UNIVERSITA' DEGLI STUDI DI VERONA
FACOLTA' DI SCIENZE MM.FF.NN.
DIPARTIMENTO SCIENTIFICO E TECNOLOGICO

DOTTORATO DI RICERCA IN
BIOTECNOLOGIE INDUSTRIALI - AMBIENTALI
CICLO XIX

PROTEOMIC AND BIOCHEMICAL STUDIES OF MICROORGANISMS EXPLOITABLE IN BIO-RESTORATION AND BIO-REMEDIATION, AND DEVELOPMENT OF NEW STRATEGIES TO COPE WITH THE "SMALL AND LOW-ABUNDANCE PROTEOME"

S.S.D. BIO 10

Coordinatore: Prof. Luis Hugo Monaco

Tutor: Prof. Pier Giorgio Righetti

Dottorando: Paolo Antonioli

28 febbraio 2007

Alla mia famiglia,

In realtà si sa solo quando si sa poco,
col sapere aumenta il dubbio.
PLAUTO

POLITECNICO DI MILANO
DIPARTIMENTO DI CHIMICA, MATERIALI ED INGEGNERIA CHIMICA "Giulio Natta"
Prof. P.G. Righetti, Laboratory of Proteome Analysis, e-mail: Piergiorgio.righetti@polimi.it
Tel.: +39-02-23993045; cell phone: 329-6508572

Alla cortese attenzione del Collegio dei Docenti:

Oggetto: profilo del Dott. Paolo Antonioli, dottorando di ricerca in Biotecnologie Industriali Ambientali, XIX ciclo, Dipartimento Scientifico e Tecnologico.

Il Dott. Antonioli, nel corso del triennio di attività scientifica svolta nell'ambito del Dottorato di Ricerca in Biotecnologie Industriali-Ambientali, parte presso il Dipartimento Scientifico e Tecnologico dell'Università di Verona, e parte presso il Dipartimento di Chimica del Politecnico di Milano, ha condotto un brillante ed approfondito lavoro sia per quanto riguarda l'analisi proteomica di diversi ceppi batterici coinvolti in processi di bioremediation e di recupero del patrimonio culturale, sia per quanto riguarda nuove metodiche innovative per l'analisi del proteoma di piccola taglia.
Nello svolgimento del lavoro di ricerca il Dott. Antonioli ha sviluppato una metodica altamente innovativa per l'analisi del proteoma di piccola taglia, consistente in una separazione di complessi proteina/detergente anionico in condizioni non di elettroforesi normale, ma di focalizzazione. Tale metodica permette alta risoluzione anche di digeriti triptici di proteine, processo questo non implementabile con la metodica tradizionale.
Dal punto di vista del proteoma batterico, sono state affrontate due tematiche di grande attualità: da una parte è stato rivelato il meccanismo di azione di un ceppo di *Pseudomonas stutzeri*, in grado di recuperare un affresco (irrimediabilmente perduto) strappato dalle pareti del Camposanto di Pisa; dall'altra parte, con un complesso lavoro di indagine scientifica, è stato elucidato il meccanismo di riduzione di Se^{IV}, altamente tossico, disperso nei terreni, a Se metallico, ad opera di un ceppo di *Stenotrophomonas maltophilia.*

I temi sono stati affrontati con rigore scientifico, portando alla messa a punto di nuove metodologie per l'analisi proteomica ad alta risoluzione e alla dettagliata spiegazione delle dinamiche di azione dei due ceppi batterici analizzati.
La produzione scientifica svolta dal Dott. Paolo Antonioli in questi tre anni, in seguito ai progetti attuati, ha portato alla pubblicazione dei seguenti articoli:

1. Characterization of a new *Stenotrophomonas maltophila* strain (SeITE02) exploitable in bioremediation of contaminated soils. Antonioli P, Lampis S, Chesini I, Vallini G, Rinalducci S, Zolla L, Righetti PG. *J Bacteriol.* 2007; submitted.

2. Capturing and amplifying impurities from purified recombinant monoclonal antibodies via Equalizer Beads: a proteomic study. Antonioli P, Fortis F, Guerrier L, Rinalducci S, Zolla L, Righetti PG, Boschetti E. *Proteomics*. 2007; in press.

3. SDS-PAGE under Focusing Conditions: An Electrokinetic Transport Phenomenon Based on Charge Neutralization. Zilberstein G, Korol L, Antonioli P, Righetti PG, Bukshpan S. *Anal. Chem.* 2007; 79: 821-827.

4. Mass distribution, polydispersity and focusing properties of carrier ampholytes for IEF II: pH 4-6 intervals. Simo C, Mendieta ME, Antonioli P, Sebastiano R, Citterio A, Cifuentes A, Peltre G, Righetti PG. *Electrophoresis.* 2006; 27: 4849-4858.

5. Mass distribution and focusing properties of carrier ampholytes for isoelectric focusing: novel and unexpected results. Sebastiano R, Simó C, Mendieta ME, Antonioli P, Citterio A, Cifuentes A, Peltre G, Righetti PG. *Electrophoresis.* 2006; 27: 3919-3934.

6. A new approach for the detection and identification of protein impurities using combinatorial solid phase ligand libraries. Fortis F, Guerrier L, Areces LB, Antonioli P, Hayes T, Carrick K, Hammond D, Boschetti E, Righetti PG. *J Proteome Res.* 2006; 5: 2577-2585.

7. A new approach for the removal of protein impurities from purified biologicals using combinatorial solid-phase ligand libraries. Fortis F, Guerrier L, Areces LB, Antonioli P, Hayes T, Carrick K, Hammond D, Boschetti E, Righetti PG. *Electrophoresis.* 2006; 27: 3018-3027.

8. Quasi-isoelectric buffers for protein analysis in a fast alternative to conventional capillary zone electrophoresis. Antonioli P, Mendieta ME, Sebastiano R, Citterio A, Peltre G, Busnel JM, Descroix S, Candiano G, Righetti PG. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2006; 833: 19-25.

9. Organic and inorganic di-cations for capillary silica coating and EOF modulation in CE: Example of application in PEG analysis. Sebastiano R, Mendieta ME, Antonioli P, Bossi A, Righetti PG, Citterio A. *Electrophoresis.* 2006; 27: 1495-1501.

10. A new approach for the removal of protein impurities from purified biologicals using combinatorial solid-phase ligand libraries. Fortis F, Guerrier L, Righetti PG, Antonioli P, Boschetti E. *Electrophoresis.* 2006; 27: 3018-3027.

11. Formation of truncated proteins and high molecular-mass-aggregates upon soft illumination of photosynthetic proteins. Rinalducci S, Campostrini N, Antonioli P, Righetti PG, Roepstorff P, Zolla L. *J Proteome Res.* 2005; 4: 2327-2337.

12. Proteome analysis in the clinical chemistry laboratory: myth or reality? Righetti PG, Castagna A, Antonucci F, Piubelli C, Cecconi D, Campostrini N, Rustichelli C, Antonioli P, Zanusso G, Monaco S, Lomas L, Boschetti E. *Clin Chim Acta.* 2005; 357: 123-39.

POLITECNICO DI MILANO
DIPARTIMENTO DI CHIMICA, MATERIALI ED INGEGNERIA CHIMICA "Giulio Natta"
Prof. P.G. Righetti, Laboratory of Proteome Analysis, e-mail: Piergiorgio.righetti@polimi.it
Tel.: +39-02-23993045; cell phone: 329-6508572

13. Proteomic approaches for studying chemoresistance in cancer. Righetti PG, Castagna A, Antonioli P, Cecconi D, Campostrini N, Righetti SC. *Expert Rev Proteomics.* 2005; 2: 215-228.

14. Prefractionation techniques in proteome analysis: the mining tools of the third millennium. Righetti PG, Castagna A, Antonioli P, Boschetti E*. Electrophoresis.* 2005; 26: 297-319.

15. Art-loving bugs: the resurrection of Spinello Aretino from Pisa's cemetery. Antonioli P, Zapparoli G, Abbruscato P, Sorlini C, Ranalli G, Righetti PG. *Proteomics*. 2005; 5, 2453-2459.

16. A proteomic approach to cisplatin resistance in the cervix squamous cell carcinoma cell line A431. Castagna A, Antonioli P, Astner H, Hamdan M, Righetti SC, Perego P, Zunino F, Righetti PG. *Proteomics.* 2004; 4, 3246–3267.

17. Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches. Righetti PG, Castagna A, Antonucci F, Piubelli C, Cecconi D, Campostrini N, Antonioli P, Astner H, Hamdan M. *J. Chromatogr. A.* 2004; 1051: 3-17.

Il numero delle pubblicazioni prodotte certifica l'importante impegno profuso dal Dott. Paolo Antonioli nell'adempimento del suo lavoro. L'elaborato della tesi è sicuramente collocabile al livello di ottimo.

Il responsabile Scientifico
Prof. Pier Giorgio Rigetti

Visto
Il Coordinatore del Dottorato di Ricerca
in Biotecnologie Industriali ed Ambientali
Prof. Hugo L. Monaco
Verona, 28 Febbraio 2007

TABLE OF CONTENTS

RIASSUNTO

Il lavoro di ricerca svolto nell'ambito del corso di dottorato in Biotecnologie Industriali ed Ambientali è stato effettuato in parte nel Laboratorio di Metodologie Biochimiche del Dipartimento Scientifico e Tecnologico dell'Università di Verona, e in parte nel Laboratorio di Proteomica del Dipartimento di Chimica, Materiali e Ingegneria Chimica "Giulio Natta" del Politecnico di Milano, sotto la guida e supervisione del Prof. Pier Giorgio Righetti. Sono state inoltre instaurate alcune collaborazioni occasionali o continuative con altre strutture universitarie o con aziende private.

In particolare, il lavoro di caratterizzazione biochimica e proteomica del ceppo A29 di *Pseudomonas stutzeri*, precedentemente impiegato in un peculiare caso di biorestauro operato su un antico affresco del Camposanto di Piazza dei Miracoli di Pisa, è stato condotto in collaborazione con il laboratorio di microbiologia della Prof. Sorlini (Dipartimento di Scienze degli Alimenti e Microbiologia, Università di Milano, Milano) e con il gruppo del Prof. Ranalli (Dipartimento di Scienze Agro-alimentari, Ambientali e Microbiologiche, Università del Molise).

Per quanto riguarda invece le indagini microbiologiche, biochimiche e proteomiche condotte sul ceppo selenito-riduttore SeITE02 di *Stenotrophomonas maltophila*, il lavoro è stato svolto in collaborazione con il gruppo del Prof. Vallini (Laboratorio di Microbiologia, Dipartimento Scientifico e Tecnologico, Università di Verona).

In aggiunta al lavoro di ricerca realizzato in ambito microbiologico, durante i tre anni del corso di dottorato sono stati trattati anche alcuni aspetti metodologici della tecnica principe dell'analisi proteomica: l'elettroforesi bidimensionale. A tal proposito, si è tentato di sviluppare nuove metodiche al fine di risolvere due delle maggiori limitazioni di tale tecnica: il ristretto intervallo di concentrazioni proteiche compatibili con i sistemi di rivelazione classici (tre-quattro ordini di grandezza), e l'impossibilità di effettuare separazioni soddisfacenti con masse molecolari inferiori a 10 000 Dalton. Relativamente al primo problema, la recentissima tecnologia delle Sferette Equalizzatrici è stata applicata allo studio dei contaminanti proteici contenuti in preparati teoricamente puri, dimostrando in tal

modo la capacità di detta tecnica di portare specie proteiche poco abbondanti a livelli di concentrazioni compatibili con la metodica bidimensionale. Per tale studio si è instaurata una collaborazione con Ciphergen Biosystems (Cergy-Pontoise, Francia), che ha eseguito il trattamento dei campioni e le analisi SELDI-MS.

Infine, per quanto concerne l'inadeguatezza della tecnica bidimensionale alla separazione di miscele di proteine con bassa massa molecolare, in collaborazione con Cleardirection Ltd. (Rehovot 76702, Israele) è stato sviluppata una nuova tecnica elettroforetica SDS-PAGE, in cui la separazione proteica avviene su gradienti di carica immobilizzati in matrici poliacrilammidiche non setaccianti. Si ritiene che tale tecnica sia in grado di risolvere catene polipeptidiche estremamente piccole, con masse fino a 200 Da.

Oltre alle sopraccitate collaborazioni "tematiche", il gruppo del Prof. Zolla (Laboratorio di Proteomica, Dipartimento di Scienze Ambientali, Università della Tuscia, Viterbo) ha contribuito in modo determinante alla produzione di tutti i dati raccolti in questa tesi, eseguendo le analisi di spettrometria di massa volte ad identificare le proteine d'interesse individuate nell'ambito dei vari studi.

La presente tesi è state articolata in due sezioni tematiche sulla base dei lavori sopraccitati e qui di seguito brevemente descritti: una prima sezione di ricerca classica, in cui si inseriscono i lavori di caratterizzazione dei ceppi microbiologici utilizzabili in intereventi di bio-restauro o di bio-decontaminazione di terreni inquinati, e una seconda sezione più spiccatamente metodologica, cui appartiene il lavoro di messa a punto della nuova tecnica di separazione elettroforetica su gradienti di carica immobilizzati e quello di arricchimento in specie proteiche rare eseguito tramite Sferette Equalizzatrici.

La caratterizzazione biochimica e proteomica del ceppo A29 di *P. stutzeri* ha prodotto importanti informazioni relativamente ai meccanismi proteolitici impiegati dal batterio durante la rimozione dello strato colloso a base proteica che oscurava l'intera superficie dell'affresco, e che risultava essere eccezionalmente resistente alle tradizionali tecniche di restauro. In particolare, si è osservato che il microrganismo, posto a contatto con il substrato colloso, sintetizza e secerne un pannello di enzimi proteolitici con masse molecolari comprese nell'intervallo 20 – 120 kDa, capaci di degradare efficientemente sia la caseina che il collagene, componenti principali della crosta collosa che si intendeva rimuovere. Tale

miscela di enzimi è risultata però essere completamente assente quando al batterio viene fornito come substrato carbonioso della colla fresca. Ciò dimostra i notevoli vantaggi conferiti dall'utilizzo di cellule vitali anziché di preparati enzimatici negli interventi di biorestauro. Le cellule, infatti, grazie alla riconosciuta capacità di adattamento alle condizioni ambientali più disparate, sono in grado di sintetizzare e secernere miscele enzimatiche particolarmente efficaci nel degradare lo specifico substrato carbonioso presente.

Anche nel caso del ceppo SeITE02 di *S. maltophila* le indagini eseguite a livello microbiologico, biochimico e proteomico hanno portato ad un considerevole progresso nella conoscenza dei meccanismi molecolari di riduzione dell'ossianione selenito. In particolare, si è visto che il sistema di detossificazione impiegato dal batterio durante la riduzione non dissimilativa del selenito coinvolge sia la via metabolica di assorbimento (ma non di riduzione) del nitrito, sia il sistema antiossidante del glutatione. A tal riguardo sono state raccolte diverse evidenze sperimentali. Per quanto concerne l'interazione fra le vie metaboliche di nitrito e selenito, si è potuto osservare che, in colture microbiche contenenti entrambi gli ossianioni, la riduzione del selenito risulta fortemente inibita; ciononostante non si è potuta rilevare alcuna attività enzimatica di riduzione del nitrito, e il medesimo effetto inibitorio lo si è ottenuto incubando i batteri in presenza di selenito e di un inibitore dell'assorbimento del nitrito. Relativamente al glutatione sono state poi raccolte diverse prove a sostegno della teoria che vede in tale composto uno dei protagonisti principali del processo di riduzione: (i) se utilizzato in saggi di riduzione enzimatica in vitro (con estratti citoplasmatici batterici), il glutatione aumenta enormemente le cinetiche di riduzione; (ii) colture batteriche cresciute in presenza di un inibitore della sintesi del glutatione sono incapaci di ridurre il selenito durante la fase di crescita esponenziale, mentre normalmente la riduzione avviene in modo simultaneo alla crescita; (iii) i dati di proteomica evidenziano che due enzimi coinvolti nella sintesi del glutatione sono significativamente indotti in presenza di selenito. Nel complesso, i dati raccolti hanno permesso di approfondire le conoscenze di alcuni importanti meccanismi molecolari coinvolti nella riduzione del selenito in condizioni aerobiche, condizioni assai frequenti negli interventi di bonifica di terreni inquinati.

Passando invece alla seconda parte del progetto, di carattere più spiccatamente metodologico, il trattamento di preparati proteici puri con le sferette equalizzatrici ha permesso di ottenere importanti informazioni circa l'effettivo grado di purezza di tali campioni, la natura dei contaminanti proteici, l'effettiva capacità equalizzatrice delle sfere e la loro applicabilità a processi di "polishing" nell'ambito di purificazioni su larga scala. I primi esperimenti pilota eseguiti su campioni artificialmente contaminati con miscele proteiche complesse hanno evidenziato la sorprendente capacità "equalizzatrice" di questa tecnologia, producendo una miscela di proteine con un intervallo di concentrazioni estremamente ristretto a partire da un campione costituito per il 95% da un'unica specie proteica (mioglobina). Lo studio si è poi sviluppato attraverso l'analisi di campioni reali, eseguendo l'equalizzazione di preparati farmacologici apparentemente puri e analizzando poi le varie specie proteiche apparse in seguito al trattamento. Tale analisi ha evidenziato che la composizione del mezzo di crescita influisce notevolmente sulla purezza dell'estratto finale, e che buona parte delle impurezze purificate assieme al prodotto d'interesse sono sia proteine dell'organismo ospite (nel caso di proteine ricombinanti), sia prodotti di degradazione o di aggregazione della prodotto stesso. Tali risultati hanno messo in evidenza come un'inadeguata fase di polishing possa portare ad un prodotto finale che, se inteso per il consumo umano, potrebbe costituire un considerevole rischio per la salute (presenza di potenziali allergeni). A tal riguardo, la vasta specificità di legame mostrata dalle librerie esapeptidiche immobilizzate sulle sferette equalizzatrici sembrerebbe costituire un valido strumento per la rimozione simultanea di tutte le specie presenti in minime quantità, essendo la loro scarsa abbondanza non un ostacolo, come solitamente accade con le tradizionali tecniche di purificazione, ma un requisito, poiché un'abbondanza relativa particolarmente elevata porterebbe alla saturazione dei siti di legame, con conseguente peggioramento nelle prestazioni della metodica.

Da ultimo è stato affrontato uno dei maggiori problemi della tecnica bidimensionale classica, ovvero l'incapacità di separare miscele di proteine di ridotta massa molecolare. Com'è ben noto, infatti, la metodica SDS-PAGE è in grado di trattare in modo adeguato masse molecolari superiori a 10 kDa o, nel caso del sistema Tris-Tricina, superiori a 3 kDa. Quest'ultimo rappresenta tuttavia

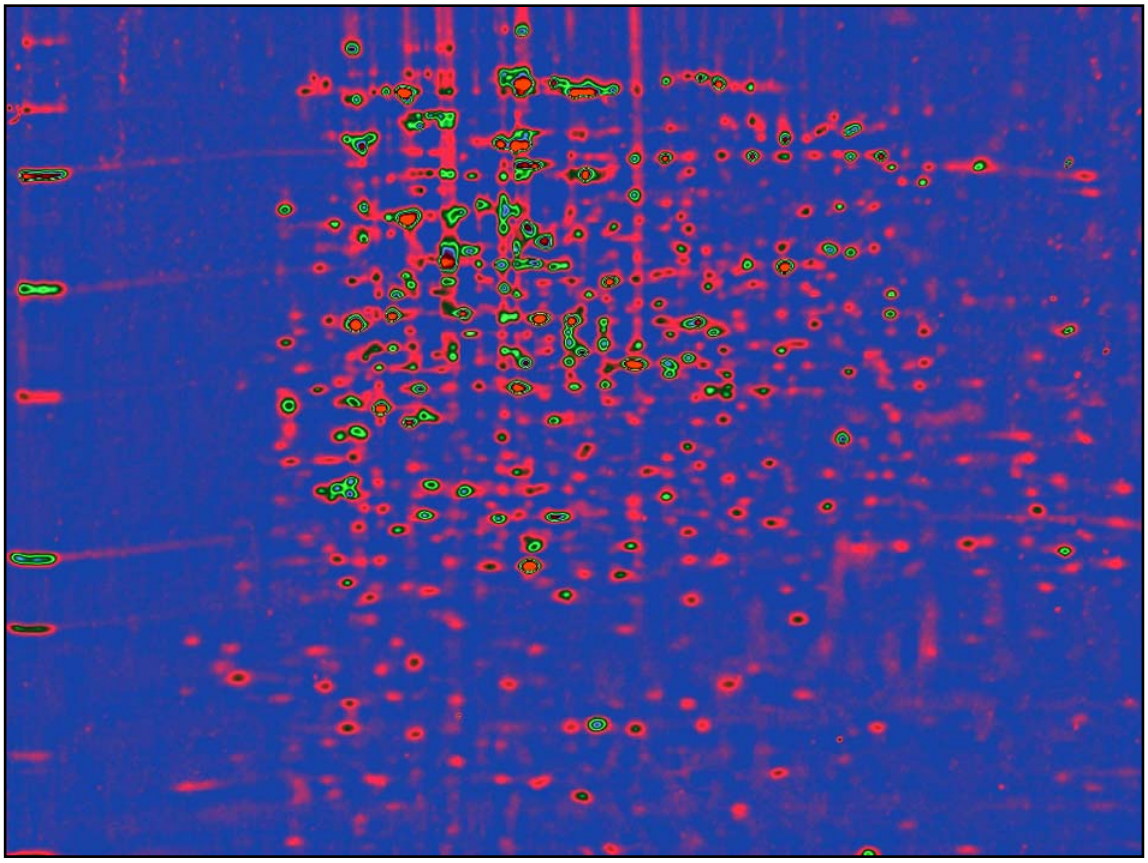
il limite estremo della metodica, che pertanto risulta totalmente inadeguata allo studio di piccole molecole e peptidi, i quali però, secondo le teorie più recenti, potrebbero costituire marcatori patologici o tumorali di estremo interesse. A tal proposito è stata sviluppata una variante della tradizionale tecnica SDS-PAGE, capace di separare anche le proteine più piccole grazie ad un sistema di neutralizzazione delle cariche presenti sul complesso polipeptide-SDS. Tale neutralizzazione avviene per mezzo di un gradiente di cariche positive immobilizzato sulla matrice poliacrilammidica in cui avviene la separazione. Tali cariche vengono fissate sulla matrice grazie all'impiego di opportune Immobiline basiche in fase di polimerizzazione del gel. La proteina, muovendosi lungo il gradiente di carica preformato, raggiunge un punto in cui la densità di carica esterna eguaglia quella presente sulla superficie del complesso stesso, dando luogo ad un fenomeno di neutralizzazione che impedisce ogni ulteriore migrazione. Gli studi condotti hanno rivelato che questa nuova tecnica, da noi chiamata "SDS-PAGE focusing" in virtù dello stato stazionario cui le proteine pervengono in fase di separazione, presenta un potere risolutivo eccezionalmente elevato, che le consente di separare in modo soddisfacente differenze di masse molecolari nell'ordine dei 150 Da. Per quanto riguarda lo studio dei piccoli polipeptidi, la tecnica "SDS-PAGE focusing" si è dimostrata particolarmente valida, originando profili elettroforetici di tre diversi digeriti triptici (lisozima, mioglobina e avidina) decisamente migliori rispetto a quelli ottenuti tramite un tradizionale SDS-PAGE su gradiente di porosità. Concludendo, la nuova tecnica qui proposta rappresenta una valida alternativa al tradizionale SDS-PAGE nel caso in cui si debbano studiare masse molecolari estremamente ridotte. Inoltre tale metodica potrebbe trovare ampio impiego anche nell'analisi di polianioni naturali quali gli acidi nucleici, i quali, posti a migrare all'interno di un gradiente di cariche positive immobilizzate, dovrebbero comportarsi in modo analogo ai complessi SDS-proteina, originando profili elettroforetici stabili nel tempo e ben risolti.

LIST OF ABBREVIATIONS

2-DE:	Two Dimensional Electrophoresis
ACN:	acetonitrile
APS:	ammonium persulfate
BSA:	bovine serum albumin
BSO:	S-n-butyl homocysteine sulfoximine
CA:	Carrier Ampholytes
CFU:	Colony-Forming Units
CHAPS:	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
Da:	Dalton
DIGE:	Differential in Gel Electrophoresis
DM A:	Defined Minimal medium modified A
DM B:	Defined Minimal medium modified B
DNA:	Deoxyribonucleic Acid
DNP:	2,4-dinitrophenol
DTE:	dithioerythritol
DTT:	dithiothreitol
EB:	Equalizer Bead
EDTA:	ethylenediamine-tetraacetic acid
FA:	formic acid
FFE:	Free-Flow Electrophoresis
fmol:	femto-mole
FTICR-MS:	Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry
GR:	Glutathione Reductase
GSH:	glutathione
GS-Se-SG:	selenodiglutathione
GST:	Glutathione S-transferase
HCIC:	Hydrophobic Charge Induction Chromatography
HCP:	Host Cell Protein
HIC:	Hydrophobic Interaction Chromatography

HSA:	Human Serum Albumin
ICAT:	Isotope Coded Affinity Tags
IEC:	Ion Exchange Chromatography
IEF:	Isoelectric Focusing
IgG:	Immunoglobulin G
IPG:	Immobilized pH Gradient
LC:	Liquid Chromatography
$LD_{50:}$	Lethal Dose 50
LDH:	Lactate Dehydrogenase
m/z:	Mass over Charge
Mab:	Monoclonal Antibodies
MALDI:	matrix assisted laser desorption ionization
MCE:	Multicompartment Electrolyzers
MEP:	4-mercaptoethyl-pyridine
MIC:	Minimum Inhibitory Concentration
Mr:	Relative Mass
MS/MS:	Tandem Mass Spectrometry
MS:	Mass Spectrometry
Mud-LC:	Multidimensional Liquid Chromatography
NADH:	Nicotinamide Adenine Dinucleotide
NCBInr:	National Center for Biotechnology Information non-redundant
NEDA:	N-(1-Naphthol)-ethylendiamine dihydrochloride
NIR:	Nitrite Reductase
PAGE:	Polyacrylamide Gel Electrophoresis
PBS:	Phosphate Buffer Saline
pI:	Isoelectric Point
PMF:	Peptide Mass Fingerprint
ppm:	part per million
PSA:	Prostate Serum Antigen
PTM:	Post Translational Modification
rDNA:	recombinant DNA
ROS:	Reactive Oxygen Species
SARS:	Severe Acute Respiratory Syndrome

SDS:	sodium dodecyl sulphate
Se^{0} (Se):	elemental selenium
Se^{-II} (H_2Se):	selenide
Se^{IV} (SeO_3^{2-}):	selenite
SELDI:	Surface-Enhanced Laser Desorption Ionization
Se^{VI} (SeO_4^{2-}):	Selenate
SILAC:	Stable Isotopic Labelling by Amino Acids in Cell Culture
SSP:	Standard Spot Protein
TAP:	Tandem Affinity Purification
TBP:	ributylphosphine
TCA:	trichloroacetic acid
TEMED:	*N*,*N*,*N*',*N*'-tetramethylethylenediamine
TFA:	trifluoroacetic acid
TOF:	Time of Flight
TR:	Thioredoxin Reductase
Vh:	Volt x hours
YE:	Yeast Extract



CHAPTER I:

PROTEOME AND PROTEOMICS TODAY

1.1. GENERAL INTRODUCTION

The terms “Proteome” and “Proteomics” were used for the first time by Wilkins [1] in 1995, to indicate, respectively, the protein equivalent of an organism genome and the scientific discipline deputed to its study. Proteomics represents today one of the most important area of research in the “post-genomic era”, so called since the scientific community realized that genomics represents only the first step in understanding the complex biological functions performed by simple or complex living organisms. This is due to the fact that the relationship between genes and proteins is quite complex, since lots of intermediate biological steps must take place in order to convert the inactive information stored in DNA sequence into biologically active products, namely proteins. In fact a gene can give rise to many mature transcripts because of alternative splicing, and each mature mRNA can originate a multitude of different final protein products because of post-translational modifications (PTMs). Thus, neither genomics nor transcriptomics can give a satisfying picture of any biological system, being the number of proteins expressed in a cell much larger than the coding potential of an organism [2]. Thus, for example, genomics cannot explain the enormous gap existing between humans and worms, since human genome sequencing has revealed that these two organisms possess barely the same number of genes (i.e. about 23 000, [3]). Therefore such diversity must rather depend on proteins, and since proteins are the unique actual actors responsible for any biological function taking place in a cell, it is highly preferable, if not mandatory, to focus the attention directly on them, by studying expression profiles, protein cellular localization and structure, PTMs and functional networks. The proteome is an extremely dynamic entity, reflecting the functional status of a cell, or of a biological compartment (organ, tissue, cell or subcellular compartments), at a particular time and under particular environmental conditions. Hence, since the first definition of proteomics given by Wilkins does not take into account such highly dynamic nature of the proteome, it should be replaced by the more recent one proposed by de Hoog and Mann in 2004, who stated that proteome is the protein complement of a given cell at a given time, including the set of all protein isoforms and modifications [4]. Hence proteomics can be defined as the large-scale study of protein expression,

structure and function, whose branches include protein separation, identification and quantification, protein sequence analysis, structural proteomics, interaction proteomics and protein modification [5].

1.2. PROTEOMICS: APPLICATIONS AND OBJECTIVES

Currently proteomic studies focus on two major research areas: expression proteomics, which aims at evaluating protein expression levels, and functional proteomics, which studies protein activities, multi-protein complexes and signalling pathways.

1.2.1. EXPRESSION PROTEOMICS

A typical application of expression proteomics is the comparative analysis of protein expression patterns in normal and tumour cells. In biomedical applications, this approach is usually employed for identifying proteins that are up- or down-regulated in a disease specific manner, for use as diagnostic markers or therapeutic targets.

1.2.1.1. Diagnosis

Nowadays disease markers for early diagnosis consist mainly in individual proteins, which are not always reliable if considered independently from other factors. For example, prostate serum antigen (PSA), used to screen for prostate cancer, sometimes results to be over-expressed also under benign conditions. On the other hand, proteomics allows for simultaneous analysis of thousands of proteins, giving a complete picture of the functional status of a biological system, thus providing a useful and powerful tool to discriminate between healthy and pathological conditions.

Because of the ability of proteomics to identify new tumour markers, in the last years there has been great interest in applying proteomics to the study of tumours. Among these, ovarian cancer rose great interest, since it is often diagnosed only at an advanced stage of the disease (5-years survival rate of 35%,

[6]). However, a proteomics studied has identified a serum protein patterns that distinguish pathologic from healthy conditions with accuracy of 94% at the first stage of the disease, when the 5-year survival rate is 90% [7].

Early diagnosis could be useful also with pathologies different from tumours, such as infectious diseases. Availability of screening test usable during pre-clinical infection would allow early treatment of potentially lethal disease, reducing the risk of transmission and the disease intensity. Thus, for example, in sera from patients affected by severe acute respiratory syndrome (SARS), a truncated form of α1-antitrypsin has been found at significantly higher concentration with respect to controls, therefore constituting a potential diagnostic marker or even a therapeutic target [8].

1.2.1.2. Therapeutics

Proteomics doesn't constitute only a valuable tool for disease marker discovery. For instance, lots of the studies carried out in the last years have concerned the molecular basis of tumour development or of cell resistance to drugs. In particular, chemoresistance phenomena in cancer cells can depend on very complex molecular mechanisms, whose understanding would improve the existing therapies or lead to new treatment strategies. Verrills *et al.* [9] performed a comparative study of drug-resistant and drug-sensitive lymphoblastic leukaemia cell lines, discovering a connection between cytoskeletal proteins (or proteins which bind to cytoskeleton) and the resistant phenotype, thus pointing out some potential new drug targets for treating resistant disease.

Drug resistance is also a major clinical problem in the treatment of many infectious diseases, and, in many cases, the mechanism is unknown. Genetic and protein-sequence data are now available for many microorganisms, thus providing a tool for understanding their resistance to drugs and for identifying novel therapeutic agents. As an example, chloroquine is one of the most exploited drugs to treat malaria, but now in many parts of the world it cannot be used any longer because of widespread chloroquine resistance. In this concern, proteomics technologies are playing a major role in identifying potential therapeutic targets in *Plasmodium* species, as well as host–pathogen interactions and protein–drug interactions. Up to date, considerable advances have been made in identifying

immune targets for vaccination and immune protection, and in better understanding the mechanisms of chloroquine resistance [10].

1.2.2. FUNCTIONAL PROTEOMICS

Functional proteomics has two major objectives: to investigate the biological function of unknown proteins and to elucidate cellular mechanisms at a molecular level. In fact, the various genome sequence projects carried out in the last years have led to an impressive amount of genes and of predicted protein sequences whose function is still totally obscure. One of the most widespread approaches used to attribute a function to such unknown proteins consists in isolating and characterizing the multi-subunit protein complexes. In fact, in the cells many proteins exert their biological functions through rapid and transient association within large protein complexes [11]. Therefore, in order to understand the protein functions and the molecular mechanisms it is necessary to proceed with the isolation and the identification of the protein partners present in multi-subunit protein complexes.

1.2.2.1. Identification of protein partners: the "Fishing for Partner Strategy"

A key contribution to the identification of proteins interacting within stable complexes in cellular systems is provided by the affinity-based approaches. The idea is to express the protein of interest fused with a suitable tag to be used as a bait to fish its specific partners out from a cellular extract. With this approach, the isolation of the entire multi-protein complex can be accomplished by using several anti-tag systems immobilised onto agarose or Sepharose supports [12].

In the so called "Fishing for Partners Strategy", commercially available protein expression systems are used to synthesise a specific protein bait as a hybrid protein fused with different tags, such as Glutathione S-transferase (GST-fused Protein), or small peptide epitopes, containing, for example, a poly-His tail or covalently modified with biotin [13]. The tagged bait can be immobilised onto agarose beads derivatised with the appropriate anti-tag ligand (glutathione, anti-epitope antibodies, nickel ions, streptavidin, etc.). The entire cellular extract or the extract from a specific organelle can then be incubated with the immobilised bait.

The protein components specifically recognised by the bait are retained, while the unbound proteins are removed by washing. Although this approach has found large application in the study of protein-protein interactions, it suffers from some drawbacks, in particular the excessive background caused by the scarce interaction specificity between tag and anti-tag molecules.

In order to overcome this problem, highly specific antibody-antigen recognition can be used to purify protein complexes by the so-called “immunoprecipitation” method. In this case, proteins of interest are synthesised fused to the desired antigen. Once protein complexes have formed, they are immunoprecipitated by anti-tag monoclonal antibodies.

Further improvements in terms of purity of the isolated complex came with the “tandem affinity purification (TAP) tag system”, which allows for high yield protein purification under native conditions [14]. With this procedure two different tags, usually separated by an enzyme-cleavable linker, are inserted on the desired proteins, and sequentially used to purify the proteins and its partners. The key feature of this approach consists in coupling two highly specific recognition systems to cleavable linkers that allow for easy protein recovery in fully native conditions. According to these features, the first TAP tag system consisted of a *Staphylococcus* protein A domain and a calmodulin binding domain, physically connected by a protease cleavable linker. In this way, protein recovery after the first purification step can be easily achieved by enzyme cleavage, while Ca^{2+} sequestrating agents (i.e. ethylenediamine-tetraacetic acid, EDTA) allows for final protein purification [15].

1.2.2.2. Post-translational modification (PTMs) analysis

PTMs introduce great complexity into the proteome, generating from a single protein dozens of different species due to phosporylation, methylation, lipidation, glycosylation and many others. PTM study is a rapidly expanding field, because protein activity and function strictly depend on the PTMs they bear.

Phosphorylation is one of the most studied PTMs, since enzyme activity is frequently regulated by adding or removing one or more phosphate groups on tyrosin, threonin or serine residues. Classical methods for studying the phosphoproteome include metabolic radiolabeling with ^{32}P [16] and affinity

chromatography with either metal affinity chromatography [17] or phospho-specific antibodies [18]. Moreover, recently, chemical tagging strategies have been developed for quantitative phosphoproteome analysis, based on the removal of phosphate groups from amino acid lateral chains followed by the addition of particular groups which can be specifically recognized and captured on a solid phase [19, 20].

1.3. ANALYTICAL METHODS FOR EXPRESSION PROTEOMICS

1.3.1. GEL-BASED AND GEL-FREE METHODS: AN OVERVIEW

The classical expression proteomics approach consists first in a protein separation step, most frequently two-dimensional electrophoresis (2-DE) or, sometimes, two-dimensional chromatography, followed by protein detection (general or specific methods) and identification, the latter usually performed by mass spectrometry (MS).

In 2-DE proteins are separated sequentially on the basis of two independent chemico-physical features, usually electric charge (isoelectric focusing, IEF) and molecular mass (sodium dodecyl sulphate electrophoresis, SDS-PAGE). IEF based on immobilized pH-gradients (IPG) coupled to SDS-PAGE and followed by protein identification through MS is currently the workhorse for proteomics [21], permitting immediate quantitative expression profiling of large sets of complex protein mixtures isolated from the most different samples. After separation and detection, proteins are usually identified by in-gel digestion with trypsin followed by peptide mass fingerprinting (PMF) or tandem mass spectrometry (MS/MS).

As for any other analytical technique, there are some limitations to the universal application of 2-DE to the analysis of biological samples. Eukaryotic cells are likely to express, at a given time and under certain conditions, up to 10^3 different protein species, the majority of which are present at extremely low concentrations (10^2 molecules per cell) as compared to the few housekeeping proteins (up to 10^6 copies per cell) usually detected in proteomics studies. Besides the small detection sensitivity range, 2-DE suffers from poor solubility of membrane

proteins, limited reproducibility of gels, relatively low throughput and it is considerably labour demanding. However, in spite of these limitations, 2-DE is currently the most rapid and effective method for the extensive study of protein expression profiles in extremely different samples. In fact also the chromatographic methods present some relevant drawbacks that limit their use in proteomics studies.

1.3.1.1. Multi-dimensional liquid chromatography coupled to mass spectrometry

In the so-called multidimensional (but actually two dimensional) nano- or capillary liquid chromatography (Mud-LC), protein digestion takes place in solution prior to any separation step, after that particular peptides are selectively enriched from the whole digest and subsequently separated by LC or two-dimensional LC coupled to electrospray ionization (ESI)-MS/MS. Figure 1.1 graphically summarizes the difference between 2-DE and two dimensional LC-based proteomics. The fact that, with the chromatographic approach, protein detection is carried out by online ESI-MS/MS has led to the development of MS-based techniques for quantification of the separated peptides. They are based on the differential chemical labelling of the protein samples under comparison with, for example, different stable isotopes ($^{12}C/^{13}C$, $^{14}N/^{15}N$, $^{1}H/^{2}H$). Such an approach allows the simultaneous analysis of the two samples in a unique chromatographic run, since coeluting differentially labelled peptides can be separated and quantified during detection by MS.

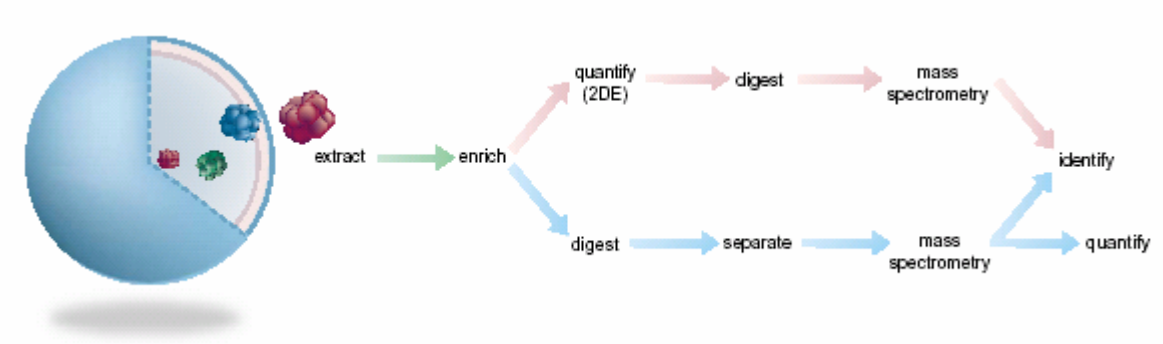


Figure 1.1. Schematic representation of proteomics investigation based on 2-DE (upper path) or two-dimensional chromatography (lower path).

With the metabolic isotopic labelling (stable isotopic labelling by amino acids in cell culture, SILAC), stable isotopes are introduced during cell growth by introducing in culture media isotope-labelled proteins or amino acids. However such an approach, based on the biological incorporation of labelled amino acids, is applicable only to cell cultures, since only proteins extracted from fast growing cells can be analyzed. On the other hand, chemical incorporation of labelled tags after protein extraction is universally usable. With this approach, named “isotope coded affinity tags” (ICAT) by Aebersold group [22], after protein extraction and disulphide bonds reduction cysteines are labelled with “light” and “heavy” ICAT reagents. Subsequently samples are mixed and the proteins are digested, the labelled peptides are isolated, by affinity capture on an avidin column, and then fractionated by Mud-LC and analyzed by MS/MS. Relative quantification is eventually achieved by measuring the peak area of each eluted peptide. ICAT-based quantitative analysis was successfully used for total proteomes comparative studies, as in the case of *Pseudomonas aeruginosa* cultured under normal conditions and under conditions that induce expression of virulence factors [23], or the case of salt stress induced proteome changes in yeast [24].
Despite the enormous potential of this technology, it is far from being able to draw a comprehensive picture of a whole proteome. The most important disadvantage of the ICAT method is that proteins lacking cysteines cannot be studied.

1.3.1.2. The SELDI-TOF platform

Another very recent platform for global proteome analysis is the so called surface enhanced laser desorption ionization-time of flight (SELDI-TOF) protein profiling. Here, either whole or pre-fractionated protein extracts are loaded onto arrays equipped with surface chemistries (e.g. ion exchange, hydrophobic interaction, metal chelate and others) or known biological ligands (e.g. antibodies), which selectively retain a fraction of the loaded proteins that are subsequently analyzed by MS [25, 26]. The functioning principle is very similar to the common adsorption-desorption mechanism of column chromatography, but here the solid phase constitutes a flat surface where a drop of protein solution is laid. While the unbound proteins are removed by repeated washes, proteins specifically interacting with the chromatographic surface are retained and finally covered with

a matrix solution which enhances protein vaporization and ionization during laser-induced desorption. Thus, differently from common chromatography, where only eluted proteins can be recovered and studied, here only retained proteins are finally analyzed. Figure 1.2 displays a schematic representation of the surface chemistries and of the process leading to MS spectrum.

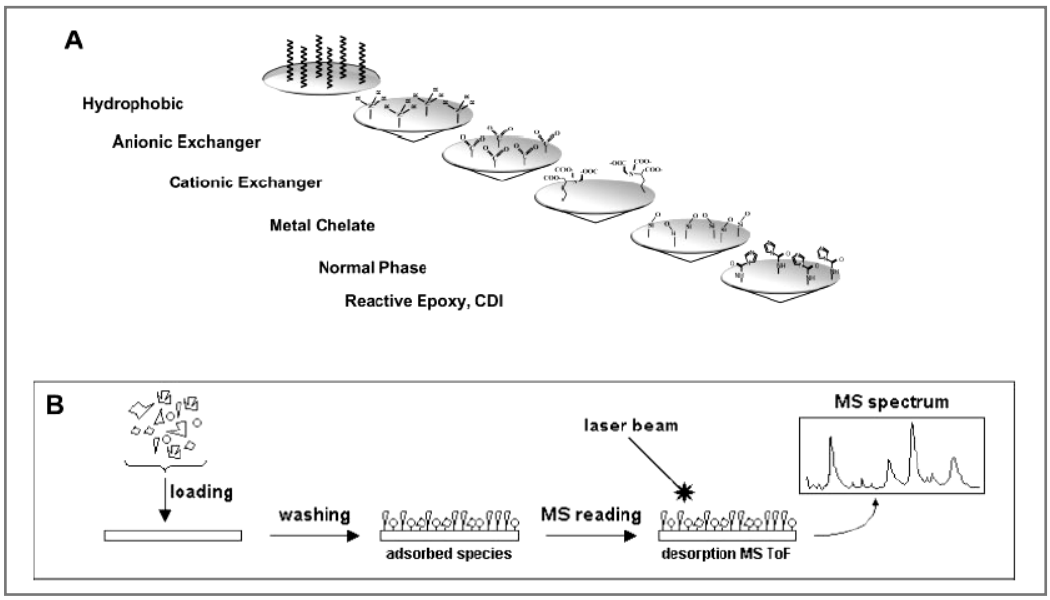


Figure 1.2. (A) Schematic representation of ProteinChip Arrays with chemical functions responsible for protein capture. (B) Mechanism of adsorption-desorption and ionization of proteins prior to entering a mass spectrometer analyzer.

This technique works extremely well with peptides and small proteins (up to 20 kDa), allowing for comparative analysis of protein expression profiles in the femto-mol (fmol) range, although with the same limitations of dynamic range affecting the other technologies. Various biological samples such as urine, cell lysates and cell secretion have been profiled using SELDI-TOF MS, and serum proteome profiling is rising great interest in the field of cancer diagnostic. Different types of cancer have been examined with this technique, including ovarian, prostate, breast, bladder and renal [27-31], and in general SELDI-TOF profiling has proved to be more reliable in terms of sensitivity and specificity as compared to classical diagnostic biomarkers [32].

1.4. TWO DIMENSIONAL ELECTROPHORESIS: BASIC METHODS

Currently 2-DE constitutes the tool of election for proteome investigations, because of the numerous advantages offered over other techniques.
First it can be employed to obtain, with a fairly standardized protocol, a panoramic picture of a complete proteome starting from almost any kind of biological sample. Thousands of proteins can be displayed simultaneously on a single 2D map, and for each protein feature the apparent molecular mass and the isoelectric point can be easily calculated from the spot position. Moreover, proteins visualized on a 2D gel are usually present in such an amount that they can be recovered for further analysis, like identification via PMF.
Another important advantage of 2-DE over chromatographic techniques consists in the fact that polypeptides are analyzed in their full-length form, without need for digestion prior to analysis, thus allowing for separation, detection and characterization of the single PTMs. Such functional studies find valuable help in western blot techniques performed after two-dimensional separation, which allows for detection of specific protein (protein specific antibodies) or specific PTMs (chemical group specific antibodies). Moreover, some hints about possible PTMs are also provided by 2-D gels stained with universal staining methods, since some PTMs usually lead to peculiar electrophoretic protein profiles (i.e. phosporylation generates horizontal spot strings).
Finally, despite being highly labour demanding, the 2-DE platform is not among the most expensive, since complete equipment can be purchased for some thousands euros.
Proteomics investigation by 2-DE involves specific sequential steps:

1. Sample preparation
2. Two dimensional electrophoresis (IEF/SDS-PAGE)
3. Protein detection and quantification
4. Computer-based 2-D image analysis
5. Further protein analysis (i.e., PMF-based identification, PTMs analysis and the like)

A typical 2-DE workflow is schematically represented in figure 1.3.

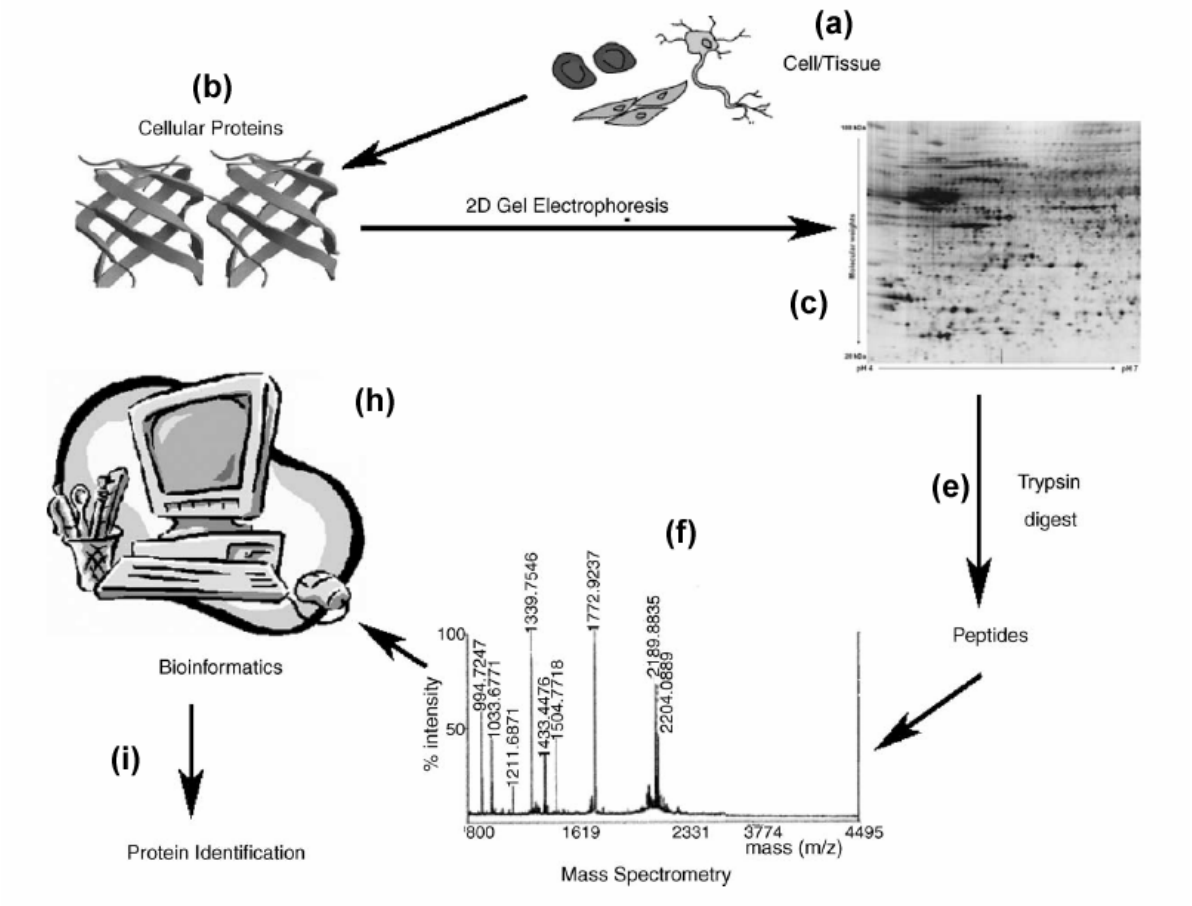


Figure 1.3. Workflow of a typical 2-DE experiment: sample collection (a) and protein extraction (b); 2-DE separation of proteins (c); image analysis followed by spots cutting and in-gel trypsin digestion (e); PMF analysis or MS/MS sequencing (f) followed by database search (h) for protein identification (i).

1.4.1. SAMPLE PREPARATION

This is maybe the most important step in a 2-DE experiment, because either the quality of the 2-D map or the specific set of proteins that are finally displayed on it depend on this phase. To take advantage of the high resolution of 2-DE, proteins have to be fully solubilised by complete disruption of molecular interactions, and this is achieved through extensive protein denaturation followed by reduction and alkylation of disulphide bonds. Sample preparation should be as simple as possible to maximize reproducibility and avoid excessive sample loss. Essentially, sample preparation can be subdivided into three or four fundamental steps: cell or

matrix disruption, protein solubilisation, removal of interfering substances and, eventually, protein prefractionation.

1.4.1.1. Cell disruption

Cell disruption can be achieved by harsh or gentle methods, depending on the fact that proteins have to be recovered in native or denatured forms. Gentle methods consist in osmotic lysis, freeze-thaw cycling, detergents lysis or enzymatic lysis of the cell wall. Harsh methods are usually required when robust cell walls are involved (i.e. bacteria and plant tissues), and are sonication, grinding with liquid nitrogen, high pressure (e.g. French press) homogenization with glass beads or rotating blade homogenizer.

1.4.1.2. Protein solubilisation

This step usually takes place in an alkaline, strongly denaturant buffer containing chaotropes (usually urea and thiourea), non-ionic or zwitterionic detergents (e.g. Triton X-100 or detergent sulphobetaines), reducing agents, carrier ampholytes (CA) and, possibly, protease inhibitors. The original solubilisation buffer proposed by O'Farrel [33] in 1975 was 9 M urea, 2-4% 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 1% dithiothreitol (DTT) and 2% CA, but this recipe has been then ameliorated in order to maximize membrane proteins solubilisation. Today, the most used solubilisation buffer contains 7 M urea, 2 M thiourea, 2-4% CHAPS, 40 mM Tris, 5 mM tributylphosphine (TBP) and 1% CA [34].

Chaotropes. Urea is quite effective in disrupting hydrogen bonds, causing protein unfolding and denaturation, while thiourea breaks hydrophobic interactions, thus facilitating solubilisation of highly hydrophobic proteins.

Detergents. Detergents are utilized for solubilising hydrophobic cores of proteins. Neutral (Triton X-100, Nonidet P40, β-dodecyl maltoside and others) or zwitterionic (CHAPS or other sulphobetains) detergents have to be used in order not to interfere with the IEF process.

Reduction/alkylation. Disulphide bonds reduction promotes protein unfolding and solubilisation. The better reducing agent seems to be TBP [35], which allows for complete –S-S- reduction at low concentrations (3-5 mM) without interfering with

thiourea, which reacts promptly with other commonly used reducing agents (e.g. DTT or dithioerythritol, DTE). Moreover, once reduced, cysteines should be blocked by alkylation prior to any further analysis, in order to prevent artefacts generation during 2-DE (i.e. intra- or inter-molecular disulphide bridge formation, [36]).

1.4.1.3. Removal of interfering substances

Compounds interfering with protein solubility, stability or with any of the separation steps must be removed from the sample. Complex proteomes usually contain also various classes of proteases that might cause extensive sample degradation. For that reason strongly denaturant alkaline buffer containing a protease inhibitor cocktail should be used for preventing protease activity.

Other interfering substances affecting 2-DE and, in particular, IEF, are salt ions, lipids, polysaccharides, polyphenols and nucleic acids. All of these substances should be carefully eliminated during sample preparation (for a review see [21]).

Salts, when present at high concentrations, keep conductivity high, thus slowing down the IEF process. In order to reduce salt concentration, either protein precipitation or dialysis can be exploited.

High amount of lipids might interact with membrane proteins and sequestrate considerable amounts of detergent, thus reducing protein solubility. Lipid removal can be effectively accomplished by extraction with organic solvents like cold acetone or a mixture of acetone, tributylphosphate and methanol [37].

Polysaccharides and nucleic acids, being large polyanionic molecules, can interact with proteins via ionic bonds, thus causing horizontal streaks in 2-D gels. Protein precipitation with trichloroacetic acid (TCA) or organic solvents is a useful practice.

Although many cleaning procedures have been already set up for various samples, nevertheless the researcher should always optimize the preparation protocol on the basis of the specific sample, in order to achieve the best results.

1.4.2. TWO DIMENSIONAL ELECTROPHORESIS (IEF/SDS-PAGE)

2-DE, as originally proposed by O'Farrell in 1975, is based on denaturing carrier ampholyte-based IEF (1^{st} dimension) coupled to SDS-PAGE (2^{nd} dimension). In spite of the great merit of this scientist, his technique suffered from a relevant drawback, namely the poor reproducibility caused by CA variability from batch to batch, cathodic drift and pH gradient instability over time [38]. Such difficulties have been largely overcome by the development of IPGs [39], based on the chemistry of Immobilines, which are modified acrylamide molecules with the general structure CH_2=CH-CO-NH-R, where –R contains either carboxyl or amino groups buffering at different pH values, thus generating pH gradients chemically grafted to the polyacrylamide matrix. IPGs allow for much greater reproducibility, good focusing of alkaline proteins and high resolution (Δ pI = 0.001). With IPGs, proteins are usually loaded by passive or active re-hydration of the dry strips with an appropriate volume of solubilisation buffer. Another possible way of loading proteins is the so called cup-loading, with which solubilised proteins are placed in a well sealed to the surface of the previously reswollen IPG and enter the gel in a specific point with the help of a low voltage.

Once loading has been completed, protein focusing takes place by applying a slow voltage gradient, until reaching 5 000 V (for 7 cm long strips) or 10 000 V (for 18 cm or longer strips) and a total Volts x hours (Vh) product of, respectively, 25 000 and 75 000. Besides the Vh product, another marker of focusing achievement is low and stable current under high voltage conditions.

After the focusing process, proteins present a neutral surface charge, thus a conditioning step with SDS is needed before further electrophoretic separation. The conditioning stage lasts for a strictly defined period of time (usually 25-30 minutes), sufficient for proper saturation of proteins with SDS but not so long to cause significant protein diffusion. Equilibration buffer usually contains 2% SDS, 6 M urea and 20% glycerol (to minimize the interaction between surface charges of gel and proteins and to improve protein transfer from the first to the second dimension) and Tris-HCl pH 8.8 (to match the conditions present in the gel of second dimension). Proper equilibration is fundamental for a good protein transfer and for minimizing protein loss (up to 20% of total proteins, [40]) due to excessive

protein adsorption to the gel matrix, insufficient equilibration time or wash-off effects.

Soon after protein equilibration with SDS the strips have to be interfaced with the second dimension gel and SDS-PAGE separation must be performed. A solution containing 0.5% agarose melt in the cathodic buffer used in the second dimension is commonly used for fixing the strip onto the gel surface. Once solidified, the agarose gel permits good protein transfer between the two gels.

SDS-PAGE separates proteins according to their molecular mass, using alternative buffer systems depending on the Mr range of the proteins under study. Usually a Tris-glycine buffer is chosen to achieve good separations of proteins with Mr values in the interval 10 - 150 kDa, although Schagger and von Jagow [41] proposed to substitute glycine with tricine to better separate molecular masses from 2 to 50 kDa. Another important aspect to consider when preparing SDS-PAGE gels is the acrylamide concentration. Soft gels with low acrylamide concentrations are suitable for high molecular masses separations, while harder gels with high acrylamide concentrations better separate small polypeptides. However in complex proteomes both low and high molecular mass proteins are usually present simultaneously, thus acrylamide gradients should be preferred over fixed concentrations. In fact an extended acrylamide gradient from 7 to 20%T guarantees proper resolution of proteins from 10 up to 200 kDa of molecular mass.

Recently, a collaboration between our group and Cleardirection Ltd. (4 Pekeris St., Rehovot 76702, Israel) has lead to the development of a new separation technique named SDS-PAGE focusing, which combines traditional SDS-PAGE with Immobiline chemistry with the aim of separating molecular masses down to a limit of ca. 500 Da, thus constituting a great improvement in the field of peptide analysis. Such a technique will be described later in this thesis.

1.4.3. PROTEIN DETECTION AND QUANTIFICATION

Protein detection procedures embrace both universal and specific staining methods. The major requirements of protein detection methods are high sensitivity (low detection limit), an extended linear dynamic range (for quantitative

accuracy), reproducibility and compatibility with MS-based post-electrophoretic procedures. Due to the extreme heterogeneity of protein concentrations that usually characterizes complex proteomes (up to 7 orders of magnitude), the ideal stain would be one able to detect as low a protein amount as few nanograms, displaying response linearity over a vast concentration range (6-7 orders of magnitude). Unfortunately such a stain doesn't exist, thus various expedients have to be enacted in order to face proteome complexity, such as protein pre-fractionation coupled to the use of ultra-narrow IPGs (zoom-in gels). Universal as well specific detection methods have been recently reviewed by Patton [42].

Among the "universal staining methods", the most popular are Coomassie Blue, silver-based methods (visible stains) and the Sypro fluorescent stain family, although many other detection techniques have been developed so far. Coomassie blue, especially in its colloidal version, is really cheap, convenient, offers a good linearity range (2 orders of magnitude) and a quite good sensitivity (about 10 ng per spot) [43]. Silver-based procedures offer the highest sensitivity among universal stains (down to 0.1 ng), nevertheless they are far from being stoichiometric and much less reproducible than Coomassie blue. Moreover they are usually not compatible with subsequent analysis via MS. Another widely used chemical is the Sypro Ruby fluorescent stain, developed by Steinberg *et al.* [44] and based on ruthenium chemistry, which conjugates wide linearity range with a considerably low detection limit (1-2 ng) and a full compatibility with further MS analysis. A great breakthrough in proteome analysis is constituted by a recently introduced method named "Differential in Gel Electrophoresis" (DIGE, [45]), which greatly reduces the problem of low reproducibility in comparative 2-DE. With this method two samples are labelled in vitro prior to any electrophoretic separation with two different fluorescent cyanine dyes (CyDyes, GE HealthCare), used on a "minimal" labelling strategy, differing in their excitation and emission wavelengths. Being stained prior to any separation and being selectively detectable on the base of specific dye excitation, equal amounts of the two samples can be mixed and simultaneously separated on a single gel, thus reducing the number of replicas required for proper statistical analysis. The experimental variability between the two samples is thus minimized, because the two samples are separated on the same IPG and second dimension gel. A schematic view of a 2-D DIGE

experiment with three fluorescent dyes (two samples and an internal standard) is shown in figure 1.4.

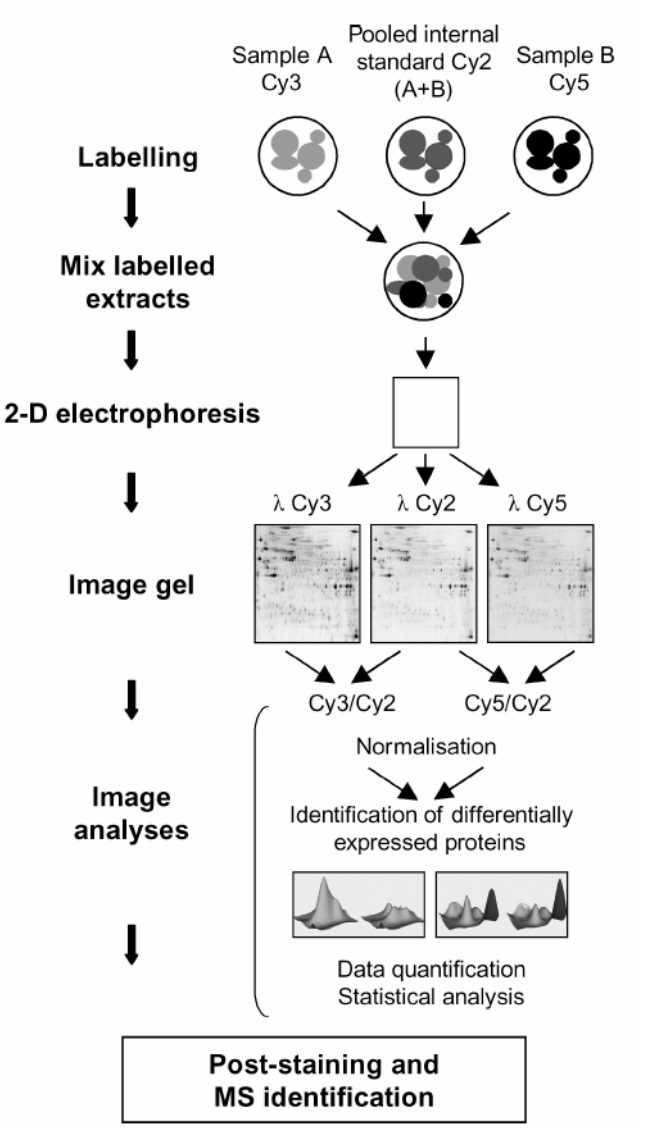


Figure 1.4. Schematic overview of 2-D DIGE technology. Samples A and B are fluorescently labelled with either Cy3 or Cy5, and a pooled internal standard (with equal amounts of the two samples under comparison) is labelled with Cy2. Samples are mixed and resolved in a single 2-D gel, after that protein spots are visualised by alternately illuminating the gel at the excitation wavelengths of each dye.

A universal protein stain is generally used for performing a wide-spectrum comparative analysis of complex proteomes, in order to visualize and quantify as many proteins as possible. However, when one is interested in a limited part of a proteome, such as specific protein classes, particular PTMs or even single proteins, specific detection methods have to be used.

As previously said in this section, PTMs studies are crucial for functional proteomics, because numerous regulatory pathways, enzyme activities, developmental changes and pathogenesis depend on them. Among the most

studied PTMs there are phosphorylation and glycosilation, for which many specific detection methods have been developed. For example, phosphorylated proteins can be selectively stained by immuno-detection with phosphoamino acid-specific antibodies after western blot. Recently, a fluorescent detection method for gel-separated phosphoproteins using Pro-Q Diamond phosphoprotein dye has been introduced. A similar fluorescent dye has been developed also for glycoproteins, namely Pro-Q Emerald 488. Another way of detecting glycoproteins is based on the great affinity and specificity displayed by particular proteins (i.e. lectins) for sugar moieties.

1.4.4. COMPUTER-BASED 2-D IMAGE ANALYSIS

As previously discussed, the main objective of expression proteomics is the simultaneous quali-quantitative analysis of as many proteins as possible from a complex proteome. Once 2-D maps have been generated, image analysis is usually performed with the help of specific softwares after image digitisation. The traditional workflow is the following: (i) pre-processing gel images, (i.e. image cropping, filtering, streaking removal and background subtraction); (ii) automated spot detection; (iii) automated and manual matching of all the spots displayed on the various gel images involved in the comparison; (iv) qualitative and quantitative analysis of matched spots (i.e. automated selection of all the spots newly expressed, repressed or significantly modulated in one sample with respect to the other); (v) data collection and representation. The whole workflow is routinely performed with the help of various dedicated softwares (PDQuest, Melanie, Z3 and Z4000, Phoretix and Progenesis), whose aim is to render the comparative analysis as fast as possible, reducing the manual intervention by the operator. A recently introduced really innovative software for 2-D map analysis is SameSpots (Nonlinear Dynamic Ltd, Newcastle upon Tyne, UK), which works on a principle quite different from the one the other softwares are based on: in practice it allows for visual, quick image distortion prior to any analytical step, thus making subsequent spot detection much more accurate, precise and reproducible from gel to gel. However, in spite of the great efforts made by the informatics companies to produce ever more reliable and automated softwares, because of

improper sample preparation or IPG choice, 2-D maps are often over-crowded with spots and proper matching is frequently a challenge. For this reason a careful, time consuming manual revision of the automatic matching is a compulsory step in comparative proteomics.

1.4.5. FURTHER PROTEIN ANALYSIS

The desired end point of any expression proteomics study is usually the identification and the characterization of all of the “selected” proteins, namely those whose expression level has been found to be significantly different (usually 2-folds or more) among the samples under comparison. MS has become the technique of choice for protein identification, because of high sensitivity and accuracy, small amount of material required and high throughput (reviewed by [46]). Moreover, MS also allows for PTMs investigation, including phosphorylation and glycosylation [47, 48]. PMF is usually the first tool for protein identification, since it can be performed also with linear MS (without the need for MS/MS), and is based on the comparison between a set of peptide masses produced by MS analysis of a protein digest (that is the protein fingerprint, usually obtained with Trypsin) and an *in silico*-built library of peptide mass profiles obtained by virtually cutting with the same enzyme all the protein sequences stored in a database. Identification is thus achieved by MS profiles matching between the unknown protein and the known proteins in the database. Despite its simplicity and user-friendliness, this method presents some relevant drawbacks: (i) it is unreliable when the organism genome has not been completely sequenced; (ii) masses profiles obtained from proteins with extensive PTMs might not match those calculated for protein sequences in databases; (iii) when more proteins are present in a single spot (spot overlap, [49]) MS spectra interpretation might become considerably difficult. In all of these cases it is necessary to obtain amino acid sequences, which can be achieved by tandem MS/MS. MS/MS takes advantage of two-stage machines like matrix assisted laser desorption ionization-time of flight/time of flight (MALDI-TOF/TOF), ESI-triple quadropole, ion-trap or quadrupole-TOF, to induce peptide bond fragmentation.

1.5. AIMS OF THE WORK

The research work carried out in the last three years and presented in this thesis has been conduced partly in the Laboratory of Biochemical Methodologies of the Scientific and Technological Department of the University of Verona, and partly in the new Laboratory of Proteomics recently set up in the Department of Chemistry, Materials and Chemical Engineering "Giulio Natta" in the Polytechnic of Milano.
The objectives of this Ph D project can be grouped under two different thematic sections, according to which the present thesis will be globally structured:

1. Section I: Characterization of microorganisms exploitable in bioremediation and biorestoration process by means of 2-DE and other microbiological and biochemical methods.
2. Section II: Development and/or validation of new proteomics methods to cope with the “small” and the “low abundance” proteome:
    - development of a new application of the recently introduced “Equalizer beads” technology for the removal of potentially harmful low abundance impurities from r-DNA products of pharmaceutical interest and the achievement of extremely high purity grades.
    - Development of a new analytical technique for the separation of small proteins and peptides in polyacrylamide-based matrices according to molecular mass (SDS-PAGE focusing).

Various collaborations were set up for the different studies.
The biochemical features of the peculiar *Pseudomonas stutzeri* strain employed for the biorestoration of an ancient fresco from Pisa’s cemetery were investigated in collaboration with Prof. Sorlini (Department of Food Science and Microbiology, University of Milan, Italy) and Ranalli’s group (Department of Agri-Food, Environmental and Microbiological Science and Technology, University of Molise, Italy).
As concerns the *Stenotrophomonas maltophilia* strain isolated from a selenite-contaminated soil and thus exploitable for bioremediation processes, all the work has been performed in collaboration with Prof. Vallini’s group (Laboratory of

Microbiology, Scientific and Technological Department, University of Verona, Italy).

Prof. Lello Zolla (Proteomics Laboratory, Department of Environmental Sciences, Tuscia University, Viterbo, Italy) provided the expertise of MS-based protein analysis and carried out all the protein identifications reported in this thesis.

The experiments aimed at validating the Equalizer beads technology and at developing a new polishing step for obtaining extremely pure, safe r-DNA products for human consumption were performed in collaboration with Ciphergen Biosystems (Cergy-Pontoise, France), which provided the samples and performed SELDI-MS analyses.

If, on one side, the Equalizer beads technology helps overcoming the age-old problem of the scarce dynamic range typically offered by 2-DE, it doesn't resolve the inability of the latter to deal with proteins smaller than 5 kDa. In this concern, the last subject discussed in this thesis regards the development of a new eletrophoretic separation technique, dubbed "SDS-PAGE focusing", where proteins and peptides focus in presence of SDS and Immobilines affixed to a neutral polyacrylamide matrix according to molecular mass. This work was carried out in collaboration with Cleardirection Ltd. (4 Pekeris St., Rehovot 76702, Israel).

1.6. BIBLIOGRAPHY

[1] Wilkins M. *Nature.* 1995; 378-653.

[2] Cobon GS, Verrills N, Papakostopoulos P, Eastwood H, Linnane AW. *Biogerontology.* 2002; **3**: 133-136.

[3] Southan C. *Proteomics.* 2004; **4**: 1712-1726.

[4] de Hoog CL, Mann M. *Proteomics. Annu. Rev. Genomics Hum Genet.* 2004; **5**: 267–293.

[5] Kavallaris M, Marshall GM. *Med. J. Aust.* 2005; **6**: 575-579.

[6] Ozols RF. *Cancer J.* 2002; **8**: S22-S30.

[7] Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. *Lancet.* 2002; **359**: 572-577.

[8] Ren Y, He QY, Fan J, Jones B, Zhou Y, Xie Y, Cheung CY, Wu A, Chiu JF, Peiris JS, Tam PK. *Proteomics.* 2004; **4**: 3477-3484.

[9] Verrills NM, Walsh BJ, Cobon GS, Hains PG, Kavallaris M. *J. Biol. Chem.* 2003; **278**: 45082-45093.

[10] Johnson JR, Florens L, Carucci DJ, Yates JR. *J. Proteome Res.* 2004; **3**: 296-306.

[11] Alberts B. *Cell.* 1998; **92**: 291–294.

[12] Terpe K. *Appl. Microbiol. Biotechnol.* 2003; **60**: 523–533.

[13] Orru S, Caputo I, D'Amato A, Ruoppolo M, Esposito C. *J. Biol. Chem.* 2003; **278**: 31766–31773.

[14] Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin, B. *Methods.* 2001; **24**: 218–229.

[15] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin, B. *Nat. Biotechnol*. 1999; **17**: 1030–1032.

[16] Immler D, Gremm D, Kirsch D, Spengler B, Presek P, Meyer HE. *Electrophoresis.* 1998; **19**: 1015–1023.

[17] Andersson L, Porath J. *Anal. Biochem.* 1986; **154**: 250–254

[18] Blaydes JP, Vojtesek B, Bloomberg GB, Hupp TR. *Methods Mol. Biol.* 2000; **99**: 177–189.

[19] Oda Y, Nagasu T, Chait BT. *Nat. Biotechnol.* 2001; **19**: 379–382.

[20] Zhou H, Watts JD, Aebersold R. *Nat. Biotechnol.* 2001; **19**: 375–378.

[21] Hamdan, M, Righetti, P.G.: *Proteomics Today: Protein Assessment and Biomarkers Using Mass Spectrometry, 2-D Electrophoresis and Microarray Technology*, Wiley-VCH, Hoboken, 2005, pp. 341-373.

[22] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R., *Nat. Biotechnol.* 1999; **17**: 994–999.

[23] Guina T, Wu M, Miller SI, Purvine SO, Yi EC, Eng J, Goodlett DR, Aebersold R, Ernst RK, Lee KA. *J. Am. Soc. Mass Spectrom*. 2003; **14**: 742–751.

[24] Li J, Steen H, Gygi SP. *Mol. Cell. Proteomics*. 2003; **2**: 1198–1204.

[25] Lueking AM, Horn H, Eickhoff H, Lehrqach H, Walter G. *Anal. Biochem.* 1999; **270**: 103–111.

[26] Borrebaeck CA, Ekstrom S, Molmborg Hager AC, Nilson J, Laurell T, Marko-Varga G. *BioTechniques.* 2001; **30**: 1126– 1132.

[27] Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. *Lancet*. 2002; **359**: 572–577.

[28] Grizzle WE, Adam BL, Bigbee WL, Conrads TP, Carroll C, Feng Z, Izbicka E, Jendoubi M, Johnsey D, Kagan J, Leach RJ, McCarthy DB, Semmes OJ, Srivastava S, Srivastava S, Thompson IM, Thornquist MD, Verma M, Zhang Z, Zou Z. *Dis. Markers.* 2003; **19**: 185–195.

[29] Paweletz CP, Trock B, Pennanen M, Tsangaris T, Magnant C, Liotta LA, Petricoin EF 3rd. *Dis. Markers.* 2001; **17**: 301–307.

[30] Zhang YF, Wu DL, Guan M, Liu WW, Wu Z, Chen YM, Zhang WZ, Lu Y. *Clin. Biochem*. 2004; **37**: 772–779.

[31] Won Y, Song HJ, Kang TW, Kim JJ, Han BD, Lee SW. *Proteomics.* 2003; **3**: 2310–2316.

[32] Powell K. *Nat. Med*. 2003; **9**, 980.

[33] O'Farrell PH, *J. Biol. Chem.* 1975; **250**: 4007–4021.

[34] Rabilloud T. *Electrophoresis.* 1998; **19**: 758-60.

[35] Herbert BR, Molloy MP, Gooley BAA, Walsh J, Bryson WG, Williams KL. *Electrophoresis.* 1998; **19***:* 845–851.

[36] Herbert B, Galvani M, Hamdan M, Olivieri E, MacCarthy J, Pedersen S, Righetti PG. *Electrophoresis*. 2001; **10**: 2046-2057.

[37] Mastro R, Hall M. *Anal Biochem.* 1999; **273**: 313-315.

[38] Righetti PG, Drysdale JW, *Ann. N. Y. Acad. Sci.* 1973; **209**: 163–186.

[39] Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W. *J. Biochem. Biophys. Methods*. 1982; **6**: 317–339.

[40] Zuo X, Speicher DW, *Electrophoresis*. 2000; **21**: 3035–3047.

[41] Schägger H, von Jagow G, *Anal. Biochem.* 1987; **166**: 368–379.

[42] Patton WF, *Electrophoresis.* 2000; **21**: 1123–1144.

[43] Neuhoff V, Arold N, Taube D, Ehrhardt W, *Electrophoresis.* 1985; **9**: 255–262.

[44] Steinberg, T.H., Chernokalskaya, E., Berggren, K., Lopez, M.F., Diwu, Z., Haugland, R.P., Patton, W.F. *Electrophoresis.* 2000 21: 486-496..

[45] Ünlü M, Morgan ME, Minden JS, *Electrophoresis* 1997; **18**: 2071–2077.

[46] Aebersold R, Mann M. *Nature.* 2003; **422**: 198–207.

[47] Kalume DE, Molina H, Pandey A, *Curr. Opin. Chem Biol.* 2003; **7**: 64–69.

[48] Mann M, Jensen ON. *Nat. Biotechnol.* 2003; **21**: 255–261.

[49] Campostrini N, Areces LB, Rappsilber J, Pietrogrande MC, Dondi F, Pastorino F, Ponzoni M, Righetti PG. *Proteomics*. 2005; **5**: 2385-2395.

SECTION I:

CHARACTERIZATION OF MICROORGANISMS EXPLOITABLE IN BIOREMEDIATION AND BIORESTORATION PROCESSES

GENERAL INTRODUCTION

Microorganisms are known for their impressive ability of growing on the most different substrates, from the richest and complex ones to the poorest and simple. Moreover they often show a much higher resistance to unfavourable environments in comparison with other more complex organisms (i.e. plants and animals). In fact some microbes, especially bacteria, can efficiently sequestrate and/or metabolize toxic heavy metals, or degrade some dangerous organic compounds (e.g polycyclic aromatic hydrocarbons, polychlorinated biphenyls, phthalate esters, nitro-aromatic compounds and industrial solvents), while simultaneously detoxifying the environment where they live, and thus representing an economically convenient device for detoxifying polluted soils or waste waters. Among the most widespread ambient pollutants there are heavy metals and other trace elements, so defined because usually present at low concentrations in most soils, plants and living organisms [1]. Various trace elements, despite being essential for normal growth of plants and animals (Cu, Zn, Fe, Mn, Mo, Co and Se), when present at higher concentrations become toxic to live species. Trace element concentrations in soils depend both on parent material and human activities, the latter being much more incisive in urban and cultivated areas, due to industrial activities (i.e. mining and smelting, atmospheric deposition and waste disposal) and to the use of metal-enriched compounds in agriculture (fertilizers, fungicides, pesticides and herbicides containing Cu, Zn, Fe, Mn and B) [2]. The term “bioremediation” is used to describe the process that uses living organisms to degrade or remove hazardous compounds present in the environment. As concerns metals, microbial transformations can be classified into two general categories: redox reactions on inorganic forms and conversion from inorganic to organic forms and *vice versa* (typically, methylation and demethylation). By reducing iron, sulphur, manganese and arsenic, microbes can obtain energy [3], while arsenic species, selenium and uranium can be used by bacteria as electron terminal acceptors in anaerobic respiration [4, 5]. In addition, microorganisms possess reduction mechanisms that are not coupled to respiration, thus constituting a genuine resistance mechanism. Microbial methylation plays an important role in the biogeochemical cycle of metals, because methylated

compounds are often volatile. Finally, metal can be also precipitated as insoluble sulphides by sulphate-reducing bacteria. To sum up, microbiological processes can either solubilise metals, thereby increasing their availability, or immobilize them, thus reducing their toxicity. In any case, whatever the mechanism involved, microbial-driven metal transformation is a very important component of biogeochemical cycles of metals, and a powerful tool for bioremediation of contaminated soils [6].

Due to their great ability of growing in the most difficult environments, microbial communities are also frequently found on the surfaces of ancient artworks, made of the most different materials, such as paintings on canvas and walls, sculptures in wood and stone, and every kind of object in paper, leather, wood and masonry [7-9]. As concerns paintings, the events in which microbial colonization has been recognized as the main cause of physical and aesthetical degradation are very numerous. Paintings, either easel or mural, contain a wide panel of organic and inorganic compounds, which provide ecological niches exploitable by a large number of microbial species. In easel painting, the support material itself (paper, canvas, wood, silk and others) could be easily degraded by microorganisms, as could the organic molecules used as thickeners and organic binders (animal or plant glues, sugars, gums, proteins, oils and waxes, [10]). Perhaps, the most emblematic example of how micro-organisms can rapidly colonize apparently invulnerable artworks is the case of the Lascaux Cave in the Dordogne region of France, designated by some the Sistine Chapel of the Paleolithic, because of the impressive amount of prehistoric mural paintings there conserved. This precious cultural heritage, after having survived for thousands of years without significant deterioration, was opened in 1948 to the visitors, but after only fifteen years it had to be definitely closed to the public because of the appearance of the *maladie verte* (or green disease), consisting of a green patina constituted by the green unicellular alga *Bracteacoccus minor* [11]. Such a rapid microbial colonization of the ancient pictorial surface was surely due to the consistent micro-climate modification brought by human presence. Although this is probably an extreme case, it is certain that outdoor artwork, especially when lithoid materials, stones, frescoes and paint are involved, is very prone to deterioration caused by ageing

and pollution, the latter being responsible for monument surface alterations such as black crusts, nitration, sulphation and hydrocarbon deposition [12].

Although usually associated with artwork degradation, in some events microbes have also been exploited for restoring and consolidating purposes, especially in the case of stoneworks. In this sense, sulphate-reducing and nitrate-reducing bacteria can be used to remove, respectively, sulphate and nitrate forming the so-called black crusts [13, 14], mostly originated from the consumption of fossil fuels, while hydrocarbon degrading bacteria can break down aliphatic and aromatic hydrocarbons laid down on stone. Biocalcifying bacteria promote stone consolidation by inducing chemical changes that promote calcite precipitation [15]. Eventually, various enzymatic activities (i.e. proteases, lipases and carbohydrase) can be exploited to easily remove various organic substances from artwork surfaces [16].

The following chapters report the results of our investigations on two peculiar microorganisms usable for biorestoration and bioremediation purposes. These two bacterial strains are *Pseudomonas stutzeri* strain A29, which has been used to efficiently degrade protein-based glue resistant to the cleaning techniques routinely employed, and *Stenotrophomonas maltophilia* strain SeITE02, a bacterium isolated from a selenite-contaminated soil capable of reducing the toxic metal to insoluble elemental selenium, thus detoxifying the surrounding environment.

BIBLIOGRAPHY

[1] Phipps DA. Chemistry and biochemistry of trace metals in biological systems. In: Lepp NW (Editor). Effect of heavy metal pollution on plants: effects of trace metals on plant function, vol. I. London and New Jersey: Elsevier Applied Sci. 1981; pp. 1–54.

[2] Webber J. Trace metals in agriculture. In: Lepp NW (Editor). Effect of heavy metal pollution on plants: Metals in the environment, vol. II. London and New Jersey: Elsevier Applied Science, 1981; pp. 159–84.

[3] Santini JM, Sly LI, Schnagl RD, Macy JM. *Appl. Environ. Microbiol.* 2000; **66**: 92-97.

[4] Stolz J, Oremland R. *FEMS Microbiol. Rev*. 1999; **23**: 615-627.

[5] Tebo BM, Obraztsova AY. *FEMS Microb Letters.* 1998; **162**:193-198.

[6] Lloyd JR, Lovley DR. *Current Opinion in Biotechnology.* 2001; **12**: 248-253.

[7] Barkay T, Schaefer J. *Current Opinion in Microbiology.* 2001; **4**: 318-323.

[8] Bravery AF. Biodeterioration of paint-a state-of-the-art comment. In Houghton DR, Smith RN and Eggins HOW (Editors). Biodeterioration, vol. 7. New York, N.Y: Elsevier Applied Science, 1988; pp. 466–485.

[9] Tiano P. Biodeterioration of stone monuments: a critical review. In Garg KL, Garg N, Mukerji KG (Editors). Recent advances in biodeterioration and biodegradation, vol. I. Calcutta, India: Naya Prokash, 1993; pp. 301-321.

[10] Strzelczyk AB. Paintings and sculptures. In Rose AH (Editor). Microbial deterioration. London, United Kingdom: Academic Press, 1981; pp. 203–234.

[11] Lefèvre M. La “maladie verte” de Lascaux. *Stud. Conserv*. 1974; **19**: 126–156.

[12] Saiz-Jimenez C. *Agrobiologia.* 1995; **11**: 161–175.

[13] Warscheid T, Braams J. *Int. Biodeterior. Biodegrad.* 2000; **46**: 343–368.

[14] Ranalli G, Matteini M, Tosini I, Zanardini E, Sorlini C. Bioremediation of cultural heritage: removal of sulphates, nitrates and organic substances. In: Ciferri O, Tiano P, Mastromei G. (Editors). Of Microbes and Art. The Role of Microbial Communities in the Degradation and Protection of

Cultural Heritage. New York: Kluwer Academic/Plenum Publishers, 2000; pp. 231–245.

[15] Castanier S, Le Métayer-Levrel G, Perthuisot JP. *Sediment Geol.* 1999; **126**: 9–23.

[16] Wolbers R. Cleaning painted surfaces: aqueous methods. London: Archetype Publications, 2000.

CHAPTER II:

PROTEOMIC AND BIOCHEMICAL CHARACTERIZATION OF *PSEUDOMONAS STUTZERI* STRAIN A29, USED IN THE BIORESTORATION OF AN ANCIENT FRESCO

2.1. INTRODUCTION: BIORESTORATION OF THE PISA'S CEMETERY FRESCOES

Spinello Aretino, along with other artists, in 1391 was asked to decorate the peripheral walls of the Campo Santo (Monumental Cemetery) in Piazza dei Miracoli (Pisa, Italy). He painted a cycle of four frescoes illustrating some episodes of the life of the saint martyrs Efisio and Potito. These frescoes underwent several restoring interventions, with the last one performed at the end of the World War II, during which the entire monumental complex was highly damaged. Thus, starting with the end of the forties, the frescoes were removed from the walls by the "tear off" technique, consisting in the application of a wide gauze onto the fresco surface, by using animal glue as a consolidating agent. Upon hardening of the glue, the covering gauze and the fresco are fused into a compact layer, which is then detached from the wall. In order to liberate the fresco from the cloth, its back is reinforced by a canvas cloth, whose adhesion is ensured by glue (e.g., casein or collagen based glues), in turn supported on rigid panels. At this point, in principle, when the front of the fresco is treated with proteolytic enzymes, the gauze obscuring it can be gently removed, thus restoring the painting to view. The second of the four frescoes depicted by Aretino, by the title "Conversione di S. Efisio e Battaglia", was removed some twenty years ago and abandoned in a store room up to recent times. Finally, when the curators attempted to remove the cloth, they found that the glue resisted any attempt to digestion, even when treated with the most aggressive proteolytic enzymes available on the market. In fact, at the time of the fresco removal, the glue had been admixed with formaldehyde, added as an antimicrobial agent, which however reacted with the ε-amino groups of Lys and, in general, with all primary amino groups of proteins. Although this reagent is not a cross-linking agent like glutaraldehyde, it is known that, upon long storage times, secondary reactions lead indeed to actual cross-linking (also irreversible when occurring on Tyr and Trp residues) [1]. Probably, the layer of glue covering the fresco (ranging in thickness from 1 to 3 mm), upon such a long storage time had been extensively cross-linked. The problem remained unsolved until Ranalli and Sorlini's research groups, who possess a considerable know-how of bioremediation intervention

with viable bacterial cells (i.e., direct application onto artwork surfaces, in particular stone and marble) [2-3], were contacted by the Opera Primaziale Pisana. A series of preliminary experiments were conducted for solving the question. Strains of *Pseudomonas* were chosen among various possible species, since these bacteria are not pathogenic, not spore-forming and are known to react readily to environmental conditions by inducing and secreting a broad range of enzymatic activities. Five of such species were tested: *P. cepacia, P. testosteroni, P. fluorescens, P. flavescens* and *P. stutzeri*. Since the latter species exhibited the best growth curves onto ground chips of the insoluble glue harvested from the clothed fresco, it was selected for the bioremediation procedure. Although full restoration of the fresco was easily achieved (see [4] for an exhaustive description of the bioremediation intervention), specific information on the molecular mechanism involved was still missing. Therefore biochemical and proteomics investigations were carried out by our research group in an attempt at collecting more details on such an innovative restoring technique.

2.2. MATERIALS AND METHODS

2.2.1. CHEMICALS

β-Casein, azo dye-impregnated collagen (Azocoll), azo dye-impregnated casein (Azocasein), Trypsin TPCK treated, Collagenase (Clostridiopeptidase A) were purchased from Sigma-Aldrich (Steinheim, Germany). Triton X-100, NADH and pyruvic acid were from Merck (Darmstadt, Germany). Acrylamide/Bis acrylamide solution and the linear IPGs (pH gradient 3-10, 17 cm long), were obtained from Bio-Rad Labs (Hercules, CA, USA). Carrier ampholytes were purchased from Pharmacia Biotech (Uppsala, Sweden). *Pseudomonas stutzeri* A29 strain, animal glue and the aged animal glue derived from the cloth employed in the fresco's restoration were kindly provided by Distam, University of Milan, Milan, Italy.

2.2.2. BACTERIAL CULTURES

Pseudomonas stutzeri A29 strain was inoculated at a starting OD_{600} of about 0.02 in M9 mineral medium, supplemented with one among the three carbon sources employed in this study at a final concentration of 5 g/L. The three carbon substrates were glucose, animal glue and the aged animal glue derived from the cloth employed in the fresco's restoration. Incubation was performed at 30°C in agitation for 10 hours. Carbon substrates were previously sterilized by filtration (glucose) or autoclave (animal glue, 121°C for 15 min), while aged animal glue was ground and added to medium without previous sterilization, so as to mimic the conditions encountered in the real application. After incubation, cells were harvested by centrifuging at 10 000 x g for 10 minutes, washed twice in saline phosphate buffer (PBS) and stored at -20°C. Cultural supernatants were also collected and immediately frozen at -20°C.

2.2.3. ENZYMATIC ACTIVITY ASSAYS

2.2.3.1. Lactate dehydrogenase (LDH) activity

LDH activity in cultural supernatants was assayed in order to detect possible cell lysis. The reaction was carried out in 1 mL cuvettes at 20°C in the following conditions: 60 mM phosphate buffer pH 7.00, 0.2 mM NADH, 1 mM pyruvic acid. 100 µL of supernatant were eventually added, after that absorbance at 340 nm was immediately monitored by reading every 20 seconds for a total of 6 minutes. As positive control we used cultural supernatant of *Pseudomonas* cells lysed by sonication, while the blank was the same sample but thermally treated in order to completely inactivate LDH.

2.2.3.2. General proteolytic activity

Cultural supernatants were assayed for the ability of hydrolysing casein. An azo dye-impregnated casein (Azocasein) was used for a colorimetric test. For positive control Trypsin solution (10 µg/mL) was employed, and it was also used to create a four-points calibration curve (0 for blank, 10, 25, 50, 100 ng of enzyme). A three-point curve (25, 50 and 100 µL) for each supernatant sample was also traced at

different incubation times. The proteolytic reactions were carried out in 1.5 mL tubes at 37°C in 1.2 mL of solution (100 mM phosphate buffer pH 8.00, 0.25% w/v Azocasein) by adding appropriate volumes of supernatant (if needed adding milli-Q water to obtain a total sample volume of 100 µL) After pre-established incubation times, volumes of 200 µL were withdrawn from each tube and the reaction was stopped by adding TCA (6% v/v final concentration). After 10-minute incubation at 4°C, the residual intact protein was precipitated by a 10 min centrifugation at 20 000 x g, and 180 µL of supernatant were finally transferred to a microtitration plate well. Sample colour was enhanced by rising the pH to alkaline values with 0.5 M NaOH (final concentration), after that the absorbance at 415 nm was immediately measured.

2.2.3.3. Specific collagenolytic activity

Collagenolytic activity in cultural supernatants was evaluated by an azo dye-impregnated collagen substrate (Azocoll). Collagen degradation kinetics were followed by using only one concentration of sample supernatant and of reference enzyme solution (*Chlostridium histolyticum* pure Collagenase, 1 mg/mL), by varying the reaction time. Reactions were carried out in 15-mL sterile tubes in a total volume of solution of 3 mL (0.5% w/v solid Azocoll, 100 mM phosphate buffer pH 8.0, 1mM $CaCl_2$), adding 0.2 mL of sample. As reference 10 µg of pure Collagenase were used, while 0.2 mL of water were used as blank. The proteolytic reaction was carried out at 37°C in fast agitation. After pre-established incubation times, volumes of 700 µL were withdrawn from each tube and the reaction was stopped by adding TCA (6% v/v final concentration). Each sample was immediately centrifuged at 20 000 x g for 5 min, after that 500 µL of the supernatant were transferred to a 1-mL cuvette and eventually diluted 1:2 (v/v) with water. Absorbance at 520 nm was measured, and the values plotted against incubation time.

2.2.4. ZYMOGRAM

A casein zymogram was used in order to assess how many and what kind of proteolytic components were present in our cultural supernatants. Samples were

opportunely concentrated and prepared for SDS-PAGE separation in Laemmli buffer without reduction and boiling. After an incubation at 20°C for 20 minutes samples were loaded onto a 13-cm long 10% T polyacrylamide gel copolymerized with 0.1% (w/v) β-casein. Electrophoresis was performed at 4°C in Tris-glycine buffer pH 8.3, by applying 60 V for 1 h, followed by 130 V until the end of the run. The gel was then incubated overnight in 2% (v/v) Triton X-100 dissolved in 100 mM phosphate buffer pH 8.00, under gentle shaking, in order to remove SDS and allow for protein refolding and activity restoration. Finally the gel was stained with colloidal Coomassie blue and destained with 7% acetic acid water solution.

2.2.5. TWO DIMENSIONAL ELECTROPHORESIS

The supernatants of *Pseudomonas stutzeri* grown on glucose, on the fresco glue or on fresh glue were desalted by dialysis, lyophilized and then dissolved in a denaturing IEF buffer (7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP and 0.5% carrier ampholytes pH 3-10). After centrifugation (4 000 x g, 4°C, 10 minutes) for eliminating any residual particulate material, the samples were alkylated for 1 hour with 10 mM acrylamide, and the reaction was stopped with 10 mM DTT. The first dimension run was performed on 17 cm-long IPGs (linear pH 3-10 gradients). Isoelectric focusing (IEF) was carried out with a Protean IEF Cell (Bio-Rad, Hercules, CA). At the end of the focusing phase each strip was re-conditioned with a solution containing 2% SDS, 6 M urea, 20% glycerol and 0.375 M Tris-HCl (pH 8.8). The second dimension was performed on 7-20% polyacrylamide gradient gels by using a PROTEAN II XL Multi-Cell (Bio-Rad, Hercules, CA) and maintaining a controlled temperature of 11°C. All the gels were then fixed with a solution containing 40% ethanol and 10% acetic acid for 30 minutes prior to staining with Sypro Ruby. Image digitisation was carried out with the Versa Doc Image System (Bio-Rad, Hercules, CA), and spots of interest were excised from the gel, in-gel digested with trypsin and subject to PMF by MALDI-TOF MS. Measurements were performed using a TofSpec 2E MALDI-TOF instrument, operated in reflectron mode. The final mass spectra were produced by averaging 50-200 laser shots. Peptide masses were searched against SWISS-

PROT and TrEMBL databases, by using the search engines Profound, ProteinLynx and Mascot.

2.3. RESULTS

2.3.1. GROWTH CURVES AND INTEGRITY OF *P. STUTZERI* CELLS

For the present laboratory study, we had to select optimal growth times for control cells and cells grown on fragments of the original formaldehyde cross-linked glue, so as to ensure that all of the samples would be harvested, for analysis, not later than the late exponential growth phase, so as to avoid interferences in the following analysis due to potential cell lysis after reaching a growth plateau [5]. Figure 2.1 shows the growth curves of *P. stutzeri* strain A29 cultivated in broths containing glucose or a suspension of shredded glue from the fresco (growth on animal glue is not shown).

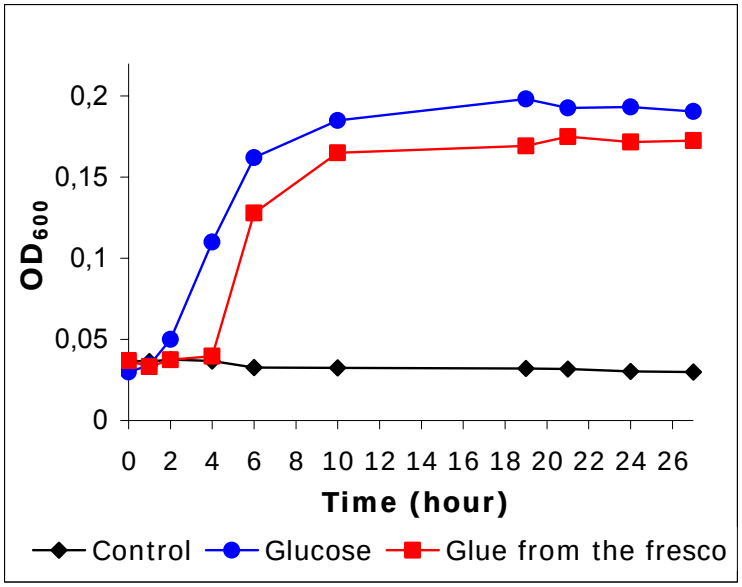


Figure 2.1. Growth curves of *P. stutzeri* grown on glucose (blu circles) and on the glue from the fresco (red squares). Black diamonds represent the negative control.

As expected, the onset of growth occurs at earlier times in the glucose–grown cell cultures, but, after a lag time, also the cells with glue debris as the sole energy source appear to grow at a similar rate and reach a confluence plateau at just about 10 hours. Therefore this time was selected for all cultures for harvesting. In order to make sure that no cell lysis had occurred prior to harvesting supernatant,

we performed the LDH test. The results demonstrated that LDH activity was completely absent in all of the cultural supernatants, while being clearly present in the positive control (results not shown).

2.3.2. 2-DE ANALYSIS OF THE *P. STUTZERI* SUPERNATANT

In a number of experiments, the supernatants of the controls and of the cells grown on the aged glue were subjected to classical 2-DE analysis. The maps developed with Sypro Ruby revealed ca. 70 spots of polypeptide chains in the supernatants of the glue from the fresco, while in the case of cells cultured in glucose a lower number of different proteins could be detected. Moreover, while many spots appeared to be common to the two samples, others seemed to be synthesised exclusively by the cells grown on the cross-linked glue. Figure 2.2 displays the 2-D maps of the two samples. A comprehensive quali-quantitative comparative analysis of the two protein profiles was performed by *in silico* analysis with PDQuest software, which confirmed that a substantial number of proteins were exclusively or preferentially secreted by the bacteria growing on the aged glue in respect to those grown on glucose. Of these, the more abundant ones were chosen for identification via peptide-mass fingerprinting, as described in the previous paragraphs. However, elution of major spots, trypsin digestion and MS analysis in the MALDI-TOF mode did not lead us to any spot recognition, partly because several spots contained too little material, partly because, even in presence of good spectra, we could not find corresponding proteins in the SWISS PROT and TrEMBL data bases (Swiss Prot lists only 59 proteins for *P. stutzeri*, among which not a single protease).

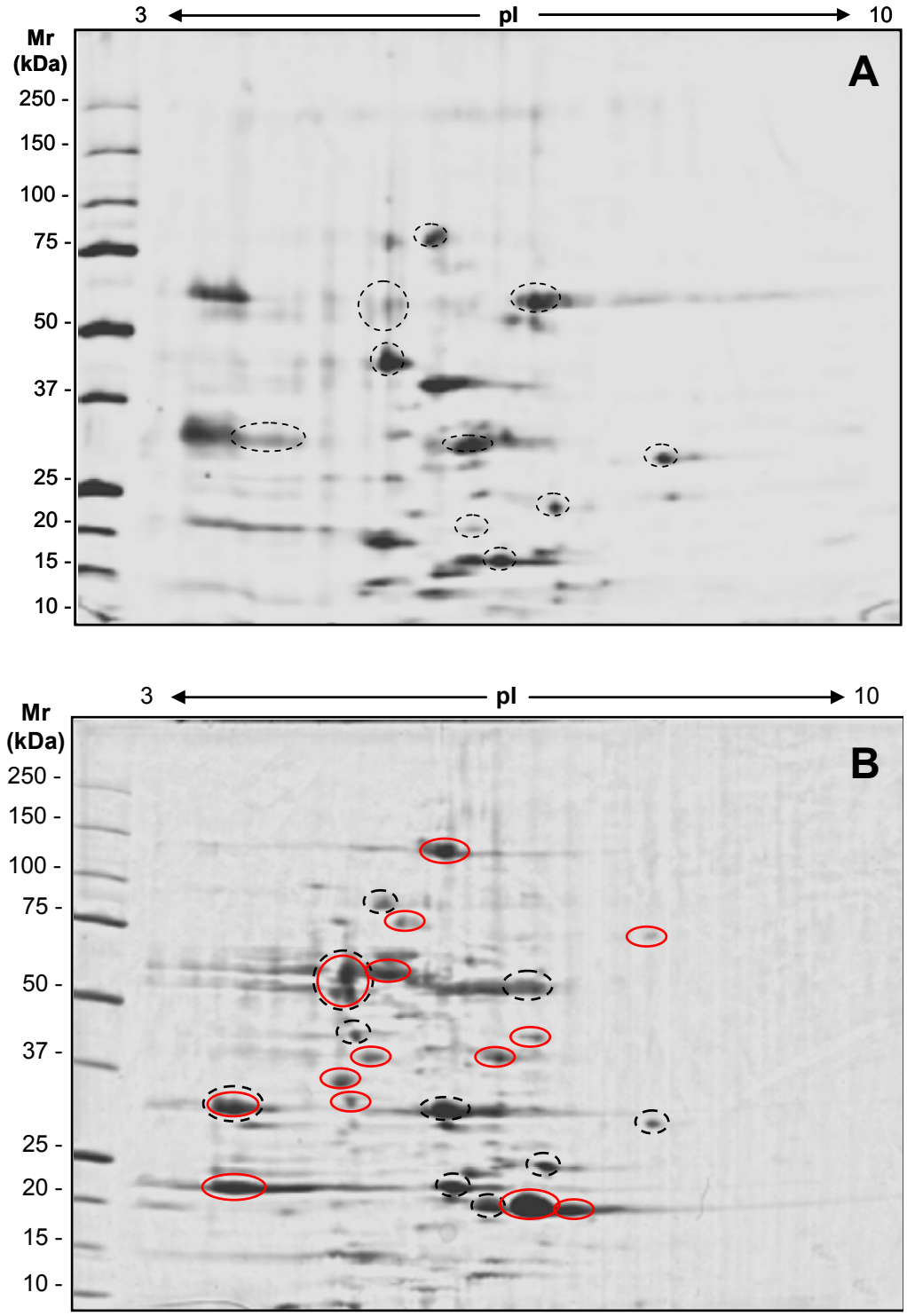


Figure 2.2. Two dimensional maps of supernatants of *P. stutzeri* grown on glucose as sole carbon source (A) and on debris of glue from Spinello's fresco (B). Zones indicated by dotted black circles represent some spots present in both the samples and provide a spatial reference for image comparison. Red circles point out the spots of the aged-glue sample that have been chosen for identification via PMF.

2.3.3. TESTS FOR PROTEOLYTIC ACTIVITIES IN THE CULTURAL SUPERNATANTS

A spectrophotometric assay for general caseinolytic activity was set up, by exploiting Azocasein as substrate. Figure 2.3 shows the proteolytic activity of the various supernatants (glucose, animal glue or the original cross-liked glue detached from the fresco surface), as measured at a fixed incubation time of 15 minutes as a function of increasing amounts of standard trypsin solution or of supernatant volumes.

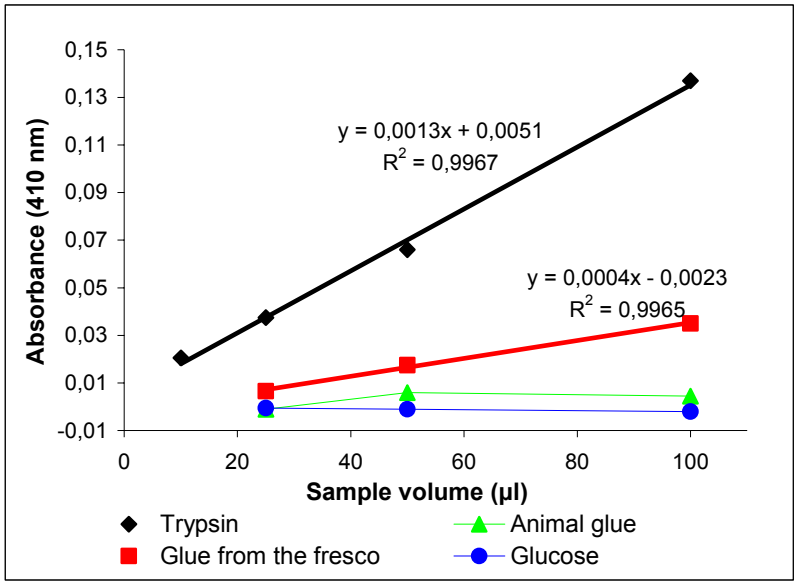


Figure 2.3. Caseinolytic activity in supernatants as measured via the rate of release of coloured peptides from Azocasein (substrate).
Quantification of the collagenolytic activity was obtained by comparing the slopes of the linear regression equations.

It is seen that such enzyme activity, in the different supernatants, is exhibited only by the bacteria grown in the original glue from the frescoes. Its specific enzyme activity, as compared with that of the standard trypsin solution, could be calculated as 3.63 mg/(mL*h). Also the more specific collagenolytic activity was assessed, using Azocoll as specific substrate. Figure 2.4 shows the enzyme activity profiles of a purified Clostridiopeptidase A from *C. histolyticum* (upper curve), of a *P. stutzeri* supernatant (grown on original glue from frescoes, middle curve), as well as of other supernatants of bacteria grown on fresh glue or on glucose (the bottom curves). It can be appreciated that, whereas the cells grown on the original glue show activities almost approaching those of the purified collagenase, all other supernatants exhibit essentially no activity, except for some

faint one in those supernatants originating from fresh glue. By calculating the slopes from linear regression analysis, it was possible to obtain the specific enzymatic activities (in mg of substrate converted per µL of supernatant per hour) for the various supernatants, by comparison with that of the purified collagenase from *C. histolyticum*. These activities are: 2,31 mg/(mL*h) for the supernatant of the cells grown on the glue from the fresco and 5,91 x 10^{-3} mg/(mL*h) for the supernatant obtained from a culture on fresh animal glue.

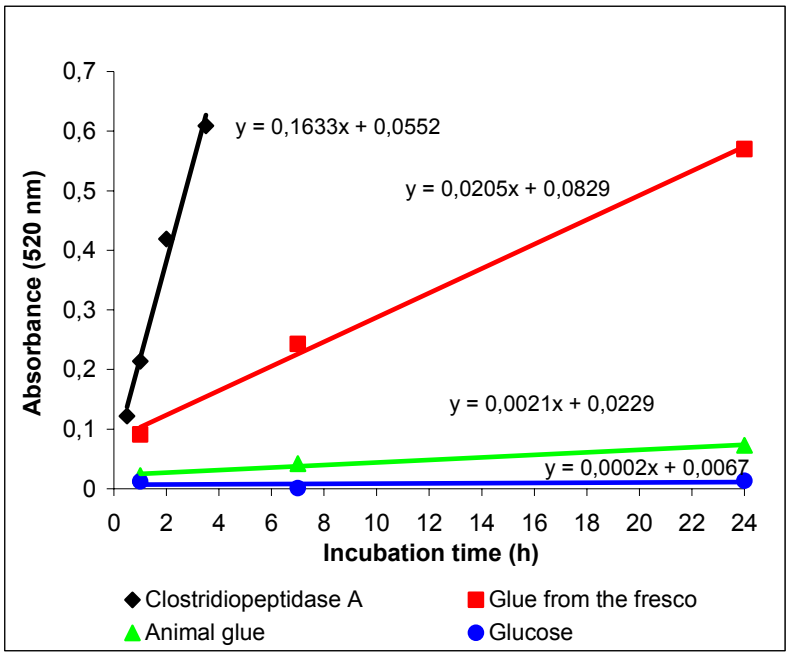


Figure 2.4. Collagenolytic activity in supernatants as measured via the rate of release of coloured peptides during proteolysis of Azocoll as substrate. Quantification of specific activities was calculated by comparing the slopes of the linear regression equations.

2.3.4. PROTEASE ZYMOGRAM

In an attempt at obtaining more detailed information about the set of proteins secreted by *P. stutzeri* strain A29 in order to grow on such a challenging protein substrate, we resorted to a peculiar type of SDS-PAGE, by which the various supernatants were analysed in a 10%T, 2.5%C polyacrylamide gel, cast in presence of 0.1% casein as a substrate. The supernatants were pre-concentrated 300 folds and incubated with SDS under mild conditions (no boiling, absence reducing agents) prior to electrophoretic separation. After the run, the polypeptide chains were allowed to refold by overnight incubation in 2% Triton X-100 dissolved in phosphate buffer pH 8.0, followed by staining and destaining in

colloidal Coomassie blue. As shown in figure 2.5, only the supernatant from cultures on Spinello's glue gave a ladder of caseinolytic activity bands, extending from *ca*. 20 kDa up to 250 kDa. Also a commercial preparation of collagenase purified from *C. hystoliticum* (tracks 7 and 8) was analysed by this technique along with the supernatant samples, producing an intense band at 120 kDa, that apparently coincides with the strongest activity bands in the supernatant from Spinello's glue (track 2).

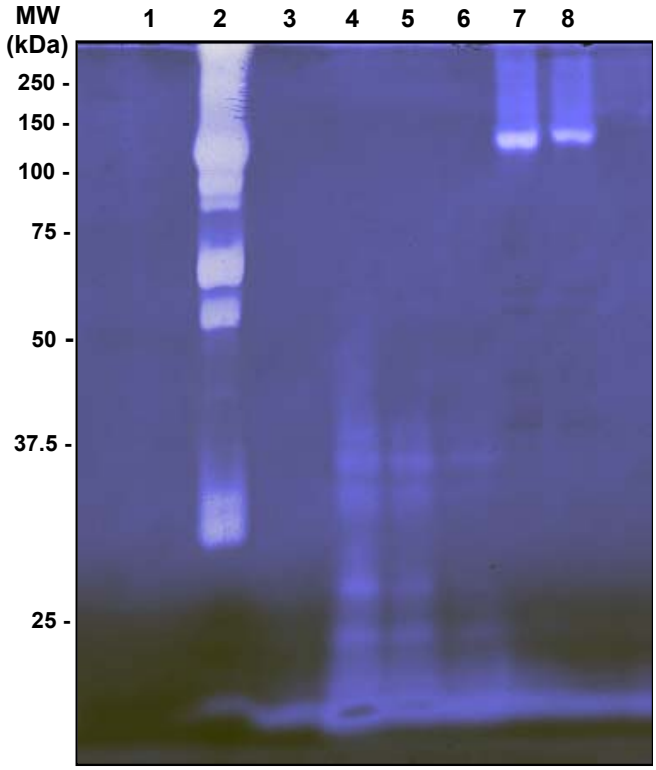


Figure 2.5. Casein zymograms *via* SDS-PAGE. The first three samples are supernatants (300 X concentrated) of *P. stutzeri* grown on glucose (track 1), on Spinello's glue (2) and on fresh animal glue (3). Tracks 4-6 are positive controls with purified trypsin (10, 5 and 1 µg, respectively, in tracks 4 to 6). Tracks 7 and 8 are purified collagenase from *C. hystoliticum* (20 µg in 7, 10 µg in 8).

2.4. DISCUSSION

Bioremediation of cultural heritage by means of "friendly" bacteria is gaining wide acceptance in present times. Outdoor art work, especially where lithoid materials, marble, frescoes and paint are involved, is very susceptible to deterioration brought about by ageing and, in recent decades, by pollution. In addition to air pollutants, the surface of man-made artistic work can also be altered by organic matter that has been applied, but then not promptly removed, during restoration. This was the case of the present report, since the gauze covering Spinello's art

work detached from the walls of Pisa's Monumental Cemetery could not be removed any longer from the face of the fresco, due to hardening and cross-linking of the animal glue during the 20 years of storage. No single enzyme, or concoctions of the most powerful proteases available on the market, could act on the layer of glue, and the fresco seemed to be irremediably lost. It is remarkable that *P. stutzeri* strain A29 could extensively digest this intricate mass of untreatable proteinaceous glue. For the *in situ* bioremediation process, the fresco had been entirely covered with a thin layer of hydrophilic cotton wool impregnated with the bacterial cell suspension at a high cell density (8.5 log CFU/mL). Contact times of 12-14 hours at a temperature of 30ºC had ensured almost full digestion (100% in the thinner areas, 80-90% in the thicker, up to 3 mm, layers), to the point that the gauze coating the fresco could be easily removed

With the present laboratory study, we wanted to investigate more in depth the molecular mechanism responsible for such a good performance. In this concern, we had to select optimal growth times for both control cells and cells grown on fragments of the original formaldehyde cross-linked glue, so as to ensure that all of the samples would be harvested, for analysis, not later than the late exponential growth phase, so as to avoid interferences in the following analysis due to potential cell lysis after reaching a growth plateau. In fact we started from the reasonable assumption that the bacterium, when attempting to digest such a recalcitrant protein substrate, had to secrete in the surrounding environment one or more proteolytic enzymes, whose synergic action would have been necessary for an effective degradation of the glue. Growth curves revealed that, although the onset of growth occurs at earlier times in the case of glucose, in presence of the aged-glue as the sole carbon source, after a short lag time, the growth rate approaches the one observed in the case of glucose, and a confluence plateau is reached by both cultures after about 10 hours. The fact that, after a short lag time, probably necessary for protease production and secretion, the growths proceed at similar rates with both glucose and the cross-linked glue, demonstrates the great ability of this strain to degrade the otherwise untreatable protein material, thus confirming its suitability for difficult biorestoration interventions. The time required for reaching the growth plateau (10 hours) is in full agreement with the incubation

period necessary to achieve the gauze detachment during the *in situ* intervention (12-14 hours), thus this incubation time was selected for all our experiments.
In order to make sure that no contamination between secreted and cytoplasmic proteins occurred during cell culture preparation, LDH activity assay was carried out. In fact LDH is known to be a cell sap enzyme that should not be found in the culture supernatant. This assay definitely proved that no cell lysis took place during sample preparation, thus confirming the extracellular nature of the enzymes under study.
The entire sets of proteins secreted by *P. stutzeri* cells under the two different growth conditions were also analysed by 2-DE. Although numerous proteins appeared to be expressed and secreted exclusively in cultures with the aged glue, thus suggesting that a quite complex set of enzymatic activities might operate in a synergic way to achieve the purpose, nevertheless PMF analysis by MALDI-TOF MS didn't lead to any protein identification. This failure had two concomitant causes: too little amount of material present in each spot, which in most cases prevented us from obtaining good spectra, and the few entries in the SWISS PROT and TrEMBL data bases for this species (Swiss Prot lists only 59 proteins, among which not a single protease, for *P. stutzeri*). Because not enough material for further gel analysis was available, we resorted to other classical biochemical studies, which required a much smaller amount of sample proteins.
In the first place, the general proteolytic activity in the culture supernatants was assessed by measuring the rate of coloured peptide release in solution in presence of an azo-derivative of β-casein (Azocasein). A general, easily degradable protein substrate was chosen in order to have a comprehensive picture of the panel of proteases present in the samples. This simple assay revealed that a considerable proteolytic activity (3.63 mg of substrate converted *per* mL of supernatant *per* hour) is secreted by *P. stuzteri* cells in the case of aged glue-cultures, but not when glucose or fresh animal glue were provided as the sole carbon sources. Because the glue employed for the fresco detachment was of animal origin, thus containing a large amount of collagen, the more particular collagenolytic activity was evaluated as well. Again, only the aged-glue sample showed a good proteolytic activity (2.31 mg/mL*h), as compared to pure Clostridiopeptidase A purified from *C. hystoliticum*. The two values of activity of

3.63 and 2.31 mg/(mL*h) for, respectively, general and collagenolytic activities, suggest that the majority of the proteases secreted by this bacterial strain are collagenases, while a few components of the proteolytic machinery are not able to digest such a substrate. Also the supernatant obtained from a culture with fresh-glue showed the ability to partially digest Azocoll, but at a much lower extent (5,91 x 10^{-3} mg/mL*h). This fact implies two consequences: first, the proteolysis of common animal glue doesn't require the special set of enzymes secreted by *P. stutzeri* strain A29 in order to grow on the Spinello's glue; in the second place, and as a corollary of the previous consideration, in absence of the original glue it was impossible for us to simulate the original environmental conditions in which *P. stutzeri* had produced such a unique set of proteases.

Although these data fully confirm the presence of proteolytic activities (at least against collagen and casein) essentially only in the supernatants of intact *P. stutzeri* grown on the original glue from Spinello's fresco, they could hardly give us any clue for the identification of such enzyme(s) (no data exist on proteases in *P. stutzeri* strains). For this reason we resorted to a classical zymogram technique, a peculiar type of SDS-PAGE by which the various supernatants were analysed in a 10%T, 2.5%C polyacrylamide gel cast in presence of 0.1% casein as a substrate for proteases. With this technique, after electrophoretic separation in moderately denaturing conditions, the proteins are generally allowed to re-fold by overnight detergent exchange with Triton X-100. In this particular case, during incubation for surfactant exchange, the re-folding proteases also degrade the surrounding casein, thus creating a white band over a dark-blue background when the entire gel is stained with Coomassie blue. As shown in Fig. 2.5, only the supernatant obtained from a culture with Spinello's glue gave a ladder of caseinolytic activity bands, extending from *ca*. 20 kDa up to 250 kDa. Interestingly, a commercial preparation of Collagenase from *C. hystoliticum* (tracks 7 and 8) gave an intense band at 120 kDa, which apparently coincides with the strongest activity band in the supernatant from Spinello's glue (track 2). This suggests that the *P. stutzeri* collagenase might have some homology with the *C. hystoliticum* one. We do not know, at present, if the other major activity bands of lower Mr values, extending from 100 kDa down to as low as 20 kDa, represent novel types of proteolytic enzymes or merely some degradation products of the 120 kDa parental species,

as reported for other types of collagenases from diverse bacterial species [6]. We believe, though, that at least some of these bands should represent unique enzyme activities, each endowed with specific proteolytic performance, and that only their combined action is able to attack the hardened glue till its complete digestion. In fact, when originally treated with purified enzyme preparations (including the same 120 kDa *C. hystoliticum* collagenase), the 20-year-aged glue could not be digested at all, as clearly demonstrated by the researchers who carried out the intervention [4].

2.5. CONCLUSIONS

The positive results achieved with the fresco of Pisa's cemetery demonstrate how soft biotechnologies can play an important role in the field of artwork restoration. Although we could not completely identify all the enzymes involved in the peculiar process because of material shortage, nevertheless we have demonstrated that a complex set of enzymes, with molecular masses ranging from 20 up to 120 kDa, are secreted by the *P. stutzeri* strain A29 when the highly cross-linked glue is provided as the sole carbon source. The fact that these particular enzymes could not be detected when the bacterial cells were cultivated on fresh animal glue stresses the importance of using viable cells in challenging biorestoration cases, where the great ability of microorganims to produce specific mixes of synergic enzymatic activities ensures good results in very brief times.

2.6. BIBLIOGRAPHY

[1] March J. *Advanced Organic Chemistry* (2nd Ed.). Tokyo: McGraw-Hill Kogakusha, 1977; pp. 502-503 and 810-811.

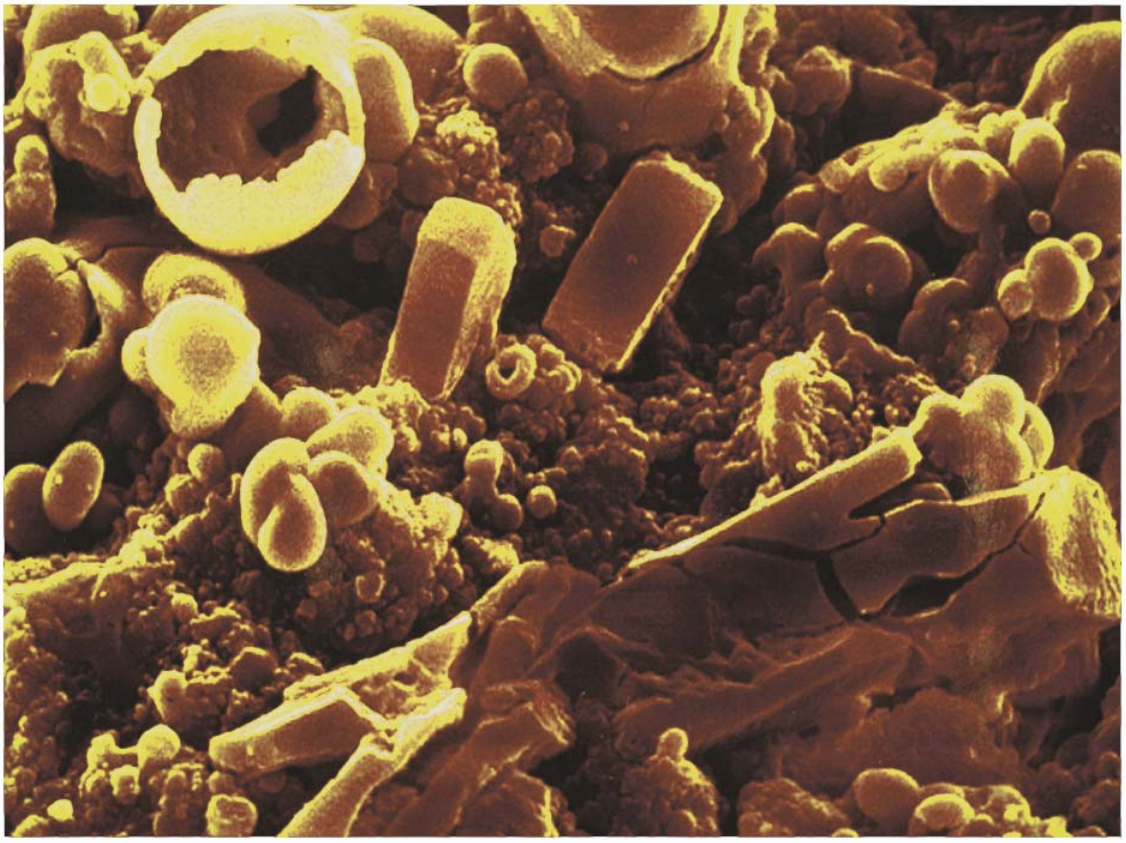
[2] Ranalli G, Matteini M, Tosini I, Zanardini E, Sorlini C. Bioremediation of cultural heritage: removal of sulphates, nitrates and organic substances. In: Ciferri O, Tiano P, Mastromei G (Editors). Of Microbes and Art. The Role of Microbial Communities in the Degradation and Protection of Cultural Heritage. New York: Kluwer Academic/Plenum Publishers, 2000; pp. 231–245.

[3] Ranalli G, Belli C, Baracchini C, Caponi G, Pacini P, Zanardini E, Sorlini C. In Saiz-Jimenez C. (Editor). Molecular Biology and Cultural Heritage. Lisse: Balkema Publ., 2003; pp. 243-246.

[4] Ranalli G, Alfano G, Belli C, Lustrato G, Colombini MP, Bonaduce I, Zanardini E, Abbruscato P, Capitelli F, Sorlini C. *J. Appl. Microbiol*. 2005; **98**: 73-83.

[5] Pessione E, Giuffrida MG, Prunotto L, Barello C, Mazzoli R, Fortunato D, Conti A, Giunta C. *Proteomics.* 2003; **3**, 1070-1076.

[6] Matsushita O, Jung CM, Katayama S, Minami J, Takahasi Y, Okabe A. *J. Bacteriol.* 1998; **181**, 923-933.



CHAPTER III:

CHARACTERIZATION OF A NEW *STENOTROPHOMONAS MALTOPHILIA* STRAIN (SelTE02) EXPLOITABLE IN BIOREMEDIATION OF CONTAMINATED SOILS

3.1. INTRODUCTION

Selenium is a metalloid element absolutely essential – in trace amounts - for life in many organisms, from bacteria to mammals [1]. In particular, selenium is integrated into proteins in the form of selenocysteine, a rare amino acid that helps promoting antioxidant reactions. These selenocysteine-containing proteins are called selenoproteins. In contrast to the relatively few eubacterial selenoproteins, there are several dozens of beneficial selenoproteins in archaea and eukaryotes [2]. On the other hand, the amino acid selenomethionine occurs mostly in plants, where it plays however no catalytic function [3]. Nevertheless, selenium has a very narrow margin of safety. In fact, the span between minimum (*i.e.* deficiency) and maximum (*i.e.* toxicity) tolerable levels of this element within the different organisms is often a matter of parts per million (ppm) or even less [4]. For instance, while in humans the recommended daily intake of selenium is in the range of 55-70 μg, at doses above 350 μg per day this element starts to become toxic [5], or possibly to exert mutagenic [6] and carcinogenic [7] effects. Therefore, selenium contamination represents an actual health concern, not only in humans, but also for livestock and wildlife. Although selenium can enter the environment as a consequence of natural erosion and leaching mechanisms affecting seleniferous parental rocks, other sources of contamination for soil, freshwater, groundwater, and sediments include, among others: coal, gold, silver, nickel, and phosphate mining, metal smelting, municipal landfills, oil transport, refining, and utilization, combustion of fossil fuels in power plants, pigments production, pharmaceutical preparations, glass manufacturing as well as repeated irrigation of selenium-rich agricultural soils [8].

Se can occur in a variety of oxidation states (selenate [Se^{VI}], selenite [Se^{IV}], elemental selenium [Se^{0}], and selenide [Se^{-II}]), whose toxicity is strictly related to their degree of water solubility and, hence, to their bioavailability [9]. Selenite (SeO_3^{2-}) and selenate (SeO_4^{2-}) are the two soluble species of selenium that can be mostly found in oxygenic habitats. Both oxyanions are toxic and tend to bioaccumulate [10]. Selenium in the form of Se^{IV} is more toxic to most organisms than selenium in the form of Se^{VI}, while Se^{0} is insoluble and thus unavailable to biological systems.

Microbial activity is thought to be the primary means by which soluble, chemically active forms of selenium can be reduced to the elemental state and thus precipitated [11]. Since elemental selenium is insoluble, this means that it is less of a threat to move up through the soil into the food chain, or down through the soil into the groundwater [12]. Microbes that can reduce selenium oxyanions are not restricted to any particular group/subgroup of prokaryotes, and numerous examples are found throughout the bacterial domain [13]. Among the various selenium species, in the last two decades selenite - due to its high toxicity - has greatly attracted attention as a potential substrate for microbial reduction. Evidence has been collected for microbial selenite reduction both in anaerobic and aerobic conditions. Various enzymatic systems, such as nitrite reductase (NIR), sulfite reductase and glutathione reductase (GR), have been proposed for, or suspected to be involved in, the reduction of selenite in bacteria. In particular, great attention has been given to dissimilatory reduction. On the other hand, evidences have been found also for microbial reduction of selenite and selenate in aerobic conditions. Nevertheless this reduction mechanism, generally referred to as 'detoxification mechanisms', is poorly understood. The role of a periplasmic dissimilatory NIR in selenite reduction has been hypothesized in *Thauera selenatis,* were mutants lacking this enzymatic activity were shown not to be able to reduce either nitrite or selenite [14]. Starting from the consideration that dissimilatory nitrite reductase activity had been identified in purple non-sulfur bacteria such as *Rhodopseudomonas palustris* [15], *Rhodobacter sphaeroides* [16], and *Rhodobacter capsulatus* [17], recently Kessi [18] has carried out investigations to clarify the possible reciprocal influence between nitrite dissimilation and selenite reduction in *Rhodospirillum rubrum* and *Rb. capsulatus*. While in *Rs*. *rubrum* the reduction pathway of selenite resulted to be completely separated from that of nitrite, in *Rb*. *capsulatus* both nitrite and selenite were metabolized simultaneously with these oxyanions interfering on the level of transport system rather than on the level of the reductase. However, selenite reductase activity seems to be constitutively expressed in either *Rb*. *capsulatus* or *Rs*. *rubrum*. On the other hand, nitrite reduction is supposed not to be constitutively expressed in these bacteria, with the nitrite reductase lacking any selenite reductase activity. Furthermore, an inducible dissimilatory sulfite

reductase has been found to be capable of reducing selenite in *Clostridium pasteurianum* [19]. Again, working with *Rb*. *capsulatus* and *Rs*. *rubrum*, Kessi [18] has demostrated that none of these phototrophic α-proteobacteria rely on a sulfite reductase for the reduction of selenite. Finally, glutathione (GSH) has been proposed to be implicated in bacterial selenite reduction, assuming a mechanism originally described by Ganther [20] for Se^{IV} reduction in mammalian systems, based on the formation of selenodiglutathione (GS-Se-SG) which is then reduced to Se^0 by GR. As GSH occurs at mM concentrations in a wide range of bacteria including cyanobacteria and α-, β-, and γ-proteobacteria [21], this compound, highly reacting with selenite [22], seems to be a possible ingredient of both dissimilatory and aerobic reduction of selenite in microbes able to synthesize this metabolite. The proof that selenite can induce GR in *Rb. Sphaeroides* and *Escherichia coli* [23] makes this hypothesis stronger. Moreover, since selenodiglutathione has been demonstrated to efficiently oxidize reduced thioredoxin in bacteria [24], it may be realistic to assume a role of thioredoxin and thioredoxin reductase (TR) in selenite reduction within those bacteria capable of synthesizing GSH. Selenite has been also shown to stimulate the synthesis of thioredoxin and TR in cultures of *E. coli* [25].

Starting from all the assumptions and considerations discussed above, in the present work we have used proteomic, microbiological and biochemical approaches for investigating the mechanisms responsible for selenite reduction in a new *Stenotrophomonas maltophilia* strain (SeITE02), a γ-proteobacteria capable of resisting to high concentrations of selenite, reducing it to non-toxic elemental selenium in aerobic conditions. Such a peculiar bacterial strain was kindly provided by Prof. Giovanni Vallini, who had previously isolated it from a mining soil contaminated with selenite [26]. Both nitrite reductase-based and glutathione-based mechanisms have been investigated by either *in vivo* or *in vitro* (subcellular protein fractions) studies.

3.2. MATERIALS AND METHODS

3.2.1. CHEMICALS

Fluka (Buchs, Switzerland) provided sodium phosphate dibasic, sodium phosphate monobasic, magnesium sulphate, manganese sulphate, calcium chloride, zinc sulphate, nitrilotriacetic acid, boric acid, cobalt sulfate, sodium molybdate, sodium fluoride, glycine, sodium dodecyl sulphate, urea, thiourea, tributylphosphine (TBP), 3-[3-cholamidopropyl dimethylammonia]-1-propansulfonate (CHAPS) and carrier ampholytes (Ampholine 3-10).

Nutrient Broth and Bacteriological Agar were furnished by Oxoid Italia Spa (Garbagnate Milanese, Italy).

Sodium chloride, ammonium sulphate, D-glucose, aluminium potassium sulfate, 2,3-diaminonaphtalene, nicotinamide adenine dinucleotide reduced form (NADH), 2,4-dinitrophenol (DNP), S-n-butyl homocysteine sulfoximine (BSO), erythromycin, ascorbic acid, hydroquinone, reduced glutathione (GSH), sucrose, Tris, mineral oil, phosphoric acid, Coomassie Brilliant Blue G-250, DL-dithiothreitol (DTT), β-mercaptoethanol, bromophenol blue, agarose, cupric sulfate as well as reduced methyl viologen were purchased from Sigma-Aldrich (Steinheim, Germany).

Riedel-de Haën provided sodium selenite, ferrous sulfate, disodium oxalate, ethylenediamine-tetraacetic acid (EDTA) sodium salt, N-(1-Naphthol)-ethylendiamine dihydrochloride (NEDA), sulphanilamide.

Merck (Darmstadt, Germany) furnished chloridric acid, citric acid, sodium hydroxide, glycerol, methanol, ethanol, acetic acid, acetone, isopropanol, cyclohexane, Triton X-100 and orthophosphoric acid.

Modified porcine trypsin (sequencing grade) was obtained from Promega (Madison, WI).

Dc-Protein Assay, acrylamide solution, *N*,*N*,*N*',*N*'-tetramethylethylenediamine (TEMED), ammonium persulfate (APS), Sypro Ruby and the immobilized pH 3-10 gradient strips (IPGs) were from Bio-Rad Laboratories (Hercules, CA).

Complete protease inhibitor was purchased from Roche Diagnostic.

The bacterium *Stenotrophomonas maltophilia* strain SeITE02 was kindly supplied by Prof. Giovanni Vallini (Laboratory of Microbiology, Scientific and Technological Department, university of Verona, Verona, Italy).

3.2.2. MICROBIOLOGICAL TECHNIQUES

3.2.2.1. Cultural media

The following cultural media were prepared as described by Frassinetti *et al.* [27] and sterilized by temperature treatment (121°C for 15 minutes):

Defined Minimal medium modified A (DM A): disodium phosphate, potassium phosphate, yeast extract, vitamin solution, Wolfe solution.

Defined Minimal medium modified B (DM B): disodium phosphate, potassium phosphate, ammonia sulphate, yeast extract, vitamin solution, Wolfe solution.

Agarized Nutrient Broth: nutrient broth, agar.

Vitamin solution, Wolfe mineral solution and carbon substrates were sterilized by filtration with cellulose acetate filters (0.2 µm, Millipore).

3.2.2.2. Determination of nitrite LD_{50}

Stationary phase pre-inoculums were inoculated with 250 mL of DM A medium supplemented or not with sodium nitrite (final concentrations 2.9, 7.2, 11.6, 18.7, 29.0 and 72.5 mM). Incubation took place at 28°C under fast agitation (180 rpm) in an orbital shaker, and the bacterial growth was monitored by counting the colony-forming units (CFU) in agarised Nutrient Broth plates seeded with aliquots of the bacterial cultures. Each analysis was performed in duplicate.

3.2.2.3. Cultures with nitrite: nitrite pre-induction and nitrite-selenite mixed cultures

First, the effect of nitrite pre-induction on the ability of bacterial cells to grow on and reduce selenite was evaluated. Stationary phase pre-inoculums were inoculated with 250 mL of DM A medium supplemented or not with sodium nitrite at a final concentration of 10 mM. After 24-hour incubation 25 mL of each culture were recovered and inoculated with a smaller volume (100 mL final volumes) of fresh medium, and this procedure was repeated twice. Eventually, both the control

and the pre-induced bacterial cells were inoculated with DM A medium supplemented with 0.5 mM sodium selenite, and incubation was allowed to proceed at 28 °C under agitation (180 rpm). The bacterial growth was monitored by counting the CFU on agarised Nutrient Broth plates seeded at increasing incubation times. The selenite content of each culture aliquot taken at various incubation times was also determined as described later in this section.

With another approach, the effects of the simultaneous presence of nitrite and selenite on the viability and the activity of the strain SeITE02 were investigated. In this case, stationary phase pre-inoculums were inoculated with 250 mL of DM A medium supplemented or not with selenite (0.2 mM final concentration) or nitrite and selenite together (18.7 mM and 0.2 mM final concentrations, respectively). Incubation took place at 28 °C under fast agitation (180 rpm). The bacterial growth was monitored by counting the CFU on agarised Nutrient Broth plates seeded at increasing incubation times. The culture samples collected at different times were also analysed for their selenite content.

3.2.2.4. Cultures under nitrite intake suppression

Stationary phase pre-inoculums were inoculated with 100 mL of DM A medium containing selenite (0.2 mM final concentration) and supplemented or not with the nitrite intake inhibitor 2,4-dinitrophenol (DNP), which was added at the three final concentrations of 0.1, 0.5 and 1.0 mM. Incubation took place at 28 °C under fast agitation (180 rpm). Culture samples were withdrawn at various incubation times and the residual selenite was quantified as described later in this section.

3.2.2.5. Cultures under glutathione synthesis suppression

Stationary phase pre-inoculums were inoculated with 100 mL of DM A medium supplemented or not with S-n-butyl homocysteine sulfoximine (0.5, 1.0 or 3.0 mM final concentrations). Each culture was then supplemented with selenite (0.2 mM final concentration), which was added either at the beginning of the bacterial growth or during the exponential growth (after 18-hour incubation) or the stationary phase (42-hour incubation). Incubation took place at 28 °C under fast agitation (180 rpm). Culture samples were taken at various incubation times and

the selenite content determined by the spectrophotometric assay subsequently described.

3.2.2.6. Preparation of bacterial cells for enzymatic assays and proteomic studies

Stationary phase pre-inoculums were infused with 400 mL of DM A medium supplemented or not with 16 mM sodium nitrite or 0.2 mM sodium selenite. Incubation took place at 28 °C under fast agitation (180 rpm) in an orbital shaker until reaching the stationary phase; after that the bacterial cells were recovered by centrifugation (10 000 x g, 10 minutes) and further processed.

3.2.3. SEPARATION OF THE SUB-CELLULAR PROTEIN FRACTIONS

The subcellular protein fractions (cytoplasm, periplasm and membrane proteins) were separated in order to localize the enzyme(s) responsible for selenite reduction at a subcellular level. The bacterial cells previously recovered by centrifugation were washed twice with 400 mL of physiological solution (0.9% NaCl), after that the cells were pelletted again and subjected to periplasmic protein solubilisation as previously described by Osborn and Munson [28]. In practice, cells were suspended and left for 5 minutes in an ice-cold buffer solution containing 30 mM Tris-HCl pH 8.0 and 30% sucrose, after that lysozime was added to a final concentration of 0.4 g/L. The solution was kept in ice for another 2 minutes, after that 3 volumes of an ice-cold solution of 15 mM EDTA pH 8.0 were slowly added in an interval of 10 minutes. Eventually, the spheroplasts (*i.e.* cells without wall) were pelletted by centrifuging at 25 000 x g for 20 minutes and then suspended in 10 mL of a solution containing 50 mM NaCl and one tablet of Complete protease inhibitor, while the supernatant, which contained the periplasmic protein fraction, was recovered, filtered (0.2 µm, Millipore) and immediately frozen in liquid nitrogen and stored at -20°C. The spheroplasts disruption was achieved by means of a sonicator equipped with a steel tip, by repeating 7 sonication cycles (40 seconds sonication alternated with 40 seconds of rest in ice), while keeping the samples always in ice. After sonication, the solution was centrifuged at 200 000 x g for 75 minutes. The soluble cytoplasmic

proteins present in the supernatant were recovered, filtered and immediately frozen in liquid nitrogen, while the membranes fragments, visible as a brown pellet at the bottom of the centrifuge tubes, were solubilised in 10 mL of a 50 mM phosphate buffer solution pH 7.4 containing 0.5% Triton X-100, frozen and stored at -20 °C.

3.2.4. SPECTROPHOTOMETRIC ASSAYS

3.2.4.1. Selenite content determination

Selenite concentration in cultural supernatants and reaction mixtures was determined spectrophotometrically according to Kessi *et al.* [29]. First, 10 mL of 0.1 M HCl, 0.5 mL of 0.1 M EDTA, 0.5 mL of 0.1 M NaF, and 0.5 mL of 0.1 M of disodium oxalate were mixed in a 50 mL glass bottle. Then a volume of 250 µL of sample (containing approximately 100 to 200 nmol of selenite and centrifuged so as to eliminate the particles in suspension) was added to the solution, and finally 2.5 mL of a solution containing 0.1% 2,3-diaminonaphthalene in 0.1 M HCl was admixed. The bottles were incubated at 40°C for 40 minutes and cooled to room temperature for 15 minutes. The selenium-2,3-diaminonaphthalene complex was extracted with 6 mL of cyclohexane by vigorously shaking the two-phase system for about 1 min in a separating funnel, and the absorbance at 377 nm of the recovered organic phase was measured. All the manipulations were done in the dark, due to the complex light-sensitivity. Calibration curves were constructed by using 0, 50, 100, 150, 200 nmol of selenite dissolved in the same solution as that of the sample.

In some cases, selenite conversion to elemental selenium was followed only at a semi-quantitative level by reading the solution absorbance at 415 nm, a wavelength where the red colloidal suspension of elemental selenium absorbs most. In this event, the kinetics of red elemental selenium formation were obtained by subtracting the absorbance variations observed in the various negative controls (non enzymatic reduction) to that measured in the complete reaction mixture.

3.2.4.2. Nitrite content determination

Nitrite was quantified using the method of Arnold and colleagues [30]. In practice, the sample was centrifuged in order to remove the particles in suspension (in the case of cell cultures), after that the supernatant was diluted to obtain final nitrite concentrations in the range 0.1-1.0 µg/mL. 800 µL of each supernatant were taken and mixed with 200 µL of a solution obtained by mixing 1 volume of the NEDA solution (5.6 mg/mL of N-(1-Naphthyl)-ethylenedyamine dihydrochloride in H_2O milli-Q) with 10 volumes of the sulphanilamide solution (8 mg/mL sulphanilamide in 1.8 M phosphoric acid). The reactants were ice-cold, and the reaction took place in ice for 5 minutes before reading the absorbance at 545 nm. The calibration curve was prepared with the following nitrite concentrations: 0.2, 0.4, 0.6, 0.8 and 1.0 µg/mL.

3.2.4.3. Protein content determination

Protein concentration was determined by means of the Dc-Protein Assay (Bio-Rad, Hercules, CA) as recommended by the producer. Three dilutions of the protein extract were prepared in duplicate for each sample, so as to have concentrations ranging approximately in the interval 0.1-2.0 mg/mL. The reference was the buffer used for protein solubilisation, while the calibration curve was obtained by diluting a 2 mg/mL concentrated bovine serum albumin standard solution to the following concentrations: 0.4, 0.8, 1.2, 1.6 and 2.0 mg/mL. Measurements were done by reading the absorbance at 750 nm.

3.2.5. IN VITRO ENZYMATIC ASSAYS

3.2.5.1. Subcellular localization of the selenite reduction activity and determination of the optimal pH and electron donor

The various protein fractions were incubated in presence of selenite and various electron donors (NADH, NADPH, reduced ascorbate, reduced hydroquinone) at various pH values (pH 5.0, 6.0, 6.5, 7.0 and 8.0). The buffers used were either McIlvaine (250 mM, pH 5.0, 6.0 and 6.5) or phosphate buffers (250 mM, pH 7.0 and 8.0). In order to avoid possible interference due to bacterial growth in the assay environment, the protein samples (periplasm and cytoplasm fractions), the

buffers and the electron donor solutions were sterilized by filtration, while in the case of membrane proteins the bacterial growth was prevented by the addition of 250 µg/mL erythromycin. The reaction was carried out in 96-well microtitration plates in a total volume of 300 µL, and for each sample 1 complete reduction system and 3 negative controls were prepared as follows:

- Complete reduction system:
  - (+): 150 µL of protein sample (P) + 1mM electron donor (D) + 0.5 mM sodium selenite (S) + buffer solution (to 300 µL total volume)
- Negative controls:
  - (-P): 150 µL H_2O + 1mM electron donor + 5 mM sodium selenite + buffer solution (to 300 µL total volume)
  - (-S): 150 µL sample + 1mM electron donor + buffer solution (to 300 µL total volume)
  - (-D): 150 µL of sample + 5 mM sodium selenite + buffer solution (to 300 µL total volume).

Each mixture was prepared in duplicate, and the formation of elemental selenium was measured at a semi-quantitative level by reading the absorbance at 415 nm, the wave length of maximum absorption of the red colloidal suspension of selenium.

3.2.5.2. Selenite reduction by GSH–NADPH coupled systems

In this case the reaction was carried out in 1.5 mL plastic vials, and the conversion of selenite was monitored directly by quantifying the residual selenite by means of the 2,3-diaminonaphthalene-based method. The following reaction conditions were used: 1.3 mL total volume, 400 µL of cytoplasm protein extracts, McIlvaine buffer pH 6.0, 0.5 mM NAPDH as electron donor, 0.5 mM GSH and 0.25 mM sodium selenite. In order to avoid accidental GSH oxidation the solutions were deoxygenated by fluxing argon prior to use. For each sample 1 complete reduction system and 3 negative controls were prepared as follows:

- Complete reduction system:
  - (+): 400 µL of sample (P) + 0.5 mM NADPH (D) + 0.5 mM GSH (G) + 0.25 mM sodium selenite (S) + buffer solution (to 1.3 mL total volume)
- Negative controls:

- (-P): 400 µL of H_2O + 0.5 mM NADPH + 0.5 mM GSH + 0.25 mM sodium selenite + buffer solution (to 1.3 mL total volume)
- (-G): 400 µL of sample + 0.5 mM NADPH (D) + 0.25 mM sodium selenite (S) + buffer solution (to 1.3 mL total volume)
- (-PG): 100 µL of H_2O + 0.5 mM NADPH + 0.25 mM sodium selenite + buffer solution (to 300 µL total volume)

The cytoplasm protein extracts were prepared starting from cultures grown either in absence of sodium selenite or in presence of 0.2 mM or 0.5 mM sodium selenite. The protein content of the various extracts was measured in order to level the protein concentrations, so as to properly quantify potential activity differences among the samples. For this purpose the most diluted samples were concentrated by centrifugation with Microcon concentrators (5 kdD MWCO) until the concentrations resulted equalized. For each sample, aliquots of 250 µL were withdrawn from the reaction environment at different times (5, 10, 15, 25 and 40 minutes) in order to measure the selenite content. The proteins were removed by cold-acetone precipitation, after that the supernatant was allowed to evaporate in a speed-vac device. Finally, the residual selenite was recovered and quantified as described in 3.2.4.1.

3.2.5.3. Nitrite reduction assays

Nitrite reduction activity was assayed in all of the subcellular fractions, either by monitoring methyl viologen oxidation or by measuring the residual nitrite present in solution.

In the first event, the assay was carried out in microtitration plates in a total volume of 200 µL. The reactions took place at room temperature and at different pH values, using either McIlvaine buffers (20 mM final concentration, pH 5.0, 6.0 and 6.5) or phosphate buffers (20 mM final concentration, pH 7.0 and 8.0). For each subcellular fraction (cytoplasm, periplasm and membrane) obtained from cells pre-induced with nitrite and for each pH value the following reaction mixtures were analyzed:

- Complete reduction system:
  - (+): 50 µL of sample (P) + 2.5 mM (final concentration) methyl viologen (D) + 2.5 mM nitrite (N) + 20 mM buffer solution (to 200 µL total volume)

- Negative controls:
  - (-P): 2.5 mM (final concentration) methyl viologen (D) + 2.5 mM nitrite (N) + 20 mM buffer solution (to 200 µL total volume)
  - (-N): 50 µL of sample (P) + 2.5 mM (final concentration) methyl viologen (D) + 20 mM buffer solution (to 200 µL total volume)
  - (-PN): 2.5 mM (final concentration) methyl viologen (D) + 20 mM buffer solution (to 200 µL total volume).

Since reduced methyl viologen strongly absorbs at 655 nm, while the oxidized form doesn't, nitrite reduction was followed indirectly by monitoring the net decrease of absorbance at 655 nm due to methyl viologen oxidation. Such net decrease was calculated by subtracting the absorbance variations observed in the various negative controls (methyl viologen oxidation caused by abiotic reactions or by enzymatic reduction of something different from nitrite) to that measured in the complete reaction mixture.

In the case of direct quantification of nitrite, the same experimental setup (*i.e.*, with the same complete mixture and negative controls) was performed in a total volume of 1.3 mL, from which aliquots of 200 µL were withdrawn at pre-fixed incubation times. In this case NADH (0.5 mM final concentration) was used as electron donor, while the buffer was phosphate pH 7.0 (20 mM final concentration). Nitrite was supplemented at a final concentration of 90 µM, so as to have a final concentration compatible with the calibration curve (0.2-22.0 µM) after diluting with water by a factor of 4 (800 µL). Nitrite content was then determined as described in the paragraph 3.2.4.2.

3.2.6. PROTEOMICS ANALYSIS

3.2.6.1. Sample preparation

Cytoplasm protein extracts from three different cell cultures were analyzed:

- cells cultured in DM A medium
- cells cultured in DM A medium + 0.2 mM sodium selenite
- cells cultured in DM A medium + 16 mM sodium nitrite.

Cytoplasm proteins, extracted as described previously, were first quantified and then precipitated by incubation with trichloroacetic acid (TCA, 15% final

concentration) for 90 minutes in ice. The protein pellet formed after 15-minute centrifugation at 4°C and 20 000 x g was washed twice with cold acetone (-20°C), after that it was solubilised in the IEF buffer (7M urea, 2M urea, 3% CHAPS, 40 mM Tris, 5 mM TBP) so as to obtain a final protein concentration of 2 mg/mL. Cysteine alkylation was performed by adding 10 mM free acrylamide in presence of 5 mM fresh TBP, and the reaction continued for 1 hour. The residual acrylamide monomers were quenched by adding 10 mM DTT (final concentration). Finally 0.5% carrier ampholytes pH 3-10 and traces of bromophenol blue were added just before loading the samples.

3.2.6.2. 2-DE analysis

For each sample 5 large-size (18 x 20 cm) replica maps were prepared. Proteins were loaded on 18-cm long non linear IPGs 3-10 by passive rehydration with a total volume of 450 µL, so that about 900 µg of total protein were analyzed in each map. After a 8-hour rehydration the focusing was performed by applying a voltage program as follows: from 100 to 1000 V linearly in 5 hours; 1000 volts constant for 5 hours; logarithmic voltage ramp from 1000 V to 10 000 V in 1 hour; 10 000 V constant until reaching 75 000 Vh. At the end of the focusing the strips were incubated with the equilibration solution (6M urea, 2% SDS, 20% glycerol in 185 mM Tris buffer pH 8.8) for 25 minutes under gentle agitation in order to saturate the proteins with SDS prior to the second dimension. The latter was performed on gels with a linear polyacrylamide gradient from 10% to 20% T, and the electrophoretic run took place in a Protean® Plus Dodecacell (Bio-Rad, Hercules, CA) with glycine-Tris buffer pH 8.3 (192 mM glycine, 0.1% SDS and Tris to pH 8.3). A current program was set in this case: 50 total mA for 1 hour; 100 total mA for 1 hour; 150 total mA overnight; 250 total mA until the end of the run.

All the gels were then fixed with a solution containing 40% methanol and 7% acetic acid for 1 hour before staining with the Sypro Ruby fluorescent stain. Image digitisation was carried out with the VersaDoc 4000 device (Bio-Rad, Hercules, CA), and a three-way comparative analysis was performed simultaneously with the 15 maps of the three samples, considering the proteins extracted from the cells grown in simple DM A medium as the reference for the other “treated” samples

(DM A + 0.2 mM selenite and DM A + 16 mM nitrite). The spots of interest (*i.e.* those modulated by at least two-fold) were analyzed by MS/MS.

3.2.6.3. MS/MS analysis

The protein spots were carefully excised from stained gels and subjected to in-gel trypsin digestion according to Shevchenko *et al.* [31] with minor modifications. The gel pieces were swollen in a digestion buffer containing 50 mM NH_4HCO_3 and 12.5 ng/µL of trypsin (modified porcine trypsin, sequencing grade, Promega, Madison, WI) in an ice bath. After 30 min the supernatant was removed and discarded, 20 µL of 50 mM NH_4HCO_3 were added to the gel pieces and digestion continued at 37°C overnight. The supernatant containing tryptic peptides was dried by vacuum centrifugation. Prior to mass spectrometric analysis, the peptide mixtures were redissolved in 10 µL of 5% formic acid.

Peptide mixtures were then separated by using a nanoflow-HPLC system (Ultimate, Switchos, Famos, LC Packings, Amsterdam, Netherlands). A sample volume of 10 µL was loaded by the autosampler onto a homemade 2-cm fused silica precolumn (75 µm I.D.; 375 µm O.D.; Resprosil C18-AQ, 3 µm (Ammerbuch-Entringen, DE)) at a flow rate of 2 µL/min. Sequential elution of peptides was accomplished using a flow rate of 200 nL/min and a linear gradient from 100% of solution A (2% acetonitrile, 0.1% formic acid) to 50% of solution B (98% acetonitrile, 0.1% formic acid) in 40 minutes over the precolumn in-line with a homemade 10-15 cm resolving column (75 µm I.D.; 375 µm O.D.; Resprosil C18-AQ, 3 µm, Ammerbuch-Entringen, Germany).

Peptides were eluted directly into a High Capacity ion Trap HCTplus (Bruker-Daltonik, Germany). The capillary voltage was 1.5-2 kV and a dry gas flow rate of 10 L/min was used with a temperature of 200°C. The scan range used was from 300 to 1800 m/z. Protein identification was performed by searching in the National Center for Biotechnology Information non-redundant database (NCBInr) using the Mascot program (http://www.matrixscience.com). The following parameters were adopted for database searches: complete carbamidomethylation of cysteines and partial oxidation of methionines, peptide Mass Tolerance ± 1.2 Da, Fragment Mass Tolerance ± 0.9 Da, missed cleavages 2. For positive identification, the score of the result had to be over the significance threshold ($P < 0.05$).

The MS analysis was carried out in the laboratory of Prof. L. Zolla (Laboratory of Proteomics, Department of Environmental Sciences, Tuscia University, Viterbo, Italy).

3.3. RESULTS

3.3.1. SELENITE REDUCTION IN VIVO

3.3.1.1. The effect of nitrite on cell growth and selenite reduction in vivo

As reported by some authors, nitrite reductase can play a role in selenite reduction, as in the case of *Thauera selenatis* [14]. Moreover, Vallini and coworkers have observed that a nir- mutant of *Rhizobium hedysari* strain HCNT1 (*i.e.* with a non-functioning nitrite reductase) was initially unable to reduce selenite, although the ability was slowly recovered with longer culture times [32]. In this work we have taken into consideration the possible interactions between the selenite and the nitrite reduction pathways, and in particular we have studied the effect of nitrite pre-induction on cell ability to reduce selenite and the effect of nitrite presence during selenite reduction performed either by living cells (*in-vivo* experiments) or by enzyme extracts (*in-vitro* experiments).

In the first place we have tried to determine the specific nitrite lethal dose 50 (LD_{50}) for our strain and, possibly, a minimum inhibitory concentration (MIC). *S. maltophilia* strain SeITE02 was able to grow in presence of all the nitrite concentrations assayed, although a sensible decrease in total cells was observed starting from the concentration of 29 mM, as shown in figure 3.1. Growth tests were carried out in a total volume of 100 mL of DM A medium added with increasing nitrite concentrations. Since in presence of 29 mM nitrite the total amount of cells present in the stationary phase was about half of that present in absence of nitrite (2.44×10^9 and 4.84×10^9, respectively), this concentration was considered an optimum approximation of the real LD_{50} value. Instead no MIC value could be determined at this dose, neither at considerably higher concentrations (up to 72 mM), which still allowed for a slow moderate cell growth.

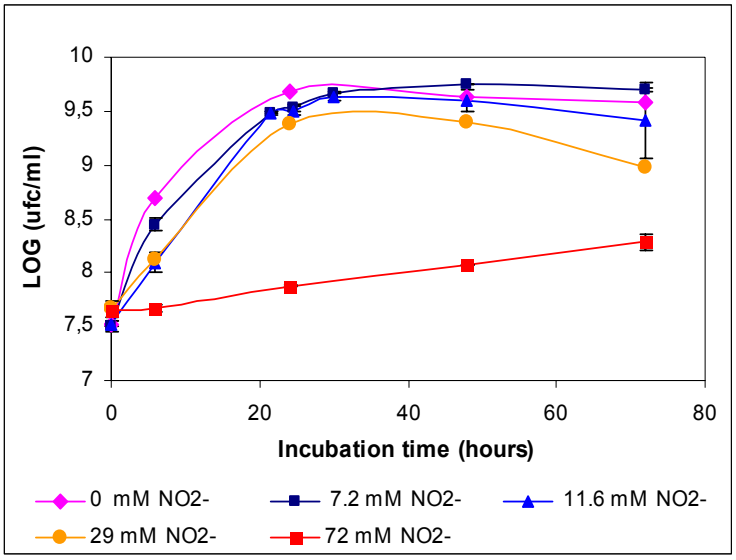


Figure 3.1. Nitrite effect on bacterial growth: determination of LD_{50} and MIC. Cells were cultured in DM A medium supplemented with the nitrite concentrations reported in the graph legend. Vertical bars indicate the standard deviations as determined from measurements in duplicate.

With the aim of evaluating whether nitrite pre-induction had any effect on the bacterial ability to grow on and reduce selenite, two pre-inocula were prepared in DM A medium with or without nitrite, which was added at a final concentration of 11.6 mM, a level sufficient to induce nitrite-related enzymes but not so high to inhibit cell growth. After three 24-hour cycles of growth the two pre-inocula were infused with fresh media supplemented with 0.2 mM selenite, and the cell rate growth was monitored by counting the CFU formed on agarised Nutrient Broth plates. As one can observe from figure 3.2, nitrite pre-induction had no appreciable effect either on bacterial cell growth or on the rate of selenite reduction. The interaction between nitrite and selenite has been also studied by performing mixed-cell cultures in presence of both nitrite and selenite together, in order to verify whether there was competition between the two oxyanions for the same enzymes. For this purpose three cultures were prepared, differing for the nitrogen source: (i) only yeast extract (YE, DM A medium), (ii) ammonium sulphate (DM B medium), and (iii) sodium nitrite, added to such a final concentration as to provide the same nitrogen amount as that provided with DM B medium. A small amount of yeast extract was necessary in all of the cultures, because in its absence no cell growth could be attained. Selenite was added to the three cell cultures at a final concentration of 0.2 mM.

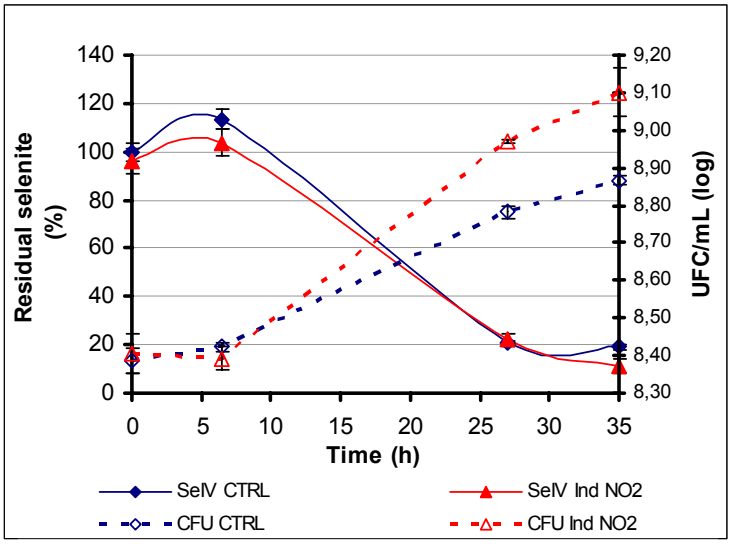


Figure 3.2. Comparison of growth rates (dotted lines) and selenite reduction activities between control cells (CTRL) and cells pre-induced with nitrite (Ind NO2). Selenite is expressed as a percentage of the initial amount (0.2 mM). Vertical bars indicate the standard deviations as determined by measurements in duplicate.

As shown in figure 3.3, the nitrogen source had a considerable impact on both the total amount of cells obtained in the stationary phase and the selenite reduction rates. In fact, quite unexpectedly, the best cell growth was observed when only YE was provided as nitrogen source, while both nitrite and ammonium affected moderately the cell growth rate, and more consistently the total amount of cells present in the stationary phase. In this concern, ammonium performed worst, allowing for a total bacterial population ten times minor (1 logarithmic point) as compared to the control. In particular, nitrite allowed for a higher cell population in the stationary phase as compared to ammonium, although a more consistent cell death was observed after 72 hours. However, the most impressive difference was observed in the selenite reduction kinetics, which was really rapid in the case of yeast extract and ammonia, where 90% of the selenite initially present had already been reduced after 20-hour incubation (*i.e.* during the exponential growth phase), while being very slow and partial when nitrogen was supplied in the form of nitrite. In fact, in this case, the reduction started at the end of the exponential growth phase, and proceeded slowly during the stationary phase, until reaching a total conversion of about 20% of the initial selenite after a period of 50 hours. Such results were also evident from the intense red colour conferred by the elemental selenium to the culture media.

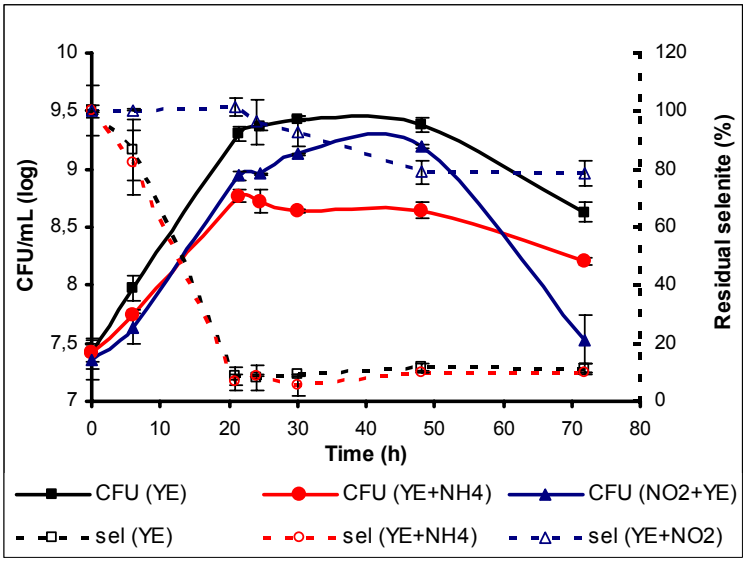


Figure 3.3. Growth rates (CFU, full lines) and selenite reduction kinetics (sel, dotted lines) of cells cultured in DM A media added with three different nitrogen sources: sole yeast extract (YE), YE + ammonium (YE+NH4) and YE + nitrite (YE+NO2). Vertical bars indicate the standard deviations as determined by measurements in duplicate.

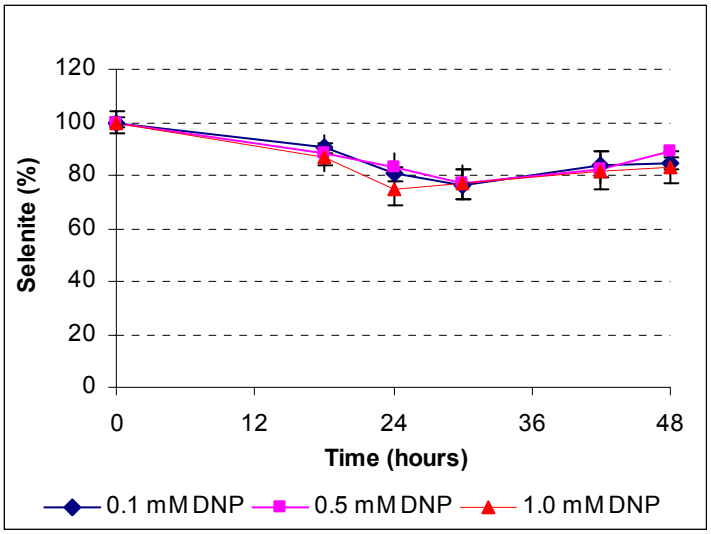
Figure 3.4 displays a picture of the three cell cultures after 24-hour incubation, where selenite has already been extensively converted into elemental selenium only in presence of yeast extract or ammonia (intense red colour), while in the case of nitrite no selenite reduction has occurred (unstained medium).

DM A medium + NO_2 (+ selenite 0.2 mM) | **DM B medium (+ selenite 0.2 mM)** | **DM A medium (+ selenite 0.2 mM)**

Figure 3.4. Picture of the three cell cultures in presence of three different nitrogen sources: sole yeast extract (CTRL, on the right), yeast extract + ammonium sulphate (in the centre) and yeast extract + nitrite (on the left).

As cell pre-treatment with nitrite had no effect on cell growth and selenite reduction rate, we have hypothesised that the point of interaction between the two reduction pathways is at the transport level rather than at the reduction level. In fact the nitrite reduction system is usually inducible, thus, if it was directly involved in selenite reduction, cell adaptation to nitrite prior to incubation with selenite should confer some advantage in terms of growth rate and selenite reduction efficiency. In order to assess whether nitrite delayed selenite reduction onset by competition at the transport level, we performed an experiment where cells were grown in presence of selenite and a specific inhibitor of nitrite intake (2,4-dinitrophenol, DNP) [33, 34]. As displayed in figure 3.5, in presence of DNP selenite reduction was very limited, reaching a total conversion of about 20% of the initial amount of selenite after 24 hours, while in the remaining incubation time no further conversion of selenite to elemental selenium took place.

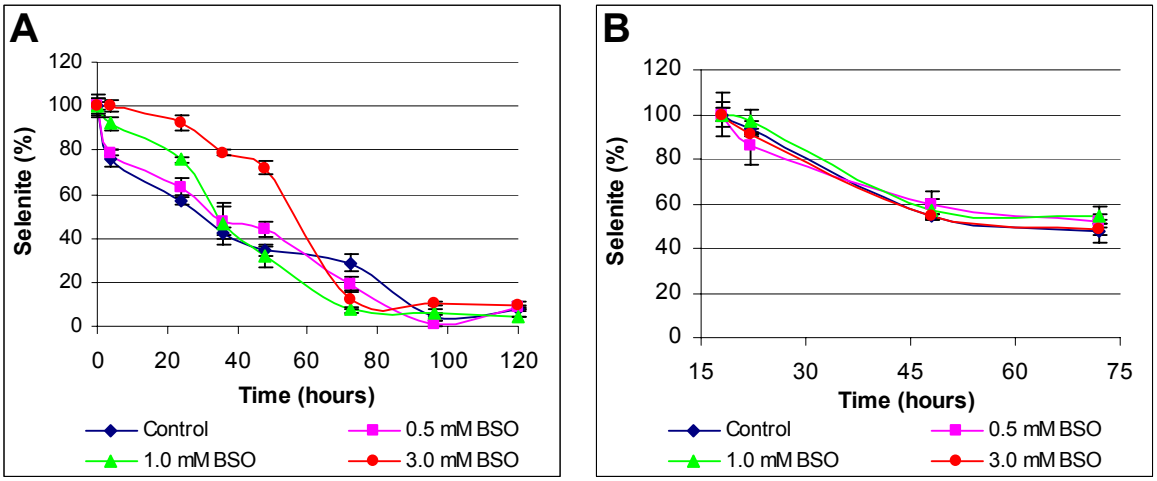
Figure 3.5. Selenite reduction in presence of DNP (0.1, 0.5 and 1.0 mM).



The inhibitory effect was not dependent on DNP concentration (0.1 - 1.0 mM) thus suggesting that the lowest concentrations was sufficient to achieve a saturation effect. These results clearly suggest that selenite enters the cells *via* the nitrite absorption pathway. Further *in vitro* experiments have been performed in order to better understand the role of other nitrite-related enzymes (e.g. nitrite reductase). The results of these assays will be reported in the next paragraphs.

3.3.1.2. The role of GSH in selenite reduction in vivo

As already discussed in the introduction, GSH has been proposed to be implicated in bacterial selenite reduction, based on the formation of selenodiglutathione (GS-Se-SG), which is then reduced to elemental selenium (Se^0) by glutathione reductase [20]. In order to evaluate whether our strain uses a detoxification mechanism based on reduced thiols, bacterial cells were cultured in presence of a well known inhibitor of GSH synthesis, and selenite reduction was monitored by direct measurements of residual selenite. The inhibitor was S-n-butyl homocysteine sulfoximine (BSO), which has already been used by Kessi to evaluate the influence of GSH level on bacterial growth and selenite reduction rate [18]. BSO was added to three different final concentrations (0.5, 1.0 and 3.0 mM), at the beginning of cell growth, while selenite (0.2 mM final concentration) was admixed at different incubation times (0, 18 and 42 hours for, respectively, lag, exponential and stationary phase). As clearly shown in figure 3.6, BSO considerably inhibited selenite reduction when selenite was admixed at the beginning of the experiment (panel A), while no difference between control and cultures amended with BSO could be observed when selenite was added during the exponential growth (18 h, panel B) or during the stationary phase (42 h, panel C). Conversely, in the first event, BSO significantly delayed the onset of selenite reduction until the stationary phase, when also in presence of the highest BSO concentration the selenite content diminished quickly to levels comparable with those achieved in cultures without BSO. This suggests that GSH plays its role during the very first steps of cell response to the toxic oxyanion, which are likely to be the most important of the entire process, since full conversion to elemental selenium can be achieved only when selenite is present since the beginning of cell growth.



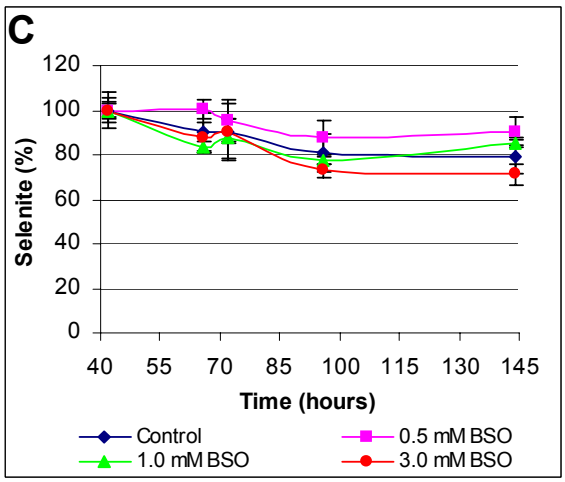


Figure 3.6. Selenite reduction in presence of BSO (0.5, 1.0 or 3.0 mM). Selenite was added at different growth stages: lag phase (0 hours, A), exponential phase (18 hour, B) or stationary phase (42 hours, C). Residual selenite is expressed as percentage of the initial amount. Vertical bars indicate the standard deviations as determined by measurements in duplicate.

3.3.2. SELENITE REDUCTION IN VITRO

3.3.2.1. Subcellular localization and optimal reaction conditions

In order to better understand the biological mechanism of selenite reduction we performed protein extraction in native conditions from cells grown in absence of selenite, followed by separation of the subcellular protein fractions and enzymatic activity assays.

In the first place native proteins were fractionated according to their sub-cellular localization, in the attempt to determine where selenite reduction took place in *S. maltophilia* strain SeITE02. Periplasmic proteins were first isolated as previously described by Osborn and Munson [28], after that the spheroplast were disrupted by sonication and the membrane proteins were separated from the cytoplasm

ones by ultra-centrifugation. First the three protein extracts were incubated with selenite solutions buffered at various pH values (5.0, 6.0, 6.5, 7.0 and 8.0) in presence of NADH as electron donor, and for each sample three negative controls were prepared, in order to evaluate the single contribution given by each factor (protein extracts, selenite and electron donor) to the red-coloured suspension developed in the reaction environment. As shown in figure 3.7 panel A, the unique sub-cellular fraction capable of reducing selenite was cytoplasm, while neither membranes nor periplasm developed any red suspension of elemental selenium after a 24-hour incubation. The optimal conditions seemed to be at pH 6.0, where there was the maximum colour difference between the reaction mixtures and the negative controls, as quantified by reading the absorbance at 415 nm. Some colour developed also in absence of the electron donor at pH 6.0 and 6.5, but not in absence of the protein extracts, thus demonstrating that the reaction is specifically catalyzed by one or more cytoplasmic enzymes. The graph displayed in figure 3.7 panel B shows selenite reduction kinetics in presence of cytoplasm extracts at different pH values as measured by reading absorbance at 415 nm. Net enzymatic activities were calculated by correcting the values read in the complete reaction mixtures (+) with those registered in the various negative controls (-P, -S, -D). The histogram displayed in panel C compares the enzymatic activities of the various sub-cellular fractions at different pH values. Such activities are expressed as percentages of the highest one (cytoplasm pH 6.0, 100%), and were calculated by comparing the slopes of the regression lines displayed in panel B (no detectable activity was considered to give slope 0). After having determined that selenite reduction takes place in cytoplasm and that reduction proceeds better at pH 6.0, also other electron donors were evaluated. Cytoplasm proteins were incubated with NADH, NADPH, reduced ascorbate or reduced hydroquinone as electron donors, in presence of selenite at pH 6.0. The formation of elemental selenium was followed by reading the absorbance at 415 nm. Figure 3.8 shows the results of this experiment: the curves displayed in panel A represent the reaction kinetics with the various electron donors, while in panel B the relative catalytic efficiencies are expressed as percentages with respect to the best electron donor (hydroquinone).

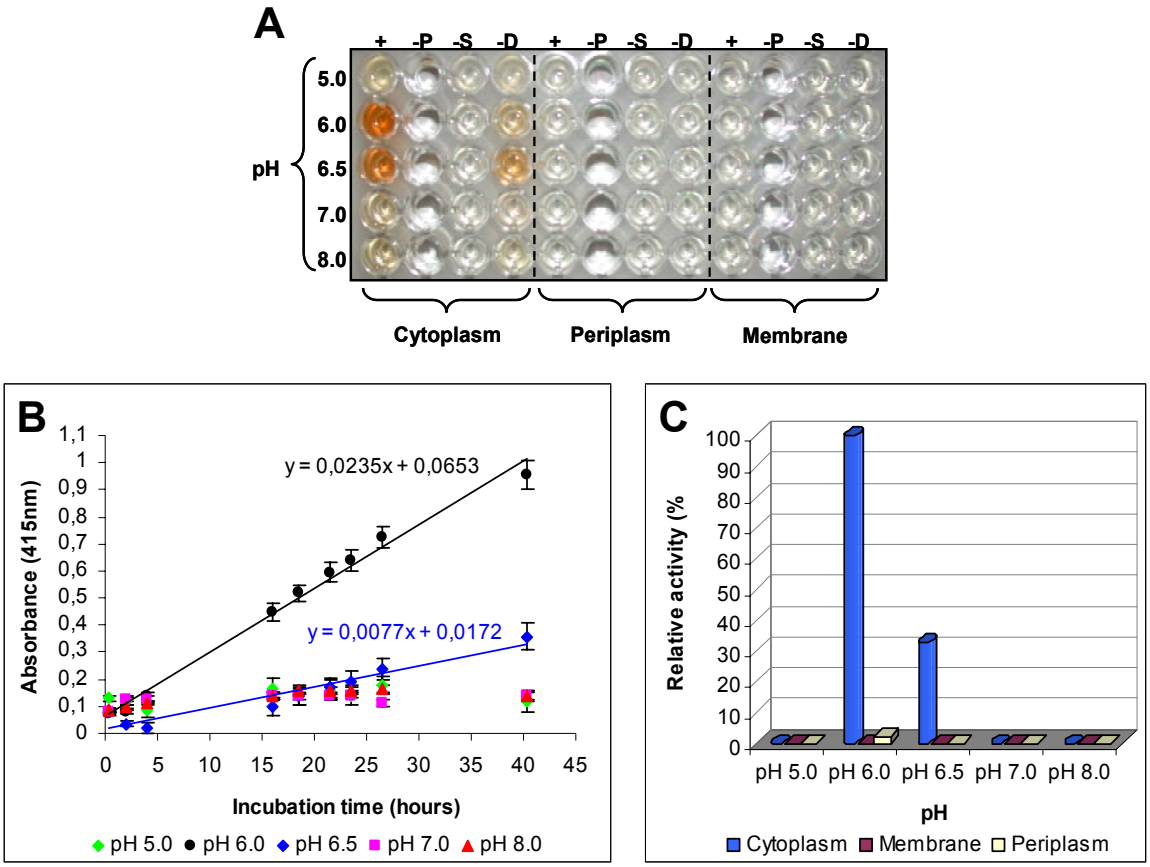


Figure 3.7. *In vitro* selenite reduction in presence of sub-cellular fractions (cytoplasm, membrane and periplasm) and NADH (electron donor) in different pH conditions. (A) picture of the microtitration plate used for the assay; note the presence of red elemental selenium in the wells containing cytoplasm at pH 6.0 and, to a lower extent, at pH 6.5. (B) Selenite reduction kinetics, as measured by reading the absorbance at 415 nm (*i.e.* the wavelength where the red elemental selenium generated by selenite reduction absorbs most) of the reaction mixtures displayed in panel A. Vertical bars indicate the standard deviations as determined by measurements in duplicate. (C) Histogram representing the catalytic activities of the various sub-cellular fractions at different pH values, expressed in percentage with respect to the highest one (cytoplasm pH 6.0, 100% activity).

The various kinetics values were calculated by subtracting the absorbance of the control (without enzyme extract) from that of the complete reaction mixture, containing cytoplasm proteins, selenite and the electron donor. The relative activities shown in panel B were calculated by comparing the slopes of the linear regression curves approximating the real kinetics shown in panel A. Although hydroquinone seemed to be the best donor, allowing for a much more rapid

appearance of the red colloidal suspension of selenium, NADPH was preferred, since in presence of hydroquinone also non specific, non enzymatic conversion of selenite occurred. On the other hand NADPH allowed for a selenite conversion twice more rapid than that observed with NADH, without giving abiotic selenite reduction in absence of the enzymatic component.

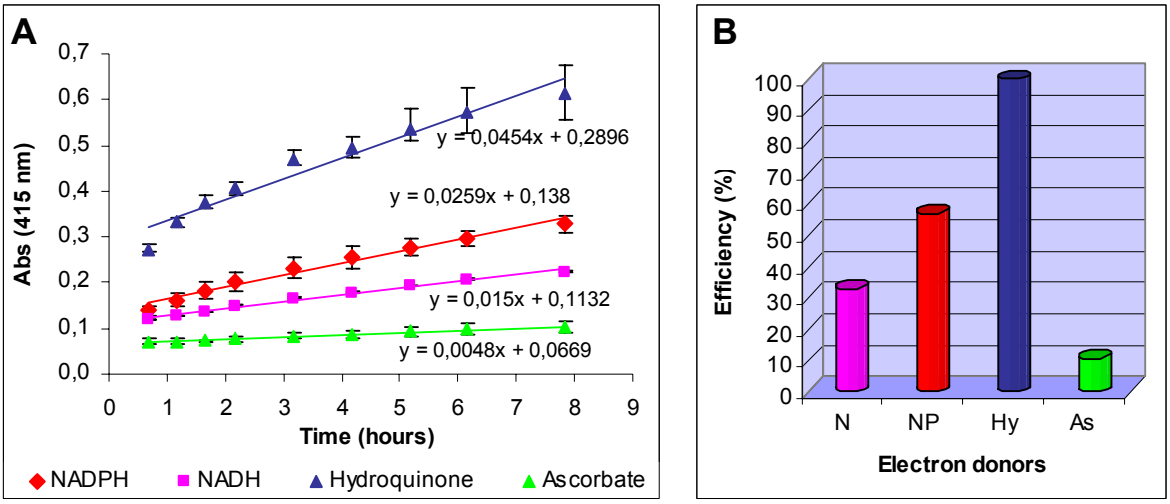


Figure 3.8. Comparison of different electron donors for selenite reduction *in vitro*: NADH (N), NADPH (NP), reduced ascorbate (As) and reduced hydroquinone (Hy). (A) Selenite reduction kinetics, as measured by reading the absorbance at 415 nm of the reaction mixtures. Vertical bars indicate the standard deviations as determined by measurements in duplicate. (B) Histogram representation of the various donors efficiencies, expressed in percentage with respect to the highest one (hydroquinone, 100% activity).

3.3.2.2. The effect of GSH

The role of GSH in the detoxification mechanism used by *S. maltophilia* strain SeITE02 in response to toxic concentrations of selenite was evaluated by *in vitro* experiments as well. In this concern, we have tried to add reduced GSH to the reaction environment, in order to verify whether it had any effect on the selenite reduction kinetics. For this purpose we used three different cytoplasm protein extracts, deriving from cultures grown in presence (0.2 or 0.5 mM) or in absence of selenite. By comparing these three samples we have obtained some information about the possible involvement of this detoxification pathway in the adaptation of the bacterial cells to selenite. The protein contents of the sample

were levelled before the assay, in order to have comparable data. Selenite reduction was followed by direct quantification of the selenite present in solution as described in the paragraph 3.2.4.1, by taking aliquots of the reaction mixtures and of the three negative controls without GSH (-G), proteins (-P), or both of them (-GP) after prearranged incubation times. Figure 3.9 shows the reduction kinetics of the six systems over a total incubation time of 40 minutes.

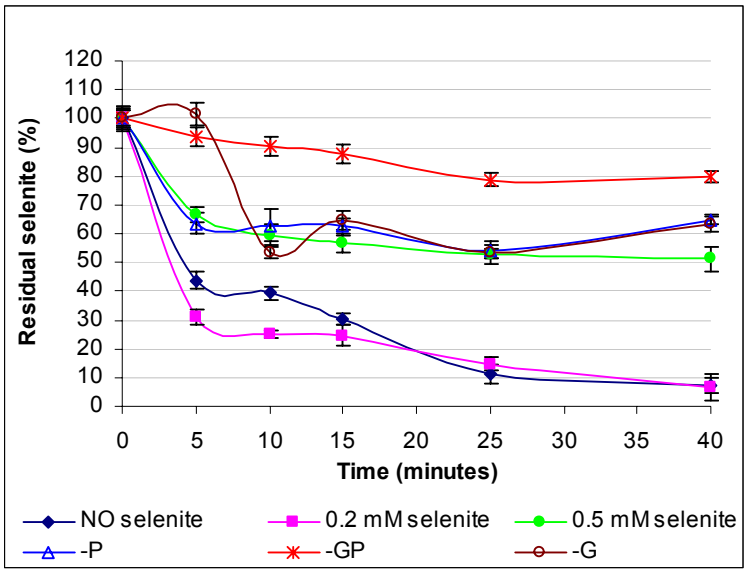


Figure 3.9. *In vitro* selenite reduction in presence of cytoplasm protein extracts, NADPH and GSH. Complete reduction systems contained cytoplasm protein extracts from cultures either without selenite (NO selenite) or with 0.2 or 0.5 mM selenite. Negative controls were without the protein factor (-P), without GSH (-G) or without both of them (-GP). Residual selenite (initial amount 0.2 mM) was determined directly by spectrophotometric quantification. Vertical bars indicate the standard deviations as determined by measurements in duplicate.

While when both the proteins and GSH were missing only a modest reduction occurred (10% of the initial amount) over the total reaction time, the other two controls (-G and -P) showed a more consistent selenite reduction in the first 10 minutes (30%), although afterwards the conversion didn't proceed any longer. On the other hand, in the case of complete reaction mixtures with proteins from cultures without selenite or with 0.2 mM selenite, the reaction developed more rapidly at the beginning (65% and 50% of conversion, respectively), and then continued slowly until reaching a total conversion of about 90% of the initial selenite at the end of the experiment. Selenite conversion was accompanied by

the appearance of the red colloidal suspension of elemental selenium, which developed much more rapidly (minutes) than in the case in which only the electron donor was present (hours, see figure 3.7). On the other hand, the proteins extracted from cultures grown in presence of 0.5 mM selenite behaved like the controls “-G” and “-P”, thus suggesting that high concentrations of selenite during cell growth suppress this mechanism rather than inducing it.

3.3.2.3. The effect of nitrite

Since nitrite proved to greatly affect the bacterial cells ability of reducing selenite *in vivo* during the exponential growth phase, the influence of nitrite on the reduction process was assessed *in vitro* as well, by performing reactions in presence of protein extracts, NADPH, GSH and nitrite. In this case only one protein sample was assayed (cell culture in absence on selenite) with different nitrite concentrations. In particular, nitrite was added to the reaction mixtures in the following ratio with respect to selenite: 1:1, 5:1, 10:1 and 75:1, the last one being the ratio present in the case of nitrite-selenite mixed cultures. Two negative controls were used: a mixture deprived of proteins (-P) and one without GSH (-G), both with the highest nitrite-selenite ratio. As usual, selenite conversion was followed by direct quantifications at fixed incubation times, over a total period of 40 minutes. Figure 3.10 displays the reduction kinetics of the various mixtures. It is possible to observe that the two negative controls (-G and –P) behaved exactly like the correspondent controls in the experiment with GSH but no nitrite (see figure 3.9), giving a total conversion of 10% and 30% of the initial selenite, respectively. On the other hand, nitrite seemed to moderately affect the enzymatic reduction of the oxyanion only at the highest concentration assayed, where the total conversion of selenite decreased from 100% (nitrite/selenite 1:1, 5:1 and 10:1) to 70% (nitrite/selenite 75:1). These results suggest that nitrite exert some negative effect on the enzymatic reduction of selenite in presence of GSH, although it cannot absolutely explain the total absence of reduction observed during the exponential growth phase of cells cultured in nitrite-containing media.

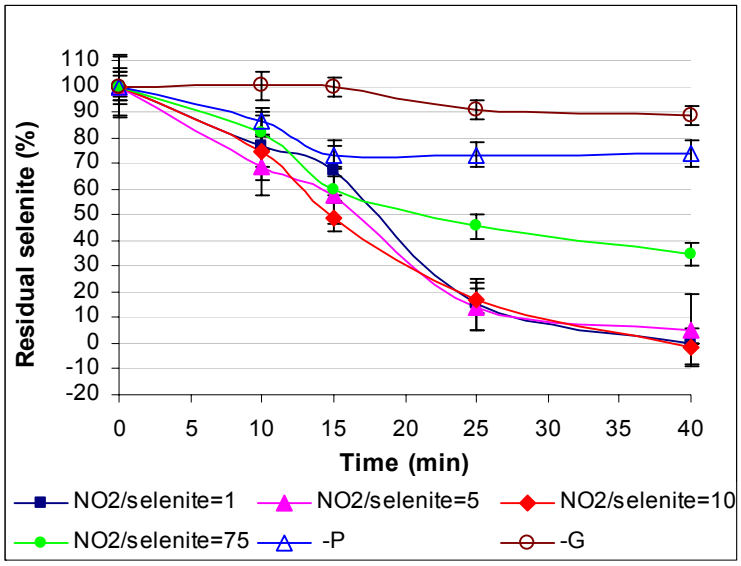
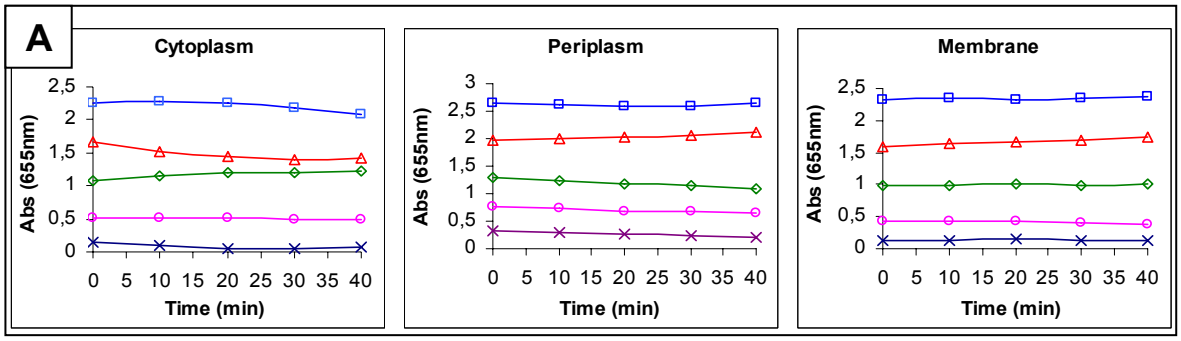


Figure 3.10. *In vitro* selenite reduction in presence of cytoplasm protein extracts (from cultures without selenite), NADPH, GSH and various concentrations of nitrite (as indicated in the graph legend). Negative controls were without the protein component (-P) or without GSH (-G). Residual selenite was determined directly by spectrophotometric quantification. Vertical bars indicate the standard deviations as determined by measurements in duplicate.

3.3.3. NITRITE REDUCTION IN VITRO

NIR activity in sub-cellular protein extracts was evaluated as well. In this case, a cell culture pre-induced with nitrite was used, and the three sub-cellular fractions were isolated as previously described. Nitrite reduction *in vitro* was assessed either indirectly, by monitoring methyl viologen oxidation in presence of nitrite, or directly, using NADH as electron donor and quantifying the residual selenite. Figure 3.11 panel A displays the kinetics of methyl viologen oxidation (net values, obtained by subtracting the absorbance variations observed in the negative controls to those measured in the complete reaction mixtures) in presence of nitrite buffered at different pH values (pH 5.0, 6.0, 6.5, 7.0 and 8.0), while panel B shows the kinetics of nitrite reduction as determined by direct measurements of the residual nitrite in solution. Definitely, no significant net reduction of nitrite could be observed in these conditions, neither with methyl viologen nor with NADH as sole electron donors.



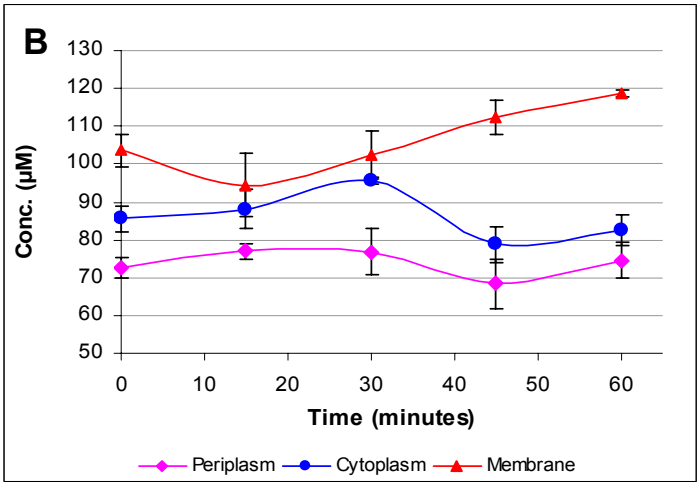


Figure 3.11. *In vitro* nitrite reduction in presence of the three sub-cellular protein fractions (from cultures without selenite), NADH and nitrite. (A) kinetics of methyl viologen oxidation in various pH conditions: light blue, pH 8.0; red, pH 7.0; green, pH 6.5; pink, pH 6.0; blue, pH 5.0. (B) direct determination of nitrite content (pH 7.0).

3.3.4. GLOBAL 2-D ANALYSIS OF THE PROTEOMIC RESPONSE TO NITRITE AND SELENITE

As selenite reduction was observed exclusively in presence of the cytoplasm fraction, and since prefractionation of complex samples is highly recommended prior to any analysis, we have decided to focus our attention on this protein fraction for evaluating selenite and nitrite global effects on the proteome of *S. maltophilia* strain SeITE02. We have considered these two treatments because previous analysis had suggested that some connection exists between the two detoxifying pathways. Three cell cultures were prepared in simple DM A medium (which was our reference), or in media infused with either 0.2 mM selenite or 16 mM nitrite (final concetrations). At the beginning of the stationary phase the cells were collected and processed as described in paragraph 3.2.3 in order to separate the various subcellular protein fractions. Possible contaminants

interfering with the separation processes were removed by protein precipitations with TCA, after that 5 replica maps of each one of the three samples were prepared and compared in a single three-way analysis. We decided to process all the maps in a unique match-set in order to have a direct comparison of all the protein spots detected in the various samples. By staining with the highly sensitive Sypro Ruby fluorescent stain we were able to detect in each map about 900 protein features. Selenite was found to significantly modulate 96 protein zones (23 induced, 28 repressed, 17 positively modulated and 28 negatively modulated by a factor of at least two), while nitrite caused expression variations in 85 different spots (62 induced, 8 repressed, 8 positively modulated and 7 negatively modulated by a factor of at least two). Although nearly equal amounts of proteins resulted to be modulated by the two treatments (about 10% of the total spots visualized on the gels), only a few of them were found to be common elements of the SeITE02 responses to selenite and nitrite. Figure 3.12 shows the master map that summarizes the comparison between the control and the two treated samples, while figure 3.13 displays representative maps of each group, that is control (panel A), selenite (panel B) and nitrite (panel C). In each map the modulated spots are highlighted by coloured symbols and standard spot protein (SSP) numbers, the latter being listed also in table 3.1, which reports the protein identities as obtained by MS/MS analysis. Although only a few of the proteins identified presented the same modulation trend in the two treatments, they could be grouped into functional classes which revealed the existence of some common defence mechanisms activated by our strain in response to both the oxyanions here investigated. Such functional classes are reported in table 3.2, with the most interesting ones highlighted in blue: damaged protein catabolism, DNA metabolism and cell division, and oxidative stress response. Some of these proteins will be described better in the discussion.

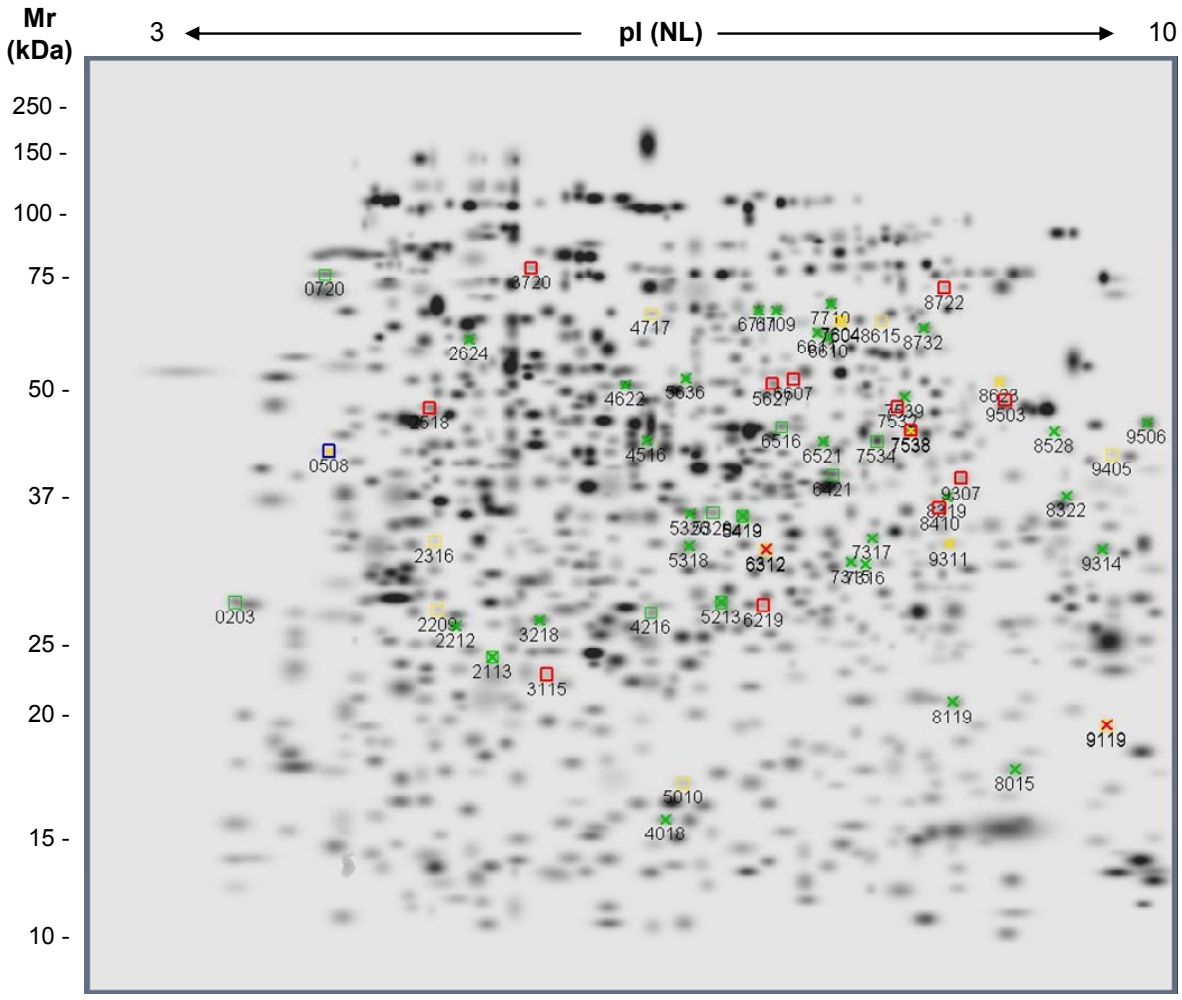
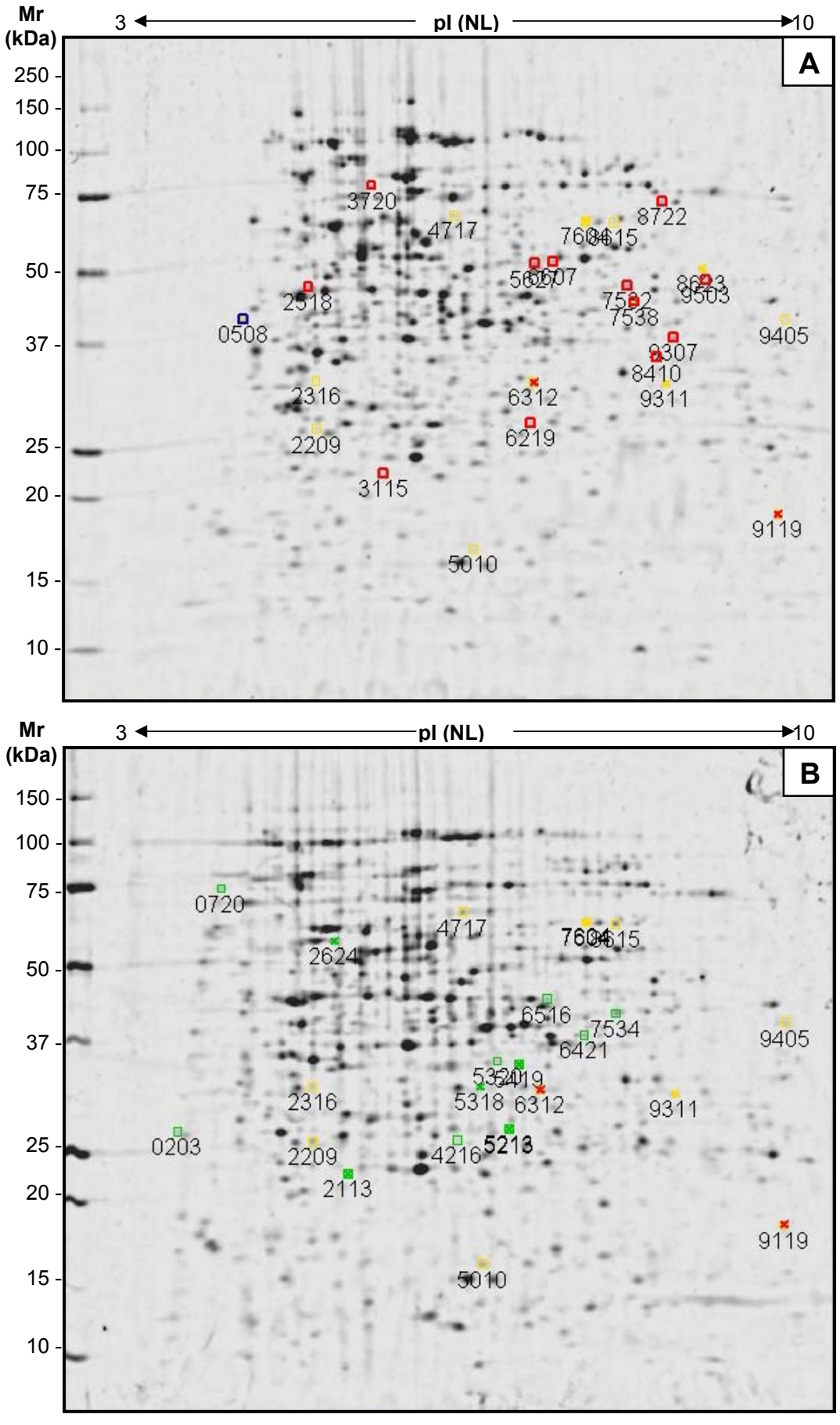


Figure 3.12. Master map of the three-way comparison among cells non-treated (control) and treated with 0.2 mM selenite or 16 mM nitrite. Squares highlight spots modulated by selenite, while crosses indicate nitrite-induced modulations. Green, newly induced spots; yellow, positively modulated spots; blue, negatively modulated spots; red, completely repressed spots.


Mr (kDa)
3
pI (NL)
10
A
250
150
100
75
50
37
25
20
15
10
3720
4717
8722
2518
0508
7538
9405
9307
8410
2316
6312
9311
6219
2209
3115
9119
5010
Mr (kDa)
3
pI (NL)
10
B
150
100
75
50
37
25
20
15
10
0720
4717
2624
6516
7534
9405
6421
2316
5318
6312
9311
0203
2209
4216
2113
9119
5010

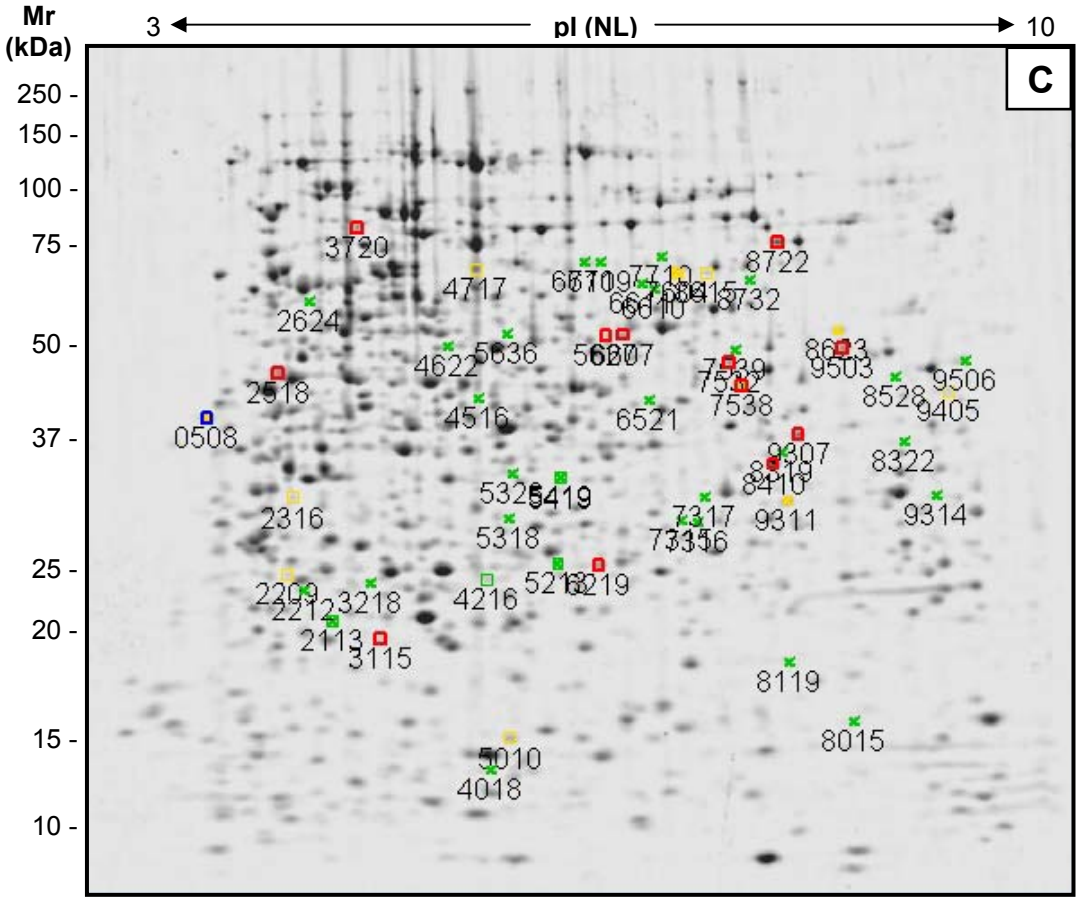


Figure 3.13. Representative maps of the three cytoplasm protein samples: control (A), 0.2 mM selenite (B) and 16 mM nitrite (C). Squares highlight spots modulated by selenite, while crosses indicate nitrite-induced modulations. Green, newly induced spots; yellow, positively modulated spots; blue, negatively modulated spots; red, silenced spots.

Table 3.1. Proteins identified by LC-MS/MS.

SSP	Sel	NO_2	Coverage (identified peptides)	Mascot Score	Mr (kDa) pI (observed)	Protein name [homologous organism]
0203	ON	/	1	56	25,5 3,40	**Purine nucleoside phosphorylase** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21112337)
0508	/	UP	4	150	38,1 4,10	**Conserved hypothetical protein** [Xanthomonas axonopodis pv. citri str. 306] (gi|21108564)
0720	***ON***	***/***	***2***	***80***	66,7 4,07	***Catalase [Xanthomonas axonopodis pv. citri str. 306]*** (gi|21107359)

2113	***ON***	***/***	***1***	***60***	***21,6 5,26***	***ATP-dependent Clp protease proteolytic subunit [Xanthomonas campestris pv. campestris str. ATCC 33913]*** (gi|21112001)
2209	UP	/	1	60	24,9 4,89	**Citrate synthase** [Xanthomonas campestris pv. vesicatoria str. 85-10] (gi|78037490)
2212	/	ON	1	60	23,7 5,02	**Cxytidylate kinase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21108544)
2316	***UP***	***/***	***2***	***68***	***30,6 4,88***	***Glutamate-cysteine ligase precursor [Xanthomonas campestris pv. campestris str. ATCC 33913]*** (gi|21114650)
2518	OFF	/	4	241	42,7 4,84	**Enolase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107915)
2624	/	ON	3	118	53,5 5,11	**Elongation factor G** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21107104)
3115	OFF	/	2	60	20,4 5,63	**Hypothetical protein AcidDRAFT_7986** [Solibacter usitatus Ellin6076] (gi|67933900)
3218	/	ON	3	187	24,1 5,58	**Elongation factor Tu** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21111917)
3720	***OFF***	***/***	***2***	***76***	***69,2 5,53***	***Glycyl-tRNA synthetase beta chain [Xanthomonas axonopodis pv. citri str. 306]*** (gi|21110642)
4018	***/***	***ON***	***1***	***62***	***14,0 6,36***	***Glutathione synthetase [Xanthomonas axonopodis pv. citri str. 306]*** (gi|21109427)
4216	***ON***	***/***	***2***	***116***	***24,7 6,27***	***Phenol hydroxylase [Xanthomonas campestris pv. campestris str. 8004]*** (gi|77761325)
4516	/	ON	5 2 1	339 93 64	38,7 6,25	**Acyl-CoA dehydrogenase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107744) **Oxidoreductase, zinc-binding** [Pseudomonas syringae pv. tomato str. DC3000] (gi|28869885) **Xaa-Pro aminopeptidase** [Hahella chejuensis KCTC 2396] (gi|83643941)
4622	/	ON	3	143	46,0 6,16	**Kynurenine 3-monooxygenase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107785)
4717	UP	/	2	70	58,2 6,27	**Polynucleotide phosphorylase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21108963)
5010	UP	/	1	40	15,8 6,46	**Conserved hypothetical protein** [Xanthomonas axonopodis pv. citri str. 306] (gi|21106892)
5213	ON	/	1	75	25,5 6,67	**Enoyl-CoA hydratase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107472)
5318	***ON***	***ON***	***2***	***84***	***30,1 6,50***	***Exodeoxyribonuclease III [Legionella pneumophila subsp.***

						pneumophila str. Philadelphia 1] (gi|21106384)
5320	ON	/	1	55	33,3 6,62	**2-dehydro-3-deoxyphosphooctonate aldolase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107913)
5326	/	ON	5	302	33,2 6,50	**Putative glycerol-3-phosphate dehydrogenase** [Xanthomonas campestris pv. vesicatoria str. 85-10] (gi|78034193)
5419	ON	/	3	171	32,9 6,84	**2-dehydro-3-deoxyphosphooctonate aldolase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107913)
5627	OFF	/	3	134	46,2 7,07	**Putative phosphoglucomutase** [Shigella flexneri 2a str. 301] (gi|24053647)
5636	/	ON	2	98	47,0 6,48	**Glycerol kinase** (ATP:glycerol 3-phosphotransferase) [Xylella fastidiosa] (gi|24636898)
6219	OFF	/	1	57	25,2 7,01	**Hypothetical protein** [Nocardia farcinica IFM 10152] (gi|54017403)
6312	UP	/	1	70	29,8 7,02	**eEsterase** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21114640)
6421	ON	/	1	71	36,7 7,52	**UDP-N-acetylmuramate-alanine ligase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21106901)
6516	ON	/	2	99	40,2 7,14	**Acetoacetyl-CoA thiolase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107510)
6521	/	ON	6	290	38,7 7,44	**Twitching motility protein** [Xanthomonas axonopodis pv. citri str. 306] (gi|21109227)
6607	OFF	/	2	88	46,9 7,23	**Conserved hypothetical protein** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21112741)
6610	/	ON	2	144	53,9 7,48	**Topoisomerase IV subunit B** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107909)
6611	/	ON	7	275	54,6 7,41	**Peptide chain release factor 3** [Xanthomonas axonopodis pv. citri str. 306] (gi|21109358)
6709	/	ON	3	105	58,8 7,10	**Amidophosphoribosyltransferase** [Xylella fastidiosa 9a5c] (gi|9107049)
6711	/	ON	2	126	58,8 6,97	**Amidophosphoribosyltransferase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107173)
7315	/	ON	1	52	28,7 7,65	**Inorganic polyphosphate/ATP-NAD kinase** [Xanthomonas campestris pv. vesicatoria str. 85-10] (gi|78035622)
7316	/	ON	1	45	28,5 7,76	**Extragenic supressor protein** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21113417)
7317	/	ON	1	62	30,9	**Bifunctional purine biosynthesis**

					7,81	**protein** [Xanthomonas axonopodis pv. citri str. 306] (gi|21106603)
7532	OFF	/	1	61	42,9 8,00	**Conserved hypothetical protein** [Pseudomonas fluorescens Pf-5] (gi|68347656)
7534	ON	/	3	120	38,7 7,85	**DNA topoisomerase I** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21115033)
7538	/	UP	1	45	39,7 8,10	**Sugar ABC transporter ATP-binding protein** [Xanthomonas axonopodis pv. citri str. 306] (gi|21108297)
7539	/	ON	2	96	44,3 8,05	**Conserved hypothetical protein** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21112741)
7604	UP	/	2	86	56,8 7,58	**Conserved hypothetical protein** [xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21115308)
7710	/	ON	3	120	60,0 7,50	**RNA polymerase beta subunit** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107099)
8015	/	ON	1 2	72 70	16,2 8,84	**ATP synthase alpha chain** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21111547) **Transcription termination factor NusB** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21111708)
8119	/	ON	3	118	19,0 8,40	**Septum site-determining protein** [Xanthomonas oryzae pv. oryzae MAFF 311018] (gi|84368934)
8319	/	ON	3	118	34,9 8,35	**Inositol-5-monophosphate dehydrogenase** [Xylella fastidiosa Temecula1] (gi|28199329)
8322	/	ON	4 1	202 63	35,0 9,20	**Diadenosine tetraphosphatase** [Xylella fastidiosa 9a5c] (gi|9107285) **Regulatory protein, LysR:LysR, substrate-binding** [Ralstonia metallidurans CH34] (gi|68556888)
8410	OFF	/	2	82	33,7 8,31	**GTP binding protein** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21112324)
8528	/	ON	2	118	39,6 9,11	**TolB protein** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21114229)
8615	UP	/	1	63	56,8 7,88	**Conserved hypothetical protein** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21115308)
8623	/	UP	4	194	46,5 8,73	**Chromosomal replication initiator** [Xanthomonas axonopodis pv. citri str. 306] (gi|21106039)
8722	***OFF***	***/***	***1***	***56***	63,2 8,34	***Exodeoxyribonuclease V alpha chain [Xanthomonas axonopodis pv. citri***

						str. 306] (gi|21110777)
8732	/	ON	4	193	55,5 8,19	**GTP-binding protein** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21113105)
9119	***UP***	***/***	***1***	***50***	18,0 9,48	***Single-stranded DNA binding protein [Xylella fastidiosa 9a5c] (gi|9106400)***
9307	OFF	/	4	171	36,5 8,46	**Transcriptional regulator** [Xanthomonas axonopodis pv. citri str. 306] (gi|21107845)
9311	/	UP	1	58	30,3 8,37	**Shikimate kinase** [Pelodictyon phaeoclathratiforme BU-1] (gi|68550241)
9314	/	ON	2	55	29,8 9,45	**Chromosomal replication initiator** [Xanthomonas axonopodis pv. citri str. 306] (gi|21106039)
9405	UP	/	2	82	37,9 9,51	**Dihydrolipoamide acetyltranferase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21110045)
9503	OFF	/	2	59	43,8 8,77	**Trehalose-6-phosphate synthase** [Xanthomonas campestris pv. campestris str. ATCC 33913] (gi|21114302)
9506	/	ON	2	88	40,7 9,67	**6-phosphofructokinase** [Xanthomonas axonopodis pv. citri str. 306] (gi|21109796)

Table columns report (in the order, from left to right) the SSP number of the protein spots, the modulation trend in the samples treated with selenite and nitrite, the protein coverage, the Mascot score, the experimental values of Mr and pI, and the protein name (with NCBI accession numbers). Proteins of interest are in italic blue.

Table 3.2. Functional protein classes.

Functional classes	Selenite modulations	Nitrite modulations
Nucleotide biosynthesis and metabolism	• Purine nucleoside phosphorylase • Polynucleotide phosphorylase • GTP binding protein	• Amidophosphoribosyltransferase • Cytidylate kinase • Bifunctional purine biosynthesis protein • Inositol-5-monophosphate dehydrogenase • Inorganic polyphosphate/ATP-NAD kinase • ATP synthase alpha chain • Diadenosine tetraphosphatase

Protein catabolism	• *ATP-dependent Clp protease proteolytic subunit* • *Prolyl oligopeptidase*	
Protein synthesis		• Elongation factor G • Elongation factor Tu • Peptide chain release factor 3
Peptide and aminoacid metabolism	• Glycyl-tRNA synthetase beta chain	• Kynurenine 3-monooxygenase • Xaa-Pro aminopeptidase • Shikimate kinase
Lipid metabolism	• Enoyl-CoA hydratase • Esterase • Acetoacetyl-CoA thiolase	• Acyl-CoA dehydrogenase
Sugar metabolism	• Citrate synthase • Dihydrolipoamide acetyltranferase Enolase • Putative phosphoglucomutase • Trehalose-6-phosphate synthase	• Putative glycerol-3-phosphate dehydrogenase • Glycerol kinase • 6-phosphofructokinase
Sugar transport (and regulation)		• Sugar ABC transporter ATP-binding protein • HPr kinase/phosphatase
DNA related proteins and cell division	• *DNA topoisomerase I* • *Exodeoxyribonuclease III* • *Exodeoxyribonuclease V alpha chain* • *Single-stranded DNA binding protein*	• RNA polymerase beta subunit • Transcription termination factor NusB • Regulatory protein, LysR:LysR, substrate-binding • *Exodeoxyribonuclease III* • *Topoisomerase IV subunit B* • Septum site-determining protein • Chromosomal replication initiator
Oxidative stress related proteins	• *Catalase* • *Glutamate-cysteine ligase precursor*	• *Glutathione synthetase*
Aromatic compound metabolism	• *Phenol hydroxylase*	
Motility proteins		• Twitching motility protein
Cell wall synthesis	• 2-dehydro-3-deoxyphosphooctonate	

	aldolase • UDP-N-acetylmuramate-alanine ligase	

“Interesting” functional classes are highlighted in blue.

3.4. DISCUSSION

3.4.1. NITRITE EFFECT ON CELL GROWTH AND SELENITE REDUCTION IN VIVO

The microbial detoxification mechanisms exploited by bacteria in the attempt to resist to high concentrations of selenite are still poorly understood. In particular, great attention has been given to dissimilatory reduction, although there are evidences for microbial reduction of selenite and selenate in aerobic conditions as well. In some events, periplasmic dissimilitory nitrite reductase has been proposed to be able to reduce selenite, as in the case of *Thauera selenatis,* where a mutant lacking such an enzyme was unable to metabolize selenite [14]. On the other hand Kessi [18] has recently reported that, while in *Rb*. *capsulatus* both nitrite and selenite are simultaneously metabolized, with these oxyanions interfering at the transport level rather than at the reduction one, in *Rs*. *rubrum* the reduction pathway of selenite seems to be completely separated from that of nitrite. Moreover, while selenite reductase activity was found to be constitutively expressed in either *Rb*. *capsulatus* or *Rs*. *Rubrum*, nitrite reductase is known to be an inducible enzyme in these bacteria. Therefore, the involvement of nitrite reductase in microbial selenite reduction is highly probable, at least in some species, although it’s evident that other independent mechanisms exist as well. As regards the bacterial strain here investigated, the possible interactions between selenite and nitrite reduction pathways have been studied either *in vivo* or *in vitro*. First we have analyzed the influence of nitrite on cell growth and selenite reduction activity by performing experiments of cells pre-induction with nitrite and nitrite-selenite mixed cultures. As nitrite reductase is usually not expressed in bacteria in absence of its specific susbstrate, cells pre-induction with nitrite should

have given an advantage over non pre-induced cells in the event that nitrite reductase plaid a role in selenite reduction. However no appreciable differences of growth rates and reduction kinetics could be observed among pre-induced and non induced cells cultivated in presence of selenite, and this suggests that the detoxification mechanism is independent from nitrite reduction. This hypothesis is also corroborated by the fact that the strain SeITE02, when cultivated in a medium added with 0.5 mM selenite, presents the same lag time as that observed in cultures not supplemented with anything [26], thus further suggesting that selenite is reduced by e constitutively expressed mechanism different from the nitrite reduction pathway. On the other hand, the form in which nitrogen was supplied to the selenite-nitrite mixed cultures greatly affected both the cell growth and the selenite reduction kinetics. In fact, quite unexpectedly, cells grew up more rapidly and reached the highest cell density in the stationary phase when nitrogen was provided as sole yeast extract (organic form). Nitrite allowed for a lower cell population in the stationary phase, which was, however, considerably higher than that obtained in cultures with ammonium sulphate. As concerns selenite reduction, the cell cultures supplemented with only yeast extract or yeast extract and ammonium sulphate behaved almost identically, converting 90% of the initial selenite during the exponential growth phase, while in the case of nitrite the reduction started at the beginning of stationary phase, proceeding slowly until reaching a total conversion of only 30% of the selenite initially present. These results are in total agreement with those recently reported by Kessi [18] about *Rb. capsulatus*, which performs selenite reduction during the exponential growth phase, without showing any growth start delay due to need for adaptation in presence of nitrite. Moreover, also in *Rb. capsulatus* nitrite greatly affects selenite reduction during the exponential growth phase. A possible interpretation of this behaviour is that the reduction pathways of nitrite and selenite may share one or more enzymes, most probably at the level of transport. In fact, since in absence of nitrite selenite clearly enters the cells immediately after its addition, in the event that the competition between these two oxyanions was at the level of reduction, selenite would accumulate in the cell cytoplasm during nitrite reduction in the exponential growth phase, thus rapidly oxidizing inorganic and organic –SH groups (glutathione, proteins). However, because nitrite-selenite mixed cultures

don't seem to be greatly affected by the presence of selenite as compared to cultures supplemented with only nitrite, it is likely that the consistent delay observed in selenite reduction onset is caused by the fact that selenite is not imported into the cells in presence of nitrite. This hypothesis is also strongly supported by the results obtained with cultures fortified with DNP, a specific inhibitor of nitrite intake, which caused a consistent decrease of selenite reduction *in vivo*, allowing for a total conversion of only 20% of the initial selenite. This is perfectly comparable with what is observed in cultures reinforced with both nitrite and selenite (see figure 3.3). To sum up, there is great evidence that the enzyme(s) shared by the reduction pathways of nitrite and selenite are involved in the oxyanion transport rather than in their reduction, as also proposed by Hudman and Glenn [33].

3.4.2. THE ROLE OF GSH IN SELENITE REDUCTION *IN VIVO*

Non-protein thiols, such as thioredoxin and glutathione, are widely known for playing a fundamental role in cell protection from toxic oxidizing compounds. GSH has been proposed to drive also selenite reduction in bacterial species, according to the reaction scheme hypothesised by Kessi and Hanselman [22] and displayed in figure 3.14. With this reaction mechanism, GSH functions as the main electron donor, reacting with selenite and forming selenodiglutathione (GS-Se-SG, reaction 1). This is then converted into selenopersulfide of glutathione ($GS\text{-}Se^-$, reaction 2) by glutathione reductase (GR). Finally, the last intermediate dismutates into reduced glutathione and elemental selenium (reaction 3). Such a reduction pathway seems to apply well also to our bacterium. In fact, selenite reduction appeared to be considerably delayed in cultures supplemented with BSO and selenite since the very beginning of cell growth. As shown in figure 3.6, panel A, the inhibitory effect is directly proportional to BSO concentration, which, when present in sufficient amounts (3 mM), almost completely prevents selenite reduction during the lag and the exponential growth phases. However, no difference between control and "inhibited" cultures could be observed when selenite was supplemented after 18 or 42 hours of incubation, although the

reduction degree was considerably lower as compared to that observed in cultures supplemented with selenite since the beginning of growth.

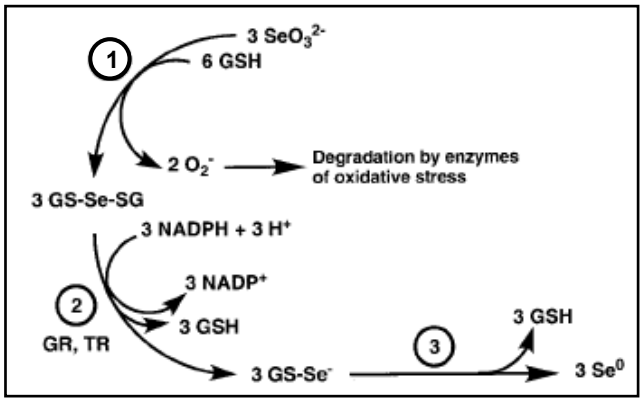


Figure 3.14. Mechanism of biological reduction of selenite. (1) the reaction between GSH and selenite leads to the formation of selenodiglutathione (GS-Se-SG); (2) conversion into selenopersulfide of glutathione (GS-Se^-); (3) the last intermediate spontaneously dismutates into GSH and elemental selenium (Se^0).

These findings allow us to draw some conclusions: (i) GSH is clearly involved in the selenite reduction mechanism of the present bacterial strain, since the inhibition of its synthesis leads to a considerable delay in the reduction onset; (ii) more mechanisms are likely to coexist, since BSO has no effect when cultures are reinforced with selenite during the exponential or the stationary growth phases, and also when the inhibitor is present from the very beginning (lag phase), selenite is successfully reduced only when the cells have entered the stationary phase; (iii) as a corollary of the previous point, it is likely that GSH plays its role during the very first steps of the cell response to the toxic oxyanion (lag and exponential growth stages), which are also the most important, accounting for a total conversion of abut 90% of the selenite initially present (see figure 3.3).

3.4.3. SELENITE REDUCTION *IN VITRO*

When cultured in presence of selenite, *S. maltophilia* strain SeITE02 gives rise to an intense coloration of the culture medium, caused by the formation of red colloidal particles made of elemental selenium that accumulate either inside the cells (cytoplasm) or in the medium, as described previously by Di Gregorio and coworkers [26]. On the other hand, the selenium bodies have never been found in the periplasmic space, thus leaving doubts about the site of selenite reduction

(proposed to be at the periplasmic level by many) and the provenience of the extracellular particles. In the attempt to unravel the riddle we have performed subcellular compartment fractionation and *in vitro* enzymatic assays aimed at localizing the selenite reduction activity. Our results have confirmed that the unique site of selenite reduction is cytoplasm, at least in the conditions that we used for the assays. According to these findings, the presence of free selenium bodies in the medium could be explained by either the existence of an active secretion mechanism or by cell lysis.

As regards the selenite reduction assays, various electron donors have been taken into account (*i.e.* NADH, NADPH, reduced ascorbate and reduced hydroquinone). Moreover different pH conditions have been investigated (pH 5.0, 6.0, 6.5, 7.0 and 8.0). The fastest selenite conversion was observed at pH 6.0 and in presence of NAPDH as electron donor. The fact that NADPH is preferred to NADH by the enzyme(s) involved in selenite reduction supports the hypothesis that GSH and glutathione reductase (GR) play an important role. In fact, NADPH is the preferential electron donor of GR, which, according to the mechanism displayed in figure 3.14, would restore the pool of reduced GSH by oxidizing NADPH and reducing selenodiglutathione.

In order to verify such a hypothesis, various in vitro assays have been performed in presence of cytoplasm protein fraction, NADPH and GSH. As described in the results, GSH allows for a much more rapid selenite reduction in vitro, decreasing the time needed for the formation of a detectable amount of selenium suspension from hours to minutes (compare figures 3.7 and 3.9). As concerns the reduction kinetics, we have observed that when all of the functional components (*i.e.* cytoplasm fraction, NADPH and GSH) are present, selenite reduction proceeds rapidly until reaching a total conversion of about 90% of the initial amount in 40 minutes. When proteins are not added to the reaction mixture, the reduction proceeds rapidly at the beginning, converting about 40% of the initial amount in the first 10 minutes, but afterwards the reaction completely stops and no further conversion takes pace. This can be easily explained if we consider that, according to the reaction scheme shown in 3.15, the first reduction step between GSH and selenite takes place spontaneously without the need for enzymatic catalysis. However, once the GSH pool has been converted, it cannot be restored by GR,

and the reaction stops. The need for both GSH and cytoplasm proteins is also evidenced by the fact that, when none of them is present in the reaction mixture, spontaneous selenite reduction is minimal.

Also the specific cultural conditions seem to have an effect on the reduction mechanism. In fact, while both the samples grown in absence of selenite and in presence of 0.2 mM selenite reduced the oxyanion at an almost identical rate, the enzymes extracted from cells grown in presence of 0.5 mM selenite performed a slower and partial conversion (40% of the initial amount after 40 minutes), although the total protein amount used for the assay was identical in the three conditions. This may suggest that, in presence of selenite concentrations higher than 0.2 mM, other protective strategies could be induced.

The inhibitory effect of nitrite was also assessed in these experimental conditions, in the attempt at better understanding the interaction between the two pathways. As shown in figure 3.10 nitrite, when present at the same ratio as that used in the selenite-nitrite mixed cultures (75:1), sensibly inhibits selenite reduction *in vitro*, allowing for a total conversion of about 70% of the initial selenite after a 40-minute incubation. However this is far from being comparable with what observed *in vivo*, where the reaction doesn't start until cells enter the stationary phase. Thus, while the modest inhibition observed *in vitro* could be due to enzyme inactivation caused by oxidative events (at the beginning the various reactions proceed at a similar rate) or to competition between nitrite and selenite for the same enzyme (which should have, however, an affinity much greater for selenite than for nitrite, thus excluding a nitrite reductase), in the case of viable cells the competition is likely to take place at the transport level.

3.4.4. NITRITE REDUCTION IN VITRO

In order to definitely ascertain whether nitrite reductase plays some role in selenite reduction, the cytoplasm fraction obtained from cells pre-induced with nitrite has been assayed for nitrite reduction activity. The results confirmed what we had already hypothesised on the basis of previous experiments: nitrite reductase is not likely to be involved in the cytoplasmic reduction of selenite, since

no activity of such an enzyme could be detected in conditions analogous to those used for *in vitro* selenite reduction.

3.4.5. GLOBAL PROTEOMIC ANALYSIS OF THE BACTERIAL RESPONSE TO NITRITE AND SELENITE

Global proteome analysis by 2-DE has been used to get a comprehensive picture of the proteomic response of *S. maltophilia* strain SeITE02 to the two oxyanions here investigated. Both 0.2 mM selenite and 16 mM nitrite caused a significant modulation of about 10% (96 and 85 spots, respectively) of the total proteins visualized on the respective 2-D maps. These protein spots were excised from gels and subjected to in-gel trypsin digestion and MS/MS analysis, aimed at identifying all of the modulated proteins. As frequently occurs in these experiments, where organisms are challenged with environmental conditions highly different from those they are adapted to, the proteins modulated are numerous and belong either to specific response mechanisms or to general biochemical pathways. As concerns this study, the modulated proteins, in spite of being considerably different in the two comparisons, can be grouped into common classes according to their functional role in the cell: nucleotide synthesis and metabolism, damaged proteins catabolism, protein and amino acid metabolism, lipid metabolism, carbohydrate metabolism, DNA-related proteins and protein involved in cell division, and oxidative stress-response proteins. This classification is also reported in table 2.2, where the most “interesting” classes (damaged proteins catabolism, DNA-related proteins and cell division, and oxidative-stress-response) are highlighted in blue. Some of these proteins are discussed next.

SSP 0720 (induced by selenite): Catalase. This enzyme converts hydrogen peroxide to oxygen and water, thus constituting a central enzyme in the oxidative-stress response. Because H_2O_2 is a natural by-product of the abiotic reduction of selenite catalyzed by GSH [22], the fact that this enzyme is expressed only during growth in presence of selenite strongly supports the hypothesis that GSH takes part in the first steps of selenite reduction *in vivo*.

SSP 2113 (induced by selenite): ATP-dependent Clp protease proteolytic subunit. This proteolytic enzyme plays an important role in the degradation of damaged proteins. Protein damage frequently occurs in presence of reactive oxygen species (ROS), which can introduce amino acid lateral chain modifications that can completely inhibit the enzymatic function of the protein.

SSP 2316 (positively modulated by selenite): Glutamate-cysteine ligase precursor. Along with the enzyme glutathione synthetase, this enzyme drives the biosynthetic pathway of glutathione in most of the eukaryotes, cyanobacteria and purple bacteria [35].

SSP 4018 (induced by nitrite): Glutathione synthetase. Glutathione (the most abundant non-protein thiol compound present in a cell) and its related enzymes (among which glutamate-cysteine ligase, glutathione synthase, glutathione reductase, glutathione-S transferase) constitute an extremely important protection mechanism against oxidant agents both in eukaryotic and prokaryotic cells. GSH has been reported to provide high resistance to acidic environments, osmotic and oxidative stress [36].

SSP 5318 (induced by selenite and nitrite): Exodeoxyribonuclease III. There are two families of AP endonucleases, each sharing homology with one of the two *E. coli* prototypes: exodeoxyribonuclease III, encoded by the *xthA* gene, and endonuclease IV, encoded by the *nfo* gene. 5′ AP endonucleases are central enzymes in the base excision repair pathway, and besides their participation in the repair of abasic sites, they play an important role in the repair of oxidative DNA lesions by removing 3′ blocks formed upon DNA strand breakage. In fact, mutants lacking a functional xthA gene have been shown to be minimally resistant to H_2O_2 [37].

SSP 9119 (positively modulated by selenite): Single-stranded DNA binding protein. This protein is essential for all the processes of chromosome replication, recombination and DNA reparation.

In conclusion, the global proteomics analysis, despite not providing further details on the specific mechanism responsible for selenite reduction in the bacterial strain SeITE02, has revealed that both nitrite and selenite probably exert their toxicity by generating ROS during their conversion into less dangerous or harmless compounds. In fact, various enzymes involved in the oxidative stress response and in the processes of protein and DNA reparation have been found to be newly expressed, or at least positively modulated, upon cell treatment with one of the two oxyanions. Moreover, there is some further evidence that glutathione actually plays a role in the first steps of selenite reduction *in vivo*. In fact selenite seems to strongly induce the expression of at least two different GSH-related enzymes (Glutamate-cysteine ligase and Glutathione synthetase), which are responsible for the biosynthesis of this compound in prokaryotes.

3.5. CONCLUSIONS AND FINAL REMARKS

Various enzymatic systems, such as nitrite reductase, sulfite reductase and glutathione reductase, have been proposed to be involved in the reduction of selenite in bacteria. Differently from most of the bacterial species able to reduce toxic selenite to insoluble elemental selenium, the strain SeITE02 of *S. maltophilia* has been reported to be able to efficiently grow in selenite-containing media, simultaneously reducing it to non-toxic elemental selenium in aerobic conditions [26]. Various enzymatic pathways involved in the selenite detoxification mechanisms of other bacteria have been taken into account in this study, and their involvement in the molecular strategy used by the strain SeITE02 to cope with high concentrations of selenite has been ascertained. Although in some cases nitrite reductase (especially the periplasmic form) has been shown to be able to reduce selenite as well, in our strain the enzymes responsible for nitrite reduction don’t seem to take part in the detoxification process, since cell pre-induction with nitrite doesn’t confer any advantage in selenite reduction as compared to non-adapted cells, and because no nitrite activity could be detected in *in vitro* experiments. However, our results demonstrated that nitrite strongly interferes with selenite removal when the two oxyanions are simultaneously

present in the cultural medium, thus suggesting that the two reduction/detoxification pathways share a common enzymatic step, probably at the level of cellular transport. This hypothesis is strongly corroborated by the fact that inhibition of nitrite intake by means of DNP perfectly mimics what observed in cells cultured in presence of both selenite and nitrite, where only a very modest conversion of selenite takes place.

A further proof of the fact that nitrite reductase is not responsible for selenite reduction comes from the extremely modest effect of nitrite on selenite reduction *in vitro*. In fact, a limited decrease of the selenite reduction rate can be observed only with the highest nitrite/selenite ratio (75:1), thus demonstrating a higher specificity for selenite rather than for nitrite.

While nitrite related enzymes are likely to simply regulate selenite absorption from the environment, without taking part directly in the reduction process, glutathione and glutathione-related enzymes are more likely to play a central role in the detoxification process. Various evidences of GSH involvement have been collected. Firstly, the inhibition of GSH synthesis during cell growth in presence of selenite delays the reduction onset to the stationary phase. Secondly, GSH strongly reduces the incubation time needed for the development of a red colloidal suspension of selenium *in vitro*. Moreover, NADPH, which is the natural electron donor of glutathione reductase, works better than NADH in supporting selenite reduction. Lastly, also the proteomics data support the hypothesis that GSH takes part in selenite reduction. Two different enzymes of the biosynthetic pathway of glutathione have been found to be up-regulated: glutamate-cyteine ligase and glutathione synthetase. Besides these two enzymes, which are reported to be important oxidative-stress-related enzymes (GSH being one of the major protective mechanism against ROS and other oxidant species), also catalase has been found to be strongly induced by selenite. Catalase drives the conversion of dangerous H_2O_2 into harmless H_2O and O_2, and since hydrogen peroxide is one of the by-products of the GSH-based selenite reduction mechanism as proposed by Kessi and Hanselman (see figure 3.14), the fact that this enzyme results to be induced by selenite is in full agreement with this hypothesis.

3.6. BIBLIOGRAPHY

[1] Shamberger RJ. The Biochemistry of Selenium. New York: Plenum Press, 1983.

[2] Rother M, Resch A, Wilting R, Böck A. *Biofactors.* 2001; **14**: 75-83.

[3] Sors TG, Ellis DR, Salt DE. *Photosynt. Res.* 2005; **86**: 373-389.

[4] Daniels LA. *Biol. Trace Elem. Res*. 1996; **54**: 185-199.

[5] Rayman MP. *Lancet*. 2001; **356**: 233-241.

[6] Shamberger RJ. *Mutat. Res.* 1985; **154**: 29-49.

[7] Duffield-Lillico AJ, Reid ME, Turnbull BW, Combs GF Jr, Slate EH, Fischbach LA, Marshall JR, Clark LC. *Cancer Epidemiol. Biomarkers Prev.* 2002; **11**: 630-639.

[8] Bocovay G. *Can. Minerals Yearb.* 1995; **55**: 28-31.

[9] Barceloux DG. Selenium. *J. Toxicol. Clin. Toxicol.* 1999; **37**: 145-172.

[10] Dungan RS, Frankenberger WT Jr. *Bioremediation Journal*. 1999; **3**: 171-188.

[11] Stolz JF, Basu P, OremLand RS. *Int. Microbiol.* 2002; **5**: 201-207.

[12] Lovley DR. *J. Ind. Microbiol.* 1995; **14**: 85-93.

[13] Vallini G, Di Gregorio S, Lampis S. *Z. Naturforsch.* 2005; **60**: 349-356.

[14] DeMoll-Decker H, Macy JM. *Arch Microbiol.* 1993; **160**: 241–247.

[15] Preuss M, Klemme JH. *Z Naturforsch.* 1983; **38**: 933–938.

[16] Richardson DJ, Bell LC, Moir JWB, Ferguson SJ. *FEMS Microbiol Lett.* 1994; **120**: 323–328.

[17] Schwintner C, Sabaty M, Berna B, Cahors S, Richaud P. *FEMS Microbiol Lett.* 1998; **165**: 313–321.

[18] Kessi J. *Microbiology.* 2006; **152**: 731-743.

[19] Harrison G, Curle C, Laishley EJ. *Arch Microbiol.* 1984; **138**: 72-78.

[20] Ganther HE. *Biochemistry.* 1971; **10**: 4089–4098.

[21] Newton GL, Fahey RC. Glutathione in prokaryotes. In: Viña JMD. (Editor) Glutathione: Metabolism and Physiological Functions. Boca Raton, Florida: CRC Press. 1989; pp. 69–77.

[22] Kessi J, Hanselmann KW. *J Biol Chem.* 2004; **279**: 50662–50669.

[23] Bébien M, Chauvin JP, Adriano JM, Grosse S, Verméglio A. *Appl Environ Microbiol.* 2001; **67**: 4440–4447.

[24] Bjornstedt M, Kumar S, Holmgren A. *J Biol Chem.* 1992; **267**: 8030-8034.

[25] Bébien M, Lagniel G, Garin J, Touati D, Verméglio A, Labarre J. *J Bacteriol.* 2002; **184**: 1556–1564.

[26] Di Gregorio S, Lampis S, Vallini G. *Environment International*. 2005; **31**: 233– 241.

[27] Frassinetti S, Setti L, Corti A, Farrinelli P, Montevecchi P, Vallini G. *Can. J. Microbiol.* 1998; **44**: 289– 297.

[28] Osborn MJ, Munson R, *Methods Enzymol.* 1974; **31**: 642-653.

[29] Kessi J, Ramuz M, Wehrli E, Spycher M, Bachofen R. *Appl Environ Microbiol.* 1999; **65**: 4734-4740.

[30] Arnold EG, Lenere SCl, Andrew DE (Editors). Standard methods for the examination of water and wastewater, 18th edition. Washington D.C.: American Public Health association, 1992.

[31] Shevchenko A, Wilm M, Vorm O, Mann M. *Anal. Chem*. 1996; **68**: 850–858.

[32] Basaglia M, Toffanin A, Baldan E, Bottegal M, Shapleigh JP, Casella S. *FEMS Microbiol. Lett.* 2007, Jan 15, Epub ahead of print.

[33] Hudman JF, Glenn AR. *Arch Microbiol.* 1984; **140**: 252-256.

[34] Gharieb MM, Gadd GM. *Mycol. Res.* 2004; **108**: 1415-1422.

[35] Vergauwen B, De Vos D, Van Beeumen JJ. *J. Biol. Chem.* 2006; **281**: 4380-94.

[36] Riccillo PM, Muglia CI, de Bruijn FJ, Roe AJ, Booth IR, Aguilar OM. *J Bacteriol.* 2000; **182**: 1748–1753.

[37] Galhardo RS, Almeida CE, Leitao AC, Cabral-Neto JB. *J Bacteriol.* 2000; **182**: 1964-1968.

SECTION II:

DEVELOPMENT AND VALIDATION OF NEW METHODS TO COPE WITH THE “SMALL” AND THE LOW ABUNDANCE PROTEOME

INTRODUCTION: CURRENT LIMITATIONS IN 2-DE

The scientific community has lamented for years that no analytical technique, taken alone, is powerful enough to deal with the extreme complexity of a proteome. This is due to the fact that the current analytical techniques, such as MS and 2-DE, when performed using non-fractionated protein mixtures generate patterns that are frequently too complex for a proper interpretation. Moreover, these approaches are said to be incapable of having a full vision of a whole proteome. For instance, strongly alkaline proteins are poorly represented using classical 2-DE, as underlined by Bae *et al.* [1], while highly hydrophobic proteins cannot be properly solubilised. Electrophoretic methods are neither appropriate for polypeptides with masses lower than 5 000 Da, and only MS contributes significantly to the analysis of small polypeptides. Another noxious phenomenon affecting 2-DE, maybe the most powerful technique in terms of number of proteins detectable in a single experiment, is the so called “spot overlapping” [2], which is caused by the use of too wide IPGs with much too complex protein mixtures. Although the use of narrow or ultra-narrow pH gradients helps minimizing the problem, nevertheless to load an entire proteome leads inevitably to massive protein precipitation of the components focusing out of the pH interval, with the further drawback that the proteins focusing in the desired pH interval are only a small fraction of the total, and therefore under-represented. Another relevant problem is the extremely huge range of concentrations among proteins present in complex biological samples (up to 7 orders of magnitude), that is not matched by the limited dynamic range of the current analytical techniques (3 orders of magnitude). Therefore, the unique way to follow in the attempt to have a more representative picture of a complex proteome is protein pre-fractionation.

One of the oldest but still most valid methods for simplifying a proteome is the separation of cellular sub-structures by the centrifugal cell-fractionation scheme, as originally developed by de Duve’s group [3]. By this procedure, with a series of runs at increasing centrifugal forces, it is possible to purify subcellular organelles such as nuclei, mitochondria, lysosomes, peroxisomes and other microbodies.

A well-consolidated pre-fractionation technique is chromatography, which provides a very large choice of solid-phase sorbents. Various classical

chromatographic separations have been used to obtain simpler protein fractions: size-exclusion, anion and cation-exchange chromatography, lectin affinity chromatography and hydrophobic interaction have been reported along with others. By this approach Fountoulakis's group managed to identify from a 2-D map about 250 new proteins, which had never been detected previously, using in a sequential manner affinity chromatography on heparin gels, chromatofocusing on a polybuffer exchanger and hydrophobic interaction chromatography [4-6].

A proteomics platform functioning on a similar principle is the so-called "Surface-Enhanced Laser Desorption Ionization-Mass Spectrometry" (SELDI-MS), where surface chemistries are used to specifically retain a group of proteins which are subsequently analysed.

Another well established methodology is preparative electrophoresis in gel phases or in liquid media. Free-flow electrophoresis (FFE) was initially proposed by Hannig as a method for purifying cells and subcellular organelles [7]. However, due to the high diffusion coefficient typical of proteins, this tool was not directly applicable to their pre-fractionation, thus another device equipped with multiple collection chambers separated by porous settings rotating around its longitudinal axis was developed by Milan Bier [8]. Bier's device was the first protein pre-fractionation tool performing preparative IEF in absence of anticonvective media and supported by carrier ampholytes. Although this method, in its today's form (Rotofor, Bio-Rad Laboratories, Hercules, CA), allows for the rapid obtainment of 20 pI fractions out of any desired pH interval, its accuracy is quite low, due to the fact that it still works with CA-based pH gradients. In this concern, multicompartment electrolyzers (MCE) operated with immobiline membranes offer much higher pI accuracy [9]. By this technique Pedersen and co-workers managed to visualize on a 2-D map also very alkaline yeast membrane proteins [10], a protein class particularly difficult to investigate with 2-DE.

Besides classical pre-fractionation, which leads to the selection of a particular class of proteins, one might also be interested in analyzing a whole proteome, including both low and high abundant proteins. Protein equalization by solid phase ligand libraries is the revolutionary answer to the well-known problem of excessive concentration differences among proteins constituting complex proteomes. With this technology, the proteins present in a complex sample can be brought to

similar quantity levels, so as to allow for loading detectable levels of the low abundance species as well. The Equalizer Bead technology, whose theory and functioning will be described better in the next chapter, although being in its early stage, has already given promising results. For example, in the last few years considerable research efforts have been devoted to mapping the human urinary proteome, since this is perhaps the unique biological fluid that can be collected in a fully non-invasive manner and in large volumes. When large urine volumes were treated with one of such ligand libraries, not less than 471 unique proteins could be identified by Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry (FTICR-MS) analysis [11], a very impressive result if compared with the data available in the literature up to now, which report a maximum of 150 unique protein annotations obtained via extensive sample pre-fractionation and 2-DE analysis [12]. Another extensively investigated (but extremely challenging) biological fluid is serum, which is thought to contain most of, if not all, the human proteins, including important disease markers. Also in this case the equalization treatment has led to an extremely large number of protein identifications, *i.e.* about 3200 unique gene products [13], a figure totally comparable with that reported by the HUPO Plasma Protein Project, which has published a database containing 3020 human serum proteins [14]. If, on one side, the equalization technology allows for an effective recover of low abundance species from very complex proteomes, on the other side it can also be exploited for the opposite aim, that is to remove low abundance contaminants from quasi-pure protein products, thus leading to virtually 100% pure protein preparations. This would be the case, for example, of recombinant DNA (rDNA) products, which are rapidly becoming very important therapeutic agents, but whose production is challenged by the impossibility of obtaining perfectly pure preparations. In practice, even under the best conditions, it is very laborious to achieve purity grades better than 99%, and the last purification step is generally prohibitively expensive and leads to dramatic loss of valuable bio-pharmaceutical product. In the following chapter, the removal of impurity traces from pseudo-pure protein preparations for human consumption by means of the Equalizer Bead technology will be described.

Another important limitation of the 2-DE technique is the impossibility of studying low molecular mass proteins. In fact, in spite of the enormous progress made in

the field of protein separation since the introduction of SDS-PAGE in 1967 [15], such as the use of discontinuous buffers (e.g. the conventional glycine/HCl system proposed by Laemmli [16] or the boric acid/sulphate system introduced by Neville [17]), which allow for a much higher resolution with respect to zone electrophoresis, or the use of polyacrylamide gradients instead of constant porosity gels, by which one can separate on a single gel proteins with molecular masses spanning the range 10 - 1 000 kDa [18], proteins with molecular masses lower than 10 kDa cannot be efficiently analyzed in polyacrylamide-based matrixes yet. Only the Tris-Tricine system proposed by Schägger and von Jagow [19] permits resolution of molecular masses down to 3 kDa, but this is, up to now, the actual lower limit. In this regard, in collaboration with Cleardirection Ltd. (4 Pekeris St., Rehovot 76702, Israel) we have developed a new electrophoretic method capable of separating extremely small proteins (possibly down to 0.5 kDa) with molecular masses differing by as low as 0.2 kDa. This technique, that we called “SDS-PAGE focusing”, will be described in details in chapter 5.

BIBLIOGRAPHY

[1] Bae SH, Harris AG, Hains PG, Chen H, Garfin DE, Hazell SL, Paik YK, Walsh B, Cordwell SJ. *Proteomics* 2003; **3**: 569–579.

[2] Campostrini N, Areces LB, Rappsilber J, Pietrogrande MC, Dondi F, Pastorino F, Ponzoni M, Righetti PG. *Proteomics*. 2005; **5**: 2385-2395.

[3] de Duve C, Pressman BC, Gianetto R, Wattiaux R, Appelmans F. *Biochem. J.* 1955; **60**: 604–612.

[4] Fountoulakis M, Takacs B, *Prot. Expr. Purif.* 1998; **14**: 113–119.

[5] Fountoulakis M, Langen H, Gray C, Takacs B. *J. Chromatogr. A* 1998: **806**; 279–291.

[6] Fountoulakis M, Takacs MF, Takacs B, *J. Chromatogr. A.* 1999; **833**: 157–168.

[7] Hannig K, *Electrophoresis.* 1982; **3**: 235–242.

[8] Bier M. *Electrophoresis*. 1998; **19**: 1057-63.

[9] Righetti PG, Faupel M, Wenisch E. in: Chrambach A, Dunn MJ, Radola BJ. (Editors). *Advances in Electrophoresis*, Vol. *5*, VCH, Weinheim 1992, pp. 159–200.

[10] Pedersen SK, Harry JL, Sebastian L, Baker J, Traini MD, McCarthy JT, Manoharan A, Wilkins MR, Gooley AA, Righetti PG, Packer NH, Williams KL, Herbert B. *J. Proteome Res.* 2003; **2**: 303–312.

[11] Castagna A, Cecconi D, Sennels L, Rappsilber J, Guerrier L, Fortis F, Boschetti E, Lomas L, Righetti PG. *J. Proteome Res.* 2005; **4**: 1917–1930.

[12] Pieper R, Gatlin CL, McGrath AM, Makusky AJ, Mondal M, Seonarain M, Field E, Schatz CR, Estock MA, Ahmed N, Anderson NG, Steiner S. *Proteomics.* 2004: **4**: 1159–1174.

[13] Sennels L, Salek M, Lomas L, Boschetti E, Righetti PG, Rappsilber J. *Mol. Cell. Proteomics*. 2007; in press.

[14] Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, Kapp EA, Moritz RL, Chan DW, Rai AJ, Admon A, Aebersold R, Eng J, Hancock WS, Hefta SA, Meyer H, Paik YK, Yoo JS, Ping P, Pounds J, Adkins J, Qian X,

Wang R, Wasinger V, Wu CY, Zhao X, Zeng R, Archakov A, Tsugita A, Beer I, Pandey A, Pisano M, Andrews P, Tammen H, Speicher DW, Hanash SM. *Proteomics.* 2005; **5**: 3226–3245.

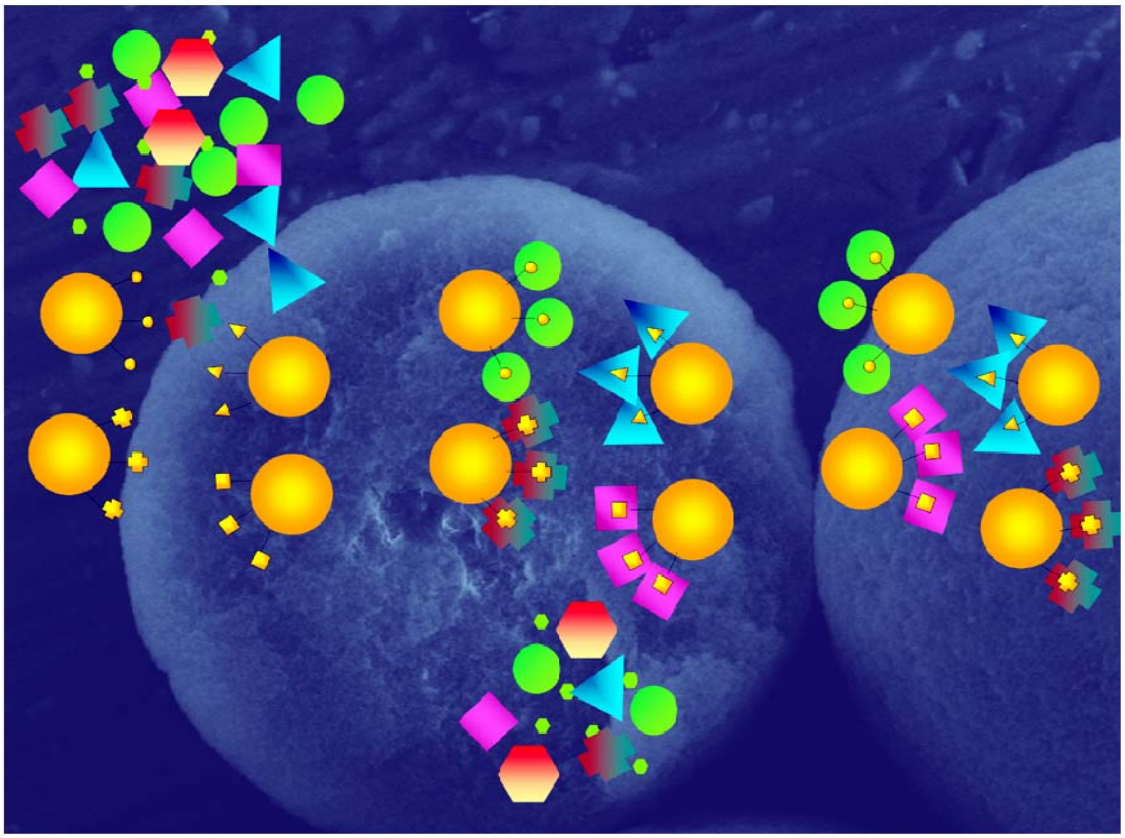
[15] Shapiro AL, Vinuela E, Maizel JV. *Biochem. Biophys. Res. Commun*. 1967; **28**: 815-829.

[16] Laemmli UK. *Nature* 1970; **227**: 680-685.

[17] Neville DM. *J. Biol. Chem*. 1971; **246**: 6328-6334.

[18] Margolis J, Kenrick KG. *Anal. Biochem.* **1968**; **25**: 347-354.

[19] Schägger H, von Jagow G. *Anal. Biochem*. 1987; **166**: 368-379.



CHAPTER IV:

A NEW APPROACH FOR THE REMOVAL OF PROTEIN IMPURITIES FROM PURIFIED BIOLOGICALS USING THE EQUALIZER BEAD TECHNOLOGY

4.1. INTRODUCTION

One of the major dilemmas associated with the production of r-DNA proteins for human consumption is the extraction and purification from very crude feed stocks. In principle, such recombinant proteins should be purified to homogeneity, so as to avoid undesired side effects when they are injected with traces of protein impurities that can be immunogenic, especially if derived from host cell proteins (HCPs) of non-human origin. In practice, even under the best conditions, it is very laborious to achieve purity levels better than 99% [1, 2]. The last purification step, for removing remaining impurity traces, is generally prohibitively expensive and leads to considerable loss of bio-pharmaceutical product. This is the reason why a protein purification process design for biopharmaceuticals starting from crude feed stocks is relatively complicated; it involves several unitary separation steps and implies to consider important parameters such as the relative amount of the target protein in the crude extract, its physicochemical properties and the desired final purity goal. The analysis of these properties is at the basis of the selection of chromatographic resins having orthogonal properties so as to increase the separation efficiency from one step to another. The first chromatographic separation step of the process is based on the selectivity of the resin for the target protein. By that way it is possible to concentrate the protein of interest and to get rid of the maximum amount of protein impurities. This step is generally named “capture” [3]. If the capturing resin is based on a biological affinity ligand, the resulting purification factor in a single step could be very substantial. This is the case of antibody capture using a Protein A column [4] where the purity of the target protein can go up to 95-98%.

The pre-purified protein from the capture phase is then submitted to an intermediate separation process capable of producing fractions where the protein of interest is collected in one or two of them [5-8]. Resins used for this purpose are regular or high capacity ion exchange chromatography (IEC) as well as hydrophobic interaction chromatography resins (HIC). In spite of a stringent selection of chromatographic conditions for the first two steps, the final purity of the target proteins from the second step is never perfect; it contains traces of a number of protein impurities coming from the expression system or from the

biological crude extract. In order to reach purity levels capable to meet regulatory requirements it is necessary to add at least another step called “polishing”, able to eliminate all of the remaining impurity traces. For accomplishing this task the ideal chromatographic resin should be capable of adsorbing all impurities and leaving the target protein in the flowthrough. This configuration would allow reducing the size of the column as compared to a situation in which the target proteins would be adsorbed specifically and impurities eliminated in the flowthrough. Such a definition corresponds to an affinity column; unfortunately protein traces to be removed are constituted of a large number of species and bioaffinity adsorbents corresponding to classical definitions are unpractical. Considering the above context, the most popular approach for polishing is designed around gel permeation supports based on molecular size discrimination [9-11]. Most of the times gel permeation appears as the best choice, because it is orthogonal towards the previous separation steps. However, it suffers from a number of productivity issues that limit the applicability at preparative large scale. In a large number of cases gel filtration had proven good performance in separating impurities essentially because they were of different molecular size. Nevertheless mass similarities between the target protein and impurities create a situation that cannot be resolved by this approach.

Other methods have also been described, based either on HIC [12, 13] or IEC [14, 15]; both methods are process-specific and must be utilized under the best possible conditions for accommodating the adsorption of most of the impurities and the collection of the target protein in the flowthrough.

In the attempt to find a generic process for the removal of impurity traces, a new approach based on the use of a multitude of combinatorial hexapeptide ligands immobilized on polymethacrylate beads (one ligand structure per bead) is reported in this chapter. It demonstrates the ability of the Equalizer Beads to capture impurities yielding the target proteins significantly purer than before the treatment. This approach has been first applied to the polishing of pure sperm whale myoglobin artificially contaminated with either albumin-depleted human serum proteins or *E. coli* protein whole extract, in order to validate the polishing method, after that it has been used for removing low abundance impurities from three tentatively pure protein preparations: recombinant Protein A, expressed in

Escherichia coli and supplied as 99% pure by the manufacturer, human serum albumin (HSA), expressed in *Pichia pastoris* and supplied as a >96% pure product, and injectable monoclonal antibodies (Mabs), produced by hybridoma cells cultured in either standard or protein-free culture medium.

4.1.1. THE "EQUALIZER BEADS": THEORY AND FUNCTIONING

Affinity adsorption is the most selective way for capturing single proteins from very crude dilute feed stocks, adding concentration to a purification effect. After the binding capacity of the solid phase is reached, the process is stopped, the adsorbed protein is recovered and a second cycle can be activated. If this simple mechanism is extended to a large number of different affinity solid phase ligands mixed together, many proteins would be captured at once by their own complementary ligand. Each unique affinity ligand within the pool would bind and concentrate its specific protein partner up to the point of ligand saturation. If in such a mixed bed one loads a complex mixture composed of many proteins, one of them being largely dominant (e.g. 99% purity), this latter will saturate its corresponding solid phase ligand very rapidly, while impurity traces will not. Using loading conditions where the impurities do not saturate the solid phase ligands, the flowthrough of the column will be composed of the dominant protein of purity substantially higher than before the process. Therefore, this process can be used for the removal of impurity traces at the end of downstream separation processes. It is obvious that the bottleneck of this approach consists in the possibility of producing a sufficiently ample spectrum of different affinity ligands. And in fact no one had ever been able to achieve this purpose, until a few years ago, when Thulasiraman and co-workers [16] described the synthesis of solid-phase ligand libraries of huge diversity *via* the "split, couple, recombine" combinatorial chemistry [17-19]. In order to produce a sufficient diversity of affinity ligands, a library of 64 millions of hexapeptides (made by combining the 20 natural amino acids in all the possible ways) was synthesised on polymethacrylate beads by Peptides International (Louisville, Kentucky). Each bead, with an average diameter of 65 μm, carries about 50 pmoles of the same hexapeptide distributed throughout its core. The functioning mechanism of such a technology operated

either in overloading or not overloading conditions (the latter being the one required for achieving a highly purified protein in the flowthrough) is depicted in figure 4.1 (panels A and B, respectively, for the two conditions).

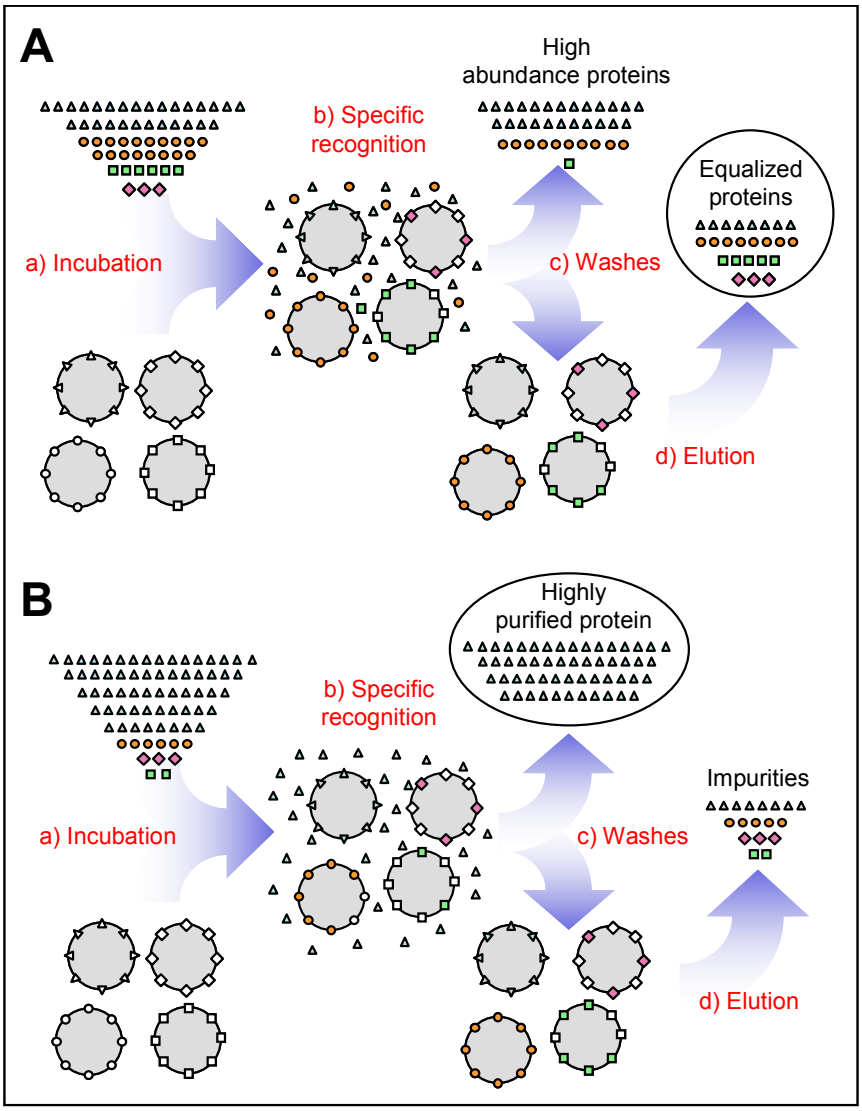


Figure 4.1. Principle of functioning of the Equalizer Bead technology in overloading (A) or not overloading (B) conditions. Samples are first incubated with the Equalizer Bead library (a), and specific recognition takes place (b). Next the unbound proteins are washed away (c) and the equalized sample are finally recovered from the beads.

4.2. MATERIALS AND METHODS

4.2.1. CHEMICALS, BIOLOGICALS AND MATERIALS

The solid phase combinatorial peptide library was prepared by Peptide International, Louisville, KY, according to instructions of American Red Cross. This was first described by Buettner et al. [20] and then modified as recently reported [21].

ProteinChip® Arrays were supplied by Ciphergen Biosystems (Fremont, CA, USA). Biochemicals such as recombinant albumin expressed in *P. pastoris* (95% purity grade), sperm whale myoglobin, molecular mass standards, as well as chemicals such as urea, sodium dodecyl sulfate, glycine, thiourea, 3-[3-cholamidopropyl dimethylammonio]-1-propansulfonate, acetonitrile, trifluoroacetic acid, the fluorescent stain Sypro Ruby, Tris, dithiothreitol, tributylphosphine, bromophenol blue and carrier ampholytes were purchased from Sigma-Aldrich (Steinheim, Germany).

Recombinant Protein A expressed in *E. coli* (99% purity grade) was from Repligen (Waltham, MA).

Injectable 5% HSA solution of 96-98% purity (Albumarc) was originally marketed by American Red Cross, Rockville, MD.

Purified mouse Mabs were from hybridoma cells cultured in two different media: Mab 1 preparation was obtained by culturing hybridoma cells in a basal medium supplemented with bovine serum albumin, bovine insulin and human transferrin. Mab 2 preparation was obtained from hybridoma cultures using a protein-free medium.

Ten-well pre-cast 1 mm-thick 16% Tricine polyacrylamide gel plates and colloidal Coomassie were supplied by Invitrogen (Carlsbad, CA, USA).

Chromatography sorbents for preparative applications such as MEP HyperCel (sorbent for hydrophobic charge induction chromatography) and CM HyperD were from Life Technologies (Grand Island, NY, USA).

The chromatographic columns CM Ceramic HyperD and MEP HyperCel were purchased from BioSepra-Pall (Cergy-Pontoise, France).

Acrylamide/Bis solution (40%), *N*,*N*,*N*',*N*'-tetramethylethylenediamine, acrylamide powder, ammonium persulfate as well as the linear IPGs were obtained from Bio-Rad Laboratories (Hercules, CA).

4.2.2. SAMPLE PREPARATION

4.2.2.1. Mab purification

Mab 1 was purified by hydrophobic charge induction chromatography (HCIC) on a MEP HyperCel column containing 4-mercaptoethyl-pyridine (MEP) [22]. Briefly, the filtered cell culture supernatant was directly loaded onto a MEP HyperCel column pre-equilibrated with physiological PBS. Elution of antibody was performed by lowering the pH with 50 mM acetate buffer pH 4.0. The collected IgG fraction was then stored frozen for further analysis. The purity was estimated as 96% by HPLC on a TSK G4000 SWXL analytical column.

Mab 2 preparation was purified by cation exchange chromatography on DEAE Ceramic HyperD (CM HyperD) [23]. Briefly, the filtered cell culture supernatant was adjusted to pH 4.5 and the ionic strength to a concentration of 0.12 M NaCl. The column was also equilibrated with 50 mM acetate buffer pH 4.5, 0.12 M sodium chloride. Elution was performed by 50 mM acetate buffer pH 4.5 containing 1 M sodium chloride. The purity was estimated of 89%.

4.2.3. POLISHING METHODS

4.2.3.1. Purified sperm whale myoglobin

Commercial myoglobin from sperm whale was first purified by anion-exchange chromatography for removing a number of impurities naturally present in this preparation; after that it was mixed either with human serum proteins (at a ratio of 1.8 or 3.7 mg of protein per 100 mg of pure myoglobin) or with an *E. coli* protein whole extract (at a ratio of 5 mg of protein per 100 mg of pure myoglobin). The buffer used was 25 mM phosphate, pH 7.0. Subsequently, 400 mL of each solution was mixed with 80 mL of Equalizer Beads and incubated under gentle shaking for 20 min at room temperature. The mixture was then centrifuged in a spin column to drain out and collect the liquid phase containing polished

myoglobin. The collected polished myoglobin was then analyzed as described in the following paragraphs.

4.2.3.2. Recombinant HSA, Protein A and injectable HSA

A solution of the purified protein to be polished was prepared in PBS. A column of solid-phase combinatorial peptide library was packed separately (1 mL volume), equilibrated with PBS and connected to a chromatographic setup comprising a pumping system and UV/pH detection unit for recording both events at the column outlet. The column was loaded continuously with the protein solution to be polished at a linear flow rate of 50 cm/h. The flowthrough was collected in fractions of a few millilitres each to analyze the capability of the solid phase to remove protein impurities, likewise a frontal analysis applied for the determination of the binding capacity of sorbents [24]. Once the load was terminated, a PBS solution was introduced for washing out the excess of proteins. Eventually, the captured proteins were desorbed from the beads by a solution of 9 M urea-citric acid pH 3.5, containing 2% CHAPS. In the case of injectable HSA, the recovery before electrophoresis was done using Laemmli buffer at 100°C for 5 minutes. The protein concentration of each loaded solution was 15 mg/mL for Protein A, 10 mg/mL for recombinant HSA, and 50 mg/mL for injectable HSA; the total volumes loaded were 200, 100, and 10 mL, respectively. The desorbed proteins were collected as a single fraction.

4.2.3.3. Mab preparation

One hundred and fifty mg of each purified and lyophilized Mabs were dissolved in PBS at a concentration of 3 mg/mL. Separately, a small chromatographic column was filled with 0.5 mL of solid phase combinatorial peptide library and then equilibrated with PBS. The column was connected to a chromatographic system comprising a pump and a UV/pH detection unit for recording both events at the column outlet. The column was loaded continuously with protein solutions at a linear flow rate of 50 cm/hour. The flow-through was eliminated and, once the loading was completed, a PBS solution was introduced for washing out the proteins in excess. The captured proteins were finally desorbed using a solution of 9 M urea-citric acid, pH 3.3, containing 2% CHAPS. The fractions collected from

both columns, containing 7 mg of amplified impurities, were then dialyzed and lyophilized prior to analysis.

4.2.4. ELECTROPHORETIC ANALYSIS

4.2.4.1. SDS gel electrophoresis

Mono-dimensional SDS-PAGE of collected protein impurities was performed by using 10-well pre-cast 1-mm-thick 16% polyacrylamide Tricine gel plates. Samples of appropriate protein concentration were diluted 2-fold in sample buffer. Thirty µL of diluted sample were loaded per lane and electrophoretic migration was carried out at 150 volts for about 90 minutes. Staining and de-staining were performed with Colloidal Coomassie Blue and 7% acetic acid water solution, respectively.

4.2.4.2. 2-DE analysis

The same protein amount (*i.e.* 150 µg) was used for all the samples investigated (starting material and Equalizer Bead eluates). The protein samples, after being dialyzed against ammonium acetate and subsequently lyophilized, were directly solubilized in IEF buffer (7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP and 10 mM acrylamide) and allowed to be alkylated at room temperature for 60 minutes. In order to stop the alkylation reaction, 10 mM DTT was added to the solution, followed by 0.5% Ampholine and a trace amount of bromophenol blue. Seven-cm long IPG strips pH 3-10 (for myoglobin and recombinant HSA) or pH 4-7 (for recombinant protein A) were rehydrated with 150 µL of protein solution for 4 hours. IEF was carried out with an initial voltage gradient from 100 up to 1000 V, followed by 1000 volts constant for 5 hours. The voltage was then increased again rapidly up to 5000 volts in 30 min, and kept at such a value until reaching 30 000 Vh. For the second dimension the IPG strips were laid on a 10-20% acrylamide gradient gel, and the electrophoretic run was performed by setting a current of 5 mA/gel for 1h, followed by 10 mA/gel for 1h and 20 mA/gel until the dye front reached the bottom of the gel. Gels were then immediately stained with colloidal Coomassie Blue. Destaining was performed in 7% acetic acid. The 2-DE gels were scanned with a Versa-Doc Imaging System (Model 3000, Bio-Rad, Hercules

CA), and the images were treated with the software PDQuest version 8.0 (Bio-Rad, Hercules CA).

4.2.5. SELDI-MS ANALYSIS

This analysis, based on both protein interaction with a chemically modified surface and MS for the discrimination of masses captured by the surface, was performed according to a previously described method [25]. Briefly, each spot of a ProteinChip® Array (referred to as biochip throughout the text) was equilibrated twice with 5 µL of the indicated array-specific binding buffer for 5 min according to the manufacturer's instructions. Each spot was then loaded with 6 µL of the sample previously diluted with the array binding buffer so as to have the appropriate final protein concentration (i.e. 20 ng/µL). After an incubation period of 30 min with constant shaking, the sample was carefully removed and each spot was washed three times with 5 µL of the binding buffer for 5 min to eliminate non-adsorbed proteins, followed by a quick rinse with deionized water. All the surfaces were air-dried followed by two applications of a saturated solution of sinapinic acid in a mixture of 50% acetonitrile (ACN) in water containing 0.5% aqueous trifluoroacetic acid (TFA) (1 µL of this solution per application with air drying in between). The biochips were then analyzed by SELDI-TOF MS using a Ciphergen PCS 4000 ProteinChip Reader operated in a positive-ion mode, with an ion acceleration potential of 20 kV and a detector voltage of 2.8 kV. The molecular mass range investigated was from 1 kDa to 300 kDa. Time-lag focusing was optimized at either 5 kDa or 70 kDa for low and high mass ranges, respectively. Processing of data included baseline subtraction and external calibration using a mixture of known peptide and protein calibrants. Peak detection (S/N>3) and peak clustering was performed automatically using ProteinChip® Software 3.2. Biochips used throughout this study were CM10 (weak cation exchange), Q10 (strong anion exchange), H50 (hydrophobic surface) and IMAC30 (metal ion chelating surface) loaded with Cu^{++} ions.

4.2.6. PROTEIN IDENTIFICATION *VIA* MS/MS ANALYSIS

4.2.6.1. In-gel digestion

Spots were cut out from Colloidal Coomassie blue stained gels and subjected to in-gel trypsin digestion according to Shevchenko *et al.* [26] with minor modifications. The gel pieces were swollen in a digestion buffer containing 50 mM NH_4HCO_3 and 12.5 ng/µL of trypsin (modified porcine trypsin, sequencing grade, Promega, Madison, WI) in an ice bath. After 30 min the supernatant was removed and discarded, 20 µL of 50 mM NH_4HCO_3 were added to the gel pieces and digestion was allowed to proceed at 37° C overnight. The supernatant containing tryptic peptides was dried by vacuum centrifugation. Prior to mass spectrometric analysis, the peptide mixtures were redissolved in 10 µL of 5% formic acid (FA).

4.2.6.2. Protein identification by MS/MS

Peptide mixtures were separated by using a nanoflow-HPLC system (Ultimate; Switchos; Famos; LC Packings, Amsterdam, The Netherlands). A sample volume of 10 µL was loaded onto a home-made 2 cm long fused silica pre-column (75 µm inner diameter, 375 µm outer diameter; Resprosil C18-AQ, 3 µm, Ammerbuch-Entringen, DE) at a flow rate of 2 µL/min. Sequential elution of peptides was accomplished using a flow rate of 200 nL/min and a linear gradient from solution A (2% acetonitrile, 0.1% FA) to 50% of solution B (98% acetonitrile, 0.1% FA) in 40 minutes over the pre-column in line with a homemade 10-15 cm resolving column (75 µm inner diameter; 375 µm outer diameter, Resprosil C18-AQ, 3 µm, Ammerbuch-Entringen, DE).

Peptides were eluted directly into a High Capacity ion Trap HCTplus (Bruker-Daltonik, Germany). Capillary voltage was 1.5-2 kV and a dry gas flow rate of 10 L/min was used at a temperature of 200°C. The scan range used was from 300 to 1800 m/z. Protein identification was performed by searching in the National Center for Biotechnology Information non-redundant database (NCBInr) using the Mascot program (http://www.matrixscience.com). The following parameters were adopted for database searches: complete carbamidomethylation of cysteines and partial oxidation of methionines, peptide Mass Tolerance ± 1.2 Da, fragment mass tolerance ± 0.9 Da, missed cleavages 2. For positive identification, the score had

to be over the significance threshold level ($P < 0.05$). The MS analysis was carried out in the laboratory of Prof. Zolla (Laboratory of Proteomics, Department of Environmental Sciences, Tuscia University, Viterbo, Italy).

4.3. RESULTS

4.3.1. VALIDATION OF THE POLISHING METHOD

4.3.1.1. Polishing of artificially spiked sperm whale myoglobin

As previously described, when a purified protein still containing a certain number of impurities is put in contact with the library, each bead should adsorb its specific partner (if present). The target protein should be also co-adsorbed, but because of the limited binding capacity of the beads for a single protein species, this would result in a minimal reduction of the overall recovery. The limited binding capacity for a single protein means that the described process should be very performant when a large number of protein impurities are present, but each one in a small amount, as it is very frequent with HCPs. In order to verify this assumption, we performed a preliminary polishing experiment with pure myoglobin artificially contaminated with either human serum proteins (added to the two final ratios of 1.8% and 3.7% with respect to myoglobin) or a whole extract of *E. coli* (large diversity of molecular masses present, 5% final ratio). Figure 4.2 represents the SDS-PAGE analytical data of such a case model. Whatever the impurities present, the combinatorial ligand library was capable of removing or significantly reducing the foreign proteins. In fact, all the serum bands detectable in the contaminated myoglobin samples (lanes “a” and “c”) are not visible any longer after polishing (lanes “b” and “d”). Elimination of 5% *E. coli* proteins was also very efficient, as demonstrated in figure 4.2 B and in figure 4.3, the latter showing the results of a 2-DE analysis (7 cm-long IPGs 3-10 and 10-20% polyacrylamide gels) performed on myoglobin before and after contamination (panels A and B, respectively), and after polishing (panels C). Panel D shows the two-dimensional profile of the *E. coli* impurities recovered after the polishing step.

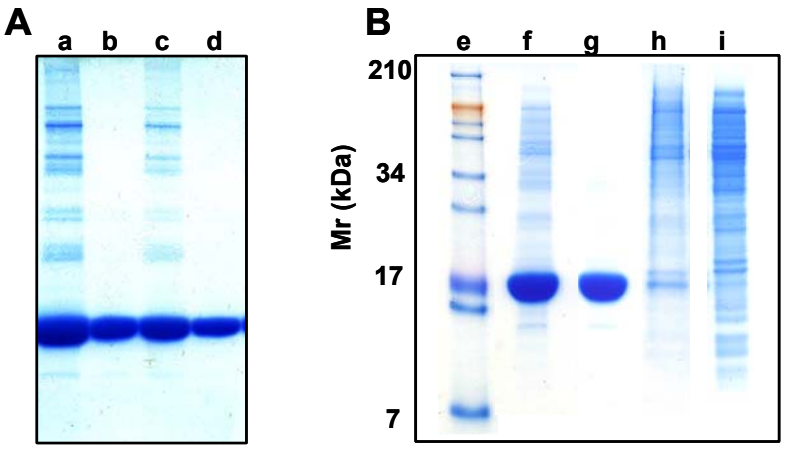


Figure 4.2. SDS-PAGE of myoglobin polished by using an Equalizer Bead library after contamination by, respectively, human serum proteins (A) or 5% of a soluble whole extract of *E. coli* proteins (B). (A) Lanes “a” and “c” represent myoglobin contaminated with, respectively, 3.7% and 1.8% serum proteins; lanes “b” and “d” represent myoglobin after being processed with the ligand library. (B) lane “e” is a molecular mass ladder; lane “f” is myoglobin containing 5% *E. coli* proteins; lane ”g” is polished myoglobin; lane ”h” is the eluate showing the removed contaminating proteins and some traces of myoglobin; lane “i” represents the whole *E. coli* extract.

Before the addition of *E. coli* soluble proteins, myoglobin showed a pretty high level of purity. Few isoforms were present and three main impurities traces were visible at, respectively, p*I* values of about 7.8–8.3, 7.1 and 5.8 and *M*r values of about 35–40 kDa, 6–7 kDa and 12 kDa (figure 4.3, panel A). When supplemented with 5% *E. coli* protein extract, a large number of spots of impurities were visible, with various pIs and masses ranging from 10 to 200 kDa (panel B). After the polishing process, myoglobin appeared roughly as pure as before the contamination: as shown in panel C, only five minor extra spots were still present. Conversely, removed impurities were very numerous (panel D). Traces of myoglobin were also present, because concomitantly captured by the ligand library, as expected (see the arrow). Additional investigations have been performed by SELDI-MS analysis of a Mr window between 1 and 20 kDa, where 2-DE is less performant (especially below 10 kDa). This was done using selected chromatographic chelating surfaces with copper ions (IMAC-Cu^{++} interacting surface).

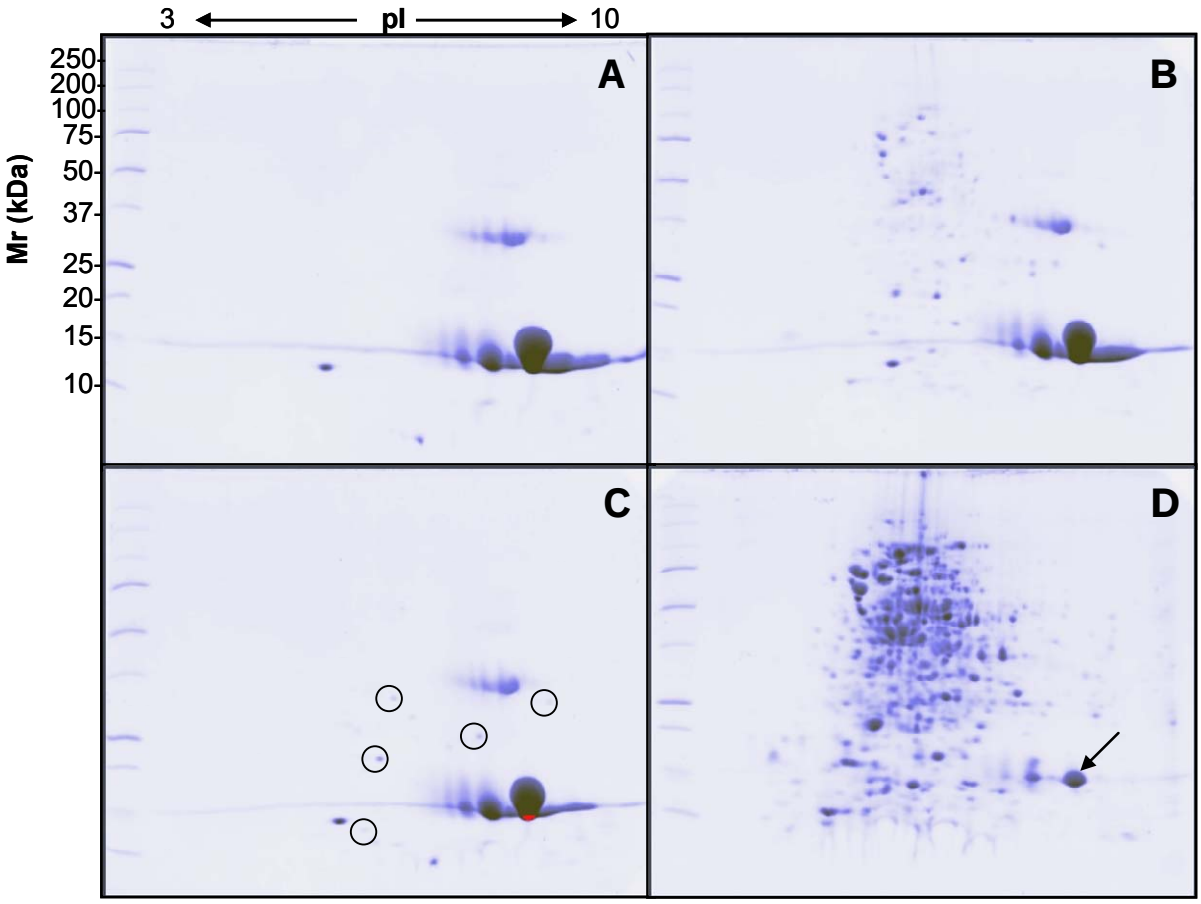


Figure 4.3. 2-DE of samples of myoglobin before and after polishing out *E. coli* protein impurities. p*I* range was between 3 and 10; *M*r analyzed was between 250 and 10 kDa as indicated by the marker ladder. (A) Initial pure myoglobin; (B) myoglobin containing 5% of soluble *E. coli* proteins; (C) myoglobin after having polished out protein impurities (remaining *E. coli* protein traces are highlighted by circles); (D) captured protein impurities by the ligand library (proteins are mostly from *E. coli* extract; the myoglobin trace coadsorbed by the beads is highlighted by an arrow).

As shown in figure 4.4, myoglobin spiked with *E. coli* proteins (A) presented plenty of impurities in addition to regular myoglobin (17.8 kDa, single arrow) and its doubly-charged form (double arrow), confirming thus the data from 2-DE. However, after treatment with the Equalizer Beads, most, if not all, of these protein impurities completely disappeared from the MS spectrum, as displayed in panel B. By comparing panels A and C one can appreciate how the contaminants initially present in the spiked sample were efficiently removed during the polishing step.

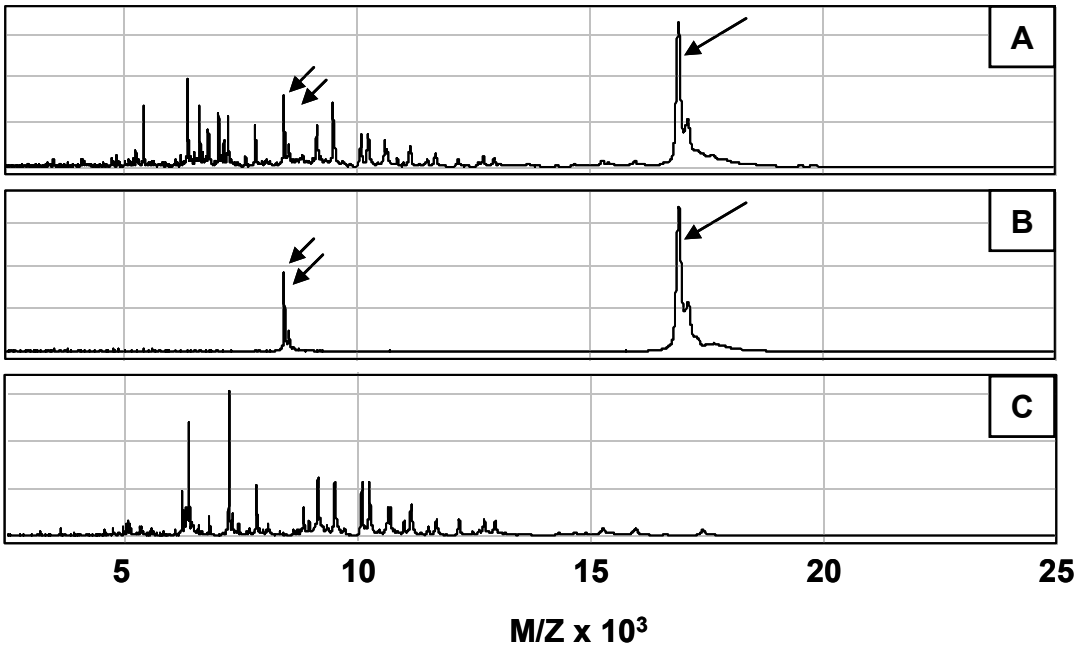


Figure 4.4. SELDI-MS analysis on IMAC-Cu^{++} interacting surface of myoglobin fractions contaminated with 5% of soluble *E. coli* protein extract. (A) Myoglobin contaminated with 5% of foreign proteins; (B) polished myoglobin (single arrow); double arrow represents doubly charged myoglobin. (C) Contaminants removed during the polishing process. All proteins represent only species that interact with copper ions.

The present results clearly demonstrate the Equalizer Bead ability to bind a large number of very diverse proteins in terms of pI and mass, thus confirming the suitability of such a technology for polishing in a purification process.

4.3.1.2. Chromatographic frontal analysis with recombinant HSA

In spite of the fact that the Equalizer Bead technology proved to be able to remove a wide spectrum of different proteins each one present in very small amounts, a very different situation occurs when a similar impurity amount (about 5% w/w) is constituted by only one or two different proteins. In fact, in this case, they may exceed the binding capacity of the partner beads, and thus would not be completely removed. Because each bead present in the library has a binding capacity which is estimated to be approximately 1-3 ng of protein per bead, by calculation 1 mL of beads should be enough to polish 15 mg of protein, but only if the impurity diversity is sufficiently large. Therefore, because the situation of protein polishing is different from case to case, it is highly advisable to proceed by

a preliminary set of experiments to determine the optimum ratio between the sample and the column volume. In practice, optimization of the polishing process design has to be approached by a trial-and-error way, in order to define the best capturing configuration in terms of bead-sample volume ratio. In this respect we have carried out a frontal chromatographic experiment aimed at determining the maximum protein amount to load on a certain beads volume without reaching saturation. In this experiment recombinant HSA (95% pure, as stated by the manufacturer) was used. Here, the number of detectable protein impurities was relatively limited, thus representing a difficult case of polishing, because of the large amount and small number of protein impurities to remove. The result of this experiment is displayed in figure 4.5, which shows the chromatographic profile of recombinant HSA run on a column packed with Equalizer Beads as described in the paragraph 4.2.3.2.

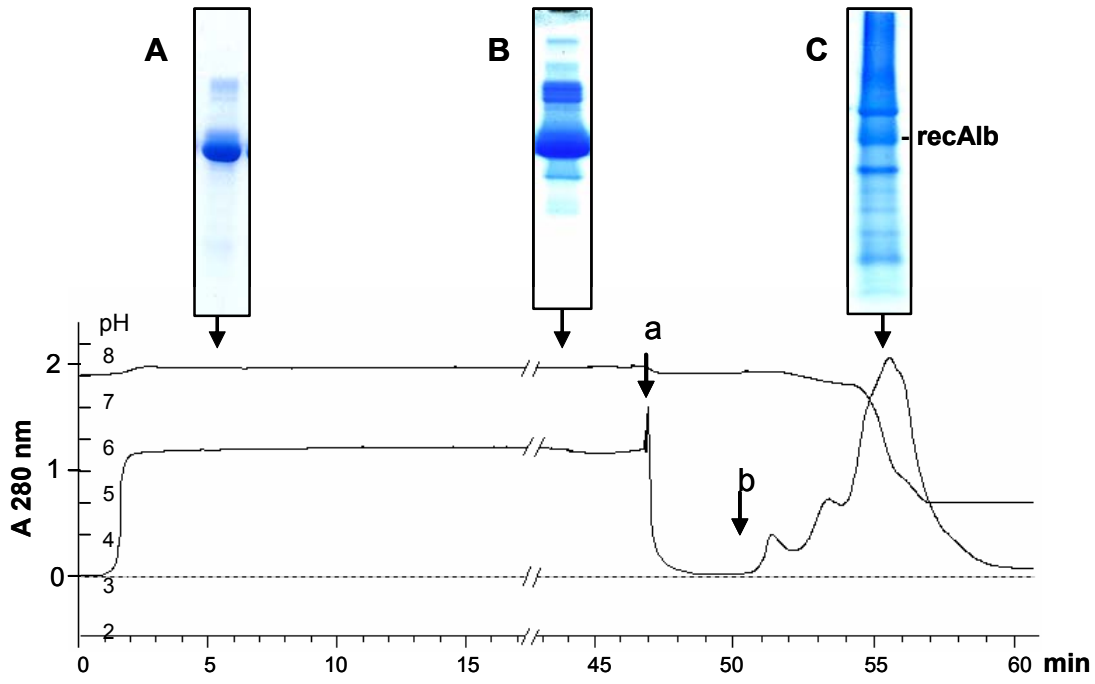


Figure 4.5. Chromatographic frontal analysis with purified recombinant albumin and elution of captured protein impurities. (A) SDS-PAGE analysis of a fraction from the beginning of the chromatographic separation. (B) SDS-PAGE analysis of an intermediate fraction. (C) analysis of the eluate (captured protein impurities) after a washing step with PBS solution (a) and elution by means of a solution of acidic urea (b). Albumin position is indicated (recAlb). Dotted line: base line in plain PBS buffer.

Inserts display the SDS-PAGE analysis of three discrete fractions taken from the flowthrough (two fractions) and from the eluate (1 fraction). In the first part of the flowthrough, HSA showed better purity (insert A), while at the end of the frontal analysis the apparent purity was similar to the initial sample (insert B). However, when analysing the proteins adsorbed onto the combinatorial ligand column, it appeared that a lot of impurities were present, which were not visible in the crude initial sample (insert C). This experiment demonstrates that the column packed with the ligand library has the ability to efficiently adsorb impurities until the limit of capacity dictated by the availability of impurity-specific beads. Although the impurity binding capacity hasn't been measured very accurately, it appears that the column volume for the removal of *P. pastoris* impurities starting from 0.5 gram of total proteins should be at least 2 mL.

MS analysis (SELDI-MS) unambiguously demonstrated that the initial product was pretty rich in impurities: they were dispersed over a wide range of masses, ranging from few thousand Daltons (figure 4.6 panel A) up to about 120 kDa (figure 4.6 panel B).

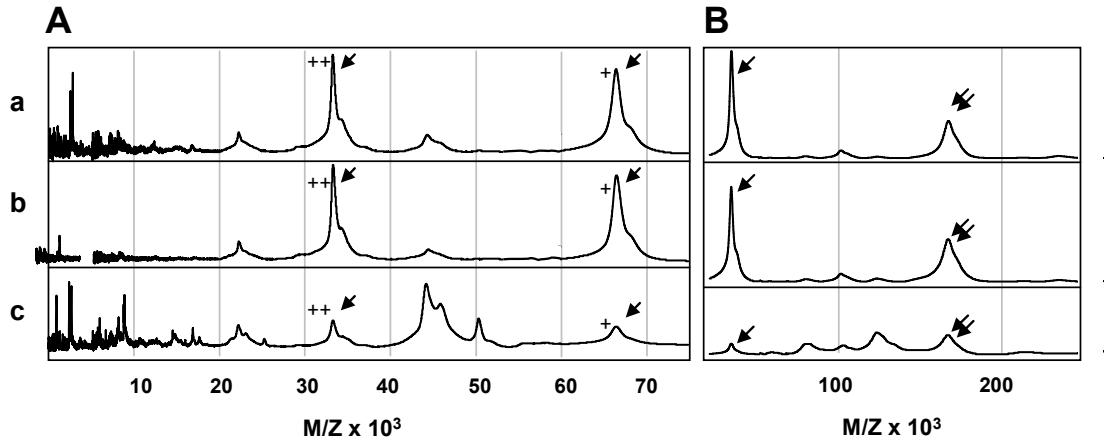


Figure 4.6. SELDI-MS analysis of recombinant HSA from *P. pastoris*. The analysis has been performed by using a CM10 ProteinChip array under classical conditions, and the samples analysed were recombinant HSA before (a) and after (b) polishing, in addition to the protein impurities recovered at the end of the process (c). (A) Molecular mass range up to 75 kDa. Albumin signal is indicated by the arrow; "+" indicates single charged albumin ions, "++" indicates the doubly-charged form. (B) Molecular masses between 60 and over 200 kDa. Single arrow indicates albumin monomer and double arrows indicate the signal of albumin dimer.

After the polishing process the MS spectra revealed that albumin was almost the unique species present, while many different impurities were found in the elution fraction.

4.3.2. AMPLIFICATION AND IDENTIFICATION OF PROTEIN IMPURITIES FROM REAL BIOLOGICAL PRODUCTS

4.3.2.1. Recombinant Protein A expressed in E. coli

We have first analysed the impurities extracted from recombinant Protein A, expressed in *E. coli* and supplied as 99% pure by the manufacturer. Protein A is used mainly as a ligand for the purification of therapeutic Mabs. Fig. 4.7 shows SELDI-MS and SDS-PAGE profiles of the product as commercially available and of the eluate collected after treatment with the Equalizer Beads. It can be appreciated that, in the first case, one main band is visible, with traces of very few minor components. Conversely, in the eluate from the ligand library where all impurities are collected and "amplified" (in that they are highly concentrated from a dilute solution), many more peaks are visible in both tracings.

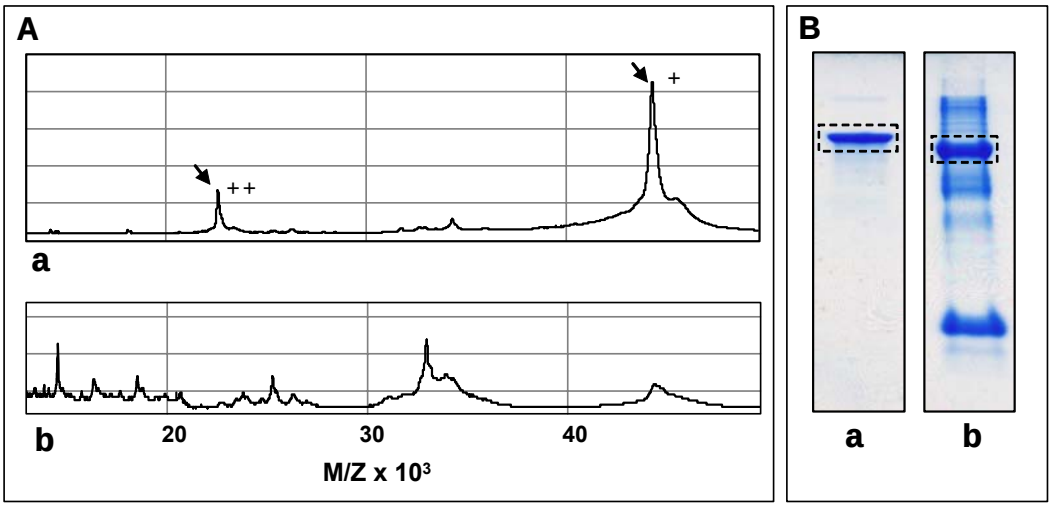


Figure 4.7. SELDI-MS (A) and SDS-PAGE (B) analysis of Protein A (a) and of HCPs impurities extracted using the peptide combinatorial library (b). The dotted boxes in the SDS-PAGE image represent the position of Protein A, while arrows indicate the single (+) and the double charge (++) Protein A ions in the MS spectra.

As revealed by these analyses, the commercial product is purified to a quite good level *via* ion-exchange chromatography, which allows for removing all of the impurities with pI values sufficiently different from that of Protein A. This can also be appreciated in the 2D maps, where all the impurities are seen to be centered around a very narrow pI distribution (figure 4.8). Here too, whereas the control shows the main Protein A component, with at most 3-4 minor impurities, the eluate from the ligand library shows, in addition to the target protein (indicated in both cases with a red arrow), at least 20 and more other species. The main spots in the 2D map have been cut and subjected to MS identification. Table 4.1 gives the protein identities of the 18 spots analysed. It appears that the three most abundant spots visualized after protein equalization represent a tetramer (spot n. 1) a dimer (spot n. 2) and the monomer (spot n. 3). Spots 4-5 and 7-8 are fragments of protein A, probably co-purified due to their similar pI values with the intact protein.

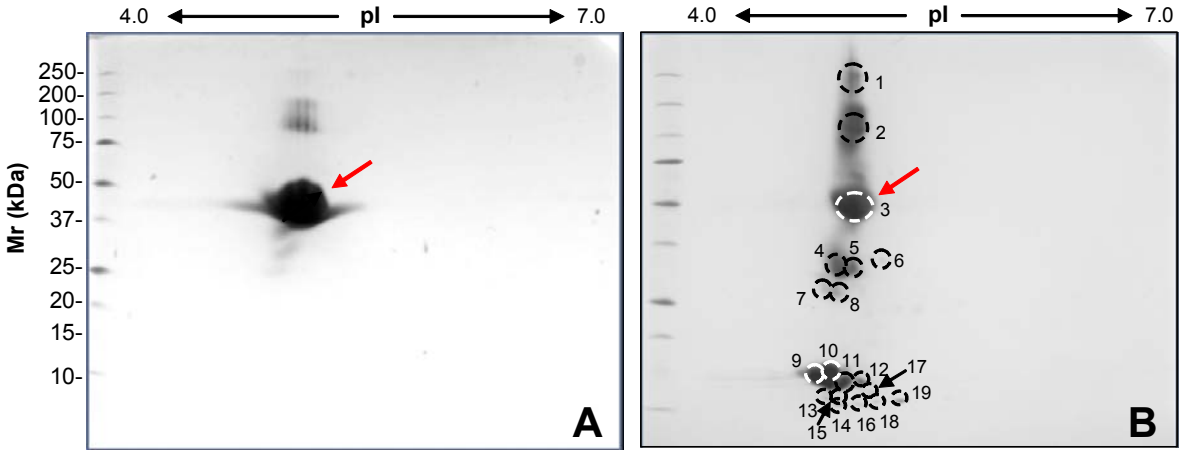


Figure 4.8. Two-dimensional electrophoresis maps of native Protein A (A) and of HCPs impurities extracted using the Equalizer Beads (B). The pI gradient was from 4.0 to 7.0. It is to be noticed that impurities and Protein A have relatively similar isoelectric points, suggesting a pH-driven purification process. Red arrows indicate the positioning of Protein A. The numbers in panel B indicate the spots eluted and subjected to MS analysis (see data in Table 4.1).

The cluster of polypeptide chains centred at pI values from 4.5 to 5.8 and mass values from 10 to 15 kDa are HCPs carried over during the purification process

(probably here too due to the fact that their pI values are clustered around the pI of protein A). It is of interest to note that no one of them was visible in the control sample.

Table 4.1. Protein Identification by LC-MS/MS analysis of the spots cut from 2-D maps of recombinant protein A. Table columns list (from left to right, in the order): spot identity number, NCBI accession number, protein name, sequence coverage, theoretical and experimental pI and Mr values, putative protein form (full length or fragment).

Gel Spot	NCBI Access	Protein ID	Cover. %	Theor. pI & Mr (Da)	Exper. pI & Mr (kDa)	Molecular Forms
1	gi|586027	SPA2_STAAU (Staphylococcal protein A)	36	5.53 55.4	5.3 200.0	PA tetramer
2	gi|586029	SPA2_STAAU (Staphylococcal protein A)	36	5.53 55.4	5.3 100.0	PA dimer
3	gi|586031	SPA2_STAAU (Staphylococcal protein A)	41	5.53 55.4	5.3 52.0	PA monomer
4	gi|586033	SPA2_STAAU (Staphylococcal protein A)	38	5.53 55.4	5.3 33.0	PA frag 1
5	gi|586035	SPA2_STAAU (Staphylococcal protein A)	39	5.53 55.4	5.3 39.0	PA frag 2
6		Unidentified				
7	gi|586026	SPA1_STAAU (Staphylococcal protein A)	26	5.54 57.3	4.8 28.0	PA frag 3
8	gi|586026	SPA1_STAAU (Staphylococcal protein A)	23	5.54 57.3	5.0 27.0	PA frag 4
9	gi|16130920	**Hypothetical protein b3024_E coli K12**	53	5.08 14.1	4.7 15.0	B3024 monomer
10	gi|16130920	Hypothetical protein b3024_E coli K12	51	5.08 14.1	4.9 15.0	B3024 monomer
11	gi|16130920	Hypothetical protein b3024_E coli K12	53	5.08 14.1	5.1 14.0	B3024 monomer
12	gi|16128400	**Riboflavin synthase. beta chain_E coli K12**	32	5.15 16.1	5.4 14.0	RSBC
13	gi|16130920	Hypothetical protein b3024_E coli K12	32	5.08 14.1	4.8 13.0	B3024 frag 1
14	gi|16130920	Hypothetical protein b3024_E coli K12	32	5.08 14.1	5.0 13.0	B3024 frag 2

15	gi|16130920	Hypothetical protein b3024_E coli K12	41	5.08 14 059	5.0 12 000	B3024 frag 3
16	gi|16130920	Hypothetical protein b3024_E coli K12	32	5.08 14.1	5.5 12.0	B3024 frag 4
17	gi|16131554	**Heat shock protein_E coli**	34	5.19 16.3	5.6 13.0	HSP frag 1
18	gi|732038	UPF0076 protein yjgF	34	5.36 13.7	5.8 12.5	YJGF frag 1
19	gi|732038	**UPF0076 protein yjgF**	64	5.36 13.7	6.0 13.0	YJGF monomer

4.3.2.2. HSA expressed in P. pastoris or purified from serum

The second sample analysed was recombinant HSA expressed in *P. pastoris* and supplied as a >96% pure product. Figure 4.9 shows the SDS-PAGE and SELDI-MS profiles of the product as commercially available and of the eluate collected after treatment with the Equalizer Beads.

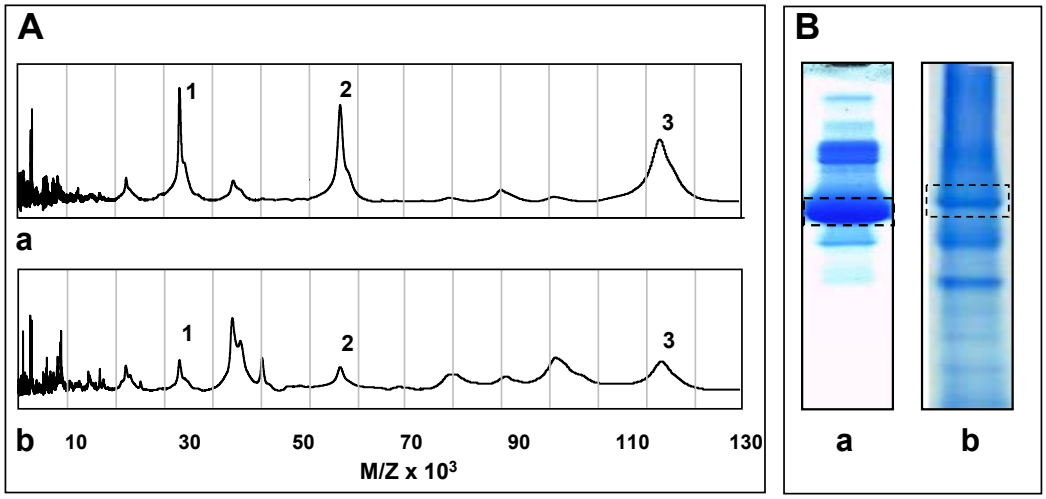


Figure 4.9. SELDI MS (A) and SDS-PAGE (B) analysis of initial recombinant HSA (a) and of HCP impurities (b) extracted by using the peptide combinatorial library. The MS analysis has been done by using a Q10 ProteinChip array. Peaks 1, 2, and 3 represent, respectively, double-charge albumin, single-charge albumin, and albumin dimer. The dotted boxes in the SDS-PAGE images represent the positioning of recombinant HSA in the initial product (a) and in the extract obtained from treatment with the Equalizer Beads (b).

Here, the two most intense peaks in the SDS tracing (as also confirmed by SELDI-MS analysis) represent albumin monomer and dimer (marked as 1 and 2, double and single charge albumin, and 3, representing the dimer). Conversely, the SDS-tracing of the eluate from the Equalizer Beads shows a much higher number of impurities, as also detectable in the SELDI-MS spectra (b). The data displayed in figure 4.9 have been fully validated by the 2-D maps shown in figure 4.10, here too representing the commercial product and the eluate from the ligand library. In this last case, many more protein spots are visible, scattered in the 2-D plane. A total count done with the PDQuest software has revealed more than 50 polypeptide chains. A substantial number (26) of the visible spots in the 2-D map has been excised, digested and analysed by MS/MS. The data are listed in Table 4.2: here the situation is quite different from that of Protein A. In fact, it would appear that most of these spots represent albumin fragment, ranging in mass from the native species (Mr 71 317 Da) to as low as ca. 8 000 Da, as assessed from the second dimension of the 2-D gel.

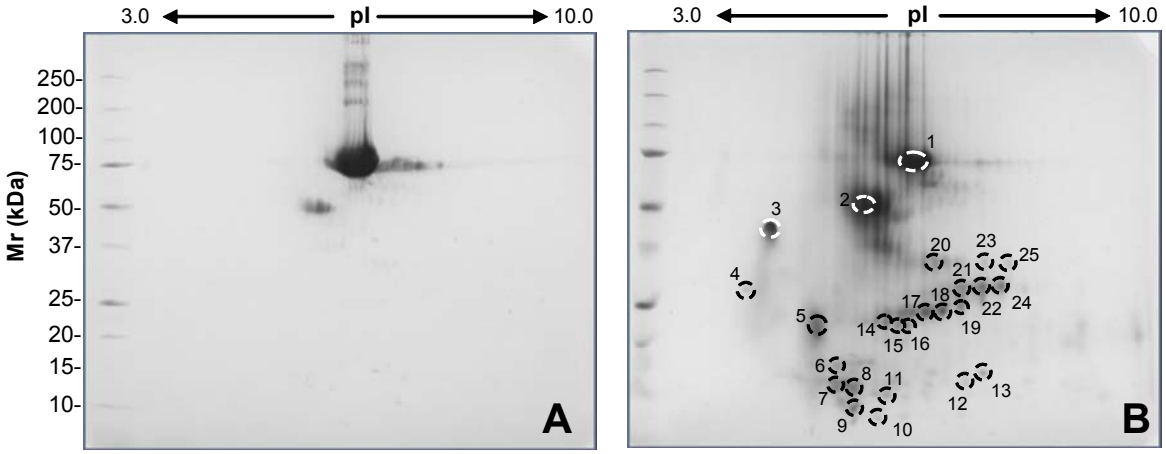


Figure 4.10. Two-dimensional maps of initial purified recombinant HSA (A) and of HCP impurities extracted using the peptide library (B). The pH gradient was from 3.0 to 10.0. Impurities are extended over a large zone of isoelectric points and molecular masses. The numbers in panel B indicate the spots eluted and subjected to MS analysis.

This large number of fragments should have been generated by cleavage induced by proteases present in the culture media. A number of *P. pastoris* proteins have

also been identified among the spots analysed. Here too it is of interest to note that, in the control, essentially no one of these spots could be detected, except for the large albumin fragments with a mass of ca. 50 kDa (spot n. 2).

Table 4.2. Protein Identification by LC-MS/MS analysis of the spots cut from 2-D maps. Table columns list (from left to right, in the order): spot identity number, NCBI accession number, protein name, sequence coverage, theoretical and experimental pI and Mr values, putative protein form (full length or fragment).

Gel Spot	NCBI Access	Protein ID	Cover. %	Theor. pI & Mr (Da)	Exper. pI & Mr (Da)	Molecular Forms
1	gi|113576	**ALBU_HUMAN (Serum albumin precursor)**	35	5.92 71 317	6.5 72 000	HSA monomer
2	gi|113576	ALBU_HUMAN (Serum albumin precursor)	37	5.92 71 317	5.8 52 000	HSA frag 1
3	gi|28590	ALBU_HUMAN (Serum albumin precursor)		5.92 71 246	4.2 45 000	HSA frag 2
4		Unidentified				
5	gi|28590	ALBU_HUMAN (Serum albumin precursor)	23	5.92 71 246	5.0 23 000	HSA frag 3
6	gi|28590	ALBU_HUMAN (Serum albumin precursor)	9	5.92 71 246	5.3 17 000	HSA frag 4
7	gi|28590	ALBU_HUMAN (Serum albumin precursor)	12	5.92 71 246	5.2 14 000	HSA frag 5
8	gi|2104961	**Alcohol oxidase (*P. pastoris*)**	3	6.04 74 479	5.7 12 000	AO frag 1
9	gi|28590	ALBU_HUMAN (Serum albumin precursor)	14	5.92 71 246	5.6 10 000	HSA frag 6
10	gi|28590	ALBU_HUMAN (Serum albumin precursor)	12	5.92 71 246	5.9 8 000	HSA frag 7
11	gi|28590	ALBU_HUMAN (Serum albumin precursor)	13	5.92 71 246	6.1 12 000	HSA frag 8
12	gi|28590	ALBU_HUMAN (Serum albumin precursor)	5	5.92 71 246	7.3 14 000	HSA frag 9
13		Unidentified				
14	gi|28590	ALBU_HUMAN (Serum albumin precursor)	22	5.92 71 246	6.1 23 000	HSA frag 10
15	gi|28590	ALBU_HUMAN (Serum albumin precursor)	19	5.92 71 246	6.2 22 000	HSA frag 11
16	gi|28590	ALBU_HUMAN (Serum albumin precursor)	11	5.92 71 246	6.4 21 500	HSA frag 12
17	gi|28590	ALBU_HUMAN (Serum albumin precursor)	25	5.92 71 246	6.6 24 000	HSA frag 13
18	gi|2495229	**Imidazoleglycerol-phosphate dehydratase (*P. pastoris*)**	18	6.14 24 322	6.8 24 000	IGPD monomer
19	gi|2495229	Imidazoleglycerol-phosphate dehydratase (*P.* pastoris)	18	6.14 24 322	7.2 24 000	IGPD monomer

20	gi|28592	ALBU_HUMAN (Serum albumin precursor)	32	6.05 71 319	6.8 35 000	HSA frag 14
21	gi|28592	ALBU_HUMAN (Serum albumin precursor)	16	6.05 71 319	7.3 30 000	HSA frag 15
22	gi|46981803	**Ornithine carbamoyltransferase (*P. pastoris*)**	18	6.46 37 832	7.5 30 000	OCMT frag 1
23	gi|28592	ALBU_HUMAN (Serum albumin precursor)	27	6.05 71 319	7.5 34 000	HSA frag 16
24	gi|28592	ALBU_HUMAN (Serum albumin precursor)	28	6.05 71 319	7.5 30 000	HSA frag 17
25	gi|28592	ALBU_HUMAN (Serum albumin precursor)	23	6.05 71 319	8.0 34 000	HSA frag 18
26	gi|46981803	Ornithine carbamoyltransferase (*P. pastoris*)	11	6.46 37 832	10.0 23 000	OCMT frag 2

Similarly, a third protein, injectable HSA purified from serum and supplied as 96-98% pure, has been analyzed before and after impurity "amplification" with the hexapeptide combinatorial library beads. SDS-PAGE analysis demonstrated that the product was only apparently pure, whereas, following treatment, many additional protein traces could be observed. By the same token, LC-MS/MS analysis of the product identified only a single impurity (haptoglobin), whereas, after treatment, many impurities were identified (figure 4.11). Table 4.3 assembles the data obtained from LC-MS/MS analysis of amplified HCPs.

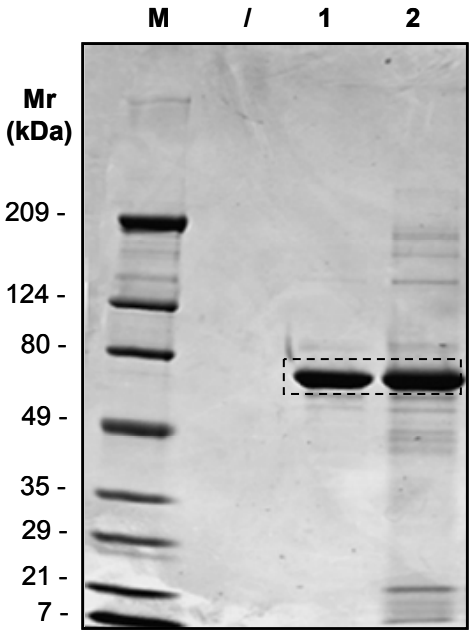


Figure 4.11. SDS-PAGE analysis of injectable HSA purified from human serum (lane 1) and of the proteins captured by treatment with the Equalizer Beads (lane 2). M is a ladder of standard molecular masses. The dotted boxes indicate the positioning of albumin.

Table 4.3. Protein Identification by LC-MS/MS analysis of the bands cut from SDS-PAGE separation of injectable albumin purified from human serum. Table columns list (from left to right, in the order): protein name, Mascot score, sequence coverage and the number of unique peptide sequences.

Identified proteins	Mascot Score	Coverage %	Unique peptides
Before treatment			
Human serum albumin	418.3	41	21
Haptoglobin	20.2	4	1
After treatment with Equalizer Beads			
Human serum albumin	520.3	44	18
Haptoglobin	214.2	26	7
Hemopexin	108.2	18	4
Alpha2-HS glycoprotein	60.2	2	2
Ferroxidase	50.3	7	4
Afamin precursor	40.2	9	3
Human beta-2-glycoprotein-I	40.1	10	2
Peptidoglycan recognition protein L precursor	30.2	7	2
C1 esterase inhibitor	30.2	9	3
Alpha-1B-glycoprotein precursor	20.3	6	2
Transthyretin	20.3	41	2
Hemoglobin chain D	10.2	11	1
Vitamin D-binding protein	10.2	5	1
Immunoglobulin lambda chain	10.1	8	1

4.3.2.3. Monoclonal antibodies expressed in mouse hybridoma cells

The field of antibody-based therapeutics is of growing interest, as attested by the increasing number of development projects at various stages of clinical investigation. In this concern, we have chosen two types of Mabs, differing for the culture conditions in which mouse hybridoma cells were grown (protein-rich medium, Mab 1, or protein-free medium, Mab 2) and for the chromatographic approach used for purification. In fact, these antibodies were purified either with standard IEC on a CM HyperD column (Mab 2) [23] or with a novel type of resin (Mab 1), containing 4-mercaptoethyl pyridine (MEP) [22] and working on the principle of hydrophobic charge induction chromatography (HCIC) [27]. In the case of Mab 1 purification, the MEP HyperCel column was preferred because it

had demonstrated much better performance in antibody purification with respect to CM HyperD. Also in this case we have used the hexapeptide ligand library in order to amplify and study the residual protein impurities present in the Mab preparations that could represent a threat for human health. Fig. 4.12 shows the SDS-PAGE profiles (under non-reducing conditions) of the two Mab preparations. It can be appreciated that the main IgG band (ca. 160 kDa) appears as the prominent species, with a few contaminants exhibiting higher and lower Mr values. The monoclonals purified *via* CM-HyperD, additionally, appear to be somewhat cleaner that those treated with MEP-HyperCel. This was surely due to the fact that, in the first case, the culture medium was protein-free. However, also with these samples the situation significantly changed after incubation with the Equalizer Beads. Here now many more impurities appeared distributed all along the electrophoretic tracks.

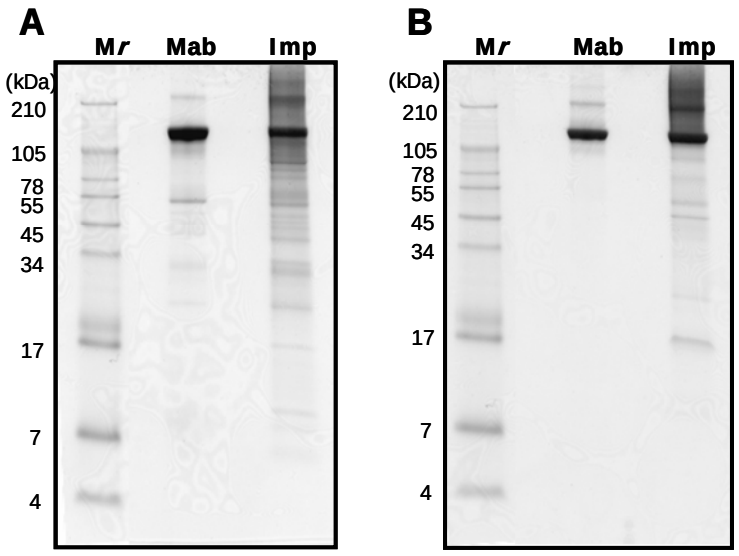


Figure 4.12. SDS-PAGE of Mabs purified with MEP-HyperCel (A, Mab 1) or CM-HyperD (B, Mab 2). Mr: molecular mass ladder; Mab: monoclonal antibody preparation (control); Imp: eluates from the Equalizer Bead library.

As shown in figure 4.13, when the MEP-HyperCel purified monoclonals are run in a 2D map under reducing conditions, the bands exhibited by the control (panel A) are mostly due to the heavy (ca. 50 kDa) and to the light (ca. 25 kDa) immunoglobulin chains, each one exhibiting a horizontal string of bands, as customary in IgG spectra. Additionally, a few of these bands appear to aggregate into dimers and higher-order aggregates, notably in the case of heavy chains. Non-IgG bands are only a few spots scattered in the pI/Mr plane.

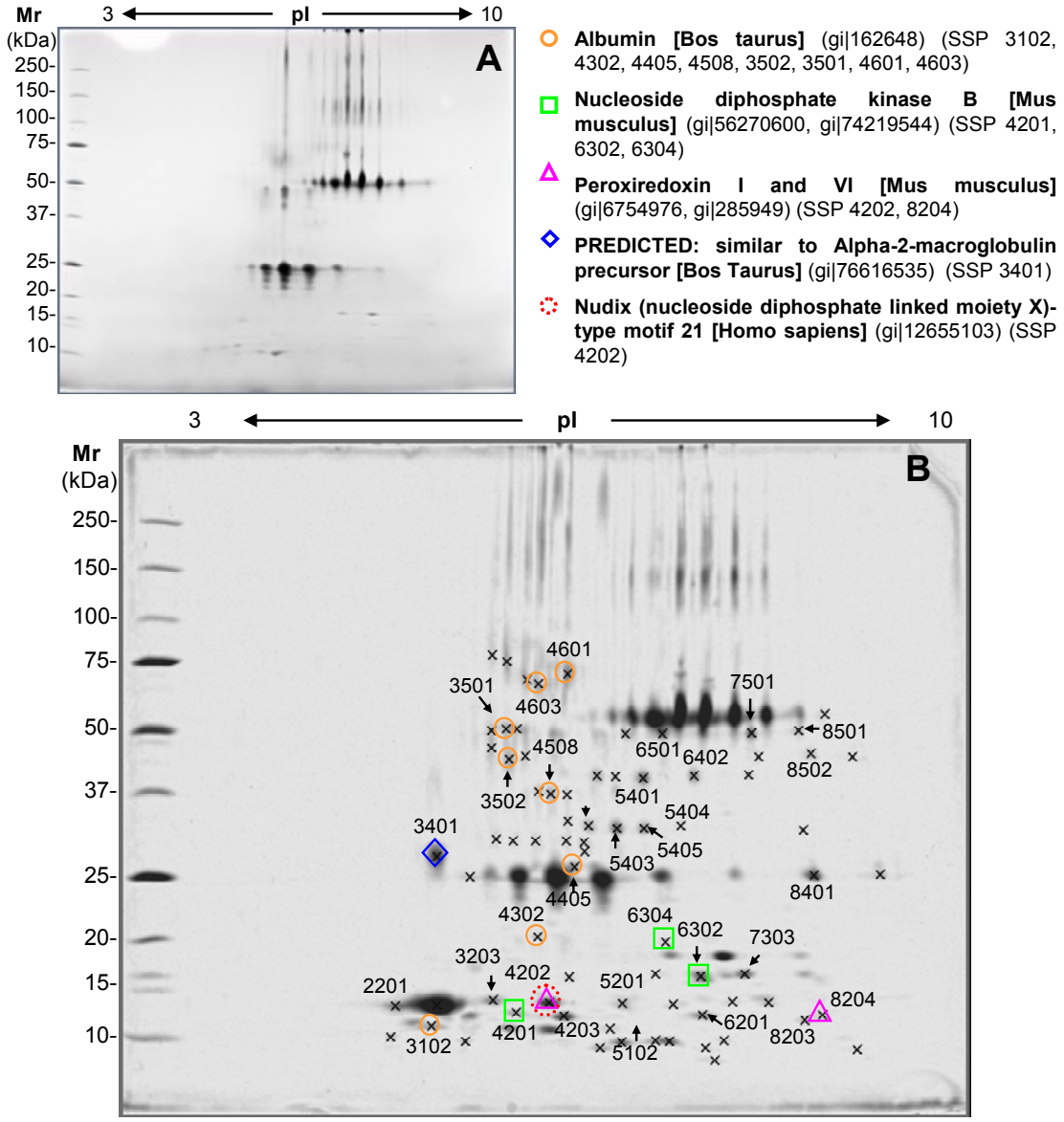


Figure 4.13. 2D patterns of the Mab 1 preparation, purified with a MEP-HyperCel resin from cells cultured in a protein-rich medium. (A) control (untreated) antibodies; (B) eluate from the Equalizer Beads, with crosses marking the new spots detected after equalization, their SSP identification numbers (see Table 4.4) and other symbols indicating some of the impurities identified by MS-MS.

On the contrary, the Equalizer Bead eluate shows a much increased number of spots, mostly distributed in the Mr region below the 25 kDa chains and in between the light and heavy chains (panel B). By PDQuest analysis at least 75 additional spots were counted in the equalized Mab 1. The major ones and some other spots present both in start and equalized material have been eluted and sent for

MS identification. The results are shown in Table 4.4. Analogous data could be obtained in the case of antibodies purified by means of CM-HyperD resin (Mab 2, figure 4.14). However, in this case, a lower number of extra spots could be detected in the equalized sample.

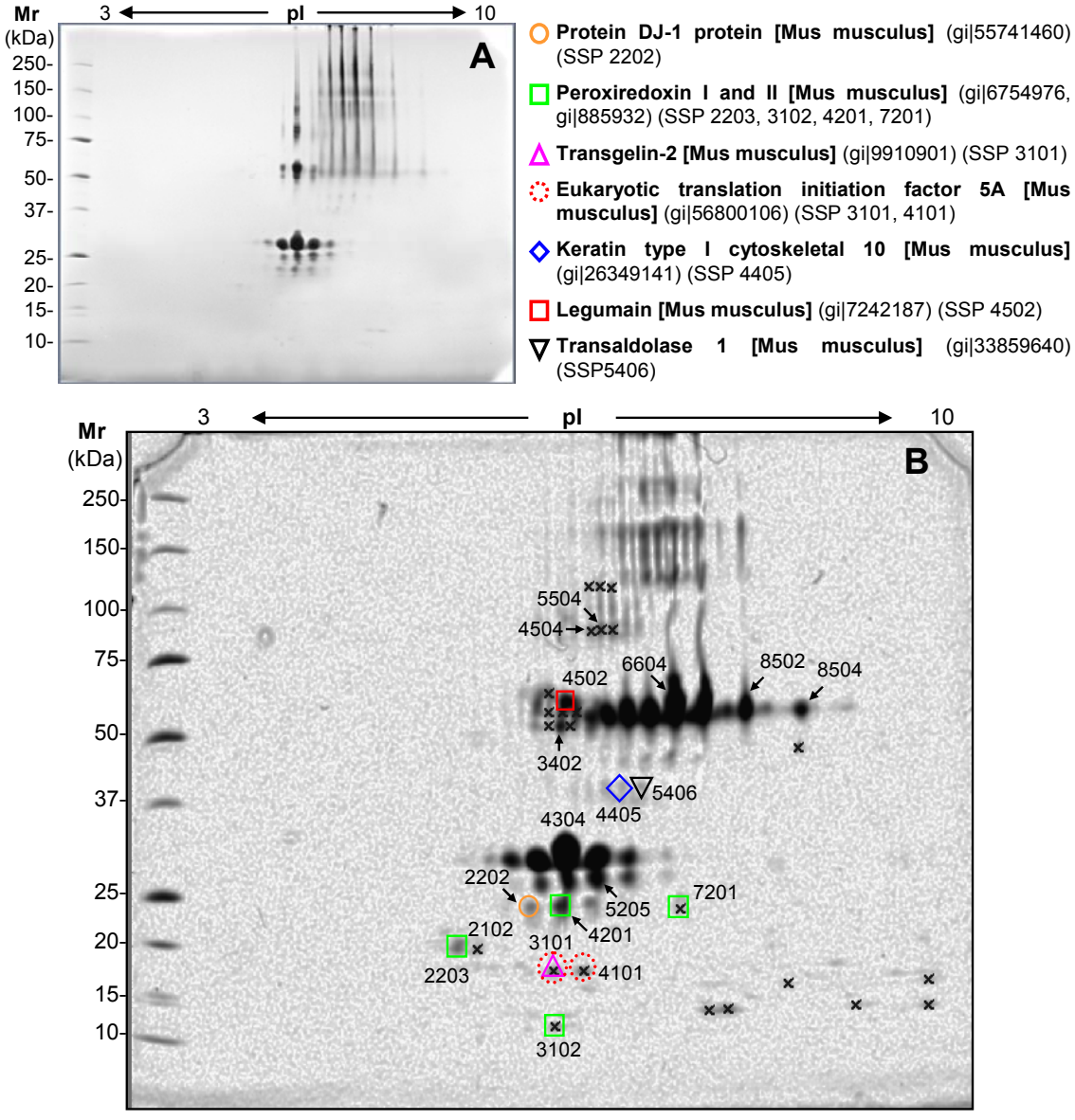


Figure 4.14. 2D patterns of Mab 2 preparation, deriving from cells cultured in protein-free media and purified with a CM-HyperD resin. (A) control (untreated) antibodies; (B) eluate from the Equalizer Beads with crosses marking the new spots detected after equalization, their SSP identification numbers (see Table 4.5) and other symbols indicating some of the impurities identified by MS-MS.

In fact this preparation appeared to be decidedly more pure than Mab 1, thus confirming that a smaller amount of protein impurities can be obtained by using protein-free culture media. The spots identified by MS/MS analysis are listed in table 4.5.

Table 4.4. Proteins identified by LC-MS/MS from the MEP-HyperCel purified antibody preparation (Mab 1).

SSP	N. of identif. Pept.	Mascot Score	Cover. (%)	NCBI Accession number	Theor. pI Mr (kDa)	Exp. pI Mr (kDa)	Protein Definition (NCBI)
2201	4	224	-	Multiple entries	-	4.73 12.7	Immunoglobulin [*Mus musculus*][a]
3102	6	242	10	gi|162648	5.82 69.3	5.02 10.9	**Albumin** [*Bos taurus*] (fragment)
3203	3	149	7	gi|21071089	6.33 50.9	5.51 13.2	Monoclonal antibody BBK-1 heavy chain [*Mus musculus*] (fragment)
3401	2	86	1	gi|76616535	5.64 168.1	5.06 27.4	**PREDICTED: similar to Alpha-2-macroglobulin precursor (Alpha-2-M)** (Protease inhibitor, able to inhibit all four classes of proteinases) **[*Bos Taurus*]** (fragment)
3501	19	967	27	gi|30794280	5.82 69.3	5.68 51.0	**Albumin** [*Bos taurus*] (fragment)
3502	18	941	29	gi|30794280	5.82 69.3	5.72 44.0	**Albumin** [*Bos taurus*] (fragment)
4201	2	74	20	gi|56270600	6.97 17.4	5.81 12.2	**Nme2 protein** (Nucleoside diphosphate kinase B) **[*Mus musculus*]**
4202	3	147	10	gi|74151031	8.85 26.2	6.17 13.0	**Unnamed protein product (cleavage and polyadenylation specific factor 5) [*Mus musculus*]**
	3	122	16	gi|4558336	5.37 23.6		Chain L. Igg1 Fab Fragment (58.2) Complex With 12-Residue Cyclic Peptide (Including Residues 315-324 Of Hiv-1 Gp120) (Mn Isolate) [*Mus musculus*]
	2	88	8	gi|285949	6.00 25.0		**KIAA0106** (Peroxiredoxin-6 is involved in redox regulation of the cell)

							[Homo sapiens]
4203	1	110	16	gi|7438683	9.39 11.5	6.28 11.9	Ig kappa chain [Mus musculus]
4302	3	119	4	gi|162648	5.82 69.3	6.1 20.0	**Albumin** [*Bos taurus*] (fragment)
4405	4	177	5	gi|162648	5.82 69.3	6.36 26.0	**Albumin** [*Bos taurus*] (fragment)
4508	10	452	14	gi|30794280	5.82 69.3	6.19 37.2	**Albumin** [*Bos taurus*] (fragment)
4601	18	1032	32	gi|162648	5.82 69.3	6.3 64.9	**Albumin** [*Bos taurus*]
4603	18	924	27	gi|30794280	5.82 69.3	6.1 62.1	**Albumin** [*Bos taurus*]
5102	1	85	16	gi|52603	9.20 13.1	6.77 9.6	Immunoglobulin heavy chain (V-region) [*Mus musculus*] [b]
5201	1	61	-	Multiple entries	-	6.78 12.9	Immunoglobulin [*Mus musculus*] [a]
5401	6	311	-	Multiple entries	-	6.48 31.8	Immunoglobulin [*Mus musculus*] [a]
5403	6	345	-	Multiple entries	-	6.73 31.5	Immunoglobulin [*Mus musculus*] [a]
5404	4	300	18	gi|5114265	7.17 48.1	6.98 40.1	Gamma1 heavy chain of Mab7 [*Mus musculus*]
5405	6	320	-	Multiple entries	-	6.99 31.4	Immunoglobulin [*Mus musculus*] [a]
6201	1	59	16	gi|1407764	6.84 11.8	7.57 12.0	IgK light chain [*Mus musculus*]
6304	5	220	39	gi|74219544	6.97 17.5	7.21 19.4	**Unnamed protein product (expressed in non-metastatic cells)** (Nucleoside diphosphate kinase B) **[*Mus musculus*]**
6302	5 3	285 153	- 23	Multiple entries gi|74219544	- 6.97 17.5	 7.56 15.5	Immunoglobulin [*Mus musculus*] [a] **Unnamed protein product (expressed in non-metastatic cells)** (Nucleoside diphosphate kinase B) **[*Mus musculus*]**
6402	3	197	7	gi|5114265	7.17 48.1	7.5 40.6	Gamma1 heavy chain of Mab7 [*Mus musculus*]
6501	7	427	19	gi|5114265	7.17 48.1	7.18 49.9	Gamma1 heavy chain of Mab7 [*Mus musculus*]
7303	2	103	-	Multiple entries	-	7.94 15.7	Immunoglobulin [*Mus musculus*] [a]
7501	7	428	18	gi|5114265	7.17 48.1	8.00 50.0	Gamma1 heavy chain of Mab7 [*Mus musculus*]
8203	2	86	-	Multiple entries	-	8.46 11.5	Immunoglobulin [*Mus musculus*] [a]
8204	2	70	10	gi|6754976	8.26 22.2	8.6 12.0	**Peroxiredoxin 1 [*Mus musculus*]** (fragment)
8401	5	332	35	gi|90724	9.43 13.0	8.53 24.9	Ig heavy chain V region [*Mus musculus*]

8501	2	113	-	Multiple entries	-	8.4 50.7	Immunoglobulin [*Mus musculus*] [a]
8802	1	54	15	gi|74147403	4.85 12.9	8.51 45.2	**Immunoglobulin epsilon heavy chain variable region [*Homo sapiens*]**

[a] Since immunoglobulins are very well conserved, several protein hits match the same set of peptides and, for this reason, the identification cannot be related to a specific protein.
[b]Identification has been obtained by matching also theoretical and experimental Mr and pI values.

Table 4.5. Identified protein spots by LC-MS/MS from the CM-HyperD purified antibody preparation.

SSP	N. of identif. Pept.	Mascot Score	Cover. %	NCBI Accession number	Theor. pI Mr (kDa)	Exp. pI Mr (kDa)	Protein Definition (NCBI)
2102	-	-	-	-	-	5.45 19.0	Not identified
2202	4 4 2	194 175 85	22 11 5	gi|7024437 gi|21312436 gi|55741460	5.39 24.0 5.78 20.6 6.32 20.0	5.95 24.1	Immunoglobulin light chain [*Mus musculus*] **Hypothetical protein LOC69816** [*Mus musculus*] **Protein DJ-1 protein** (May function as redox-sensitive chaperone and as sensor for oxidative stress) **[*Mus musculus*]**
2203	2 2	78 70	- 9	Multiple entries gi|885932	- 5.20 21.8	5.30 19.3	Immunoglobulin [*Mus musculus*] [a] **Peroxiredoxin II [*Mus musculus*]**
3101	3 2	158 103	16 10	gi|9910901 gi|56800106	6.60 23.6 4.85 16.3	6.18 16.9	**Transgelin-2** (Belongs to the calponin family, thin filament-associated proteins) **[*Mus musculus*]** **Eukaryotic translation initiation factor 5A [*Mus musculus*]**
3102	2	70	10	gi|6754976	8.26 22.2	6.2 12.4	**Peroxiredoxin 1 [Mus musculus]**
3204	3 2	122 121	15 -	gi|2392096 Multiple entries	5.45 23.7 -	6.26 22.8	Chain M, Hiv-1 Capsid Protein (P24) Complex With Fab25.3 [*Mus musculus*] [b] Immunoglobulin [*Mus*

							musculus] [a]
3304	3	200	18	gi|7024437	5.39 24.0	6.29 33.6	Immunoglobulin light chain [*Mus musculus*]
3402	3 3	150 136	- 14	Multiple entries gi|4930001	- 6.22 23.7	 6.1 54.1	Immunoglobulin [*Mus musculus*] [a] Chain A, Idiotope-Anti-Idiotope Fab-Fab Complex [*Mus musculus*]
4101	4	146	14	gi|56800106	4.85 16.3	6.4 16.9	**Eukaryotic translation initiation factor 5A [*Mus musculus*]**
4201	4 5	243 181	24 25	gi|7024437 gi|6754976	5.39 24.0 8.26 22.2	6.24 24.2	Immunoglobulin light chain [Mus musculus] **Peroxiredoxin I [*Mus musculus*]**
4304	4	293	22	gi|7766934	6.23 24.0	6.3 31.0	Chain A, Structure Of An Activity Suppressing Fab Fragment To Cytochrome P450 Aromatase [*Mus musculus*] [b]
4305	6	369	30	gi|7766934	6.23 24.0	6.3 31.0	Chain A, Structure Of An Activity Suppressing Fab Fragment To Cytochrome P450 Aromatase [*Mus musculus*] [b]
4405	7 5	405 305	10 -	gi|26349141 Multiple entries	4.91 52.7 -	6.7 40.5	**Keratin type I cytoskeletal 10 [*Mus musculus*]** Immunoglobulin [*Mus musculus*] [a]
4502	4 2 2	251 135 113	22 5 -	gi|7024437 gi|7242187 Multiple entries	5.39 24.0 5.92 49.4 -	6.29 62.4	Immunoglobulin light chain [*Mus musculus*] **Legumain** (Cleaves asparaginyl bonds. Perhaps involved in MHC II processing) **[*Mus musculus*]** Immunoglobulin [*Mus musculus*] [a]
4504	5	253	-	Multiple entries	-	6.45 93.1	Immunoglobulin [*Mus musculus*] [a]
5205	5 3	351 172	30 16	gi|7024437 gi|19527116	5.39 24.0 5.81 25.1	6.48 27.2	Immunoglobulin light chain [*Mus musculus*] **Phosphoserine phosphatase [*Mus musculus*]**

	1	62	3	gi|31565266	6.09 26.0		**Rpia protein** (Belongs to the ribose 5-phosphate isomerase family) **[*Mus musculus*]**
5406	4 2	220 70	- 6	Multiple entries gi|33859640	- 6.57 37.4	6.86 40.6	Immunoglobulin [*Mus musculus*] [a] **Transaldolase 1** (acts in the pentose-phosphate pathway) ***Mus musculus*]**
5504	4 3	238 112	- 15	Multiple entries gi|7024437	- 5.39 24.0	6.51 93.9	Immunoglobulin [*Mus musculus*] [a] Immunoglobulin light chain [*Mus musculus*]
6604	8	574	-	Multiple entries	-	7.17 60.9	Immunoglobulin [*Mus musculus*] [a]
7201	4 2	188 106	18 -	gi|6754976 Multiple entries	8.26 22.1 -	7.2 24.0	**Peroxiredoxin I [*Mus musculus*]** Immunoglobulin [*Mus musculus*] [a]
8502	6	324	-	Multiple entries	-	7.76 60.2	Immunoglobulin [*Mus musculus*] [a]
8504	6	341	-	Multiple entries	-	8.23 59.5	Immunoglobulin [*Mus musculus*] [a]

[a] Since immunoglobulins are very well conserved, several protein hits match the same set of peptides and, for this reason, the identification cannot be related to a specific protein.

[b] Identification has been obtained by matching also theoretical and experimental Mr and pI values.

4.4. DISCUSSION

4.4.1. VALIDATION OF THE POLISHING METHOD

The present approach of removing protein traces from purified proteins demonstrates for the fist time the applicability of combinatorial ligand libraries to downstream processing applications. To date, these libraries had been extensively described for drug discovery [28, 29] or, more recently, for amplifying the signal of low-abundance proteins in proteomics investigations [30-32, see 33

for a review]. Due to their large spectra, these libraries can easily be used under mixed bed configuration for the adsorption of many proteins at a time; consequently they can be used for the adsorption of HCPs present in trace amounts in purified biological preparations. The large number of ligands and their diverse specificity makes it possible to adsorb proteins of different species, independently from their composition and origin, conferring to the process a wide applicability. This has been demonstrated by the removal of *E. coli* protein traces and human plasma polypeptides from artificially spiked myoglobin. Interestingly, in this first pilot experiment, the equalization effect was quite emblematic, since we could obtain a strong enrichment of the low abundance species (*E. coli* proteins) along with an impressive decrease of myoglobin. However, the conditions of polishing have to be accurately adjusted in order to avoid saturation of the ligands by protein impurities, which would render the process less effective. A point to consider, in fact, is that the binding capacity for each individual protein to be removed depends on the number of beads carrying the appropriate ligand, since one single protein interacts with a limited number of beads. The larger the number of different ligands, the lower is the number of individual beads carrying the same ligand in a given volume of beads. While the binding capacity is estimated to be close to 1–3 ng per bead, variations may exist from one bead to another, due to individual kinetics and thermodynamic situations dependent, among other things, on the individual structure and the size of proteins to be removed. As a general rule, the binding capacity for protein impurities has to be determined case by-case by, for instance, chromatographic frontal analysis. For preparative repetitive polishing applications, a few additional considerations must be mentioned. The first is to ascertain that the binding capacity of the ligand library does not decrease over runs. To this end it should always be necessary to completely remove adsorbed proteins using strong desorption agents (e.g. chaotropes and detergents). A second question is related to ligand leakage that would possibly contaminate the polished protein. However, the latter is always collected in the flowthrough, with no change in physicochemical conditions of the solid phase. Therefore, there's no reason for ligand release during this step, while it might occur at the desorption stage, when pH and/or ionic strength are modified, thus weakening the whole construct [32]. Moreover, as concerns possible peptide

leakage due to proteolytic cleavage (e.g. exopeptidases), previous studies had proved that proteases are totally ineffective on the small hexapeptides, for reasons related to the structure of the peptide library, as already reported [29, 31]. Nonetheless, in case of ligand leakage the elimination of the released peptides should be an easy task, because of the very low molecular mass (a few hundred daltons) compared to the size of the polished protein.

4.4.2. ON THE TYPE AND ORIGIN OF IMPURITIES PRESENT IN R-DNA PRODUCTS

Recombinant-DNA proteins represent the latest generation of pharmaceutical products, since quite a few of them (e.g., erythropoietin, human growth hormone, superoxide dismutase, tissue plasminogen activator) are used as therapeutic agents. One of the major dilemmas associated with the production of r-DNA proteins for human consumption is their extraction and purification from very crude feed stocks. In principle, such recombinant proteins should be purified to homogeneity, so as to avoid undesired side effects when they are co-injected with traces of protein impurities that can be immunogenic, especially if deriving from HCPs of non-human origin. In the analysis of r-DNA proteins, two types of impurities are in general expected, which are classified as process-related and product-related compounds.

The process-related impurities derive from the manufacturing process and are classified into three major categories: cell substrate-derived (A), cell culture-derived (B) and downstream-derived (C).

A) Cell substrate-derived impurities include, but are not limited to, proteins and nucleic acids derived from the host organism. For HCPs, a sensitive assay capable of detecting a wide range of protein impurities (e.g., immunoassay) is generally utilized. The level of DNA from the host cells can be detected by direct analysis on the product (PCR or hybridization techniques).

B) Cell culture-derived impurities include inducers, antibiotics, culture proteins (e.g. serum or single proteins), and other components of the medium.

C) Downstream-derived impurities include enzymes, chemical and biochemical processing reagents (e.g., cyanogen bromide, guanidine, oxidizing and reducing

agents), inorganic salts (e.g., heavy metals, arsenic, non metallic ions), solvents, carriers, ligands (e.g., Mabs), and other compounds, particularly from chromatographic columns.

The product-related impurities represent the most frequently encountered molecular variants of the desired compound. Degradation products arising during manufacture and/or storage could often be present in significant amounts. They can be classified as truncated forms, post-translationally modified forms and aggregates.

Although all these contaminants are expected in r-DNA products, they are rarely seen in current practice, because they are in general present at very-low levels, far below the detection limits. The unique performance of the Equalizer Bead technology allows such impurities to be “amplified” to the point of becoming detectable in 2D maps and to appear in sufficient amounts for characterization by MS analysis. In this study we have chosen some r-DNA products of high (~96%) or very high (~99%) purity grade, and we have explored them in the attempt to discover such traces of protein contaminants. The following proteins have been taken into consideration: recombinant Protein A expressed in *E. coli* (99% pure), recombinant HSA expressed in *P. pastoris* (95% pure), injectable HSA purified from human plasma (96-98% pure) and monoclonal IgG antibodies expressed in mouse hybridomas and purified by chromatography on a MEP-HyperCel column or on a CM-HyperD column. The results of our analyses revealed that, in spite of the apparently high purity grade of the protein preparations, all of them, once treated with the hexapeptide ligand library, contain quite a high number of co-purified protein impurities. Such impurities belong either to the class of process-related impurities (in particular the cell substrate-derived and the cell culture-derived) or to the class of product-related impurities (in particular truncated forms and aggregates). For instance, a Protein A preparation, which had been purified from cells cultured in protein-free media, contained mostly HCP impurities (UPF0076 protein yjgF, Heat shock protein, Riboflavin synthase beta chain, Hypothetical protein b3024) along with product-related aggregates and truncated forms. Similar results were obtained with recombinant HSA expressed in *P. pastoris* (the HCPs identified were: Alcohol oxidase, Imidazoleglycerol-phosphate dehydratase, Ornithine carbamoyltransferase), but in this case plenty of albumin

fragments scattered in the 2-D plane were detected as well. This suggests that another very important aspect to consider in the field of purification processes is the intrinsic proteolytic activity. In fact proteases may be present in trace amounts, therefore not easily detectable. Their presence might be evidenced by the formation of protein fragments after having stored the biopharmaceutical for quite long periods of times. In this concern, the amplification of these few protein fragments might help discovering unexpected residual proteolytic activity. The present results also demonstrate the extra valence that 2-D maps have over classical 2D chromatography. In the last case, samples are typically pre-digested and then analysed by MS, therefore, by this technique, there's no way to detect possible fragments in protein preparations.

As concerns the purified IgG Mabs, protein equalization helped evaluating the efficacy of the two columns used for purification and the impact of the culture medium composition on the final product impurity content. In fact we have clearly shown that components of the culture medium result to be present (notably albumin) in the antibodies prepared with MEP-HyperCel column starting from cells grown in a medium containing foreign proteins (Mab 1), together with a high number of truncated forms and a few aggregates. Also various HCPs were detected in the two preparations, although CM-HyperD column seemed to be quite more selective than MEP-HyperCel towards IgGs, as the latter led to final products with many more HCP spots. Thus, in spite of the fact that the total purity grade had been found to be higher for Mab 1 preparation (96% for Mab 1 and 89% for Mab 2, as determined by HPLC on a TSK G4000 SWXL analytical column), the Equalizer Bead treatment has revealed that a major number of different HCPs co-purify with Mabs when MEP HyperCel is used, thus originating a more dangerous hazard for human health (more possibilities of immunogenic response), which could not have been revealed without the use of this innovative technology.

4.4.3. ON THE QUANTITATION OF TRACE IMPURITIES

Although we haven't focused our attention on the possible quantification of the low-level impurities present in our purified monoclonals, nevertheless we can

attempt to make some estimates. Most of the impurities we have identified clearly fall below the sensitivity level of our staining technique (colloidal Coomassie), therefore, considering that such a staining method has a sensitivity of about 10 ng of protein *per* spot and that 150 µg of proteins were loaded on each map, before equalization each impurity had to be present at an approximate concentration lower than 0.007% with respect to the principal protein. Assuming an average molecular mass value of 40 kDa, this would mean that most of the impurities are below the pico-mole level (50-100 femtomoles). A more precise method for estimating the initial impurity level could be the following: the purified sample should be spiked with progressive dilutions of known (and highly purified) protein markers (one would have to use more than one, and in extended Mr and pI ranges, so as to mimic properly the behaviour of the impurities in the sample under investigation). After equalization of the samples containing progressive dilutions, one could see at which dilution the spiked impurities become visible and identifiable. The concentration of such spiked proteins would give a reasonable estimate of the levels of impurities present in the original sample under investigation.

4.5. CONCLUSIONS

In conclusion, the present method of treating r-DNA proteins with ligand libraries achieves two main goals:

i) firstly, it allows for "amplification" of impurities, in that they are so highly concentrated upon adsorption onto the Equalizer Beads that they become visible after elution from the adsorptive material. Such a process, of course, depends on the initial amount of sample treated: the higher the amount of r-DNA product treated, the higher the level of impurities captured and thus detectable upon subsequent analysis;

ii) secondly, it should allow to achieve extremely pure r-DNA products to be released on the market, with minimal loss of the therapeutic product. In fact, just as impurities are captured, also the main product (present, of course, in much larger amounts) will be adsorbed. However, since each

protein present in the sample has, in principle, only one or a few main baits acting as adsorption center, also the main r-DNA product will be captured in very limited amounts, till saturation of its binding sites.

As an additional, important finding, the present results demonstrates that r-DNA products should be always screened via 2-D mapping and MS, so as to detect the presence of cleaved species, which seem to be widely present in r-DNA products (at least in the ones investigated by us). This in turn corroborates the important advantages of the 2-DE technique over current chromatographic methods, which, working with pre-digested proteins, cannot distinguish between full-length proteins and the proteolytic fragments deriving from them.

4.6. BIBLIOGRAPHY

[1] Janson JC. In: Janson JC and Rydén L. (Editors), *Protein Purification*, New York, Wiley–VCH, 1998; pp. 3–40.

[2] Simpson RJ. In Simpson RJ. (Editor), *Purifying Proteins For Proteomics*, Cold Spring Harbor, Cold Spring Harbor Laboratory Press, 2004; pp. 17–40.

[3] Wisniewski R, Boschetti E, Jungbauer A. *Process design considerations for large scale chromatography of biomolecules; in Biotechnology and Biopharmaceutical Manufacturing Processing and Preservation* (volume 2). In Avis KE and Wu VL. (Editors). Interpharm Press, 1996; pp 61-181.

[4] Östlund C, Borwell P, Malm B, *Dev Biol Stand.* 1987; **66**: 367-375.

[5] Duffy S, Moellering BJ, Prior GM, Doyle KR, Prior CP. *BioPharm.* 1989; June: 35-47.

[6] Abendroth J, Chatterjee S, Schomburg D, *J Chromatogr*. *B*. 2000; **737**: 187-194.

[7] Gibert S, Bakalara N, Santarelli X. *J. Chromatogr*. *B*. 2000; **737**: 143-150.

[8] Brion F, Rogerieux F, Noury P, Migeon B, Flammarion P, Thybaud E, Porcher JM, *J Chromatogr. B.* 2000; **737**: 3-12.

[9] Laurent TC, Kilander J. *J Chromatogr.* 1964; **14**: 317-330.

[10] Zanette D, Soffientini A, Sottani C, Sarubbi E. *J. Biotechnol*. 2003; **101**: 275-287.

[11] Dieryck W, Noubhani AM, Coulon D, Santarelli X. *J. Chromatogr. B.* 2003; **786**: 153-159.

[12] Lauer I, Bonnewitz B, Meunier A, Beverini M. *J. Chromatogr. B*. 2000; **737**: 277-284.

[13] Roder C, Krusat T, Reimers K, Werchau H. *J. Chromatogr. B.* 2000; **737**: 97-101.

[14] Mrabet NT, *Biochemistry*. 1992; **31**: 2690-2702.

[15] Herzer S, Kinealy K, Asbury R, Beckett P, Eriksson K, Moore P. *Protein Expr. Purif*. 2003; **28**: 232-240.

[16] Thulasiraman V, Lin S, Gheorghiu L, Lathrop J, Lomas L, Hammond D, Boschetti E. *Electrophoresis.* 2005; **26**: 3561–3571.

[17] Lam KS, Salmon SE, Hersch EM, Hruby VJ. *Nature.* 1991; **354**: 82–84.

[18] Furka A, Sebestyen F, Asgedom M, Dibo G. *Int. J. Pept. Protein Res.* 1991; **37**: 487–493.

[19] Watts AD, Hunt NH, Hambly BD, Chaudhri G. *Electrophoresis.* 1997; **18**: 1086–1091.

[20] Buettner JA, Dadd CA, Baumbach GA, Masecar BL, Hammond DJ. *Intl*. *J. Pept. Protein Res*. 1996; **47**: 70-83.

[21] Hammond DJ, Lathrop JT, US Patent Application 20030-211471: "Method for Detecting Ligands and Targets in a Mixture", November 13, 2003.

[22] Boschetti E, Guerrier L. *Int. J. BioChromatogr*. 2002; **6**: 269-283.

[23] Voute N, Fortis F, Guerrier L, Girot P. *Int. J. Biol. C*. 2000 ; **5** : 49-65.

[24] Bjorklund M, Hearn MT. *J. Chromatogr. A* 1996; **743**: 145–162.

[25] Wiesner A, *Curr. Pharm. Biotechnol.* 2004; **5**: 45–67.

[26] Shevchenko A, Wilm M, Vorm O, Mann M. *Anal. Chem*. 1996; **68**: 850–858.

[27] Boschetti, E., *TRENDS Biotech*. 2002; **20**: 333-337.

[28] Lam KS, Hruby VJ, Lebl M, Kazmierski WM, Hersh EM, Salmon SE. *Bioorg. Med. Chem. Lett.* 1993; **3**: 419–424.

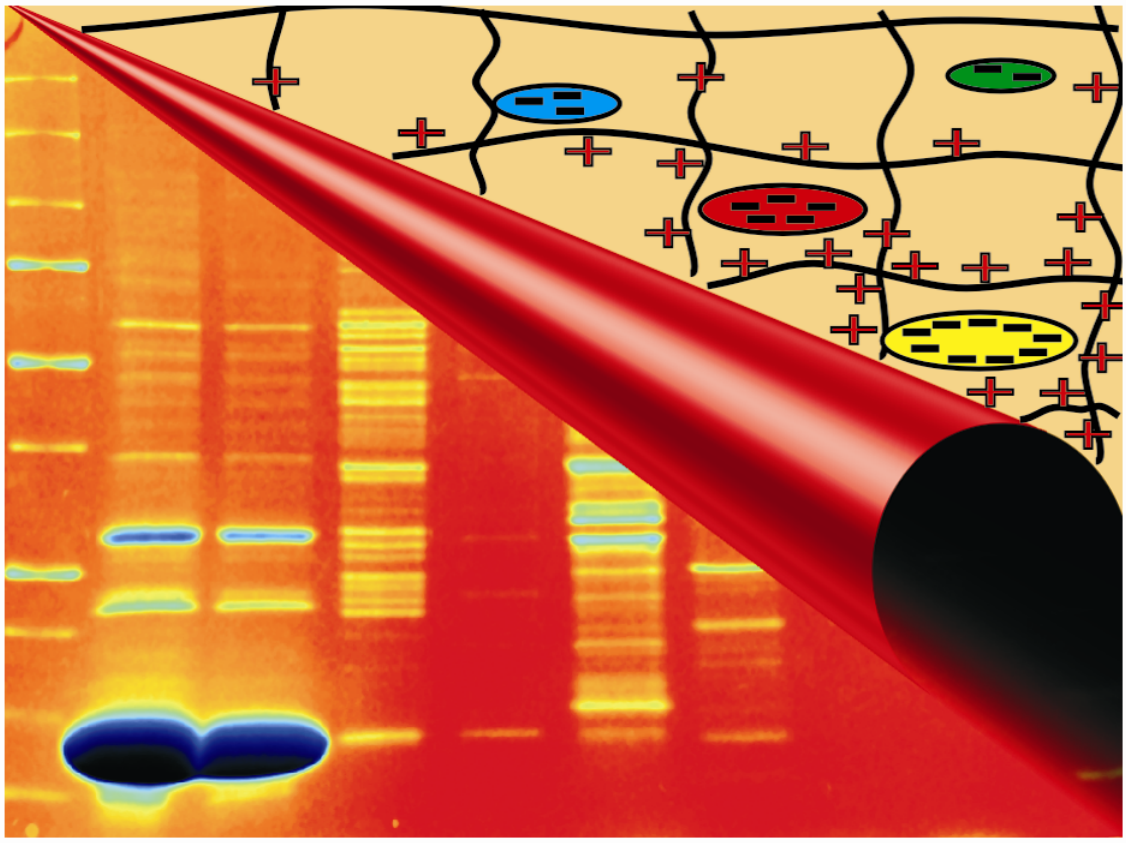
[29] Gordon EM, Barrett RW, Dower WJ, Fodor SPA, Gallop MA*. J. Med. Chem.* 1994; **37**: 1386–1401.

[30] Castagna A, Cecconi D, Sennels L, Rappsilber J, Guerrier L, Fortis F, Boschetti E, Lomas L, Righetti PG. *J. Proteome Res.* 2005; **4**: 1917–1930.

[31] Guerrier L, Thulasiraman V, Castagna A, Fortis F, Lin S, Lomas L, Rigetti PG, Boschetti E. *J. Chromatogr. B.* 2006; **833**: 33–40.

[32] Jungbauer A, Boschetti E, *J. Chromatogr. B.* 1994; **662**: 143–179.

[33] Righetti PG, Boschetti E, Lomas L, Citterio A. *Proteomics*. 2006; **6**: 3980-3992.



CHAPTER V:

SDS-PAGE UNDER FOCUSING CONDITIONS: A NOVEL ELECTROKINETIC TRANSPORT PHENOMENON BASED ON CHARGE NEUTRALIZATION

5.1. INTRODUCTION

In the 1960s, two major events drastically boosted the performance of electrokinetic methodologies. The first one (1960-1964) was the introduction of IEF via the theoretical work of Svensson [1] and the synthesis of carrier ampholytes by Vesterberg [2]. The second one was the discovery that polypeptide chains, when saturated with micelles of an the anionic surfactant sodium dodecyl sulphate (SDS) would reach a nearly constant charge to mass ratio, thus allowing their separation, in a sieving matrix (by polyacrylamide gel electrophoresis, PAGE), mostly according to their molecular mass [3]. The figure of 1.4 g of SDS bound per gram of protein is often quoted as a typical stoichiometric value [4], and this means that the number of SDS molecules bound is of the order of half the number of amino acid residues in the polypeptide chain. This amount of highly charged surfactant molecules is sufficient to overwhelm effectively the intrinsic charges on the polypeptide chain, so that the net charge per unit of mass becomes approximately constant. Electrophoretic migration is then proportional to the effective molecular radius or approximately to the molecular mass of the polypeptide chain. Although it is an oversimplification to which a number of exceptions are to be found [5, 6], the relationship does in fact hold true for a very large number of proteins. Both techniques represented a quantum jump over conventional zone electrophoresis, including disk electrophoresis [7, 8] in that they not only brought about a remarkable resolving power, but also permitted for the first time assessment of some important physicochemical parameters of polypeptide chains, such as their pI value (with a precision of 0.02 pH unit) and their relative molecular mass (with a lower precision, of the order of 3-5 kDa). Just like IEF, also SDS-PAGE was refined over the years, to become perhaps the most popular method today for fast determination of *M*r values of proteins and for assessing their purity during a purification protocol (and, of course, for use as a second dimension in 2-D maps). Its resolving power was substantially augmented by the introduction of discontinuous buffers systems, such as those of Laemmli [9] (the conventional glycine-HCl, terminating/leading ion boundary) and Neville [10] (a boric acid/sulfate boundary), permitting zone sharpening and stacking into very thin zones. When SDS-PAGE is run in constant-concentration gels, the range of

linearity in *M*r assessment (by the standard semilog plot of log Mr *vs.* Rf) spans typically the *M*r 15 000-70 000 Da interval, with changes of slope occurring both below and above these values. In order to greatly expand this Mr interval, polyacrylamide concentration gradients were introduced [11-13]. Gradient gels enabled proteins of widely different sizes to be separated in a single gel, while simultaneously offering superior resolution and sharper bands. The method was thus widely applicable for Mr measurements over the range 13 000-1 000 000 Da by using a 3-30% T gel gradient [12]. Also, the models for the structure of complexes between proteins and SDS have been refined over the years. One of the earliest models proposed on the basis of hydrodynamic measurements is the so-called “rod-like particle model”, in which it is hypothesized that, upon binding of SDS, the polypeptide would form rod-like structures, about 3.6 nm in diameter and 0.074 nm/amino acid residue in length [14]. In contrast with the rod-like particle, a “necklace model” was also proposed. In this second model, the polymer chain is flexible and micelle-like clusters of SDS are scattered along the chain [15, 16]. In an improved version, SDS binds to the protein in the form of spherical micelles and the polypeptide forms α-helices mostly in the hydrophobic region of the micelles. A flexible “α-helices/random coil” structure has also been suggested by Mattice *et al.* [17] on the basis of circular dichroism changes of proteins upon SDS binding. Finally, a flexible helix model, in which the polypeptide chain is helically coiled around an SDS micelle, attached by hydrogen bonds between sulphate group oxygens and peptide bond nitrogens was proposed by Lundahl *et al.* [18]. Over the years, Lundahl’s group has refined this last model [19-22], which has now become the “protein-decorated, micelle model”. Notwithstanding the numerous improvements and refinements achieved in the SDS-based electrokinetic methodologies during the last decades, it’s now quite a long time since this technique has reached a development plateau, and standard protocols have been adopted in most laboratories around the world. As already mentioned in the introduction of this section, such a technique, in spite of the unique performance in protein separation, presents a relevant drawback: it cannot deal properly with protein smaller than 10 kDa (traditional Glycine-HCl system) or 3-4 kDa (Schägger and von Jagow’s Tris-Tricine system, [23]). In this concern, a novel method based on the well-known stochiometry in the SDS-protein complex

is presented here. This method, called “SDS-PAGE focusing”, consists of performing the SDS run in the presence of a gradient of positive charges grafted onto the polyacrylamide gel. As the various protein-SDS micelles, moving towards regions of increasing charge densities, reach a gel zone in which the surrounding charge matches the total amount of negative charges on the surface of the SDS-polypeptide chain complex, the proteins are arrested by a neutralization mechanism that resembles a focusing process, in that zero net charge conditions are reached, impeding further migration of the polypeptide chains. Being a focusing process based on surface charge rather than on particle size (it is performed in a non-sieving matrix), SDS-PAGE focusing allows for separating very small polypeptides as well. Some unique advantages of this novel technique are presented in this chapter.

5.2. MATERIALS AND METHODS

5.2.1. CHEMICALS, BIOLOGICALS AND MATERIALS

Basic Immobiline with various pK values (8.5, 9.3, 10.3 and >12.0) and the peptide marker kit (2.5-16.9 kDa, Catalog No. 80-1129-83) were purchased from GE-Healthcare (Chalfont St. Giles, UK). The pure proteins insulin (5.7 kDa), aprotinin (6.5 kDa), cytochrome (12.3 kDa), lysozyme (14.6 kDa), avidin (16 kDa), myoglobin (17.0 kDa), β-lactoglobulin (18.4 kDa) and bovine serum albumin (BSA, 69.3 kDa) were from Sigma-Aldrich (Steinheim, Germany). Fluka (Buchs, Switzerland) provided acetic acid, methanol, Coomassie Brilliant Blue R, Tris, Tricine and sodium acetate. Acrylamide/bisacrylamide solution (Cat. No. 161-0156), Laemmli Sample Buffer (Cat. No. 161-0737) and Tricine Sample Buffer (Cat. No. 161-0739) were from Bio-Rad (Hercules, CA).

5.2.2. CONVENTIONAL SDS-PAGE

Electrophoresis of marker proteins was performed in classical conditions using 15-well precast 15% T polyacrylamide gel plates (Tris-Tricine) or 10-well

laboratory-made 4-20% T polyacrylamide gel gradients (Tris-Gycine). Protein samples were diluted 2-fold in the appropriate sample buffer (Laemmli or Tricine Sample Buffer for, respectively, Tris-Glycine and Tris-Tricine runs), after that five microliters of diluted protein and peptide marker solutions were loaded per lane, and electrophoresis migration was performed at 200 V for 35 minutes.

5.2.3. SDS-PAGE FOCUSING

Gels were cast with a two-vessel gradient mixer, as routinely performed for pouring IPGs [24]. Mini gels (8 x 7 cm, 0.75 mm thick) were cast in the Bio-Rad system either singly or in a six-unit casting chamber. All gels were supported by a Mylar film (Gel Bond) for easy handling when opening the cassette. The typical gel formulations consist of 4% T polyacrylamide (3.3% cross-linker) in the presence of a gradient of positively charged ions, i.e. basic Immobilines. We have tried different basic species, but the preferred gel formulation adopts the pK 10.3 Immobiline, since this species is fully protonated at the standard running pH of SDS-PAGE (pH 8.3 in 100 mM Tris-Tricine buffer and 0.1% SDS). Different slopes of basic Immobiline gradients have been tested, in the intervals 0-8, 0-12, 0-18 and 0-25 mM. Immediately after casting, the gels were placed in an oven at 50 °C and polymerization proceeded at this temperature for 1 hour. This step is very important, since it ensures good conversion of monomers into the polymeric form and efficient incorporation of charged Immobilines (~90% conversion, [25]). For running, we have tried both vertical and horizontal systems. The preferred configuration was a horizontal setup in the Multiphor II chamber (GE-Healthcare, Chalfont St. Giles, UK). The cathodic electrolyte was a 0.1 M Tris-Tricine solution (pH 8.3) containing 0.1% SDS, while at the anode 0.1 M Tris-acetate (pH 6.4) with 0.1% SDS was used. Contact between electrolyte solutions and gel was ensured by filter pads. Gels were run for various periods of time, from 45 minutes up to 24 hours, in order to test for steady-state “focusing” conditions. In some cases, the SDS-PAGE system was run not in presence of a gradient of a basic Immobiline but by adopting a plateau (constant concentration, 12 mM Immobiline). Typical running conditions are 300 V with forced cooling at 10 °C.

5.2.4. STAINING AND DESTAINING

Staining was by colloidal Coomassie Brilliant Blue R for 1 hour, while destaining was performed with a water solution containing 40% methanol and 7% acetic acid, by incubating in a microwave oven for about 2 minutes with repeated changes of destaining solution.

5.3. RESULTS

5.3.1. SDS-PAGE FOCUSING: THEORY AND PRACTICE

According to the theory of SDS-PAGE focusing, separation should be achieved on the basis of total surface charge, with the smallest polypeptides entering minimally the charge gradient and the largest ones penetrating most, until reaching a surrounding environment able to neutralize the total charge of the SDS-protein complex, which is directly proportional to the protein size. This principle is graphically explained in figure 5.1.

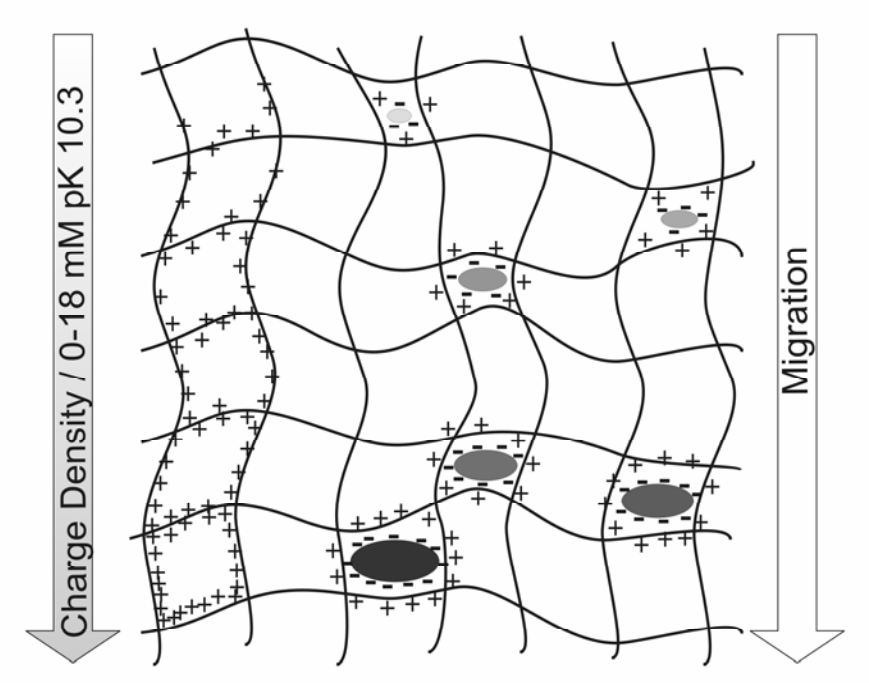


Figure 5.1. Pictogram of the focusing of protein-SDS complexes along a polyacrylamide gel containing a gradient of positive charges.

With this technique, small proteins or peptides are focused in gel regions of low charge density (i.e., close to the application point), whereas larger proteins migrate progressively down the gel length toward regions of higher positive charge densities until reaching an “isoelectric point” due to balancing of the negative charges of the various SDS-protein complexes with positive gel charges. In order to ascertain whether genuine focusing occurs and final steady state is achieved in these conditions, we performed a time course experiment with running times going from 6.5 up to 24 hours, as shown in figure 5.2. Different charge gradients were assayed (0-8, 0-12, 0-18 and 0-25 mM), but real final steady state could be achieved only with the two steepest gradients (0-18 and 0-25 mM).

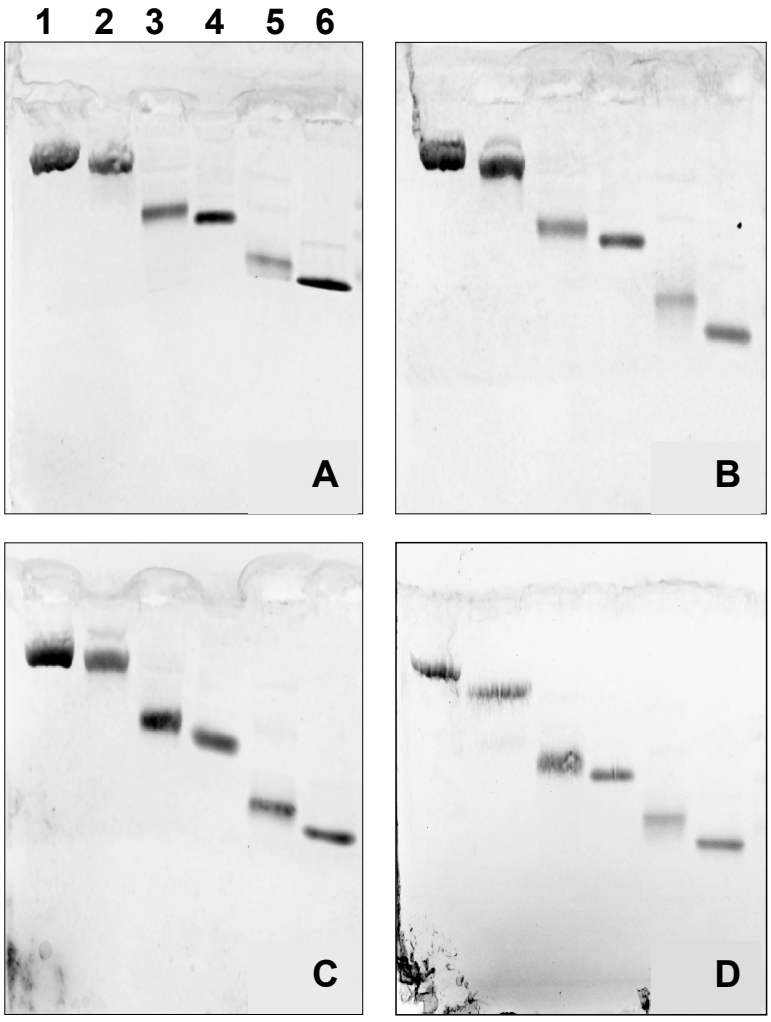


Figure 5.2. Time course of SDS-PAGE focusing. 7-cm long, 4% T polyacrylamide gels with a linear gradient of 0-18 mM pK 10.3 Immobiline. Running conditions: 300 V for 6.5 (A), 10 (B), 18 (C), and 24 (D) hours. Protein bands: (1) insulin, 5.7 kDa; (2) aprotinin, 6.5 kDa; (3) cytochrome, 12.3 kDa; (4) avidin, 16.0 kDa; (5) myoglobin, 17.0 kDa; (6) β-lactoglobulin 18.4 kDa.

As it is visible in the different time panels, with a 0-18 mM Immobiline pK 10.3 gradient, steady-state conditions were reached already after 6.5 hours by the smallest proteins, although, for the largest ones, there was still an asymptotic migration toward the anodic end of the gel. However, the patterns visualized after 10 and 18 hours were identical to that obtained after 24 hours, suggesting that true steady-state conditions had been reached. Here it can also be appreciated that it is the largest proteins that migrate much further down the electrophoretic path, thus confirming what hypothesised in our model.

5.3.2. UNUSUAL ELECTROPHORETIC MIGRATION

As in SDS-PAGE focusing large proteins migrate more than those with low molecular masses through the immobilized charge gradient, this technique exhibits the peculiar feature of producing molecular mass *vs*. migration distance plots with positive slopes. In order to determine the precise function relating these two figures a well-defined polypeptide marker mixture was analyzed on a 0-18 mM Immobiline pK 10.3 gradient. Figure 5.3 shows the results of this experiment: panel A displays the separation of a set of proteins with molecular masses encompassing the 2.5-17.0 kDa range, while panel B gives the corresponding plot for Mr calculation according to relative migration distances. The unique features of this novel system are apparent here: first, and contrary to current dogma in SDS-PAGE, small polypeptides are the ones that exhibit the lowest migration distance, with the larger polypeptides travelling much further down the migration path. As a result, the plot shows a positive slope, whereas in conventional SDS-PAGE, this plot has a negative slope. Additionally, the plot is fully linear on axes that also contain a linear scale (figure 5.3 panel B). Conversely, in standard SDS-PAGE, the plot for Mr assessment is semilog, since typically the Mr axis is on a log scale (for gels of constant porosity). For gels containing porosity gradients, moreover, the Mr plot is performed on a double-log scale (usually log Mr *vs.* log %T, where T is the total concentration (w/v) of acrylamide and bis-acrylamide monomers).

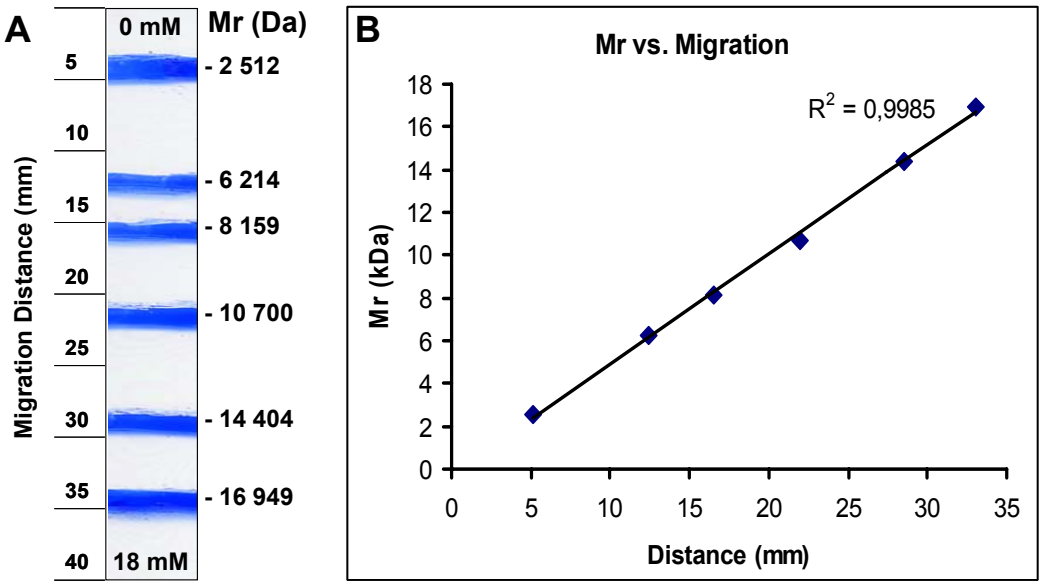


Figure 5.3. Mr *vs.* migration distance. (A) Separation of SDS-polypeptide chains in the Mr 2.5-17.0 kDa. (B) Plot of Mr value *vs.* migration distance for Mr determination. Note the positive slope.

5.3.3. POSSIBLE APPLICATIONS

5.3.3.1. Separation of protein digests

As this peculiar technique provides great advantages in peptide separation with respect to traditional methods, we have tried to apply it to the study of some real samples frequently encountered in the laboratory practice. In the first place, we have compared the performance of traditional SDS-PAGE (performed in a polyacrylamide gradient) and of SDS-PAGE focusing in separating tryptic digests. Three proteins were chosen for this study (lysozyme, myoglobin and avidin), and their tryptic digests were separated either on a traditional porosity gradient (4-20% T) or on a charge density gradient (0-12 mM Immobiline pK 10.3). As shown in figure 5.4, the traditional method is unable to efficiently resolve all the peptides, the majority of which escape the separation range offered by the technique (panel A). Conversely, all the tryptic fragments are easily detected after separation on a 0-12 mM Immobiline pK 10.3 gradient (panel B), where the residual intact proteins focus in front of their tryptic peptides, thus ensuring that no one of them is lost during migration. In this case the focusing effect can be appreciated as well, since

after 18 hours the various bands look as if they were considerably sharper than after 6 hours, and many more components are visualized by Coomassie staining.

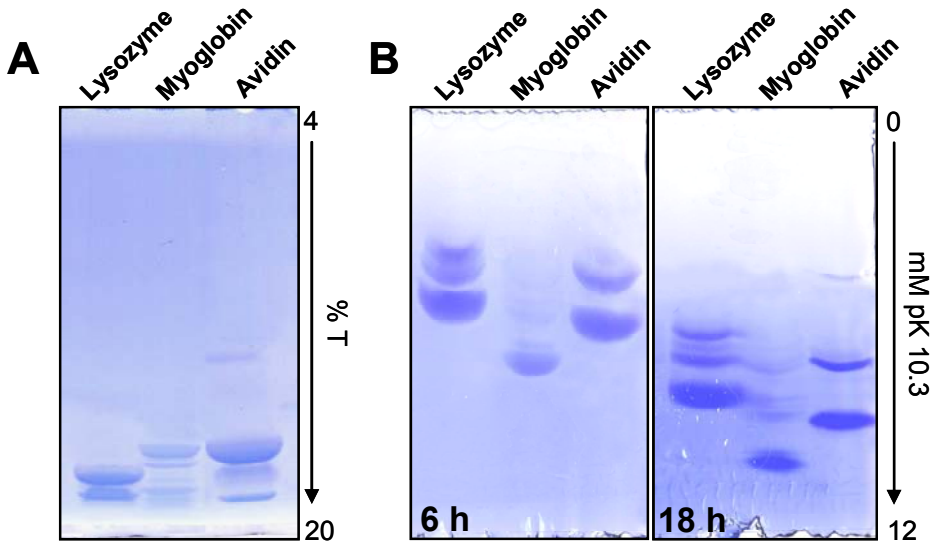


Figure 5.4. Separation of tryptic digests on a traditional SDS-PAGE gel (A, 4-20% T polyacrylamide) and on a positive charge density gradient (B, 0-12 mM Immobiline pK 10.3). In the case of SDS-PAGE focusing, the running times were 6 (panel B, left) or 18 hours (panel B, right).

5.3.3.2. Peptides separation in discontinuous systems

In the attempt to render the electrophoretic profiles of protein digests clearer and "readable", an alternative gel configuration has been assayed. With this configuration, a strongly sieving (e.g., 15% T polyacrylamide) loading zone without Immobiline is cast on the top of the running gel containing the desired charge gradient. According to present knowledge on gel porosity [26], this part of gel should completely retain all the indigested protein, whereas the tryptic peptides should migrate down in the SDS-PAGE gel and focus at appropriate distances along the migration path. In this case, bovine serum albumin (BSA) was analysed on a gel with an upper 15% T loading zone and a lower 4% T running gel (with a 0-25 mM Immobiline pK 10.3 gradient). The gel was loaded with native BSA, with a mixture of native BSA and its tryptic fragments, or with BSA-depleted tryptic fragments. As shown in Figure 5.5, our assumption is fully verified in practice, since no bands could be detected in lane 1 (only BSA), whereas in lanes

2 and 3 (loaded with BSA + tryptic fragments and with only tryptic fragments, respectively) a large number of tryptic peptides focused along the migration path. The two marked black lines visible in the 15% T stacking gel in lanes 1 and 2 represent trapped native BSA.

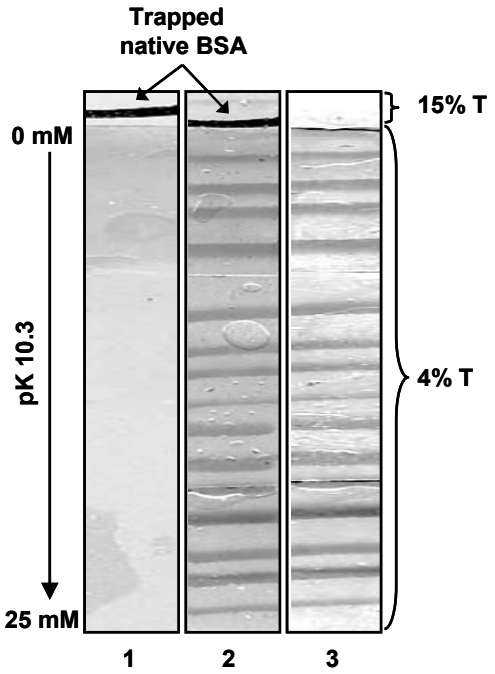


Figure 5.5. Peptide analysis. The upper sieving layer (15% T) served as the trapping zone for the high molecular mass protein; the underlaying 4% T charge gradient gel (0-25 mM Immobiline) served for the separation of peptide fragments. Lane 1, native BSA; lane 2, native BSA and its tryptic fragments; lane 3, tryptic fragments of BSA alone.

5.3.3.3. Separation of closely spaced proteins by "resolving gels"

Another unique aspect of SDS-PAGE focusing can also be appreciated in Figure 5.6. It is well known that, in conventional SDS-PAGE, two neighbouring protein bands cannot be resolved unless their difference in molecular mass is of the order of at least 3 000 Da [27]. Thus, for closely related proteins, envelops of bands of slightly different Mr values are usually experienced along the migration path, this leading to spot overlap in two-dimensional map analysis [28, 29]. However, this situation can be completely resolved by using SDS-PAGE focusing performed in a charge plateau (resolving gel) instead of a charge gradient. By this way we have managed to completely separate two closely related protein species with a molecular mass difference of only 1.5 kDa (7.6% mass difference): myoglobin (17.0 kDa) and β-lactoglobulin (18.5 kDa). As shown in figure 5.5, the two coloured bands appear to be entirely separated in our focusing system (left strip), while in a conventional SDS-PAGE gel (right strip) the two proteins migrate in a single zone, although there is a slight hint at a separation, since the upper zone

boundary is seen in bluish and the lower boundary somewhat reddish. Conversely, in the focusing system, the two proteins are seen separated by 1-cm distance. Thus, it would appear that the present system has the potential of separating zones with molecular masses differing by as little as <1%, considering that a mass difference of 7.6% allows for about 1-cm free gel space and assuming a 1-mm free gel space between two adjacent zones to be an acceptable resolving power. This unique resolving ability was obtained by running our focusing system not in the presence of a charge gradient, as customarily run, but in the presence of a plateau of charges (a gel section with constant concentration of 12 mM p*K* 10.3 Immobiline), this leading to differential charge neutralization along the migration path.

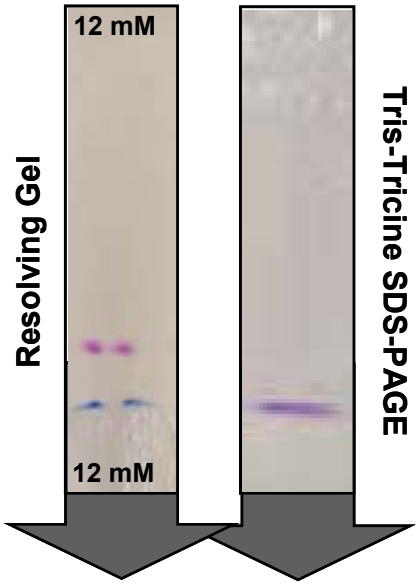


Figure 5.6. Separation of a mixture of myoglobin and β-lactoglobulin conventional SDS-PAGE (right strip) *vs.* resolving gel (left strip, 4% T polyacrylamide gel with a 12 mM constant plateau of pK 10.3 Immobiline). The Mr markers are prestained. Red fraction (1) myoglobin; blue zone (2), β-lactoglobulin.

5.4. DISCUSSION

It could be of interest to review here briefly the historical and scientific background of this novel isotachophoretic technique. IEF was born with the theoretical work of Svensson-Rilbe [1], which disagreed with the wrong notion by Kolin [30] of focusing in non-amphoteric buffers. For proper IEF, the buffers had to be amphoteric, so that they would reach a steady-state position along the migration path, and “carrier” (of buffering power and conductivity at pH = pI) [31]. The second generation of focusing went against the dogma of Svensson-Rilbe, in that

it adopted buffers comprising a series of weak acids and weak bases, all monoprotic, but certainly non-amphoteric. The violation of Svensson-Rilbe's law was only apparent, though: when these buffers were incorporated into the polyacrylamide matrix, and thus insolubilized, each infinitesimal gel layer perpendicular to the migration path was indeed amphoteric, isoelectric, and endowed with strong buffering power. Our present method represents the third way of performing focusing, and thus, perhaps, it closes the circle. In our system, we move from fully soluble buffers (conventional IEF) and fully insoluble species (IPGs) to a hybrid system, in which half of the charges are insolubilized (the basic Immobilines) into the polyacrylamide matrix, whereas the other half (the counterions) are soluble species. In the amphoteric, "isoelectric" complex, thus, a hybrid system, partly soluble, partly insoluble is formed.

5.4.1. ON THE SEPARATION MECHANISM

The novel method of SDS-PAGE focusing presented here offers some unique features worth discussing. First, it is of interest to speculate on the mechanism of separation. The pictogram of figure 5.1 gives an idealized vision of such a mechanism. Along the migration path, from cathode to anode, the polyacrylamide matrix exhibits an increasing concentration of positive charges, grafted onto the neutral matrix backbone, much like IPG matrices do, except that in the latter case the negatively charged, titrating counterion is grafted as well. As the negatively charged SDS-polypeptide chain complex migrates down the electrophoretic path, it encounters a linearly increasing density of positive charges. It is likely that, once a polypeptide chain of a given size meets along the path a density of positive charges matching its negative surface charges, there will be charge neutralization and the macromolecule will come to a stop. This explains the odd migration behaviour, by which, in a sieving matrix, the smaller polypeptides are the ones that are first arrested, whereas the larger peptides/proteins keep migrating toward the anode in regions of higher positive charge densities. Clearly, the smaller polypeptides have a lower total surface charge as compared to larger polypeptide chains, thus their migration is arrested, via this mechanism of charge neutralization, shortly after exiting from the application pocket. As demonstrated

with the time course experiment performed with different charge gradients, the latter should be always chosen on the basis of the specific protein mixture, because larger proteins require higher charge densities in order to reach a final steady state, but the resolution of two neighbouring species is inversely proportional to the gradient steepness. The gradient may also be cast with a starting Immobiline concentration different from zero, so as to obtain shallow gradients capable of separating quite large proteins with small mass differences. At the charge matching point, it is quite likely that the polypeptide chains, especially the largest ones, will distort locally the polyacrylamide fibers, attracting them to their surface, so that charge neutralization will occur efficiently. In fact, ion-to-ion interactions are among the strongest ones in non-covalent links among macromolecules [32]. In this regard, only monovalent ions were used as electrolytes, so as to minimally interfere with the ion-ion interactions and ensure proper focusing of the SDS-protein complex. Given the fact that we use rather dilute matrices (down to only 4% T) endowed with large pores, such local fiber distortion has to occur, since the average pore size will be of the order of a few nanometers. Additionally, gels are considered as viscoelastic media, fluctuating in the surrounding space [33], and this is especially true for dilute matrices, thus such local fiber distortion is a most likely event. The fact that our system works on a focusing principle is also implicit on the fact that the polyacrylamide matrices, being highly diluted, are minimally sieving; thus, they could not arrest the migration of small polypeptides via an approach to a pore limit, as customary in porosity gradient gels [34, 35].

There are other advantages of the present system worth mentioning. First there's the fact that the relationship between Mr and migration distance is a linear function (see figure 5.3) as opposed to a semilog or even double-log function in conventional SDS-PAGE. The replacement of the logarithmic scale with the linear scale provides significant advantages, such as improved accuracy of mass determination and the ability for optimizing the separation for different mass regions, including very low and very high molecular mass ranges. Moreover, such a novel method could also permit estimation of the total surface charge of a given SDS-protein complex. For example, in the case of Figure 5.2, when measuring the distance migrated into the gel by, for example, β-lactoglobulin, and by

assuming a 90% incorporation of the basic Immobiline into the polyacrylamide gel fibers, full protonation of almost all positive charges (given a pK of 10.3 and a running buffer pH of 8.3), and minimal pK variation of the free and bound Immobiline (as verified in practice by Bjellqvist *et al.*, [36]), from its steady-state position one could calculate that the total negative surface charge molarity of the SDS-β-lactoglobulin complex is 9 mM, matching the molarity of Immobiline in the surrounding gel fibers. By the same token, for the smallest polypeptide insulin, one can calculate a negative surface charge molarity of its SDS complex of 3 mM. To our knowledge, conventional SDS-PAGE could not give this type of information on the migrating protein-SDS complexes.

5.4.2. INNOVATIVE APPLICATIONS IN PEPTIDE ANAYSIS

Another interesting aspect of the SDS-PAGE focusing system is that it opens up a window for simple and efficient separation of tryptic digests and peptides in general, as shown in Figures 5.4 and 5.5. From this point of view, it is remarkable that a quite complex tryptic digest could be sharply separated in a strongly dilute polyacrylamide matrix. This has the advantage of allowing for easy extraction of the peptides for further analysis, for example, by mass spectrometry. This is really important indeed, because “peptidomics”, a branch of proteomics devoted to peptide fractionation and analysis, is now considered as central in biomarker discovery, since it is believed that a host of markers for diseases could be fragments of normal proteins abnormally processed during the onset of pathological conditions [37]. Moreover, with the discontinuous configuration presented here, one could easily select a molecular mass window of interest without any need for sample pre-fractionation, thus making the entire study faster and easier.

Another interesting aspect is the possibility of greatly modulating the migration distance of closely spaced proteins by a method that we have called “resolving gel”, as illustrated in figure 5.6. Here a constant charge plateau was used instead of a gradient of charges as typical of the present methodology. In a properly selected charge plateau, the difference of migration between two proteins with small differences in Mr values is magnified as the protein-SDS complexes are

progressively titrated along their migration toward the anode. This method is not a steady-state, focusing method, as is customary of the present technique, but represents a case analogue to the “nonequilibrium IEF” of O’Farrell *et al.* [38], adopted for focusing in very alkaline regions where the lack of suitable carrier ampholytes impeded reaching a true steady state. The possibility of fine-tuning the present method to the separation of closely related species greatly enhances the flexibility of our technique, since it could, in principle, allow for separation of species differing by as little as 150 Da in their Mr values, a resolving power that is at least 1 order of magnitude greater than that exhibited by conventional SDS-PAGE.

5.5. CONCLUSIONS

In conclusion, although we have only partially explored the capability and flexibility of the present methodology, it is likely that further important applications can be realized. SDS-PAGE focusing may become a key technique in the field of peptide analysis. Moreover, it could substitute SDS-PAGE as second dimension in some particular 2-D analyses involving, for example, ultra-narrow molecular mass intervals or very small polypeptides, where its excellent resolving power would be of great help. Lastly, this methodology may find large applications in the separation of natural polyanions like DNA, which would focus in a charge gradient just like protein-SDS complexes do.

5.6. BIBLIOGRAPHY

[1] Svensson H. *Acta Chem. Scand.* 1961; **15**: 325-341.

[2] Vesterberg O. *Acta Chem. Scand.* 1969; **23**: 2653-2666.

[3] Shapiro AL, Vinuela E, Maizel JV. *Biochem. Biophys. Res. Commun.* 1967; **28**: 815-829.

[4] Pitt-Rivers R, Impiombato FSA. *Biochem. J.* 1968; **109**: 825-833.

[5] Weber K, Osborn M. *J. Biol. Chem.* 1969; **244**: 4406-4412.

[6] Dunker AK, Rueckert RR. *J. Biol. Chem.* 1969; **244**: 5074-5080.

[7] Davis BJ. *Ann. N. Y. Acad. Sci*. 1964; **121**: 404-427.

[8] Ornstein L. *Ann. N. Y. Acad. Sci*. 1964; **121**; 321-349.

[9] Laemmli UK. *Nature* 1970; **227**; 680-685.

[10] Neville DM. *J. Biol. Chem*. 1971; **246**: 6328-6334.

[11] Studier FW. *J. Mol. Biol.* 1973; **79**: 237-242.

[12] Lambin P. *Anal. Biochem.* 1978; **85**: 114-125.

[13] Poduslo JF, Rodbard D. *Anal. Biochem*. 1980; **101**: 394-406.

[14] Reynolds JA, Tanford C. *Proc. Natl. Acad. Sci. USA* 1970; **66**: 1002-1007.

[15] Shirahama K, Tsujii K, Takagi T. *J. Biochem*. *(Tokyo).* 1974; **75**: 309-319.

[16] Takagi T, Tsujii K, Shirahama K. *J. Biochem. (Tokyo)*. 1975; **77**: 939-947.

[17] Mattice WL, Riser JM, Ckark DS. *Biochemistry.* 1976; **15**: 4264-4272.

[18] Lundahl P, Greijer E, Sandberg M, Cardell S, Eriksson KO. *Biochim. Biophys. Acta.* 1986; **873**: 20-26.

[19] Mascher E, Lundahl P. *J. Chromatogr.* 1989; **476**: 147-158.

[20] Wallstén M, Lundahl P. *J. Chromatogr*. 1990; **512**: 3-12.

[21] Lundahl P, Watanabe Y, Takagi T. *J. Chromatogr*. 1992; **604**: 95-102.

[22] Ibel K, May RP, Sandberg M, Mascher E, Greijer E, Lundahl P. *Biophys. Chem*. 1994; **53**: 77-84.

[23] Schägger H, von Jagow G. *Anal. Biochem*. 1987; **166**: 368-379.

[24] Righetti PG. I*mmobilized pH Gradients: Theory and Methodology,* Elsevier, Amsterdam, 1990, pp. 127-139.

[25] Righetti PG, Ek K, Bjellqvist B. *J. Chromatogr*. 1984; **291**: 31-42.

[26] Fawcett JS, Morris CJOR. *Sep. Stud.* 1966; **1**: 9-16.

[27] Andrews AT. Electrophoresis: Theory, Techniques, Biochemical and Clinical Applications; Clarendon Press: Oxford, 1986; pp 119-136.

[28] Pietrogrande MC, Marchetti N, Dondi F, Righetti PG. *Electrophoresis* 2002; **23**: 283-291.

[29] Pietrogrande MC, Marchetti N, Dondi F, Righetti PG. Electrophoresis 2003; **24**: 217-224.

[30] Kolin A. In *Electrofocusing and Isotachophoresis*; Radola, B. J., Graesslin, D., Eds.; de Guyter: Berlin, 1977; pp 3-34.

[31] Righetti PG. *Isoelectric Focusing: Theory, Methodology and Applications*; Elsevier: Amsterdam, 1983; pp 15-39.

[32] van Holde KE, Johnson WC, Ho PS. *Principles of Physical Biochemistry*; Prentice Hall: Upper Saddle River, NJ, 1998; pp 95-99.

[33] Bode HJ. *Anal. Biochem*. 1977; 83: 204-210.

[34] Margolis J, Kenrick KG. *Anal. Biochem.* 1968; **25**: 347-354.

[35] Margolis J. *Lab. Pract.* 1973; **22**: 107-112.

[36] Bjellqvist B, Ek K, Righetti PG, Gianazza E, Görg A, Westermeier R, Postel W. *J. Biochem. Biophys. Methods.* 1982; **6**: 317-339.

[37] Srinivas PR, Verma M, Zhao Y, Srivastava S. *Clin. Chem.* 2002; **48**: 1160-1169.

[38] O'Farrell PZ, Goodman HM, O'Farrell PH. *Cell* 1977; **12**: 1133-1141.

PUBLICATIONS

REGULAR ARTICLE

Art-loving bugs: The resurrection of Spinello Aretino from Pisa's cemetery

Paolo Antonioli[1], Giacomo Zapparoli[1], Pamela Abbruscato[2], Claudia Sorlini[2], Giancarlo Ranalli[3] and Pier Giorgio Righetti[1]

[1] Department of Industrial and Agricultural Biotechnologies, University of Verona, Verona, Italy
[2] Department of Food Science and Microbiology, University of Milan, Milan, Italy
[3] Department of Agri-Food, Environmental and Microbiological Science and Technology, University of Molise, Campobasso, Italy

Received: October 15, 2004
Revised: November 26, 2004
Accepted: December 1, 2004

During 1391–1392, Spinello Aretino painted a cycle of frescoes in Pisa's cemetery on the theme "*Storie dei Santi Martiri Efisio e Potito*", highly prized by contemporaries and by Vasari. Twenty years ago, one of these frescoes, "*Conversione di S. Efisio e Battaglia*" (Conversion of S. Efisio and Battle), because of discolouring and bad damage caused by humidity and atmospheric pollution, was removed from the walls using the "tear-off" technique, consisting in covering the surface with a strong cloth bound to the surface with generous layers of formaldehyde-treated glue. As luck would have it, this large fresco (3.50 × 7.80 m) was abandoned in a storehouse for more than 20 years. When the curators attempted to remove the cloth, much to their dismay they found that the glue resisted any attempt at digestion, even when treated with concoctions of the most aggressive proteolytic enzymes available on the market. It is likely that during the long storage the glue became slowly cross-linked by the formaldehyde to the point of forming an intricate mass of untreatable proteinaceous material. Thus, although poor Spinello died presumably as a *bona fide* Christian, his painting was condemned to wear the burka, Muslim-fashion, for the rest of its life. When we recently treated the fresco with a suspension of viable *Pseudomonas stutzeri* cells, these bacteria, although agnostic, were able to fully digest the hardened glue and restore to life Spinello's glorious painting. We show here how proteomics helped us solve the riddle of how these bacteria acted on the burka-obscured fresco.

Keywords:
Art biorestoration / Bacteria / Collagenase / Cultural heritage / Pseudomonas stutzeri / Zymograms

1 Introduction

Until World War II, one of the most frequently visited monuments in Pisa was the Campo Santo (Monumental Cemetery) in Piazza dei Miracoli, a huge, rectangular structure, whose portico-shielded peripheral walls were almost entirely covered by 1500 m^2 of frescoes. These frescoes dated back to the fourteenth century and were executed by famous painters such as Antonio Bonaiuti, Antonio Veneziano, Benozzo Gozzoli, Taddeo Gaddi, Francesco Traini, Bonamico Buffalmacco and Spinello Aretino. Spinello was commissioned to paint eight large frescoes (each of slightly different size, on average 3.50 × 7.80 m) on the south wall, illustrating some episodes of the life of the saint martyrs Efisio and Potito, whose mortal remains were buried in the cathedral. Spinello started his work in September 1391 and terminated it in March of the following year, to the satisfaction of the commissioners, who paid him 50 florints *per* panel, a large sum at that time. The hordes of visitors dwindled to a trickle after World War II, partly because, during an air raid by the

Correspondence: Professor Pier Giorgio Righetti, Department of Agricultural and Industrial Biotechnologies, Strada le Grazie 15, I-37134 Verona, Italy
E-mail: righetti@sci.univr.it
Fax: +39-045-8027929

Abbreviations: AA, acrylamide; **APS**, ammonium persulphate, **CBBG**, Coomassie Brilliant Blue G; **LDH**, lactate dehydrogenase; **TBP**, tributylphosphine

allies in 1944, some bombs fell on the cemetery causing further damage to the frescoes that had been partly degraded over the centuries. Thus, starting at the end of the 1940s, the frescoes were removed from the walls using the "tear-off" technique, consisting in the application of a wide gauze onto the fresco surface, by using animal glue as a consolidating agent. Upon hardening of the glue, the covering gauze and fresco are fused into a compact layer, which is then detached from the wall. In order to liberate the fresco from the cloth obscuring it, its back is reinforced by a canvas cloth, whose adhesion is ensured by casein-based glue, in turn supported on asbestos panels, again coated with casein. At this point, in principle, when the front of the fresco is treated with proteolytic enzymes, the gauze obscuring it can be gently removed, thus restoring the painting to view.

The second of the four frescoes depicted by Spinello, with the title "*Conversione di S. Efisio e Battaglia*" (Conversion of S. Efisio and Battle), was removed 20 years ago and abandoned in a storeroom until recently. When the curators attempted to remove the cloth, much to their dismay they found that the glue resisted any attempt at digestion, even when treated with concoctions of the most aggressive proteolytic enzymes available on the market. At the time of the fresco removal, the glue had been admixed with generous doses of formaldehyde, added as an antimicrobial agent. Perhaps the curators had no idea that formaldehyde reacts readily with the α-amino groups of Lys in proteins and, in general, with all primary amino groups. Although this reagent is not, *per se*, a cross-linking agent like glutaraldehyde, since it contains only a single reacting group, it is known that after long storage times, secondary reactions do indeed lead to cross-linking [1] (irreversible when occurring on Tyr and Trp residues). It is likely that the layer of glue covering the fresco (ranging in thickness from 1 to 3 mm), after such a long storage time, had been extensively cross-linked to the point of forming an intricate mass of untreatable proteinaceous material. The situation appeared hopeless.

When our group, which had already gathered expertise on biorestoration intervention, *i.e.*, direct application of viable bacterial cells onto an artwork surface (especially stone and marble) [2–4], was contacted by the Opera Primaziale Pisana, a series of preliminary experiments were conducted to solve the riddle. Strains of *Pseudomonas* were selected, since these bacteria are non-pathogenic, not spore-forming and are known to react readily to environmental conditions by inducing and exporting extracellularly a broad range of enzymatic activities. Five such strains were tested: *P. cepacia, P. testosteroni, P. fluorescens, P. flavescens* and *P. stutzeri*. Since the last strain exhibited the best growth curves on ground chips of insoluble glue harvested from the clothed fresco, it was selected for the biorestoration procedure [4]. Although full restoration of the fresco was achieved about 1 year ago, we had the desired results but lacked specific information on the molecular mechanism of the process. With a proteomic approach, we offer here a plausible route to the mechanism of action of this novel biorestoration process.

2 Materials and methods

2.1 Materials

Tris, mineral oil, dl-DTT, Tween 20, β-casein, azo dye-impregnated collagen, azo-casein, TPCK-treated trypsin, collagenase (Clostridiopeptidase A) and EDTA were purchased from Sigma-Aldrich (Steinheim, Germany). Glycerol, methanol, ethanol, acetic acid, acetone, Triton X-100 detergent, NADH and pyruvic acid were obtained from Merck (Darmstadt, Germany). TEMED, 40% acrylamide/Bis solution, acrylamide (AA), ammonium persulphate (APS), the Protean IEF Cell, a GS-710 densitometer, a Versa Doc Scanner, the software PDQuest version 7.3 as well as linear Immobiline dry strips pH gradient 3–10 (17 cm) were obtained from Bio-Rad (Hercules, CA, USA). Glycine, SDS, iodoacetamide, urea, thiourea, tributylphosphine (TBP) and CHAPS were purchased from Fluka (Buchs, Switzerland). Bromophenol blue, carrier ampholytes and agarose were purchased from Amersham Bioscience (Uppsala, Sweden). *Pseudomonas stutzeri* A29 strain, animal glue and the aged animal glue derived from the cloth employed in the fresco's restoration were kindly provided by Distam, University of Milan, Milan, Italy.

2.2 Strains and growth conditions

Pseudomonas stutzeri A29 strain was inoculated at a starting OD_{600} of about 0.02 in M9 mineral medium, supplemented with one of three carbon sources employed in this study at a final concentration of 5 g/L. The three carbon substrates were glucose, animal glue and the aged animal glue derived from the cloth employed in the fresco's restoration. Incubation was performed at 30°C under agitation for 10 h. Carbon substrates were previously sterilized by filtration (glucose) or autoclaving (animal glue, 121°C for 15 min), while aged animal glue was ground and added to medium without previous sterilization. After incubation, cells were harvested, washed twice in PBS and stored at −20°C. Culture supernatant was also collected and immediately frozen at −20°C.

2.3 Enzymatic activity assay

Lactate dehydrogenase (LDH) activity in cultural supernatants was assayed in order to detect cell lysis. The reaction was carried out in a 1 mL cuvette at 20°C under the following conditions: 60 mM phosphate buffer pH 7.00, 0.2 mM NADH, 1 mM pyruvic acid. In total, 100 μL of supernatant was added. Immediately after pyruvate addition, absorbance at 340 nm was monitored by reading every 20 s for a total of 6 min. As positive control, we used culture supernatant of *Pseudomonas* cells lysed by sonication, while for blank reference we used the same sample but thermally treated in order to completely inactivate LDH.

General proteolytic activity in culture supernatant was assayed by a colorimetric test based on azo-casein (azo dye-impregnated casein). As positive control, trypsin aqueous solution (10 μg/mL) was used, and it was also used to create a four-point calibration line (0 for blank, 10, 25, 50, 100 μL of enzyme solution). A three-point curve (25, 50 and 100 μL) for each sample was also traced at different incubation times. The proteolytic reactions were carried out in a 1.5 mL test tube at 37°C in 1.2 mL of solution (100 mM phosphate buffer pH 8.00, 0.25% w/v azo-casein) by adding 100 μL of supernatant sample (opportunely diluted with Milli-Q water). After pre-established incubation times, volumes of 200 μL were drawn from each tube and the reaction was stopped by adding 6% TCA. After 10 min of incubation at 4°C, the residual intact protein was precipitated by a 10 min centrifugation at 20 000 × *g*, and 180 μL of supernatant was finally transferred into a microplate well. Sample colour was enhanced by rising the pH to an alkaline value with 500 mM NaOH, after which the absorbance at 415 nm was immediately measured.

Specific collagenolytic activity in culture supernatant was evaluated by an azo dye-impregnated collagen substrate (azocoll). Collagen degradation kinetics were followed by using only one concentration of sample supernatant and reference enzyme solution (*Clostridium histolyticum* pure collagenase, 1 mg/mL) and by varying the reaction time. Reactions were carried out in 15 mL sterile tubes in a total volume of 3 mL of solution (0.5% w/v of solid azocoll, 100 mM phosphate buffer pH 8.0 and 1 mM $CaCl_2$) by adding 0.2 mL of sample. As reference, 10 μL of pure collagenase solution (1 mg/mL) was used, while 0.2 mL of water was added for the blank. The proteolytic reaction was carried out at 37°C under fast agitation. After pre-established times of incubation, volumes of 700 μL were drawn from each tube and the reaction was stopped by adding 6% w/v TCA. Each sample was immediately centrifuged at 20 000 × *g* for 5 min; 500 μL of the supernatant was then transferred into a 1 mL cuvette and diluted 1:2 with water. Absorbances at 520 nm were read, and the values plotted against incubation time.

2.4 Zymograms

Casein zymograms were used in order to assess how many, and what kind of, proteolytic components could be present in cultural supernatants. Concentrated samples were prepared for SDS-PAGE separation in Laemmli Buffer 2X (Bio-Rad) without reduction and boiling. After an incubation at 20°C for 20 min, samples were loaded onto a 13 cm long 10%T polyacrylamide gel co-polymerized with 0.1% β-casein. Electrophoresis was performed at 4°C in Tris-glycine buffer pH 8.3 by applying 60 V for 1 h followed by 130 V until the end of the run. The gel was then incubated overnight in 2% v/v Triton X-100 in 100 mM phosphate buffer pH 8.00, under gentle agitation, in order to remove SDS and allow for protein refolding and activity restoration. Finally, the gel was stained with colloidal CBBG and destained with 7% acetic acid aqueous solution.

2.5 Two-dimensional gel electrophoresis

The supernatants of *P. stutzeri* grown on glucose, Spinello's glue or fresh glue were desalted, lyophilised and dissolved in a denaturing solution containing 7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP and 0.5% carrier ampholytes pH 3–10. After centrifugation at 4°C at 4000 × *g* to eliminate any residual particulate material, the samples were alkylated for 1 h with 10 mM acrylamide (1 M stock solution) and the reaction was stopped with 10 mM DTT (from 1 M stock solution) [5].

The first-dimension run was performed on strips (17 cm length, 0.5 mm thickness) with a linear pH gradient from 3 to 10 [6]. IEF was carried out with a Protean IEF Cell (Bio-Rad) with a low initial voltage and then by applying a voltage gradient up to 10 000 V with a limiting current of 50 μA/strip. The total product time multiplied by the voltage applied was 75 000 Vh for each strip and the temperature was set at 20°C.

Each strip was then equilibrated with an SDS denaturing solution containing 6 M urea, 2% SDS, 20% glycerol and 0.375 M Tris-HCl (pH 8.8). The contact lasted for 26 min in tubes containing 10 mL each of the equilibration solution. Each strip was then interfaced with a gel slab using 0.5% agarose solubilised in the cathodic buffer (192 mM glycine, 0.1% SDS, Tris to pH 8.3). All the gel slabs were cast with a two-vessel gradient mixer, with total sufficient volumes for polymerising gel slabs of 1.5 mm in thickness and with a porosity gradient from 7%T to 20%T. Polymerisation took place overnight. The second-dimension run was performed by using a Protean II xl Multi-Cell (Bio-Rad). The cathodic buffer was the same as in the previous step; the anodic buffer was a solution of 0.375 M Tris-HCl, pH 8.8. The electrophoretic run was performed by setting a current of 2 mA for each gel for 1 h, then 5 mA/gel for 2 h and 10 mA/gel until the end of the run. During the whole run the temperature was set at 11°C. All gels were then fixed with a solution containing 40% ethanol and 10% acetic acid for 30 min. After this step the gels were stained with SYPRO Ruby (70 mL for each gel) overnight and then destained with 10% methanol, 7% acetic acid for 1 h. The gels were finally washed with Milli-Q water. All the samples were then digitised with the Versa Doc Scanner. Spots of interest were then cut out from the gels, digested with trypsin and subjected to MALDI-TOF MS analysis. Measurements were performed using a Tof-Spec 2E MALDI-TOF instrument (Micromass, Manchester, UK) operated in reflectron mode. The final mass spectra were produced by averaging 50–200 laser shots. Peptide masses were searched against Swiss-Prot and TrEMBL databases, by using the search engines ProFound, ProteinLynx and MASCOT.

3 Results

3.1 Growth curves and integrity of *P. stutzeri* cells

For the *in situ* biorestoration process, the fresco was entirely covered with a thin layer of hydrophilic cotton wool impregnated with the bacterial cell suspension at a high cell density (8.5 log CFU mL^{-1}). Contact times of 12–14 h at a temperature of 30°C ensured almost full digestion (100% in the thinner areas, 80–90% in the thicker, up to 3 mm, layers) of the glue, to the point that detachment of the fresco surface from the gauze obscuring it was readily achieved. For the present laboratory study, we had to select optimal growth times of controls and cells grown on tiny fragments of original formaldehyde cross-linked glue, so as to ensure that all samples would be harvested for analysis at the late exponential growth phase, in order to avoid differences due to potential lysis after reaching a growth plateau [7]. Figure 1 shows the growth curves of *P. stutzeri* in broths containing glucose or a suspension of shredded Spinello's glue. As expected, the onset of growth occurs at earlier times in the glucose-grown cell cultures, but after a lag time, the cells in glue debris as the sole energy source also appear to grow at a similar rate and reach a confluence plateau at about 10 h. Thus, this time was selected for all cultures for harvesting. Based on the plausible assumption that bacteria acted on the layer of glue by exuding proteases in the culture broth, one should also ascertain that cell integrity was maintained during *P. stutzeri* action on the layer of glue, and that no accidental lysis occurred. For this, we performed the LDH test shown in Fig. 2. LDH is known to be a cell sap enzyme that should not be found in the culture supernatant. In Fig. 2 the culture broths of intact cells and of one test sample disrupted by sonication are assayed for LDH activity. This activity is readily observed on the supernatant of the lysed sample, but seems to be completely lacking on all other cell cultures.

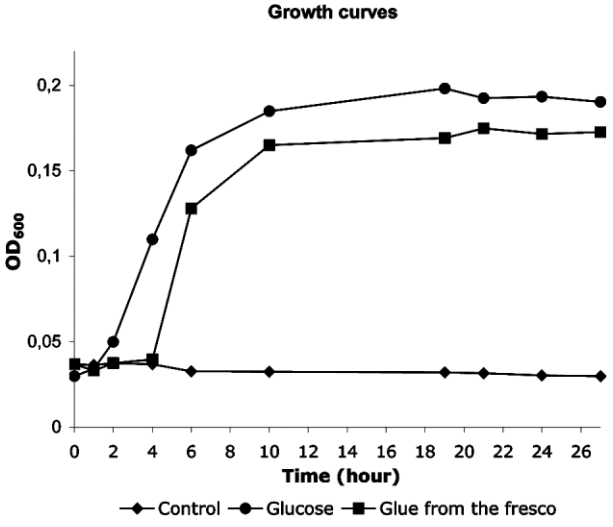


Figure 1. Growth curves pf *P. stutzeri* grown on glucose (circles) and on Spinello's glue (squares).

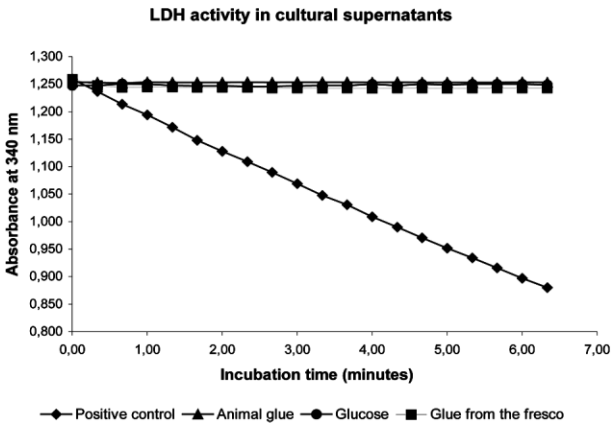


Figure 2. LDH activity test for cell integrity. The supernatant exhibiting strong LDH activity (diamonds) was a positive control of cells broken by sonication.

3.2 Attempts at spot identification *via* 2-D maps of supernatants

In a number of experiments, the supernatants of controls and of cells grown on aged glue were subjected to 2-D map analysis *via* pH 3–10 immobilized pH gradients in the first dimension, followed by SDS-PAGE. The maps developed with SYPRO Ruby revealed *ca.* 20 spots of polypeptide chains in the supernatants of Spinello's glue, which increased by about 1 order of magnitude when the sample was concentrated ten-fold (Fig. 3A).

The supernatant of the *P. stutzeri* culture grown on glucose gave maps with a similar total amount of spots but different 2-D patterns (Fig. 3B). Quantitative and qualitative differences between the electrophoretic patterns of the two samples were investigated by *in silico* analysis with PDQuest software, which revealed a substantial number of proteins exclusively or preferentially secreted by the bacteria growing on aged glue in respect to those grown on glucose. Of these, the more abundant ones were chosen for identification *via* PMF, as described in Section 2. However, elution of major spots, trypsin digestion and MS analysis in the MALDI-TOF mode did not lead to any spot recognition, partly because several spots contained too little material, and partly because even in the presence of good spectra, we could not find corresponding proteins in the Swiss-Prot and TrEMBL databases (Swiss-Prot lists only 59 proteins, among which there was not a single protease for *P. stutzeri*). 2-D map analysis was, therefore, no longer pursued and we thus resorted to classic biochemical studies.

3.3 Tests for proteolytic activities on the bacterial supernatants

A spectrophotometric assay for caseinolytic activity was set up, by exploiting azo-casein as a substrate. Figure 4 shows the proteolytic activity, as measured at a fixed time of incubation (15 min), as a function of increasing amounts of standard trypsin solution or of supernatants of *P. stutzeri*.

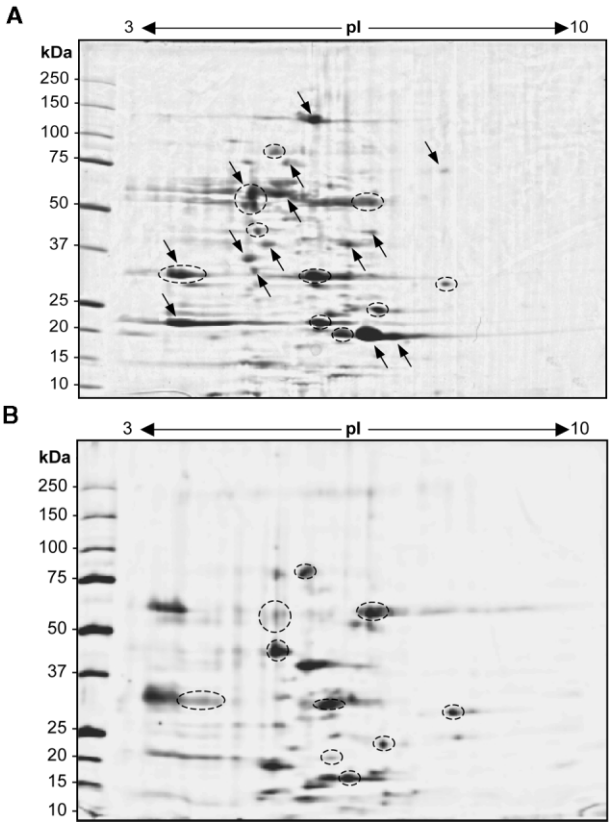


Figure 3. 2-D maps of supernatants of *P. stutzeri* grown on debris of glue from Spinello's fresco (A) and on glucose as sole carbon source (B). Circled regions represent some spots present in both samples and provide a spatial reference for image comparison. Arrows indicate the spots of the aged-glue sample chosen for identification *via* PMF. The first dimension was a pH 3–10 IPG strip; the second dimension a 8–20%T SDS-PAGE gel. Staining was with SYPRO Ruby.

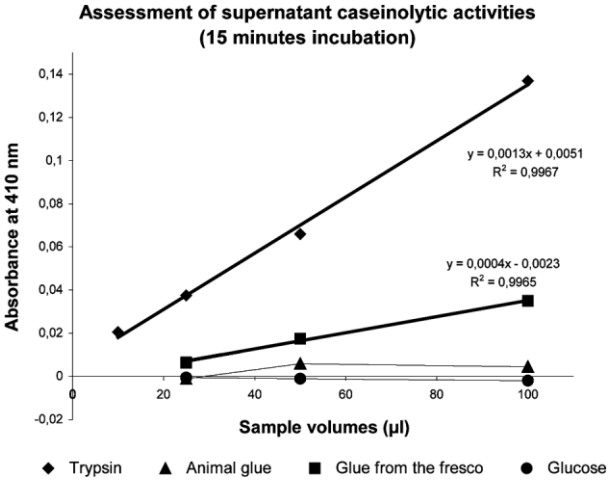


Figure 4. Caseinolytic activity in supernatants as measured *via* the rate of release of coloured peptides during proteolysis of an azo-casein substrate. For the control trypsin and the glue from the fresco, both the actual curves and their linear regression lines are plotted.

It can be seen that such enzyme activity in the different supernatants is exhibited only by the bacteria grown in the original glue from the frescoes. Its specific enzyme activity, compared to that of the standard trypsin solution, could be calculated as 3.63 10^{-3} mg/(µL*h).

In order to assess a more specific proteolytic activity against some glue proteins, another assay was employed. Figure 5A shows the enzyme activity profiles, on azo-collagen as substratum, of a purified collagenase from *C. histolyticum* (upper curve), of a *P. stutzeri* (grown on original glue from frescoes) supernatant (middle curve), as well as of other supernatants of bacteria grown on fresh glue or on glucose (bottom curves). It can be appreciated that, whereas the cells grown on original test glue show activities almost approaching those of the purified collagenase, all other supernatants exhibit essentially no activity, except for some faint activity in those supernatants originating from fresh glue, although the assay was extended up to 30 h. Figure 5B shows the same data from which, by calculating the slopes from linear

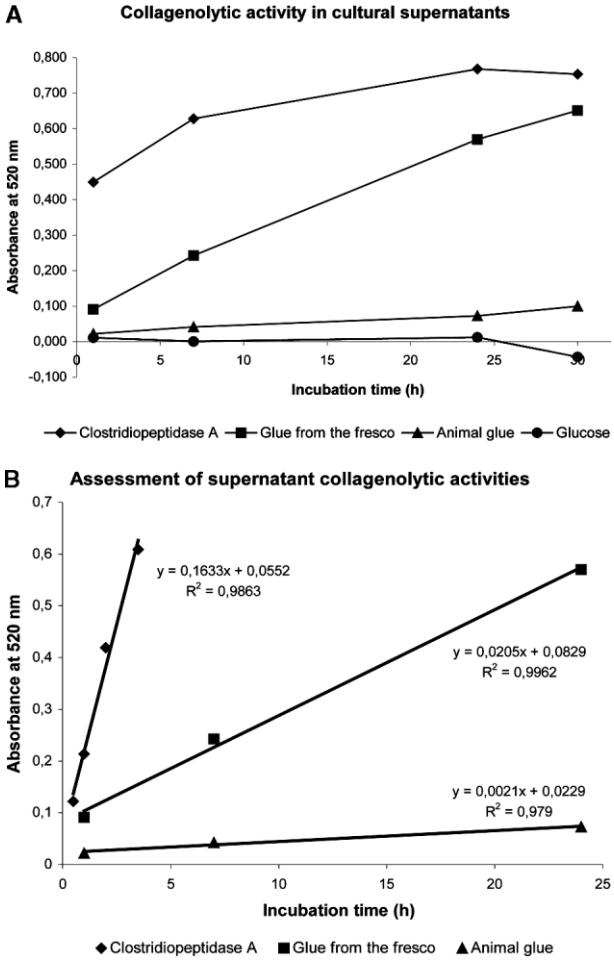


Figure 5. (A) Collagenolytic activity in supernatants as measured *via* the rate of release of coloured peptides during proteolysis of an azo-collagen substrate. (B) Quantitation of the collagenolytic activity by comparing the slopes of the linear regression lines from (A).

regression analysis, it was possible to obtain specific enzyme units (in milligrams of substrate converted *per* microlitre of supernatant *per* hour) for the various supernatants, compared to that of the purified collagenase from *C. histolyticum*. These activities are 2.31×10^{-3} mg/(μL*h) for the supernatant of cells grown on glue from the fresco and 5.91×10^{-6} for supernatants grown on fresh glue. It can thus be appreciated that the cross-linked glue used as growth substratum elicits enzyme activities that are on average some three orders of magnitude greater than those appearing on other supernatants.

Although these data fully confirm the presence of proteolytic activities (at least against collagen and casein) essentially only in the supernatants of intact *P. stutzeri* grown on the original glue from Spinello's fresco, they can hardly give us any clue for the identification of such enzyme(s) in these bacterial cells (no data exist on proteases in *P. stutzeri* strains). We thus resorted to a peculiar type of SDS-PAGE, by which the various supernatants were analysed in a 10%T, 2.5%C polyacrylamide gel, cast in the presence of 0.1% casein as a substrate. The supernatants were concentrated 300-fold and denatured in SDS under mild conditions (no boiling, absence of disulphide bridges reducing agents). After the run, the polypeptide chains were renatured by overnight incubation in 2% Triton X-100 in pH 8.0 phosphate buffer, followed by staining and destaining in colloidal CBB. As shown in Fig. 6, only the supernatant grown on Spinello's glue gave a ladder of caseinolytic activity bands, extending from *ca.* 20 kDa to 250 kDa.

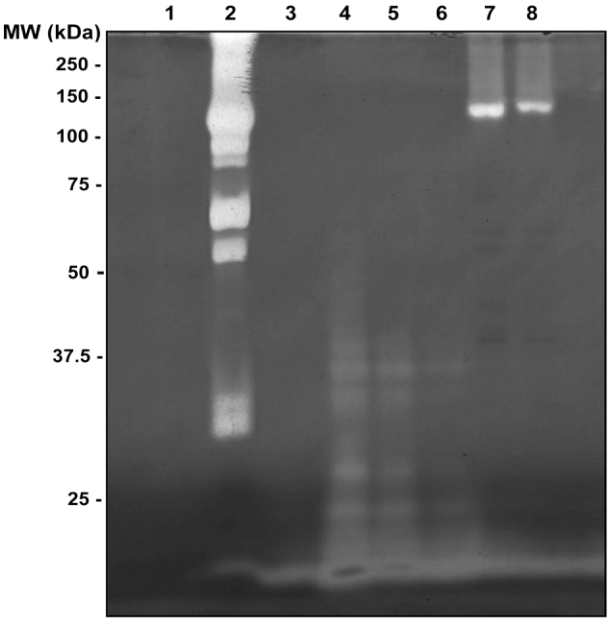


Figure 6. Casein zymograms *via* SDS-PAGE. The first three samples are supernatants (300× concentrated) of *P. stutzeri* grown on glucose (track 1), on Spinello's glue (2) and on fresh animal glue (3). Tracks 4–6 are positive controls with purified trypsin (10, 5 and 1 μg, respectively, in lines 4 to 6). Tracks 7 and 8 are purified collagenase from *C. histolyticum* (20 μg in 7, 10 μg in 8).

Interestingly, a commercial preparation of collagenase from *C. histolyticum* (tracks 7 and 8) gave an intense band at 120 kDa, whose apparent M_r coincides with the strongest activity bands in the supernatant from Spinello's glue (track 2). This suggests that the *P. stutzeri* collagenase could have a strong homology with *C. histolyticum*. At present, we do not know if the other major activity bands of lower M_r values, extending from 100 kDa down to as low as 20 kDa, represent novel types of proteolytic enzymes or if they could be degradation products from the carboxy-terminus of the 120 kDa parental species, as reported for other types of collagenases from diverse bacterial strains [8]. We believe, however, that at least some of these bands should represent unique enzyme activities, each endowed with specific proteolytic performance and that only their combined action is able to attack to some extent the hardened glue, until its complete digestion. In fact, when originally treated with purified enzyme preparations (including the same 120 kDa *C. histolyticum* collagenase), the 20-year-old glue could not be digested at all.

The results of the *P. stutzeri* treatment can be appreciated in Fig. 7. Panel A shows the final steps in overlaying the shrouded fresco with cotton strips impregnated with live bacteria; panel B shows a detail of the recovered fresco (central part of the "*battaglia*" with the angel leading the fight).

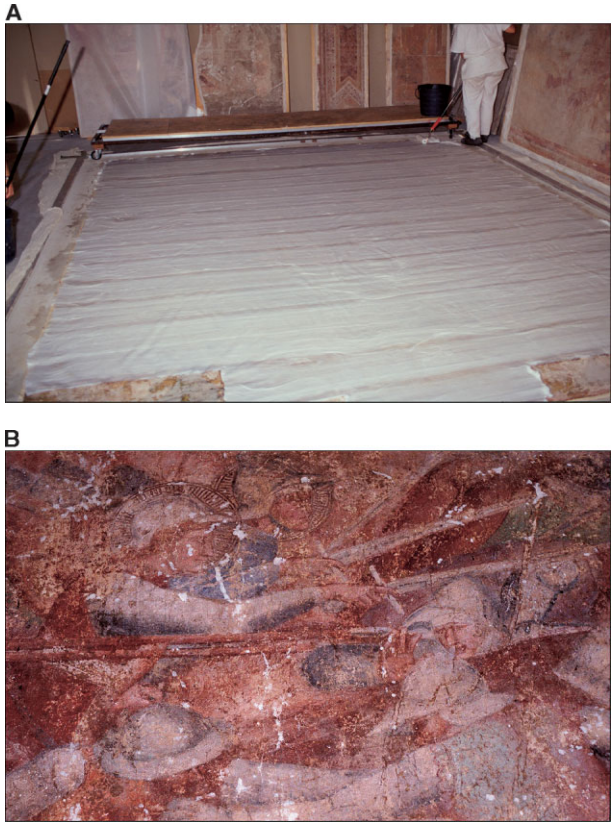


Figure 7. Spinello Aretino fresco during the *P. stutzeri* treatment (A) and after removal of the shroud (B; detail of the "*battaglia*").

4 Discussion

Biorestoration of cultural heritage *via* the use of "friendly" bacteria is gaining wide acceptance at present. Outdoor art work, especially where lithoid materials, marble, frescoes and paint are involved, is highly susceptible to deterioration brought about by ageing and, in recent decades, by pollution. For instance, in urban areas, pollution-engendered damage on monument surfaces is manifested by black crusts, nitration, sulphation and deposition of dust and residual hydrocarbons. In addition to air pollutants, the surface of man-made artistic work can also be altered by organic matter that has been applied, but then not promptly removed, during restoration. This was the case of the present report, since, in an unfortunate situation, the gauze covering Spinello's art work detached from the walls of Pisa's Monumental Cemetery could no longer be removed from the face of the fresco, due to hardening and cross-linking of the animal glue, admixed with formaldehyde, during the 20 years of storage. No single enzyme or concoctions of the most powerful proteases available on the market could act on the layer of glue, and the fresco seemed to be irremediably lost. It is remarkable that the *P. stutzeri* strain, applied at high density to the fresco surface, could extensively digest this intricate mass of untreatable proteinaceous glue in only 10–12 h, to the point that the gauze clothing the fresco could be easily removed. In the process of analysing the molecular mechanism of this action, we have detected, only in the supernatant of *P. stutzeri* grown on Spinello's glue, a strong band of collagenase activity at 120 kDa, possessing also a marked caseinolytic activity, that should be one of the principal enzymes responsible for the glue digestion (the most abundant components of animal glues are collagen and casein). Whether the spectrum of lower M_r species detected in Fig. 6 could be novel proteases or *C*-terminus cleavage products of the parental 120 kDa polypeptide remains to be seen. We tend, however, to favour the hypothesis that the spectrum of activities in Fig. 6 (track 2) represent a panel of different proteases, each endowed with some peculiar and unique activity, so that their combined and cooperative action is able to fully digest the insoluble glue. This is corroborated by the fact that, when first treated with mixtures of different, commercially available enzymes (including the 120 kDa collagenase from *C. histolyticum*), the hardened glue binding the cloth to the fresco could not be digested at all.

In conclusion, even today, when touring the Southwest of the USA, one can stumble upon tents of itinerant priests boasting the following: "daily religious services; miracles only on Sunday". It is remarkable that our humble bugs, quietly and without much ado, perform their miracles on a daily basis!

Supported by the MIUR (FIRB 2001, Cod. RBNE01KJHT; PRIN, 40%, 2003), Rome and by Cariverona (grant 2004).

5 References

[1] Marsch, J., *Advanced Organic Chemistry* (2nd Ed.). McGraw-Hill Kogakusha, Tokyo 1977, pp. 502–503, 810–811.

[2] Ranalli, G., Matteini, M., Tosini, I., Zanardini, E., Sorlini, C., in: Ciferri, O., Tiano, P., Mastromei, G. (Eds.), *Of Microbes and Art. The Role of Microbial Communities in the Degradation and Protection of Cultural heritage*. Kluwer/Plenum, New York 2000, pp. 231–245.

[3] Ranalli, G., Belli, C., Baracchini, C., Caponi, G. *et al.*, in: Saiz-Jimenez, C. (Ed.), *Molecular Biology and Cultural Heritage*. Balkema, Lisse 2003, pp. 243–246.

[4] Ranalli, G., Alfano, G., Belli, C., Lustrato, G. *et al., J. Appl. Microbiol.* 2005, *98*, 73–83.

[5] Herbert, B., Galvani, M., Hamdan, M., Olivieri, E. *et al., Electrophoresis* 2001, *22*, 2046–2057.

[6] Righetti, P. G., *Immobilized pH Gradients: Theory and Methodology*, Elsevier, Amsterdam 1990, pp. 181–254.

[7] Pessione, E., Giuffrida, M. G., Prunotto, L., Barello *et al., Proteomics* 2003, *3*, 1070–1076.

[8] Matsushita, O., Jung, C. M., Katayama, S., Minami *et al., J. Bacteriol.* 1998, *181*, 923–933.

Frederic Fortis[1]
Luc Guerrier[1]
Pier Giorgio Righetti[2]
Paolo Antonioli[2]
Egisto Boschetti[1]

[1]Ciphergen Biosystems Inc., Fremont, CA, USA
[2]Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Milan, Italy

Received November 18, 2005
Revised January 5, 2006
Accepted January 26, 2006

Research Article

A new approach for the removal of protein impurities from purified biologicals using combinatorial solid-phase ligand libraries

The removal of last impurity traces from a purified protein is generally called polishing. It is an important step in downstream processing since protein impurities may generate undesirable side effects when the preparation is intended for research, diagnostic and more importantly therapeutic applications. Polishing is generally achieved by using orthogonal separation methods to previous steps, the most common being gel permeation chromatography. In spite of its polishing effectiveness, this technique suffers from a poor separation capacity and modest productivity as a result of low speed. Other approaches, for instance, based on anion exchange or on hydrophobic chromatography, that may be optimized for a given process cannot be used as generic methods. This document reports for the first time the use of a combinatorial solid-phase peptide library as a general method for the removal of impurity traces. Several examples of impurity trace removal are reported; starting material is either a pure protein spiked with serum proteins or with *Escherichia coli* extracts or current purified proteins still containing a small percentage of impurities. Among polished proteins are recombinant human albumin expressed in *Pichia pastoris* and human transferrin purified from whole plasma. This new method is used in neutral or even physiological pH and ionic strength conditions, with a remarkable capability to remove impurities. The process is as rapid as current adsorption chromatography procedures usable for the removal of a large number of protein impurities, with each one present in small amounts, such as host cell proteins.

Keywords: Ligand library / Peptide ligands / Polishing / Protein purification / Proteomics
DOI 10.1002/elps.200500847

1 Introduction

One of the major problems associated with the production of rDNA proteins for human consumption is the extraction and purification from very crude feed stocks. In principle, such recombinant proteins should be purified to homogeneity, so as to avoid undesired side effects when they are injected with traces of protein impurities that can be immunogenic, especially if derived from host cell proteins (HCP) of non-human origin. In practice, even under the best conditions, it is very laborious to achieve purity levels better than 99% [1, 2]. The last purification step for removing remaining traces of impurity is generally prohibitively expensive and leads to significant loss of valuable biopharmaceutical product.

Therefore, a protein purification process design for biopharmaceuticals starting from crude feed stocks is relatively complicated; it involves several unitary separation steps and several important parameters must be considered, such as the relative amount of the target protein in the crude extract, its physicochemical properties and the desired final purity goal. The analysis of these parameters is at the basis of the selection of chromatographic resins having orthogonal properties in

Correspondence: Dr. Egisto Boschetti, Ciphergen Biosystems Inc., Fremont, CA 95555, USA
E-mail: eboschetti@ciphergen.com
Fax: +33-134207878

Abbreviations: CM10, carboxymethyl weak cation exchange biochip; **GPC**, gel permeation chromatography; **HCP**, host cell proteins; **H50**, hydrophobic surface biochip; **IMAC30**, immobilized metal ion chelating biochip; **Q10**, strong anion exchange biochip; **SELDI**, surface-enhanced laser desorption/ionization

order to increase the separation efficiency from one step to another. The first chromatographic separation step in the process is based on the selectivity of the resin for the target protein. In this way it is possible to concentrate the protein of interest and to get rid of the maximum amount of protein impurities. This step is generally called "capture" [3]. If the capturing resin is based on a biological affinity ligand, the resulting purification factor in a single step could be very substantial. This is the case, for instance, in the capture of antibodies when using a Protein A column [4] where the purity of the target protein can go up to 95–98%.

The pre-purified protein from the capture phase is then submitted to an intermediate separation process capable of producing fractions where the protein of interest is collected in one or two of them [5–8]. Resins used for this purpose are regular or high-capacity ion exchange chromatography (IEC) as well as hydrophobic interaction chromatography (HIC) resins. In spite of a stringent selection of chromatographic conditions for the first two steps, the final purity of the target proteins from the second step is never perfect; it contains traces of a number of protein impurities coming from the expression system or from the biological crude extract. In order to reach purity levels capable of meeting regulatory requirements it is necessary to add at least another step called "polishing" to eliminate all remaining impurity traces. For accomplishing this task the ideal chromatographic resin should be capable of adsorbing all impurities and leaving the target protein in the flowthrough. This configuration would allow the size of the column to be reduced since impurities are present in small amount compared to the target protein. Such a definition corresponds to an affinity column; unfortunately, protein traces to be removed are constituted of a large number of species and bioaffinity adsorbents corresponding to classic definitions are unpractical. Considering the above context, the most popular approach for polishing is designed around gel permeation supports based on molecular size discrimination [9–11]. Most of the time, gel permeation chromatography (GPC) appears as the best choice because it is orthogonal towards the previous separation steps. However, it suffers from a number of productivity issues that limit the applicability at a preparative large scale. In a large number of cases, GPC proved to have good performance in separating impurities essentially because they were of different molecular size. Nevertheless, mass similarities between the target protein and impurities create a situation that cannot be resolved by GPC.

Other methods have also been described based either on HIC [12, 13] or IEC [14, 15]; both methods are process-specific and must be used under the best optimized conditions for accommodating the adsorption of the largest number of impurities and the collection of the target protein in the flowthrough.

In an attempt to find a generic process for the removal of impurity traces, a new approach based on the use of a multitude of discrete combinatorial ligands (one ligand structure per bead) is reported in this paper. It demonstrates the ability to capture impurities yielding the target proteins significantly purer than before the treatment. The principle of the multiselective adsorption and several concrete examples are extensively reported.

LC is a well-established process for protein adsorption and separation [16]. Affinity adsorption is the most selective way for capturing single proteins from very crude dilute feed stocks adding concentration to a purification effect. Once the binding capacity of the solid phase is reached, the process is stopped, the adsorbed protein is recovered and a second cycle can be activated. If this simple mechanism is extended to a large number of different affinity solid-phase ligands mixed together, many proteins would be captured at a time by their own complementary ligand. Each unique affinity ligand bead within the pool would bind and concentrate its specific protein partner up to the point of ligand saturation.

If in such a mixed bed one loads a complex mixture composed of many proteins, one of them being largely dominant (*e.g.*, 99% purity), the latter will saturate its corresponding solid phase ligand very rapidly, while impurity traces will not. When using loading conditions where the impurities do not saturate the solid-phase ligands, the flowthrough of the column will be composed of the dominant protein of purity substantially higher than before the process. This is because all other protein traces are captured by their corresponding ligand bead.

Under optimized conditions, with a solid-phase pool containing a large diversity of ligands for virtually all proteins available, the process can be used for the removal of impurity traces at the end of downstream separation processes.

A similar principle has been recently described for decreasing the concentration difference of proteins from a complex biological fluid (*e.g.*, serum or urine) [17, 18]. The ligand library was produced by a combinatorial synthesis of hexapeptides comprising several millions of different peptide ligands.

In the present work a similar library is used under an inversed manner to take advantage of the selective adsorption of low-abundance proteins constituting the impurity pool.

2 Materials and methods

2.1 Chemicals and biologicals

The solid-phase combinatorial peptide library as well as ProteinChip® Arrays were supplied by Ciphergen Biosystems (Fremont, CA, USA). Biochemicals such as human proteins (pure human transferrin, recombinant human albumin), myoglobin, molecular mass standards, as well as chemicals such as urea, SDS, thiourea, CHAPS) and the fluorescent stain Sypro Ruby were all obtained from Sigma-Aldrich (St. Louis, MO, USA).

Ten-well pre-cast 1-mm-thick 16% Tricine polyacrylamide gel plates, Coomassie and silver staining reagents were supplied by Invitrogen (Carlsbad, CA, USA). Q HyperD 20 μm was purchased from BioSepra-Pall (Cergy-Pontoise, France).

2.2 General polishing method

A solution of the purified protein to be polished was prepared in PBS. A column of solid-phase combinatorial peptide library was packed separately, equilibrated with PBS and connected to a chromatographic setup comprising a pumping system and UV/pH detection unit for recording both events at the column outlet.

The column was loaded continuously with the protein solution to be polished at a linear flow rate of 50 cm/h. The flowthrough was collected under fractions of a few milliliters each to analyze the capability of the solid phase to remove protein impurities, likewise a frontal analysis applied for the determination of binding capacity of sorbents [19]. Once the load was terminated, a PBS solution was introduced for washing out excess of proteins. Finally, the captured proteins were desorbed from the beads by a solution of 9 M urea-citric acid, pH 3.5. The desorbed proteins were collected as a single fraction. This general procedure was applied to recombinant human albumin expressed in *Pichia pastoris* after purification by anion-exchange chromatography. In this case the protein concentration of the initial sample was 10 mg/mL and the volume bead library was 1 mL. The procedure was also applied to a commercial purified transferrin from human plasma; the sample volume was 10 mL at a concentration of 5 mg/mL.

2.3 Single-step polishing for myoglobin

Commercial myoglobin from sperm whale was first purified by anion-exchange chromatography to remove a number of impurities naturally present in this preparation. Pure myoglobin was then added either with human serum proteins (at a rate of 1.8 and 3.7 mg of protein per 100 mg of pure myoglobin) or with an *E. coli* protein whole extract at a rate of 5 mg of protein per 100 mg of pure myoglobin. The buffer used was 25 mM phosphate, pH 7.0.

Subsequently, 400 μL of either solution was mixed with 80 μL of peptide library and incubated under gentle shaking for 20 min at room temperature. The mixture was then centrifuged in a spin column to drain out and collect the liquid phase containing polished myoglobin. The collected polished myoglobin was then analyzed according to various methods described in the following sections.

2.4 SDS-PAGE

Electrophoresis of polished protein fractions compared to control and removed impurities was performed by using 10-well pre-cast 1-mm-thick 16% Tricine polyacrylamide gel plates. Samples of appropriate protein concentration were diluted twofold in sample buffer. Subsequently, 30 μL of diluted sample was loaded per lane and electrophoresis migration was performed at 150 V for 90 min. Staining and destaining were achieved using the method described by the supplier of reagents.

2.5 2-DE

Desalted myoglobin fractions were solubilized in the 2-DE sample buffer, containing 7 M urea – 2 M thiourea – 3% CHAPS – 0.5% Pharmalyte pH 3–10 and bromophenol blue. Volumes of 150 μL of the obtained protein solutions were then used to rehydrate 7 cm long IPG 3–10 strips for 4 h. IEF was carried out with a low initial voltage and then by applying a voltage gradient up to 5000 V, with a limiting current of 50 μA/strip. The total product time × voltage applied was 25 000 Vh for each strip, and the temperature was set at 20°C. For the second dimension, the IPG strips were equilibrated for 26 min in a solution of 6 M urea, 2% SDS, 20% glycerol, 375 mM Tris-HCl, pH 8.8, under gentle agitation. The IPG strips were then cemented with 0.5% agarose on a 10–20% gradient SDS-PAGE in the cathode buffer (192 mM glycine, 0.1% SDS and Tris to pH 8.3). The anodic buffer was a solution of 375 mM Tris-HCl, pH 8.8. The electrophoresis run was performed by setting a current of 5 mA for each gel slab for 1 h, then 10 mA/gel for 1 h and finally 20 mA/gel until the end of the run. During the entire run the temperature was set at 11°C. At the end of the run the gels were washed in a water solution containing 40% methanol and 7% acetic acid, after that they were stained overnight with the Sypro Ruby fluorescent stain. Two-hour destaining was finally performed with the same 40% methanol – 7% acetic acid solution.

2.6 Surface-enhanced laser desorption/ ionization (SELDI) MS analysis

This analysis, based on both protein interaction with a chemically modified surface and MS for the discrimination of masses captured by the surface, was performed according to a well-known method [20]. Briefly, each spot of a ProteinChip® Array (referred to as biochip throughout the text) was equilibrated twice with 5 μL of the indicated array-specific binding buffer for 5 min according to the manufacturer's instructions. Each spot was then loaded with 6 μL of the sample previously diluted sixfold in the array binding buffer. After an incubation period of 30 min with constant shaking, the sample was carefully removed and each spot was washed three times with 5 μL of the binding buffer for 5 min to eliminate non-adsorbed proteins, followed by a quick rinse with deionized water.

All surfaces were air-dried followed by two applications of a saturated solution of sinapinic acid in a mixture of 50% ACN in water containing 0.5% aqueous TFA (1 μL of this solution per application with air drying in between). The biochips were then analyzed by laser desorption TOF-MS using a linear reader (Ciphergen Biosystems) in positive-ion mode (SELDI-MS), with an ion acceleration potential of 20 kV and a detector voltage of 2.8 kV. The molecular mass range investigated by MS was from 1 kDa to 300 kDa. Time-lag focusing was optimized at either 5 kDa or 70 kDa for low and high mass ranges, respectively.

Processing of data included baseline subtraction and external calibration using a mixture of known peptide and protein calibrants. Peak detection (S/N >3) and peak clustering was performed automatically using ProteinChip® Software 3.2. Biochips used throughout this study were CM10 (cation exchange), Q10 (anion exchange), H50 (hydrophobic surface) and IMAC30 (metal ion chelating surface) loaded with Cu^{++} ions.

3 Results and discussion

Most of injectable therapeutic recombinant proteins are expressed either in eukaryotic systems such as CHO cells or simpler microorganisms like yeasts (*e.g.*, *Pichia pastoris)* and prokaryotes (*e.g. E. coli*). In purification processes from very crude fermentation broths, not only isolation of the target protein but also elimination of HCP impurities must be considered. When these latter are at a concentration below ~3% a particular polishing step is specifically designed. This step is not based on a well-defined procedure and does not involve a unique approach. Therefore, several strategies have been described whose efficiency depends on the nature of the impurities to be removed. When polishing is intended to remove DNA-related molecules and endotoxins, anion exchangers are most generally used [21, 22]. The success of this operation is contingent upon the net charge of the target protein; if the protein has an acidic p*I* then the separation process may be much less efficient in terms of overall yield. The protein of interest may in fact partially be co-adsorbed on the chromatographic media thus significantly reducing the recovered amount.

A widely used polishing process involves GPC [23, 24]. This separation method is based on the difference in molecular size between the target protein and the protein impurities. However, by definition, impurities to be removed are very diverse and a number of them may have a similar molecular size as the target protein and thus will not separate well or even not separate at all. In spite of the possibility to accommodate the method to the formulation buffer of interest, preparative scale GPC suffers from several serious limitations: (i) the reduced amount of sample processed (volume and concentration) for a given volume of chromatographic media, (ii) the very low speed of the process and (iii) the fact that the operation adds a significant dilution factor.

Other polishing methods have been described in the literature based on the use of hydrophobic sorbents [12, 13] or pseudo-affinity media [15]. In spite of their reported efficacy, these approaches remain of narrow specificity. They apply to a selected category of proteins and when the target protein has a different hydrophobic behavior compared to the impurities.

The use of the combinatorial ligand library described in this document proposes a more generic solution for polishing. Combinatorial peptide libraries used as ligands for protein interaction offer a very large diversity covering a variety of protein–ligand affinity situations. This can be in fact compared to a multitude of affinity columns, each of them dedicated to the capture of a single protein. Although it is not yet formally demonstrated that the collection of ligands covers completely the full spectrum of possible protein impurities present, examples reported in this paper and other experienced cases, not reported here, unambiguously showed that the overall purity of the treated sample is largely improved in a single step under physiological pH and/or ionic strength conditions.

As previously described [17], a peptide ligand library has the property of acting as an assembly of an extremely large number of small solid-phase affinity adsorbents for a large variety of proteins. When a sample of a purified protein still containing a certain number of impurities is put in contact with the library, each bead adsorbs its specific partner (if present). The target protein is also co-

adsorbed when in the presence of its partner ligand bead; however, due to the limited binding capacity of the beads for a single protein species, this results in a minimal reduction of the overall recovery. The limited binding capacity for a single protein means that the described process is very performant when a large number of protein impurities are present, but each one in a small amount, as is very frequently the case for HCP [25]. Conversely, if for a similar level of impurity rate, only one or two protein impurities are present, they may exceed the binding capacity of partner beads and thus would not be completely removed. Because the situation of protein polishing is different from case to case, it is necessary to proceed by a preliminary set of experiments to determine the ratio between the sample and the column volume.

Figure 1 represents SDS-PAGE analytical data of a case model where the protein of interest was myoglobin (17.8 kDa) and protein impurities added were either human serum proteins or a whole extract of *E. coli* (large diversity of molecular masses present).

Whatever the impurities present, the combinatorial ligand library was capable of removing or significantly reducing added foreign proteins in cases where the added human serum protein impurities represented, respectively, 3.7 and 1.8% of protein mass.

Elimination of 5% added *E. coli* proteins was also very efficient as demonstrated in Fig. 2. Before the addition of *E. coli* soluble protein, myoglobin showed a pretty high level of purity (panel a). Few isoforms were present and three main impurities traces were visible at, respectively, p*I* values of about 7.8–8.3, 7.1 and 5.8 and M_r values of about 35–40 kDa, 6–7 kDa and 12 kDa. When added with 5% of *E. coli* protein extract (panel b), a large number of spots of impurities were visible of various p*I*s and masses ranging from 10 to 200 kDa. After the polishing process, myoglobin appeared almost as pure as initially: as highlighted, only five minor spots were still present (panel c). Conversely, removed impurities shown in panel d are very numerous. Traces of myoglobin were concomitantly captured by the ligand library as expected (see arrow). Data clearly illustrate the capability of the described process to remove a large number of very diverse proteins in terms of p*I* and M_r.

Additional analysis of myoglobin polishing experiments has been performed by MS in a molecular mass window between 1 and 20 kDa where 2-DE is less performant especially below 10 kDa. This was done using selected chromatographic chelating surfaces with copper ions (see Fig. 3). Contaminated myoglobin (panel a) shows a large number of low-molecular-weight protein impurities that are then captured by the ligand library column as shown in panel c, thus yielding a polished myoglobin as shown in panel b (17.8 kDa, single arrow and double-charged, double arrow) and confirming and complementing the data from the 2-DE. Many of the removed protein impurities were identified as belonging to the whole *E. coli* extract by 2-DE (data not shown). This experiment demonstrates the capability of the combinatorial library material to remove efficiently species of high and low molecular mass.

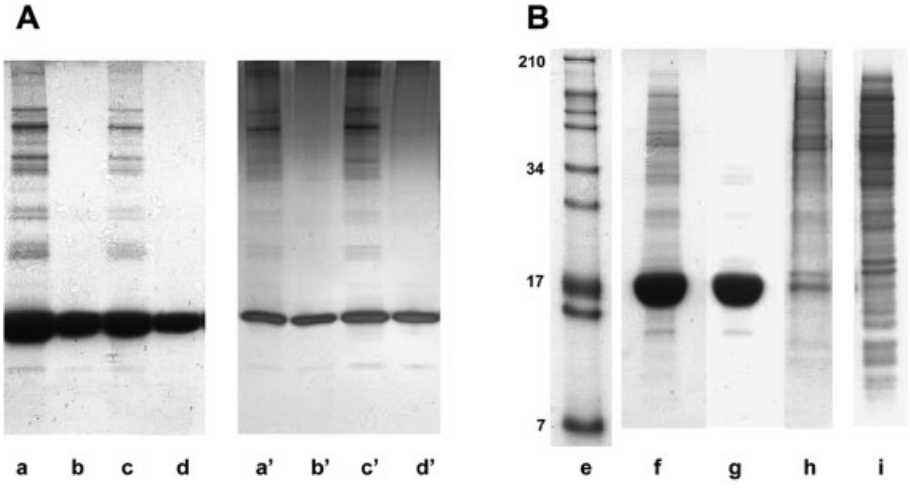


Figure 1. SDS-PAGE of myoglobin, polished by using a bead ligand library after contamination by, respectively, human serum proteins (A) or 5% of a soluble whole extract of *E. coli* proteins (B). (A) Lanes a and c represent the myoglobin contaminated with, respectively, 3.7% and 1.8% serum proteins; lanes b and d represent myoglobin after having been processed on the ligand library according the description in Section 2. a'–d' are the same samples but silver-stained. (B) e is a molecular mass ladder; lane f is myoglobin containing 5% of *E. coli* proteins; lane g is the polished myoglobin; lane h is the eluate showing removed contaminating proteins and some traces of myoglobin; lane i represents the whole *E. coli* extract.

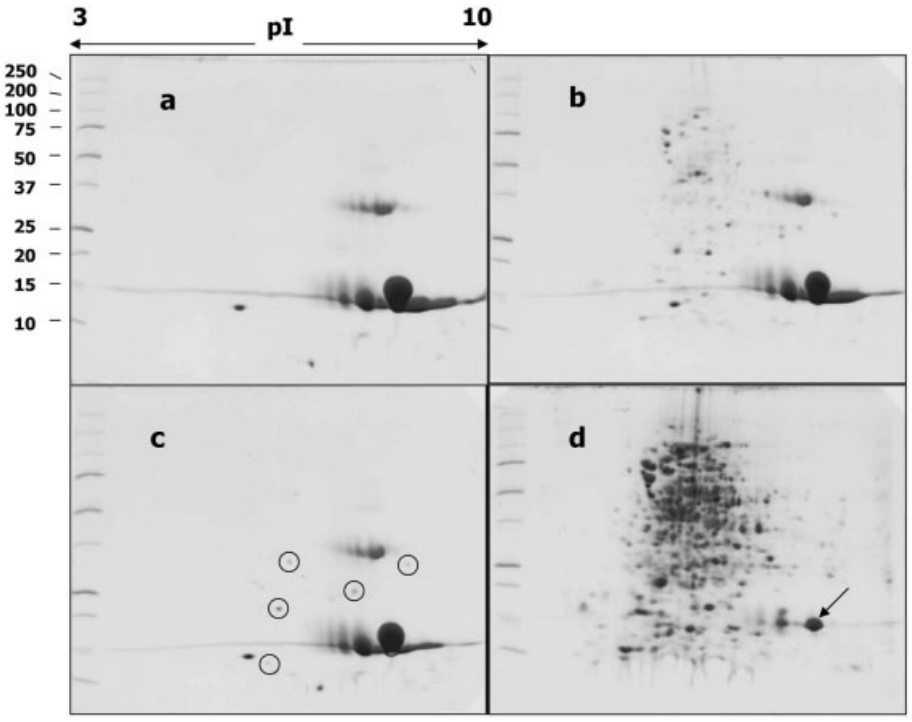


Figure 2. 2-DE of samples of myoglobin before and after polishing out *E. coli* protein impurities. p*I* range was between 3 and 10; M_r analyzed was between 250 and 10 kDa as indicated by the marker ladder. (a) Initial pure myoglobin; (b) myoglobin containing 5% of soluble *E. coli* proteins; (c) myoglobin after having polished out protein impurities (remaining *E. coli* protein traces are highlighted by circles); (d) captured protein impurities by the ligand library (proteins are mostly from *E. coli* extract; the myoglobin trace coadsorbed by the beads is highlighted by an arrow).

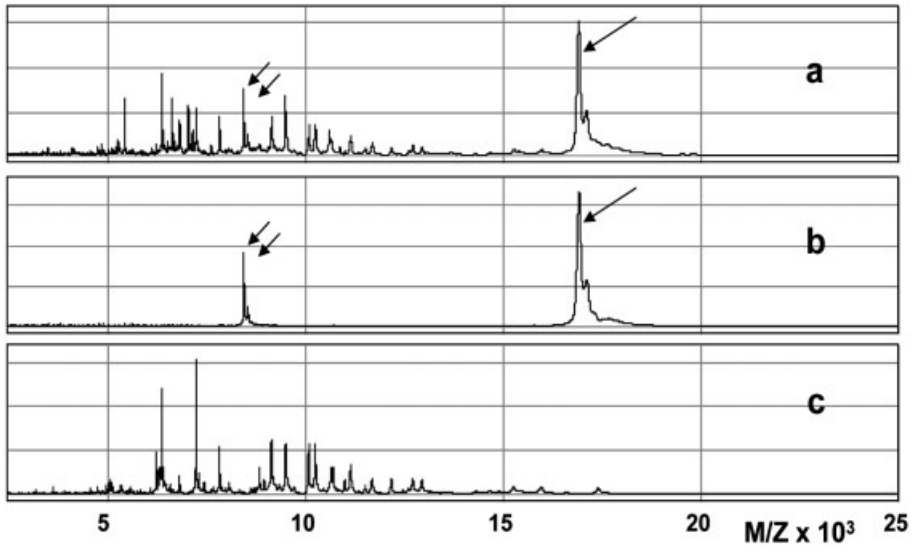


Figure 3. SELDI MS analysis on IMAC-Cu^{++} interacting surface of myoglobin fractions contaminated with 5% of soluble *E. coli* protein extract. (a) Contaminated myoglobin with 5% of foreign proteins; (b) polished myoglobin (single arrow); double arrow represents double-charged myoglobin. (c) Separated contaminants during the polishing process. All proteins represent only species that interact with copper ions (IMAC-Cu^{++}).

Both experiments with human serum proteins and bacterial proteins demonstrated the "universal" applicability of the described library in removing protein impurities from different species. The capture of *E. coli* proteins is by itself of high practical interest since many recombinant proteins are expressed in this microorganism (see, for example, [6, 7, 11, 14, 26, 27]); such impurities may potentially represent a danger for receiving patients when present even in trace amounts.

Although the polishing process showed an interesting capability to remove protein impurity traces, the principle itself is subjected to questions. The overall amount associated to the number of protein impurities to be eliminated could be critical to the success of the procedure, because each bead-ligand present in the library has a limited binding capability which is estimated as close to 1–3 ng of protein per bead. Calculated on the basis of milligram per milliliter of packed bed beads, this corresponds to an

average of about 15 mg of proteins per milliliter of packed bed when the diversity of protein impurities present is very large. On this basis a rational experiment design to remove a given amount of unknown impurities could be relatively complex, because it involves probability calculations as a consequence of both the ligand diversity and the number of protein impurities, along with the amount of each of them. In practice, optimization of the polishing process design has to be approached by a trial-and-error fashion in order to define the best capturing configuration in terms of bead-sample volume ratio. In this respect, frontal chromatography provides a good approach when facing complex mixtures.

The experiment shown in Fig. 4 represents the chromatographic profile of a recombinant human albumin expressed in *P. pastoris*; its purity estimated by electrophoresis was about 95%. The number of visible bands outside albumin (main band) was relatively limited: four main bands and three less concentrated ones of various masses were visible. This represents a difficult case of polishing because of the apparent overall large amount and small number of protein impurity to remove.

With this sample, a frontal analysis followed by an elution step of captured proteins was performed. The flow-through was separated into discrete fractions and then analyzed by SDS-PAGE in comparison to both the initial sample and the captured impurity species.

In the first part of the flowthrough (see Fig. 4, insert 1), recombinant albumin showed better purity compared to the fraction collected at the last part of the frontal analysis (insert 2). The analysis of the elution fraction from the combinatorial ligand column showed the presence of a large number of impurities that were not visible in the crude initial sample. A band positioned as albumin is clearly visible as one would expect (see insert 3). This experiment demonstrated that the ligand library column has the capability to adsorb impurities until the limit of capacity dictated by the available beads specific to a particular impurity. Although in the present case the calculation of impurity binding capacity was not measured, it appeared that the column volume for the removal of *P. pastoris* impurity starting from 0.5 g total proteins would have been 2 mL. Since SDS-PAGE analysis of the initial recombinant albumin showed the presence of a dominant band (albumin) and few other major bands, additional MS and immunochemical investigations were undertaken. MS analysis unambiguously demonstrated (Fig. 5) that the initial product was quite rich in impurities: they were dispersed over a wide range of masses from a few thousand dalton (Fig. 5A) up to about 120 kDa (Fig. 5B). After polishing, the signals showed that albumin was almost the only species present: this included a large amount of dimer and, as usual, a double-charged signal. Impurities, on the other hand, were found in the elution fraction. They were of diverse molecular masses from few thousand Da up to the albumin dimer mass. MS signal, however, cannot identify the nature of proteins; therefore, immunochemical-specific identifications of human albumin, on the one hand, and *P. pastoris* proteins, on the other hand, were undertaken. This was performed by Western blot after electrophoresis separation. Experimental results (data not shown) indicated that initial

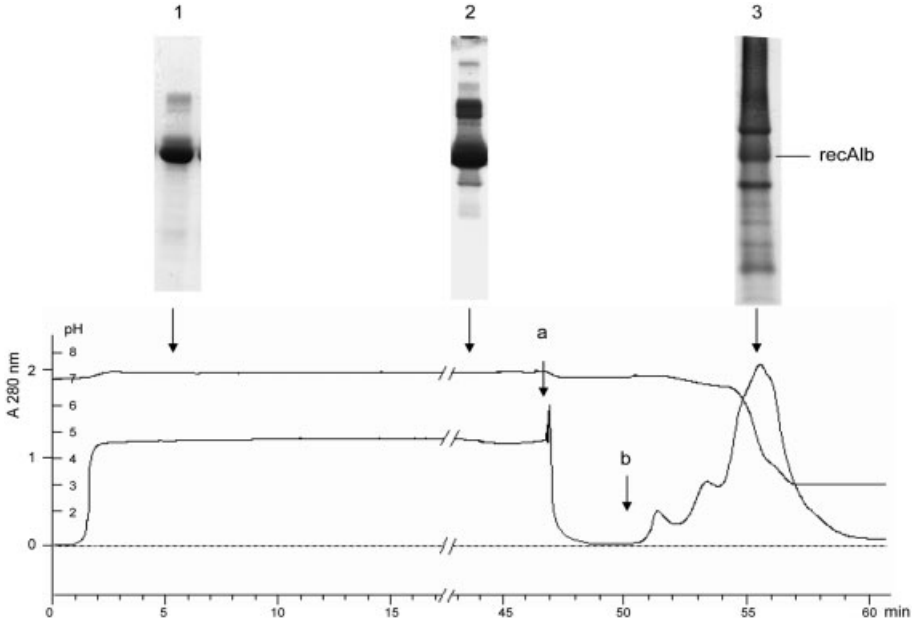


Figure 4. Chromatographic frontal analysis with purified recombinant albumin and elution of captured protein impurities. (1) SDS-PAGE analysis of fraction from the beginning of the chromatographic separation. (2) SDS-PAGE analysis of intermediate fraction, (3) analysis of elution fraction (captured protein impurities) after a washing step with PBS solution (a) and elution (b) by means of a solution of acidic urea. Albumin positioning is indicated (recAlb).

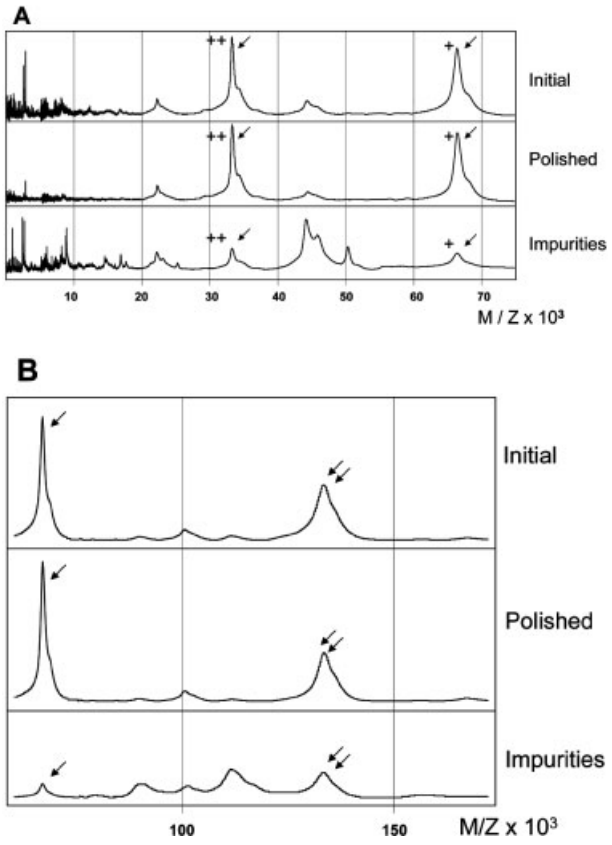


Figure 5. SELDI MS analysis of polished recombinant albumin from *Pichia pastoris*. Analysis performed by using a CM10 ProteinChip array under classical conditions (see Section 2). (A) Comparative analysis of molecular mass range up to 75 kDa. Albumin signal is indicated by the arrow; "+" indicates single charged albumin ions, "++" indicates double-charged albumin ions. (B) Comparative analysis of molecular masses between 60 and over 150 kDa. Single arrow indicates albumin ions and double arrows indicate the signal of albumin dimer.

material contained several bands including the major one responding positively to anti-albumin antibodies, suggesting that fragments were present. Moreover, when the blot was made using antibodies against total *P. pastoris* proteins, it was clearly seen that no impurities were left after polishing. Furthermore, all impurities collected from the elution following frontal analysis (Fig. 4) were *P. pastoris* proteins except one minor band of co-adsorbed albumin.

In another example, the reduction of protein impurities from serum transferrin under conditions described in Section 2, was also unambiguously shown (see Fig. 6). The initial purity of this protein was estimated close to 95%; after polishing under the described conditions the purity was significantly enhanced. The purity was, nevertheless, a function of the column load; when the load was limited to 50 mg of proteins per milliliter of ligand library beads, the purity obtained was significantly higher and probably close to 99%. When the load was increased to 1000 mg of transferrin per milliliter of ligand library beads, the impurities could not be substantially reduced. This relatively limited column loading compared to the myoglobin case doped with 5% of *E. coli* protein is explainable by the low number of impurities present in the estimated 5%. Figure 6 represents the analysis data of the situation, where impurities of large masses were clearly visible by SDS-PAGE and low masses were better shown by MS on IMAC-Cu^{++} ProteinChip array. Both data unambiguously showed a strong reduction of impurities that were found in the captured fraction.

Overall, the described method is applicable to a variety of situations rendering the polishing step of wide general applicability. The diversity makes the library rather universal for the adsorption of protein impurities independently from their composition and origin. However, the conditions of polishing have to be accurately adjusted in order to avoid saturation of the ligands by protein impurities that would render the process less effective. A point to consider, in fact, is that the binding capacity for each individual protein to be removed depends on the number of beads carrying the appropriate ligand, since one single protein interacts with a limited number of beads. The larger the number of different ligands, the lower is the number of individual beads carrying the same ligand in a given volume of beads. While the binding capacity is estimated close to 1–3 ng per bead, variations may be present from one bead to another due to individual kinetics and thermodynamic situations dependent, among other things, on the individual structure and the size of proteins to be removed. As a general rule, the binding capacity for protein impurities has to be determined case-by-case by, for instance, chromatography frontal analysis as described in the Section 2.

Comparing the described polishing technology with GPC, one can argue about the capability of the latter to formulate the protein solution in the final storage buffer. This feature is also possible with the use of combinatorial ligand beads described in this paper. The composition of the buffer for polishing can in fact be selected in order to meet formulation requirements in terms of pH, ionic strength and additives, providing these latter do not compete with the impurities to be eliminated.

For preparative repetitive polishing applications, a few additional considerations must be mentioned. The first is to ascertain that the binding capacity of the ligand library does not decrease over runs. To this end it is necessary to completely remove adsorbed proteins using appropriate desorption agents as described in other articles dealing

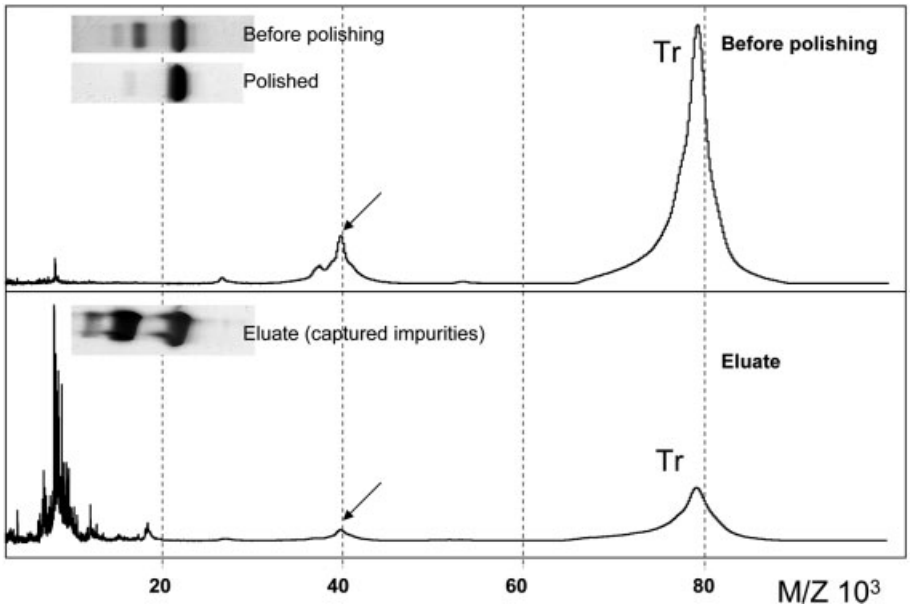


Figure 6. SELDI MS and SDS-PAGE analysis (inserts) from polishing experiment with human transferrin. Trace before polishing shows transferrin (Tr) as a major component and relatively few impurities. The arrow indicates a peak of double-charged transferrin. Eluted fraction of polished out species is illustrated at the bottom. Most of the impurities showed a mass below 20 kDa. Transferrin (Tr) was also slightly co-adsorbed, as expected, as shown by peaks at 80 kDa and 40 kDa (double charge; arrow).

with proteome analysis [17] (Righetti, P. G., Boschetti, E., Lomas, L., Citterio, A., submitted for publication). A second question is related to ligand leakage that would possibly contaminate the polished protein. This latter is always collected, as described, in the flowthrough with no change in physicochemical conditions of the solid phase; therefore, no major reason for ligand release can be invoked. In fact, ligand leakage generally occurs at the desorption stage when pH and/or ionic strength are modified, thus weakening the whole construct [28]. Proteases, such as exopeptidases, are ineffective for reasons related to the structure of the peptide library, as already reported [17, 29]. Nonetheless, in case of ligand leakage the elimination of released peptides is a particularly easy task because of the very low molecular mass (a few hundred dalton) compared to the size of the polished protein. Investigations on these considerations are in progress.

4 Concluding remarks

The approach of removing protein traces from purified proteins described in this report demonstrates for the fist time the applicability of combinatorial ligand libraries to downstream processing applications. To date, these libraries have been extensively described for drug discovery [30, 31] or for the discovery of ligand structures applicable to the selective capture of well-identified proteins [32, 33]. More recently, similar libraries have been described as a powerful means to amplify the signal for low-abundance proteins in proteomics investigations [17, 18, 29]. Due to their large spectra, these libraries can easily be used under mixed bed configuration for the adsorption of many proteins at a time; consequently they can be used for the adsorption of many diverse proteins such as HCP present in trace amount in purified biological preparations.

The large number of ligands and their diverse specificity makes possible the adsorption of proteins of different species, conferring to the process a wide applicability. This was demonstrated with the removal of *E. coli* protein traces as well as *P. pastoris* protein impurities and human plasma proteins.

Compared to the well-established GPC method, the ligand library can remove proteins of similar molecular mass and does not significantly dilute the target protein. Additionally, the process is applied as a regular adsorption chromatography column in terms of flow rate and handleability. Used as a capturing method for protein impurity traces, the column volume is significantly smaller than the volume of the sample to be treated; nevertheless, its size depends on the amount and the diversity of protein impurities to be eliminated.

In the last few years processes for protein purification have been modified or optimized for better productivity and applicability at large scale. The notion of a purification process based on three main steps (initial “capture”, “intermediate fractionation” and final “polishing”) introduced about 10 years ago is becoming reality with the direct loading of the feed stock under physiological conditions out of the fermentor (this is generally an affinity capture column). For certain proteins (especially for anti-

bodies), the capturing column is already well defined. Intermediate purification is generally constituted of ion exchangers or of hydrophobic interaction resins; they are both also relatively well standardized. What was still missing was the use of a polishing step of large specificity applicable to a large number of situations. With the use of a combinatorial ligand approach, it is believed that a complete standardization of a protein purification process from very crude feed stocks will shortly become a reality. In a full purification process constituted of the above-described three main steps, only one of them – the capture – would be case-by-case dependent. Under this configuration the design of a downstream process could be limited to the discovery or the selection of a proper ligand for the initial capture phase while all other steps would be generic.

The application of the described polishing process could be extended to other domains such as the preparation of pure markers still required for analytical methods such as very pure proteins of either different p*I* for IEF analysis or well-defined masses as markers for electrophoresis and MS analysis.

Finally, a complementary application to the polishing process is the amplification, and identification of impurity traces from biopharmaceuticals that are not detectable using current analytical methods (Fortis, F., Guerrier, L., Hayes, T., Antonioli, P. *et al.*, submitted for publication).

PGR is supported in part by grants from MURST (PRIN, progetto coordinato 40%, 2005), by FIRB 2001 (grant No. RBNE01KJHT) and by the European Community, Contract No. 12793, project "Allergy Card".

5 References

[1] Janson, J. C., in: Janson, J. C., Rydén, L. (Eds.), *Protein Purification*, Wiley-VCH, New York 1998, pp. 3–40.

[2] Simpson, R. J., in: Simpson, R.J. (Ed.), *Purifying Proteins for Proteomics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor 2004, pp. 17–40.

[3] Wisniewski, R., Boschetti, E., Jungbauer, A., in: Avis, K. E., Wu, V. L. (Eds.), *Biotechnology and Biopharmaceutical Manufacturing Processing and Preservation*, vol. 2, Interpharm Press, Buffalo Grove, IL 1996, pp. 61–181.

[4] Östlund, C., Borwell, P., Malm, B., *Dev. Biol. Stand.* 1987, *66*, 367–375.

[5] Duffy, S., Moellering, B. J., Prior, G. M., Doyle, K. R. *et al.*, *BioPharm.* 1989, *June*, 35–47.

[6] Abendroth, J., Chatterjee, S., Schomburg, D., *J. Chromatogr. B* 2000, *737*, 187–194.

[7] Gibert, S., Bakalara, N., Santarelli, X., *J. Chromatogr. B* 2000, *737*, 143–150.

[8] Brion, F., Rogerieux, F., Noury, P., Migeon, B. *et al.*, *J. Chromatogr. B* 2000 *737*, 3–12.

[9] Laurent, T. C., Kilander, J., *J. Chromatogr.* 1964, *14*, 317–330.

[10] Zanette, D., Soffientini, A., Sottani, C., Sarubbi, E., *J. Biotechnol.* 2003, *101*, 275–287.

[11] Dieryck, W., Noubhani, A. M., Coulon, D., Santarelli, X., *J. Chromatogr. B* 2003, *786*, 153–159.

[12] Lauer, I., Bonnewitz, B., Meunier, A., Beverini, M., *J. Chromatogr. B* 2000, *737*, 277–284.

[13] Roder, C., Krusat, T., Reimers, K., Werchau, H., *J. Chromatogr. B* 2000, *737*, 97–101.

[14] Mrabet, N. T., *Biochemistry* 1992, *31*, 2690–2702.

[15] Herzer, S., Kinealy, K., Asbury, R., Beckett, P. *et al.*, *Protein Expression Purif.* 2003, *28*, 232–240.

[16] Scopes, R. K., *Protein Purification; Principles and Practice*, Springer Verlag, New York 1987, pp. 133–142.

[17] Lomas, L., Lathrop, J., Thulasiraman, V., Hammond, D. *et al.*, *Electrophoresis* 2005, *26*, 3561–3571.

[18] Castagna, A., Cecconi, D., Rappsilber, J., Guerrier, G. *et al.*, *J. Proteome Res.* 2005, *4*, 1917–1930.

[19] Bjorklund, M., Hearn, M. T., *J. Chromatogr. A* 1996, *743*, 145–162.

[20] Wiesner, A., *Curr. Pharm. Biotechnol.* 2004, *5*, 45–67.

[21] Petsch, D., Anspach, F. B., *J. Biotechnol.* 2000, *76*, 97–119.

[22] Kepka, C., Lemmens, R., Vasi, J., Nyhammar, T. *et al.*, *J. Chromatogr. A* 2004, *1057*, 115–124.

[23] Eriksson, K-O., in: Vijayalakshmi, M. A. (Ed.), *Biochromatography: Theory and Practice*, Taylor & Francis Publisher, New York 2002, pp. 9–23.

[24] Boschetti, E., Jungbauer, A., in: Ahuja, S. (Ed.), *Handbook of Bioseparations*, vol. 2, Academic Press, San Diego, CA 2000, pp. 535–632.

[25] Eaton, L. C., *J. Chromatogr. B* 1995, *705*, 105–114.

[26] Haslbeck, M., Schuster, I., Grallert, H., *J. Chromatogr. B* 2003, *786*, 127–136.

[27] Takakura, T., Ito, T., Yagi, S., Notsu, Y. *et al.*, *Appl. Microbiol. Biotechnol.* 2006, *70*, 183–192.

[28] Jungbauer, A., Boschetti, E., *J. Chromatogr. B* 1994, *662*, 143–179.

[29] Guerrier, L., Thulasiraman, V., Castagua, A., Fortis, F. *et al.*, *J. Chromatogr. B* 2006, *833*, 33–40.

[30] Lam, K. S., Htuby, V. J., Lebl, M., Knapp, R. J. *et al.*, *Bioorg. Med. Chem. Lett.* 1993, *3*, 419–424.

[31] Gordon, E. M., Barret, R. W., Dower, W. J., Fodor, S. P. A. *et al.*, *J. Med. Chem.* 1994, *37*, 1386–1401.

[32] Buettner, J. A., Dadd, C. A., Baumbach, G. A., Masecar, B. L. *et al.*, *Int. J. Pept. Protein Res.* 1996, *47*, 70–83.

[33] Bastek, P. D., Land, J. M., Baumbach, G. A., Hammond, D. H. *et al.*, *Sep. Sci. Technol.* 2000, *35*, 1681–1706.

research articles Journal of proteome research

A New Approach for the Detection and Identification of Protein Impurities Using Combinatorial Solid Phase Ligand Libraries

Fréderic Fortis,[†] Luc Guerrier,[†] Liliana B. Areces,[‡] Paolo Antonioli,[§] Tim Hayes,[||] Kevin Carrick,[||] David Hammond,[||] Egisto Boschetti,[†] and Pier Giorgio Righetti*[,‡]

Ciphergen Biosystems, Cergy-Pontoise, France, Institute FIRC for Molecular Oncology Foundation (IFOM), Via Adamello 16 and European Institute of Oncology (IEO), Via Ripamonti 435, 20129 Milano, Italy, Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Via Mancinelli 7, Milano 20131, Italy, and American Red Cross, Biomedical R&D, Rockville, Maryland, U.S.A.

Received March 14, 2006

We propose a novel method for detection of protein impurities present in plasma-derived and recombinant purified injectable biopharmaceuticals by enhancing the concentration of protein impurities, in essence "amplifying" their presence to detectable levels. The method is based on the capture of proteins using a combinatorial solid-phase hexapeptides ligand library previously described for the reduction of protein concentration difference in biological fluids. Three proteins have been investigated: *Staphylococcus aureus* Protein A, expressed in *Escherichia coli* and supplied as 99% pure, recombinant human albumin, expressed in *Pichia pastoris* and certified as 95% pure, and therapeutic albumin supplied as 96−98% pure injectable solution. In all cases, after treatment with the ligand libraries, a number of additional polypeptide chains, not visible in the control, could be detected and obtained in sufficient amounts for MS analysis. In the cases of the two recombinant proteins, it could be demonstrated that a number of these polypeptide chains were host cell proteins still present in the purified product. In addition, a substantial number of these spots were found to be cleavage products of the original recombinant DNA species. Such cleavage products were particularly abundant in the recombinant human albumin preparation. From pure injectable serum albumin, a number of human plasma protein impurities were also identified by LC−MS/MS analysis. Treatment with ligand libraries of purified proteins is thus seen as a very powerful method of capture and concentration of host proteins and cleaved products for further analysis to control better the quality of industrial biotechnology products.

Keywords: host cell proteins • impurity analysis • mass spectrometry • ligand library • peptides ligands

Introduction

Most of biopharmaceuticals today are products of recombinant DNA technology or are derived from human plasma. Recombinant proteins are expressed in selected host cells under controlled conditions, whereas human plasma-derived products are extracted from pooled human plasma. Both are complex starting materials full of thousands of proteins that are potential impurities for the final product that may, in rare cases, cause adverse events in the patient ranging from a slight fever to long-term immunogenicity to toxic and, in even rarer cases, fatal events. Host cells used for the biosynthesis of recombinant proteins are relatively complex systems extending from bacteria (e.g., *Escherichia coli*),[1−5] to yeasts (e.g., *Pichia pastoris*)[6−10] and to eukaryotic cells such as chinese hamster ovarian cells (CHO).[11] During culture, these cells secrete a very large number of their own proteins that can easily contaminate the recombinant DNA product.[12−16] Even after sophisticated purification steps, significant levels of host cell proteins (HCP) may still remain present in the final purified biopharmaceutical.[17−23] Although host cell impurities are mostly innocuous to the receiving patient, regulatory agencies require demonstrations that HCPs are not only minimized but also analyzed with the most sensitive available methods. Current analytical methods are limited in number and also not sufficiently sensitive for the detection of trace levels of HCPs. Current HPLC techniques have good resolution; however, they suffer from their low sensitivity, the possibility of non-specific binding, and subjective interpretation. Electrophoretic analytical methods (e.g., SDS-PAGE with silver staining) also offer good resolution, but sensitivity is low and the interpretation remains very subjective. Immunological determinations are more specific than HPLC and electrophoresis; however, they suffer from the limited specifications and varied affinities of the selected antibodies. Moreover, they may not be reproducible from lot to lot, especially when very low concentration species must be detected. In addition, detection of HCPs is only possible if the

* To whom correspondence should be addressed. E-mail: piergiorgio.righetti@polimi.it.
[†] Ciphergen Biosystems.
[‡] Institute FIRC for Molecular Oncology Foundation (IFOM).
[§] Politecnico di Milano.
[||] American Red Cross.

10.1021/pr060090s CCC: $33.50

Published on Web 08/24/2006

whole antigen repertoire is covered by antibodies. If antibodies to a specific antigen are missing, the impurity simply escapes detection. Immunodetection can be separated into two groups: Western blots and Immunoassays. The former has a sensitivity of about 10 ppm and the latter a sensitivity lower than 1 ppm. In conclusion, all detection methods for HCPs face a most challenging problem: how to deal with very low concentrations of contaminating proteins present in "pure" biopharmaceuticals after separation/purification with current processing techniques.

We have recently described a novel approach for capturing the "hidden proteome", i.e. those rare and very rare proteins that constitute the vast majority in any cell or tissue lysate and in biological fluids.[24−29] It is based on a combinatorial library of hexameric peptide ligands bound to porous polyacrylate beads. Each bead contains a single hexapeptide sequence distributed throughout its porous structure reaching concentrations of about 50 pmol per bead. This amounts to a ligand density of ca. 40−60 μmoles per mL of bead volume (average bead diameter of about 60 μm). Each single bead has millions of copies of a single, unique ligand and each bead potentially has a different ligand from every other bead. Since the 20 natural amino acids were used for the synthesis, the library would contain a population of 20^6 i.e., 64 million different linear hexapeptides. With such vast populations of heterogeneous ligands, an appropriate volume of beads could theoretically interact with most if not all proteins present in a complex proteome (be it a biological fluid or a tissue or cell lysate of any origin). This ligand library has been efficiently used for capturing and revealing a very large population of previously undetected urine[27] and serum[30] proteins. Although these combinatorial ligand libraries have been used, so far, for classical proteomic approaches, purification of proteins, and identification of new activities, another important application would be the one described in the present report, i.e. capturing and "amplifying" trace amounts of impurities present in all r-DNA proteins available on the market. This approach would have a double advantage: on the one hand, it would provide sufficient amounts of such impurities to allow for their chemical characterization; on the other hand, it would permit purifying r-DNA products to an extreme degree, not achievable with present-day technologies.

Materials and Methods

Chemicals, Biologicals, and Materials. Recombinant albumin expressed in *P. pastoris* and purified at 95%, molecular mass standards, as well as chemicals such as urea, sodium dodecyl sulfate, thiourea, 3-[3-cholamidopropyl dimethylammonio]-1-propansulfonate (CHAPS), and the fluorescent stain Sypro Ruby were all from Sigma-Aldrich, St Louis, Mo. Recombinant Protein A expressed in *E. coli* and purified at 99% level was from Repligen, Waltham, MA. Injectable 5% human serum albumin solution of 96−98% purity (Albumarc) was originally marketed by American Red Cross, Rockville, MD. The solid-phase combinatorial peptide library was prepared by Peptide International, Louisville, KY, according to instructions of American Red Cross. This was first described by Buettner et al.[31] and then modified for proteomics research as recently reported.[32]

ProteinChip Arrays were supplied by Ciphergen Biosystems Inc., Fremont, CA. Ten-well, pre-cast, 1-mm-thick 16% polyacrylamide Tricine gel plates, Colloidal Coomassie and silver staining reagents were supplied by Invitrogen, Carlsbad, CA. Rabbit antibodies against *P. pastoris* proteins and goat anti mouse IgG labeled (horseradish peroxidase) antibodies were supplied by Biodesign International, Saco, Maine. Tris, mineral oil, DL-dithiothreitol (DTT), Tween 20, and EDTA were purchased from Sigma-Aldrich Chemie Gmbh (Steinheim, Germany). Glycerol, methanol, ethanol, acetic acid, and acetone were from Merck (Darmstadt, Germany). Acrylamide/Bis solution (40%), *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TEMED), acrylamide (AA), ammonium persulfate (APS), the Protean IEF Cell, the GS-710 Densitometer, the Versa Doc Scanner, the software PDQuest Version 7.2, as well as the linear Immobiline dry strips pH gradient 3−10 and 4−7 (11 cm) were obtained from Bio-Rad Labs, Hercules, CA; glycine, sodium dodecyl sulfate (SDS), iodoacetamide, urea, thiourea, tributylphosphine (TBP), and CHAPS were from Fluka. Bromophenol blue, carrier ampholytes, and agarose were purchased from Pharmacia Biotech (Uppsala, Sweden).

HCP Amplification with Combinatorial Ligand Library. A solution of the purified protein from which impurities were to be recovered was prepared in phosphate buffered saline (PBS). Separately, a small chromatographic column filled with solid phase combinatorial peptide library (1 mL for both recombinant Protein A and recombinant albumin and 0.5 mL for injectable serum albumin) was packed, equilibrated with physiological saline, and connected to a chromatographic system comprising a pump and a UV/pH detection unit for recording both events at the column outlet. The column was loaded continuously with protein solutions at a linear flow rate of 50 cm/hour. The flowthrough was eliminated, and when the loading was completed, a PBS solution was introduced for washing out the excess proteins. Captured proteins were desorbed from the beads and then analyzed. In the case of recombinant proteins, the elution solution was 9 M urea−citric acid, pH 3.5 containing 2% Chaps; in the case of injectable albumin, the recovery before electrophoresis was performed using Laemmli buffer at 100 °C for 5 min. The protein concentration of each loaded solution was 15 mg/mL for Protein A, 10 mg/mL for recombinant albumin, and 50 mg/mL for injectable albumin; the total volumes loaded were 200, 100, and 10 mL, respectively.

Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE). Electrophoresis of protein impurities collected compared to control was performed using 10-well, pre-cast, 1-mm-thick 16% polyacrylamide Tricine gel plates. Samples of appropriate protein concentration were diluted 2-fold in sample buffer. Diluted sample (30 μL) was loaded per lane and electrophoresis migration was performed at 150 V for 90 min. Staining and destaining were performed according to the supplier's instructions.

Immunoblot Analysis. After electrophoresis migration according to the above-described SDS-PAGE method, the plate was submitted to an electrotransfer of proteins on nitrocellulose of 0.45 mμ under 20 V, 600 mA for 4 h. Total proteins were revealed by using amidoblack, and proteins from *P. pastoris* were revealed by immunostaining. The latter was accomplished by a first incubation of nitrocellulose membrane in a solution of antibodies (500 μL) diluted to 4.5 mL of a solution of 3% of bovine albumin in physiological saline for 30 min at room temperature. After an extensive rinsing with physiological saline, 12.5 μL of secondary antibody (goat IgG against rabbit IgG labeled with horseradish peroxidase, HRP) diluted with 5 mL of physiological saline was added, and the cellulose membrane was incubated for an additional of 30 min. Finally,

the membrane was extensively washed to eliminate the excess of antibody solution and HRP activity revealed with the specific substrate supplied by the antibody provider.

2D-PAGE Analysis. For proper comparisons, the same protein amount, i.e., 200 μg, was used for all the samples investigated (starting material and bead eluates). The protein samples, after being dialyzed against ammonium acetate and subsequently lyophilized, were directly solubilized in "2D sample buffer" (7 M urea, 2 M thiourea, 3% CHAPS, 40 mM Tris, 5 mM TBP, and 10 mM acrylamide) and alkylated at room temperature for 90 min. To stop the alkylation reaction, 10 mM DTT was added to the solution followed by 0.5% Ampholine and a trace amount of bromophenol blue to the solution. IPG strips (7-cm long) (Bio-Rad) pH 3−10 or 4−7 were rehydrated with 150 μL of protein solution for 4 h. Isoelectric focusing (IEF) was performed with an initial voltage gradient from 100 up to 1000 V in 5 h, followed by 1000 V constant for 5 h. The voltage was then increased again rapidly up to 5000 V in 30 min and maintained at such a value until reaching 30 kV/hours. For the second dimension, the IPG strips were laid on an 8−18% acrylamide gradient SDS-PAGE. The electrophoretic run was performed by setting a current of 5 mA/gel for 1 h, followed by 10 mA/gel for 1 h and 20 mA/gel until the dye front reached the bottom of the gel. Gels were then immediately stained in colloidal Coomassie Blue. Destaining was performed in 7% acetic acid until the background become completely transparent. The 2DE gels were scanned with a Versa-Doc Imaging System (Model 3000, Bio-Rad, Hercules CA), and the images were treated with the software PDQuest version 8.0 (Bio-Rad, Hercules CA).

SELDI MS Analysis. Protein solutions at appropriate concentration, i.e., 0.02 μg/μL, were deposited upon ProteinChip Array surfaces (Ciphergen Biosystems Inc.), using a Bioprocessor. Three types of arrays were selected: CM10 (weak cation exchanger), Q10 (strong anion exchanger), and IMAC 30 (Immobilized Metal ions Affinity Capture) loaded with copper ions. Each array contained eight distinct spots over which the adsorption of protein could be performed. After applying the three samples, all surfaces were dried and prepared for SELDI-TOF MS analysis by applying 1 μL of matrix solution composed of a saturated solution of sinapinic acid in 50% acetonitrile and 0.5% trifluoroacetic acid. All arrays were then analyzed with a Ciphergen PCS 4000 ProteinChip Reader. The instrument was used in a positive ion mode with an ion acceleration potential of 20 kV and a detector gain voltage of 2 kV. The mass range investigated was from 3 to 20 kDa. Laser intensity was set between 200 and 250 units according to the sample tested. The instrument was calibrated with "All-in-1 protein standard" mixture (Ciphergen Biosystems Inc.).

Digestion of Proteins and Sample Preparation. For recombinant proteins, the classical "in gel digestion" was performed based on a standard protocol. In brief, the gel pieces excised from the 2D maps were transferred manually into an individual well of a V-bottom polypropylene 96-well plate (Pink Microtiter plates-pierced, Genomic Solutions) and washed sequentially once with 100 μL of water and 100 μL of 25 mM NH_4HCO_3 and then dehydrated by adding 100 μL of acetonitrile. The gel pieces were subsequently swollen in 100 μL of 10 mM DTT for 30 min at RT. The gel pieces were shrunk with acetonitrile, and the proteins were alkylated with 100 μL of 55 mM iodoacetamide for 20 min at RT in the dark. All the liquid material was transferred by centrifugation at 1000 rpm for 2 min. After drying completely, the gel pieces were reswollen with 20−25 μL of digestion buffer (12.5 ng/μL trypsin in 25 mM NH_4HCO_3) for 30 min at 4°C. Afterward, the digestion buffer was replaced by the same buffer without trypsin to ensure that the gel plugs were kept wet during digestion process. The plate was covered and incubated at 37 °C for 12−18 h. Following digestion, the peptide mixture was desalted and concentrated using home-made C_{18}−Stage tips. The peptides were eluted directly onto a collection plate using 10 μL of 80% acetonitrile/1% TFA by vacuum (4−6 Hg). Following elution, the volume of the extracted peptides was reduced to 2 μL using a centrifugal vacuum concentrator (Eppendorf, concentrator 5301).

Digestion of captured proteins from injectable serum albumin was performed directly on the beads without elution. Briefly, the hexapeptide library beads with captured proteins were removed from the column, washed with 50 mM ammonium carbonate pH 7.8 and centrifuged for supernatant removal. A solution of 0.3% of rapigest was then added to the beads along with 50 mM of dithiothreitol and heated at 100 °C for approximately 5 min in a boiling water bath. The resulting sample was added to 150 mM solution of iodoacetamide and stored in the dark for approximately 60 min. Trypsin (0.5 mg/mL) was added to the resulting solution, and digestion of proteins was carried out at 37 °C for 16 h. HCL (500 mM, 10 μL) was added and incubated for 40 min. Finally, the sample was centrifuged at 13 000 rpm for 10 min. The volume was then reduced by speed vacuum and 0.1% formic acid was added prior to analysis by LC/MS/MS.

Protein Identification by Mass Spectrometry and Data Analysis. For recombinant proteins, the total tryptic digest samples were analyzed by reverse-phase nano-LC−MS/MS using an Agilent 1100 nanoflow LC system (Agilent Technologies Inc.). The LC−system was coupled to a QSTAR XL hybrid QqTOF mass spectrometer (Applied Biosystem, MDS Sciex). Binding and chromatographic separation of peptides was achieved in a 15 cm fused silica emitter (100 μm ID from Picotip New Objective) packed in-house with reverse phase C18 resin (Maisch, Germany). The tryptic peptide mixtures were loaded using an auto-sampler and at a flow rate of 700 nL/min into the packed column. The peptides were separated in a linear gradient of acetonitrile in acetic acid: solution A was 0.5% acetic acid in water containing 5% (v/v) acetonitrile and solution B was 0.5% acetic acid in water containing 24% (v/v) acetonitrile. The gradient was developed in 15 min and then a solution of 0.5% acetic acid containing 84% acetonitrile was injected for 5 min at a flow rate of 300 nL/min. The QSTAR was operated in data-dependent acquisition mode to automatically switch between MS and MS/MS. Protein identification was performed using Mascot (Matrix Science, London, UK) to search the nonredundant protein database. Search parameters were: MS and MS/MS tolerance 0.2 Da, tryptic specificity allowing for up to 1 missed cleavage, fixed modification: carbamidomethylation of cysteine, variable modification: oxidation of methionine.

In the case of injectable albumin, 5 μL of the digest was loaded onto a nanospray tip loaded with C18 resin. The sample was eluted using a continuous gradient and directly infused into a LCQ DECA XP mass spectrometer. The data were searched against a human protein database using Sequest.[33]

Results

Initially, we analyzed the impurities extracted from recombinant Protein A, expressed in *E. coli* supplied as 99% pure by the manufacturer. This protein binds specifically to the Fc

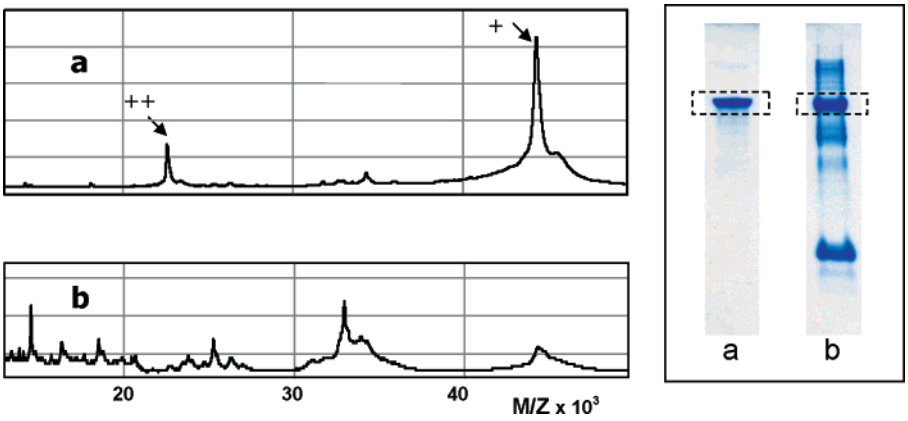


Figure 1. (a) SELDI MS and SDS-PAGE analysis of native Protein A and (b) host cell protein impurities extracted using peptide combinatorial library. The boxes in SDS-PAGE images represent the positioning of Protein A and arrows show the single (+) and double charged (++) Protein A ions. *m*/*z* = mass over charge.

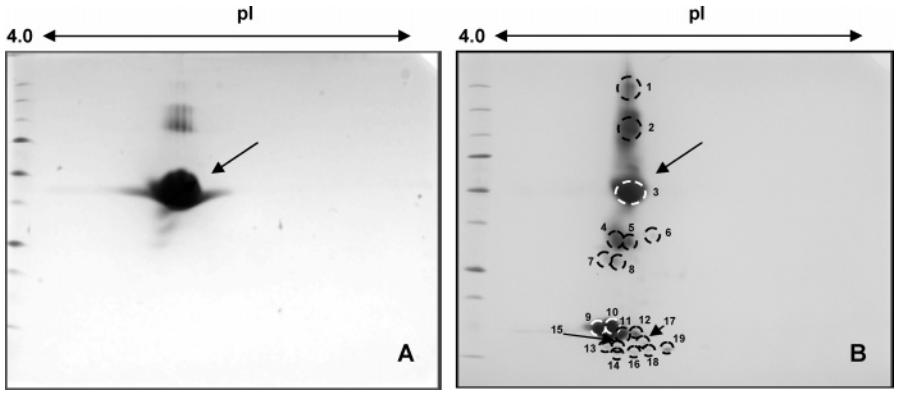


Figure 2. (A) Two-dimensional electrophoresis maps of native Protein A and (B) HCP impurities extracted using peptide combinatorial library. p*I* gradient was from 4.0 to 7.0. It is to be noticed that impurities and Protein A have relatively similar isoelectric points, suggesting a pH-driven purification process. Arrow indicates the positioning of Protein A. The numbers in panel B indicate the spots eluted and subjected to MS analysis (se data in Table 1). The arrows indicate the position of Protein A.

region of immunoglobulin through interaction with the heavy chain, and thus, it is extensively used as a ligand for the purification of therapeutic monoclonal antibodies. Figure 1 shows the SDS-PAGE and SELDI MS profiles of the product as commercially available (upper panel) and of the eluate collected after treatment with the combinatorial library beads (lower tracing). In the first case, mainly one band is visible, with only traces of very few, minor components. Conversely, in the eluate from the ligand library, where all impurities are collected and "amplified" (highly concentrated from a dilute solution), many more peaks are visible in both tracings. Differences between the initial product and the beads eluate are also demonstrated in the 2D map analysis (Figure 2) where a significant amplification of the present impurities is shown (panel B). Most proteins present have isoelectric points within a narrow pH range (between 4.5 and 5.5); this is due to the purification of recombinant Protein A from the crude extract by anion exchange chromatography. Whereas the control (panel A) shows the main Protein A component, with at most 3−4 minor impurities, the eluate from the ligand library (panel B) shows, in addition to the target protein, 20 or more other species, of which 7−8 appear to be most abundant. As schematically indicated in the figure, the main spots from the 2D map have been cut and subjected to proteolysis prior to MS identification. Table 1 gives the IDs of the 19 spots analyzed. It appears that the 3 most abundant spots (1−3) represent a tetramer (1), a dimer (2), and the monomer (3) of Protein A. Spots 4, 5, 7, and 8 are fragments of protein A, probably co-purified due to their similar p*I* values with the intact protein. The cluster of polypeptide chains centered at p*I* values 4.5−5.8 and mass values from 10 to 15 kDa are host (*E. coli*) proteins carried over during the purification process (probably due to the fact that their p*I* values are clustered around the p*I* of protein A). It is of interest to note that not one of them was visible in the control sample.

The second sample analyzed is human albumin, expressed in *P. Pastoris* and supplied as a 95% pure product. Figure 3 shows the SDS-PAGE and SELDI MS profiles of the product as commercially available (upper panel) and of the eluate collected after treatment with the hexapeptide library beads (lower panel). Here, the two most intense peaks in the SDS profile (confirmed by SELDI MS analysis) represent the albumin monomer and dimer, respectively (marked 1 and 2, double and single charged albumin, and 3, representing the dimer). In the lower panel, the SDS-tracing of the eluate from the ligand beads show much higher number of impurities, as also detectable by the SELDI MS signals. SDS-PAGE and SELDI data are not redundant, but rather complementary because the first is a better indicator for high molecular mass components and the second is more sensitive for low molecular masses. Immunoblot analysis demonstrated that these impurity components could largely be host cell proteins. The extract from the beads

Table 1. Protein Identification by LC−MS/MS from Spots Cut on 2D Maps: The Case of Recombinant *Staphylococcus aureus* Protein A

gel spot	NCBI access	protein ID	cover. %	theor. p*I*	theor. Mr (Da)	exper. p*I*	exper. Mr (Da)	molecular forms
1	gi|586027	SPA2_STAAU(Staphylococcal protein A)	36	5.53	55406	5.3	200000	PA tetramer
2	gi|586029	SPA2_STAAU(Staphylococcal protein A)	36	5.53	55406	5.3	100000	PA dimer
3	gi|586031	SPA2_STAAU(Staphylococcal protein A)	41	5.53	55406	5.3	52000	PA monomer
4	gi|586033	SPA2_STAAU(Staphylococcal protein A)	38	5.53	55406	5.3	33000	PA frag 1
5	gi|586035	SPA2_STAAU(Staphylococcal protein A)	39	5.53	55406		39000	PA frag 2
6		unidentified						
7	gi|586026	SPA1_STAAU(Staphylococcal protein A)	26	5.54	57286	4.8	28000	PA frag 3
8	gi|586026	SPA1_STAAU(Staphylococcal protein A)	23	5.54	57286	5.0	27000	PA frag 4
9	gi|16130920	hypothetical protein b3024_E coli K12	53	5.08	14059	4.7	15000	B3024 monomer
10	gi|16130920	hypothetical protein b3024_E coli K12	51	5.08	14059	4.9	15000	B3024 monomer
11	gi|16130920	hypothetical protein b3024_E coli K12	53	5.08	14059	5.1	14000	B3024 monomer
12	gi|16128400	riboflavin synthase. beta chain_E coli K12	32	5.15	16147	5.4	14000	RSBC
13	gi|16130920	hypothetical protein b3024_E coli K12	32	5.08	14059	4.8	13000	B3024 frag 1
14	gi|16130920	hypothetical protein b3024_E coli K12	32	5.08	14059	5.0	13000	B3024 frag 2
15	gi|16130920	hypothetical protein b3024_E coli K12	41	5.08	14059	5.0	12000	B3024 frag 3
16	gi|16130920	hypothetical protein b3024_E coli K12	32	5.08	14059	5.5	12000	B3024 frag 4
17	gi|16131554	heat shock protein_E coli	34	5.19	16315	5.6	13000	HSP frag 1
18	gi|732038	UPF0076 protein yjgF	34	5.36	13660	5.8	12500	YJGF frag 1
19	gi|732038	UPF0076 protein yjgF	64	5.36	13660	6.0	13000	YJGF monomer

contained many more signals than the initial product while the first eluate was negative toward the immunoblot, suggesting a high degree of purity of purified albumin.

Data from Figure 3 have been fully validated by the 2D maps of Figure 4, here representing the commercial product (panel A) and the eluate from the ligand library (panel B). In this last case, many more protein spots are visible, scattered in the 2D plane. A total count after background cleaning and streak subtraction, performed with the PDQuest software, revealed more than 50 polypeptide chains.

As further proof of the above data, 26 of the visible spots in the 2D maps of Figure 4 have been excised, digested and analyzed by mass spectrometry on a Q-STAR instrument. The data are listed in Table 2: here, the situation is quite different from that of Protein A. It would appear that most of these spots represent albumin fragments, ranging in mass from the native (Mr 71317) species to as low as ca. 8000 Da, as assessed from the SDS dimension of the 2D gel. The large number of fragments observed may be a result of cleavages by proteases present in the culture media. A number of *P. pastoris* proteins have also been identified among the spots analyzed. Here too it is of interest to note that, in the control, essentially not one of these spots could be detected, except for the large albumin fragments with a mass of ca. 50 kDa (spot No. 2, Figure 4).

Similarly, a third protein, injectable human albumin from serum supplied as 96−98% pure solution, has been analyzed before and after impurity "amplification" with the hexapeptide combinatorial library beads. SDS-PAGE analysis demonstrated the apparent purity of the product (Figure 5). However, following treatment, many additional protein traces were observed. LC−MS/MS analysis of the product identified only a single impurity, haptoglobin, whereas, after treatment, many impurities have been identified (Figure 5).

Table 3 assembles data from identification of serum proteins by LC−MS/MS.

Discussion and Conclusions

The mechanism by which combinatorial ligand libraries enhances low and very low abundance species from complex mixtures has been described elsewhere.[24] Briefly, each protein of the complex mixture binds a few beads carrying the

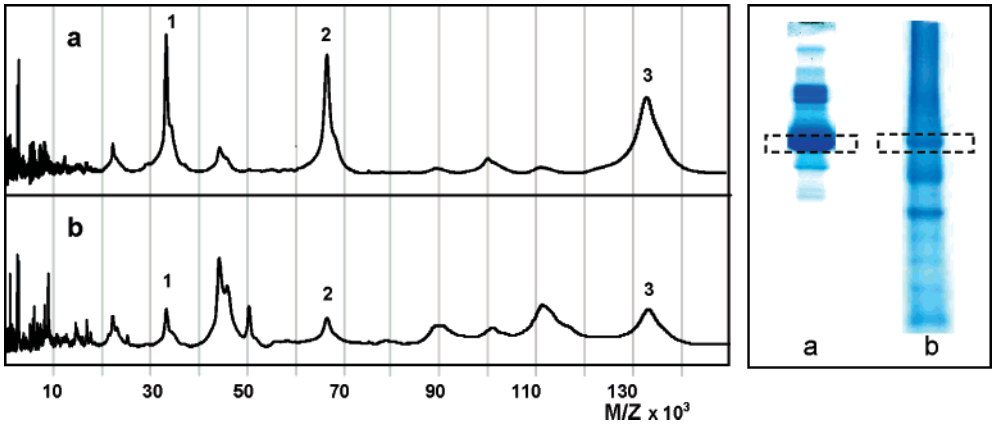


Figure 3. (a) SELDI MS and SDS-PAGE analysis of initial recombinant human albumin and (b) HCP impurities extracted using peptide combinatorial library. The mass spectrometry analysis has been performed using a Q10 ProteinChip array. 1, 2, and 3 represent ions of respectively double-charged albumin, single-charged albumin, and albumin dimmer. The boxes in SDS-PAGE images represent the positioning of recombinant albumin in respectively the initial purified protein (95% purity) and the extract from hexapeptide combinatorial library. *m*/*z* = mass over charge.

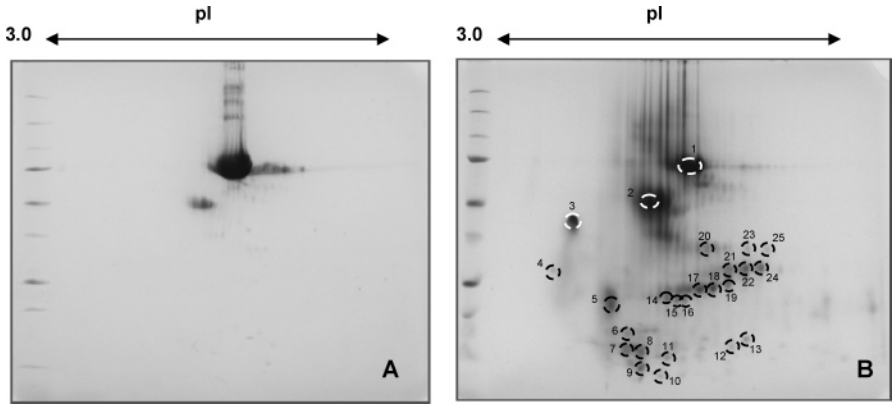


Figure 4. (A) Two-dimensional electrophoresis maps of native purified recombinant human albumin (95% purity) and (B) host cell protein impurities extracted using peptide combinatorial library. The pH gradient was from 3.0 to 10.0. Impurities are extended over a large zone of isoelectric points and molecular masses. The numbers in panel B indicate the spots eluted and subjected to MS analysis (see data in Table 2).

Table 2. Identification of Protein Impurities Contained in Recombinant Human Albumin Expressed in *Pichia pastoris*[a]

gel spot	NCBI access	protein ID	cover. %	theor. p*I*	theor. Mr (Da)	exper. p*I*	exper. Mr (Da)	molecular forms
1	gi|113576	ALBU_HUMAN (Serum albumin precursor)	35	5.92	71317	6.5	72000	HSA monomer
2	gi|113576	ALBU_HUMAN (Serum albumin precursor)	37	5.92	71317	5.8	52000	HSA frag 1
3	gi|28590	ALBU_HUMAN (Serum albumin precursor)		5.92	71246	4.2	45000	HSA frag 2
4		unidentified						
5	gi|28590	ALBU_HUMAN (Serum albumin precursor)	23	5.92	71246	5.0	23000	HSA frag 3
6	gi|28590	ALBU_HUMAN (Serum albumin precursor)	9	5.92	71246	5.3	17000	HSA frag 4
7	gi|28590	ALBU_HUMAN (Serum albumin precursor)	12	5.92	71246	5.2	14000	HSA frag 5
8	gi|2104961	alcohol oxidase (Pichia pastoris)	3	6.04	74479	5.7	12000	AO frag 1
9	gi|28590	ALBU_HUMAN (Serum albumin precursor)	14	5.92	71246	5.6	10000	HSA frag 6
10	gi|28590	ALBU_HUMAN (Serum albumin precursor)	12	5.92	71246	5.9	8000	HSA frag 7
11	gi|28590	ALBU_HUMAN (Serum albumin precursor)	13	5.92	71246	6.1	12000	HSA frag 8
12	gi|28590	ALBU_HUMAN (Serum albumin precursor)	5	5.92	71246	7.3	14000	HSA frag 9
13		unidentified						
14	gi|28590	ALBU_HUMAN (Serum albumin precursor)	22	5.92	71246	6.1	23000	HSA frag 10
15	gi|28590	ALBU_HUMAN (Serum albumin precursor)	19	5.92	71246	6.2	22000	HSA frag 11
16	gi|28590	ALBU_HUMAN (Serum albumin precursor)	11	5.92	71246	6.4	21500	HSA frag 12
17	gi|28590	ALBU_HUMAN (Serum albumin precursor)	25	5.92	71246	6.6	24000	HSA frag 13
18	gi|2495229	imidazoleglycerol-phosphate dehydratase (Pichia pastoris)	18	6.14	24322	6.8	24000	IGPD monomer
19	gi|2495229	imidazoleglycerol-phosphate dehydratase (Pichia pastoris)	18	6.14	24322	7.2	24000	IGPD monomer
20	gi|28592	ALBU_HUMAN (Serum albumin precursor)	32	6.05	71319	6.8	35000	HSA frag 14
21	gi|28592	ALBU_HUMAN (Serum albumin precursor)	16	6.05	71319	7.3	30000	HSA frag 15
22	gi|46981803	ornithine carbamoyltransferase (Pichia pastoris)	18	6.46	37832	7.5	30000	OCMT frag 1
23	gi|28592	ALBU_HUMAN (Serum albumin precursor)	27	6.05	71319	7.5	34000	HSA frag 16
24	gi|28592	ALBU_HUMAN (Serum albumin precursor)	28	6.05	71319	7.5	30000	HSA frag 17
25	gi|28592	ALBU_HUMAN (Serum albumin precursor)	23	6.05	71319	8.0	34000	HSA frag 18
26	gi|46981803	ornithine carbamoyltransferase (Pichia pastoris)	11	6.46	37832	10.0	23000	OCMT frag 2

[a] Individual spots from 2D maps have been cut and the plugs digested with trypsin prior submission to ID by LC−MS/MS.

complementary ligands. Therefore, a larger percentage of each trace protein in the complex mixture is bound by the library than its more abundant counterparts. The amount of protein traces is increased by simply loading larger samples volumes of the initial product up to the saturation point. Experiments performed in this respect by using human urine and human serum demonstrated the strong capability of these beads to detect extremely low abundance proteins when the sample volume is large.[27,30] Washing the library followed by elution results in a sample enriched in trace components.[24] This process means that trace component proteins could be identified and possibly linked to function. With the present work, we effectively detected and identified protein impurities from naturally derived and recombinant pure products.

Plasma-derived products were the first generation biotherapeutic products (since 1941) primarily used for replacement therapy. Human plasma as a starting material for the production of biotherapeutics is very complex with an estimated of hundreds of thousands of different protein species present in a concentration range spanning several orders of magnitude. Therefore, when purifying a protein from plasma, traces of numerous other proteins could still be present. In the described process it has been demonstrated that even a highly purified albumin (96−98% pure) from plasma still carries trace contaminants as revealed by its trace impurity profile unmasked by treatment with combinatorial ligand libraries. Preliminary data on detection of protein impurities from purified human serum albumin were presented by T.

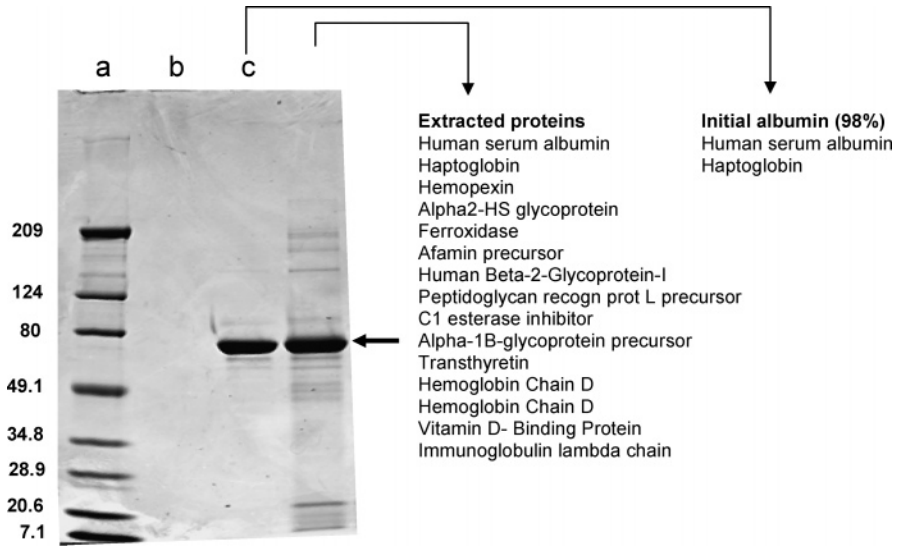


Figure 5. Sodium dodecyl sulfate electrophoresis of (Lane c) injectable human pure albumin from serum and (Lane d) captured proteins by the ligand library. The inserted summary table indicates the list of impurities found respectively in the initial sample and in the sample where impurity traces have been amplified. (Lane a) Ladder of standard protein masses. (Lane b) Void. The arrow indicates the position of albumin.

Table 3. Protein Impurity Identification by LC−MS/MS from Injectable Human Serum Albumin Before and After Treatment with Peptide Ligand Library

identified proteins	score	sequence coverage %	unique peptides
Before ligand library treatment			
human serum albumin	418.3	41	21
haptoglobin	20.2	4	1
After ligand library treatment			
human serum albumin	520.3	44	18
haptoglobin	214.2	26	7
hemopexin	108.2	18	4
alpha2-HS glycoprotein	60.2	2	2
ferroxidase	50.3	7	4
afamin precursor	40.2	9	3
human beta-2-glycoprotein-I	40.1	10	2
peptidoglycan recognition protein L precursor	30.2	7	2
C1 esterase inhibitor	30.2	9	3
alpha-1B-glycoprotein precursor	20.3	6	2
transthyretin	20.3	41	2
hemoglobin chain D	10.2	11	1
vitamin D-binding protein	10.2	5	1
immunoglobulin lambda chain	10.1	8	1

Hayes at BioCrossroads, February 2005, Montreal (personal communication).

Recombinant DNA proteins represent the latest generation of biotherapeutic products, since a number of them (e.g., erythropoietin, human growth hormone, superoxide dismutase, tissue plasminogen activator) are used as therapeutic agents. One of the major dilemmas associated with the production of r-DNA proteins for human therapy is the progressive reduction of the content of HCP by chromatographic purification steps. In principle, such recombinant proteins should be purified to homogeneity, in order to avoid undesired side effects when they are co-injected with traces of protein impurities that can be immunogenic, especially if derived from host proteins of nonhuman origin. In practice, even under the best conditions, it is very laborious to achieve purity levels better than 99%.[34,35]

In general, pharmaceutical companies are able to identify one, or a few, major impurities, and this governs the pedigree of the products available on the market for human therapy. The described experimental data demonstrate the capability of these libraries to extract and identify most, if not all, HCPs from both recombinant albumin expressed in *P. pastoris* and Protein A expressed in *E. coli*.

Impurities that were difficult to detect by current methods were concentrated by the described treatment and reduced observable by several analytical methods. Current analysis and western blot of initial products did not detect significant amount of protein impurities. Similarly, neither electrophoresis-based analysis nor mass spectrometry of pure recombinant Protein A expressed in *E. coli* revealed impurities. However, after concentration using the hexapeptides ligand library, proteins from *P. pastoris* on one side and *E. coli* on the other side were detected by electrophoresis and mass spectrometry. Additionally, degradation products of the expressed r-DNA protein were observed. Although some of the impurities present have been successfully identified, many visible spots from 2D analysis from both recombinant Protein A and recombinant albumin remained unidentified. In practice, the amount of protein present in very weak spots was insufficient for a proper identification.

The combined 2D-MS/MS data are illuminating: they show (in both cases, but especially for the albumin sample) that quite a number of truncated fragments are co-purified and carried over in the commercial products. In the case of Protein A, in which the final purification step was ion-exchange, fragments having similar p*I* values have been carried over along with a number of host proteins having similar p*I* values. Even if the capturing step had been performed via an immuno-affinity column, clearly a number of fragments would have been co-purified and their presence would not have been detected by classical ELISA-based assay methods. The analytical approach described in this paper also demonstrates the extra valence that 2D maps have over classical 2D chromatography. In the latter case, samples are typically pre-digested and then analyzed by MS and therefore fragments are not detected, suggesting a

much higher degree of purity than that found by our described approach. Detection of protein fragments indicates the presence of proteases; this is an important aspect in the detection of HCPs. Proteases in fact can be present in such small trace amounts that even the most sensitive available methods are ineffective for detection. Their presence is only evidenced by the formation of protein fragments during long-term storage.

Concentration and detection of HCPs applies to all types of host cell proteins with no distinction of species or origin. It does not replace the quantitative approaches used today, but it introduces a way of obtaining samples with a very significant enhancement of HCP concentration. Such a detection process for trace impurities depends, naturally, on the initial amount of sample treated: the higher the volume (and concentration) of treated product, the higher the number and concentration of protein impurities captured and thus detectable upon subsequent analysis. In this respect the enhancement of trace impurities might be, quite naturally, an impossible dream, because it is to be expected that a large number of impurities will be present in traces amounts, such that they might never be able to saturate the binding sites on the beads.

An additional important finding is that, even after extensive purification processes, the final isolated injectable biopharmaceuticals, such as r-DNA and natural extracted products, should be screened by 2D mapping and LC−MS/MS, to detect the presence of cleaved species. When used for human treatments, these fragments could become, with long-term use, immunogenic.

As a consequence of the above, we anticipate that purer, better characterized, and hence hopefully safer r-DNA products will be developed and marketed in the future.

Acknowledgment. Supported in part by grants from FIRB 2001 (Grant No. RBNE01KJHT), by the European Community, Contract No. 12793, project “Allergy Card”, and by Fondazione Cariplo (Milano).

References

(1) Whitmire, M. L.; Eaton, L. C. An immunoligand assay for quantitation of process specific *Escherichia coli* host cell contaminant proteins in a recombinant bovine somatotropin. *J. Immunoassay* **1997**, *18*, 49−65.
(2) Garke, G.; Deckwer, W. D.; Anspach, F. B. Preparative two-step purification of recombinant human basic fibroblast growth factor from high-cell-density cultivation of *Escherichia coli*. *J. Chromatogr., B: Biomed. Sci. Appl.* **2000**, *737*, 25−38.
(3) Dutta, S.; Lalitha, P. V.; Ware, L. A.; Barbosa, A.; Moch, J. K.; Vassell, M. A.; Fileta, B. B.; Kitov, S.; Kolodny, N.; Heppner, D. G.; Haynes, J. D.; Lanar, D. E. Purification, characterization, and immunogenicity of the refolded ectodomain of the *Plasmodium falciparum* apical membrane antigen 1 expressed in *Escherichia coli*. *Infect. Immun.* **2002**, *70*, 3101−3110.
(4) Hewitt, L.; McDonnell, J. M. Screening and optimizing protein production in *E. coli*. *Methods Mol. Biol.* **2004**, *278*, 1−16.
(5) Persson, J.; Lester, P. Purification of antibody and antibody-fragment from *E. coli* homogenate using 6,9-diamino-2-ethoxyacridine lactate as precipitation agent. *Biotechnol. Bioeng.* **2004**, *87*, 424−434.
(6) Piefer, A. J.; Jonsson, C. B. A comparative study of the human T-cell leukemia virus type 2 integrase expressed in and purified from *Escherichia coli* and *Pichia pastoris*. *Protein Expr. Purif.* **2002**, *25*, 291−299.
(7) Steinlein, L. M.; Graf, T. N.; Ikeda, R. A. Production and purification of N-terminal half-transferrin in *Pichia pastoris*. *Protein Expr. Purif.* **1995**, *6*, 619−624.
(8) Ouyang, J.; Wang, J.; Deng, R.; Long, Q.; Wang, X. High-level expression, purification, and characterization of porcine somatotropin in *Pichia pastoris*. *Protein Expr. Purif.* **2003**, *32*, 28−34.
(9) Murasugi, A.; Tohma-Aiba, Y. Production of native recombinant human midkine in the yeast, *Pichia pastoris*. *Protein Expr. Purif.* **2003**, *27*, 244−252.
(10) Gurkan, C.; Ellar, D. J. Expression in *Pichia pastoris* and purification of a membrane-acting immunotoxin based on a synthetic gene coding for the Bacillus thuringiensis Cyt2Aa1 toxin. *Protein Expr. Purif.* **2003**, *29*, 103−116.
(11) Kreusch, A.; Lesley, S. A. High-throughput cloning, expression and purification technologies, In *Genomics, Proteomics and Vaccines*; Grandi, G., Ed.; John Wiley & Sons: Chichester, 2004; pp 171−182.
(12) Anicetti, V. R.; Fehskens, E. F.; Reed, B. R.; Chen, A. B.; Moore, P.; Geier, M. D.; Jones, A. J. Immunoassay for the detection of *E. coli* proteins in recombinant DNA derived human growth hormone. *J. Immunol. Methods* **1986**, *91*, 213−224.
(13) Merrick, H.; Hawlitschek, G. A complete system for quantitative analysis of total DNA, protein impurities and relevant proteins. *Biotech. Forum E.U.* **1992**, *6*, 398−403.
(14) Chen, A. B.; Championsmith, A. A.; Blanchard, J.; Gorrell, J.; Niepelt, B. A.; Federici, M. M.; Formento, J.;, Sinicropi, D. V. Quantitation of *E. coli* protein impurities in recombinant human interferon-gamma. *Appl. Biochem. Biotechnol.* **1992**, *36*, 137−152.
(15) Eaton, L. C. Host cell contaminant protein assay development for recombinant biopharmaceuticals. *J. Chromatogr.* **1995**, *705*, 105−114.
(16) Dagoussat, N.; Haeuw, J. F.; Robillard, V.; Damien, F.; Libon, C.; Corvaia, N.; Lawny, F.; Nguyen, T. N.; Bonnefoy, J. Y.; Beck, A. Development of a quantitative assay for residual host cell proteins in a recombinant subunit vaccine against human respiratory syncytial virus. *J. Immunol. Methods* **2001**, *251*, 151−159.
(17) Wan, M.; Wang, Y.; Rabideau, S.; Moreadith, R.; Schrimsher, J.; Conn, G. An enzyme-linked immunosorbent assay for host cell protein contaminants in recombinant PEGylated staphylokinase mutant SY161. *J. Pharm. Biomed. Anal.* **2002**, *28*, 953−963.
(18) Phillips, M.; Cormier, J.; Ferrence, J.; Dowd, C.; Kiss, R.; Lutz, H.; Carter, J. Performance of a membrane adsorber for trace impurity removal in biotechnology manufacturing. *J. Chromatogr., A* **2005**, *1078*, 74−82.
(19) Laird, M. W.; Sampey, G. C.; Johnson, K.; Zukauskas, D.; Pierre, J.; Hong, J. S.; Cooksey, B. A.; Li, Y.; Galperina, O.; Karwoski, J. D.; Burke, R. N. Optimization of BLyS production and purification from *Escherichia coli*. *Protein Expr. Purif.* **2005**, *39*, 237−246.
(20) Miles, A. P.; Zhu, D.; Saul, A. Determining residual host cell antigen levels in purified recombinant proteins by slot blot and scanning laser densitometry. *Methods Mol. Biol.* **2005**, *308*, 233−242.
(21) Zhu, D.; Saul, A. J.; Miles, A. P. A quantitative slot blot assay for host cell protein impurities in recombinant proteins expressed in *E. coli*. *J. Immunol. Methods* **2005**, *306*, 40−50.
(22) Krawitz, D. C.; Forrest, W.; Moreno, G. T.; Kittleson, J.; Champion, K. M. Proteomic studies support the use of multi-product immunoassays to monitor host cell protein impurities. *Proteomics* **2006**, *6*, 94−110.
(23) Yigzaw, Y.; Piper, R.; Tran, M.; Shukla, A. A. Exploitation of the Adsorptive Properties of Depth Filters for Host Cell Protein Removal during Monoclonal Antibody Purification. *Biotechnol. Prog.* **2006**, *22*, 288−296.
(24) Thulasiraman, V.; Lin, S.; Gheorghiu, L.; Lathrop, J.; Lomas, L.; Hammond, D.; Boschetti, E., Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands. *Electrophoresis* **2005**, *26*, 3561−3571.
(25) Righetti, P. G.; Castagna, A.; Antonioli, P.; Boschetti, E., Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* **2005**, *26*, 297−319.
(26) Righetti, P. G.; Castagna, A.; Antonucci, F.; Piubelli, C.; Cecconi, D.; Campostrini, N.; Rustichelli, C.; Antonioli, P.; Zanusso, G.; Monaco, S.; Lomas, L.; Boschetti, E. Proteome analysis in the clinical chemistry laboratory: myth or reality? *Clin. Chim. Acta* **2005**, *357*, 123−139.
(27) Castagna, A.; Cecconi, D.; Sennels, L.; Rappsilber, J.; Guerrier, L.; Fortis, F.; Boschetti, E.; Lomas, L.; Righetti, P. G. Exploring the hidden human urinary proteome via ligand library beads. *J. Proteome Res.* **2005**, *4*, 1917−1930.
(28) Righetti, P. G.; Boschetti, E.; Lomas L.; Citterio, A. Protein Equalizer Technology: the quest for a “democratic proteome”. *Proteomics* **2006**, *6*, 3980−3992.
(29) Lathrop, J. T.; Hayes, T. K.; Carrick, K.; Hammond D. Rarity gives a charm: evaluation of trace proteins in plasma and serum. *Expert Rev. Proteomics* **2005**, *2*, 393−406.
(30) Sennels, L.; Salek, M.; Lomas, L.; Boschetti, E.; Righetti, P. G.; Rappsilber, J. Proteomic analysis of human blood serum using peptide library beads. *Mol. Cell. Proteomics* **2006**, submitted.

(31) Buettner, J. A.; Dadd, C. A.; Baumbach, G. A.; Masecar, B. L.; Hammond, D. J. Chemically derived peptide libraries: a new resin and methodology for lead identification. *Intl. J. Pept. Protein Res.* **1996**, *47*, 70−83.
(32) Hammond, D. J.; Lathrop, J. T.; US Patent Application 20030-211471: "Method for Detecting Ligands and Targets in a Mixture"; November 13, 2003.
(33) http://fields.scripps.edu/sequest/
(34) Janson, J. C. In *Protein Purification*; Janson, J. C., Rydén, L., Eds.; Wiley-VCH: New York, 1998; pp 3−40.
(35) Simpson, R. J. In *Purifying Proteins For Proteomics*; Simpson, R. J., Ed.; Cold Spring Harbour Laboratory Press: Spring Harbour, 2004; pp 17−40.

PR060090S

Anal. Chem. **2007,** *79,* 821−827

Accelerated Articles

SDS-PAGE under Focusing Conditions: An Electrokinetic Transport Phenomenon Based on Charge Neutralization

Gleb Zilberstein,† Leonid Korol,† Paolo Antonioli,‡ Pier Giorgio Righetti,‡ and Shmuel Bukshpan*,†

Cleardirection Ltd., 4 Pekeris Street, Rehovot 76702, Israel, and Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Milano 20131, Italy

A novel method is reported for mass separation of proteins, based on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Unlike conventional SDS-PAGE, in which separation by mass of SDS-laden polypeptide chains is obtained in constant concentration or porosity gradient gels, the present method, called "SDS-PAGE focusing", exploits a "steady-state" process by which the SDS−protein micelles are driven to stationary zones along the migration path against a gradient of positive charges affixed to the neutral polyacrylamide matrix. As the total negative surface charge of such complexes matches the surrounding charge density of the matrix, the SDS−protein complex stops migrating and remains stationary, as typical of steady-state separation techniques. As a result of this mechanism, the proteins are separated in an unorthodox way, with the smaller proteins/peptides staying closer to the application point and larger proteins migrating further down toward the anodic gel end. This results in a positive slope of the M_r vs migration plot, vs a negative slope in conventional SDS-PAGE. Moreover, such a plot is linear (by design), whereas in standard SDS-PAGE it is semi- or even double logarithmic. Particularly advantageous appears the ability of the present method to fine-tune the separation of small-size fragments and tryptic digests, where conventional SDS-PAGE usually fails. Additionally, by exploiting constant plateaus of charges, rather than gradients, it is possible to amplify the separation between species having closely spaced M_r values, down to a limit of ∼150 Da. This increases the resolution by at least 1 order of magnitude as compared with standard SDS-PAGE, where for a proper separation of two adjacent species, an M_r increment of ∼3000 Da is needed.

In the 1960s, two major events drastically boosted the performance of electrokinetic methodologies. The first one (1960−1964) was the introduction of isoelectric focusing (IEF) via the theoretical work of Svensson[1] and the synthesis of carrier ampholytes by Vesterberg.[2] In IEF, proteins are captured along a particular point of their titration curve, where the local pH equals their isoelectric point (pI) value. By this method, proteins could be separated solely on the basis of their surface charge. The second one was the discovery that polypeptide chains, when saturated with micelles of an anionic surfactant (notably sodium dodecyl sulfate, SDS) would reach a nearly constant charge to mass ratio, thus allowing their separation, in a sieving matrix (by polyacrylamide gel electrophoresis, PAGE), mostly according to their molecular mass.[3] The figure of ∼1.4 g of SDS bound per gram of protein is often quoted as a typical stoichiometric value.[4] This means that the number of SDS molecules bound is of the order of half the number of amino acid residues in the polypeptide chain. This amount of highly charged surfactant molecules is sufficient to overwhelm effectively the intrinsic charges on the polypeptide chain so that the net charge per unit of mass becomes approximately constant. Electrophoretic migration is then proportional to the effective molecular radius or approximately to the molecular mass of the polypeptide chain. Although it is an oversimplification to which a number of exceptions are to be found,[5,6] the relationship does in fact hold true for a very large number of proteins. Both techniques represented a quantum jump over conventional zone electrophoresis, including disk electrophoresis,[7,8] in that they not only brought about a remarkable resolving power but also permitted for the first time assessment of some important physicochemical parameters of polypeptide chains, such as their pI value (with a precision of ∼0.02 pH unit) and their relative molecular mass (with a lower precision, of the order of ±3−5 kDa). Zone electrophoresis, on the contrary, in which contribution to migration is due to combination of protein

* To whom correspondence should be addressed. E-mail: shmuel.b@auriga.co.il.
† Cleardirection Ltd.
‡ Politecnico di Milano.

(1) Svensson, H. *Acta Chem. Scand.* **1961**, *15*, 325−341. Svensson, H. *Acta Chem. Scand.* **1962**, *16*, 456−466.
(2) Vesterberg, O. *Acta Chem. Scand.* **1969**, *23*, 2653−2666.
(3) Shapiro, A. L.; Vinuela, E.; Maizel, J. V. *Biochem. Biophys. Res. Commun.* **1967**, *28*, 815−829.
(4) Pitt-Rivers, R.; Impiombato, F. S. A. *Biochem. J.* **1968**, *109*, 825−833.
(5) Weber, K.; Osborn, M. *J. Biol. Chem.* **1969**, *244*, 4406−4412.
(6) Dunker, A. K.; Rueckert, R. R. *J. Biol. Chem.* **1969**, *244*, 5074−5080.
(7) Davis, B. J. *Ann. N. Y. Acad. Sci.* **1964**, *121*, 404−427.
(8) Ornstein, L. *Ann. N. Y. Acad. Sci.* **1964**, *121*, 321−349.

10.1021/ac0615091 CCC: $37.00
Published on Web 12/15/2006

charge, size, and shape, does not easily allow separating the different tributes. The only way of obtaining some discrimination is to resort to Ferguson plots,[9] which however require a very cumbersome experimental setup.

Just like IEF, also SDS-PAGE was refined over the years, to become perhaps the most popular method today for fast determination of M_r values of proteins and for assessing their purity during a purification protocol (and, of course, for use as a second dimension in 2D maps). Its resolving power was substantially augmented by introducing discontinuous buffers systems, such as those of Laemmli[10] (the conventional glycine-HCl, terminating/leading ion boundary) and Neville[11] (a boric acid/sulfate boundary), permitting zone sharpening and stacking into very thin zones. When SDS-PAGE is run in constant-concentration gels (T of 7.5%) the range of linearity in M_r assessment (by the standard semilog plot of log M_r vs R_f) spans typically the M_r 15 000−70 000 Da interval, with changes of slope occurring both below and above these values. In order to greatly expand this M_r interval, acrylamide concentration gradients were introduced.[12−14] Gradient gels enabled proteins of widely different sizes, which would otherwise require runs at two or more different values of T, to be separated in a single gel, while simultaneously offering superior resolution and sharper bands. The method was thus widely applicable for M_r measurements over the range 13 000−1 000 000 Da by using a 3−30% gel gradient.[13] Other improvements were the introduction of thin or ultrathin gel matrixes (in the 200−500-μm gel thickness)[15] permitting application of higher voltage gradients with shorter analysis times and staining/destaining protocols at high temperatures, allowing band visualization within minutes.[16]

Also, the models for the structure of complexes between proteins and SDS have been refined over the years. One of the earliest models proposed on the basis of hydrodynamic measurements is the so-called "rodlike particle model", in which it is hypothesized that, upon binding of SDS, the polypeptide would form rodlike structures, about 3.6 nm in diameter and 0.074 nm/amino acid residue in length.[17] In a more recent reinterpretation, the particle was described as a "short, rigid, rodlike segment, with intervening regions possessing some flexibility".[18] In contrast with the rodlike particle above, a "necklace model" was also proposed. In this second model, "the polymer chain is flexible and micelle-like clusters of SDS are scattered along the chain".[19,20] In an improved version, SDS binds to the protein in the form of spherical micelles and the polypeptide forms α-helices mostly in the hydrophobic region of the micelles. A flexible "α-helix/random coil" structure has also been suggested by Mattice et al.,[21] on the basis of circular dichroism changes of proteins upon SDS binding. Finally, a flexible helix model, in which the polypeptide chain is helically coiled around an SDS micelle, attached by hydrogen bonds between sulfate group oxygens and peptide bond nitrogens was proposed by Lundahl et al.[22] Over the years, Lundahl's group has refined this last model,[23−26] which has now become the "protein-decorated, micelle model". Based on small-angle neutron scattering data, this model proposes that adjacent, protein-decorated, spherical micelles are formed, linked with short polypeptide segments. For a protein of average size (50 000 Da), it is believed that this complex could be composed of three protein-decorated SDS micelles, not necessarily of equal size.

Over the past decade, developments on SDS-PAGE have reached a plateau and standard protocols have been adopted in most laboratories around the world. These protocols are based on the common feature of the direct stochiometry in the SDS−protein complex leading to a constant charge to mass ratio.

This property provides the basis of the new approach to SDS-PAGE here presented, called "SDS-PAGE focusing". It consists of performing the SDS run in the presence of a gradient of positive charges grafted onto the polyacrylamide gel. As the various protein−SDS micelles reach a gel zone in which the surrounding charge density matches the total amount of negative charges on the surface of the SDS−polypeptide chain complex, the proteins are arrested by a charge neutralization mechanism that resembles a focusing process, in that zero net charge conditions are reached, impeding further migration of the polypeptide chains. Some unique advantages of this novel technique will be presented and evaluated. A detailed theoretical model for the method will be published elsewhere.

EXPERIMENTAL SECTION

Chemicals and Materials. Basic Immobiline, with various pK values (8.5, 9.3, 10.3, 12), and the peptide marker kit (2.5−16.9 kDa, Catalog No. 80-1129-83) were purchased from Amersham Pharmacia Biotech (Uppsala, Sweden). The pure proteins insulin (5.7 kDa), aprotinin (6.5 kDa), cytochrome (12.3 kDa), avidin (16 kDa), myoglobin (17.0 kDa), β-lactoglobulin (18.4 kDa), and bovine serum albumin (BSA; 69.3 kDa) were from Sigma-Aldrich Chemie Gmbh (Steinheim, Germany). Fluka (Buchs, Switzerland) provided phosphoric acid, sodium hydroxide, acetic acid Coomassie Brilliant Blue R, and Tris. Acrylamide/bisacrylamide solution (Catalog No. 161-0156) was from Bio-Rad (Hercules, CA).

Conventional SDS-PAGE. Electrophoresis of marker proteins was performed in classical conditions using 15-well precast 15% polyacrylamide gel plates. Samples of appropriate protein concentration were diluted 2-fold in Laemmli sample buffer. Three microliters of diluted protein and peptide marker solutions was loaded per lane, and electrophoresis migration was performed at 200 V for 35 min. Colloidal Coomassie Brilliant Blue R staining and destaining was achieved using the method described by the supplier of reagents (Invitrogen Corp.). The migration plates were

(9) Ferguson, K. A. *Metabolism* **1964**, *13*, 985−995.
(10) Laemmli, U. K. *Nature* **1970**, *227*, 680−685.
(11) Neville, D. M. *J. Biol. Chem.* **1971**, *246*, 6328−6334.
(12) Studier, F. W. *J. Mol. Biol.* **1973**, *79*, 237−242.
(13) Lambin, P. *Anal. Biochem.* **1978**, *85*, 114−125.
(14) Poduslo, J. F.; Rodbard, D. *Anal. Biochem.* **1980**, *101*, 394−406.
(15) Radola, B. J. *Electrophoresis* **1980**, *1*, 43−56.
(16) Wong, C.; Sridhara, S.; Bardwell, J. C. A.; Jakob, U. *BioTechniques* **2000**, *28*, 426−432.
(17) Reynolds, J. A.; Tanford, C. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *66*, 1002−1007.
(18) Tanford, C. *The Hydrophobic Effect. Formation of Micelles and Biological Membranes,* 2nd ed.; J. Wiley & Sons: New York, 1980; pp 159−164.
(19) Shirahama, K.; Tsujii, K.; Takagi, T. *J. Biochem.* **1974**, *75*, 309−319.
(20) Takagi, T.; Tsujii, K.; Shirahama, K. *J. Biochem.* **1975**, *77*, 939−947.
(21) Mattice, W. L.; Riser, J. M.; Ckark, D. S. *Biochemistry* **1976**, *15*, 4264−4272.
(22) Lundahl, P.; Greijer, E.; Sandberg, M.; Cardell, S.; Eriksson, K. O. *Biochim. Biophys. Acta* **1986**, *873*, 20−26.
(23) Mascher, E.; Lundahl, P. *J. Chromatogr.* **1989**, *476*, 147−158.
(24) Wallstén, M.; Lundahl, P. *J. Chromatogr.* **1990**, *512*, 3−12.
(25) Lundahl, P.; Watanabe, Y.; Takagi, T. *J. Chromatogr.* **1992**, *604*, 95−102.
(26) Ibel, K.; May, R. P.; Sandberg, M.; Mascher, E.; Greijer, E.; Lundahl, P. *Biophys. Chem.* **1994**, *53*, 77−84.

then stored dry after image acquisition with the Versa Doc system (Bio-Rad).

SDS-PAGE under Focusing Conditions. With the novel method reported here, gels are cast with a two-vessel gradient mixer, as routinely performed for pouring immobilized pH gradient (IPG) gels.[27] Mini gels (8 × 7 cm, 0.75 mm thick) are cast in the Bio-Rad system either singly or in a six-unit casting chamber. All gels are supported by a Mylar film (Gel Bond) for easy handling when opening the cassette. The typical gel formulations consist of 4% *T* polyacrylamide (3.3% cross-linker) in the presence of a gradient of positively charged ions, i.e., basic Immobilines. We have tried different basic species, but the preferred gel formulation adopts the p*K* 10.3 Immobiline, since this species will be fully protonated at the standard running pH of SDS-PAGE (pH 8.3 in 100 mM Tris−Tricine buffer and 0.1% SDS). Different slopes of basic Immobiline gradients have been tested, in the intervals 0−8, 0−12, and 0−18 mM. Immediately after casting, the gels are placed in a oven at 50 °C and polymerization is continued at this temperature for 1 h. This step is very important, since it ensures good conversion of monomers into the polymeric form and incorporation of charged Immobilines to ∼90% conversion.[28] For running, we have tried both vertical and horizontal systems. The preferred configuration has been a horizontal setup in the Multiphor II chamber (Amersham-GE). The cathodic electrolyte was a 0.1 M Tris−Tricine solution (pH 8.3) containing 0.1% SDS, while at the anode 0.1 M Tris-acetate (pH 6.4) with 0.1% SDS was used. Contact between electrolyte solutions and gel was ensured by filter paper. Gels have been run for various periods of time, from 45 min up to 24 h, in order to test for steady-state "focusing" conditions. In some cases, the SDS-PAGE system has been run not in a presence of a gradient of basic Immobiline but by adopting a plateau (constant concentration) level. Typical running conditions: 300 V with forced cooling at 10 °C.

Staining and Destaining. Gels are stained with Coomassie Blue R, as in conventional SDS-PAGE. However, since the polyacrylamide matrix contains a gradient of positive charges, which will create a gradient of intense blue background color, for destaining, we incubate the gels in a microwave oven for ∼2 min with repeated changes of destaining solution until a clear background is achieved.

RESULTS

Figure 1A shows the separation of a set of M_r marker proteins, in the 2.5−17-kDa range, by SDS-PAGE focusing, while Figure 1B gives the corresponding plot for M_r calculation according to relative migration distances. Some unique features of this novel system are apparent: first, and contrary to current dogma in SDS-PAGE, small polypeptides are the ones that exhibit the lowest migration distance, with the larger polypeptides traveling much further down the migration path. As a result (Figure 1B), the M_r plot shows a positive slope, whereas in conventional SDS-PAGE, this plot has a negative slope. Additionally, the plot is linear on axes that also contain a linear scale. Conversely, in standard SDS-PAGE, the plot for M_r assessment is semilog, since typically the M_r axis is on a log scale (for gels of constant porosity). For gels containing porosity gradients, moreover, the M_r plot is performed on a double-log scale (usually log M_r vs log % *T*).

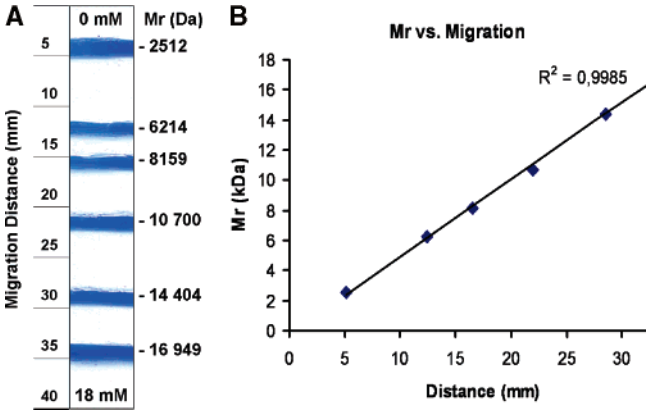


Figure 1. SDS-PAGE focusing. (A) Separation of SDS−polypeptide chains in the M_r 2.5−18 kDa. 4% *T* polyacrylamide gel with a linear gradient of 0−18 mM p*K* 10.3 Immobiline. Buffer: 112 mM Tris-acetate, pH 6.4, containing 0.1% SDS. Running conditions: 300 V for 4 h. (B) Plot of M_r value vs migration distance for M_r determination. Note the positive slope of this plot.

Since the present system is dubbed SDS-PAGE focusing, one has to prove that steady-state conditions are indeed reached and maintained over time, as typical of true IEF, especially in its IPG version. A time course was thus performed, in order to prove this assumption, and it is shown in Figure 2. As is visible in the different time panels, steady-state conditions are reached already after 6 h of migration, although, for the larger peptides, there is still an asymptotic migration toward the anodic end. The pattern obtained at 10 and 18 h, though, is identical to that after 24 h, suggesting that true steady-state conditions had been reached. Here too it can be appreciated that it is the larger proteins that migrate much further down the electrophoretic path. The focusing effect can also be appreciated by the shape of the bands: although the samples are loaded on rectangular pieces of filter paper on the gel surface, conditions leading to band spreading and decay in zone electrophoresis, here the sample zones are seen to become considerably thinner at their steady-state position.

An extra bonus of SDS-PAGE focusing can be appreciated in Figure 3. It is well known that, in conventional SDS-PAGE, proteins smaller than 10 kDa and thus all peptides in the M_r 1−6-kDa range are lost during analysis, since they migrate in the glycine-HCl boundary in the Laemmli system.[29] In order to implement separation of small peptides, one has to resort to relatively high polyacrylamide concentrations (up to 20% *T*) while employing a buffer with pH of 8.3 (the p*K* value of the basic group of Tricine) coupled with a high molarity of Tris (the so-called Tris−Tricine system of Schägger and von Jagow).[30] When comparing conventional versus "focusing" SDS-PAGE (Figure 3) for the analysis of small-size peptides, it is seen that the conventional gel (right track) is unable to separate all smaller peptides, which are on the contrary well visible in the focusing gel, although the latter is cast in a bare 4% *T* matrix versus a 12% *T* matrix in the conventional system. One should also appreciate that, in order to compare the two gel systems, they have to be aligned with inverse polarity, due to the "unorthodox" migration of polypeptides in the focusing gel.

(27) Righetti, P. G. I*mmobilized pH Gradients: Theory and Methodology;* Elsevier: Amsterdam, 1990; pp 127−139.
(28) Righetti, P. G.; Ek, K.; Bjellqvist, B. *J. Chromatogr.* **1984**, *291*, 31−42.
(29) Williams, J. G.; Gratzer, W. B. *J. Chromatogr.* **1971**, *57*, 121−130.

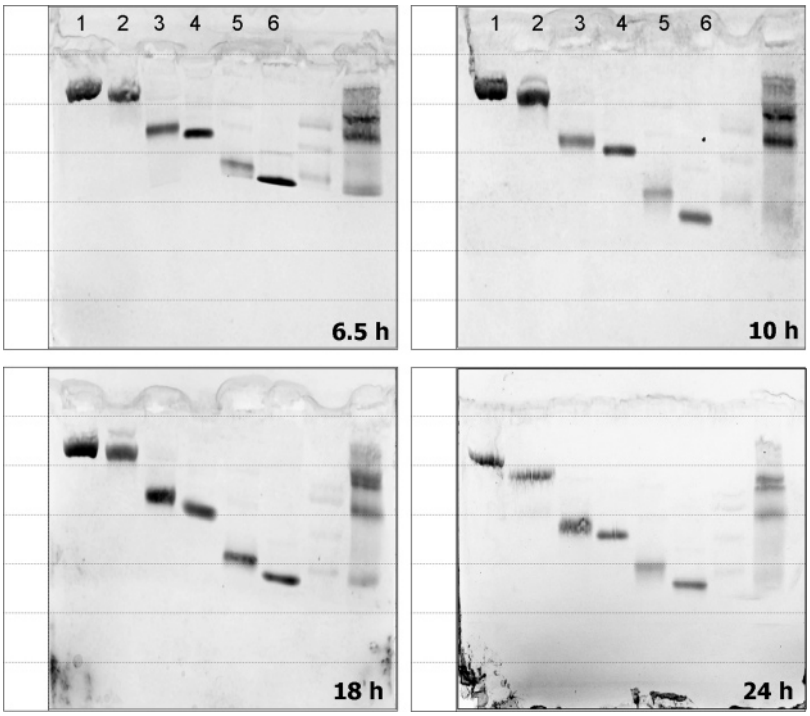


Figure 2. Time course of SDS-PAGE focusing. 4% *T* polyacrylamide gels with a linear gradient of 0−18 mM p*K* 10.3 Immobiline. Buffer: 112 mM Tris-acetate, pH 6.4, containing 0.1% SDS. Running conditions: 300 V for 6 (A), 10 (B), 18 (C), and 24 (D) h. Gel slabs run horizontally on a Multiphor II chamber with surface sample application on filter paper strips. Staining with micellar Coomassie Brilliant Blue R. Protein bands: (1) insulin, 5.7 kDa; (2) aprotinin, 6.5 kDa; (3) cytochrome, 12.3 kDa; (4) avidin, 16.0 kDa; (5) myoglobin, 17.0 kDa; (6) β-lactoglobulin 18.4 kDa.

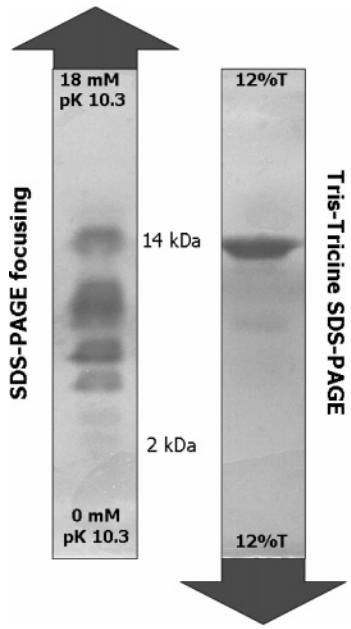


Figure 3. Comparison of the peptide separation ability of conventional SDS-PAGE (right strip) vs SDS-PAGE focusing (left strip). Gels and running conditions as in Figure 2. Note that, for proper gel alignment, the gels had to be juxtaposed with inverted polarity. The bands represent tryptic digest of myoglobin.

In terms of peptide analysis, the unique performance of the present system can also be evaluated from Figure 4. Here a gel with a strong discontinuity in porosity was cast with an upper 15% *T* loading zone and a lower 4% *T* (with a gradient of basic Immobiline) running gel. The wells were loaded with native BSA (lane 1), with a mixture of native BSA and its tryptic fragments (lane 2), and with only tryptic fragments of BSA (lane 3). According to present knowledge on gel porosity,[31] BSA should be trapped in the loading zone, whereas the tryptic peptides should migrate down in the SDS-PAGE gel and focus at appropriate distances along the migration path. As shown in Figure 4, this assumption is fully verified in practice, since no bands are seen migrating in lane 1, whereas in lanes 2 and 3, a large number of tryptic peptides are seen "focused" along the migration path (the two marked black lines in the 15% *T* stacking gel in lanes 1 and 2 representing trapped native BSA).

Another unique aspect of our SDS-PAGE focusing method can also be appreciated in Figure 5. It is well known that, in conventional SDS-PAGE, two neighboring bands cannot be resolved unless their difference in molecular mass is of the order of at least 3000 Da.[32] Thus, for closely related proteins, envelops of bands of slightly different M_r values are usually experienced along the anodic migration path, this leading to spot overlap in two-dimensional map analysis.[33,34] So far, no remedies to this drawback have been found. The example of Figure 5 shows how this situation could be dramatically changed: here two colored marker proteins, β-lactoglobulin and myoglobin (ΔM_r of 1500 Da)

(30) Schägger, H.; von Jagow, G. *Anal. Biochem.* **1987**, *166*, 368−379.
(31) Fawcett, J. S.; Morris, C. J. O. R. *Sep. Stud.* **1966**, *1*, 9−16.
(32) Andrews, A. T. *Electrophoresis: Theory, Techniques, Biochemical and Clinical Applications*; Clarendon Press: Oxford, 1986; pp 119−136.
(33) Pietrogrande, M. C.; Marchetti, N.; Dondi, F.; Righetti, P. G. *Electrophoresis* **2002**, *23*, 283−291.
(34) Pietrogrande, M. C.; Marchetti, N.; Dondi, F.; Righetti, P. G. *Electrophoresis* **2003**, *24*, 217−224.

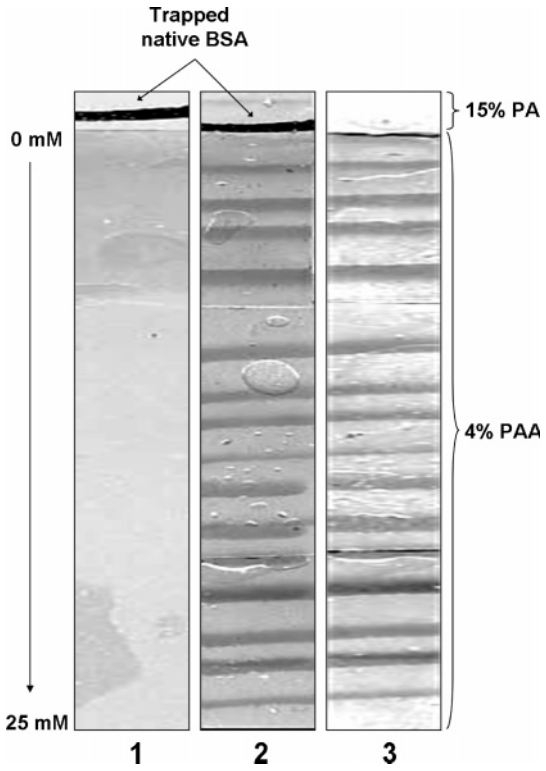


Figure 4. Peptidome analysis. The upper layer of 15% PAG (15-mm length, indicated by the arrows) served as the trapping zone for the high molecular mass proteins; the underlaying 4% PAG charge gradient gel (0−25 mM Immobiline, 77-mm length) served for the separation of peptide fragments. Lane 1, native BSA; lane 2:, mixture of native BSA and its tryptic fragments; lane 3, tryptic fragments of BSA alone. The two arrows in lanes 1 and 2 and the thick black zones indicate trapped native BSA in the 15% *T* PAG loading layer.

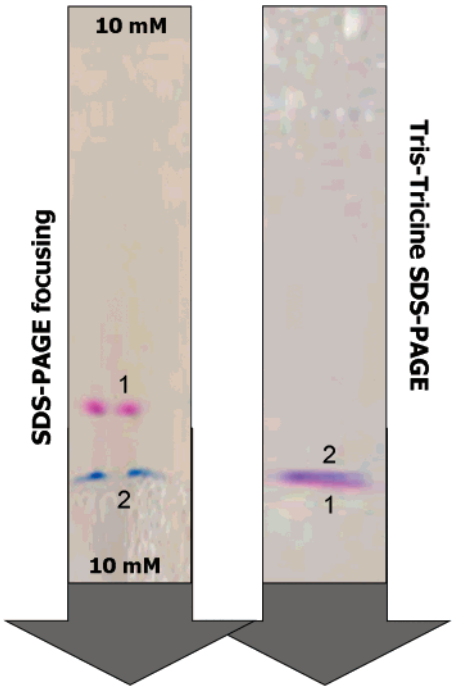


Figure 5. Separation of a mixture of myoglobin and β-lactoglobulin in conventional SDS-PAGE (right strip) vs SDS-PAGE focusing (left strip). 4% *T* polyacrylamide gel with a 10 mM constant plateau of p*K* 10.3 Immobiline (7-cm length). The M_r markers are prestained. Red fraction, myoglobin; blue zone, β-lactoglobulin.

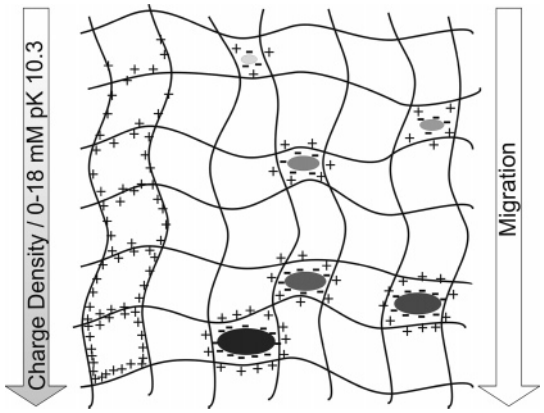


Figure 6. Pictogram of the focusing of protein−SDS complexes along a polyacrylamide gel containing a gradient of positive charges. Small proteins/peptides are focused in gel regions of low charge density (i.e., close to the application point), whereas larger proteins migrate progressively down the gel length toward regions of higher positive charge densities until reaching an "isoelectric point" due to balancing of positive gel charges with the negative charges of the various SDS−protein complexes.

have been separated in our focusing system (left strip) as compared to a conventional SDS-PAGE gel (right strip). In the latter case, the two proteins migrate in a single zone, although there is a slight hint at a separation, since the upper zone boundary is seen in bluish and the lower boundary somewhat reddish. Conversely, in the focusing system, the two proteins are seen separated by 1-cm distance. Thus, it would appear that the present system has the potential of separating zones with a ΔM_r of barely 150 Da, assuming a 1-mm free gel space between two adjacent zones to be an acceptable resolving power. This unique resolving ability was obtained by running our focusing system not in the presence of a charge gradient, as customarily run, but in the presence of a plateau of charges (a gel section with constant concentration of 10 mM p*K* 10.3 Immobiline), this leading to differential charge neutralization along the migration path.

DISCUSSION

The novel method of SDS-PAGE focusing presented here offers some unique features worth discussing. First, it is of interest to speculate on the mechanism of separation. The pictogram of Figure 6 gives an idealized vision of such a mechanism. Along the migration path, from cathode to anode, the polyacrylamide matrix exhibits an increasing concentration of positive charges, grafted onto the neutral matrix backbone, much like IPG matrixes do, except that in the latter case the negatively charged, titrating counterion is grafted as well. As the negatively charged SDS−polypeptide chain complex migrates down the electrophoretic path, it encounters a linearly increasing density of positive charges. It is tempting to speculate that, once a polypeptide chain of a given size meets along the path a density of positive charges matching its negative surface charges, there will be charge neutralization and the macromolecule will come to a stop. This explains the odd migration behavior, by which, in a sieving matrix, the smaller polypeptides are the ones that are first arrested, whereas the larger peptides/proteins keep migrating toward the anode in regions of higher positive charge densities. Clearly the smaller size polypeptides have a lower total surface charge as compared to larger

polypeptide chains; thus their migration is arrested, via this mechanism of charge neutralization, shortly after exiting from the application pocket. At the charge matching point, it is quite likely that the polypeptide chains, especially the larger ones, will distort locally the polyacrylamide fibers, attracting them to their surface, so that charge neutralization will occur efficiently. In fact, ion-to-ion interactions are among the strongest ones in noncovalent links among macromolecules.[35] Given the fact that we use rather dilute matrixes (down to only 4% *T*) endowed with larger pore sizes, such local fiber distortion has to occur, since the average pore size will be of the order of a few nanometers. Additionally, gels are considered as viscoelastic media, fluctuating in the surrounding space,[36] and this is especially true for dilute matrixes, thus such local fiber distortion is a most likely event. The fact that our system works on a focusing principle is also implicit on the fact that our polyacrylamide matrixes, being highly diluted, are minimally sieving; thus. they could not possibly arrest the migration of small polypeptides via an approach to a pore limit, as customary in porosity gradient gels.[37,38] Interestingly, our novel method could permit estimation of the total surface charge of a given complex SDS−protein. For example, in the case of Figure 2, when measuring the distance migrated into the gel by, for example, β-lactoglobulin at steady state (18 or 24 h runs against a linear gradient of 0−18 mM p*K* 10.3 Immobiline), by assuming a 90% conversion efficiency of the basic Immobiline into the polyacrylamide gel fibers, full protonation of all positive charges (given a p*K* of 10.3 and a running buffer pH of 6.4), and minimal p*K* variation of the free and bound Immobiline (as verified in practice by Bjellqvist et al.[39]), from its steady-state position, one could calculate that the total negative surface charge molarity of the SDS−β-lactoglobulin complex is 9 mM, matching the molarity of Immobiline in the surrounding gel fibers. By the same token, for the smallest polypeptide run in Figure 2 (insulin), one can calculate a negative surface charge molarity of its SDS complex of 3 mM. To our knowledge, conventional SDS-PAGE could not give this type of information on the migrating protein−SDS complexes. This information can be used to evaluate the total charge of the SDS−protein complex on the assumption that the p*K* values of SDS and Immobilines are known.

Our hypothesis of migration arrest (and thus "focusing") due to charge neutralization between the positively charged polyacrylamide matrix and the negatively charged SDS−protein micelles is further corroborated by the following findings: (a) first, the same principle applies to DNA separations, here too with excellent resolution (e.g., baseline resolution of two DNA fragments differing by only 3 bp in the 700-bp region; manuscript in preparation); (b) second, when the PAG matrix is cast with a gradient of negative charges (by incorporating acid Immobilines), no separation or focusing effects of SDS−protein micelles or DNA can be obtained.

It is also of interest to review here briefly the pathway leading to this novel way of performing a focusing experiment. IEF was born via the theoretical work of Svensson-Rilbe,[1] which dissipated the wrong notion by Kolin[40] of focusing in nonamphoteric buffers. For proper IEF, the buffers had to be amphoteric, so that they would reach a steady-state position along the migration path, and "carrier" (of buffering power and conductivity at pH = pI).[41] The second generation of focusing went against the dogma of Svensson-Rilbe, in that it adopted buffers comprising a series of weak acids and weak bases, all monoprotic, but certainly nonamphoteric. The violation of Svensson-Rilbe's law was only apparent, though: when these buffers were incorporated into the polyacrylamide matrix, and thus insolubilized, each infinitesimal gel layer perpendicular to the migration path was indeed amphoteric, isoelectric, and endowed with strong buffering power. Our present method represents the third way of performing focusing, and thus, perhaps, it closes the circle. In our system, we move from fully soluble buffers (conventional IEF) and fully insoluble species (IPG) to a hybrid system, in which half of the charges are insolubilized (the basic Immobilines) into the polyacrylamide matrix, whereas the other half (the counterions) are soluble species. In the amphoteric, "isoelectric" complex, thus, a hybrid system, partly soluble, partly insoluble is formed. One might also wonder if, upon prolonged focusing, the SDS−protein mixed micelle could be disrupted. Our experiments clearly prove that this is not the case; in addition, experiments performed not with SDS but with STS (sodium tetradecyl sulfate; not shown) gave identical results, indicating that the hybrid protein−SDS micelles are quite stable under our experimental conditions.

Another interesting aspect of our SDS-PAGE focusing system is that it opens up a window for simple and efficient separation of tryptic digests and peptides in general, as shown in Figures 3 and 4. This can only be achieved with special SDS-PAGE variants, exploiting particular buffer systems and highly dense polyacrylamide gels. From this point of view, it is remarkable that our tryptic digests (see Figure 4) could be sharply separated in a rather dilute polyacrylamide matrix, barely 4% *T*. This has the advantage of allowing easy extraction of the peptides for further analysis, for example, by mass spectrometry. The inability of conventional SDS-PAGE gels to fractionate any protein/peptides below 10 kDa is lamented as one of the main limitations of 2D map analysis.[42] "Peptidomics", a branch of proteomics devoted to peptide fractionation and analysis, is now considered as central in biomarker discovery, since it is believed that a host of markers for diseases, e.g., in sera, could be fragments of normal proteins abnormally processed during the onset of pathological conditions.[43] Up to the present, the best method for separating complex peptide mixtures and tryptic digests of proteins has been RP-HPLC, especially on C_{18} beads.[44]

There are other advantages of the present system worth mentioning. One of them is the fact that the relationship between M_r and migration distance is a linear function (see Figure 1) as opposed to a semilog or even double-log function in conventional

(35) van Holde, K. E.; Johnson, W. C.; Ho, P. S. *Principles of Physical Biochemistry*; Prentice Hall: Upper Saddle River, NJ, 1998; pp 95−99.
(36) Bode, H. J. *Anal. Biochem.* **1977**, *83*, 204−210. Bode, H. J. *Anal. Biochem.* **1977**, *83*, 364−371.
(37) Margolis, J.; Kenrick, K. G. *Anal. Biochem.* **1968**, *25*, 347−354.
(38) Margolis, J. *Lab. Pract.* **1973**, *22*, 107−112.
(39) Bjellqvist, B.; Ek, K.; Righetti, P. G.; Gianazza, E.; Görg, A.; Westermeier, R.; Postel, W. *J. Biochem. Biophys. Methods* **1982**, *6*, 317−339.
(40) Kolin, A. In *Electrofocusing and Isotachophoresis*; Radola, B. J., Graesslin, D., Eds.; de Guyter: Berlin, 1977; pp 3−34.
(41) Righetti, P. G. *Isoelectric Focusing: Theory, Methodology and Applications*; Elsevier: Amsterdam, 1983; pp 15−39.
(42) Hamdan, H.; Righetti, P. G. *Proteomics Today*; Wiley-Interscience: Hoboken, NJ, 2005; pp 341−344.
(43) Srinivas, P. R.; Verma, M.; Zhao, Y.; Srivastava, S. *Clin. Chem.* **2002**, *48*, 1160−1169.
(44) Wolters, D. A.; Washburn, M. P.; Yates, J. R., III. *Anal. Chem.* **2001**, *73*, 5683−5690.

SDS-PAGE. The replacement of the logarithmic scale of conventional electrophoretic systems with the linear scale provides significant advantages, such as improved accuracy of mass determination, and the ability to optimize the separation for different mass regions, including very low and very high molecular mass ranges. Another interesting aspect is the possibility of greatly modulating the migration distance of closely spaced proteins by a method that we call "resolving gel", as illustrated in Figure 5. Here a constant charge plateau was used, instead of a gradient of charges as typical of the present methodology. In a properly selected charge plateau, the difference of migration between two proteins with small differences in M_r values will be magnified as the protein−SDS complexes are progressively titrated along their migration toward the anode. This method is not a steady-state, focusing method, as is customary of the present technique, but represents a case analogue to the "nonequilibrium IEF" of O'Farrell et al.,[45] adopted for focusing in very alkaline regions where the lack of suitable carrier ampholytes impeded reaching a true steady state. The possibility of fine-tuning the present method to the separation of closely related species greatly enhances the flexibility of our technique, since it could, in principle, allow separation of species differing by as little as 150 Da in their M_r values, a resolving power that is at least 1 order of magnitude greater than in conventional SDS-PAGE.

(45) O'Farrell, P. Z.; Goodman, H. M.; O'Farrell, P. H. *Cell* **1977**, *12*, 1133−1141.

In conclusion, although we hope we have demonstrated the capability and flexibility of the present methodology, we believe further important potential applications can be realized and will be published in the near future. It is anticipated that preparative aspects will be just as important, by adopting a segmented setup in which conventional gel layers will be separated by charged polyacrylamide zones, in a stepwise fashion, able to capture selected M_r populations (work in progress). And this without mentioning the applications of this novel technique to DNA separations, where results just as important are presently being obtained (manuscript in preparation).

ACKNOWLEDGMENT

PGR is supported by grants from MURST (PRIN 2006) and by Fondazione Cariplo.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was released ASAP on December 15, 2006 with minor text errors throughout the paper. The correct version was posted on December 19, 2006 with the word compensation changed to neutralization.

Received for review August 14, 2006. Accepted November 17, 2006.

AC0615091

Ringraziamenti

Un ringraziamento speciale va innanzitutto al Prof. Pier Giorgio Righetti, che in tutti questi anni non ha mai mancato di darmi fiducia e completo appoggio, costituendo per me una preziosa guida.

Desidero ringraziare anche tutte le persone che, con il loro impegno, le loro conoscenze, le loro idee e il loro entusiasmo, mi hanno aiutato a compiere nel miglior modo possibile il mio lavoro. In particolare desidero ricordare la Prof. Claudia Sorlini, il Prof. Giancarlo Ranalli, Pamela Abbruscato, il Prof. Giovanni Vallini, il Prof. Lello Zolla, Egisto Boschetti e tutti i suoi collaboratori, Gleb Zilberstein e Shmuel Bukshpan.

Un ringraziamento particolare va anche a tutti i miei compagni dottorandi e ricercatori, in particolare Annalisa Castagna, le Righette e i Righetti, Giacomo Zapparoli, Gianni Zoccatelli, Sara Frigerio, Silvia Lampis e Irene Chesini, Sara Rinalducci e Gian Maria D'Amici, da cui ho ricevuto preziosi consigli e un impareggiabile supporto morale.

Voglio ringraziare anche il gruppo del Politecnico di Milano, che mi ha accolto con calorosa amicizia e grande disponibilità: il Prof. Attilio Citterio, il Prof. Roberto Sebastiano, Marco Martino, Mavi, Luca Mittino, Mirvana Lauria e tutti gli amici del "Gruppo Citterio".

Ringrazio di cuore anche Martha Mendieta e Carolina Simò, con cui ho condiviso momenti di lavoro e di svago in un clima di serena amicizia.

Grazie anche a tutti gli amici di ora e di sempre per la loro insostituibile e totale amicizia, che è stata per me un solido appoggio e un grande conforto nei momenti difficili della mia vita.

Un ringraziamento speciale va poi a Francesca, che mi ha sempre sostenuto nelle decisioni difficili, dimostrando grande disponibilità e infinita pazienza, e non mi ha mai negato amore e comprensione, anche quando pensavo di poterne fare a meno.

Infine un grazie sincero ai miei genitori, Giuseppe e Rosanna, a mia sorella Angela, a mia zia Sandra e a tutta la mia famiglia, che mi ha aiutato e incoraggiato sempre e incondizionatamente, mi ha dato valori veri in cui credere, ed è stata per me un prezioso esempio di vera umanità.