein IF1. *J. Bioenerg. Biomembr. 42*, 117–23.
28. van Raaij, M. J., Orriss, G. L., Montgomery, M. G., Runswick, M. J., Fearnley, I. M., Skehel, J. M., and Walker, J.E. (1996) The ATPase inhibitor protein from bovine heart mitochondria: the minimal inhibitory sequence. *Biochemistry 35*, 15618–15625.
29. Marino, G., Pucci, P., Birolo, L., and Ruoppolo, M. (2003) Exploitation of proteomic strategies in protein structure−function studies. *Pure Appl. Chem. 75*, 303–310.
30. Gordon-Smith, D. J., Carbajo, R. J., Yang, J. C., Videler, H., Runswick, M. J., Walker, J. E., and Neuhaus, D. (2001) Solution structure of a C-terminal coiled-coil domain from bovine IF_1: the inhibitor protein of F_1-ATPase. *J. Mol. Biol. 308*, 325–339.
31. Majava, V., Petoukhov, M. V., Hayashi, N., Pirilä, P., Svergun, D. I., and Kursula, P. (2008) Interaction between the C-terminal region of human myelin basic protein and calmodulin: analysis of complex formation and solution structure. *BMC Struct. Biol. 8*, 10.
32. Majava, V., and Kursula, P. (2009) Domain swapping and different oligomeric states for the complex between Calmodulin and the Calmodulin-binding domain of Calcineurin A. *PLoS One 4*, e5402.
33. Black, D. J., Tran, Q. K., and Persechini, A. (2004) Monitoring the total available calmodulin concentration in intact cells over the physiological range in free Ca2+. *Cell Calcium 35*, 415–425.
34. Rhoads, A. R., and Friedberg, F. (1997) Sequence motifs for calmodulin recognition. *FASEB J. 11*, 331–340.
35. Peersen, O. B., Madsen, T. S., and Falke, J. J. (1997) Intermolecular tuning of calmodulin by target peptides and proteins: differential effects on Ca^{2+} binding and implications for kinase activation. *Protein Sci. 6*, 794–807.
36. Brown, S. E., Martin, S. R., and Bayley, P. M. (1997) Kinetic control of the dissociation pathway of calmodulin-peptide complexes. *J. Biol. Chem. 272*, 3389–3397.
37. Mirzoeva, S., Weigand, S., Lukas, T. J., Shuvalova, L., Anderson, W. F., and Watterson, D. M. (1999) Analysis of the functional coupling between Calmodulin's calcium binding and peptide recognition properties. *Biochemistry 38*, 14117–14118.
38. Gomez-Fernandez, J. C., and Harris, D. A. (1978) A thermodynamic analysis of the interaction between the mitochondrial coupling adenosine triphosphatase and its naturally occurring inhibitor protein. *Biochem. J. 176*, 967–975.
39. Cortés-Hernández, P., Domínguez-Ramírez, L., Estrada-Bernal, A., Montes-Sánchez, D. G., Zentella-Dehesa, A., de Gómez-Puyou, M. T., Gómez-Puyou, A., and García, J. J. (2005) The inhibitor protein of the F1FO-ATP synthase is associated to the external surface of endothelial cells. *Biochem. Biophys. Res. Commun. 330*, 844–849.

technical notes Journal of proteome research

Identification of ΔNp63α Protein Interactions by Mass Spectrometry

Angela Amoresano,[†] Antonella Di Costanzo,[‡] Gabriella Leo,[†] Ferdinando Di Cunto,[§] Girolama La Mantia,[‡] Luisa Guerrini,[||] and Viola Calabrò*[,‡]

Dipartimento di Chimica Organica e Biologica, Università Federico II, Napoli, Italy, Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli, Federico II, Italy, Centro di Biotecnologie Molecolari, Università di Torino, Torino, Italy, and Dipartimento di Scienze Biomolecolari e Biotecnologie, Università di Milano, Italy

Received December 4, 2009

Abstract: p63, a transcription factor related to the p53 tumor suppressor, plays a key role in epidermal differentiation and limb development. The gene has two distinct promoters that allow the formation of proteins that either contain (TA) or lack (ΔN) a transactivation domain. ΔNp63α is the most widely expressed isoform, at all stages of development and in adult tissues. It supports the regenerative capacity of basal keratinocytes and its upregulation is a hallmark of human squamous carcinomas. To get insight into the complex biology of ΔNp63α, we set out to identify ΔNp63α interacting proteins by co-immunoprecipitation in mammalian cells and mass spectrometry analysis. A total of 49 potential ΔNp63α binding proteins, including several heterogeneous ribonucleoproteins (hnRNPs), were identified. Integration of the proteomic data with a Human Coexpression Network highlighted 5 putative p63 protein interactors whose expression is significantly comodulated with p63: hnRNPA/B, hnRNPK, hnRNPQ, FUS/TLS and Keratin 5. hnRNPA/B was already described as a p63 partner, but the others were novel. Interaction of ΔNp63α with hnRNPQ, hnRNPK and FUS/TLS was confirmed by reciprocal co-immunoprecipitations in human keratinocytes. The finding that ΔNp63α exists in complexes with several RNA-binding proteins lays the premises for the analysis of the role of ΔNp63α in mRNA metabolism and transport.

Keywords: Epithelial Differentiation • Mass Spectrometry • p53 Gene Family • mRNA Metabolism • Protein Interaction

Introduction

The tumor suppressor p53 controls a powerful stress response by integrating upstream signals from different types of DNA damage and oncogenic stimuli. Activated p53 elicits cell cycle arrest and apoptosis, thereby preventing the formation of tumors.[1] p53 is the founding member of a family of proteins including p63 and p73.[2] All three genes can regulate cell cycle and apoptosis after DNA damage. However, mouse knockout models revealed an unexpected functional diversity among them. Indeed, p63 and p73 null mice exhibit severe developmental abnormalities but no increased cancer susceptibility, whereas this picture is reversed for p53 knockouts.[3]

The structure of the corresponding genes, named TP53, TP63 and TP73, is evolutionarily conserved. In particular, the most conserved regions are those encoding the aminoterminal transactivation domain (TA), the central DNA binding domain (DBD), and the carboxyterminal oligomerization domain (OD).[2,3] The TP63 gene has two distinct promoters allowing the synthesis of proteins that either contain (TA) or lack (ΔN) a transactivation domain. In addition, alternative splicing at the 3′ end generates proteins with different C-termini, denoted α, β and γ. Only the α-isoforms (TA and ΔN) contain a Sterile Alpha Motif (SAM) domain,[4] which is absent in p53. Distinct p63 isoforms are expressed and differentially modulated during epithelial differentiation.[5] In particular, ΔNp63α is strongly expressed in basal keratinocytes of stratified epithelia and disappears in differentiating cells. Accordingly, human ΔNp63α is known to support the regenerative capacity of basal keratinocytes.[4] Heterozygous germ line mutations of p63 cause rare autosomal dominant developmental disorders associated with ectodermal dysplasia and limb deformities.[3] Accordingly, mice lacking p63 are severely compromised in their ability to generate limbs, and stratified epithelia such as skin.[2]

In stark contrast with the high mutation rate of p53 in a large compendium of cancer types, p63 is not mutated in tumors. Therefore, although p63 is able to activate p53 responsive genes and induce apoptosis, it is debated whether it might act as a tumor suppressor or an oncogene. Remarkably, ΔNp63α gene is upregulated or amplified in a broad range of squamous cell carcinomas, thus, suggesting that p63 might be required to provide cancer cell population with a selective advantage.[6] Studies from different groups showed that ΔNp63α overexpression contributes to cell dedifferentiation and induction of metastasis.[7,8] Conversely, it has recently been shown that complete loss of p63 is associated with cancer metastasis and poor prognosis, thus, implying that p63 can be a potential marker for cancer progression and invasion.[9]

Despite its pleiotropic involvement in multiple facets of epithelial cell biology, mechanistic explanation of how p63 accomplishes its functions remains to be defined. The study of protein−protein interaction by mass spectrometry is an increasingly important strategy to understand the role of

* To whom correspondence should be addressed. Prof. Viola Calabrò, Dipartimento di Biologia Strutturale e Funzionale, Università Federico II, Napoli. Via Cinzia, Monte S Angelo, 80126 Napoli, Italy. Phone: +39 081 679069. Fax: +39 081 679033. E-mail: vcalabro@unina.it.

[†] Dipartimento di Chimica Organica e Biologica, Università Federico II.

[‡] Dipartimento di Biologia Strutturale e Funzionale, Università Federico II.

[§] Università di Torino.

[||] Università di Milano.

10.1021/pr9011156

Published on Web 01/19/2010

multifunctional proteins. To shed light on the role of ΔNp63α in epithelial biology, we set out to identify ΔNp63α binding proteins by co-immunoprecipitation in H1299 cells followed by mass spectrometry analysis. We identified 49 putative ΔNp63α molecular partners, including several hnRNPs. The finding that ΔNp63α associates with multiple RNA-binding proteins suggests that it may play a relevant role in RNA metabolism processes.

Materials and Methods

Cell Culture and Transfections. H1299 cells derived from human lung carcinoma were provided by ATCC (CRL-5803). Cells were grown at 37 °C and 5% (v/v) CO^2 in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum (Euroclone). HaCaT cells were maintained in RPMI medium and 10% fetal calf serum. H1299 cells were seeded at a density of about 70% confluence and transfected with 0.2 μg of plasmid encoding mycΔNp63α using LipofectAMINE 2000 (Life Technologies, Inc.).

Immunoaffinity Purification of p63-Associated Proteins. p63 complexes were affinity purified from 10 mg of total extracts prepared from MycΔNp63α transfected H1299 cells. At 24 h after transfection, cells were washed with PBS and lysed in 50 mM Tris/HCl, pH 7.5, 150 mM NaCl, 0.5% NP40, 5 mM EDTA, 10% glycerol, and protease inhibitors (Sigma). Total protein in cell extracts was quantified using the BioRad protein assay and incubated with mouse anti-IgG agarose conjugated beads (Sigma) overnight (o.n.) at 4 °C. Beads were pretreated with 5% nonfat milk. After clarification, extracts were centrifuged at 3000 rpm for 5 min at 4 °C. Cell extracts were then incubated o.n. with immobilized anti-Myc 9E10 antibody (Santa Cruz). The beads were washed extensively with lysis buffer (150 mM NaCl, 10% glycerol, and 50 mM Tris/HCl, pH 7.5). Control samples were prepared in parallel using untransfected cell extracts. Protein samples were eluted with Myc competitor peptide (200 μg/mL in BC100), TCA precipitated, resuspended in 2× loading buffer (Sigma) and loaded on a 12.5% SDS polyacrylamide gel. The gel was run for 1 h in Tris-Glycine buffer at 25 mA and stained with colloidal Coomassie blue (Pierce).

Western Blotting Analysis. Each step of the experimental procedure was paralleled by Western blotting step that served as control. Samples aliquots were resuspended in 2× loading buffer (Sigma) and loaded in a SDS-10% polyacrylamide gel. The gel was run in 10× Tris-Glycine buffer at 200 V, and transferred to a PVDF membrane (Hybond-P, Amersham Biosciences). The membrane was blocked with 4% nonfat milk in TBS (25 mM Tris, pH 7.4, and 125 mM NaCl) for 1 h at room temperature, washed 3 times with TBS and then incubated with mouse anti-Myc horseradish peroxidase-conjugated secondary antibodies (Alexis Biochemicals) Protein were revealed by enhanced chemiluminescence (RPN2132; Amersham, Buckinghamshire, U.K.).

***In Situ* Digestion and MALDI Analysis.** Trypsin, dithiothreitol (DTT), iodoacetamide and α-cyano-4-hydroxycinnamic acid were purchased from Sigma. NH_4HCO_3 was from Fluka. Trifluoroacetic acid (TFA)-HPLC grade was from Carlo Erba. All other reagents and solvents were of the highest purity available from Baker.

Slices containing protein bands were excised from the gel and destained by repetitive washes with 0.1 M NH_4HCO_3, pH 7.5, and acetonitrile. Samples were reduced by incubation with 50 μL of 10 mM DTT in 0.1 M NH_4HCO_3 buffer, pH 7.5, and carboxyamidomethylated with 50 μL of 55 mM iodoacetamide in the same buffer. Enzymatic digestion was carried out with trypsin (12.5 ng/μL) in 10 mM ammonium bicarbonate buffer, pH 7.8. Gel pieces were incubated at 4 °C for 2 h. Trypsin solution was then removed and a new aliquot of the same solution was added; samples were incubated for 18 h at 37 °C. A minimum reaction volume was used as to obtain the complete rehydratation of the gel. Peptides were then extracted by washing the gel particles with 10 mM ammonium bicarbonate and 1% formic acid in 50% acetonitrile at room temperature. The resulting peptide mixtures were desalted using ZipTip pipettes from Millipore, following the recommended purification procedure.

MALDI-TOF mass spectra were recorded using an Applied Biosystem Voyager STR instrument equipped with a nitrogen laser (337 nm). One microliter of the analyte mixture was mixed (1/1, v/v) with a 10 mg/mL solution of α-cyano-hydroxycinnamic acid in acetonitrile/50 mM citrate buffer (2/3, v/v) and was applied to the metallic sample plate and dried down at room temperature. Acceleration and reflector voltage were set up as follows: target voltage at 20 kV, grid at 66% of target voltage, delayed extraction at 150 ns to obtain the best signal to-noise ratios and the best possible isotopic resolution. Mass calibration was performed using external peptide standards purchased from Applied Biosystems. Raw data were analyzed using the computer software provided by the manufacturer as monoisotopic masses.

NanoHPLC-chip MS/MS Analysis. LC/MS/MS analyses were performed on a LC/MSD Trap XCT Ultra (Agilent Technologies, Palo Alto, CA) equipped with a 1100 HPLC system and a chip cube (Agilent Technologies). After loading, the peptide mixture was first concentrated and washed in 40 nL enrichment column (Agilent Technologies chip), with 0.2% formic acid in 2% acetonitrile as the eluent. The sample was then fractionated on a C18 reverse-phase capillary column (Agilent Technologies chip) at flow rate of 300 nL/min, with a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 2% acetonitrile) from 7 to 60% in 50 min. Peptide analysis was performed using data-dependent acquisition of one MS scan (mass range from 300 to 1800 m/z) followed by MS/MS scans of the three most abundant ions in each MS scan. Dynamic exclusion was used to acquire a more complete survey of the peptides by automatic recognition and temporary exclusion (2 min) of ions from which definitive mass spectral data was previously acquired. Nitrogen at a flow rate of 3 L/min and heated to 325 °C was used as the dry gas for spray desolvation. MS/MS spectra were measured automatically when the MS signal surpassed the threshold of 50 000 counts. Double charged ions were preferably isolated and fragmented over single charged ions. The acquired MS/MS spectra were transformed in Mascot generic file format and used for peptides identification with a licensed version of MASCOT, in a local database.

Protein Identification. Spectral data were analyzed using Analyst software (version 1.4.1) and MSMS centroid peak lists were generated using the MASCOT.dll script (version 1.6b9). MSMS centroid peaks were threshold at 0.1% of the base peak. MSMS spectra having less than 10 peaks were rejected. MSMS spectra were searched against NCBInr database (2009/07/10 version, 211 309 entries) and then converted in Swiss-protein code, using the licensed version of Mascot 2.1 (Matrix Science), after converting the acquired MSMS spectra in mascot generic file format. The Mascot search parameters were taxonomy

human; "trypsin" as enzyme allowing up to 3 missed cleavages, carbamidomethyl on as fixed modification, oxidation of M, pyroGlu N-term Q, as variable modifications, 600 ppm MSMS tolerance and 0.6 Da peptide tolerance and top 20 protein entries. Spectra with a MASCOT score <25 having low quality were rejected. The score used to evaluate quality of matches for MSMS data was higher than 30.

Comparison with IntAct Data. Data from mass spectrometry analysis were compared with those provided by IntAct database (htpp://www.ebi.ac.uk/intact/search/do/search?searchString={0}) linked in the corresponding Swiss-Prot entries (htpp://www.expasy.org/sprot/). Each protein candidate was submitted to the IntAct databank obtaining a list of all proteins involved in interaction with the query protein.

HierarchView (htpp://www.ebi.ac.uk/intact/hierarchView/display.do) was used to construct a graphical representation of the interaction network centered on each protein of interest that had been used as bait in pull-down experiments. IntAct database was also searched for selecting proteins belonging to the functionally different complexes recruited by our bait to obtain graph view of the interactions between these complexes.

Coexpression Analysis. The list of human genes whose expression is comodulated with p63 was derived essentially as described.[10] Briefly, we collected 353 human normal tissue Affymetrix microarray experiments, corresponding to 65 different tissues.[11]

We then calculated the Pearson correlation coefficients of every probe set (r_1) in the expression matrix with all other probe sets (r_x), and ranked the probe sets according to the values of this index. A directed edge was established from r_1 to r_x if r_x fell within the top 1% rows in terms of correlation with r_1.

These directed networks where then converted into undirected networks by mapping the probe sets to the corresponding Entrez Gene identifiers. The list of genes whose expression is strictly comodulated with p63 is defined as the ensemble of the first-level links of p63 in this network.

Co-Immunoprecipitation. HaCaT cells from subconfluent 100 mm plates were collected for preparation of whole-cell extracts as previously described.[12] Lysates containing 1 mg of proteins were precleared with 30 μL of protein A-agarose (50% slurry; Roche) and then incubated overnight at 4 °C with 3 μg of anti-p63 monoclonal antibody (4A4; Santa Cruz), anti-hnRNPK (H-300: sc-25373 Santa Cruz), anti-hnRNPQ (I8E4; sc-56703 Santa Cruz), or anti-FUS (A300−302A Bethyl Laboratories) antibodies. Anti-GFP antibodies (Roche) were used as control of specificity. Anti-hnRNPA/B (H-200: sc-15385 Santa Cruz) antibodies were used as control of positive interaction. The immunocomplexes were collected by incubating with protein A agarose at 4 °C for 4 h. Beads were washed vigorously with 50 mM Tris-HCl, pH 7.5, 150 mN NaCl, 5 mM EDTA, 0.5% NP40, and 10% glycerol, resuspended in 4× SDS-PAGE loading buffer and loaded directly in an 8% SDS-polyacrylamide gel.

Results

Identification of Proteins Specifically Interacting with the Bait. The proteins specifically interacting with p63 were defined by functional proteomics using transfected Myc-tagged ΔNp63α protein as bait. Two biological replicates were performed using different cellular extracts. Briefly, total protein extract (10 mg) from ΔNp63α transfected H1299 cells was precleared with underivatized agarose beads to minimize nonspecific binding. The unbound material was immunoprecipitated with polyclonal anti-Myc IgG conjugated agarose beads. The proteins bound to the affinity column were extensively washed and eluted with two different concentration of Myc peptide solution (Figure 1A, sample I and sample II). Control experiments were carried out, in parallel, using extracts from untransfected H1299 cells. Following each step of the procedure, Western blotting was performed on a sample aliquot using anti-Myc antibodies as a control (Figure 1A). The protein mixture from the first myc elution was subjected to SDS-PAGE and visualized with Coomassie Blue staining (Figure 1B).

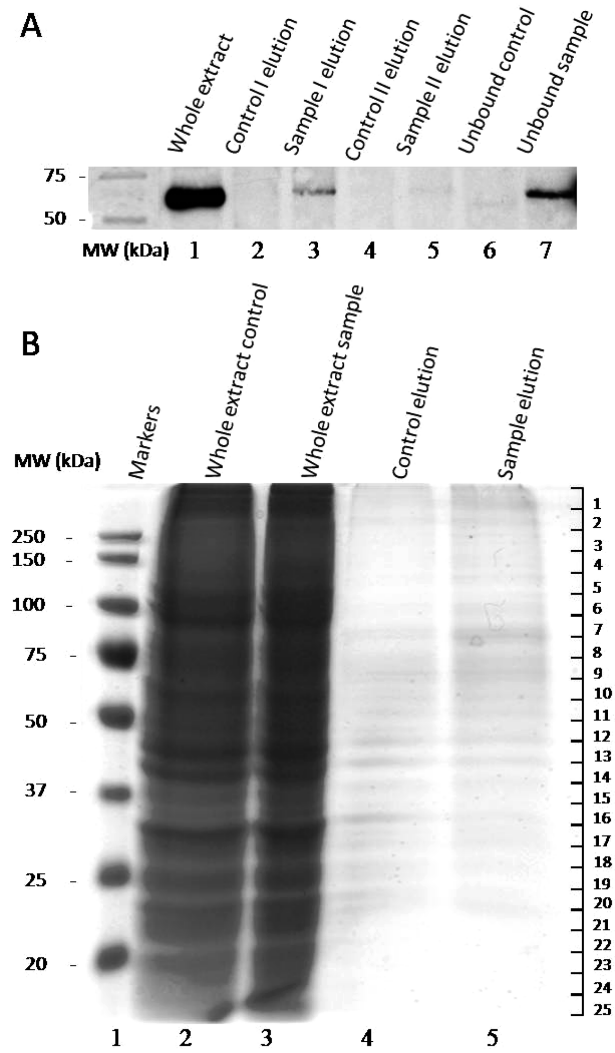


Figure 1. Western blot analysis of ΔNp63α during the affinity purification procedure. (A)Western blotting of extracts from Myc-ΔNp63α transfected H1299 cells (sample). Untransfected H1299 cell extracts were used as control. Aliquots (40 μg) corresponding to each step of the experimental procedure were subjected to Western blotting and revealed with anti-Myc and peroxidase-conjugated secondary antibodies. Whole cell extract (lane 1). Unbound control (lane 6) and sample (lane 7) after precleaning with underivatised agarose beads. Elution with myc-peptide competitor, control (lanes 2 and 4), sample (lanes 3 and 5). (B) SDS-PAGE fractionation of ΔNp63α protein complexes. ΔNp63α protein complexes were isolated by immunoprecipitation experiment. Following SDS-PAGE fractionation, Coomassie Blue stained protein bands occurring in the sample (lane 5) and in the control (lane 4) were submitted to mass spectral analyses. Corresponding gel slices from the sample and control lanes (slices 1−25) were submitted to the identification procedure. Common proteins identified in both the sample and the control gel slices were eliminated. Molecular weight markers are also shown (lane 1).

Stained gel displayed a number of discrete bands both in the control and sample lanes (lanes 4 and 5, respectively). Because of the complexity of the gel patterns and the low resolution of 1D electrophoresis, several proteins can occur in the same gel band. Therefore, protein bands specifically present in the sample lane and absent in the control lane cannot be identified by simply comparing the two gel profiles. Thus, the entire lanes from the sample and the control were cut in 25 slices each and the slices were destained and trypsin digested, *in situ*. The resulting peptide mixtures were extracted from the gel, tested by MALDI-MS, and analyzed by nanoLC/MS/MS analysis generating sequence information on individual peptides.

This information, together with the peptide mass values, was then used to search protein databases, leading to the identification of the protein components. All the gel bands from the sample and control lanes were then compared on the basis of the proteins effectively identified in each gel slice. Common proteins identified both in the control and the sample slices were discarded and only those proteins solely identified in the sample slices and absent in the control were considered. As further selection criteria, only the proteins, identified by MASCOT search with at least 3 peptides, and found exclusively in both biological replicates were selected, providing a full list of putative ΔNp63α interactors.[13] As an example, slice 11 from the sample lanes, in Figure 1B, revealed the presence of 9 proteins identified with at least 3 peptides. However, six of these putative interactors were also found in the corresponding slices from the control lanes and were then discarded. The remaining three proteins fulfilled the selection criteria and were then considered as putative p63 interactors.

This strategy let us to obtain more confident identification of truly ΔNp63α binding candidates minimizing the number of false positives.[14] It should be noted that some proteins showed a higher electrophoretic mobility, with respect to their corresponding molecular weight. This finding could be due to some proteolytic processing. On the contrary, other protein candidates exhibited a molecular mass lower than expected on the basis of the electrophoretic mobility, likely due to the occurrence of post-translational modifications. A total of 49 putative p63-interacting proteins were identified and listed in Table 1 together with their corresponding genes and the number of peptides used for their identification.

Database Analysis of Protein−Protein Interactions. To assess whether the proteins identified by the proteomic screening were novel p63 interacting partners, we extensively searched the literature and protein interaction databases for known interactors. Interestingly, hnRNPA/B (ABBP1) was already described as a p63 direct interactor.[15]

To define the functional relationships between the putative p63 interactors, we classified them according to the Gene Ontology annotation and analyzed if some of the GO keywords were significantly enriched. Six distinct functional clusters were represented: transcription, mRNA processing, mitosis and protein folding factors, ribosomal and structural proteins (Figure 2). Statistical analysis revealed that proteins involved in RNA-binding and mRNA processing were significantly enriched in our list of candidates (GO keywords: RNA binding, 1.8×10^{-13}; mRNA processing, 1.9×10^{-5}).

Integration of the Proteomic Screening with Coexpression Analysis. Integration of proteomic and transcriptomic data can be of great help to identify functionally relevant protein−protein interactions.[16]

To increase the likelihood of selecting the most functionally significant interactors, among the p63-binders, we searched microarray data for those genes having an expression profile very similar to p63. To address this question, we extracted a list of genes highly comodulated with p63 (Supporting Information) from a recently described human coexpression network obtained from a large meta analysis of normal tissues microarray data.[10] Although it is conceivable that the protein expression profiles in cancer cells could be different from normal cells, we reasoned that, in tumors, p63 should still retain functionally relevant interactions with proteins that are normally coexpressed. Moreover, integration of proteomic data from lung cancer cells with transcriptomics from normal tissues should lead to the identification of interactors that may be of interest to the p63 functions, far beyond the particular cell line used for the proteomic screening.

Interestingly, five proteins included in our list of p63 putative partners, Keratin 5, hnRNPA/B, hnRNPK, hnRNPQ (Syncrip) and FUS/TLS, are also significantly comodulated with p63, in human tissues. Keratin 5 is a structural protein; since it is highly expressed in squamous cell lung carcinomas,,[17,18 19] it might likely represent a false positive.

FUS, hnRNPQ, hnRNPK, and hnRNPA/B (ABBP1) are all RNA binding proteins involved in pre-mRNA processing, mRNA export and turnover mechanisms.[20] Since hnRNPA/B was already validated as a p63 interactor,[15] we decided to further validate the interaction of p63 with hnRNPK, hnRNPQ and FUS/TLS.

Confirmation of ΔNp63α Association with FUS/TLS, hnRNPQ, and hnRNPK. To validate the interaction of ΔNp63α with FUS/TLS, hnRNPK, or hnRNPQ, we performed co-immunoprecipitation assays in HaCaT human keratinocytes, a more physiological environment. First, we verified that hnRNPK, hnRNPQ, and FUS proteins are endogenously expressed in HaCaT cells by Western blot (data not shown). Then, we immunoprecipitated endogenous ΔNp63α, from HaCaT cells, with the 4A4 antibody. Specific immunodetection showed that FUS, hnRNPK, and hnRNPQ were present in p63 complexes (Figure 3A,B and data not shown). Conversely, p63 was specifically pulled down by antibodies against hnRNPQ, hnRNPK, or FUS (Figure 3C and data not shown). Interestingly, the amount of associated p63 was much higher for hnRNPK than for FUS or hnRNPQ, indicating that the fraction of FUS or hnRNPQ associated with p63 was a minor pool of the total FUS and hnRNPQ proteins. Neither FUS nor hnRNP proteins were immunoprecipitated using anti-GFP antibodies (Figure 3). Immunoprecipitation with antibodies against ABBP1 was performed as control of positive interaction (Figure 3D). These results, together with the GO keyword overrepresentation analysis, indicate that p63 might play a role in association with components of ribonuclear protein complexes.

IntAct Data. Finally, we compared our proteomic results with those provided by IntAct database (see Materials and Methods). Each protein candidate was submitted to the IntAct databank obtaining a list of all proteins involved in interaction with the query protein. HierarchView (htpp://www.ebi.ac.uk/intact/hierarchView/display.do) was used to construct a graphical representation of the interaction network centered on each protein of interest. IntAct database was also searched for selecting proteins belonging to functionally different complexes recruited by our bait. Total interaction graph view is reported (Supporting Information), where the putative interactors identified by our MS approach are evidenced. Similar results were

Table 1. Full List of ΔNp63α Putative Interactors Identified by Mass Spectrometry

slice[a]	molecolar weight	gene symbol	accession number	protein name	number of peptides	score	amino acid coverage
2	273400 Da	FAS	P49327 FAS	FAS [TNF receptor member 6]	17	633	10%
	242984 Da	CAD	P27708 PYR1	Cad protein	4	51	3%
	290959 Da	FLNC	Q14315 FLNC	Filamin-C	3	50	2%
4	140958 Da	DHX9	Q08211 DHX9	ATP-dependent RNA helicase A	7	124	6%
5	118315 Da	SEC24C	P53992 SC24C	Protein transport protein Sec24C	3	44	2%
6	89322 Da	VCP	P55072 TERA	Transitional endoplasmic reticulum ATPase	5	118	7%
	521126 Da	SACS	Q9NZJ4 SACS	Sacsin	8	47	3%
	100200 Da	PSMD2	Q13200 PSMD2	26S proteasome non-ATPase regulatory subunit 2	4	56	4%
7	70943 Da	HNRNPR	O43390 HNRPR	Heterogeneous nuclear ribonucleoprotein R	3	81	5%
	261140 Da	GOLGA4	Q13439 GOGA4	Golgin subfamily A member 4	9	44	4%
8	72333 Da	HSPA5	P11021 GRP78	78 kDa glucose-regulated protein [Precursor]	6	146	14%
	53426 Da	FUS	P35637 FUS	RNA-binding protein FUS	6	173	9%
10	163278 Da	A2M	P01023A2MG	Alpha-2-macroglobulin [Precursor]	5	116	4%
	69603 Da	SYNCRIP	O60506 HNRPQ	Heterogeneous nuclear-ribonucleoprotein Q	7	93	18%
	45923 Da	ZNF676	Q8N7Q3 ZN676	Zinc finger protein 676	3	39	7%
11	76785 Da	TP63	Q9H3D4 P63	Tumor protein 63	23	270	37%
	59210 Da	CPSF6	Q16630 CPSF6	Cleavage and polyadenylation specificity factor subunit 6	3	69	5%
	70436 Da	PLS3	P13797 PLST	Plastin-3	3	56	7%
12	62378 Da	KRT5	P13647 K2C5	Keratin, type II cytoskeletal 5	5	168	9%
	63460 Da	CARM1	Q86 × 55 CARM1	Histone-arginine methyltransferase CARM1	5	159	10%
	60187 Da	HNRNPL	P14866 HNRPL	Heterogeneous nuclear ribonucleoprotein L	7	139	19%
	59621 Da	CCT8	P50990 TCPQ	T-complex protein 1 subunit theta	4	101	11%
	54232 Da	NONO	Q15233 NONO	Non-POU domain-containing octamer-binding protein	4	73	9%
	50976 Da	HNRNPK	P61978 HNRPK	Heterogeneous nuclear ribonucleoprotein K	7	50	19%
13	60025 Da	KRT6C	P48668 K2C6C	Keratin, type II cytoskeletal 6C	13	797	25%
	49895 Da	TUBA1C	Q9BQE3 TBA1C	Tubulin alpha-1C chain	7	173	22%
	59751 Da	ATP5A1	P25705 ATPA	ATP synthase subunit alpha, mitochondrial [Precursor]	3	86	7%
15	46109 Da	RPL3	P39023 RL3	60S ribosomal protein L3	4	104	14%
	50141 Da	EEF1A1	P68104 EF1A1	Elongation factor α 1	4	90	9%
16	32854 Da	RPSA	P08865 RSSA	40S ribosomal protein SA	5	103	24%
	105114 Da	ZBTB4	Q9P1Z0 ZBTB4	Zinc finger and BTB domain-containing protein 4	3	42	4%
17	32142 Da	HNRPCL1	O60812 HNRCL	Heterogeneous nuclear ribonucleoprotein C-like 1	4	134	12%
	39595 Da	HNRPA3	P51991 ROA3	Heterogeneous nuclear ribonucleoprotein A3	3	125	11%
	36225 Da	HNRPAB	Q99729 ROAA	Heterogeneous nuclear ribonucleoprotein A/B	3	70	8%
	306482 Da	MAP1A	P78559 MAP1A	Microtubule-associated protein 1A	7	43	4%
19	34274 Da	RPLP0	P05388 RLA0	60S acidic ribosomal protein P0	9	374	35%
	22350 Da	HIST1H1D	P16402 H13	Histone H1.3	3	63	16%
	162505 Da	UACA	Q9BZF9 UACA	Uveal autoantigen with coiled-coil domains and ankyrin repeats	4	39	4%
20	31280 Da	SNRPA	P09012 SNRPA	U1 small nuclear ribonucleoprotein A	6	165	24%
	147810 Da	UPF2	Q9HAU5 RENT2	Regulator of nonsense transcripts 2	4	35	5%
21	316415 Da	CENPE	Q02224 CENPE	Centromeric protein E	7	44	2%
22	102642 Da	SAFB	Q15424 SAFB1	Scaffold attachment factor B1	4	42	8%
	41775 Da	EDG8	Q9H228 S1PR5	Sphingosine 1-phosphate receptor Edg-8	3	39	9%
24	17779 Da	RPL24	P83731 RL24	60S ribosomal protein L24	3	90	15%
	23577 Da	RPL13A	P40429 RL13A	60S ribosomal protein L13a	3	40	12%
25	17819 Da	RPL12	P30050 RL12	60S ribosomal protein L12	4	91	18%
	19794 Da	MRLC2	P19105 MRLC3	Myosin regulatory light chain 2, nonsarcomeric	3	58	15%
26	83264 Da	HSP90AB1	P08238 HS90B	Heat shock protein HSP 90-beta	13	413	22%

[a] Gel slice numbers correspond to those reported in Figure 1B.

obtained performing the search with different programs such as http://string.embl.de/ or http://www.thebiogrid.org/.

Discussion

ΔNp63α plays a pivotal role in epithelial cell proliferation, differentiation and transformation. However, the mechanistic aspects of how p63 accomplishes its multiple functions remain somewhat enigmatic and controversial. To address this point, we have performed a comprehensive analysis of the interactome of ΔNp63α in H1299 cells by mass spectrometry. We have chosen these cells because they express neither p53 nor p73, thus, making our analyses independent from the other members of the p53 gene family.

To identify protein specifically interacting with ΔNp63α, we performed co-immunoprecipitation and mass-spectrometry analysis. We adopted the very stringent criterion of filtering out those proteins that were not identified in two independent biological replicates as well as those proteins that were seen in the nonspecific elution sample. We identified a total of 49 proteins; 48 out of 49 were novel p63 putative partners. We reasoned that ribosomal proteins, Keratin 5 and 6 could be contaminants because of their abundance. Nevertheless, since keratins are structural proteins highly expressed in lung carcinomas and are induced by p63,[17,1819] we cannot exclude that they may actually interact with p63.

Data from the functional proteomics experiments were integrated with transcriptomics from normal tissues. The rationale of integrating proteomic data from lung cancer cells with transcriptomics from normal tissues was that we wanted to focus on the interactors that may be of interest to the p63

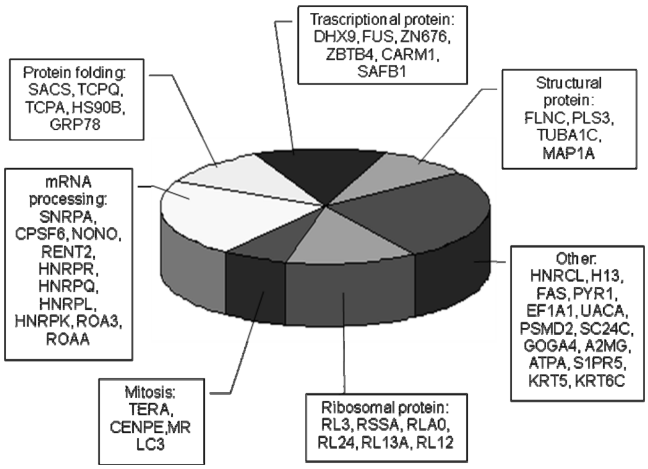


Figure 2. Functional clusters of ΔNp63α interacting partners. Schematic representation of the interactors identified in the proteomic screening. The interactors identified by proteomics are grouped in different functional categories on the basis of their reported biological activities.

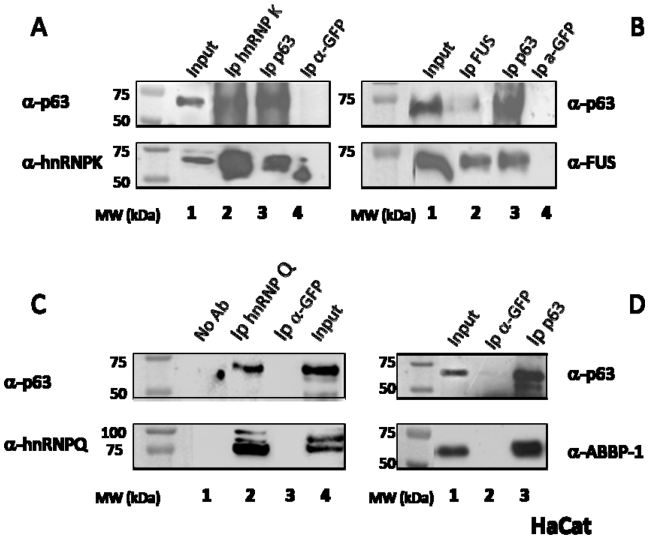


Figure 3. ΔNp63α interacts with hnRNPK, hnRNPQ, and FUS/TLS. (A) Untransfected HaCaT cell lysates were immunoprecipitated with anti hnRNPK, anti-p63, or anti-GFP antibodies. Immunocomplexes were blotted and probed with anti-hnRNPK and anti-p63 antibodies. The expression level of endogenous hnRNPK and p63 are shown in lane 1 (input). (B) Untransfected HaCaT cell lysates were immunoprecipitated with anti FUS or anti-p63 antibodies, as indicated. The immunocomplexes were blotted and probed with anti-FUS and anti-p63 antibodies. The expression level of endogenous FUS and p63 are shown in lane 1 (input). (C) Cell lysates from untransfected HaCaT cells were immunoprecipitated with anti-hnRNPQ or anti-GFP antibodies. Immunocomplexes were blotted and probed with antibodies against hnRNPQ or p63. The expression level of endogenous hnRNPQ and p63 are shown in the input lane. (D) Immunocomplexes from untransfected HaCaT cell lysates obtained with anti-p63 or anti-GFP antibodies were analyzed by SDS-PAGE and probed with anti-p63 and anti-ABBP1 antibodies. The expression levels of p63 and ABBP1 are shown in lane 1 (input).

functions, far beyond the particular cell line used for the proteomic screening.

Among others, we identified ER stress response proteins including TCPA, TCPQ, HSPA5, and HSP90B chaperones. Our list also includes Sacsin (SACS) that participates to chaperone-mediated protein folding[20] and GRP78 that facilitates the assembly of multimeric protein complexes inside the ER.[21] Interestingly, GRP78 belongs to the family of HSP70, a gene whose promoter is transactivated by ΔNp63α.[22]

Intriguingly, our list of ΔNp63α putative interactors includes a remarkable number of heterogeneous ribonucleoproteins (hnRNP). HnRNPs represent a wide gene family whose components are involved in RNA metabolism processes and transport.[23,24] We identified hnRNPR, hnRNPQ, hnRNPL, hnRNPA3, hnRNPA/B, and hnRNPK as potential components of p63 protein complexes. The presence in our list of hnRNPA/B, also known as ABBP1 (Apobec binding protein 1), already identified as ΔNp63α partner, and hnRNPQ, which interacts with ABBP1, represents a first strong validation of our approach. Remarkably, it was previously shown that the binding of ΔNp63α to ABBP1 causes a shift of the FGFR-2 alternative splicing toward its epithelial specific isoform.[15] Furthermore, it was demonstrated that ΔNp63α associates with the pre-mRNA splicing factor SRA4 which mediates p63 association with the carboxy-terminal domain of RNA polymerase II. Interestingly, AEC-associated mutations of ΔNp63α affect the binding of p63 to SRA4 causing aberrant splicing of ΔNp63α.[25] The finding of a physical association between hnRNPK and p63 was particularly attractive since it has been described as a transcriptional coactivator of p53 in response to DNA damage. HnRNPK is implicated in chromatin remodelling, transcription and translation processes.[26] Moreover, like p53, hnRNPK is ubiquitinated by the MDM2 E3 ubiquitin ligase, a major regulator of p53 activity. Depletion of hnRNPK inhibits transcriptional induction of p53 target genes and causes defects in DNA damage induced cell cycle arrest.[27]

FUS/TLS is a RNA binding protein, and is a member of the Ewing's sarcoma family of RNA binding proteins. It often occurs as a fusion protein with the transcription factor CHOP, which is oncogenic. Apparently, FUS might function in nuclear-cytoplasmic transport of mRNA. However, its specific functions are still unknown.[28]

hnRNPQ, hnRNPK, and FUS/TSL were found in the list of genes whose expression is consistently comodulated with p63 in normal human tissues. We verified their association with ΔNp63α using endogenous proteins in HaCaT keratinocytes, a more physiological context. Interestingly, when co-immunoprecipitation was performed using the anti-FUS and the Western blot was performed using the anti-p63, we could detect a low amount of p63. In this case, we noticed that the level of p63 bound to FUS was lower compared with the level of FUS in total cell lysate suggesting that a large amount of FUS might exist as a free form or engaged in distinct protein−protein complexes. Interestingly, human keratinocytes express abundant levels of FUS/TLS protein that localizes to the nuclei at all levels of the epidermis.[28] In conclusion, we have identified a large set of new potential p63 protein partners that includes a considerable number of hnRNPs. HnRNPs are component of spliceosome, the large ribonucleoprotein complex that catalyzes pre-mRNA splicing. Our data indicate that p63 might be involved in mRNA metabolism processes, an aspect of the p63 biology that is still under-investigated.

Acknowledgment. We thank Dr. A. Pollice, M. Vivo and Prof. P. Pucci for helpful discussion and critical reading. We also thank Dr. U. Ala for help with the GO keywords overrepresentation analysis. This work was supported by

AIRC and MIUR to G. La Mantia, MIUR to A. Amoresano and Telethon GGP05056 to L. Guerrini.

Supporting Information Available: Supplementary Figure, interactome of ΔNp63α protein partners showing the entire network of interacting proteins involving ΔNp63α, as available in the web by IntAct database. Supplementary Table 1, list of genes highly comodulated with p63 extracted from a human coexpression network obtained from a large meta analysis of normal tissues microarray data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pietsch, E. C.; Sykes, S. M.; McMahon, S. B.; Murphy, M. E. The p53 family and programmed cell death. *Oncogene* **2008**, *27*, 6507–6521.

(2) Levrero, M.; De Laurenzi, V.; Costanzo, A.; Gong, J.; Wang, J. Y.; Melino, G. The p53/p63/p73 family of transcription factors: overlapping and distinct functions. *J. Cell Sci.* **2000**, *113*, 1661–1670.

(3) Yang, A.; Schweitzer, R.; Sun, D.; Kaghad, M.; Walker, N.; Bronson, R. T.; Tabin, C.; Sharpe, A.; Caput, D.; Crum, C.; McKeon, F. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **1999**, *398*, 714–718.

(4) Chi, S. W.; Ayed, A.; Arrowsmith, C. H. Solution structure of a conserved C-terminal domain of p73 with structural homology to the SAM domain. *EMBO J.* **1999**, *18*, 4438–4445.

(5) Di Costanzo, A.; Festa, L.; Duverger, O.; Vivo, M.; Guerrini, L.; La Mantia, G.; Morasso, M. I.; Calabrò, V. Homeodomain protein Dlx3 induces phosphorylation-dependent p63 degradation. *Cell Cycle* **2009**, *8*, 1185–1195.

(6) Parsa, R.; Yang, A.; McKeon, F. J. Association of p63 with proliferative potential in normal and neoplastic human keratinocytes. *Invest. Dermatol.* **1999**, *113*, 1099–105.

(7) Urist, M. J.; Di Como, C. J.; Lu, M. L.; Charytonowicz, E.; Verbel, D.; Crum, C. P.; Ince, T. A.; McKeon, F. D.; Cordon-Cardo, C. Loss of p63 expression is associated with tumor progression in bladder cancer. *Am. J. Pathol.* **2002**, *161*, 1199–206.

(8) Koga, F.; Kawakami, S.; Fujii, Y.; Saito, K.; Ohtsuka, Y.; Iwai, A.; Ando, N.; Takizawa, T.; Kageyama, Y.; Kihara, K. Impaired p63 expression associates with poor prognosis and uroplakin III expression in invasive urothelial carcinoma of the bladder. *Clin. Cancer Res.* **2003**, *15*, 5501–7.

(9) Adorno, A.; Cordenonsi, M.; Montagner, M.; Dupont, S.; Wong, C.; Hann, B.; Solari, A.; Bobisse, S.; Rondina, M. B.; Guzzardo, V.; Parenti, A. R.; Rosato, A.; Bicciato, S.; Balmain, A.; Piccolo, S. A mutant p53/Smad complex opposes p63 to empowered TGBβ induced metastasis. *Cell* **2009**, *137*, 87–98.

(10) Ala, U.; Piro, R. M.; Grassi, E.; Damasco, C.; Silengo, L.; Oti, M.; Provero, P.; Di Cunto, F. Prediction of human diseases genes by human-mouse conserved coexpression analysis. *PLoS Comput. Biol.* **2008**, *28*, 37–43.

(11) Roth, R. B.; Hevezi, P.; Lee, J.; Willhite, D.; Lechner, S. M.; Foster, A. C.; Zlotnik, A. Gene expression analyses reveal molecular relationship among 20 regions of the human CNS. *Neurogenetics* **2006**, *7*, 67–80.

(12) Rossi, M.; De Simone, M.; Pollice, A.; Santoro, R.; La Mantia, G.; Guerrini, L.; Calabrò, V. Itch/AIP4 associates with and promotes p63 protein degradation. *Cell Cycle* **2006**, *5*, 1816–1822.

(13) Köcher, T.; Superti-Furga, G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* **2007**, *4*, 807–815.

(14) Monti, M.; Orrù, S.; Pagnozzi, D.; Pucci, P. Interaction proteomics. *Biosci. Rep.* **2005**, *25*, 45–56.

(15) Fomenkov, A.; Huang, Y. P.; Topaloglu, O.; Brechman, A.; Osada, M.; Fomenkova, T.; Yuriditsky, E.; Trink, B.; Sidransky, D.; Ratovitski, E. p63α mutations lead to aberrant splicing of keratinocyte growth factor receptor in the Hay-Wells syndrome. *J. Biol. Chem.* **2003**, *278*, 23906–23914.

(16) Lemeer, S.; Pinkse, M. W.; Mohammed, S.; van Breukelen, B.; den Hertog, J.; Slijper, M.; Heck, A. J. On line automated in vivo zebrafish phosphoproteomics. *J. Proteome Res.* **2008**, *7*, 1555–1564.

(17) Camilo, R.; Capelozzi, V. L.; Siqueira, S. A.; Del Carlo Bernardi, F. Expression of p63, keratin 5/6, keratin 7, and surfactant-A in non-small cell lung carcinomas. *Hum. Pathol.* **2006**, *37*, 542–546.

(18) Boldrup, L.; Coates, P. J.; Gu, X.; Nylander, K. DeltaNp63 isoforms regulate CD44 and keratins 4, 5/6, 14 and 19 in squamous cell carcinoma of head and neck. *J. Pathol.* **2007**, *213*, 384–391.

(19) Romano, R. A.; Birkaya, B.; Sinha, S. A functional enhancer of keratin 14 is a direct transcriptional target of ΔNp63. *J. Invest. Dermatol.* **2008**, *127*, 1175–1186.

(20) Takiyama, Y. Sacsinopathies: sacsin-related ataxia. *Cerebellum* **2007**, *28*, 1–7.

(21) Fu, Y.; Wey, S.; Wang, M.; Ye, R.; Liao, C. P.; Roy-Burman, P.; Lee, A. S. Pten null prostate tumorigenesis and AKT activation are blocked by targeted knockout of ER chaperone GRP78/BiP in prostate epithelium. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19444–519449.

(22) Wu, G.; Nomoto, S.; Hoque, M. O.; Dracheva, T.; Osada, M.; Lee, C. C.; Dong, S. M.; Guo, Z.; Benoit, N.; Cohen, Y.; Califano, J.; Moon, C.S,; Ratoviski, E.; Jen, J.; Sidranski, D.; Trink, B. ΔNp63α and TAp63α regulate transcription of genes with distinct biological functions in cancer development. *Cancer Res.* **2003**, *63*, 2351–2357.

(23) Kim, J. H.; Hahm, B.; Kim, Y. K.; Choi, M.; Jang, S. K. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* **2000**, *298*, 395–405.

(24) Martinez-Contreras, R.; Cloutier, P.; Shkreta, L.; Fisette, J. F.; Revil, T.; Chabot, B. hnRNP protein and splicing control. *Adv. Exp. Med. Biol.* **2007**, *623*, 123–147.

(25) Huang, Y. P.; Kim, Y.; Li, Z.; Fomenkov, T.; Fomenkov, A.; Ratovitski, E. A. AEC-associated p63 mutations lead to alternative splicing/protein stabilization of p63 and modulation of Notch signaling. *Cell Cycle* **2005**, *4*, 1440–1447.

(26) Bomsztyk, K.; Denisenko, O.; Ostrowski, J. hnRNP K: one protein multiple processes. *BioEssays* **2004**, *26*, 629–638.

(27) Moumen, A.; Masterson, P.; O'Connor, M. J.; Jackson, S. P. hnRNP K: An HDM2 target and transcriptional coactivator of p53 in response to DNA damage. *Cell* **2005**, *123*, 1065–1078.

(28) Champliaud, M. F.; Champliaud, D.; Albalat, R.; Burgeson, R.; Magro, C.; Baden, D. Localization and characterization of the RNA binding protein TLS in skin and stratified mucosa. *J. Invest. Dermatol.* **1998**, *110*, 277–281.

PR9011156

Anal Bioanal Chem (2009) 395:2269–2280
DOI 10.1007/s00216-009-3185-y

ORIGINAL PAPER

Proteomic strategies for the identification of proteinaceous binders in paintings

Gabriella Leo · Laura Cartechini · Piero Pucci · Antonio Sgamellotti · Gennaro Marino · Leila Birolo

Received: 7 August 2009 / Revised: 18 September 2009 / Accepted: 21 September 2009 / Published online: 6 October 2009

Abstract The identification of proteinaceous components in paintings remains a challenging task for several reasons. In addition to the minute amount of sample available, complex and variable chemical composition of the paints themselves, possible simultaneous presence of several binders and contaminants, and degradation of the original materials due to aging and pollution are complicating factors. We proposed proteomic strategies for the identification of proteins in binders of paintings that can be adapted to overcome the requirements and difficulties presented by specific samples. In particular, we worked on (1) the development of a minimally invasive method based on the direct tryptic cleavage of the sample without protein extraction; (2) the use of microwave to enhance the enzymatic digestion yield, followed by the analysis of the peptide mixtures by nanoLC-MS/MS with electrospray ionization (ESI). Moreover, as an additional tool to tackle the problem of contaminating proteins, we exploited the possibility of generating an exclusion list of the mass signals that in a first run had been fragmented and that the mass spectrometer had to ignore for fragmentation in a subsequent run. The methods, tested on model samples, allowed the identification of milk proteins in a sample from paintings attributed to Cimabue and Giotto, thirteenth-century Italian masters, decorating the vaults of the upper church in the Basilica of St. Francis in Assisi, Italy.

Keywords Proteinaceous binders · Enzymatic digestion · LC-ESI-MS/MS · Paintings · Paraloid B72

Electronic supplementary material The online version of this article (doi:10.1007/s00216-009-3185-y) contains supplementary material, which is available to authorized users.

G. Leo · P. Pucci · G. Marino · L. Birolo
Dipartimento di Chimica Organica e Biochimica,
Università di Napoli "Federico II",
80126 Naples, Italy

L. Cartechini
Istituto di Scienze e Tecnologie Molecolari, CNR,
c/o Dipartimento di Chimica, Università degli Studi di Perugia,
via Elce di Sotto 8,
06123 Perugia, Italy

A. Sgamellotti
Dipartimento di Chimica, Università degli Studi di Perugia,
via Elce di Sotto 8,
06123 Perugia, Italy

G. Marino · L. Birolo
Facoltà di Scienze Biotecnologiche,
Università di Napoli "Federico II",
80126 Naples, Italy

L. Birolo (✉)
Dipartimento di Chimica Organica e Biochimica, Università degli Studi Federico II, Complesso Universitario Monte S. Angelo,
via Cynthia 4,
80126 Naples, Italy
e-mail: birolo@unina.it

Introduction

Science can interact with art in many ways. One way, and the most important one currently, is providing tools to safeguard our cultural heritage for future generations.

Art paintings have traveled through the ages and the artistic techniques and materials employed were directly associated with the geographical, social, and time frames in which the artists lived and created.

The characterization of painting materials using analytical methods contributes not only to our knowledge of the artists' techniques, but also to our understanding of the evolution of painting styles. Furthermore, it is useful for

dating artworks as well as in authenticating of historic paintings. Determination of chemical composition of painting materials is an essential step that restorers require to choose the best preservation procedure. In fact, because aging can cause irreversible damage to ancient masterpieces, curators will not understand the sources of deterioration of artworks and cannot arrest degradative processes without a prior knowledge of the painting composition.

From a technical point of view, paintings are complex materials characterized by intrinsic heterogeneity on a microscale. They usually consist of multiple layers of 10–200 μm thickness made of organic (lakes) and/or inorganic pigments dispersed in an organic matrix (binder), applied on a preparation substrate, and coated with a varnish protective layer. This complexity makes the approach to their scientific investigation quite challenging, particularly when considering that analytical studies in the science of cultural heritage often involve invaluable works of art, for which destructive sampling techniques should be avoided or severely limited wherever possible. Consequently, scientists have made great efforts to put forward either non-destructive or micro-destructive analytical approaches to characterize painting materials and their alteration products [1]. This goal is especially true for the wide class of natural organic substances historically used in art (oils, resins, gums, wax, and proteinaceous material), due to their variable composition and their propensity to undergo chemical and biological degradation processes [2].

Of the many proteinaceous substances found in nature, animal glue, egg (both yolk and albumen), and casein from milk have long been used for preparation of binders, coatings, and adhesives, used alone, mixed together, or in combination with oils and other organic materials. Different methods have been proposed in literature to characterize proteinaceous components in paintings, and, so far, identification of the specific proteins occurring in binders has been essentially performed using high-performance liquid chromatography (HPLC), gas chromatography (GC)-mass spectrometry (MS), and pyrolysis (Py)-GC-MS [3, 4].

GC-MS and HPLC methods are based on the quantitative determination of the amino acids profile following protein hydrolysis. In Py-GC-MS analysis, recognition of proteins is achieved by observing specific markers formed during the pyrolytic step. However, a number of drawbacks can affect these methods, i.e., presence of contaminants, similar amino acid composition of proteinaceous binders from different biological sources (i.e., collagen from bovine or rabbit), and, particularly, alteration of amino acid ratios during sample preparation procedures or modification of the residues due to natural reactions as the artwork ages [3].

In the last few years, several new methods for reliable differentiation of proteinaceous materials have been developed, including ELISA, immunofluorescence and chemiluminescence, but they still need further refinement for routine laboratory applications [5–8].

More recently, protein mass spectrometry was successfully used for painting analysis, providing promising results [9–11]. As in a classical proteomic procedure, the proteinaceous binders were enzymatically digested with specific proteases, typically trypsin, which selectively cleaves the peptide bond following basic amino acids (namely lysine and arginine). From each individual protein occurring in the binders, a specific set of peptides is released following the enzymatic digestion that constitutes a molecular "fingerprint" of that particular protein. This peptide fingerprint is by far more specific than amino acids profile in identifying proteins. The peptide mixture originated by enzyme digestion can be analyzed by different mass spectrometry procedures leading to the unambiguous identification of proteins. The highly sensitive MALDI-TOF analysis generates a mass spectrum characterized by the accurate mass values of the peptides that often enables a reliable identification by comparison with reference samples in databases without further analysis [11, 12]. With samples containing several proteins in mixture, where the above procedure can fail to provide confident results, conversely, nanoliquid chromatography coupled to tandem mass spectrometry analysis (LC-MS/MS) provides sequence information on some peptides in the mixture, thus allowing unambiguous identification of the proteins in artworks [10].

However, application of mass spectrometry to conservation science still deserves specific considerations arising from the intrinsic nature of the samples themselves, These include a minute quantity of available sample, the need for minimally invasive analysis, the complex and quite variable chemical composition of the paints themselves, because of the possible simultaneous presence of several binders and contaminants, and degradation of the original materials affected by aging and pollution. Thus, the field of art restoration demands analytical methods that offer high sensitivity and are unaffected by endogenous as well as exogenous interferences.

Our attention focused on protein identification by an analytical approach exploiting the potentials of proteomic techniques. In particular, we developed a minimally invasive method based on the direct enzymatic digestion without protein extraction, and the use of microwaves to enhance the hydrolysis reaction, followed by the analysis of the obtained peptide mixture by nanoLC-MS/MS with electrospray ionization (ESI). Moreover, as additional tool to tackle the problem of contaminating proteins, we proposed a strategy based on creating an exclusion list of mass signals and running replicates of the sample where the precursor ions detected in the initial run(s) are excluded for MS/MS in the subsequent run.

We investigated the possible use of proteinaceous binders in the stunning ensemble of the thirteenth-century

paintings attributed to Cimabue and Giotto decorating the vaults of the upper church in the Basilica of St. Francis in Assisi. Following the earthquake that damaged the church on September 26, 1997, the two transept vaults showing St. Matthew the Evangelist and St. Hieronymus collapsed, crashing the masterpieces into tens of thousands pieces. During the restoration campaign, some tiny fragments of the collapsed paintings became available for diagnostic purposes. Unfortunately, the lack of a systematic analytical campaign, combined with the uncertain localization of the painting fragments, and tiny amount of analyzed material, not representative of the whole painting surface, will hinder any art history conclusions based on our results.

Before analyzing the samples from the vaults, we tested and optimized the procedures for a reliable and efficient identification of proteinaceous materials on several painting models. Furthermore, to evaluate possible analytical interferences of the acrylic resin used to consolidate the fragments, laboratory models treated with the fixative Paraloid B72 were also considered during experimentation.

Materials and methods

Reagents

Ammonium hydrogen carbonate (AMBIC), ethylenediaminetetraacetic acid (EDTA) and iodoacetamide were purchased from Fluka; Tri(hydroxymethyl)aminomethane (Tris), urea, dithiothreitol (DTT), Proteomics sequencing grade trypsin were from Sigma; formic acid from Baker and acetonitrile (ACN) were purchased from Romil. Deionized water was obtained from Millipore cartridge equipment.

Painting samples

Pictorial models were prepared using carbonated plaster as support to simulate mural paintings. The plaster was obtained by carbonation of a mixture of lime and sand with a weight ratio of about 1:3 and the painting layers were applied directly onto the dried surface (from hours to days to get dried). Either malachite ($CuCO_3{\cdot}Cu(OH)_2$) or hematite (Fe_2O_3) were used as pigments while egg, skimmed milk, and animal glue were used as binders with a pigment to binder ratio of 3:1. Additional standards homologous of those containing malachite were also prepared and treated with a solution of Paraloid B-72 (CTS s.r.l., Vicenza, Italy) in acetone. A sample of carbonated plaster of the support without painting layers was used as laboratory blank and processed in the same way of the pictorial models.

The analysis of historic painting materials from the Basilica of St. Francis in Assisi concerned four micro-samples (labeled GA01, GA07, CA14, and CA17) obtained from the painting fragments recovered from the collapsed vaults. The fragments were characterized by white and yellow (GA01) and red pigments (GA07, CA14, CA17) and were subjected to surface elemental analyses by non-invasive X-ray fluorescence spectroscopy (see ref. [13] for instrumental details), preliminarily to LC-MS/MS investigations. Elemental profiles showed the major presence of Ca from the substrate and Fe in all the samples, while Pb was found only in CA14 and CA17.

All samples of painting layers were available as few milligrams fragments (1–3 mg).

Sample treatment Samples and blank reference sample were subjected to different treatments depending on the following analytical procedures.

Alkaline extraction Micro-samples scraped off the painting models were subjected to protein extraction by adding 500 μl of 2.5 M NH_3 aqueous solution in an ultrasonic bath for 45 min. The extraction cycle was repeated three times. The extraction solutions were collected and dried to remove NH_3. The dried residue was resuspended in a denaturant buffer (guanidine 6 M, Tris 0.3 M, EDTA 10 mM, pH 8.0), and incubated with 40 mM dithiothreitol (final concentration) at 37 °C for 2 h to reduce disulfide bonds. The alkylation reaction was carried out with iodoacetamide at 845 mM final concentration at room temperature in the dark for 30 min. The samples were desalted with a PD10 column in ammonium bicarbonate 10 mM following absorbance at 280 nm to single out protein-containing fractions. The collected fractions were subjected to enzymatic hydrolysis.

Enzymatic digestion Samples were digested by sequencing-grade trypsin. In solution, digestion was carried out in ammonium bicarbonate 10 mM by addition of 1 μg of trypsin at 37 °C for 16 h. The reaction was stopped by lowering pH to ~1 with formic acid and the peptide mixtures were concentrated and purified using a reverse-phase C18 Zip Tip pipette tips (Millipore). The peptides were eluted with 20 μl of a solution made of 50% acetonitrile, 0.1% formic acid in MilliQ water.

The digestion in heterogeneous phase was carried out incubating for 16 h, approximately 1–3 mg of solid sample, in ammonium bicarbonate 10 mM pH 7.5, containing trypsin 0.16 μM. The supernatant was then recovered by centrifugation and the peptide mixture analyzed by LC-MS/MS.

Microwave-assisted digestion [14] was carried out in heterogeneous phase on samples prepared as above for standard heterogeneous phase digestion in solution. The sample tube was then placed in an ice bath at the center of a domestic microwave oven (700 W), with the cap open, and

irradiated for 15 min at the highest power. The supernatant was then recovered by centrifugation and the peptide mixture analyzed by LC-MS/MS.

LC-MS/MS analysis

The peptide mixtures were analyzed using a CHIP MS Ion Trap XCT Ultra equipped with a capillary 1100 HPLC system and a chip cube (Agilent Technologies, Palo Alto, CA). After loading, the peptide mixture (8 μl in 0.2% formic acid) was first concentrated and washed at 4 μl/min in 40 nl enrichment column (Agilent Technologies chip), with 0.2% formic acid in 2% acetonitrile as eluent. The sample was then fractionated on a C_{18} reverse-phase capillary column (75 μm×43 mm in the Agilent Technologies chip) at flow rate of 300 nl/min, with a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 2% acetonitrile) from 7% to 60% in 50 min.

Peptide analysis was performed using data-dependent acquisition of one MS scan (mass range from 400 to 2000 *m*/*z*) followed by MS/MS scans of the three most abundant ions in each MS scan.

Alternatively, the peptide mixtures were analyzed using a 4000 QTrap mass spectrometer (Applied Biosystem) interfaced with a nano HPLC 1100 system (Agilent Technologies). A volume of 8 μl of each sample was loaded at 10 μl/min with 0.1% formic acid on an Agilent reverse-phase pre-column cartridge (Zorbax 300 SB-C18, 5×0.3 mm, 5 μm) to desalt and concentrate the mixture. The peptides were separated on an Agilent reverse-phase column (Zorbax 300 SB-C18, 150 mm×75 μm, 3.5 μm) and the elution was performed at 0.3 μl/min using a linear gradient of solvent A (0.1% formic acid, 2% ACN in water) and solvent B (0.1% formic acid, 2% water in ACN) with an increase of solvent B from 5% to 65% in 60 min. Spectra were acquired on a *m*/*z* 400–1400 range and the MS/MS spectra were acquired on the five most abundant ions in each MS scan, using the best collision energy calculated on the bases of *m*/*z* values and charge state (rolling collision energy).

Raw data from nanoLC-MS/MS analyses were employed to query non-redundant protein databases (NCBI, either on all entries or with the appropriate taxonomy restriction) using in-house MASCOT software (Matrix Science, Boston, USA).

Results

In developing of our approach to define a general route leading to the identification of proteinaceous binders in paintings we took into account several aspects of the analytical procedure: (1) protein extraction; (2) protein hydrolysis, (3) peptide MS analysis; (4) interference from other components or contaminating background; and (5) database search.

Methodological development

First, we adopted the most widely used sample pretreatment for chromatographic analysis, i.e. protein alkaline extraction [15, 16], to verify the applicability of our proteomic analyses and to test the interference of other components, such as pigments or contaminant proteins. As second step, we developed the strategy to reduce the invasiveness of the analysis and to be ready to afford artistic samples.

Pigment interference An overview of the literature on the study and characterization of proteinaceous material in paintings indicates that analysis is often hindered by the co-presence in the samples of mineral pigments whose chemical nature could strongly affect the analytical results because of formation of complexes between amino acids and metals [15, 17]. We first tested whether our proteomic approach was also affected by metals contained in the pigments. We therefore analyzed painting models prepared with proteinaceous binders (egg, milk, or animal glue) and either hematite (a quite commonly used pigment), or malachite (a copper-containing pigment reported to be among the most interfering ones). Proteins were extracted from the painting models with ammonia, as in several experiments reported in the literature to address for metal interference [15–17]. Ammonia extraction is not a procedure for suppression of interference when copper is present.

The dried extracted proteins were submitted to reduction and alkylation steps, enzymatic hydrolysis with trypsin, followed by LC-MS/MS analysis, as in a routine proteomics experiment and applied to artistic samples by Tokarski et al. [10]. In this integrated strategy, the peptides generated in the enzymatic hydrolysis were separated by capillary chromatography in function of their hydrophobicity, and then analyzed on-line by tandem mass spectrometry, revealing information on the aminoacidic sequence beside the peptide molecular mass.

The spectra obtained were then submitted to MS/MS Ion Search in the MASCOT software, that allows protein identification based on raw MS/MS data from one or more peptides. As an example, in the painting model containing egg and malachite, five chicken proteins with more than two peptides each were identified.

As shown in Table 1, the proteinaceous binders were successfully identified in all the cases, regardless of the pigment used except for the sample with egg in presence of hematite, where only keratins could be assigned with a reasonable score (see below).

Table 1 Identification of proteins in painting models

Sample	Identified protein (accession number)	Total score[a]	Matched sequence
Egg + haematite	/	/	/
Egg + malachite	Apovitellenin-1(P02659)	276	LAEQLMEK
			LAEQLMEK+M(OX)
			NFLINETAR
			AGQFLLDVSQTTVVSGIR
			DWLVIPDAAAAYIYEAVNK
			RDWLVIPDAAAAYIYEAVNK
			DWLVIPDAAAAYIYEAVNKVSPR
			RDWLVIPDAAAAYIYEAVNKVSPR
	Vitellogenin-1(P87498)	193	LALIEVQK
			NVNFDGEILK
			LTELLNSNVR
			QKLALIEVQK
			NVPLYNAIGEHALR
			VADPIEVGIAAEGLQEMFVR
			VADPIEVGIAAEGLQEMFVR+M(OX)
			KLSDWKALPRDKPFASGYLK
			ALKLMHLLRAANEENYESVWK+M(OX)
	Vitellogenin-2(P02845)	167	YVIQEDR
			MVVALTSPR
			MVVALTSPR+M(OX)
			LPLSLPVGPR
			MTPPLTGDFR
			VGATGEIFVVNSPR
			SPQVEEYNGVWPR
			EALQPIHDLADEAISR
			NAHKAVAYVKWGWDCR
			NPVLQQVACLGYSSVVNR
Milk + haematite	Alpha-S1-casein (P02662)	28	FFVAPFPEVFGK
			YLGYLEQLLRLKK
Milk + malachite	Beta-lactoglobulin (P02754)	180	ALPMHIR
			TKIPAVFK
			TPEVDDEALEK
			TPEVDDEALEKFDK
			VYVEELKPTPEGDLEILLQK
	Alpha-S1-casein (P02662)	130	YLGYLEQLLR
			HIQKEDVPSER
			FFVAPFPEVFGK
			HQGLPQEVLNENLLR
	Beta-casein (P02666)	39	AVPYPQR
			HKEMPFPK
			KVLILACLVALALAR
			DMPIQAFLLYQEPVLGPVR
Animal glue + haematite	Collagen alpha-1 (I) bovine (P02453)	110	GLPGTAGLPGMK
			GANGAPGIAGAPGFPGAR
			SGDRGETGPAGPAGPIGPVGAR
			GPPGSAGSPGKDGLNGLPGPIGPPGPR
	Collagen alpha-2 (I) bovine (P02465)	100	IGQPGAVGPAGIR
			GEAGPAGPAGPAGPR

Table 1 (continued)

Sample	Identified protein (accession number)	Total score[a]	Matched sequence
			GEPGAPGENGTPGQTGAR
			GLPGVAGSVGEPGPLGIAGPPGAR
			GENGPVGPTGPVGAAGPSGPNGPPGPAGSR
	Collagen alpha-1 (III) bovine (P04258)	35	DGASGHPGPIGPPGPR
			SGDRGETGPAGPSGAPGPAGSR
Animal glue + malachite	Collagen alpha-1 (I) bovine (P02453)	42	GQAGVMGFPGPK+M(OX)
			GFPGLPGPSGEPGK
			SGDRGETGPAGPAGPIGPVGAR
			GARGEPGPAGLPGPPGER
			GAPGADGPAGAPGTPGPQGIAGQR
			GAPGPAGPKGSPGEAGRPGEAGLPGAK

Different models were tested containing combinations of either malachite or hematite with the most common natural binders, i.e. egg, milk, or bovine glue. Proteins were alkaline extracted, trypsin digested and analyzed by capillary LC-MS/MS. Proteins were identified searching NCBI database, with *Chordata* as taxonomy restriction, with MS/MS ion search MASCOT software (Matrix Science). Only proteins identified with at least two peptides were considered

[a] Ions score is $-10\times\mathrm{Log}(P)$, where P is the probability that the observed match is a random event. Individual ions scores >38 indicate identity or extensive homology. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. http://www.matrixscience.com/help/interpretation_help.html

Contaminant proteins interference In the case of the sample prepared with egg and hematite, MASCOT identified many signals arising from the digestion of contaminating keratins. The problem of protein contamination is also very critical in ordinary proteomic analysis, but it is usually overcome by operating all the chemical manipulations in controlled conditions, thus preventing the sample from contacting any possible source of contamination from sampling to sample loading onto the mass spectrometer, as well as preparing appropriate blank control samples. Samples coming from artwork have, instead, an intrinsic contamination problem that originates from environmental exposure, restoration interventions, or the artist's tools, such as the bristle brushes. Moreover, field control samples are generally not available for samples from artworks.

To circumvent contaminations, we adopted an exclusion list of the peptides that, in a first LC-MS/MS run allowed for identification of keratins or other protein contaminants and that in a further subsequent run have to be ignored by the mass spectrometer for fragmentation. In classical proteomic analysis, this expedient avoids "waste" of the mass spectrometer's time in fragmenting peptides derived from trypsin autodigestion or common protein contaminants and to mitigate undersampling problems [18]. Our list was created ad hoc when identification of the binder was unsuccessful in the first run because of the presence of other protein contaminants and included all signals that have been identified as peptides from contaminating sources, i.e., keratins. In our case, the list contained 64 m/z values, but, in principle, many more signals could be included.

We therefore repeated the analysis of the standard sample containing egg and hematite, following the same LC-MS/MS protocol, but for the addition of the exclusion list. Databases were interrogated with the new set of data obtained, and Apovitellenin-1 was identified with a 62 MASCOT score (Table 2), by means of two peptides, respectively the 63–70 (LAEQLMEK) and the 51–59 (NFLINETAR), thus successfully assigning egg as binder. Good quality of the fragmentation spectra was obtained in these conditions (see Fig. 1, which shows the spectrum originating from the fragmentation of the doubly charged ion at m/z of 539.30, corresponding to peptides 51–59 of apovitellin-1). To fully test the strategy, the same procedure was also successfully carried out on the sample containing animal glue and hematite that had already provided a satisfactory identification with the standard LC-MS/MS analysis, but where some keratins had also been detected. As result, a general increase of the identification score has been observed for the proteins already identified in the run without the ad hoc exclusion list (Table 2), and, moreover, one more protein was identified (Collagen alpha-2 (XI) bovine, acc. n. Q32S24) confirming the effectiveness of such a strategy.

Sample treatment Several different extraction methods have been proposed in literature, ranging from water to acidic or alkaline extraction, microwave- or ultrasonic-assisted [3, 10, 19, 20], which, in most cases provided similar results. We also tested several protocols, without observing any remarkable difference among the different extraction procedures (data not shown). The extraction of

Table 2 Ad hoc exclusion list as a tool for reducing the incidence of contaminant proteins interference for the identification of proteinaceous binder in painting models

Sample	Results from Table1			Second run with the ad hoc exclusion list		
	Identified protein (accession number)	Total score[a]	Matched sequence	Identified protein (accession number)	Total score[a]	Matched sequence
Egg + haematite	/	/	/	Apovitellenin-1 (P02659)	62	LAEQLMEK NFLINETAR
Animal glue + haematite	Collagen alpha-1 (I) bovine (P02453)	110	GLPGTAGLPGMK GANGAPGIAGAPGFPGAR SGDRGETGPAGPAGPIGPVGAR GPPGSAGSPGKDGLNGLPGPIGPPGPR	Collagen alpha-1 (I) bovine (P02453)	176	GQAGVMGFPGPK+M(OX) GPPGPMGPPGLAGPPGESGR+M(OX) SGDRGETGPAGPAGPIGPVGAR GAPGADGPAGAPGTPGPQGIAGQR GDRGETGPAGPPGAPGAPGAPGPVGPAGK QGPSGASGERGPPGPMGPPGLAGPPGESGR +M(OX)
	Collagen alpha-2 (I) bovine (P02465)	100	IGQPGAVGPAGIR GEAGPAGPAGPAGPR GEPGAPGENGTPGQTGAR GLPGVAGSVGEPGPLGIAGPPGAR GENGPVGPTGPVGAAGPSGPNGPPGPAGSR	Collagen alpha-2 (I) bovine (P02465)	100	IGQPGAVGPAGIR GEAGPAGPAGPAGPR GEPGAPGENGTPGQTGAR GLPGVAGSVGEPGPLGIAGPPGAR GENGPVGPTGPVGAAGPSGPNGPPGPAGSR
	Collagen alpha-1 (III) bovine (P04258)	35	DGASGHPGPIGPPGPR SGDRGETGPAGPSGAPGPAGSR	Collagen alpha-1(III) bovine (P04258)	147	GETGPAGPSGAPGPAGSR GEVGPAGSPGSSGAPGQR SGDRGETGPAGPSGAPGPAGSR GAAGPPGPPGSAGTPGLQGMPGER
	/	/	/	Collagen alpha-2 (XI) bovine (Q32S24)	41	QGEKGAKGDPGAVGAPGK+PyroGlu GEPGESGSPGVQGEPGVK GEKGETGQAGEAGPPGPKGPTGDDGPK

A second LC-MS/MS run was carried out with a program including an exclusion list created on the basis of the results obtained in the first run: all the signals that had been identified as peptides from contaminating sources in the first LC-MS/MS were ignored by the mass spectrometer for fragmentation. Proteins were identified searching NCBI database, either without any taxonomy restriction in order to allow contaminant identifications, or with *Chordata* as taxonomy restriction, with MS/MS Ion search MASCOT software (Matrix Science). Only proteins identified with at least two peptides were considered

[a] Ions score is $-10\times\mathrm{Log}(P)$, where P is the probability that the observed match is a random event. Individual ions scores >38 indicate identity or extensive homology. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. (http://www.matrixscience.com/help/interpretation_help.html)

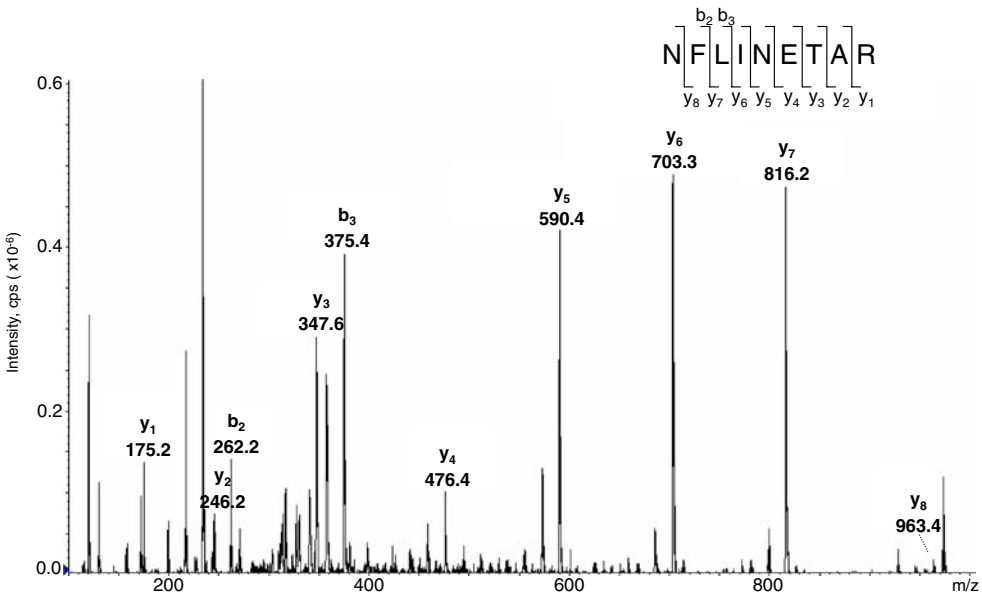


Fig. 1 MS/MS spectrum of the doubly charged ion at *m/z* 539.30, from the hydrolyzed extract of the egg-based painting model, presenting the *y* and *b* fragments of the apovitellenin-1 (P02659) peptide (NFLINETAR)

proteins can be difficult because of the low solubility of some of them. Taking advantage of the intrinsic ability of the proteomic approach to identify proteins from individual peptides without the need to extract the intact protein molecule, we also comparatively tested the results obtained digesting the proteinaceous binder directly on the sample without any pretreatment. This approach was extremely interesting, because it can be considered a minimally invasive one, since the protease in aqueous solution, at neutral pH, can be deposited on the surface of the sample and then gently removed without significantly affecting the sample itself. Moreover, by this approach, even when the protein is embedded in the painting layer, few peptides protruding from it can be digested and released by the protease in solution, and it provides enough material for a reliable identification.

Another set of solid micro-samples from the pictorial models prepared as reported in the materials and methods section, with malachite dispersed in the three considered binders was analyzed without any protein extraction or sample pretreatment but gently washing the samples with few microliters of distilled water (about 50 μl per 1–3 mg of paint fragments), followed by overnight incubation of the solid sample in heterogeneous phase, at 37 °C in a solution of ammonium bicarbonate 10 mM, pH 7.5, containing trypsin 0.16 μM. As shown in Table 3, the results obtained with this minimally invasive procedure are extremely promising as egg and milk binders were unambiguously identified in two of three samples. In one sample, 12 proteins from chicken egg were identified by means of a sequence determination of 98 peptides in total. While in the other sample, the sequence information of 17 peptides allowed the identification of six proteins from bovine milk.

The same micro-samples containing malachite were subjected to microwave-assisted digestion, always in heterogeneous phase and without any sample pretreatment other than gentle washing with distilled water. Microwave irradiation was recognized in the mid-1980s to be an efficient heating source for chemical application, where reactions that require several hours under conventional conditions can often be completed in a few minutes with very high yields and reaction selectivity [21]. In recent years, several groups have applied the microwave technique to protein digestion and shortened the digestion process to several minutes [14, 22, 23].

Table 3 reports in parallel the results obtained on the same samples treated with trypsin either with conventional incubation in water bath overnight at 37 °C or with the assistance of the exposure to microwave irradiation. The results obtained were similar but not completely superimposable. An increase in the identification score was observed in most cases when hydrolysis was microwave-assisted. Most importantly, the sample prepared with rabbit glue, for which the overnight digestion had not provided any consistent result, was successfully recognized when the digestion was performed in the microwave, with two proteins identified from seven peptides sequences. Possibly, the increase in thermal motion correlated with microwave irradiation, as well as the increase in temperature in the sample itself, improved accessibility of the substrate protein to the enzyme in the heterogeneous phase.

Finally, both conventional and microwave-assisted trypsin digestion provided satisfactory results also for the proteomic analysis of samples treated with Paraloid B72 (Table 3). Paraloid B72 is an ethyl methacrylate co-polymer often used by restorers as paintings fixative due to its properties of being durable, reversible, and non-yellowing. The deposition of Paraloid on the painting surface created a thin, transparent film that could, in principle, impair any

Table 3 Identification of the proteins in the binder of painting models by trypsin digestion in heterogeneous phase and with microwave irradiation

Sample	Identified protein (accession number)	Sample treatment			
		o/*n* trypsin digestion		Microwave-assisted trypsin digestion	
		Total score[a]	Number of identified peptides	Total score[a]	Number of identified peptides
Egg + malachite	vitellogenin 2 precursor (P02845)	1166	25	1573	40
	apolipoprotein B (P11682)	839	19	406	9
	Vitellogenin-1 precursor (P87498)	633	14	669	14
	Ovalbumin (P01012)	579	8	814	16
	ovotransferrin BB type (Q4ADJ7)	428	7	421	8
	Apovitellenin-1 precursor (P02659)	323	5	425	7
	Lysozyme (P00698)	276	5	260	4
	Ovalbumin-related protein Y (P01014)	227	5	179	4
	Vitellogenin-2 (P02845)	177	3	/	/
	Ovoglycoprotein (Q8JIG5)	156	2	/	/
	TENP protein(O42273)	111	3	/	/
	Unnamed protein product [*Gallus gallus*] (NCBI code: 63524)	80	2	165	3
	clusterin (Q9YGP0)	/	/	53	2
Milk + malachite	casein alpha-S1 (P02662)	211	7	213	5
	Serum albumin precursor (P02769)	166	3	/	/
	Beta-casein precursor (P02666)	112	3	96	4
	casein alpha-S2 (P02663)	96	2	103	2
	kappa-casein precursor (P02668)	92	1	221	5
	beta-lactoglobulin (P02754)	67	1	303	7
	Alpha-2-HS-glycoprotein (P12763)	/	/	123	3
Rabbit glue + malachite	Chain A, The Structure Of Collagen Type I (P02453)	/	/	211	5
	Collagen alpha-2 (I) chainI (P02466)	/	/	110	2
Egg + malachite + Paraloid B-72	Ovalbumin (P01012)	/	/	784	16
	Vitellogenin-2 (P02845)	264	7	755	19
	Lysozyme (P00698)	65	2	237	3
	Vitellogenin-1 precursor (P87498)	62	1	204	5
	apolipoprotein B (P11682)	57	1	375	9
	Apovitellenin-1 precursor (P02659)	/	/	240	4
	Ovotransferrin (P02789)	/	/	182	4
	Unnamed protein product [*Gallus gallus*] (NCBI code: 63524)	/	/	179	4
Milk + malachite + Paraloid B-72	casein alphaS1 (P02662)	265	5	229	7
	beta-casein (P02666)	199	5	325	9
	beta-lactoglobulin (P02754)	132	3	280	7
	kappa-casein precursor (P02668)	124	2	204	3
	alpha s2 casein (P02663)	86	1	163	3
Rabbit glue + malachite + Paraloid B-72	Collagen alpha-1 (I) (P02454)	233	5	282	6
	Collagen alpha-2 (I) chainI (P02466)			152	3

Proteins were identified searching NCBI database, with *Chordata* as taxonomy restriction, with MS/MS Ion search MASCOT software (Matrix Science). Only proteins identified with at least two peptides were considered as significant. An extended version of the table, with the sequence of the identified peptides, is reported in the Supplementary material

[a] Ions score is $-10 \times \mathrm{Log}(P)$, where P is the probability that the observed match is a random event. Individual ions scores >38 indicate identity or extensive homology. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits http://www.matrixscience.com/help/interpretation_help.html

subsequent analysis unless specific treatment was used to remove it. In all experiments, the proteinaceous binders could be successfully and unequivocally identified as egg, milk, and rabbit glue without removing the protective film of the acrylic polymer. The observation that the model sample containing iron and animal glue was not characterized while that containing also Paraloid B72 was successfully characterized (Table 3), is , however, indicative of the fact that the unquestionable potentiality of proteomics (with sensibility in the femtomolar range) is always subjected to the *alea* of sampling: the accessibility of the binder within the sample is a variable that can effectively affect the results obtained. In the former case, the use of microwave has been instrumental in producing positive results, possibly because heating and vibration have made the binder accessible to the protease. Conversely, the surface created by Paraloid B72 might not be homogenous enough to prevent the access of the enzyme to the substrate in the sample.

Application to artistic samples

The experience on laboratory pictorial models prompted us to propose a general workflow to operate with artistic samples, where the conventional digestion overnight with trypsin at 37 °C is the first choice, as the mildest one and that could be eventually considered as a minimally invasive treatment of the sample.

The four micro-samples obtained from the fragments recovered by the collapsed vaults of the Basilica of St. Francis in Assisi were carefully washed with MilliQ H_2O and incubated overnight with 200 ng of trypsin in ammonium bicarbonate 10 mM, pH 7.5. The supernatants were recovered by centrifugation and loaded for the nanoLC-MS/MS analysis. In the GA07 sample, three milk proteins could be identified with reasonable score from the sequence of 11 peptides (Table 4). An example of the fragmentation spectra that allowed milk binder identification is given in Fig. 2. On the contrary, in the other samples, no significant identification could be obtained.

The latter samples were then subjected to microwave-assisted trypsin digestion in heterogeneous phase in ammonium bicarbonate 10 mM, pH 7.5 with 200 ng of trypsin. The supernatants were recovered by centrifugation and loaded for the nanoLC-MS/MS analysis. Not even in this case was significant identification obtained, but for the GA04 sample, that returned the sequence of a peptide with a reasonable score spanning from Ala67 to Arg74 of Apovitellenin-1 precursor. One protein from the sequence of one peptide. Was that enough to unequivocally identify the binder? Probably not. According to the still-open discussion about the guidelines for peptide and protein identification data, a reliable identification of a protein requires at least two peptides with a reasonable score [24].

Any further treatment of the samples (exclusion lists, alkaline or acidic extraction) did not produce any additional successful identification.

Discussion

With the work presented in this paper, we propose proteomic strategies for the identification of proteins in binders of paintings that can be helpful to overcome requirements and difficulties presented by specific samples.

Table 4 Protein identification in the binder of fragments recovered by the collapsed vaults of the Basilica of St. Francis in Assisi

SAMPLE	Identified protein (Accession number)	Total score[a]	Matched sequence
GA07	casein alphaS1 (P02662)	113	FFVAPFPEVFGK
			VPQLEIVPNSAEER
			HQGLPQEVLNENLLR
	kappa-casein C (P02668)	91	YIPIQYVLSR
			SPAQILQWQVLSNTVPAK
			SPAQILQWQVLSNTVPAK + Deamidation (NQ)
	Beta-casein precursor (P02666)	81	DMPIQAFLLYQEPVLGPVR
			DMPIQAFLLYQEPVLGPVR + Oxidation (M)
GA04	Apovitellenin-1 precursor (P02659)	78	AGQFLLDVSQTTVVSGIR
CA14	No identification	/	/
CA17	No identification	/	/

Proteins were identified searching NCBI database, with *Chordata* as taxonomy restriction, with MS/MS Ion search MASCOT software (Matrix Science). Only proteins identified with at least two peptides were considered as significant

[a] Ions score is $-10\times\mathrm{Log}(P)$, where P is the probability that the observed match is a random event. Individual ions scores >38 indicate identity or extensive homology. Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. http://www.matrixscience.com/help/interpretation_help.html

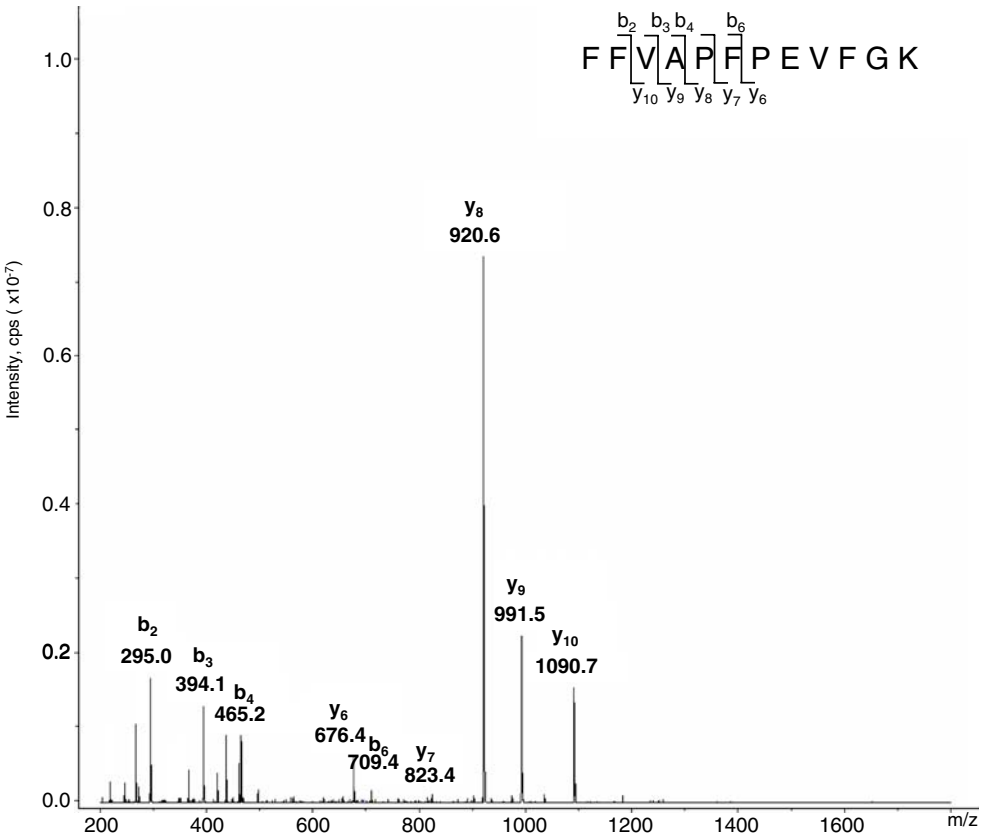


Fig. 2 MS/MS spectrum of the doubly charged ion at *m/z* 693.5 from the analysis of the microwave-assisted trypsin digestion in the heterogeneous phase of the sample GA07 from the Vault of Basilica di Assisi, presenting the *y* and *b* fragments of the casein alphaS1 (P02662) peptide (FFVAPFPEVFGK)

Until recently, the most common and reliable approach to the specific identification of protein media in old paintings was generally based on acid hydrolysis, followed by derivatization and quantitative determination of the relevant amino acids either by GC-MS or liquid chromatography [3]. Proteinaceous binders are, then, identified on the basis of the relative abundance of the quantified amino acids. With the amino acid distribution susceptible to significant changes owing to several effects (i.e., pigment interference, degradation process, and contamination), this approach may lead to erroneous results [3, 15].

The proteomic approach to properly identify a protein requires that the sequence of a few peptides is determined irrespective of the integrity of the full protein, and it does not suffer from the problems derived from quantitative determination of the aminoacidic profile. Indeed, since peptide sequences and not simple amino acid composition are determined, they are a much more specific fingerprint of a protein. Moreover, even few peptides could be enough to reliably identify a protein [24].

Most importantly, the proteomic approach can overcome the intrinsic complexity of the protein mixture in a sample. Conversely, binder identification gains reliability from the identification of more than a single protein from the same source (Tables 1, 2, and 3). What artists used were not single proteins, but natural material, such as egg, milk, or animal glue, which are composed by multiple proteins in mixture. The possibility of discriminating and identifying them constitute an added value of the proteomic approach for the proper identification of complex mixture of proteinaceous binders.

The standard proteomic approach had already been proven successful for the unambiguous and rapid identification of proteins in binding media [10]. However, samples from artistic masterpieces pose challenges, requiring specific strategies to be adopted to be successfully investigated. Because of the preciousness of the material, ideally, non-destructive treatment of the sample is mandatory. The proteomic analysis herein proposed cannot be considered as fully non-invasive as a spectroscopic one, but the heterogeneous enzymatic digestion in aqueous buffer was the softest treatment that could be proposed, akin to a simple cleaning procedure. Another advantage of the proteomic approach is that the integrity of the protein molecule is not required for reliable results. Just a few peptides protruding from the material can be recognized by the used enzyme and released in solution, and their aminoacidic sequences are sufficient to correctly identify the protein. We therefore suggest as first choice, especially for surface/top paint layer analysis, the heterogeneous enzymatic hydrolysis of the sample. The extension of this strategy to multilayer works is more complicated, and deserves further attention. The possibility

of an imaging mass spectrometry [25] of artworks is extremely attractive and should be considered in the future.

In recent years, several groups have applied the microwave technique to protein digestion to improve the digestion process and to shorten it to several minutes [14, 22, 23]. We observed a general increase of the number of peptides useful to the identification when the enzymatic digestion was microwave-assisted (see Table 3), and, most importantly, one artistic sample provided satisfactory identification of the binders when hydrolysis was carried out in the microwave oven.

We therefore suggest, as a general practice, to treat samples from artworks with heterogeneous hydrolysis at 37 °C at neutral pH. When not successful, the same sample can be subsequently submitted to more invasive procedure, such as microwave-assisted proteolysis.

It is worth mentioning that, to verify whether the negative results obtained even with microwave-assisted proteolysis were due to the procedure adopted or to the sample itself, unsuccessful samples were also submitted to the most traditional extraction methods for proteins (i.e., alkaline or acidic treatment) prior to classical enzymatic digestion and LC-MS/MS analysis. However, positive identification of the binders was never produced.

Another source of complication for this kind of sample arises from environmental contaminations. We propose a strategy of general applicability to circumvent the problem, by elaborating an exclusion list of the contaminating peptides that, in a first run have been fragmented and that the mass spectrometer had to ignore for fragmentation in a further subsequent run. Such a strategy can constitute an additional tool, to be used in difficult cases where abundant contaminant proteins have been detected.

However, although definitively successful with 90% of the model samples we tested, negative results were also obtained with artistic samples, possibly because of sampling (i.e., in a piece of the painting where the searched binder was less abundant or even absent, or where more contaminants were present). Samples from masterpieces like those we worked on, which could not be provided in several replicas, have experienced different and severe contaminations, and aging of the organic material thus dramatically increasing the complexity of the sample and complicating the identification of the components.

Acknowledgment V. Garibaldi, Soprintendente ai Beni A.P.P.S.A. D. dell'Umbria, is gratefully acknowledged for making this study possible.

References

1. Domènech-Carbò MT (2008) Anal Chim Acta 621:109–139
2. Mills JS, White R (1994) The organic chemistry of museum objects, 2nd edn. Butterworths-Heinemann, London
3. Colombini MP, Modugno F (2004) J Sep Sci 27:147–160
4. Chiavari G, Lanterna G, Luca C, Matteini M, Sandu ICA (2003) Chromatographia 57:645–648
5. Heginbotham A, Millay V, Quick M (2006) J Am Inst Conserv 45:89–106
6. Raminez-Barat B, de la Viña S (2001) Stud Conserv 46:282–288
7. Vagnini M, Pitzurra L, Cartechini L, Miliani C, Brunetti BG, Sgamellotti A (2008) Anal Bioanal Chem 392:57–64
8. Dolci LS, Sciutto G, Guardigli M, Rizzoli M, Prati S, Mazzeo R, Roda A (2008) Anal Bioanal Chem 392:29–35
9. Hynek R, Kuckova S, Hradilova J, Kodicek M (2004) Rapid Commun Mass Spectrom 18:1896–1900
10. Tokarski C, Martin E, Rolando C, Cren-Olive C (2006) Anal Chem 78:1494–1502
11. Kuckova S, Hynek R, Kodicek M (2007) Anal Bioanal Chem 388:201–206
12. Kuckova S, Nemec I, Hynek R, Hradilova J, Grygar T (2005) Anal Bioanal Chem 382:275–282
13. Ricci C, Borgia I, Brunetti BG, Miliani C, Sgamellotti A, Seccaroni C, Passalacqua P (2004) J Raman Spectrosc 35:616–621
14. Pramanik NB, Mirza UA, Ning YH, Liu YH, Bartner PL, Weber PC, Bose AK (2002) Protein Sci 11:2676–2687
15. Colombini MP, Modugno F, Giacomelli A (1999) JChromatogr A 846:101–111
16. Gauitier G, Colombini MP (2007) Talanta 73:95–102
17. De la Cruz-Cañizares J, Doménech-Carbò MT, Gimeno-Adelantado JV, Mateo-Castro R, Bosh-Reig F (2004) J Chromatogr A 1025:277–285
18. Wang N, Li L (2008) Anal Chem 80:4696–4710
19. Ronca F (1994) Stud Conserv 39:107–120
20. Wouters J, Van Bos M, Lamens K (2000) Stud Conserv 45:106–116
21. Lidstrom P, Tierney J, Wathey B, Westman J (2001) Tetrahedron 57:9225–9283
22. Sun W, Gao SJ, Wang LJ, Chen Y, Wu SZ, Wang XR, Zhang DX, Gao YH (2006) Mol Cell Proteomics 5:769–775
23. Lin S, Yao G, Qi D, Li Y, Deng C, Yang P, Zhang X (2008) Anal Chem 80:3655–3665
24. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A (2004) Mol Cell Proteomics 3:531–533
25. MacAleese L, Stauber J, Heeren RMA (2009) Proteomics 9:819–834