

Sommario

ABSTRACT	4
INTRODUZIONE.....	8
1.1 La vaccinologia: cenni storici.....	8
1.1.1 <i>microrganismi uccisi o attenuati.....</i>	<i>8</i>
1.1.2 <i>tossine batteriche e proteine coniugate a polisaccaridi.</i>	<i>9</i>
1.2 L'era genomica.....	10
1.2.1 <i>DNA-microarrays</i>	<i>13</i>
1.2.2 <i>La reverse vaccinology</i>	<i>13</i>
1.2.3 <i>La proteomica</i>	<i>15</i>
1.3 Il Batterio.....	19
1.3.1 <i>Generalità</i>	<i>20</i>
1.3.2 <i>Fattori di virulenza.....</i>	<i>21</i>
1.3.3 <i>Patogenesi.....</i>	<i>23</i>
1.3.4 <i>Manifestazioni cliniche</i>	<i>24</i>
1.4 Valutazioni per lo sviluppo di un vaccino.....	25
1.4.1 <i>Vaccini e singoli antigeni in fase di studio:</i>	<i>26</i>
1.5 Approccio Novartis allo sviluppo di un vaccino contro S. pyogenes : Reverse Vaccinology	27
1.5.1 <i>Identificazione di proteine di superficie attraverso analisi computazionale</i>	<i>27</i>
1.5.2 <i>Clonaggio ed espressione degli antigeni selezionati.....</i>	<i>30</i>
1.5.3 <i>Purificazione delle proteine espresse in E.coli</i>	<i>31</i>
1.6 Scopo della tesi.....	34
RISULTATI.....	36
Parte 1	36

2.1 Purificazione delle proteine mediante automazione	36
2.2 Analisi della purezza delle proteine purificate	50
2.3 Test immunologici degli antigeni espressi	51
Parte 2	55
2.4 Approccio proteomico applicato al ceppo GAS M1_3348	55
2.4.1 Comparazione dei ceppi analizzati	58
2.4.2 Efficienza della digestione superficiale per selezionare vaccini	63
Parte 3	64
2.5 Purificazione della proteina 57-His	64
DISCUSSIONE E CONCLUSIONI	69
MATERIALI E METODI	73
4.1 Ceppi di GAS utilizzati	73
4.2 Terreni e tamponi	73
4.2.1 Terreno di coltura LB (Luria Bertani) per la crescita di cellule batteriche:	73
4.2.2 Terreno di coltura THB per la crescita dello <i>S.pyogenes</i>	74
4.2.3 Terreno di coltura THYE per la crescita dello <i>S.pyogenes</i>	74
4.2.4 Tampone PBS (Dulbecco's Phosphate Buffered Saline).....	75
4.3 Clonaggio ed espressione delle proteine ricombinanti dello <i>S. pyogenes</i>	75
4.3.1 Vettori plasmidici.....	75
4.3.2 Ligazione e trasformazione.....	78
4.3.3 Selezione dei cloni positivi.....	78
4.3.4 Espressione	78
4.4 Purificazione in colonna di affinità	79
4.4.1 Proteine His-tag da <i>E. coli</i>	79
4.4.2 Proteine His -Tag da Corpi Inclusi di <i>E. coli</i>	81
4.4.3 Proteine GST-Tag da <i>E. coli</i>	81
4.5 Determinazione della concentrazione proteica	82

4.6 Analisi elettroforetica su gel di poliacrilamide sodio dodecilsolfato (SDS-PAGE)	83
4.7 Analisi d'immagine.....	85
4.8 Bioinformatica	86
4.9 Digestione superficiale dei batteri.....	86
4.9.1 Protocollo di desalificazione	87
4.10 Ricerca in banca dati	87
4.11 Purificazione di proteine mediante l'automazione.....	90
4.12 Purificazione della proteina 57-His.....	95
4.12.1 Clonaggio ed espressione	95
4.12.2 Estrazione e chiarificazione	95
4.12.3 Fast Protein Liquid Chromatography.....	96
4.12.4 Cromatografia di affinità.....	96
4.12.5 Cromatografia a scambio ionico	97
4.12.6 Cromatografia di gel filtrazione	98
4.13 Analisi citofluorimetrica-Fluorescence activated cell sorter- (FACS)	99
 APPENDICE	 101
 Parte 2: Approccio proteomico - breve prefazione da dati ritrovati in letteratura.....	 101
5.1.1 Analisi del ceppo GAS M1_SF370	102
5.1.2 Conferma dell'esposizione delle proteine in superficie.....	105
5.1.3 Confronto dei risultati sperimentali con le predizione di PSORT.....	107
 BIBLIOGRAFIA.....	 110

Abstract

After 200 years of practice, vaccinology has proven to be very effective in preventing infectious diseases. However, many human and animal pathogens exist for which vaccines have not yet been developed. As for other fields of medical sciences, it is expected that vaccinology will greatly benefit from the emerging genomics technologies such as bioinformatics, proteomics and DNA microarrays. In this thesis, the potential of these technologies to aid understanding of bacterial pathogens is analyzed, taking into account the few existing examples of their application in vaccine discovery. The genomics era offers exciting new opportunities for vaccine research, opportunities that are expected to shorten the time of vaccine discovery and to stimulate new interest in vaccine research projects that have been partially or totally abandoned because of the failure of previous approaches. *Streptococcus Pyogenes* (Group A Streptococci, GAS) is a causative agent of pharyngitis, skin infections, cellulitis, necrotizing fasciitis and septicemia. Pharyngeal and skin infections can lead to the life-threatening post-infectious sequelae of rheumatic heart disease and glomerulonephritis, which cause widespread morbidity, especially in developing countries.

GAS express a number of virulence factors such as adhesins, toxins and proteases. However, considerable antigenic diversity exists between strains and thus the identification of protective antigens has proved to be difficult. Thus, new approaches are needed to find common antigens that could form the basis of a vaccine that protects against a broad range of GAS strains.

The complete knowledge of the gene repertoire of a given pathogen can become the launching platform for three experimental strategies, all aimed at the identification of new vaccine candidates. The three strategies share the same four basic steps: (1) selection of a set of genes from the whole genome; (2) high throughput cloning and expression of the selected genes; (3) purification of the recombinant proteins; and (4) *in vitro* and *in vivo* assays for identification of candidates. The strategies differ in the methods used for gene selection.

According to one strategy, genes are selected *in silico* utilizing algorithms that can predict genes encoding secreted and surface-associated antigens and virulence factors. These proteins are considered the most relevant for the induction of a protective immune response. Using the second approach, surface-exposed antigens are selected experimentally through the characterization of protein fractions using proteomics techniques such as two-dimensional gel electrophoresis (2D-GE) and mass spectrometry. The third approach is aimed at the identification of invasion and virulence-associated antigens and makes use of several new technologies including *in vivo* expression technology (IVET), signature-tagged mutagenesis (STM) and DNA microarray technology.

The work, developed at the Research Center of Novartis Vaccines (Siena), is based onto 2 strategies: genomic and proteomic approaches with the aim of identifying surface proteins belonging to GAS and in particular, a protein which could be a potential candidate vaccine. The first strategy follows the Reverse Vaccinology approach, already successfully applied to 2 different pathogens: Meningococcus B (Pizza et al, 2000) and Group B Streptococcus (Maione et al, 2005). At Novartis a careful *in silico* analysis was done of the GAS M1_SF370 genome identifying open reading frames and sequence analysis performed to predict cellular localization of proteins.

Almost 600 genes were identified and selected on the basis of the predicted surface localization of their products. The project can be divided into 3 parts.

Multiple recombinant streptococcal proteins purified as 6xHis tagged or glutathione-S-transferase (GST) fusion proteins, previously cloned and over expressed in *E. coli*. This work focused on the development of an automated system in which, starting from the bacterial lysate, it was possible to obtain highly purified proteins, through affinity chromatography, and the protein concentration determined by use of the Bradford test, through to the preparation of samples for SDS-PAGE electrophoresis.

This thesis describes the use of a robotic system that was employed to purify recombinant proteins, determine protein concentration and subsequently prepare

samples for SDS-PAGE analysis. This automated system was more precise than manual work and ensured an improvement of the purification reproducibility. In addition, this system allowed the simultaneous purification of 24 samples every two days, reducing the time required to produce highly pure recombinant protein.

In this work we studied GAS M1_SF370 and purified 327 His-tagged proteins (of which 267 were soluble and 60 were insoluble). In addition, we purified 106 Gst-fusion proteins.

Purified proteins were subsequently utilized in experiments to determine their surface localization and then tested for protection in a mouse model of infection. Experiments to determine exposure on the bacterial surface were done by citofluorimetric analysis (FACS) on M1_SF370 and M1_3348 using antibodies generated to the recombinant proteins; the second strain being highly capsulated and virulent in our mouse model. By FACS we identified 86 surface proteins starting from 186 analyzed samples. A parallel approach to experimentation must be taken in proteomics. The principle of parallel processing is a typical engineering solution to addressing throughput problems. This common approach to manufacturing can be applied to scientific problems as well.

We used, based on Novartis experience, (Rodriguez-Ortega et al., 2006) a proteomic approach for identifying bacterial surface-exposed proteins quickly and reliably for their use as vaccine candidates using the M1_3348 strain. Whole cells were treated with proteases to selectively digest surface exposed proteins that were subsequently identified by mass spectrometry of the released peptides. By surface digestion it was possible to identify 14 proteins, of which 13 were common with genomic approaches and 11 of those were FACS positive.

Starting from 600 genes (435 proteins) using the first approach and 14 proteins found by the second method we identified 2 common proteins which, tested in mouse model, gave a protection level:

- M Protein (90%)
- Gas 57-His (70%).

GAS possess numerous virulence factors but their M protein is regarded as one of the most important. Streptococcal M proteins appear as long hairlike filaments on the bacterial surface and are composed of two α -helical chains which are predominantly arranged in a coiled coil conformation. All of the M proteins contain a conserved C-terminal region near the cell surface and a hypervariable N-terminal region that provides the basis for serological typing. By contrast, GAS 57 is a previously undescribed surface protein, with high-levels of expression. To date, we have identified GAS-57 in 2 M1 strains. Further work is needed to determine the presence of this protein on the surface of other strains.

In conclusion, our surface digestion, results are preliminary and still under investigation. Some proteins may not be sensitive to proteolytic activity and their structure can be compact and not easily penetrable by the enzymes.

On the contrary, the limitation of *in silico* approach is to determine, in a prediction phase, if a protein is well-expressed and its cellular localization. This strategy is longer in its development, than the rapid proteomic approach.

Both methods are used together in parallel and have a common perspective: identify surface proteins, potential vaccine candidates. Of course, our experiments were performed with bacteria grown in culture, which may be very different to those found in active infections.

In the last part of the work we studied, in detail, GAS 57-His. It was purified, through 3 chromatography steps (affinity, ionic exchange, gel filtration) to obtain a good yield and a pure protein (88%); the purity was calculated by electrophoresis gel densitometry on Novartis experience.

The high protection level of this protein suggests that GAS 57-His may be a good candidate for a vaccine against *Streptococcus Pyogenes*.

Introduzione

1.1 La vaccinologia: cenni storici

1.1.1 microrganismi uccisi o attenuati

Gli approcci utilizzati per la produzione dei vaccini hanno compiuto notevoli progressi da quando Jenner sviluppò, nel 1796, il primo vaccino contro il vaiolo. Questa vaccinazione, venne effettuata inoculando in un uomo una piccola quantità di essudato di lesioni di una malattia molto simile al vaiolo, il vaiolo vaccino, benigna per l'uomo. Anni più tardi si sviluppò un metodo che permise la conservazione del vaccino, permettendone il trasporto in località distanti. Pasteur proseguì le ricerche, arrivando ad inoculare prodotti patologici della malattia, con conseguente creazione nell'organismo di anticorpi contro la malattia stessa. Le sue conclusioni portarono alla creazione dei vaccini costituiti da germi attenuati, di cui il vaccino antirabbico fu il primo (1885). Tuttora, la maggior parte dei vaccini diretti contro le malattie virali è costituita da virus vivi attenuati, come il vaccino contro il morbillo, la parotite e la rosolia, la poliomielite (tipo Sabin, per via orale) (Johnson, 2005), la febbre gialla e la varicella. Per i batteri e i micobatteri esempi analoghi sono rappresentati dal vaccino vivo attenuato contro la febbre tifoide. Altri vaccini sono preparati partendo da agenti uccisi, cioè inattivati, come il vaccino intero contro la pertosse, contro il colera, il vaccino contro l'influenza, contro l'epatite A e il vaccino inattivato contro la poliomielite (tipo Salk).

1.1.2 tossine batteriche e proteine coniugate a polisaccaridi.

Il vaccino batterico contro la difterite fu prodotto da Ramon nel 1923 e consisteva in un supernatante di coltura di *Corynebacterium diphtheriae* inattivato con formaldeide. Gli attuali vaccini contro la difterite sono realizzati in modo simile in quanto la tossina batterica viene parzialmente purificata dal supernatante di coltura. Verso la metà del 1900 (1925) fu sviluppato il vaccino inattivato contro *Bordetella pertussis*, che è ancora oggi utilizzato in diversi paesi in via di sviluppo. Le correnti tecnologie, per la preparazione di vaccini glicoconiugati e per l'inattivazione di tossine batteriche, furono sviluppate all'inizio del ventesimo secolo. Tra i vaccini commerciali, vaccini costituiti da componenti del microrganismo o da sostanze da esso sintetizzate come le tossine, possiamo anche citare quello contro il tetano. I vaccini contro l'*Haemophilus influenzae* di tipo b, lo pneumococco e il meningococco A, C, W, e Y sono costituiti da polisaccaridi coniugati su proteine "carrier". È necessario far presente che alcuni di questi vaccini, sia per i metodi usati nelle dimostrazioni della loro efficacia sia per le tecnologie delle loro preparazioni, sono notevolmente datati. Antiquato però non è necessariamente sinonimo di "inefficace", infatti, la maggior parte dei vaccini tradizionali è sicura e valida, come dimostrato dalla recente eradicazione del vaiolo (Belongia et al., 2003) o quella prevista tra alcuni anni del virus della poliomielite. Per identificare i componenti del batterio adatti allo sviluppo del vaccino, il patogeno viene cresciuto in laboratorio e tutti i componenti che lo costituiscono vengono identificati e analizzati tramite metodi biochimici, sierologici e genetici. L'identificazione di antigeni protettivi che possono essere potenziali candidati per il vaccino comporta la loro purificazione dal patogeno stesso. Questo approccio richiede molto tempo e permette l'identificazione solo di quegli antigeni che possono essere purificati in quantità accettabili da poter essere analizzati (Rappuoli, 2000). L'avvento di tecniche a DNA ricombinante ha comportato la necessità e l'utilizzo di nuovi strumenti per la ricerca dei vaccini. Specifici antigeni sono selezionati sulla base dei dati immunologici dei pazienti; dopo la loro purificazione dai patogeni o dai sistemi eterologhi in cui i corrispondenti geni sono stati clonati, gli antigeni sono testati per validità ed efficacia. Questo approccio ha generato due vaccini

ricombinanti molto efficaci. Il primo consiste nel vaccino contro il virus dell'epatite B basato sulla proteina del capsido virale purificata (Valenzuela et al., 1982); mentre il secondo è il vaccino a-cellulare contro *Bordetella pertussis* contenente tre diverse proteine batteriche ad alto livello di purezza (Greco et al., 1996) Nello sviluppo di quest'ultimo vaccino è stato applicato, per la prima volta, uno studio sulla relazione tra la struttura e la funzione, al fine di produrre una tossina funzionalmente inattiva ma che mantenesse inalterata la conformazione antigenica (Pizza et al., 1989; Pizza et al., 2003).

Altri vaccini sono tuttora ancora da testare in fase di studi preclinici e clinici (Mora et al, 2003). Inoltre la selezione dei possibili candidati è fastidiosa e molto lunga: infatti gli antigeni non è detto che siano necessariamente protettivi, o anche se protettivi, non possono essere usati in formulazioni di vaccini a causa della variabilità nella sequenza, nella difficoltà di espressione e di purificazione e degli alti costi di produzione.

1.2 L'era genomica

Dopo 200 anni di pratica, la vaccinologia è stata confermata essere l'approccio maggiormente efficace nella prevenzione delle malattie infettive anche se tuttora ci sono molte malattie infettive ancora in attesa di una cura efficace e molti patogeni emergenti. Per queste ragioni risultano necessari non solo nuovi vaccini ma anche nuovi approcci per il loro sviluppo. Il passo maggiore da fare è stato quello di selezionare al più presto possibile, un ampio set di proteine da testare per verificare la validità e l'efficacia come vaccini.

Con il completamento dell'identificazione di un numero sempre crescente di sequenze genomiche, l'attenzione della comunità scientifica è attualmente focalizzata su come i dati contenuti nelle banche-dati delle sequenze possano essere interpretati in termini di struttura, funzione e controllo dei sistemi biologici.

La possibilità di sequenziare un intero genoma batterico ha completamente cambiato l'approccio allo sviluppo dei vaccini dando inizio ad una nuova era: "l'era

genomica” che ha spostato l’attenzione dei ricercatori verso nuovi approcci per il disegno di vaccini per il trattamento di diverse patologie.

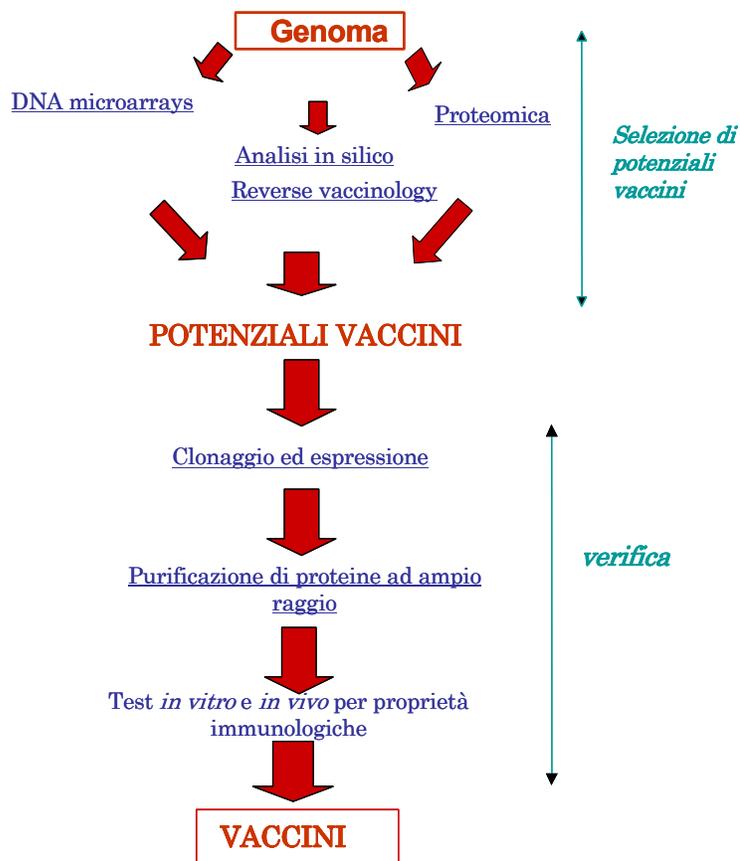


Fig 1: nuove piattaforme volte alla conoscenza dei genomi batterici e usate per l’identificazione a partire da nuovi potenziali vaccini.

L’era genomica si sviluppa su 3 fondamentali piattaforme per l’identificazione di possibili vaccini. Queste 3 piattaforme sono: i microarray di DNA, la reverse vaccinology e la proteomica (Figura 1).

Per l'identificazione di potenziali vaccini, ciascuno di questi metodi presenta i suoi vantaggi e i suoi limiti come riassunto nella tabella 1:

Tabella. 1 *Vantaggi e limiti degli nuovi approcci per lo sviluppo di un vaccino*

<u>Vantaggi</u>		<u>Limiti</u>
-Selezione di Proteine di associate alle membrane Veloce e sufficientemente affidabile.	<i>Reverse Vaccinology</i>	-Non è possibile sapere SE, QUANDO e QUANTO una proteina è espressa. -Possibile perdita di Candidati. - Incertezza Algoritmi
-Identificazione dei geni più espressi. -Identificazione dei geni trascritti durante l'infezione	<i>DNA Microarrays</i>	-No correlazione diretta tra livelli di mRNA e abbondanza proteine corrispondenti -No informazioni sulla localizzazione
- Identificazione delle proteine di membrana più abbondanti	<i>Proteomics</i>	-Limitazioni tecniche dell'approccio <i>classico</i> per le proteine di membrana proteine poco abbondanti, poco solubili, con elevato pI.

Una volta che un ipotetico antigene viene identificato è necessario purificarne una notevole quantità per confermare o rifiutare la sua possibilità di essere un vaccino. L'antigene viene purificato a partire dal patogeno stesso o attraverso un sistema di espressione in *E. coli*. L'antigene nativo o ricombinante è usato come vaccino in un modello animale, qualora quest'ultimo sia disponibile. Altrimenti, viene utilizzato per esperimenti *in vitro*, al fine di produrre anticorpi in topi che a loro volta sono usati per saggi di battericidia che riflettono una possibile protezione o per esperimenti di FACS (*Fluorescence activated cell sorter*) in cui gli anticorpi prodotti confermano la localizzazione superficiale dell'antigene.

Allorchè questi esperimenti siano concludenti, la nuova molecola può entrare nello sviluppo del vaccino.

Vediamo adesso queste tre piattaforme

1.2.1 DNA-microarrays

I DNA microarrays rappresentanti l'intero genoma del batterio studiato, sono sviluppati per analizzare i cambiamenti nell'espressione genica a diverse condizioni di crescita o indotti per differenti fattori. Applicati alla scoperta di un vaccino sono generalmente considerati i seguenti fattori:

- l'effetto del ferro
- l'effetto della riduzione dell'ossigeno
- l'effetto del contatto con la cellula dell'ospite

Per esempio, lo studio dei livelli di espressione dei geni coinvolti nella relazione ospite-patogeno, in Meningococco B, ha permesso di identificare diversi candidati con elevati livelli di trascrizione che sono stati considerati come potenziale vaccino contro il batterio (Grifantini et al, 2004). Tuttavia, forti limitazioni nella tecnologia risiedono nella non diretta correlazione tra i livelli di espressione degli mRNA e l'abbondanza delle corrispondenti proteine prodotte (Gygy et al, 1999). Inoltre, tale approccio non fornisce informazioni riguardanti la localizzazione delle proteine prodotte dai geni in esame (**Tab. 1**).

1.2.2 La reverse vaccinology

La disponibilità del genoma completo del microrganismo ha reso possibile la nascita di un nuovo approccio per lo studio e lo sviluppo di vaccini. Questa "*Reverse Vaccinology*", è stata ideata presso Novartis e applicata per la prima volta allo sviluppo di un vaccino contro meningococco B (Pizza et al, 2000; Giuliani et al 2006). E' stata successivamente applicata con successo allo sviluppo di un vaccino contro Streptococco di gruppo B (Maione et al, 2005). Il metodo non si basa sulla considerazione del microrganismo vivente, ma su algoritmi che sono in grado di trattare l'informazione contenuta nel genoma del batterio. Si usano software che predicono la

localizzazione cellulare di una detta proteina. Applicata allo sviluppo di un vaccino si selezionano le proteine indicate nella figura 2 in base al fatto che il microorganismo sia un Gram-negativo o un Gram-positivo.

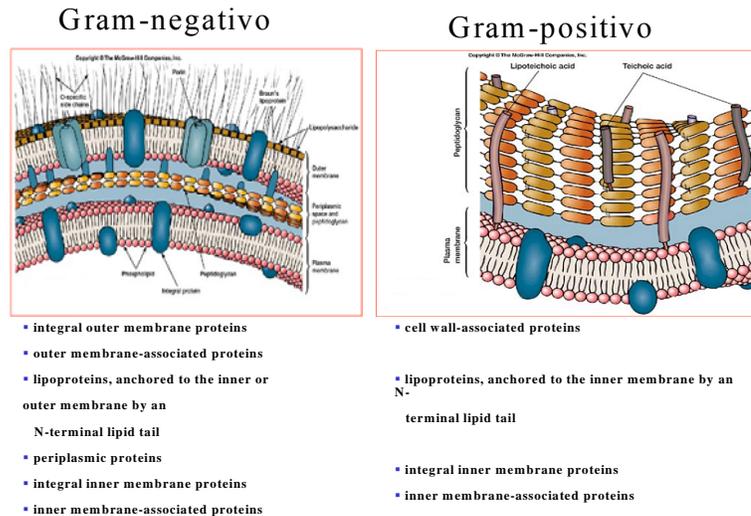


Fig 2: Rappresentazione di batteri Gram-negativo e Gram-positivo.

Il software presenta un'alta probabilità di predire una localizzazione cellulare giusta e il PSORT (<http://psort.nibb.ac.jp/form.html>), che attribuisce una localizzazione cellulare, combina una "informazione" della struttura primaria della proteina come *peptide signal membrane-spanning alpha-helices*, o il confronto della sequenze proteica, con un database di più di 1400 proteine con localizzazione note. Quando analizza un batterio Gram-negativo, le possibili localizzazioni sono il citoplasma, la membrana citoplasmatica, il periplasma, la membrana esterna e l'ambiente extracellulare. Quando analizza un batterio Gram-positivo, le possibili localizzazioni sono il citoplasma, la membrana citoplasmatica, la parete cellulare, e l'ambiente extracellulare.

Il software prende anche in considerazione la definizione dei motivi funzionali perché esiste una correlazione fra localizzazione cellulare e funzione. MenB e GBS rispettivamente a partire da 600 potenziali candidati selezionati del ceppo MC58 di

MenB (Pizza et al, 2000) e 589 potenziali candidati selezionati su 8 ceppi di GBS (Maione et al, 2005) .

La Reverse Vaccinology presenta molte caratteristiche innovative. Infatti, non è necessario coltivare il patogeno ed è quindi possibile usare anche quei microrganismi che non possono essere coltivati. Però, il limite dell'approccio *in silico* risiede nell'impossibilità di stabilire l'effettiva espressione di un antigene da parte del microrganismo, l'abbondanza dell'antigene in questione e la fase di crescita-infezione in cui tale antigene venga espresso (**Tab.1**). Inoltre, per quanto i sistemi di predizione impieghino algoritmi evoluti, il margine di errore persistente può dar luogo alla perdita di candidati utili, e alla permanenza di incertezze sulla effettiva localizzazione superficiale di alcuni antigeni.

1.2.3 La proteomica

Un terzo approccio per stabilire quali siano gli antigeni superficiali più espressi vede l'impiego delle tecnologie legate all'Analisi Proteomica, che potremmo definire come "sistematica caratterizzazione su larga scala delle proteine presenti in una cellula, tessuto o organismo" (Gorg et al, 2000). Una volta che la frazione proteica di interesse è preparata, il contenuto proteico può essere identificato seguendo due piattaforme.

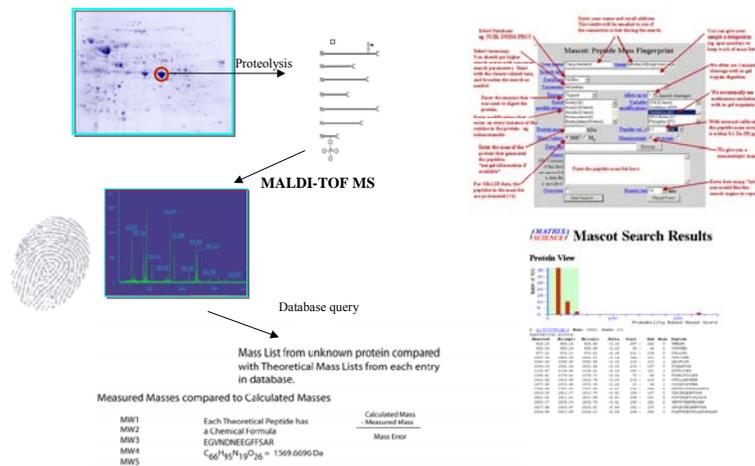


Fig. 3: Descrizione schematica della separazione della complessa miscela mediante elettroforesi bidimensionale (2D-PAGE) e successiva analisi di spettrometria di massa.

La prima (figura 3) si basa sulla separazione della complessa miscela mediante elettroforesi bidimensionale (2D-PAGE), dove le proteine sono separate in base al loro punto isoelettrico (pI) in una “prima dimensione” e al loro peso molecolare (MW) nella “seconda dimensione”. Il gel ottenuto viene sottoposto a *staining*, in modo da visualizzare gli *spot* relativi alle proteine separate. Gli *spot* proteici vengono prelevati e trattati con tripsina, al fine di ottenere la digestione *in-gel* delle proteine contenute negli *spot*; i peptidi risultanti vengono analizzati tramite spettrometria di massa. La spettrometria di massa MALDI-TOF (*Matrix Assisted Laser Desorption Ionization – Time Of Flight*), risulta al momento la tecnica più risolutiva che si possa affiancare alla 2D-PAGE: si basa sul desorbimento e la ionizzazione di specie chimiche (i peptidi triptici) mediante impulsi laser. Il loro rapporto massa/carica, e di conseguenza la loro massa, sarà individuata in base al *tempo di volo* (*Time Of Flight*), che impiega la specie chimica per raggiungere il *detector*. L’insieme delle masse dei peptidi di una particolare proteina rappresenta il cosiddetto *Peptide Mass Fingerprint* sperimentale, che confrontato con quelli teorici appresi dal genoma e disponibili in *database*, porta all’identificazione della proteina.

Questo approccio basato sulla separazione delle proteine da gel bi-dimensionale è limitante per separare e identificare proteine di alto o basso peso molecolare, estremi pI e di poca abbondanza.

E’ stata sviluppata una seconda piattaforma (figura 4) dove le proteine non vengono più separate da gel bi-dimensionale, ma vengono digerite con proteasi (generalmente la tripsina) e i peptidi risultanti dalla digestione vengono separati mediante cromatografia bi-dimensionale; la tecnica prevede una prima cromatografia dove i peptidi vengono separati mediante una cromatografia cationica e sulla seconda corsa, i peptidi vengono separati mediante una colonna a fase inversa (C18).

I peptidi eluiti dalla cromatografia in fase inversa si trovano in uno spettrometro di massa ESI Q TOF, dove vengono ionizzati al fine di potere esser separati in un quadropolo in funzione del loro rapporto massa/carica. Singoli peptidi vengono selezionati e frammentati.

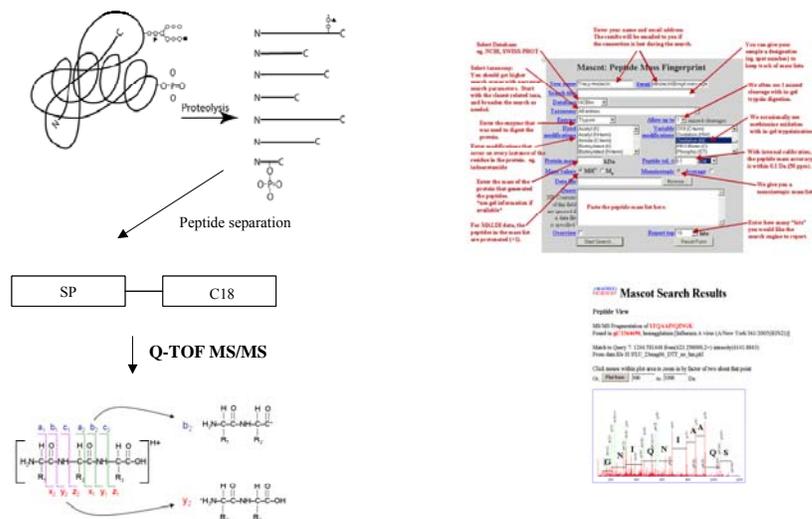


Fig 4: Rappresentazione della tecnica MS/MS

L'insieme delle masse dei frammenti di un particolare peptide viene confrontato con i frammenti teorici di sequenze proteiche generate dal genoma e disponibili in *database*, in modo tale da stabilire la sequenza del peptide e così risalire all'identificazione della proteina.

In teoria, questo approccio dovrebbe fornire un'informazione completa del contenuto proteomico nei campioni di interesse.

Applicato alla scoperta dei vaccini, l'approccio proteomico consiste nel definire quali sono le proteine contenute nella membrana esterna di un batterio gram-negativo.

Diversi protocolli sono stati descritti in letteratura per la preparazione della membrana esterna .

Fra questi protocolli, citiamo:

- L'estrazione delle proteine della membrana esterna per l'uso di detergente (sarcosina), (Rhomberg et al, 2004)
- lavaggio di un estratto batterico con carbonato (Molloy et al, 2000)
- uso di mutanti che esprimono vescicole soltanto costituite da membrane esterne e da periplasma (Ferrari et al, 2006)

Per quanto riguarda i batteri gram positivi, pochissimi lavori sono riportati in letteratura, citiamo comunque l'uso combinato di triton-X114 e lavaggio al carbonato per isolare proteine di membrana di *Streptococcus mutans* (Zuobi-Hasona et al, 2005).

Tutti questi protocolli permettono di ottenere una preparazione della frazione desiderata più o meno contaminata da proteine citoplasmatiche. In alcuni casi, queste contaminazioni impediscono un'accurata definizione del contenuto proteico relativo alla frazione in esame.

Inoltre, questi approcci non permettono di definire con sicurezza quali siano le proteine che realmente sono situate sulla superficie del batterio. Infatti, una proteina associata alla membrana può essere completamente "nascosta" all'interno del batterio stesso.

Per identificare proteine di superficie, sono stati riportati in letteratura esperimenti di biotinizazione sia su batterio gram-negativo (Sabarth et al, 2002) che su gram-positivo (Cole et al, 2004). Però questi lavori riportano una lista molto limitata di proteine di superficie.

Di recente è stata messa a punto una nuova procedura che permette la rapida e selettiva identificazione di proteine batteriche esposte sulla superficie, il pool di proteine che sono interamente o solo parzialmente esposte all'esterno della parete batterica.

Il metodo utilizza degli enzimi proteolitici, che agiscono sulla superficie del batterio e i peptidi generati sono separati dalle cellule intere e identificate mediante la spettrometria di massa.

La veridicità di questo metodo è stata dimostrata dalla caratterizzazione di un set completo di proteine esposte sulla superficie del ceppo M1_SF370 dello *Streptococcus pyogenes* (GAS) (Rodriguez-Ortega et al., 2006).

1.3 Il Batterio

Streptococcus pyogenes o Streptococco di gruppo A (GAS) è uno dei batteri patogeni più frequenti nell'uomo. Si calcola che in circa il 10% degli individui sani il batterio sia presente a livello del tratto respiratorio, senza manifestazioni cliniche.

Come la normale flora batterica, GAS può diventare infettivo quando il sistema immunitario è compromesso e può causare un ampio spettro di malattie, da leggere infezioni superficiali della pelle o della gola, a gravi infezioni invasive e complicazioni post-streptococciche, come febbre reumatica acuta e glomerulonefrite (Carapetis, 2004).

Prima della seconda guerra mondiale questo microrganismo era responsabile di più della metà delle morti *post-partum* ed era la maggior causa di morte nei pazienti ustionati. L'eradicazione dello Streptococco di gruppo A iniziò con l'uso della penicillina. Tuttavia verso la fine degli anni '80, a causa della ricomparsa di frequenti casi di febbre reumatica e di malattie sistemiche con gravi complicazioni, non solo negli stati in via di sviluppo ma anche nelle nazioni industrializzate, lo studio di GAS come patogeno ha di nuovo suscitato un notevole interesse.

Numerosi riferimenti in letteratura hanno documentato cambiamenti nell'epidemiologia delle malattie causate da *S. pyogenes*. Comprendere tali cambiamenti potrà essere molto utile per il riconoscimento ed il controllo delle infezioni e per ottenere preziose informazioni sulla patogenesi.

Ad oggi le infezioni da Streptococco di gruppo A costituiscono un grave problema per la salute pubblica.

1.3.1 Generalità

Streptococcus pyogenes appartiene al gruppo degli streptococchi β emolitici classificati nel 1933 da Rebecca Lancefield (1895-1981) che ne studiò la natura dal punto di vista antigenico (Lancefield, 1933).

Grazie a tale classificazione, basata sulle differenze immunologiche dei polisaccaridi presenti sulla parete cellulare, si sono potuti distinguere diversi sierotipi ognuno dei quali catalogato con una lettera dell'alfabeto (da A a O). E' inoltre possibile suddividere ulteriormente il gruppo A degli streptococchi sulla base della specificità della proteina M di superficie (Prescott et al., 1995).

GAS è un cocco gram-positivo, aerobio-anaerobio facoltativo, capsulato, non mobile, asporigeno che si sviluppa in catene di varia lunghezza anche se è possibile ritrovarlo da solo o in piccoli gruppi; è inoltre catalasi negativo e ossidasi positivo.

E' un batterio piuttosto esigente, infatti, per crescere necessita di terreni ricchi, a base di sangue o siero e di una temperatura di 37°C. Proprio l'agar sangue permette di distinguere lo Streptococco di gruppo A dagli altri streptococchi, quando viene cresciuto su terreno solido. Le sue colonie, infatti, appaiono di colore bianco-grigio, circondate da un alone di β -emolisi dovuto ad una lisi completa dei globuli rossi che le circondano (<http://medicine.ucsd.edu/NizetLab/streptococcipage/streptococci.html>).

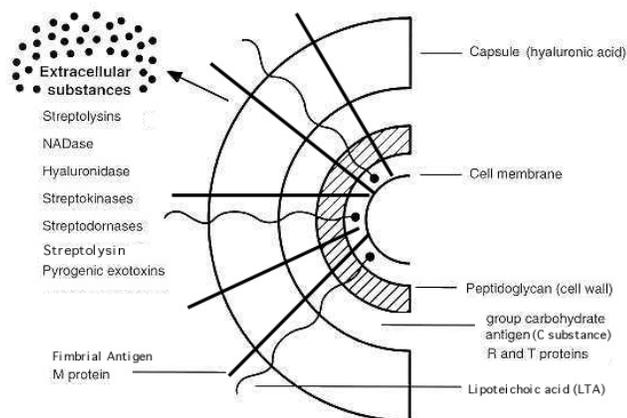


Fig. 5: Superficie cellulare di *Streptococcus pyogenes* e secrezione di prodotti di virulenza

Analizzando l'ultrastruttura (Fig. 5) di tale microrganismo possiamo osservare dall'interno verso l'esterno: la membrana cellulare, su cui sono inseriti l'acido lipoteicoico e le proteine, uno strato di peptidoglicano, in cui sono ancorate proteine con motivo LPXTG al C-terminale, uno strato formato dal carboidrato di gruppo A ed infine la capsula di acido ialuronico che circonda tutto il batterio.

1.3.2 Fattori di virulenza

Tra i fattori di virulenza maggiore attenzione è focalizzata su:

Capsula: costituita da acido ialuronico formata da unità ripetute di acido glucuronico e N-acetilglucosamina. La sua presenza ritarda la fagocitosi da parte dei leucociti polimorfonucleati e dei macrofagi dell'ospite. Ceppi mutanti acapsulati presentano capacità di colonizzazione e virulenza alterata nel modello animale. La capsula, inoltre, può essere un importante fattore di adesione nella faringe in quanto lega il CD44 delle cellule epiteliali (Ouskova et al., 2004).

L'evidenza epidemiologica che lega ceppi altamente capsulati a febbre reumatica ed a gravi malattie invasive suggerisce che la capsula giochi un ruolo fondamentale nelle infezioni invasive umane e che sia, assieme alla proteina M, uno dei fattori di virulenza più importanti.

Proteina M: è la proteina di superficie più abbondante e il fattore di virulenza più importante del batterio. L'estremità carbossi-terminale è ancorata alla parete cellulare tramite il motivo LPXTG, mentre l'estremità amino-terminale si estende dalla parete sottoforma di un dimero di α -eliche ed è molto simile alla struttura di proteine extracellulari di vari tessuti dell'ospite, quali tropomiosina, keratina-desmiventina, o la famiglia di molecole della keratina-miosina-epidermina-fibrinogeno.

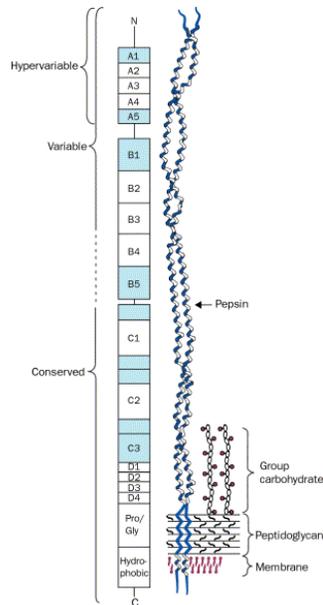


Fig. 6: Struttura della proteina M (Bisno et al., 2003).

La regione N-terminale, chiamata la regione ripetuta A, determina la specificità sierotipica del batterio, a valle della quale si trova un'altra regione della proteina variabile, la *mid-molecule region*, infine, la regione conservata all'altezza dell'estremità C-terminale (Cunningham, 2000).

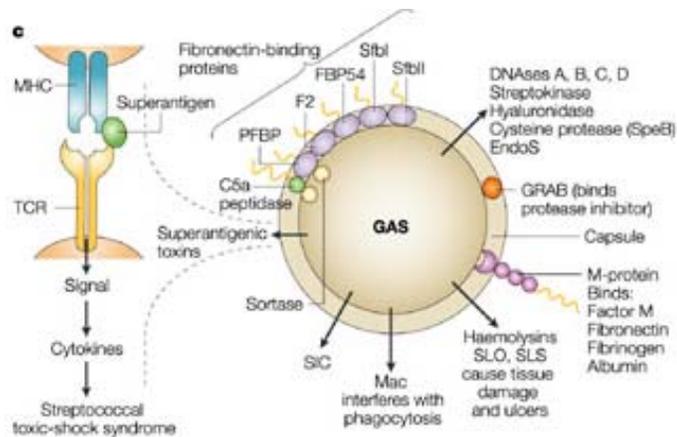


Fig. 7: principali fattori di virulenza di S. pyogenes (Mitchell, 2003)

1.3.3 Patogenesi

La tappa determinante per la colonizzazione dell'ospite è il legame tra ligandi di superficie dello Streptococco e specifici recettori presenti sulle cellule epiteliali della faringe e della cute. Senza meccanismi di adesione specifica, GAS non potrebbe invadere i tessuti dell'ospite e verrebbe rimosso dal flusso mucosale e salivare e dall'esfoliazione dell'epitelio. L'adesione specifica determina l'impossibilità della flora normale di legarsi ai siti tessuto-specifici. Lo studio di fattori di adesione è importante per capire i meccanismi di patogenesi e per lo sviluppo di terapie anti adesione o di vaccini per prevenire la colonizzazione.

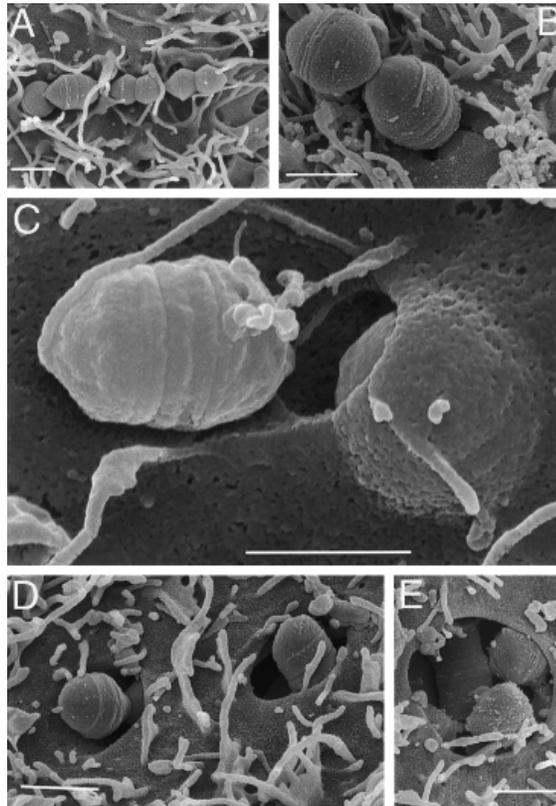


Fig. 8: analisi dell'interazione tra *S. pyogenes* e le cellule HEp2 derivate da carcinoma della laringe. (A):i batteri dopo un'ora dall'infezione sono adesi alle cellule eucariotiche. (B-E):formazione dell'invaginazione della membrana a livello del batterio. Le barre rappresentano 1 μm (Molinari et al., 2000)

Il meccanismo di interazione tra patogeno ed ospite consta di due passaggi consecutivi: un primo legame debole tra proteine del batterio e recettori presenti sul tessuto ospite, seguito da un immediato legame forte, tessuto specifico. I protagonisti delle due fasi sono l'acido lipoteicoico per la prima adesione e diversi fattori, tra cui la proteina M, per la seconda fase. L'invasione intracellulare di GAS è stata confermata grazie a studi su pazienti con tonsilliti ricorrenti, nei quali l'eradicazione dello Streptococco di gruppo A dalla gola fallisce in circa il 30% dei casi.

1.3.4 Manifestazioni cliniche

GAS colonizza la gola e la pelle ed è responsabile di infezioni suppurative e di sequele. Il concetto di distinguere ceppi che infettano la gola e ceppi che infettano la pelle nacque in seguito a decenni di studi epidemiologici in cui era evidente che alcuni sierotipi fossero correlati ad infezioni lievi quali tonsilliti, mentre altri fossero spesso associati all'impetigine (Cunningham, 2000).

L'infezione più comune causata dallo Streptococco di gruppo A è l'angina streptococcica, dovuta alla colonizzazione del tratto orofaringeo da parte del batterio e facilmente identificabile da un punto di vista clinico mediante tampone faringeo. La malattia che ne consegue è una rinofaringite e solitamente determina un quadro patologico lieve nei bambini, più acuto invece negli adulti. L'infezione solitamente non interessa i polmoni, sebbene la polmonite da *S. pyogenes* è rapidamente evolutiva e molto grave, ma sembra che sia più spesso una complicanza di infezioni virali precedenti. L'infezione localizzata negli strati cutanei superficiali prende il nome di piodermite streptococcica. E' costituita da vescicole superficiali rosse e dolenti, circondate da pelle arrossata, che si rompono e lasciano scoperte delle lesioni le cui superfici si ricoprono di pus o di croste. L'infezione, più comune nei bambini dai 2 ai 6 anni, si diffonde per contiguità ed è altamente contagiosa.

1.4 Valutazioni per lo sviluppo di un vaccino

La terapia profilattica di elezione in una infezione da *S. pyogenes* è, sin dal 1940, l'uso della penicillina. Questo antibiotico viene anche usato per prevenire il riaccutizzarsi di episodi di febbre reumatica e di glomerulonefrite.

I problemi di resistenza all'uso della penicillina si ritrovano solo in casi di invasione dei batteri nei tessuti epiteliali o in concomitanza di una coesistenza a livello del tessuto orofaringeo di batteri che producono lattamasi (Cunningham, 2000). Questa quasi totale recettività del batterio alla penicillina, unita al fatto che per la maggior parte dei casi le infezioni del batterio determinano blande infiammazioni o malattie facilmente curabili per l'uomo, non significa però che si debba ignorare la possibilità di mettere a punto modalità di prevenzione per le infezioni provocate da GAS. Non bisogna infatti dimenticare le malattie fortemente invasive che può procurare e i costi sociali associati anche a malattie più blande (Grandi, 2004).

L'ostacolo più difficile da superare per lo sviluppo di un vaccino consiste nell'evasione da una possibile risposta autoimmune. Le basi di questo concetto risiedono sulla reattività immunologica crociata che esiste tra epitopi di streptococco e tessuti umani, quali cuore, reni, cartilagini articolari e gangli basali del cervello. Il ruolo degli auto-anticorpi, indotti per esempio in caso di febbre reumatica acuta, non è ancora stato spiegato completamente. Volontari, con febbre reumatica acuta che avevano ricevuto nel 1976 un vaccino basato sulla proteina M, sono stati il bersaglio di studi clinici finalizzati all'identificazione di candidati per un vaccino (Polly SM et al., 1975).

Il passo scientifico più importante è stato l'identificazione degli epitopi protettivi della proteina M e della loro reazione incrociata con i tessuti umani. Questo ha potuto fornire una base per l'inclusione o l'esclusione di tali epitopi nello sviluppo di un vaccino. Negli anni '80 gli sforzi relativi allo sviluppo di un vaccino per GAS si sono focalizzati sulla protezione sierotipo-specifica della proteina M. Recentemente lo sviluppo del vaccino è stato esteso alla valutazione di molecole di superficie comuni tra i vari ceppi per produrre una risposta immunitaria dopo l'immunizzazione, largamente protettiva.

Lo scopo di un vaccino contro lo *S. pyogenes* sarà quello di generare una risposta immunitaria mucosale, per prevenire la colonizzazione del batterio ed evitare patologie locali come le faringiti ed allo stesso tempo stimolare anticorpi opsonizzanti circolanti per prevenire malattie sistemiche come shock, fasciti necrotizzanti e malattie cardiache.

1.4.1 Vaccini e singoli antigeni in fase di studio:

StreptAvax: è un vaccino a subunità costituito dai frammenti amino-terminali della proteina M, costruito per indurre una risposta immunitaria contro 26 ceppi di GAS, in modo da coprire gli isolati clinici responsabili del 95% delle malattie streptococciche nel Nord America. Il vaccino è stato sviluppato in primo luogo per i bambini dai 3 ai 6 anni (Hu et al., 2002).

StreptImmune: è basato su un antigene comune alle maggiori forme di GAS che causano gola infiammata e febbre reumatica e deriva dalla regione C-terminale della proteina M che è conservata all'interno di sierotipi diversi e non comporta problemi di cross-reattività. Una porzione conservata della proteina M6 è stata introdotta in un batterio commensale, lo *Streptococcus gordonii* che la presenta come antigene sulla propria superficie rendendola disponibile al riconoscimento da parte del sistema immunitario (Byrd et al., 2002).

Vaccine solutions: in questo caso il vaccino è formato da un piccolo peptide della proteina M sviluppato per l'induzione di una risposta mucosale locale tramite IgA e di una risposta immune sierologica al fine di prevenire la colonizzazione e l'infezione primaria da GAS (Dunn et al., 2002).

Intercell: è un vaccino a subunità che consiste di 2 o 3 antigeni ricombinanti scelti attraverso un programma di identificazione e validazione che induce anticorpi e stimola una risposta immunitaria mediata da linfociti T (Intercell Press release Feb 17, 2004).

Recombinant M protein fragment: consiste di 15 aminoacidi N-terminali della proteina M5 legati alla coda C-terminale di una subunità di una tossina di *E. Coli*.

Shp: l'analisi di una sequenza genomica di un sierotipo M1 identificò un gene codificante una proteina esposta sulla superficie cellulare. Questa proteina fu associata all'eme in rapporto 1: 1 e fu prodotta *in vitro*, localizzandosi sulla superficie cellulare di GAS e rendendosi accessibile a specifici anticorpi diretti contro la proteina ricombinante purificata (Lei et al., 2002).

Sfb1: è un vaccino formato da una forma ricombinante della proteina purificata o da una miscela di questa con la subunità B della tossina del colera e induce una risposta umorale specifica, mediata da IgG, una risposta mucosale polmonare e la protezione contro l'infezione da *S. pyogenes* (Schulze et al., 2003).

1.5 Approccio Novartis allo sviluppo di un vaccino contro *S. pyogenes* : Reverse Vaccinology

In base alle recenti esperienze positive ottenute con altri patogeni, Novartis ha utilizzato la Reverse Vaccinology come strategia iniziale per l'identificazione di potenziali antigeni candidati, che come descritto in precedenza si articola nelle diverse fasi, di seguito descritte in dettaglio.

1.5.1 Identificazione di proteine di superficie attraverso analisi computazionale

L'analisi computazionale è inizialmente stata condotta sul genoma del ceppo SF370 (M1) di cui era disponibile in banca dati la sequenza completa e l'annotazione (Ferretti et al., 2001). Il genoma del ceppo SF370, serotipo M1, è stato analizzato inizialmente utilizzando algoritmi compresi nel software della Genetics computer group

(Wisconsin package version 10.0) in grado di consentire una selezione virtuale delle sequenze che hanno capacità codificante. Successivamente tutte le open reading frames (ORFs) predette sono state inserite in un altro software, BLASTX, BLASTN o TBLASTX, che ricerca le omologie fra le sequenze inserite e quelle del database. Infine un terzo screening *in silico*, tramite l'uso di programmi e database, tipo BLAST, FASTA, MOTIFS, FINDPATTERNS, PSORT, ProDom, Pfam e Blocks, che consentono di identificare le caratteristiche tipiche delle proteine associate alla membrana come il peptide segnale, i domini di transmembrana, i motivi tipici delle lipoproteine (LXXC), di ancoraggio alla membrana (LPXTG) e di legame alla cellula ospite (RGD), oltre che l'omologia a proteine di superficie e a fattori di virulenza noti.

Inoltre è da sottolineare che il motivo RGD è in grado di interagire con le integrine presenti sulla superficie delle cellule eucariotiche e il motivo LPXTG permette di ancorare la proteina alla parete cellulare batterica attraverso legami covalenti.

Questo tipo di analisi ha permesso di identificare un pannello di 680 antigeni localizzati sulla superficie batterica.

Fra gli antigeni predetti di superficie è stata effettuata un'ulteriore selezione basata sui risultati ottenuti precedentemente con altri patogeni. In primo luogo sono state prese in considerazione per l'espressione in *E. coli* unicamente le 469 proteine contenenti non più di due domini transmembrana, in quanto era stato precedentemente osservato che proteine con 3 o più domini transmembrana sono difficilmente esprimibili in *E.coli*. La sequenza proteica di antigeni con più di due regioni transmembrana è stata successivamente analizzata mediante bioinformatica in modo da identificare domini proteici, potenzialmente esposti sulla superficie del batterio, da esprimere come proteine ricombinanti. L'analisi condotta sulle 211 proteine escluse dalla selezione iniziale ha portato alla definizione di un pannello di 58 domini di almeno 50 aminoacidi, appartenenti a 56 diverse proteine.

In secondo luogo, l'analisi retrospettiva dei risultati ottenuti in altri progetti indicava che gli antigeni protettivi sono generalmente codificati da geni altamente espressi. E' stata quindi intrapresa l'analisi dei livelli trascrizionali dei geni di *Streptococcus pyogenes* al fine di definire quali fossero i trascritti più abbondanti, particolarmente

durante lo stadio di infezione in un modello animale. Tale approccio sperimentale, ha portato all'identificazione di 770 geni altamente espressi, dei quali 258 codificanti per proteine predette di superficie e di 63 geni la cui espressione risultava aumentata successivamente ad inoculazione del patogeno in topi adulti.

E' tuttora in corso la selezione bioinformatica di altri antigeni di superficie di potenziale interesse, scelti fra gli antigeni unici presenti in nuovi genomi sequenziati appartenenti a ceppi di GAS con diverso serotipo M.

Si riportano in grafico le proteine appartenenti al genoma SF370 in seguito alla predizione *in silico*; le proteine sono suddivise in quattro sottogruppi:

- Proteine contenenti il motivo LPXTG (17)
- Proteine di membrana (561)
- Lipoproteine (28)
- Proteine secrete (83)

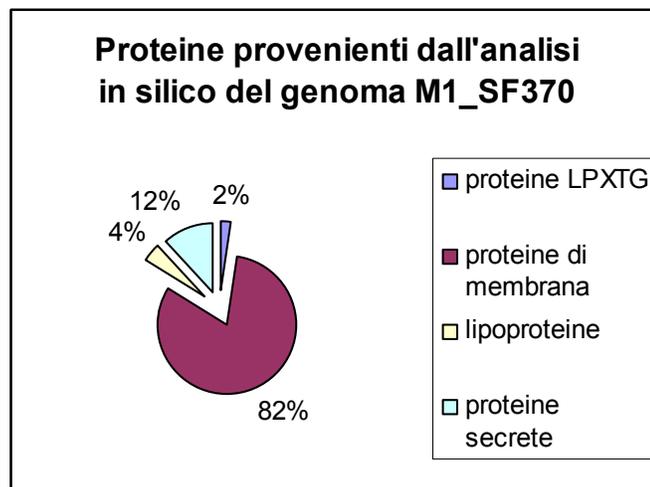


Fig. 9: rappresentazione schematica delle proteine FACS positive del genoma M1_SF370.

1.5.2 Clonaggio ed espressione degli antigeni selezionati

535 dei 543 dei geni selezionati sono stati clonati in plasmidi di espressione per *E.coli*. Tutti i geni sono stati clonati in modo da produrre sia delle proteine di fusione contenenti una coda di polyHis oppure la proteina Glutathione transferasi (GST) in modo da facilitarne la purificazione.

E' stata inoltre eseguita la verifica dell'espressione dei geni selezionati mediante analisi elettroforetica di estratti di cloni di *E.coli* trasformati con i plasmidi ricombinanti corrispondenti ai diversi geni clonati. Nell'87% dei cloni ottenuti, si è osservata l'espressione di un prodotto proteico con un peso molecolare in accordo con quello atteso per la proteina ricombinante considerata.

Un esempio è dato nella figura 10 che mostra l'analisi elettroforetica di otto estratti batterici totali isolati in cui è stata indotta l'espressione delle proteine clonate.

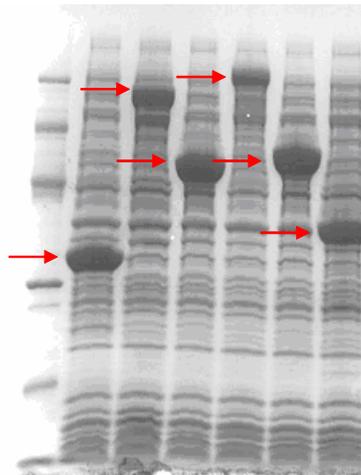


Fig.10 Esempio di SDS-PAGE di estratti batterici totali di cloni ricombinanti esprimenti proteine fuse con His o GST tag, indicate dalla freccia.

In accordo con l'analisi elettroforetica, a partire da 535 cloni, è stato possibile ottenere 381 proteine espresse contenenti la coda di sei istidine e 395 contenenti il gst come riportato in figura 11.

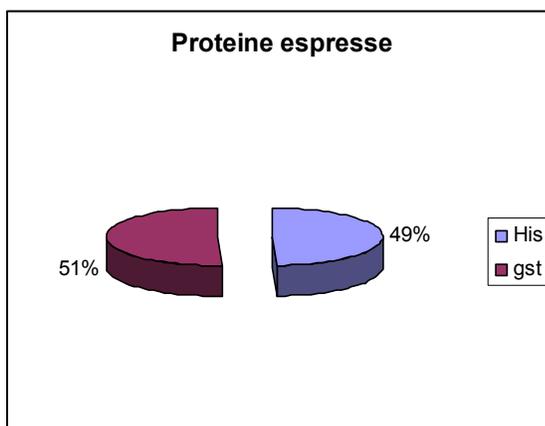


Fig. 11 Rappresentazione grafica del numero di proteine espresse

1.5.3 Purificazione delle proteine espresse in *E.coli*

Per identificare i potenziali antigeni utilizzabili per la formulazione di un vaccino, è necessario purificare un elevato numero di proteine. Questo necessita di un metodo che mantenga contenuti i tempi ed i costi. La scelta Novartis è stata indirizzata all'uso di proteine di fusione usando una colonna per cromatografia di affinità come efficace metodo di purificazione.

I due tag principalmente utilizzati sono il *glutathione S-transferasi* (GST tag) e la coda di sei istidine. Le proteine fuse con GST sono state purificate mediante la resina Glutathione-Sepharose. Il glutathione è legato alla resina mediante il legame con il gruppo ossidano (Figura 12b). La struttura del glutathione è complementare al sito di legame del glutathione-S-trasferasi.

La resina, *chelating sepharose*, usata per la cromatografia di affinità attivata con il Nichel, presenta un complesso di chelazione tra il supporto (Sepharose) e il metallo a formare un acido imminodiacetico come mostrato nella figura 12a; è comunemente nota come IMAC (*ion immobilized affinity chromatography*).

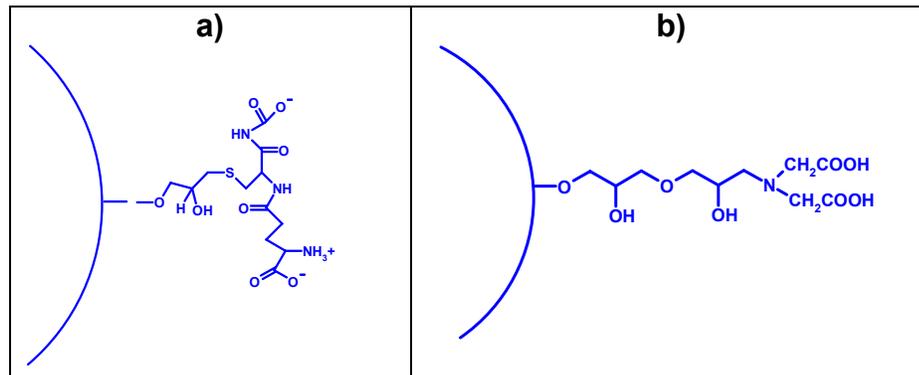


Fig. 12: a) Struttura parziale della resina attivata mediante la formazione del complesso imminodiacetico. b) Struttura terminale di glutatione sepharose. Il glutatione è legato alla resina mediante il legame con il gruppo ossirano. La struttura del glutatione è complementare al sito di legame del glutatione S-trasferasi

La strategia di purificazione consta di tre passaggi: caricamento del lisato su colonna, lavaggio ed eluizione della proteina, come descritto schematicamente in figura 13.

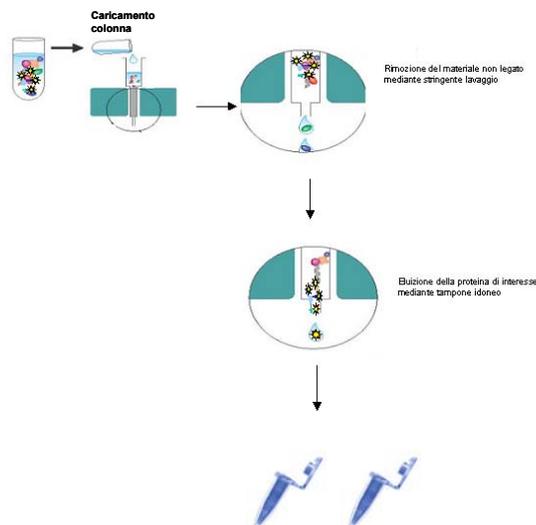


Fig:13 Strategia di purificazione delle proteine fuse ricombinanti. Le proteine sono state purificate da colture batteriche, per cromatografia di affinità.

L'approccio Novartis per quanto riguarda il progetto GAS richiedeva la messa a punto di un sistema di purificazione composto da cromatografia di affinità per uno svariato numero di proteine ricombinanti.

In primo luogo si è cercato di avere proteine solubili o con His-tag o con Gst; in secondo luogo le proteine His-tag provenienti da corpi inclusi, in cui la struttura della proteina viene denaturata. Si è scelto questo criterio perchè le proteine ricombinanti ottenute in forma solubile possiedono un "folding" uguale o comunque più simile, a quello delle corrispondenti proteine native.

La scelta di questo tipo di approccio facilita inoltre un futuro *scale-up* della purificazione.

Mediante questo sistema di screening ogni operatore era in grado di processare dalle 15 alle 20 proteine per settimana. Circa l'ottanta per cento delle proteine clonate sono state purificate in quantità e purezza sufficienti all'immunizzazione.

Dato l'elevato numero di proteine da purificare nel contesto dell'ampio progetto GAS, in Novartis l'attenzione è stata volta all'introduzione di un sistema di automazione, che consentisse la purificazione in parallelo di più proteine ricombinanti di interesse in maniera analoga al protocollo di purificazione eseguito manualmente, finora utilizzato.

1.6 Scopo della tesi

Il lavoro descritto in questa tesi rientra nell'ambito del progetto Novartis di ricerca di candidati antigenici in grado di attivare una risposta immunitaria protettiva contro infezioni da Streptococco di gruppo A. Questo progetto adotta due strategie, l'approccio genomico e l'approccio proteomico, la cui base comune prevede l'analisi *in silico* del genoma di un ceppo di sierotipo M1 di *Streptococcus pyogenes* per identificarne le ORFs e predire la localizzazione delle relative proteine.

In particolare il progetto di tesi può essere suddiviso in tre parti, come indicato nella figura 14:

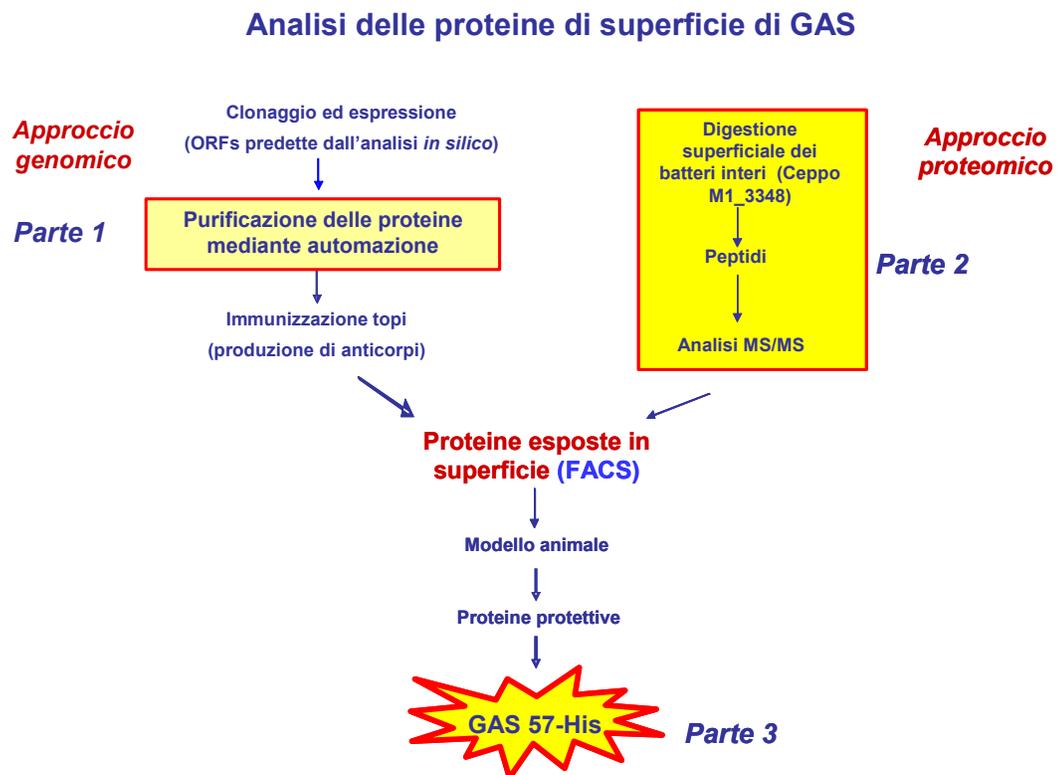


Fig. 14 Schema del progetto di ricerca

Parte 1: elaborazione di un sistema di high throughput per la purificazione dei candidati; la messa a punto di un sistema di automazione atto alla purificazione di una vasta gamma di proteine ricombinanti con il quale si potesse effettuare un rapido ed efficace screening; in tal modo si potrà avere una diffusa serie di proteine delle quali studiare le caratteristiche, le quantità prodotte e la qualità. Successivamente, avvalendosi dell'esperienza acquisita in Novartis, le proteine purificate, sono state utilizzate per immunizzare topi; i sieri anti-proteina così ottenuti, sono stati impiegati per l'analisi FACS per verificare quali di queste proteine siano esposte sulla superficie. Gli stessi sono stati successivamente saggiati nel modello murino per determinare un'eventuale protezione dall'infezione.

Parte 2: identificazione d'antigeni usando la proteomica. Parallelamente lo stesso obiettivo da me seguito nella prima strategia è stato considerato applicando anche la seconda; si effettuerà infatti un'attenta analisi proteomica su un ceppo (M1_3348), altamente capsulato e infettivo nel modello murino, mediante un nuovo approccio messo a punto in Novartis.

Il metodo prevede la digestione superficiale del batterio vivo. Dopo digestione, i peptidi generati sono purificati e la loro sequenza è determinata mediante l'utilizzo della spettrometria di massa MS/MS. La sequenza del peptide permette di definire quali sono le proteine esposte ovvero quali sono i peptidi maggiormente accessibili sulla superficie del batterio. Il metodo è stato applicato al ceppo SF370 di GAS che è un ceppo ipocapsulato e il ceppo virulento M23_DSM2071 (Rodriguez-Ortega et al, 2006). Dato che, per essere un buon vaccino, un candidato deve essere un antigene conservato su un ampio pannello di ceppi, è stato analizzato un altro ceppo.

Parte 3: L'ultimo passaggio del mio progetto sarà paragonare i due metodi descritti cercando di identificare tra le varie proteine di superficie analizzate, un potenziale candidato antigenico, studiato maggiormente in dettaglio, che presentasse tutte le caratteristiche per essere considerato un buon candidato per lo sviluppo di un vaccino contro lo Streptococco di gruppo A.

Risultati

Parte 1

2.1 Purificazione delle proteine mediante automazione

L'automazione e la robotica comportano un aumento di produttività e costi minori per i laboratori e tempi più ristretti (Gwyne et al., 2002).

Recentemente la scienza ha iniziato a beneficiare dei progressi della meccanica con l'introduzione di automazione da laboratorio e robotica in laboratori di ricerca e sviluppo. L'uso dell'automazione in laboratorio e la crescente emergenza di robotica hanno trasformato il lavoro quotidiano per molti ricercatori; gli sperimentatori sono in grado di progettare, ultimare e analizzare i risultati degli esperimenti in un tempo sicuramente minore rispetto a quanto ne necessitavano in precedenza.

L'applicazione nel campo dell'automazione va dai pipettatori multicanali a sistemi completamente robotizzati per un'operazione ad ampio spettro. Le industrie farmaceutiche e biotecnologiche si sono indirizzate verso l'automazione per trarne vantaggi (Lesley, 2001).

E' possibile automatizzare molte procedure di base con minimo sforzo: dispensare, riempire, aliquotare, lavare, fino ad effettuare test di immunochimica per analizzare un numero elevato di campioni contemporaneamente (Draveling et al, 2001).

Molti sistemi automatizzati sono utilizzati nella genomica, quanto nella proteomica; rientrano in quei sistemi utilizzati al fine di garantire un miglioramento della riproducibilità, un aumento di produzione, un'elevata sensibilità per la separazione e la purificazione del proteoma.

Sistemi composti da moduli permettono la possibilità di analizzare differenti processi di laboratorio in un'unica piattaforma di automazione.

Ciò che altamente influenza il disegno dell'automazione è la diminuzione di quantità di volume per ciascun campione da analizzare, visto che la maggior parte delle volte si ha necessità di lavorare con microlitri o anche nanolitri.

La diretta conseguenza del minore utilizzo di reagenti è legata al disegno di sistemi di dispensazione multicanale e di stazioni di lavoro che possono maneggiare piccoli volumi con elevata precisione.

La dispensazione di liquido richiede il collegamento, affinché non compia dei grossolani errori, di sofisticate strutture elettroniche che possono stimare la rivelazione del liquido in un tubo o una piastra e cercare di elargire il campione senza perdite o probabilità di contaminazione tra un campione e l'altro attraverso la scelta di differenti pressioni e velocità di flusso.



Fig. 15: Strumento utilizzato per la purificazione automatizzata delle proteine ricombinanti.

Il sistema robotico, mostrato in figura 15, da me riprodotto per le purificazioni di proteine ricombinanti, presenta:

- un braccio manipolatore con una pinza di presa
- un sistema di otto aghi
- uno spettrofotometro per la lettura delle piastre a 96 pozzetti
- un agitatore per piastre
- una stazione di lavaggio
- una pompa per il mantenimento della stazione di lavoro a 4°C.

Il sistema robotico si interfaccia con un software in cui ciascun sistema di purificazione riflette programmi dettagliati e piuttosto complessi nella loro formulazione.

Gli *scripts* messi a punto prevedono la contemporanea purificazione di un numero di campioni variabile da 1 fino ad un massimo di 24.

Particolare attenzione è stata volta alla dispensazione dei liquidi; è stato possibile infatti, a seconda della natura del liquido utilizzato, scegliere la velocità della dispensazione del liquido evitando in tal modo formazione di bolle d'aria e perdita di campione; inoltre talvolta si è agito sulla posizione dell'ago qualora il liquido di utilizzo presentasse un'elevata viscosità come, ad esempio il reattivo Bradford per la quantificazione delle proteine.

Inoltre è stato necessario valutare la portata degli aghi presenti in questo sistema; degli otto aghi precedentemente elencati, quattro hanno una capacità massima da 500 µl e altri quattro da 5 ml; infatti proprio giostrando su tale sistema si è raggiunta una migliore prestazione della dispensazione del liquido di utilizzo.

La strategia di purificazione nel sistema di automazione riprende quello adottato nel lavoro condotto manualmente.

Successivamente alla lisi dell'estratto batterico i volumi dei campioni proteici vengono normalizzati ad un volume di 12 ml; da questo momento il sistema automatizzato ha inizio. Dopo aver scelto appositamente il programma da utilizzare a seconda delle proteine da purificare e avere elaborato la lista dei campioni collegata poi all'elaborazione dei dati, i campioni vengono caricati in colonnine per il passaggio di cromatografia di affinità.

È stato possibile estendere la metodica di purificazione sia alle proteine portanti il tag GST sia le proteine del tipo His, sia solubili che insolubili. La fase di lavaggio ed eluizione differisce ovviamente per i diversi tag nell'utilizzo dei tamponi.

Per le proteine del tipo gst la resina con il campione viene lavata con 10 ml di PBS, pH 7.4; la proteina di interesse viene poi staccata dalla resina raccogliendo 2 campioni da 2 ml di glutatione ridotto a pH 8.0. Le proteine del tipo His solubili, vengono lavate con 20 mM imidazolo e successivamente eluite due frazioni da 2 ml ciascuna mediante un tampone al 250 mM imidazolo.

Analogamente si svolge la purificazione delle proteine His insolubili utilizzando per il lavaggio e l'eluizione i tamponi descritti nella sezione Materiali e Metodi.

Una volta raccolte entrambe le frazioni per ciascuna proteina i campioni vengono dosati mediante il dosaggio Bradford che copre un vasto range di linearità (0.02-50,0 mg/ml) utilizzando una curva di albumina bovina (BSA) a concentrazioni note come standard. Il test Bradford viene condotto sui campioni utilizzando un volume minimo di 142 μ l con i quali ciascuna misurazione viene effettuata in doppio, in modo che il valore uscente sia la media aritmetica dei due.

Le piastre a 96 pozzetti vengono poste in agitazione per tre minuti e successivamente lette allo spettrofotometro a 595 nm.

Il test Bradford viene condotto sia nel range standard (1.4 mg/ml-50 mg/ml) che nel test micro (0.02 mg/ml-1.4 mg/ml) (vedi sezione Materiali e Metodi) garantendo il minimo errore, in accordo con il saggio proteico scelto (circa il 15%). Si ha così la costruzione della curva di linearità paragonata sempre alla curva standard proveniente da concentrazioni note di BSA; la valutazione dell'accuratezza del metodo in ciascuna purificazione condotta è verificata dai valori di coefficienti di correlazione relativi a entrambi i range considerati nel test Bradford.

In seguito alla lettura i calcoli vengono eseguiti ed elaborati dal software; importanti ovviamente sono i valori di concentrazione (mg/ml) in base ai quali vengono preparati i campioni per la corsa elettroforetica (SDS-PAGE).

Per constatare la correttezza e la validità del metodo robotizzato sono stati in un primo momento analizzati campioni di BSA a concentrazione nota, come descritto in tabella.

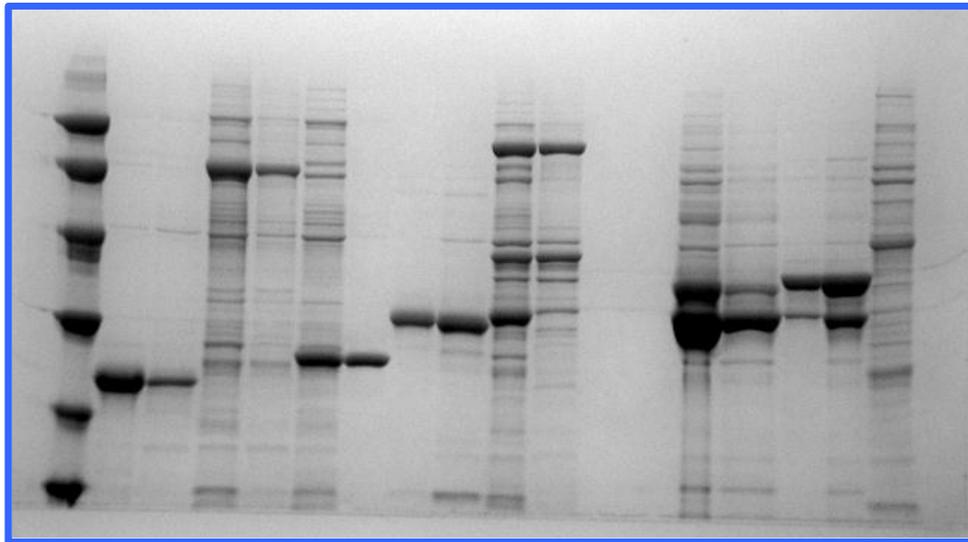
Albumina (mg/ml)	
Conc.note	Conc. ottenute (mg/ml)
50	52,76
33,3	31,06
22,2	22,21
14,8	14,54
9,87	11,50
6,58	6,45
4,38	4,27
2,92	2,98
1,95	2,98
1,3	2,59
0,86	1,18
0,57	0,89
2,5	3,13
1,66	2,78
0,5	0,59
0,3	0,47
0,04	0,05
0,02	0,02
2,5	2,02
1,66	1,19
0,5	0,34
0,3	0,25
0,04	0,05
0,02	0,02

Il secondo passaggio è stato quello di analizzare 24 proteine His diverse come descritto in tabella; Come mostrato vengono calcolate le concentrazioni, i volumi per ciascuna frazione e i milligrammi relativi; infine la concentrazione totale derivante dal pool delle due frazioni;

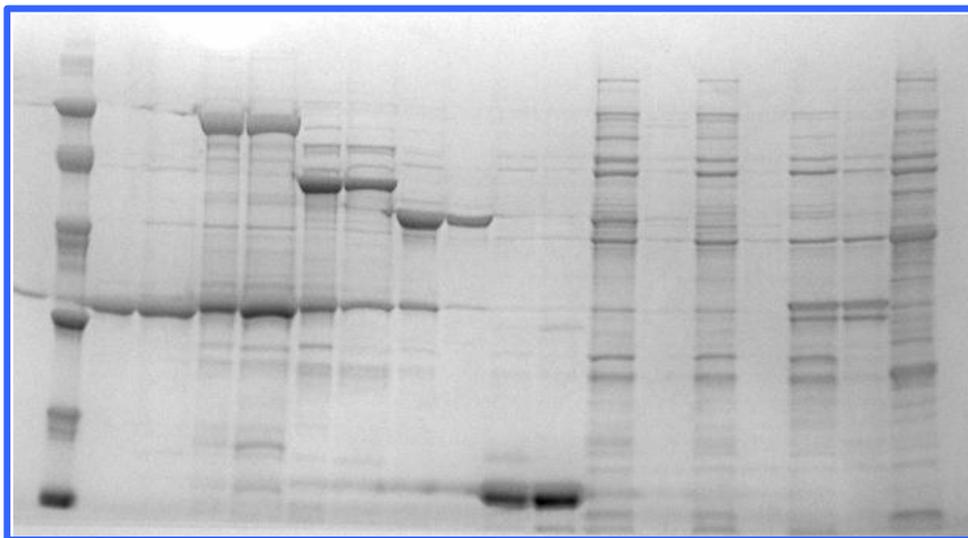
CAMPIONI	FRAZIONE 1 (mg/mL)	FRAZIONE 2 (mg/mL)	VOLUME FRAZIONE	VOLUME TOTALE	Fraz.1 (mg)	Fraz.2 (mg)	Totale prot. (mg)	Conc. mg/mL	Pool	Resa (mg/L)
18-His	6,227	0,229	1,858	3,716	11,570	0,425	12,00	3,23	1+2	24,0
25-His	1,005	0,287	1,858	3,716	1,867	0,533	2,40	0,65	1+2	4,8
30-His	1,181	2,792	1,858	3,716	2,193	5,187	7,38	1,99	1+2	14,8
65-His	9,733	0,531	1,858	3,716	18,084	0,987	19,07	5,13	1+2	38,1
75-His	1,010	0,411	1,858	3,716	1,876	0,764	2,64	0,71	1+2	5,3
bianco	0,000	0,000	1,858	3,716	0,000	0,000	0,00	0,00	1+2	0,0
77-His	1,217	0,488	1,858	3,716	2,260	0,906	3,17	0,85	1+2	6,3
81-His	10,838	1,392	1,858	3,716	20,137	2,586	22,72	6,11	1+2	45,4
82-His	10,510	0,237	1,858	3,716	19,528	0,440	19,97	5,37	1+2	39,9
88-His	4,667	0,701	1,858	3,716	8,671	1,303	9,97	2,68	1+2	19,9
97-His	1,354	0,357	1,858	3,716	2,515	0,664	3,18	0,86	1+2	6,4
99-His	4,970	2,276	1,858	3,716	9,235	4,228	13,46	3,62	1+2	26,9
108-His	6,407	1,322	1,858	3,716	11,904	2,456	14,36	3,86	1+2	28,7
130-His	0,679	0,000	1,858	3,716	1,261	0,000	1,26	0,34	1+2	2,5
142-His	0,449	0,000	1,858	3,716	0,834	0,000	0,83	0,22	1+2	1,7
178-His	0,437	0,219	1,858	3,716	0,811	0,406	1,22	0,33	1+2	2,4
187-His	2,640	0,414	1,858	3,716	4,905	0,770	5,68	1,53	1+2	11,4
195-His	1,029	0,080	1,858	3,716	1,911	0,148	2,06	0,55	1+2	4,1
206-His	1,367	0,070	1,858	3,716	2,540	0,131	2,67	0,72	1+2	5,3
218-His	0,747	0,073	1,858	3,716	1,388	0,136	1,52	0,41	1+2	3,0
249-His	0,538	0,304	1,858	3,716	1,000	0,565	1,56	0,42	1+2	3,1
277a-His	0,770	0,456	1,858	3,716	1,431	0,846	2,28	0,61	1+2	4,6
327-His	0,560	0,074	1,858	3,716	1,040	0,138	1,18	0,32	1+2	2,4
new504-His	0,778	0,353	1,858	3,716	1,445	0,656	2,10	0,57	1+2	4,2

Si riporta di seguito l'analisi elettroforetica delle proteine purificate; in ciascun gel sono riportati otto campioni (prima e seconda frazione); nella prima *lane* c'è il marker, riferimento per i pesi molecolari (LMW) e nell'ultima *lane* il solo plasmide contenente la coda di sei istidine inserito ed espresso in *E. coli* riportato come controllo negativo:

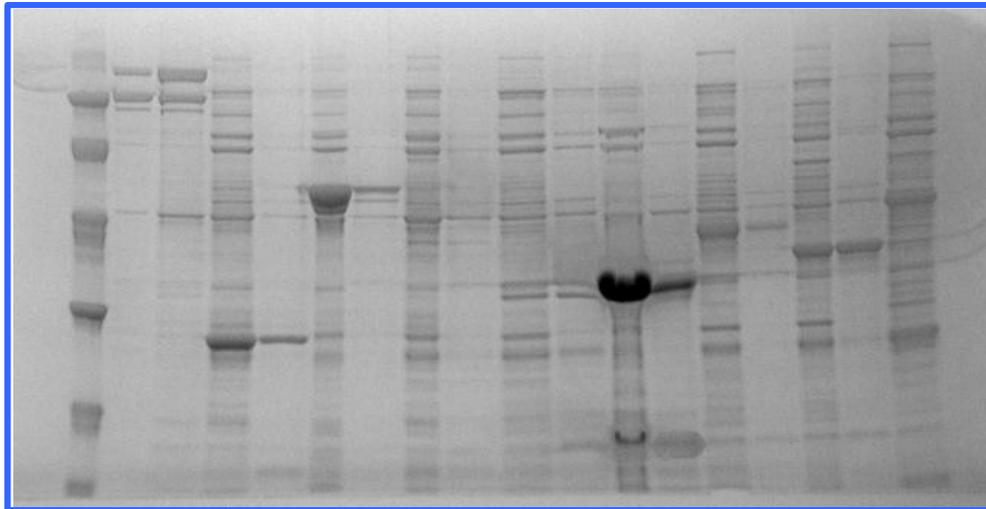
LMW 18-His 25-His 30-His 65-His 75-His Bianco 77-His 81-His Stop-His



LMW 82-His 88-His 97-His 99-His 108-His 130-His 142-His 178-His Stop-His



LMW 187-His 195-His 206-His 218-His 249-His 277a-His 327-His new504-His Stop-His



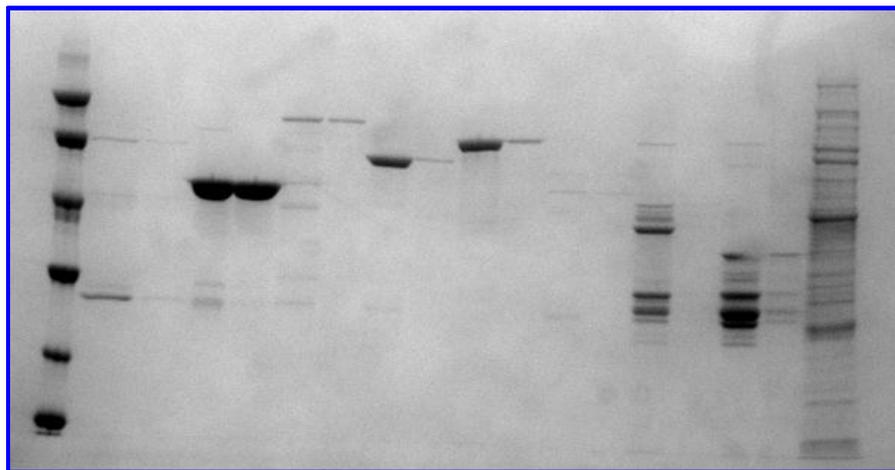
Si può notare che le proteine ricombinanti purificate presentano un'alta variabilità nel grado di purezza; quest'ultimo è calcolato mediante un metodo messo a punto in Novartis descritto nella sezione *Analisi della purezza delle proteine purificate (2.2)*.

Una volta ottimizzate le modalità di eluizione per ciascun tampone utilizzato, normalizzati i campioni per l'elettroforesi e l'analisi ed elaborazione dei dati, il passaggio successivo è stato quello di testare i coefficienti di correlazione delle curve di linearità del saggio proteico.

Si è quindi condotta la purificazione di otto gst-proteine come descritto:

CAMPION	RAZIONE1 (ng/ml)	RAZIONE2 (ng/ml)	VOLUME RAZIONE	VOLUME TOTALE	Fraz.1 (ng)	Fraz.2 (ng)	Totale prot. (ng)	Gr.: ng/mL	Rfd	Rsa (ng/L)
gst-6Hs	0,29		1,888	1,888	0,54		0,54	0,29	1	1,1
gst-60Hs	1,12	0,64	1,888	3,716	2,09	1,19	3,28	0,88	H2	6,6
gst-94Hs	0,49		1,888	1,888	0,90		0,90	0,49	1	1,8
gst-175Hs	0,22		1,888	1,888	0,41		0,41	0,22	1	0,8
gst-183Hs	2,53	0,08	1,888	3,716	4,71	0,15	4,85	1,31	H2	9,7
gst-218Hs				0,000	0,00				—	na
gst-472Hs	0,39		1,888	1,888	0,73		0,73	0,39	1	1,5
gst-630Hs	0,73		1,888	1,888	1,36		1,36	0,73	1	2,7

Si riportano anche i coefficienti di correlazione relativi alla curva standard per i valori di concentrazione compresi tra 1,4 mg/ml e 50 mg/ml e quelli relativi alla curva micro per le concentrazioni i cui valori sono compresi tra 0.02 mg/ml e 1.4 mg/ml



Coeff. Correlazione Curve STANDARD

0,9893

0,9900

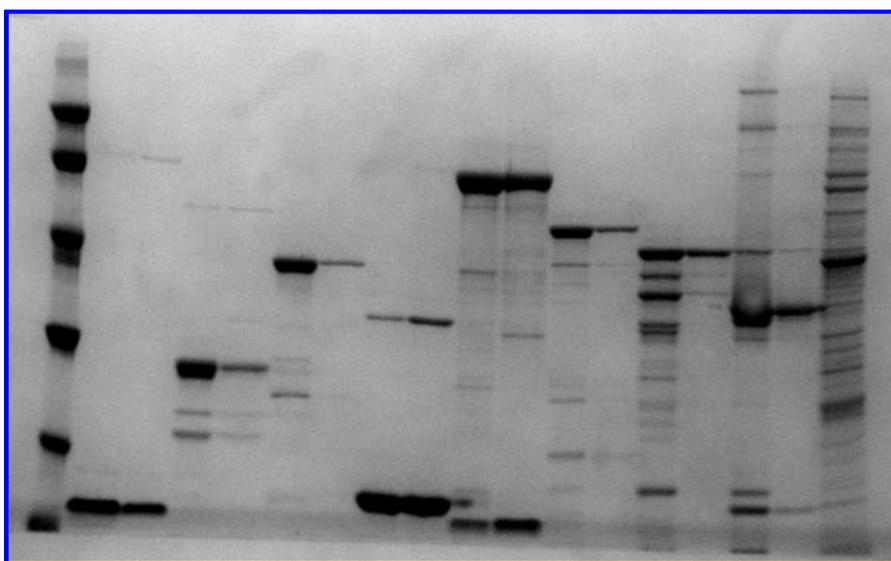
Coeff. Correlazione Curve MICRO

0,995

0,992

Per verificare la viscosità dei vari tamponi utilizzati è condotta la purificazione per otto proteine del tipo His insolubili come descritto:

CAMPIONI	FRAZIONE1 (mg/mL)	FRAZIONE2 (mg/mL)	VOLUME FRAZIONE	VOLUME TOTALE	Fraz.1 (mg)	Fraz.2 (mg)	Totale prot. (mg)	Conc. mg/mL	Pool	Resa (mg/L)
12-His	1,50	1,84	1,858	3,716	2,79	3,41	6,20	1,67	1+2	12,4
74-His	1,46		1,858	1,858	2,71		2,71	1,46	1	5,4
104-His	0,86	0,06	1,858	3,716	1,60		1,60	0,43	1+2	3,2
207-His	1,59	0,73	1,858	3,716	2,95	1,35	4,30	1,16	1+2	8,6
210-His	1,19	0,38	1,858	3,716	2,21	0,71	2,92	0,79	1+2	5,8
243-His	1,45	0,08	1,858	3,716	2,70		2,70	0,73	1+2	5,4
373-His	0,74	0,08	1,858	3,716	1,37		1,37	0,37	1+2	2,7
400-His	1,15	0,15	1,858	3,716	2,13		2,13	0,57	1+2	4,3



Coeff. Correlazione Curve STANDARD

0,99268

0,99025

Coeff. Correlazione Curve MICRO

0,983

0,971

Una volta ottimizzati i passaggi del sistema robotico sono state analizzate svariate proteine conducendo in parallelo lo stesso lavoro di purificazione manualmente.

Il primo gruppo di campioni analizzati comprende 24 proteine del tipo gst come descritto in tabella:

Campione	mg/ml manuale	mg/ml robot	mg.tot. manuale	mg. Tot. robot
gst-143-His 1	0,375	0,278	0,696	0,517
gst-143-His 2	0,295	0,25	0,549	0,465
gst-143-His 3	0,302	0,207	0,562	0,384
gst-143-His 4	0,324	0,264	0,602	0,491
gst-143-His 5	0,313	0,275	0,582	0,510
gst-143-His 6	0,356	0,289	0,661	0,538
gst-180-His 7	0,496	0,312	0,922	0,580
gst-180-His 8	0,437	0,34	0,812	0,633
gst-188-His 9	0,115	0,082	0,429	0,305
gst-188-His 10	0,116	0,083	0,433	0,310
gst-188-His 11	0,115	0,09	0,429	0,336
gst-188-His 12	0,115	0,082	0,426	0,306
gst-188-His 13	0,107	0,098	0,399	0,366
gst-188-His 14	0,116	0,095	0,431	0,351
gst-188-His 15	0,118	0,094	0,44	0,349
gst-188-His 16	0,123	0,093	0,457	0,344
gst-188-His 17	0,119	0,083	0,442	0,307
gst-188-His 18	0,123	0,085	0,456	0,317
gst-188-His 19	0,139	0,094	0,518	0,348
gst-188-His 20	0,124	0,076	0,461	0,284
gst-188-His 21	0,127	0,091	0,472	0,337
gst-188-His 22	0,146	0,098	0,544	0,364
gst-188-His 23	0,125	0,083	0,464	0,307
gst-188-His 24	0,13	0,085	0,483	0,316

Analogamente è stato condotto lo stesso esperimento anche per otto fusioni His.

Campione	mg/ml manuale	mg/ml robot	mg.Tot manuale	mg.Tot robot
2-His	0,560	0,653	2,081	2,427
2-His	0,592	0,699	2,200	2,597
57-His	4,060	2,384	15,087	8,859
57-His	4,598	3,368	17,086	12,515
57N-His	4,908	3,812	18,238	14,165
268-His	1,118	0,605	4,154	2,248

Si evince che lo strumento permette di ottenere dei dati che presentano una buona correlazione di linearità rispetto al lavoro manuale come mostrato in figura 16.

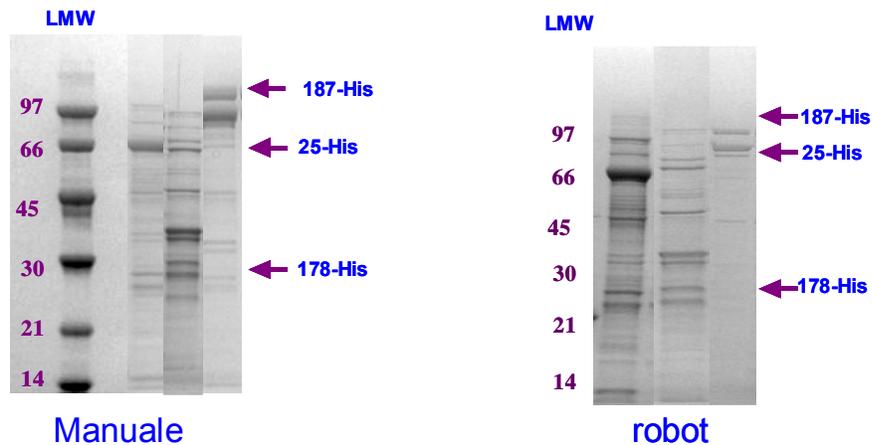


Fig.16: Analisi elettroforetica delle proteine purificate; comparazione tra lavoro manuale e robotica

Gli stessi campioni riportati in figura 16 sono stati dosati sia manualmente che avvalendosi dell'automazione e calcolate le purezze come mostrato nella seguente tabella (**Tab. 2**):

Tab. 2: dosaggio proteico e purezza dei campioni purificati sia manualmente che mediante automazione

Nome campioni	mg/L	mg/L (manuale)	Purezza robot	Purezza (man.)	MW (kDa)
25-His	4,6	6,2	52	42	59,4
178-His	2,3	1,9	28	22	26,3
187-His	10,8	8,5	45	36	124,4

Una volta garantita la messa a punto dello strumento e l'applicazione dei vari *scripts* per un numero elevato di proteine è necessario stilare delle conclusioni circa i vantaggi e gli svantaggi della automazione.

Di seguito si riporta una tabella (**Tab. 3**) dove vengono confrontati con minuzia, tutti i tempi previsti nell'esecuzione di ciascuna operazione, utilizzando l'approccio manuale e quello robotico.

Tab. 3: confronto, in termini di tempistica tra lavoro manuale e robotico.

Confronto automazione vs manuale				
Operazione	manuale		automazione	
	Tempo	operatore	Tempo	operatore
Lisi e preparazione campioni	=	=	=	=
Carico colonna	60 min.	1	60 min.	0
Lavaggio	60 min	1	60 min.	0
Raccolta frazioni	60 min	1	60 min.	1
Dosaggio proteico	4 h	1	6 h	0
Preparazione campioni SDS-PAGE	3 h	1	2 h	0
Preparazione campioni per immunizzazioni e stoccaggio	=	=	=	=

	manuale		automazione	
	Tempo tot	operatore	Tempo tot	operatore
N° campioni				
24	10h	3	11h	1

Si può notare che le operazioni eseguite mediante lo strumento impiegano un tempo pressochè paragonabile a quello effettuato manualmente.

Questo, a prima vista potrebbe sembrare uno svantaggio, ma viene facilmente soppiantato dal fatto che, a partire dallo stesso numero di proteine, c'è un minor numero di operatori impiegati nell'esecuzione. La tempistica anche se a prima vista non particolarmente vantaggiosa viene superata dal fatto che lo strumento può essere lasciato in esecuzione durante la notte, garantendo la completezza delle operazioni senza richiesta di lavoro manuale.

Inoltre questi tipi di operazione, soprattutto in una fase di selezione ad ampio raggio, essendo piuttosto tediose e di routine, facilmente comportano la presenza di errori e di imprecisioni; quest'ultime ad uno stadio di automazione raramente accadono data l'esatta riproducibilità e precisione del sistema operativo adottato.

Avvalendosi dell'aiuto di un sistema robotico per la ricerca applicativa è stato quindi possibile purificare un numero elevato di proteine ricombinanti (Figura 17). Questa attività ha permesso di ottenere 327 proteine con la coda di istidine di cui 267 solubili e 60 insolubili. Inoltre sono state purificate 106 proteine contenenti la porzione GST.

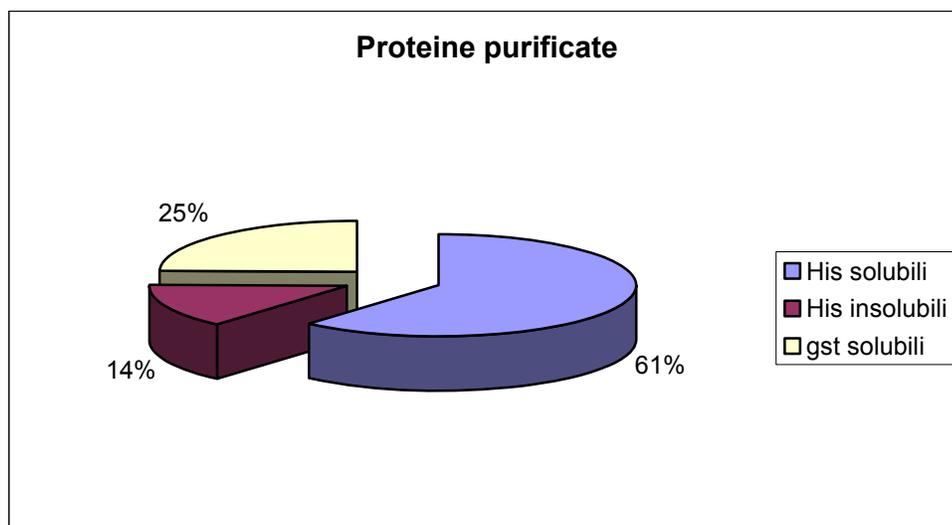


Fig. 17 Rappresentazione grafica della percentuale delle proteine purificate ottenute.

Si riporta di seguito uno schema riassuntivo (**Tab. 4**) sui risultati del genoma M1_SF370. Qui sono menzionate le quantità e le percentuali relative a ciascun passaggio previsto, a partire dalla predizione dei geni mediante analisi computazionale fino alle proteine clonate ed espresse; si riportano le proteine clonate nei due sistemi di fusione adottati (coda di poly-His e Gst), purificate, sia solubili che insolubili; inoltre si mostra la quantità di proteine disponibili all'immunizzazione legato alla quantità ottenuta.

Tab. 4 Tabella riassuntiva delle proteine purificate

Risultati del genoma SF370								
	Geni	Prot. clonate	Prot. Espresse	Purific.	Ottenute	Solubili	Non solubili	Disp. x Imm.
His	545	535	381	379	327	267	60	327
		98,2%	71,2%	69,6%	86,5%	81,6%	18,3%	86,3%
GST-His	545	424	395	186	106	106	---	83
		77,8%	72,5%	47,1%	57,5%	100,0%	---	44,6%
	Geni	Prot. clonate	Prot. Espresse	Purific.	Ottenute	Solubili	Non solubili	Disp. x Imm.
Tot.	545	480	388	565	433	374	60	410
		99,6%	86,8%	122,0%	76,8%	86,2%	13,8%	72,6%

2.2 Analisi della purezza delle proteine purificate

Dopo purificazione delle proteine ricombinanti di GAS, avvenuta come indicato, si è proceduto ad accertare la purezza di queste ultime utilizzando il programma di analisi sviluppato in Novartis. Il programma permette di parametrizzare l'immagine del gel evidenziando la "lane" dei controlli standard "markers" (lane 1) e delle proteine ("lanes" 2-17). Il grafico rappresentato in Figura 18b mostra l'andamento dell'intensità di ciascuna lane in funzione dei pixels. La banda d'interesse selezionata sul gel (Figura 18a) corrisponde al picco rosso nella figura 18b. Il valore di purezza della proteina espresso in percentuale, è dato dal rapporto tra l'area sottesa dal picco e l'area dell'intera "lane".

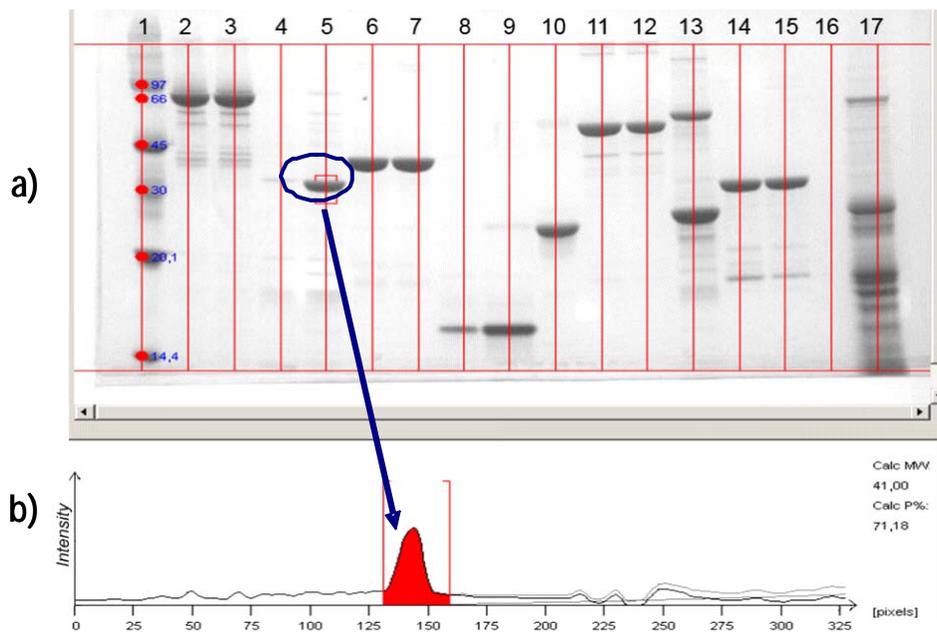


Fig 18: Analisi della purezza delle proteine su SDS-PAGE: Il programma sviluppato dai ricercatori della Novartis si basa sull'analisi densitometrica del campione in esame.

2.3 Test immunologici degli antigeni espressi

Gli antigeni identificati *in silico*, espressi e purificati, sono stati utilizzati per:

- ottenere anticorpi (sieri immuni) che permettessero lo studio dell'espressione degli antigeni stessi in ceppi di GAS diversi.
- verificare la loro capacità di indurre in animali da laboratorio una immunità protettiva nei confronti di una infezione da GAS.

Le proteine purificate secondo le modalità sopra descritte sono state utilizzate per immunizzare topi femmina adulti CD1 dell'età di 36-38 giorni. I sieri preimmuni ed immuni dei topi CD1 immunizzati sono stati utilizzati per effettuare un'analisi citofluorimetrica su batteri vitali, seguendo la metodica riportata (vedi sezione Materiali e Metodi). Questo tipo di analisi aveva come scopo quello di verificare l'esposizione dell'antigene considerato sulla superficie del batterio e il relativo livello di espressione. Antigeni rilevabili in quantità elevata sulla superficie del batterio, potrebbero essere teoricamente i bersagli preferiti di anticorpi specifici i quali potrebbero prevenire l'infezione.

Questo tipo di analisi è stata condotta con i sieri ottenuti con 100 antigeni appartenenti alle seguenti categorie:

- antigeni che avevano mostrato la migliore predizione bioinformatica della loro effettiva localizzazione sulla superficie batterica
- antigeni identificati come protettivi nei saggi di immunoprotezione *in vivo*

L'analisi è stata effettuata su 2 ceppi diversi di GAS:

- M1_SF370
- M1_3348

In particolare nel ceppo M1_SF370 sono stati analizzati in totale 116 sieri specifici anti-proteina dei quali 72 sono risultati positivi come rappresentato in figura 19.

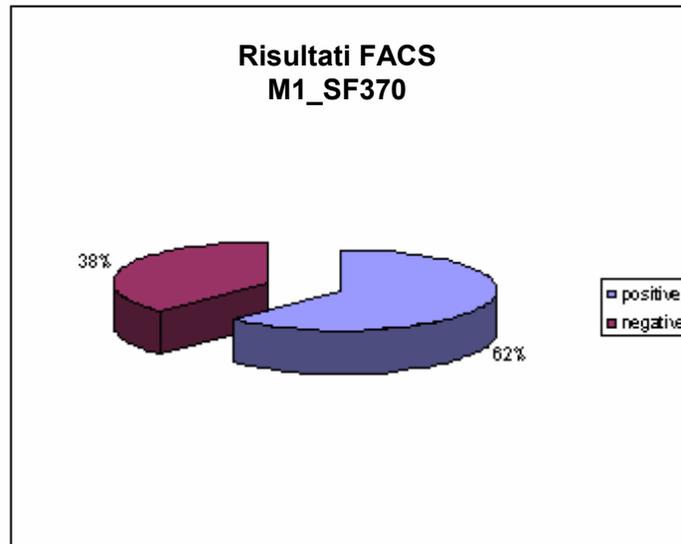


Fig.19 Analisi FACS sul ceppo M1_SF370

I risultati di ciascun esperimento di FACS provengono da un istogramma in cui vengono sovrapposti il picco di intensità di fluorescenza dato dal siero immune e quello del corrispondente siero preimmune, utilizzato per controllare il *background*. La differenza tra i due picchi, indice del livello di esposizione degli antigeni sulla superficie batterica, è indicata come Δ Mean definita come valore medio dell'intensità di fluorescenza della popolazione batterica. Si sono considerati positivi, nel caso specifico del progetto GAS, in accordo con i precedenti risultati avuti in Novartis, valori di Δ Mean > di 80 canali.

Pertanto si può concludere, considerata la positività al FACS, che 72 delle proteine inizialmente selezionate siano effettivamente sulla superficie di questo ceppo del batterio; le altre proteine, non è detto che non siano sulla superficie, ma molto probabilmente sono poco espresse.

Ciascuna proteina purificata come proteina ricombinante, a partir dalla sequenza genomica del ceppo SF370, è stata in egual modo testata anche su un altro ceppo di serotipo M1 (3348), altamente capsulato e infettivo nel modello murino.

Per il ceppo M1_3348 sono state testate al FACS 186 proteine di cui sono state trovate esposte in superficie 86 (Figura 20).

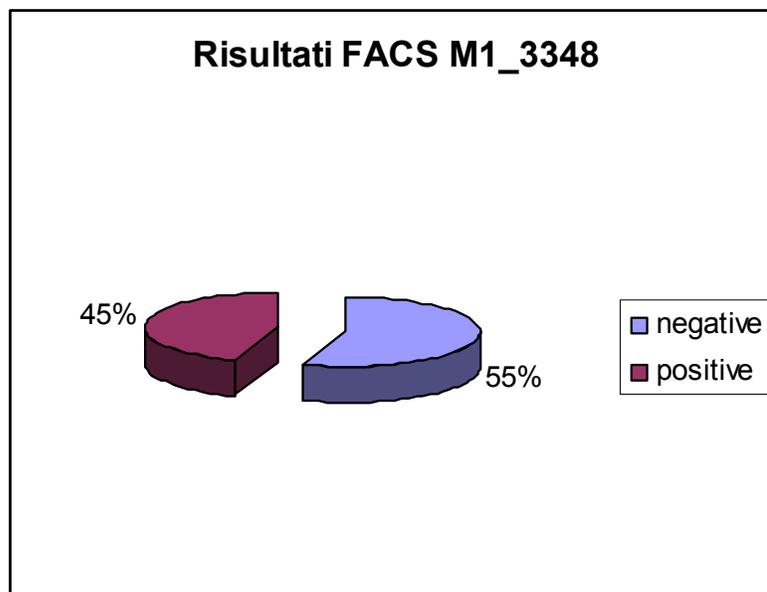


Fig. 20 Analisi delle proteine testate al FACS utilizzando il ceppo M1_3348

Si riporta di seguito la tabella (Tab. 5) delle proteine FACS positive per il ceppo M1_3348 e il valore di Δ mean maggiore di 80.

Tab. 5: Elenco delle proteine risultate positive all'analisi FACS testate sul ceppo M1_3348.

<i>Proteine GAS</i>	<i>FACS positive (Δmean)</i>	<i>Proteine GAS</i>	<i>FACS positive (Δmean)</i>
4	97	475d	436
5	322	538d	260
15	508	543d	157
16	545	553d	256
18	247	561d	201
39	403	592d	270
40	271	636d	215
56	411	643d	310
57	239	649d	271
58	309	49	149
64	258	149	148
68	229	72	147
75	182	101	146
76	325	437d	145
84	183	389	144
93	180	143	143
96	317	591d	140
98	335	663d	139
100	163	201	133
103	314	472d	128
108	175	461d	125
123	236	23	122
130	192	439d	122
142	174	89	120
157	192	426d	119
177	205	406	117
188	244	45	114
190	359	384	111
191	342	158	109
193	243	577d2	108
290	358	206	108
166	90	202	106
366	153	294	106
405	171	473d	105
433	253	495d	104
527	157	131	104
529	176	99	103
533	178	137	100
685	296	478d	99
414d	213	217	98
434d	212	433d	97
438d	178		
465d2	390		
469d	151		

Parte 2

2.4 Approccio proteomico applicato al ceppo GAS M1_3348

L'approccio genomico ha permesso di selezionare 545 geni da cui ottenere 434 proteine ricombinanti che sono state purificate con successo per essere testate mediante analisi FACS al fine di validare la loro localizzazione superficiale. Ovviamente è stato necessario effettuare uno *screening* complicato e completo di un numero elevato di proteine.

Di conseguenza si evidenzia la necessità di sviluppare un metodo complementare per la selezione di potenziali candidati.

Dalla precedente esperienza con *Meningococcus B* e lo *Streptococcus B* (Pizza et al, 2000; Maione et al 2005) sappiamo che gli antigeni protettivi devono essere ben espressi e esposti sulla superficie della cellula batterica. Inoltre devono essere presenti, in termini di sequenza e struttura in numerosi ceppi. Un vaccino per essere universale deve proteggere contro un molto ampio spettro di ceppi.

L'analisi proteomica di digestione superficiale da diversi ceppi può essere un approccio ideale all'identificazione di un candidato conservato ed esposto in superficie e tale strategia complementa la classica *reverse vaccinology*. Oltre alla analisi del ceppo M1_SF370 (vedere appendice), (Rodriguez-Ortega et al, 2006), messa a punto nei laboratori Novartis, si è deciso di investigare proteine presenti in altri ceppi che potessero dare protezione in un modello murino in seguito ad infezione e malattia.

È noto che il ceppo M1_SF370 è totalmente sequenziato, ipocapsulato e commercialmente disponibile, ma non infettivo nel modello animale; di conseguenza si è pensato di testare un altro ceppo che invece risultasse infettivo nel modello murino: M1_3348. Il ceppo è stato cresciuto fino a OD₆₀₀ 0.4. ; i batteri sono stati recuperati e digeriti con tripsina o con la proteinasi K per 30 minuti.

Dopo isolamento dei peptidi generati, questi sono stati separati mediante cromatografia bi-dimensionale, e analizzati "on line" tramite MS/MS su uno strumento ESI Q-TOF (Micromass Waters).

Le sequenze dei peptidi sono state dedotte, usando il genoma del ceppo SF370 per l'interpretazione degli spettri di frammentazione. Le sequenze proteiche sono state attribuite a 14 proteine differenti. Con l'eccezione di una proteina, tutte le altre appartengono, secondo il software PSORT ai compartimenti cellulari associati con la superficie del batterio:

Parete cellulare:

- 7 proteine con motivo LPXTG (**a**)

Membrana:

- 4 proteine di membrana (**b**)
- 1 lipoproteina (Hypothetical protein) - GAS 103

Extra-cellulare:

- 1 proteina secreta (putative 42KDa Protein) - GAS 64

Citoplasma:

- 1 proteina citoplasmatica (putative translation elongation factor EF-Tu)

Si riporta di seguito la tabella delle proteine identificate :

a) le proteine contenenti il motivo LPXTG (7 proteine)

GAS 57	putative cell envelope proteinase
GAS 158	Hypothetical protein
GAS 166	GRAB (Protein G-related alpha 2M binding protein)
GAS 188	collagen-like surface protein
GAS 190	hypothetical protein
GAS 191	C5a peptidase precursor
GAS 192	M protein

b) le proteine di membrana

GAS 39	Conserved hypothetical protein
GAS 194	putative ATP-binding cassette transporter-like protein
GAS 277	hypothetical protein
GAS 282	hypothetical protein

2.4.1 Comparazione dei ceppi analizzati

Il successivo passaggio è stato volto alla comparazione della digestione superficiale tra il ceppo M1_SF370 precedentemente analizzato (Rodriguez-Ortega et al., 2006) e il ceppo M1_3348. Il numero di proteine identificate è relativamente basso in confronto al numero di proteine identificate mediante la digestione superficiale dal ceppo M1_SF370 (14 contro 72 proteine). Per la comprensione di questo risultato, possono esservi due ragioni:

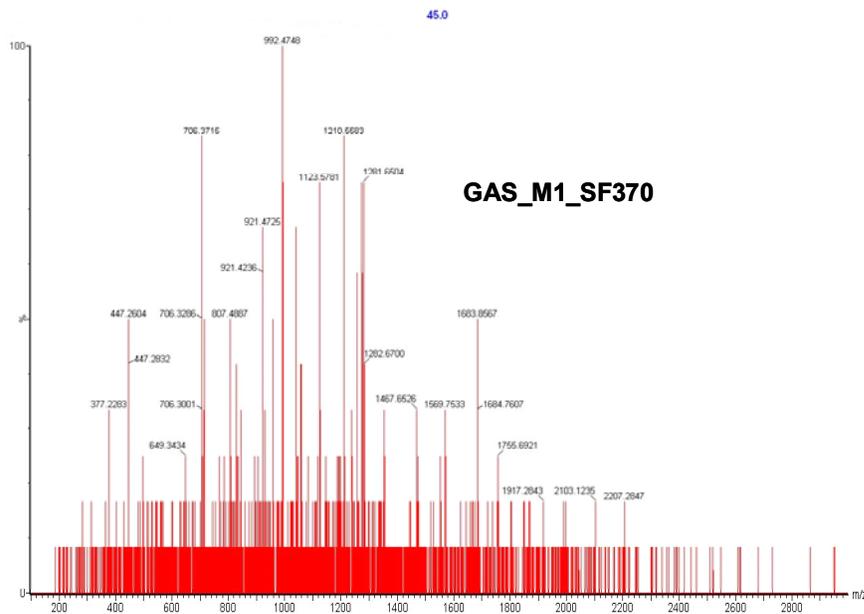
- il ceppo M1_3348 è un ceppo altamente capsulato; la capsula può diminuire l'efficienza della digestione come è stato dimostrato per il ceppo M23_DSM2071 (Rodriguez-Ortega et al, 2006)
- l'interpretazione degli spettri è stata eseguita usando informazioni del genoma del ceppo M1_SF370. Il genoma del ceppo M1_3348 non è commercialmente disponibile. E' possibile che alcuni peptidi specifici del ceppo M1_3348 non siano stati identificati.

Comunque la maggior parte dei peptidi identificati dal ceppo M1_3348 si sono trovati nel ceppo M1_SF370 indicando che le proteine e gli epitopi accessibili nel primo sono anche presenti sulla superficie del secondo.

Come esempio si riportano gli spettri di frammentazione del peptide **EKEQSQDLVSEPTTTELADTAASMANTGSDATQK** della GAS 57 isolato rispettivamente dal ceppo SF370 e 3348.

Spettro di frammentazione del peptide

EKEQSQDLVSEPTTTELADTDAASMANTGSDATQK del ceppo M1_SF370 e in tabella, la lista teorica dei frammenti del peptide, dove sono indicati in rosso i frammenti relativi allo spettro:



MS/MS Fragmentation of

EKEQSQDLVSEPTTTELADTDAASMANTGSDATQK

Found in **gij13621655**, putative cell envelope proteinase [Streptococcus pyogenes M1 GAS]

Match to Query 21: 3668.304372 from(1223.775400,3+) intensity(2156.5745) SF370\GAS_M1_SF370.pkl

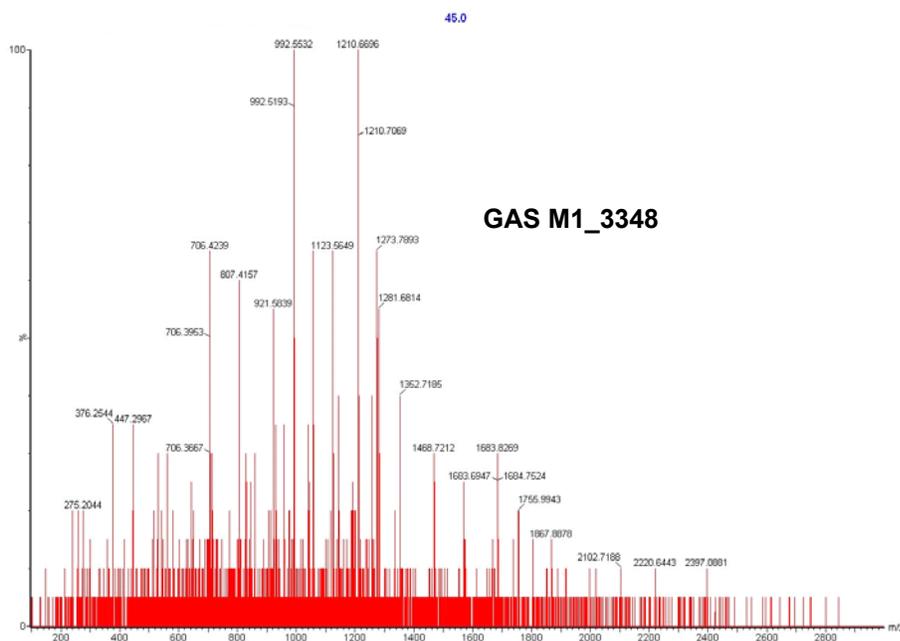
Monoisotopic mass of neutral peptide Mr(calc): 3668.65

Ions Score: 100 Expect: 5.5e-008

#	b	b ⁺⁺	b*	b ^{***}	b ⁰	B ⁰⁺⁺	Seq.	Y	y ⁺⁺	y*	y ^{***}	y ⁰	y ⁰⁺⁺	#
1	130.05	65.53			112.04	56.52	E							35
2	258.14	129.58	241.12	121.06	240.13	120.57	K	3540.61	1770.81	3523.59	1762.30	3522.60	1761.80	34
3	387.19	194.10	370.16	185.58	369.18	185.09	E	3412.52	1706.76	3395.49	1698.25	3394.51	1697.76	33
4	515.25	258.13	498.22	249.61	497.24	249.12	Q	3283.48	1642.24	3266.45	1633.73	3265.46	1633.24	32
5	602.28	301.64	585.25	293.13	584.27	292.64	S	3155.42	1578.21	3138.39	1569.70	3137.41	1569.21	31
6	730.34	365.67	713.31	357.16	712.33	356.67	Q	3068.38	1534.70	3051.36	1526.18	3050.37	1525.69	30
7	845.36	423.19	828.34	414.67	827.35	414.18	D	2940.33	1470.67	2923.30	1462.15	2922.32	1461.66	29
8	958.45	479.73	941.42	471.21	940.44	470.72	L	2825.30	1413.15	2808.27	1404.64	2807.29	1404.15	28
9	1057.52	529.26	1040.49	520.75	1039.51	520.26	V	2712.22	1356.61	2695.19	1348.10	2694.20	1347.61	27
10	1144.55	572.78	1127.52	564.26	1126.54	563.77	S	2613.15	1307.08	2596.12	1298.56	2595.14	1298.07	26
11	1273.59	637.30	1256.56	628.79	1255.58	628.29	E	2526.11	1263.56	2509.09	1255.05	2508.10	1254.56	25
12	1370.64	685.83	1353.62	677.31	1352.63	676.82	P	2397.07	1199.04	2380.05	1190.53	2379.06	1190.03	24
13	1471.69	736.35	1454.66	727.84	1453.68	727.34	T	2300.02	1150.51	2282.99	1142.00	2282.01	1141.51	23
14	1572.74	786.87	1555.71	778.36	1554.73	777.87	T	2198.97	1099.99	2181.95	1091.48	2180.96	1090.98	22
15	1673.79	837.40	1656.76	828.88	1655.78	828.39	T	2097.92	1049.47	2080.90	1040.95	2079.91	1040.46	21
16	1802.83	901.92	1785.80	893.40	1784.82	892.91	E	1996.88	998.94	1979.85	990.43	1978.87	989.94	20
17	1915.91	958.46	1898.89	949.95	1897.90	949.45	L	1867.83	934.42	1850.81	925.91	1849.82	925.42	19
18	1986.95	993.98	1969.92	985.47	1968.94	984.97	A	1754.75	877.88	1737.72	869.37	1736.74	868.87	18
19	2101.98	1051.49	2084.95	1042.98	2083.97	1042.49	D	1683.71	842.36	1666.69	833.85	1665.70	833.35	17
20	2203.02	1102.02	2186.00	1093.50	2185.01	1093.01	T	1568.69	784.85	1551.66	776.33	1550.68	775.84	16
21	2318.05	1159.53	2301.03	1151.02	2300.04	1150.52	D	1467.64	734.32	1450.61	725.81	1449.63	725.32	15
22	2389.09	1195.05	2372.06	1186.53	2371.08	1186.04	A	1352.61	676.81	1335.58	668.30	1334.60	667.80	14
23	2460.13	1230.57	2443.10	1222.05	2442.12	1221.56	A	1281.57	641.29	1264.55	632.78	1263.56	632.29	13
24	2547.16	1274.08	2530.13	1265.57	2529.15	1265.08	S	1210.54	605.77	1193.51	597.26	1192.53	596.77	12
25	2678.20	1339.60	2661.17	1331.09	2660.19	1330.60	M	1123.50	562.26	1106.48	553.74	1105.49	553.25	11
26	2749.24	1375.12	2732.21	1366.61	2731.23	1366.12	A	992.46	496.74	975.44	488.22	974.45	487.73	10
27	2863.28	1432.14	2846.25	1423.63	2845.27	1423.14	N	921.43	461.22	904.40	452.70	903.42	452.21	9
28	2964.33	1482.67	2947.30	1474.15	2946.32	1473.66	T	807.38	404.20	790.36	395.68	789.37	395.19	8
29	3021.35	1511.18	3004.32	1502.66	3003.34	1502.17	G	706.34	353.67	689.31	345.16	688.33	344.67	7
30	3108.38	1554.69	3091.35	1546.18	3090.37	1545.69	S	649.32	325.16	632.29	316.65	631.30	316.16	6
31	3223.41	1612.21	3206.38	1603.69	3205.40	1603.20	D	562.28	281.65	545.26	273.13	544.27	272.64	5
32	3294.44	1647.73	3277.42	1639.21	3276.43	1638.72	A	447.26	224.13	430.23	215.62	429.25	215.13	4
33	3395.49	1698.25	3378.46	1689.74	3377.48	1689.24	T	376.22	188.61	359.19	180.10	358.21	179.61	3
34	3523.55	1762.28	3506.52	1753.77	3505.54	1753.27	Q	275.17	138.09	258.14	129.58			2
35							K	147.11	74.06	130.09	65.55			1

Spettro di frammentazione del peptide

EKEQSQDLVSEPTTTELADTDAASMANTGSDATQK del ceppo M1_3348 e in tabella, la lista teorica dei frammenti del peptide, dove sono indicati in rosso i frammenti relativi allo spettro:



MS/MS Fragmentation of **EKEQSQDLVSEPTTTELADTDAASMANTGSDATQK** Found in **gi|13621655**, putative cell envelope proteinase [Streptococcus pyogenes M1 GAS] Match to Query 70: 3668.506872 from(1223.842900,3+) intensity(2838.7649) From data file M1_3348\GAS_M1_3348.pkl

Monoisotopic mass of neutral peptide Mr(calc): 3668.65

Ions Score: 137 Expect: 1e-011

Matches (Bold Red): 52/402 fragment ions using 104 most intense peaks

#	b	b ⁺⁺	b*	b ^{*++}	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y*	y ^{*++}	y ⁰	y ⁰⁺⁺	#
1	130.05	65.53			112.04	56.52	E							35
2	258.14	129.58	241.12	121.06	240.13	120.57	K	3540.6	1770.81	3523.5	1762.3	3522.6	1761.8	34
3	387.19	194.10	370.16	185.58	369.18	185.09	E	3412.5	1706.76	3395.4	1698.2	3394.5	1697.7	33
4	515.25	258.13	498.22	249.61	497.24	249.12	Q	3283.4	1642.24	3266.4	1633.7	3265.4	1633.2	32
5	602.28	301.64	585.25	293.13	584.27	292.64	S	3155.4	1578.21	3138.3	1569.7	3137.4	1569.2	31
6	730.34	365.67	713.31	357.16	712.33	356.67	Q	3068.3	1534.70	3051.3	1526.1	3050.3	1525.6	30
7	845.36	423.19	828.34	414.67	827.35	414.18	D	2940.3	1470.67	2923.3	1462.1	2922.3	1461.6	29
8	958.45	479.73	941.42	471.21	940.44	470.72	L	2825.3	1413.15	2808.2	1404.6	2807.2	1404.1	28
9	1057.52	529.26	1040.4	520.75	1039.5	520.26	V	2712.2	1356.61	2695.1	1348.1	2694.2	1347.6	27
10	1144.55	572.78	1127.5	564.26	1126.5	563.77	S	2613.1	1307.08	2596.1	1298.5	2595.1	1298.0	26
11	1273.59	637.30	1256.5	628.79	1255.5	628.29	E	2526.1	1263.56	2509.0	1255.0	2508.1	1254.5	25
12	1370.64	685.83	1353.6	677.31	1352.6	676.82	P	2397.0	1199.04	2380.0	1190.5	2379.0	1190.0	24
13	1471.69	736.35	1454.6	727.84	1453.6	727.34	T	2300.0	1150.51	2282.9	1142.0	2282.0	1141.5	23
14	1572.74	786.87	1555.7	778.36	1554.7	777.87	T	2198.9	1099.99	2181.9	1091.4	2180.9	1090.9	22
15	1673.79	837.40	1656.7	828.88	1655.7	828.39	T	2097.9	1049.47	2080.9	1040.9	2079.9	1040.4	21
16	1802.83	901.92	1785.8	893.40	1784.8	892.91	E	1996.8	998.94	1979.8	990.43	1978.8	989.94	20
17	1915.91	958.46	1898.8	949.95	1897.9	949.45	L	1867.8	934.42	1850.8	925.91	1849.8	925.42	19
18	1986.95	993.98	1969.9	985.47	1968.9	984.97	A	1754.7	877.88	1737.7	869.37	1736.7	868.87	18
19	2101.98	1051.49	2084.9	1042.9	2083.9	1042.4	D	1683.7	842.36	1666.6	833.85	1665.7	833.35	17
20	2203.02	1102.02	2186.0	1093.5	2185.0	1093.0	T	1568.6	784.85	1551.6	776.33	1550.6	775.84	16
21	2318.05	1159.53	2301.0	1151.0	2300.0	1150.5	D	1467.6	734.32	1450.6	725.81	1449.6	725.32	15
22	2389.09	1195.05	2372.0	1186.5	2371.0	1186.0	A	1352.6	676.81	1335.5	668.30	1334.6	667.80	14
23	2460.13	1230.57	2443.1	1222.0	2442.1	1221.5	A	1281.5	641.29	1264.5	632.78	1263.5	632.29	13
24	2547.16	1274.08	2530.1	1265.5	2529.1	1265.0	S	1210.5	605.77	1193.5	597.26	1192.5	596.77	12
25	2678.20	1339.60	2661.1	1331.0	2660.1	1330.6	M	1123.5	562.26	1106.4	553.74	1105.4	553.25	11
26	2749.24	1375.12	2732.2	1366.6	2731.2	1366.1	A	992.46	496.74	975.44	488.22	974.45	487.73	10
27	2863.28	1432.14	2846.2	1423.6	2845.2	1423.1	N	921.43	461.22	904.40	452.70	903.42	452.21	9
28	2964.33	1482.67	2947.3	1474.1	2946.3	1473.6	T	807.38	404.20	790.36	395.68	789.37	395.19	8
29	3021.35	1511.18	3004.3	1502.6	3003.3	1502.1	G	706.34	353.67	689.31	345.16	688.33	344.67	7
30	3108.38	1554.69	3091.3	1546.1	3090.3	1545.6	S	649.32	325.16	632.29	316.65	631.30	316.16	6
31	3223.41	1612.21	3206.3	1603.6	3205.4	1603.2	D	562.28	281.65	545.26	273.13	544.27	272.64	5
32	3294.44	1647.73	3277.4	1639.2	3276.4	1638.7	A	447.26	224.13	430.23	215.62	429.25	215.13	4
33	3395.49	1698.25	3378.4	1689.7	3377.4	1689.2	T	376.22	188.61	359.19	180.10	358.21	179.61	3
34	3523.55	1762.28	3506.5	1753.7	3505.5	1753.2	Q	275.17	138.09	258.14	129.58			2
35							K	147.11	74.06	130.09	65.55			1

2.4.2 Efficienza della digestione superficiale per selezionare vaccini

Delle 14 proteine identificate dalla digestione superficiale del ceppo M1_3348 tredici sono state anche selezionate dall'analisi genomica e analizzate mediante FACS per la loro localizzazione superficiale. Di queste tredici proteine analizzate, undici proteine risultano FACS positive, considerando come positivi i valori maggiori di 80 canali. Fra queste proteine si evidenzia la GAS 57, una proteina della parete cellulare, che risulta espressa sulla superficie dei 2 ceppi analizzati da cui è riportato il FACS (Figura 21):

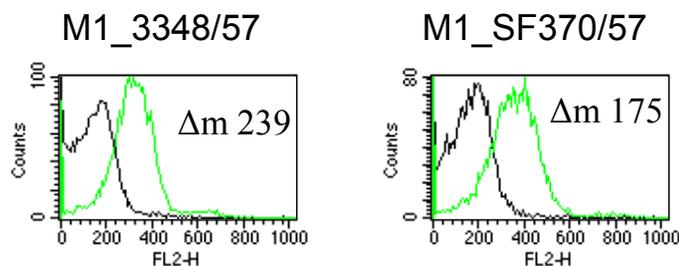


Fig. 21 Analisi citofluorimetrica del livello di esposizione dell'antigene GAS 57 sulla superficie di *S. Pyogenes*

Il livello di esposizione ($\Delta Mean$) della GAS 57 sulla superficie dei 2 ceppi analizzati è stato calcolato come la differenza tra il valore medio del picco dato dal siero immune (in verde nelle figure) e quello dato dal corrispondente siero preimmune (in nero nelle figure).

Inoltre ciascuna proteina, purificata come proteina ricombinante a partire dalla sequenza genomica del ceppo M1_SF370, è stata utilizzata per vaccinare topi. Dopo vaccinazione, i topi sono stati infettati con il ceppo M1_3348. Due proteine in particolare, sono risultate protettive in questo modello: la proteina M (90% di sopravvivenza) e la gas 57 (*putative cell envelope proteinase*) (70% di sopravvivenza).

Parte 3

2.5 Purificazione della proteina 57-His

In seguito all'identificazione di un vasto set di proteine di superficie, in accordo sia con i risultati provenienti dall'approccio proteomico che dall'approccio genomico il progetto del vaccino GAS è passato quindi in una fase in cui è stato necessario purificare una proteina, candidato potenziale; questa proteina presenta due caratteristiche fondamentali comuni ad entrambi i due approcci intrapresi:

- altamente espressa
- esposta sulla superficie del batterio

Lo scopo è stato quello di produrre la proteina in quantità maggiore e con elevata purezza.

La proteina 57-His è stata clonata ed espressa nei siti di restrizione del vettore d'espressione pET21b+ (Novagen) usato per trasformare cellule competenti di *E. Coli*. A monte del polylinker è presente una sequenza di istidine che serve per il recupero biochimico delle proteine ricombinanti tramite cromatografia di affinità.

Ogni purificazione parte dalla scelta delle colonne cromatografiche che saranno utilizzate per purificare la proteina in esame. La strategia è volta ad ottenere una purificazione completa mediante il minore numero di passaggi cromatografici. Questo è necessario per facilitare un futuro impiego della proteina nella produzione su larga scala; inoltre ha comportato la messa a punto di un protocollo di purificazione specifico, tenendo conto delle proprietà chimico fisiche come ad esempio il punto isoelettrico, l'idrofobicità (numero e distribuzione dei residui aminoacidici idrofobici), l'affinità per metalli o altri leganti e le dimensioni.

In base all'esperienza adottata in Novartis è stato possibile effettuare una strategia di purificazione come mostrato nella figura 22:

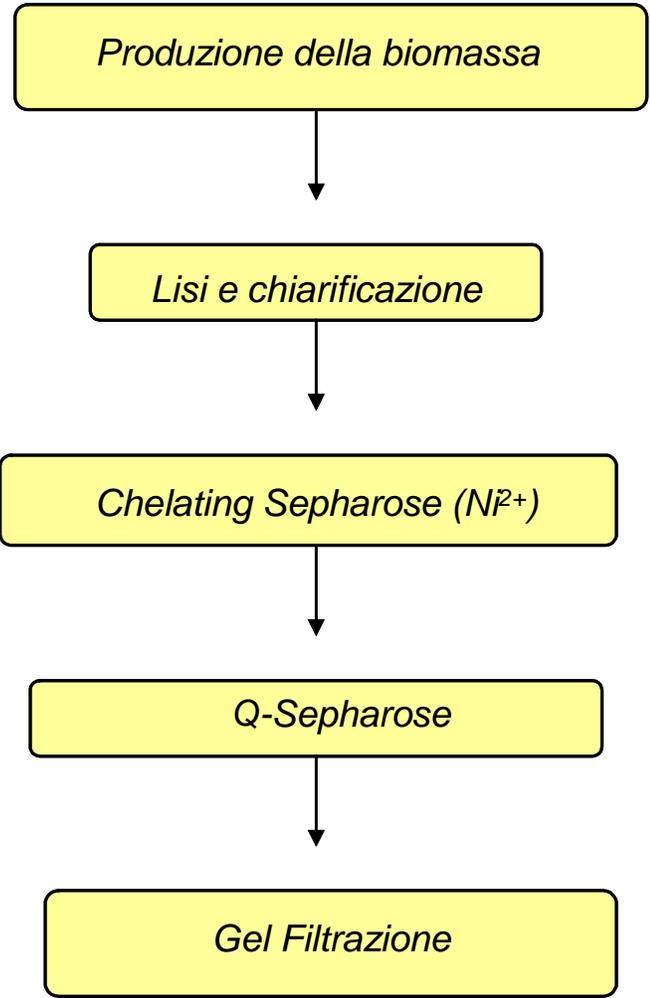


Fig. 22 Schema generale della strategia per la purificazione della proteina di interesse.

La biomassa (~12 g) è ripresa con 60 ml di B-PER contenente 10 mM imidazolo, 2 ml di deossiribonucleasi I (1000 unità Kunitz/ml), 150µl MgCl₂ 50 mM e 2 ml di lisozima (100 mg/ml) e 1 mM fenilmetilsulfonilfluoruro (PMSF) come inibitore di proteasi. La sospensione è mantenuta in agitazione per 60 minuti e centrifugata a 42500 g. Il primo passaggio di purificazione effettuato ha previsto una cromatografia di affinità (Chelating Sepharose) attivata con Ni²⁺ utilizzando un gradiente in imidazolo da 0 a 500 mM. Si identifica la proteina mediante analisi elettroforetica e si analizza il contenuto proteico effettuando il test Bradford, ottenendo 67.6 mg di proteine totali.

La proteina successivamente viene dializzata vs 30 mM Tris pH 8.0, utilizzando una membrana da 100000 MWCO come cut-off.

Il secondo passaggio di purificazione, noto il punto isoelettrico teorico, (pI = 5.85) consiste in una cromatografia a scambio anionico utilizzando una resina Q-Sepharose Fast Flow. La scelta dello scambiatore dipende dalla stabilità del campione, dal peso molecolare dei suoi componenti e dalle esigenze che la separazione presenta; si assume che il nostro campione sia più stabile a valori di pH superiori al suo punto isoionico, cioè quando ha una carica netta negativa e quindi si usa uno scambiatore di anioni. Nel nostro caso è stato utilizzato come scambiatore cationico una soluzione di Tris 30mM a pH 8.0. Si procede effettuando l'eluizione con gradiente in cui la transizione della concentrazione salina è lineare (dalla più bassa alla più alta), mediante una soluzione Tris 30 MM, 1 M NaCl. Questo passaggio è importante per rimuovere rapidamente ed efficacemente i contaminanti dalla nostra proteina.

Si identifica la proteina mediante analisi elettroforetica e quindi si analizza il contenuto proteico, mediante test Bradford, ottenendo 30.6 mg di proteine totali.

La proteina è stata poi ulteriormente purificata mediante passaggio cromatografico su gel filtrazione; questo è utilizzato come ultimo passaggio per rimuovere gli aggregati, i contaminanti più piccoli e per trasferire il campione purificato in un tampone ideale per i successivi studi. In base al peso della proteina (175,1 KDa) si sceglie la colonna maggiormente adatta; è stata utilizzata una HiLoad Superdex 200 26/60 che ci permette

di avere un'alta risoluzione con tempi di separazione relativamente brevi. Si conduce una corsa cromatografica isocratica in tampone PBS a pH 7.4

Anche in questo caso la scelta delle frazioni da raccogliere in pool, contenenti la proteina, viene effettuata analizzandole mediante analisi elettroforetica e si dosa il contenuto proteico con il metodo Bradford.

Si riportano in figura 23 l'analisi elettroforetica relativa a ciascun passaggio cromatografico effettuato:

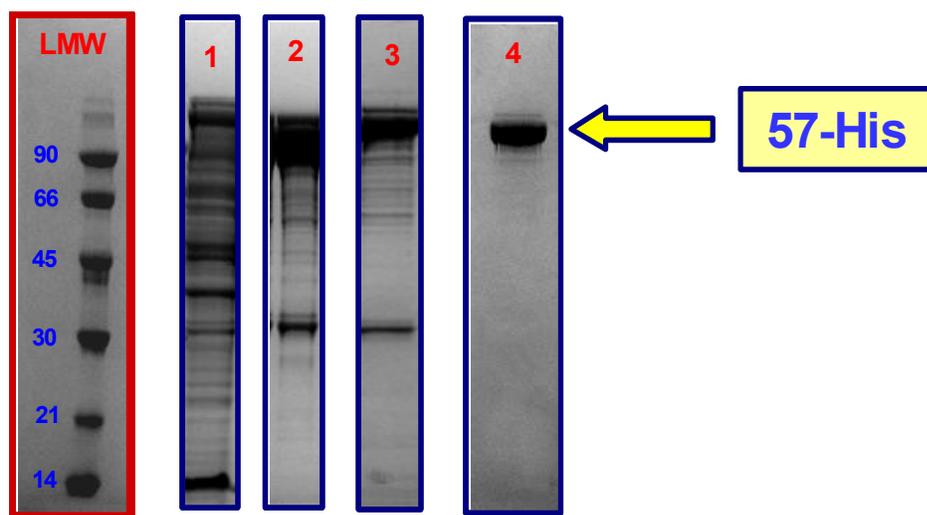


Fig. 23: analisi elettroforetica della proteina 57-His; LMW: markers; 1): materiale lisato e chiarificato; 2) cromatografia di affinità; 3) scambio anionico 4) gel filtrazione

In tabella 6 si riportano i vari passaggi cromatografici effettuati: il dosaggio proteico, le quantità totali ottenute e la purezza percentuale stimata mediante un programma sviluppato in Novartis basato sulla densitometria del gel come descritto precedentemente nella sezione Analisi della purezza delle proteine purificate (2.2):

Tab. 6 Schema riassuntivo dei vari steps previsti dalla strategia di purificazione adottata.

	Passaggio Purificazione	Volume (ml)	Dosaggio Proteico (mg/ml)	Proteine totali (mg)	Purezza Percentuale (%)
1	Chiarificazione	63	9.750	----	-----
2	Affinità su Ni²⁺	20	3.378	67.6	31
3	Scambio ionico	10	3.060	30.6	52
4	Gel filtrazione	20	0.842	16.8	88

Da come si evince dai dati riportati in tabella la strategia di purificazione ha condotto a buoni risultati sia dal punto di vista quantitativo che qualitativo (vedere tabella 6). Tale proteina è stata pertanto testata nel modello murino, messo a punto in Novartis, come possibile vaccino, conferendo un livello di protezione dall'infezione del batterio del 70%.

Discussione e conclusioni

Streptococcus pyogenes è uno dei batteri patogeni più frequenti nell'uomo. Ad oggi le infezioni provocate da questo microrganismo rimangono un grave problema per la salute pubblica. La terapia in caso di infezione da *Streptococcus pyogenes* consiste, sin dal 1940, nell'utilizzo della penicillina, usata anche per prevenire il riaccendersi di episodi di febbre reumatica e di glomerulonefrite (per rivista vedere Cunningham, 2000). Manca però una terapia preventiva e ciò ci ha quindi spinto alla ricerca di un nuovo vaccino.

In questo lavoro di dottorato, che rientra in un ampio progetto di ricerca è stato identificato un set di proteine di superficie, potenziali antigeni di GAS. Tra queste, in particolare sono state trovate due proteine, che quando usate per immunizzare topi, conferissero loro un riproducibile livello di protezione in caso di una successiva infezione con il patogeno stesso. Gran parte del lavoro effettuato nella prima parte è stato indirizzato, in accordo con la reverse vaccinology, a partire dal genoma sequenziato M1_SF370, alla ricerca di una serie di proteine ricombinanti esposte in superficie; ovviamente è stato necessario effettuare uno *screening* complicato e completo di un numero elevato di proteine. A tal fine sono state clonate, espresse e purificate; proprio quest'ultimo passaggio rappresenta gran parte del lavoro di dottorato, in cui è stato necessario avvalersi di strumenti automatizzati per effettuare la purificazione di un gran numero di proteine.

E' stato possibile ottenere con successo, a partire da più di 600 geni selezionati, un totale di 434 proteine purificate e usate per immunizzare topi. La successiva analisi al FACS ha comportato l'identificazione di 86 proteine di superficie a partire da 186 campioni analizzati. Questi dati immunologici quindi sono stati complementati dall'utilizzo della proteomica, applicando la digestione superficiale al ceppo M1_3348 e identificando 14 proteine esposte in superficie. Di queste, 13 sono state analizzate al FACS e 11 sono risultate positive (86%) confermando la loro esposizione sulla superficie batterica. (vedi Figura 24). Per queste ragioni un approccio sperimentale che

tenda a stabilire quali proteine siano veramente espresse sulla superficie del batterio, risulta imprescindibile.

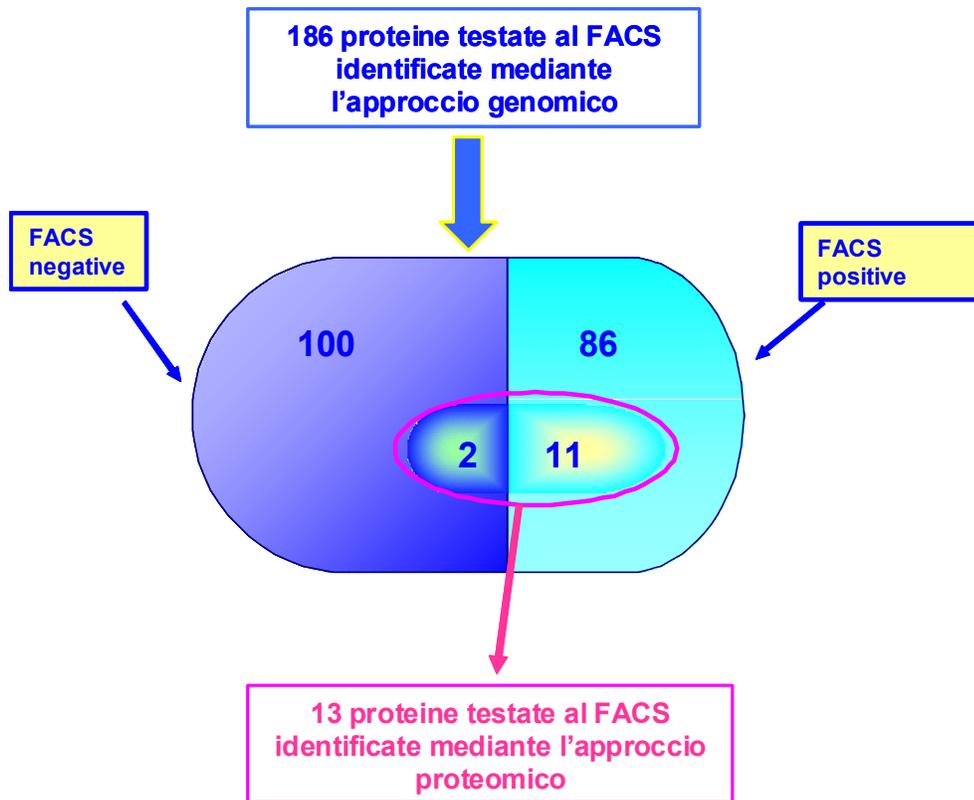


Fig. 24 Proteine di superficie comuni tra l'applicazione dell'approccio genomico e proteomico

Si può dire che a partire da 600 geni (435 proteine) per il primo approccio contro 14 proteine ritrovate mediante il secondo, è stato possibile identificare due proteine comuni che, testate nel modello animale, conferissero un elevato grado di protezione:

- Proteina M (90%)
- Gas 57 (70%).

Inoltre nella comparazione tra i due metodi i dati provenienti dall'approccio proteomico risultano numericamente inferiori rispetto a quelli ottenuti mediante l'approccio genomico; ciò può essere dovuto a due fattori:

1. i risultati ottenuti mediante digestione superficiale sono tuttora in una fase investigativa e i risultati qui riportati sono dati preliminari.

- alcune proteine identificate non possiedono dei siti di taglio facilmente riconoscibili all'azione proteolitica; la struttura di alcune proteine può opporre resistenza all'azione proteolitica qualora questa risultasse molto compatta e difficilmente penetrabile.

Contrariamente, a fronte di una rapida e sufficientemente affidabile selezione dei candidati, il limite dell'approccio *in silico* risiede nell'impossibilità di stabilire l'effettiva espressione di un antigene da parte del microrganismo, l'abbondanza dell'antigene in questione e la fase di crescita-infezione in cui tale antigene venga espresso. Inoltre, per quanto i sistemi di predizione impieghino algoritmi evoluti, si possono riscontrare incertezze sulla effettiva localizzazione superficiale di alcuni antigeni. Da qui ne scaturisce l'efficienza e l'efficacia del metodo proteomico; infatti il grosso e tedioso lavoro effettuato per l'approccio genomico comporta un lavoro sì minuzioso, ma piuttosto lungo; al contrario il *pattern* di proteine selezionate nell'approccio proteomico è sicuramente minore come mostrato in figura 25.

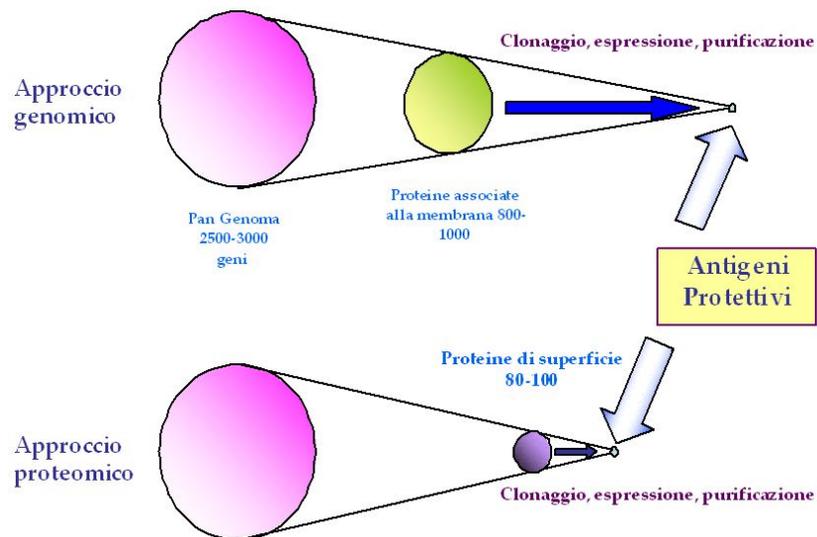


Fig. 25 Paragone tra approccio genomico e proteomico

Dall'analisi in parallelo delle due strategie adottate e in seguito alla complementazione dei risultati, è stato individuato un nuovo antigene proteico che ha dimostrato di poter indurre un elevato grado di protezione in un appropriato sistema

sperimentale *in vivo*, la GAS 57-His. La nostra analisi si è quindi concentrata sulla proteina GAS 57-His che, nel nostro modello murino di infezione, ha riproducibilmente conferito il più alto grado di protezione (70%). Questo risultato ci ha spinto a verificare che questa proteina avesse tutte quelle caratteristiche che un antigene proteico dovrebbe avere per essere incluso nella formulazione finale di un vaccino. Fra queste, la conservazione della sequenza proteica, l'espressione in un discreto numero di ceppi diversi e la localizzazione sulla superficie del batterio, che rende l'antigene più disponibile al riconoscimento da parte di costituenti del sistema immunitario, sono forse i fattori più importanti nell'ottica della definizione di un futuro vaccino. L'immunizzazione attiva di topi con la proteina ricombinante GAS 57 ha dimostrato che essa è in grado di stimolare la produzione di anticorpi che inibiscono in modo significativo la crescita del batterio nel sangue di topi immunizzati. Anche queste proprietà risultano essere estremamente importanti ai fini dello sviluppo di un efficace vaccino, poiché è noto che molti patogeni hanno sviluppato meccanismi specifici per evadere il complemento, arma fondamentale del sistema immunitario, la cui azione è probabilmente rilevante nelle due metodiche sperimentali precedentemente citate (Jarva et al., 2003).

In questo contesto, è ancora una volta importante ricordare che la GAS 57 è, ad oggi, una delle rare proteine derivate da GAS, insieme alla proteina M e a poche altre, in grado di evidenziare proprietà di questo tipo. Nell'ottica di delineare un vaccino proteico in grado di proteggere dall'infezione di molti ceppi di GAS diversi, è ragionevole supporre che un futuro vaccino anti-GAS sarà costituito da almeno 3-4 proteine, ognuna in grado di conferire protezione verso determinati sierotipi. Una strategia che si sta rivelando appropriata in altri progetti simili al nostro, è quella di costruire delle proteine di fusione formate dall'unione fisica di più proteine, o di porzioni immunologicamente importanti di esse, con lo scopo finale di ottenere un'unica proteina ricombinante che conservi le proprietà protettive delle singole componenti, conferendo una protezione ad ampio spettro.

Materiali e metodi

4.1 Ceppi di GAS utilizzati

Nel nostro studio abbiamo utilizzato diversi ceppi capsulati di *S. pyogenes*. Il ceppo altamente virulento M1_3348 principalmente usato per verificare la protezione nel nostro modello murino e per le digestioni superficiali in proteomica; M1_SF370 è stato utilizzato per effettuare saggi analitici e saggi *in vivo* e/o *in vitro*. Inoltre dal ceppo M1_SF370 è stato estratto il DNA per il clonaggio di tutte le proteine ricombinanti utilizzate nel nostro *screening*.

4.2 Terreni e tamponi

4.2.1 Terreno di coltura LB (Luria Bertani) per la crescita di cellule batteriche:

Per un litro di terreno pesare:

Tryptone	10,0 g
Yeast extract	5,0 g
NaCl	10,0 g

- sciogliere in circa 800 ml di H₂O;
- portare a pH 7,6 con NaOH 1 M;
- aggiungere H₂O fino ad un litro di soluzione;
- sterilizzare il terreno preparato in autoclave a 121°C.

4.2.2 Terreno di coltura THB per la crescita dello *S.pyogenes*

Per un litro di terreno pesare:

Todd Hewitt Broth	30,0 g
-------------------	--------

- sciogliere in circa 800 ml di H₂O;
- aggiungere H₂O fino ad un litro di soluzione;
- sterilizzare il terreno preparato in autoclave a 121°C.

4.2.3 Terreno di coltura THYE per la crescita dello *S.pyogenes*

Per un litro di terreno pesare:

Todd Hewitt Broth	30,0 g
Yeast extract	5,0 g

- sciogliere in circa 800 ml di H₂O;
- aggiungere H₂O fino ad un litro di soluzione;
- sterilizzare il terreno preparato in autoclave a 121°C.

Per preparare il terreno solido, di qualunque composizione, è necessario aggiungere agar nobile (DifcoTM), ad una concentrazione finale di 1,5%, al terreno prima di autoclavarlo. Per le piastre con THYE e THB viene aggiunto il 5% di sangue di montone dopo la sterilizzazione.

Nel caso sia necessario un terreno selettivo si aggiunge alla soluzione già autoclavata l'antibiotico alla concentrazione opportuna e si conserva a 4°C. Nel nostro lavoro abbiamo utilizzato 1 terreno selettivo: -LB + ampicillina (100 µg/ml).

4.2.4 Tampone PBS (*Dulbecco's Phosphate Buffered Saline*)

Per un litro di terreno pesare:

KCl	0,20 g
KH ₂ PO ₄	0,20 g
NaCl	8,0 g
Na ₂ HPO ₄	1,15 g

- sciogliere in circa 800 ml di H₂O;
- aggiungere H₂O fino ad un litro di soluzione;
- sterilizzare il tampone preparato in autoclave a 121°C.

4.3 Clonaggio ed espressione delle proteine ricombinanti dello *S. pyogenes*

4.3.1 Vettori plasmidici

Vengono preparati due *primers* sintetici in modo tale da indurre a monte e a valle del tratto da amplificare due brevi sequenze che, riconosciute come siti di taglio da specifici enzimi di restrizione, risultano fondamentali per il clonaggio.

Il DNA, amplificato e purificato, è stato inserito, attraverso la reazione di ligasi, nei siti di restrizione dei vettori d'espressione pET21b+ (Novagen) e pGEX (Amersham Biosciences). Il primo (Fig. 26), formato da 5442 bp, viene usato per trasformare cellule competenti di *E. coli* BL21(DE3), presenta una resistenza per l'ampicillina ed è regolato dal promotore T7. A monte del *polylinker* è presente una sequenza di Istidine che serve per il recupero biochimico delle proteine ricombinanti tramite colonne cromatografiche.

Il secondo (Figura 27) invece un vettore di 4900 bp, usato per trasformare cellule competenti di *E. coli* BL21.

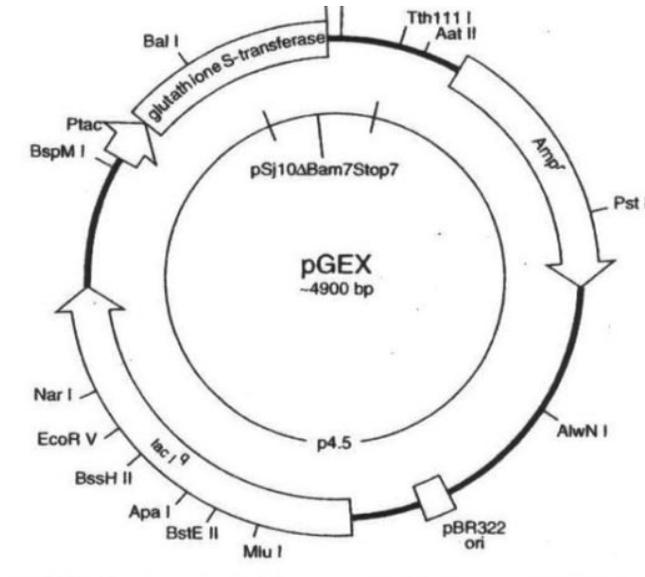


Fig. 27: pGEX (Amersham biosciences).

Il pGEX usato in questo lavoro, inoltre, ha il *polylinker* differente da quello commercializzato, è stato infatti ingegnerizzato dai ricercatori del Centro Ricerche della Novartis Vaccines introducendo il *polylinker* del vettore pET per poter clonare inserti digeriti con gli stessi enzimi di restrizione in entrambi i vettori. Presenta inoltre la Glutathione S-trasferasi che serve per il recupero biochimico delle proteine.

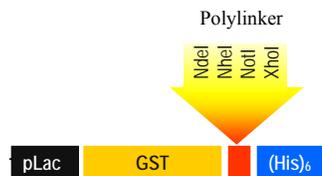


Fig. 28: Polylinker di pGEX ingegnerizzato da Novartis Vaccines.

4.3.2 Ligazione e trasformazione

Il prodotto di ligasi viene trasformato, tramite elettroporazione, nei ceppi di *E. coli*, resi competenti precedentemente descritti. 1 μ l di ligazione è stato aggiunto, quindi, a 40 μ l di cellule scongelate in ghiaccio, la soluzione è stata poi trasferita in una cuvette ed elettroporata alle seguenti condizioni: 25 mF, 200 W, 2000 volts. Immediatamente dopo le cellule vengono risospese in 1 ml del tampone SOC e lasciate per 45 min. a 37°C in rotazione.

4.3.3 Selezione dei cloni positivi

I cloni positivi vengono selezionati inoculando i batteri su piastra LB contenente ampicillina - antibiotico nei confronti del quale i plasmidi conferiscono resistenza - e verificando la presenza dell'inserto voluto tramite PCR su colonia con oligonucleotidi interni al plasmide e fiancheggianti la sequenza amplificata dal genoma.

Per amplificare il gene *gas57* su tutti i ceppi disponibili sono stati utilizzati i successivi primers, uno senso e uno antisenso:

- 57F GTGCGTCATATGGCAGATGAGCTAAGCA
- 57R GCGTCTCGAGGGCTTTTGCTGTTG

Tali primer sono stati disegnati a partire dalla sequenza pubblicata per la proteina GAS 57 in NCBI (www.ncbi.nlm.nih.gov/)

4.3.4 Espressione

I cloni positivi sono stati inoculati in 2 ml di terreno liquido (LB + ampicillina). Dopo la crescita *ON* a 37°C, ogni coltura viene diluita in 2 ml di terreno fresco e incubata fino al raggiungimento di un valore di OD₆₀₀ compreso tra 0,4 e 0,8. A questo punto l'espressione della proteina viene indotta aggiungendo isopropyl- β -D-thiogalactopyranoside (IPTG) 1 mM per il vettore pET21b+ e 0,2 mM per il vettore

pGEX. I batteri dopo 3-4 ore di incubazione sono stati centrifugati e il pellet conservato a -20°C .

4.4 Purificazione in colonna di affinità

4.4.1 Proteine His-tag da *E. coli*

Questa procedura permette di purificare elevate quantità di proteina e, contemporaneamente, ridurre i tempi di lavoro, grazie all'impiego della Cromatografia di Affinità con Ioni Metallo immobilizzati (acronimo inglese: IMAC). Si tratta in genere di ioni Nichel (Ni^{2+}) che tramite chelazione con i gruppi reattivi sono covalentemente attaccati ad un supporto solido. Per legare tale metallo alla resina si prepara una soluzione 0.1-0.2M di NiCl_2 a pH 7 o leggermente acido.

La proteina ottenuta mediante questa procedura deve essere dotata di una coda di 6 istidine. La resina (Chelating Sepharose Fast Flow) è costituita da gruppi di acido iminodiacetico (chelante) accoppiati, con legami etere stabili, a spaziatori a 7 atomi di carbonio fissati a loro volta ad una resina di agarosio. La resina, il cui volume è circa 1 ml, viene impaccata a temperatura ambiente. Per velocizzare le operazioni di lavaggio ed eluizione si può utilizzare una pompa peristaltica, regolandone il flusso a 0.5 ml/min; la stessa velocità di flusso deve poi essere mantenuta per tutti i vari passaggi in colonna.

Una volta impaccata e equilibrata la colonna, vi è caricato il lisato totale batterico, contenente la proteina da purificare.

Il *pellet* ottenuto dopo l'induzione, di 3h e 30' con IPTG 1 mM di una coltura batterica di 500 ml, viene risospeso in 10ml di Tampone di Lisi.

- Si incuba per 40', vortexando di frequente;
- L'estratto totale è centrifugato 20' a 23500 rpm;
- il surnatante è caricato in colonna di affinità;

- si lava la colonna con 2 volumi e successivamente due volte con 4 volumi di Tampone L.
- L'eluizione avviene in tre passaggi, con il Tampone E.
- Le 3 frazioni raccolte vengono addizionate con DTT (concentrazione finale 2 mM)

Tampone di Lisi: 10ml B-PER™(Bacterial-Protein Extraction Reagent, Pierce cat.782266)

- 10µl di una soluzione di MgCl₂ 100mM (concentr. Finale 1mM);
- 50µl di DNasi I (Sigma D-4263) 100 unità in PBS (1000 unità Kunitz/ml);
- 100µl di una soluzione di lisozima (Sigma L-7651) 100 mg/ml in PBS (concentrazione finale 1 mg/ml).

Tampone L di lavaggio:

- 20 mM imidazolo, 50 mM NaH₂PO₄, 300 mM NaCl, pH 8,0.

Tampone E di eluizione:

- 250 mM imidazolo, 50 mM NaH₂PO₄, 300 mM NaCl, pH 8,0.

Cromatografia d'affinità per le proteine His-tag

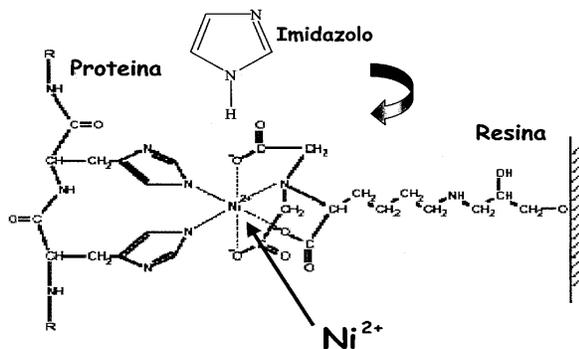


Fig. 29 Cromatografia di affinità per le proteine His-Tag.

Resina immobilizzata con ioni Ni²⁺ che formano un complesso di coordinazione con 2 anelli imidazolici appartenenti a residui di istidine della proteina legante.

4.4.2 Proteine His -Tag da Corpi Inclusi di *E. coli*

Si è applicato ai pellet la stessa procedura di lisi riportata per la purificazione della -His tag solubile. I pellet lisati sono stati centrifugati a 23500 rpm con rotore JA25.50 in centrifuga Boeringham per 20 minuti.

I pellet sono stati sciolti con Tampone G.

La procedura di purificazione è condotta a temperatura ambiente con colonnine Poly-Prep, equilibrate con 1 ml di resina Ni-Activated Chelating Sepharose Fast Flow. Le proteine legate alla colonna sono state eluite con 3 passaggi di Tampone E'. Le 3 frazioni raccolte vengono addizionate con DTT (concentrazione finale 2 mM). Il dosaggio della concentrazione proteica delle frazioni è stata eseguita con il metodo Bradford. Un'aliquota da 10 µg della proteina è stata analizzata su SDS-PAGE.

- Tampone G : 50 mM TRIS, 1Mm TCEP, 6 M Guanidinio, pH 8,5.
- Tampone E': 250 mM imidazolo, 50 mM TRIS, 6 mM Urea, 1 mM TCEP, pH 8,5.

4.4.3 Proteine GST-Tag da *E. coli*

Le cellule vengono lisate con lo stesso metodo usato per le proteine His. Il lisato viene chiarificato per centrifugazione e quindi caricato su colonnine Poly-Prep contenenti resina Glutathione-Sepharose FF equilibrata con PBS (pH 7,4). Si elimina parte delle impurità lavando la colonna con tampone PBS (pH 7,4).

Infine si eluiscono le proteine con tampone TRIS 50 mM + Glutathione ridotto 10 mM (fornito dalla SIGMA) (pH 8.0).

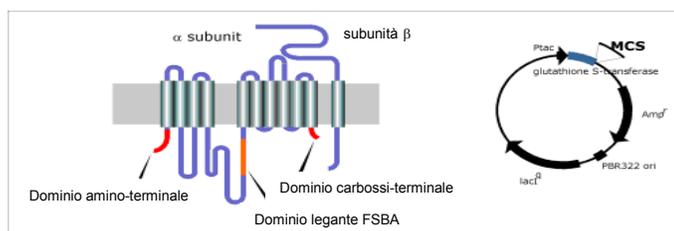


Fig. 30: Descrizione della proteina Glutathione-S-transferasi.

4.5 Determinazione della concentrazione proteica

(Bradford.MM. "A rapid and sensitive for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding". *Analytical Biochemistry* 72:248-254.1976.)

La concentrazione delle proteine è stata determinata spettroscopicamente seguendo la legge di **Lambert-Beer** e usando il coefficiente di estinzione per il colorante Comassie Brilliant Blue G-250.

Legge di Lambert-Beer:

$$C [g/100 ml] = (OD / E^{1\%}_{1 cm}) \times b$$

$$b (\text{cammino ottico}) = 1 \text{ cm}$$

Il saggio di Bradford è un dosaggio colorimetrico basato sull'osservazione che il massimo di assorbanza per la soluzione acida del colorante (100mg di Comassie Brilliant Blue G-250 in 50 ml di Etanolo al 95% e 100 ml di Acido Fosforico all'85% (v/v) in un volume totale di 1L) passa da una lunghezza d'onda (λ) di 465 ad una di 595 nm, quando si realizza il legame con gli amminoacidi basici e aromatici presenti nella proteina. Il BBC reagisce principalmente con i residui di arginina, meno con istidina, lisina, tirosina, triptofano e fenilalanina. La sensibilità del metodo abbraccia un range di concentrazione proteica che va dai 20 ai 200 μg . La concentrazione delle proteine è stimata in riferimento ad una curva standard di BSA (Siero Albumina Bovina), o di BGG (Gamma Globulina Bovina), usata come proteina campione.

Viene preparata una serie di diluizioni della proteina standard a concentrazioni note; di tali diluizioni: soluzione proteica + tampone 5mM fosfato, 30mM NaCl, pH 8, viene misurata allo spettrofotometro la corrispondente assorbanza. In un sistema Cartesiano, in cui è riportata la concentrazione delle proteine sull'asse delle ascisse (X) e la loro assorbanza sull'asse delle ordinate (Y), si traccia una retta di regressione che passa per i punti caratterizzati dalle coppie Concentrazione/Assorbanza.

Si procede successivamente a calcolare l'assorbanza dei campioni in studio e mediante la curva standard è calcolata la loro concentrazione ignota.

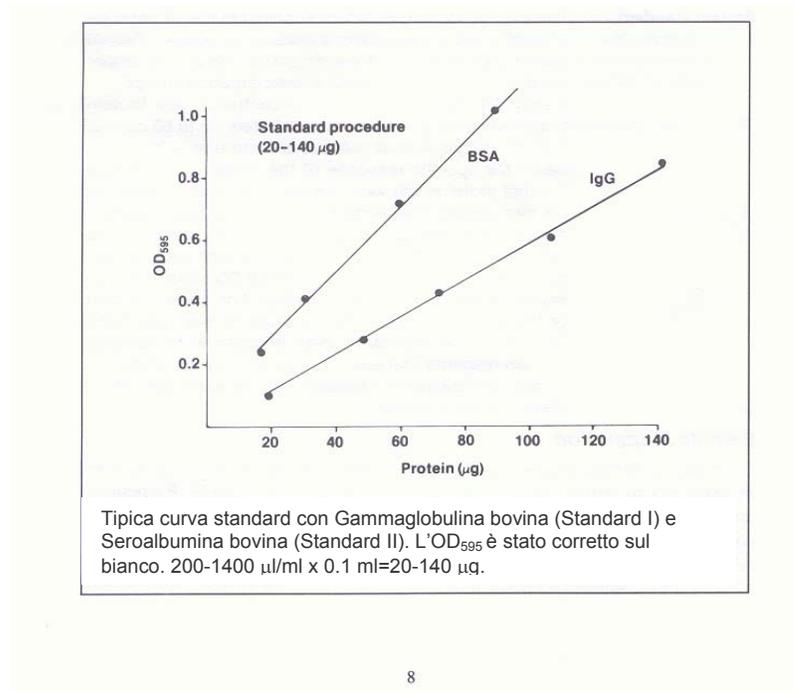


Fig. 31: Determinazione della curva standard effettuata per diluizioni seriali di γ GB (γ Globulina Bovina) Standard I e BSA (Siero Albumina Bovina) Standard II.

4.6 Analisi elettroforetica su gel di poliacrilamide sodio dodecilsolfato (SDS-PAGE)

Una volta avvenuta l'espressione in *E. coli*, le proteine sono state analizzate in gel di poliacrilamide in condizioni denaturanti.

Solitamente viene utilizzato il detergente anionico SDS (Sodio Dodecil Solfato) in associazione con agenti riducenti (DTT) e calore (95°C per 5 min.) per denaturare le proteine prima di essere caricate su gel.

L'SDS ha una coda idrofobica che interagisce fortemente con le catene polipeptidiche e quindi il numero di molecole che si lega alla proteina è proporzionale al

numero di aminoacidi che la formano. Ogni molecola di SDS conferisce due cariche negative e rompe le forze che contribuiscono al ripiegamento della proteina assicurando non solo l'uniformità di carica ma anche la linearità, in questo modo le proteine possono essere separate solo in base alla loro grandezza. Il complesso SDS-polipeptide migra lungo il gel di poliacrilamide in base alla grandezza del polipeptide stesso; utilizzando dei marker di peso molecolare noto è possibile stimare il peso molecolare della proteina e/o individuarla in un estratto totale.

I gel utilizzati per i nostri esperimenti sono gel alla concentrazione del 12% di bis/tris poliacrilamide.(Criterion Precast-Biorad). La corsa elettroforetica avviene con il gel immerso nel tampone di corsa (Mops 20X-Biorad) ad un voltaggio costante (200 V) per 55 min.

Dopo la separazione delle proteine in SDS-PAGE i gel sono colorati per immersione in una soluzione di Coomassie (*staining solution*), con cui si colorano direttamente tutte le proteine separate nel gel.

La composizione della soluzione per la colorazione è:

- Coomassie R250 0,2%
- Etanolo 20%
- Acido Acetico glaciale 5%
- H₂O 75%

Dopo la preparazione la soluzione deve essere filtrata con carta assorbente per eliminare i residui di colorante indisciolto.

Per togliere l'eccesso di colorante i gel sono immersi in una soluzione decolorante (*destaining solution*) formata da:

- Etanolo 20%
- Acido Acetico 5%
- H₂O 75%.

Si utilizza come standard il liofilo per elettroforesi SDS (Amersham Biosciences).

Si riportano i valori standard:

Proteina	PM (Da)
Fosforilasi B	97000
Albumina	66000
Ovoalbumina	45000
Anidrasi carbonica	30000
Inibitore di tripsina	20100
α -lattoalbumina	14400

4.7 Analisi d'immagine

L'immagine dei gel bidimensionali è stata acquisita mediante Personal Densitometer SI (Molecular Dynamics) a 12 bits e 50 μm /pixel di risoluzione.

L'analisi di sovrapposizione d'immagine (utilizzata sia per il confronto tra le frazioni solubile-insolubile in sodio carbonato che per la sovrapposizione delle mappe dei due vaccini OMV-based), è stata operata grazie al software ImageMaster 2D (Amersham Pharmacia Biotech), in grado di effettuare un confronto delle immagini digitalizzate dei gel bidimensionali tramite un procedimento di sovrapposizione/compensazione delle distorsioni locali (matching).

4.8 Bioinformatica

La predizione di localizzazione proteica è stata effettuata utilizzando Psort, disponibile al sito <http://psort.nibb.ac.jp> (Nakai, 1991).

4.9 Digestione superficiale dei batteri

M1_3348 sono cresciuti in THB a 37° C e in atmosfera al 5 % di CO₂, fino ad un valore di OD₆₀₀ di 0.4 (fase esponenziale). I batteri vengono centrifugati a 3500 g per 10 minuti a 4°C, e lavati 3 volte con PBS. Le cellule vengono risospese in un tampone PBS/saccarosio 40% pH 7.4 se digerita con la tripsina; se la digestione è effettuata con proteinasi K il tampone è PBS/saccarosio 40% pH 6.0.

In un'aliquota di ciascun ceppo si aggiungono 5 µg (10 µl) di tripsina (Promega) a partire da una soluzione 500 mM (1/100); in un'altra aliquota si aggiungono 5 µg di Proteinasi K (Sigma) in presenza di 5 mM DTT; le aliquote si lasciano a 37°C per 30 min.

Le colture digerite vengono centrifugate a 3500 g per 10 min a 4 °C e i supernatanti vengono filtrati su 0.22 µm (Millipore).

L'azione delle proteasi viene terminata mediante l'aggiunta di acido formico ad una concentrazione finale di 0.1%.

C'è da notare che la digestione deve essere condotta in un ambiente tale che, possa minimizzare la lisi citoplasmatica e che possano esservi in maggioranza proteine superficiali.

Una volta avvenuta la digestione è necessario analizzare e identificare le proteine superficiali.

I campioni sono caricati su colonnine Oasis (Millipore) per un passaggio cromatografico di desalificazione descritto nella procedura riportata a seguire.

4.9.1 Protocollo di desalificazione

1. Condizionare la colonna con 80% ACN
2. Equilibrare con H₂O/0.1% acido formico
3. Caricare il campione (circa 400 µl); (effettuare questo passaggio 2 volte); raccolta del FT
4. lavaggi con 2% ACN/0.1% acido formico
5. eluizione
 - 10% ACN -0.1% Acido formico
 - 20% ACN -0.1% Acido formico
 - 50% ACN -0.1% Acido formico
 - 80% ACN -0.1% Acido formico

In toto si raccolgono per ciascun aliquota 7 frazioni; successivamente si concentra in Speedvac centrifuge (Savant) fino ad un volume di 20-30µl e si conservano a -20°C.

4.10 Ricerca in banca dati

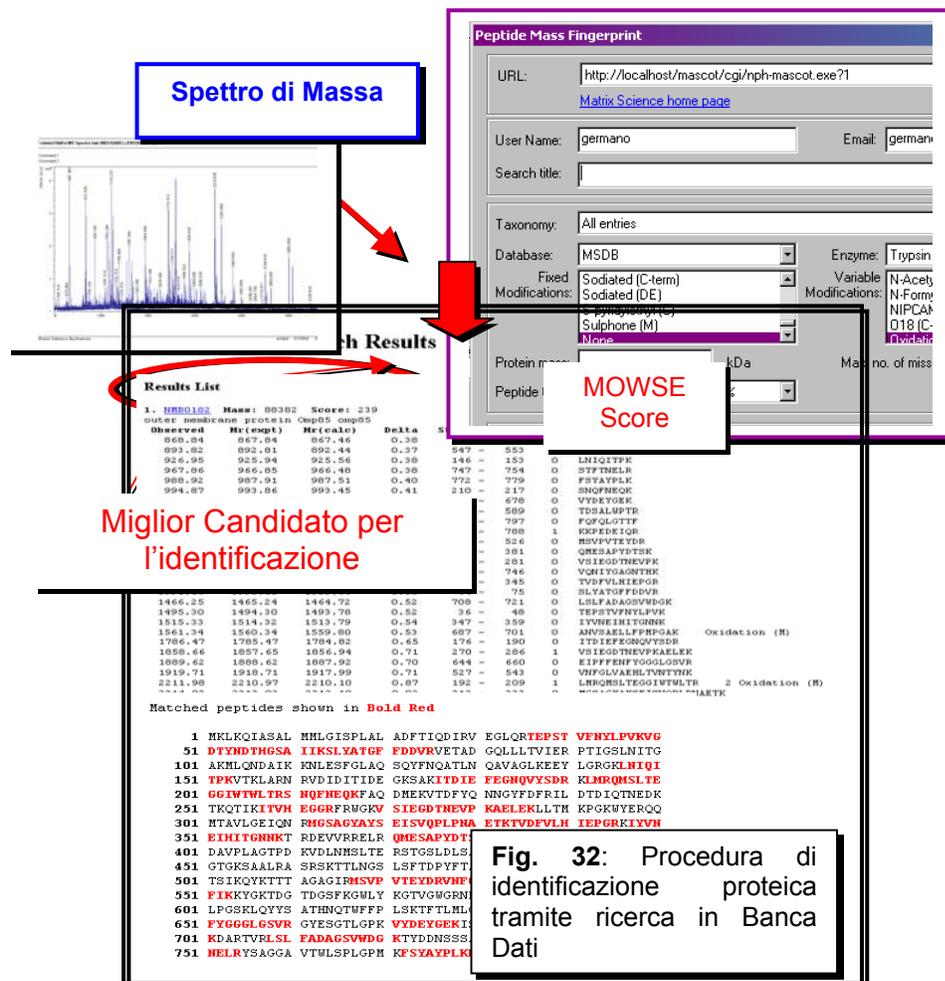
Una volta acquisito lo Spettro di Massa, adeguati software permettono di effettuare il *labelling* dei picchi, attribuendo il valore m/z ai peptidi-ioneH⁺ (XMASS, XTOF).

L'insieme dei valori delle masse costituisce il Fingerprint sperimentale, da confrontare con quelli teorici derivanti dalla disponibilità del genoma (**Fig. 32**).

Il sistema può utilizzare sia banche dati di proprietà che personalizzabili, che banche dati pubbliche (Ncbl, SwissProt).

Il software utilizzato per la ricerca è MASCOT (www.matrix-science.com), il quale permette di inserire nel confronto modificazioni di vario genere, nonché stabilire parametri di restrizione, scelta del database e della proteasi impiegata per la digestione.

Mascot fornisce un output come quello riportato in **Fig. 32**; l'identificazione è accompagnata da un MOWSE score, parametro che riflette la certezza dell'identificazione: più alto è tale valore, maggiore la sicurezza responso. Per il calcolo del MOWSE score è presa in considerazione la percentuale di copertura della sequenza della proteina identificata con quella dei vari picchi registrati allo spettrometro, il numero dei picchi che non corrispondono a nessuna porzione della sequenza presente nella proteina e la dimensione del data base utilizzato.



I peptidi vengono direttamente depositati insieme alla matrice sul *target* per l'analisi di spettrometria di massa (Anchor-Chip matrix-assisted laser desorption ionization (MALDI) target-Bruker Daltonics), preparato con un sottile strato sottile di soluzione satura di acido α -ciano-4-idrossicinnamico in acetone ogni 60 secondi usando un Proteiner FC robot (Bruker Daltonics).

L'analisi di spettrometria di massa viene effettuata con un Ultraflex MALDI-time of-flight (TOF), sotto il controllo dello software WARP LC (Bruker Daltonics)

In primo luogo gli spettri MS di tutte le frazioni analizzate per la selezione dei picchi in un range, in cui si rileva il rapporto massa/carica (m/z), da 900-4000.

Successivamente si ottengono gli spettri MS/MS dei picchi selezionati. Dopo l'acquisizione degli spettri, gli spettri individuali sono acquisiti per ciascuno ione precursore e di seguito normalizzati.

La ricerca e l'identificazione dei peptidi viene effettuata utilizzando un database MASCOT contenente 1819 proteine derivate dalla sequenza completa di *S.pyogenes* (<http://www.tigr.org/> e <http://www.ncbi.nlm.nih.gov/>).

I parametri di ricerca di MASCOT sono:

- L'identificazione della specie del ceppo
- Il numero possibile di siti proteolitici non soggetti all'azione enzimatica
- Il variabile numero di modifiche post-tradizionali (ossidazione della metionina)
- La tolleranza del peptide, ± 100 p.p.m.
- La tolleranza MS/MS, ± 0.75 Da
- La carica del peptide, +1

Successivamente i peptidi vengono separati da una nano-LC su un CapLC HPLC System (Waters) connesso ad un Q_Tof Micro electron Spray Ionization (ESI) mass spectrometer equipato con una sorgente nanospray (Waters). I campioni sono caricati su una colonna C18 (Atlantis C18 Nanoease column) (100 μ m.i.d. x 100 mm).

I peptidi eluiti vengono acquisiti automaticamente utilizzando il programma MassLynx software (Waters).

I parametri di ricerca del Mascot permettono

- L'identificazione della specie del ceppo
- Il numero possibile di siti proteolitici non soggetti all'azione enzimatica fino a 6
- Il variabile numero di modifiche post-tradizionali (ossidazione della metionina)
- La tolleranza del peptide, ± 500 p.p.m.
- La tolleranza MS/MS, ± 0.3 Da
- La carica del peptide, da +1 +4.

4.11 Purificazione di proteine mediante l'automazione

Freedom EVO (TECAN) (www.tecan.com) è uno strumento di precisione disegnato per la routine dei sistemi automatizzati per laboratorio; è una piattaforma aperta e flessibile, considerata come un assemblaggio per installazioni specifiche (Figura 33).

Lo strumento consiste in una piattaforma della lunghezza di 200 cm. La piattaforma è posizionata in una cabina al cui interno presenta:

1. un braccio manipolatore con una pinza di presa
2. un dispensatore di liquido a 8 aghi collegati a 8 siringhe (4 aventi portata massima 500 μ l e 4 dal volume massimo di 5 ml)

PRECISIONE	
Asse	Precisione
X	± 0.4 mm
Y	± 0.4 mm
Z	± 0.4 mm

RIPETIBILITÀ	
Asse	Ripetibilità
X	± 0.15 mm
Y	± 0.15 mm
Z	± 0.15 mm

Ciascun canale di liquido porta, nel nostro caso, un ago standard fisso d'acciaio inossidabile; inoltre i canali presentano la caratteristica di essere idrofilici e porosi.

Per aumentare l'idrofobicità è presente un forte rivestimento di teflon specifico per soluzioni acquose a diversa viscosità.

Gli aghi fissi da 500 µl hanno una precisione di dispensazione del volume compresa tra i 3µl e i 10 µl. Gli aghi fissi da 5000 µl presentano un intervallo compreso tra i 20 µl e i 5000µl.

3. un agitatore per piastre Te-Shake Module per piastre da 96 pozzetti
4. una stazione di lavaggio
5. una pompa per il mantenimento della stazione di lavoro a 4°C
6. spettrofotometro per lettura delle piastre a 96 pozzetti Genios Emission Filter collegato al Software Magellan versione 5.0 per la lettura dei campioni a 595 nm e elaborazione dei dati da esportare.

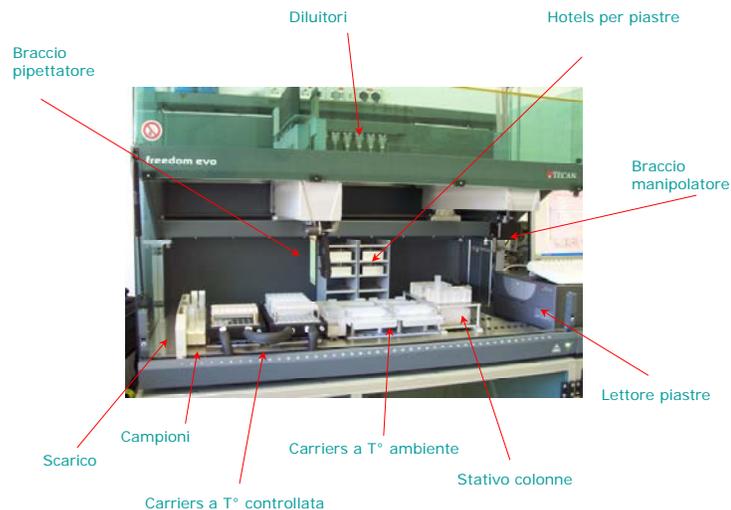


Fig. 33 Rappresentazione schematica dello strumento

Il GENios Pro combina un sistema di rivelazione avanzato: otto canali di rivelazione in una piattaforma molto compatta ad alta prestazione collegata ad un software Magellan, che offre identificazione ed elaborazione dei dati. Magellan è un software universale per la linea completa per la lettura di micropiastre di tutti i formati da 12 a 1536 pozzetti. Nel caso specifico nostro sono lette piastre da 96 pozzetti.

Un'interfaccia versatile e un'alta capacità di elaborare calcoli permettono l'ottimizzazione di una vasta gamma di lavori sperimentali da quelli di routine ai lavori maggiormente sofisticati.

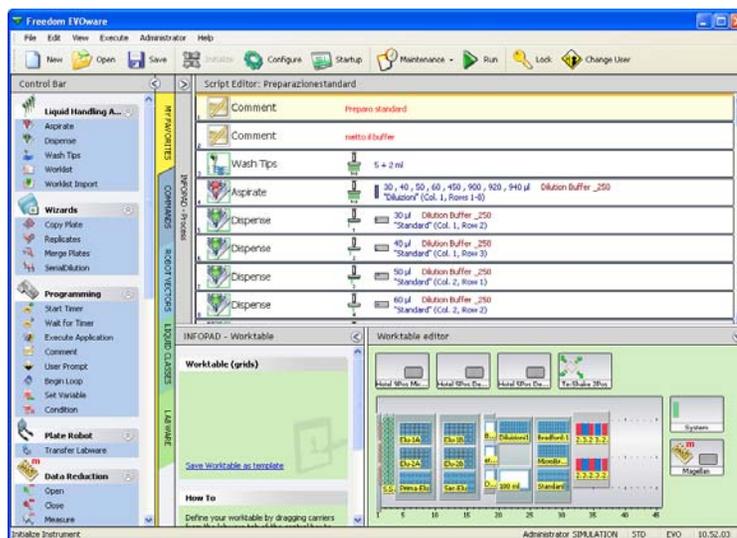


Fig. 34 Esempio di una schermata di uno script di purificazione in utilizzo

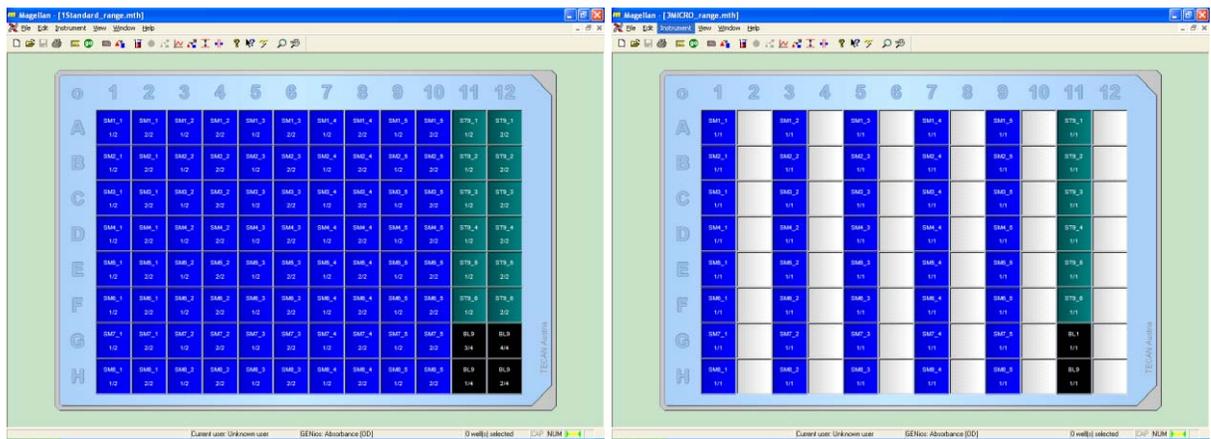
La piastra viene letta nei singoli pozzetti e nelle ultime due colonne prevede la lettura dei bianchi i cui valori vengono sottratti ai valori assoluti delle assorbanze lette e degli standard di BSA da soluzioni a concentrazione nota.

Concentrazioni note di BSA per la costruzione della curva standard	
test Standard concentrazioni (mg/ml)	test Micro concentrazioni (µg/ml)
1.4	200
1.2	160
1	120
0.8	80
0.4	40
0.2	10

Vengono riportate in figura le curve standard e micro relative alle concentrazioni di BSA precedentemente descritte.

Inoltre si riportano le figure delle piastre lette dal software utilizzato:

- a) piastra letta per la curva standard; in blu sono rappresentati tutti i campioni analizzati; in verde sono rappresentati campioni di BSA e in nero i campioni relativi ai bianchi
- b) piastra letta per la curva micro; c'è da notare che in questo caso i campioni vengono letti in singolo



a)

b)

Fig. 37 Rappresentazioni delle piastre che vengono lette dallo spettrofotometro e poi elaborate dal Magellan

4.12 Purificazione della proteina 57-His

4.12.1 Clonaggio ed espressione

Numero di amminoacidi: 1589

Peso Molecolare: 175076.3

pI teorico: 5.85

Formula: C₇₇₂₂H₁₂₀₈₇N₂₁₁₁O₂₄₇₆S₃₀

Numero totale di atomi: 24426

La proteina è stata clonata e espressa nei siti di restrizione dei vettori d'espressione pET21b+ (Novagen) usato per trasformare cellule competenti di *E. coli* BL21(DE3). A monte del *polylinker* è presente una sequenza di Istidine che serve per il recupero biochimico delle proteine ricombinanti tramite colonne cromatografiche..

Preinoculo: colonia presa da piastra in 10 ml di LB a 25°C overnight.

Inoculo: 10 ml di preinoculo (OD di partenza 0.2) messo in beute da 0.5 L con LB e Ampicillina 0.1 mg/ml a 25°C. A OD ~0.8 si aggiunge IPTG 0.2 mM finale e viene incubato a 25°C per 3 h.

4.12.2 Estrazione e chiarificazione

La biomassa (~12 g) è ripresa con 60 ml di B-PER contenente 10 mM imidazolo, 2000 µl di DNAsi, 150 µl di MgCl₂ e 2000 µl ml di lisozima a 100 mg/ml. L'estratto è chiarificato e caricato su colonna di affinità.

4.12.3 Fast Protein Liquid Chromatography

Tutti gli esperimenti di cromatografia eseguiti sono controllati da un sistema FPLC (fast protein liquid chromatography) (AKTA Purifier-Pharmacia); questa tecnica è un'applicazione della HPLC (high pressure liquid chromatography) per la purificazione di proteine. Il sistema è utilizzato come metodo di sviluppo e purificazione di grandi quantità di biomolecole, include due pompe peristaltiche capaci di garantire un flusso continuo alla colonna di purificazione scelta, una lampada UV per rivelare le biomolecole sulla colonna posta a 280 nm, lunghezza d'onda alla quale assorbono gli amminoacidi aromatici ed una valvola per determinare il gradiente adottato durante la corsa cromatografica. Questa struttura è collegata ad un sistema computerizzato mediante il software Unicorn (versione 4.11) per la visualizzazione dell'andamento del flusso delle pompe e il controllo della corsa cromatografica.

4.12.4 Cromatografia di affinità

Colonna Chelating HP (Pharmacia) Hi Trap™ 5 ml allo strumento AKTA Purifier Amersham Pharmacia Biotech. Prevede un sistema di FPLC (fast protein liquid chromatography) per una purificazione efficiente e veloce; è collegato al software Unicorn (versione 4.11) per l'elaborazione dei risultati.

Il sistema include due pompe capaci di garantire un flusso continuo alla colonna di purificazione scelta; una lampada UV per rivelare le proteine presenti sulla colonna posta a 280 nm, lunghezza d'onda alla quale assorbono gli amminoacidi aromatici ed una valvola per determinare il gradiente di concentrazione.

È stato utilizzato il seguente programma di eluizione:

- Tampone A: 50mM NaH₂PO₄, 300 mM NaCl, pH 8.0
- Tampone B: 50mM NaH₂PO₄, 300 mM NaCl, 500 mM imidazolo, pH 8.0
- Equilibrio e Carico: 2% B - Gradiente: 0-50%B in 16 CV- 50-100% in 8 CV

- Flusso 10 ml/minuto
- Lavaggio: 2 CV
- Frazioni di Eluato: 10 ml

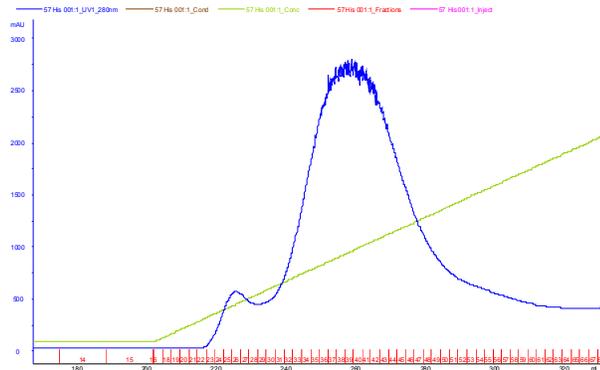


Fig 38 : cromatogramma della proteina 57-His su His-Trap Chelating Sepharose Ni-activated; la curva blu rappresenta l'assorbimento UV a 280 nm; la curva verde l'andamento della concentrazione di imidazolo; in rosso sono indicate le frazioni.

4.12.5 Cromatografia a scambio ionico

Colonna Q-Sepharose Fast Flow (Pharmacia) Q Trap™ 5 ml (colonna scambio anionico). La cromatografia è condotta con AKTA Purifier Amersham Pharmacia Biotech; è stato utilizzato il seguente programma di eluizione:

- Tampone A: 30mM TRIS, pH 8.0
- Tampone B: 30mM TRIS, 1 M NaCl, pH 8.0
- Equilibrio e Carico: 2% B - Gradiente: 0-50%B in 16 CV- 50-100% in 8 CV
- Flusso: 10 ml/minuto
- Lavaggio: 2CV
- Frazioni di eluato: 10 ml

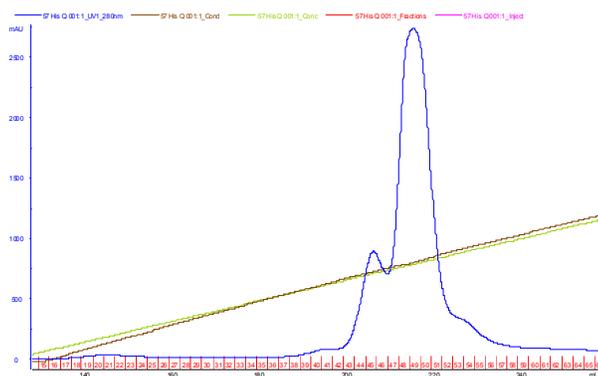


Fig 39: cromatogramma della proteina 57-His su Q-Sepharose; la curva blu rappresenta l'assorbimento UV a 280 nm; la curva verde l'andamento della concentrazione di imidazolo; la curva marrone l'andamento della conduttanza ; il rosso sono indicate le frazioni.

4.12.6 Cromatografia di gel filtrazione

Colonna HiLoad 26/60 Superdex 200 prep grade (Pharmacia) equilibrata in PBS. La cromatografia è condotta con AKTA Purifier Amersham Pharmacia Biotech; è stato utilizzato il seguente programma di eluizione:

- Tampone A: PBS (corsa in isocratica) pH 7.4
- Flusso: 2.5 ml/minuto
- Frazioni di eluato: 2.5 ml

La tecnica permette di separare macromolecole in base alla loro massa molecolare. In questo caso, grazie all'impiego di *marker* proteici di massa nota, è possibile operare una valutazione della massa molecolare delle varie frazioni separate, confrontando i tempi di ritenzione del campione con quelli dei *marker*.

Per calibrare la colonna si utilizzano gli standard Pharmacia (LMW Gel Filtration Marker) descritti di seguito:

- ribonucleasi A (13700 Da)
- chimotripsinogeno (25000 Da)
- ovoalbumina (43000 Da)
- blu destrano 2000 (2000 KDa)

Il materiale recuperato è stato risospeso in PBS e caricato su una colonna di gel-filtrazione *Superdex 200*.

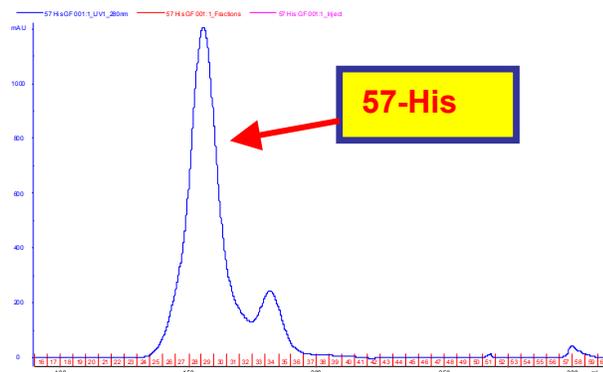


Fig. 40: cromatogramma della proteina 57-His su Gel filtrazione HiLoad Superdex 200 26/60 ; la curva blu rappresenta l'assorbimento UV a 280 nm; in rosso sono indicate le frazioni

4.13 Analisi citofluorimetrica-Fluorescence activated cell sorter-(FACS)

Il grado di esposizione delle proteine sulla superficie batterica è stato determinato mediante analisi citofluorimetrica. I sieri dei topi immunizzati con l'antigene da analizzare sono stati saggiati per verificare la capacità degli anticorpi di legarsi alla superficie di GAS. 40 μ l di una coltura batterica a OD₆₀₀ 0,32, corrispondente a circa metà fase esponenziale, sono stati lavati con PBS, centrifugati e risospesi in 20 μ l di un siero bloccante (NCS, Newborn Calf Serum; Sigma) per saturare i legami aspecifici dei sieri usati per la successiva marcatura. Dopo aver incubato la miscela per 20 min. a RT, si è proseguito con una successiva incubazione di un'ora a 4°C di una soluzione

contenente il siero policlonale di topo preimmune o immune diluito 1:200 in PBS-BSA 0,1% e i batteri risospesi in NCS, in un volume finale di 100 μ l. I controlli, positivo e negativo, usati nell'esperimento sono rispettivamente: sieri di topi immunizzati con la proteina M o sieri di topi immunizzati con His-STOP ed adiuvante. Dopo due lavaggi con 200 μ l di PBS-BSA 0,1%, i campioni vengono incubati per 30 min. a 4°C con 10 μ l di una soluzione contenente l'anticorpo secondario anti-IgG di topo coniugato con la ficoeritrina (R-phycoerithrin-conjugated F(ab)₂ goat anti-mouse IgG; Jackson ImmunoResearch Laboratories, Inc.) diluito 1:100 in PBS-BSA 0,1% in presenza di NCS al 20%. Dopo 2 lavaggi con 200 μ l di PBS-BSA 0,1% e di PBS, i batteri vengono inattivati in paraformaldeide al 2%, concentrazione che non altera la conformazione degli antigeni proteici da analizzare. La paraformaldeide, infatti, crea legami crociati tra i gruppi amminici liberi presenti nelle catene laterali degli aminoacidi che costituiscono le strutture polipeptidiche del batterio consentendo una buona conservazione della struttura cellulare e dell'antigenicità. Dopo un'incubazione di venti minuti a TA viene fatto un lavaggio con 200 μ l di PBS e infine i batteri vengono risospesi in 100 μ l di PBS. Le cellule vengono quindi analizzate utilizzando lo strumento FACS Calibur (Becton Dickinson) che possiede, tra le altre, una sorgente laser che emette alla lunghezza d'onda di 488 nm. I campioni marcati con ficoeritrina assorbono tale luce ed emettono una fluorescenza di lunghezza d'onda pari a 520 nm, rilevabile dallo strumento. Il segnale luminoso viene convertito in segnale digitale sottoforma di un diagramma di distribuzione in cui in ordinata si evidenziano il numero di cellule, mentre in ascissa l'intensità della fluorescenza.

I dati sono stati analizzati mediante Cell Quest Software (Becton Dickinson) disegnando un istogramma in cui venivano sovrapposti il picco di intensità di fluorescenza dato dal siero immune e quello del corrispondente siero preimmune, utilizzato per controllare il *background*. La differenza tra i due picchi, indice del livello di esposizione degli antigeni sulla superficie batterica, è stata indicata come Δ Mean definita come valore medio dell'intensità di fluorescenza della popolazione batterica. Si sono considerati positivi valori di Δ Mean > di 80 canali.

Appendice

Parte 2: Approccio proteomico - breve prefazione da dati ritrovati in letteratura

Le proteine di superficie di batterio giocano un ruolo fondamentale nell'interazione del batterio e il suo ambiente. Esse sono coinvolte nell'adesione alle cellule ospiti e alla possibile invasione; sono soggette alle condizioni chimico-fisiche del mezzo esterno e possono inviare numerosi segnali al compartimento citoplasmatico organizzandosi nella difesa contro le risposte dell'ospite e contro la sua eventuale tossicità.

Quindi le proteine di superficie sono potenziali *target* per farmaci deputati alla prevenzione di infezioni e malattie batteriche.

Inoltre le proteine di superficie, vista la loro possibile interazione con il sistema immunitario dell'ospite, sono candidate come potenziali componenti di un vaccino.

I vaccini basati su proteine secrete ed esposte sulla superficie sono già disponibili in commercio e altri sono in sviluppo (Grandi, 2004).

Il primo metodo è basato sulla predizione delle proteine di superficie dall'analisi del genoma usando algoritmi come il PSORT. Il metodo è rapido, ma non sempre pienamente efficace e non è assolutamente quantitativo.

Il secondo approccio impiega la separazione delle proteine di superficie dalla frazione citoplasmatica e in seguito l'identificazione di proteine mediante l'elettroforesi bidimensionale (2D) o la cromatografia bidimensionale accoppiata alla spettrometria di massa.

Questo approccio che è stato utilizzato in parecchi batteri è ragionevolmente quantitativo (Molloy et al, 2000).

Al contrario, non è sufficientemente selettivo e molte proteine, che contaminano le proteine del "rivestimento" batterico, sono identificate. Inoltre, non permette di definire realmente quali sono le proteine con domini esposti sulla superficie.

In Novartis si è messa a punto una nuova procedura, che permetta la rapida e selettiva identificazione di proteine batteriche esposte sulla superficie, il pool di proteine che sono interamente o solo parzialmente esposte all'esterno della parete batterica.

Il metodo utilizza degli enzimi proteolitici, che agiscono sulla superficie del batterio e i peptidi generati sono separati dalle cellule intere e identificate mediante la spettrometria di massa.

La veridicità di questo metodo è stata dimostrata dalla caratterizzazione di un set completo di proteine esposte sulla superficie del ceppo M1_SF370 dello *Streptococcus pyogenes* (GAS) (Rodriguez-Ortega et al., 2006).

Si mostra che il 95% delle proteine identificate appartengono alle quattro famiglie di proteine predette associate alla superficie di un batterio gram-positivo (proteine della parete cellulare con motivo di ancoraggio LPXTG, proteine integrali di membrana, lipoproteine e proteine secrete). L'efficienza dell'approccio si è verificata, dimostrando che la maggior parte delle proteine identificate sono accessibili agli anticorpi policlonali presentate contro le corrispondenti proteine ricombinanti.

Quindi quest'approccio rappresenta un valido strumento per lo studio dell'organizzazione delle proteine di superficie e per identificare nuovi candidati per un vaccino.

5.1.1 Analisi del ceppo GAS M1_SF370

Per identificare le proteine esposte sulla superficie del genoma M1_SF370 completamente sequenziato, i batteri sono fatti crescere in coltura e successivamente trattati con tripsina o con proteinasi K al fine di digerire la superficie batterica dei domini delle proteine esposte (Rodriguez-Ortega et al, 2006). Questo trattamento non inficia l'integrità della cellula; i peptidi rilasciati nel supernatante sono concentrati e analizzati alla spettrometria di massa tandem. Sono state identificate un totale di 72 proteine; 37 in seguito alla sovrapposizione di più di un peptide per proteina e 35 dal singolo peptide. Quarantatre proteine provengono dai peptidi trattati con la tripsina, diciotto proteine dal

trattamento con la proteinasi K e undici proteine dal trattamento con entrambi gli enzimi proteolitici.

Dalla analisi della sequenza, secondo il software PSORT, le 72 proteine ritrovate possono così essere raggruppate in quattro famiglie:

1. le proteine contenenti il motivo LPXTG (12 proteine) (**a**)

hypothetical protein
hypothetical protein
putative cell envelope proteinase
putative extracellular matrix binding protein
conserved hypothetical protein
hypothetical protein
protein GRAB (protein G-related alpha 2M-binding protein)
hypothetical protein
collagen-like surface protein
hypothetical protein
C5A peptidase precursor
M protein type 1

2. le lipoproteine (11 proteine) (**b**)

putative ABC transporter
oligopeptide permease
conserved hypothetical protein
hypothetical protein
putative cyclophilin-type protein
hypothetical protein
putative amino acid ABC transporter, periplasmic amino acid-binding protein
putative maltose/maltodextrin-binding protein
putative protease maturation protein
putative acid phosphatase
hypothetical protein

3. le proteine secrete (**c**)

putative secreted protein
nicotine adenine dinucleotide glycohydrolase precursor
streptolysin O precursor
putative 42 kDa protein
putative cyclomaltodextrin glucanotransferase
hypothetical protein
putative signal peptidase I
inhibitor of complement-mediated lysis

4. le proteine di membrana (**d**)

putative cell division protein (1)(d) ***
putative large conductance mechanosensitive channel (1)
hypothetical protein (1)
conserved hypothetical protein (1)
putative ABC transporter (binding protein) (1)
putative acetoin dehydrogenase (TPP-dependent) beta (1)
putative dihydrolipoamide dehydrogenase component E3 (1)
putative citrate lyase, beta subunit (1)
putative deacetylase (1)
putative cell division protein (1)
putative beta-galactosidase (1)
hypothetical protein (1)
putative penicillin-binding protein 1a (1)
hypothetical protein (1)
hypothetical protein sharing similarity with several eukaryotic proteins (1)
hypothetical protein (1)
immunogenic secreted protein precursor (1)
putative ATP-binding cassette transporter-like protein (1)
hypothetical protein (1)
mitogenic factor (1)
putative serine protease (1)
putative surface exclusion protein (2)
hypothetical protein (2)
hypothetical protein (2)
putative 1-acylglycerol-3-phosphate Oacyltransferase (2)
conserved hypothetical protein (2)
putative glutamine-binding periplasmic protein (3)
hypothetical protein (3)
putative mannose-specific phosphotransferase system component IID (3)
putative cell-division protein (4)
hypothetical protein (4)
putative ABC transporter (ATP-binding protein) (4)
putative glycine betaine binding permease protein (6)
hypothetical protein (6)
beta-glucoside permease IIABC component (7)
putative L-malate permease (10)
putative integral membrane protein (11)

*** in parenthesis sono riportati il numero di domini trans-membrana

Solo quattro delle proteine predette dal PSORT che risiedono nel compartimento citoplasmatico, sono state ritrovate nella porzione sensibile alle proteasi. Esse comprendono il fattore EF-Tu, che sembra essere associato alla superficie anche in altri batteri (Grandi et al, 2001), due proteine ribosomiali, che fanno ipotizzare alcune eventuali funzioni extracellulari, e una proteina ipotetica coinvolta nella formazione della parete cellulare (Kurar et al, 2000; Spence et al, 2000).

putative translation elongation factor EF-Tu
50S ribosomal protein L31
conserved hypothetical protein - possibly involved in cell wall localization and side chain formation
50S ribosomal protein L7/L12

5.1.2 Conferma dell'esposizione delle proteine in superficie

La quasi completa assenza di peptidi di proteine citoplasmatiche predette, ha suggerito che la procedura fosse selettiva per l'identificazione delle proteine esposte in superficie. Delle 72 proteine identificate dopo la digestione superficiale, 53 sono state clonate con successo in *E. coli*, espresse e purificate. Le proteine ricombinanti sono state usate per produrre anticorpi policlonali; questi ultimi sono stati usati per esperimenti di FACS, in modo tale da confermare l'esposizione sulla superficie.

Di queste 51 proteine testate, 43 sono state dimostrate FACS positive, ovvero di superficie. Le proteine FACS positive sono state trovate nelle 4 quattro famiglie di proteine predette associate alla superficie:

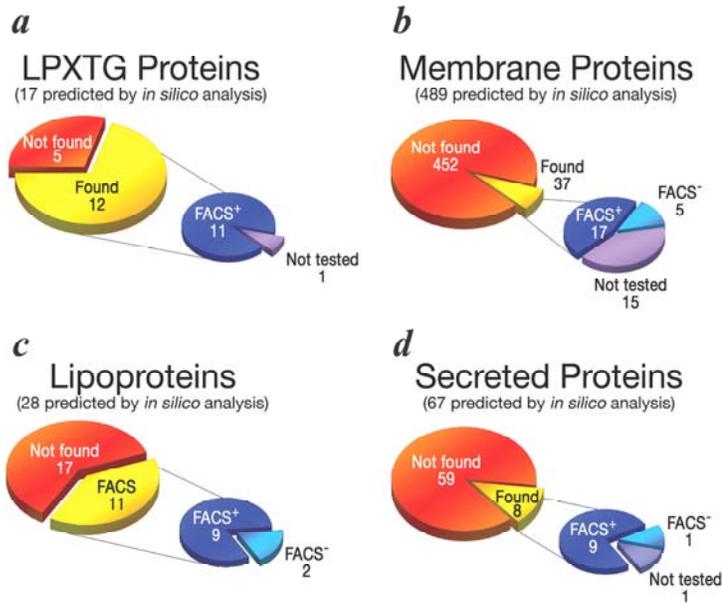


Fig. 41: Rappresentazione grafica delle proteine identificate sulla superficie di M1_SF370

Le 72 proteine appartenenti al proteoma di M1_SF370 sono raggruppate in famiglie basate sulla predizione cellulare. Le aree rosse di ciascun grafico indicano il numero di proteine predette dal PSORT che non sono state trovate sulla superficie del proteoma; le aree gialle rappresentano il numero di proteine identificate appartenente a ciascuna famiglia di proteine; i grafici blu illustrano quante proteine identificate sono state confermate essere esposte sulla superficie in comparazione con quelle ritrovate mediante analisi FACS.

Si riportano solo come esempio alcuni dei dati di analisi citofluorimetrica (descritta in dettaglio nella sezione Materiali e Metodi) in modo da rendere più chiaro il metodo di selezione delle proteine di superficie. Esempio di proteine FACS negativa (a) e FACS positiva (b) (Figura 42):

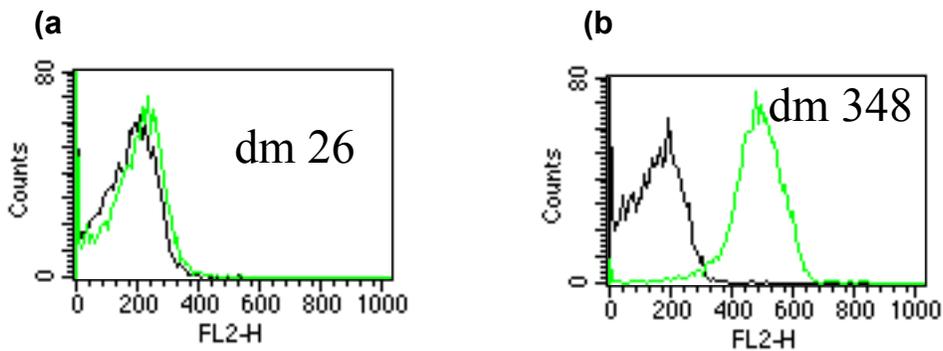


Fig.42 Esempio di analisi citofluorimetrica del livello di esposizione di proteine ricombinanti FACS negativo a) e FACS positivo b) sulla superficie di *S. pyogenes*

Il livello di esposizione ($\Delta Mean$) sulla superficie degli antigeni analizzati sul ceppo M1_SF370 è stato calcolato come la differenza tra il valore medio del picco dato dal siero immune (in verde nella figura 42) e quello dato dal corrispondente siero preimmune (in nero nella figura 42).

É da notare che molte delle proteine ben espresse ed esposte sulla superficie, come descritto dall'intensità di fluorescenza, in alcuni casi presentano lo stesso *range* avutosi nel caso della maggiore superficie di antigene, la proteina M.

L'alta percentuale di FACS positiva dimostra la validità del metodo.

5.1.3 Confronto dei risultati sperimentali con le predizione di PSORT

37 proteine transmembrana sono state identificate. Per 26 proteine su 37 è stata confermata la predizione del PSORT; per le rimanenti 11 proteine, esistono delle discordanze tra la predizione del PSORT e la digestione superficiale. Di conseguenza è stato poi necessario rivisitare alcune di queste proteine nella loro topologia effettuando ulteriori analisi. Ad esempio la GAS 163 ha 6 regioni transmembrana nella predizione (figure 10), e uno dei sei presenta uno score molto basso (dominio transmembrana 4). Assumendo che quest'ultimo non fosse un dominio transmembrana, la parte C-terminale della molecola che porta un tipico dominio di una sortasi sarebbe esposto sulla superficie e consistente con il meccanismo di azione delle sortasi.

Di seguito si mostra la topologia della proteina GAS 163:

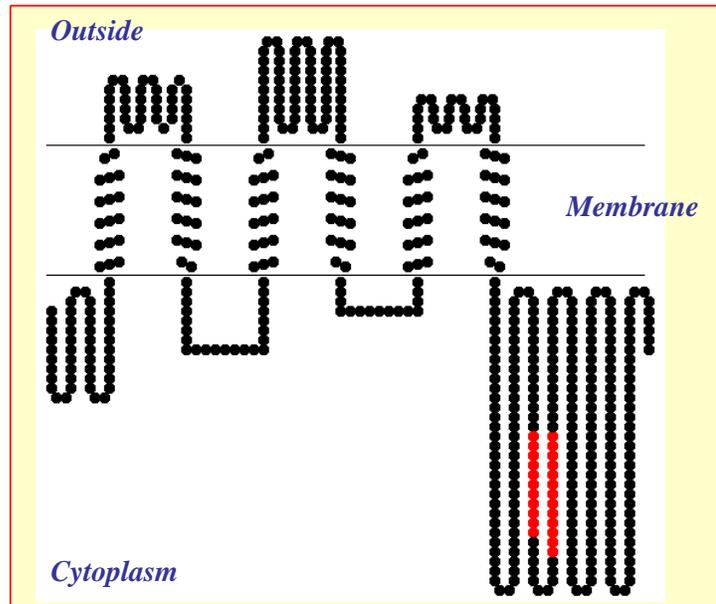


Fig. 43 Topologia basata sulla predizione del PSORT

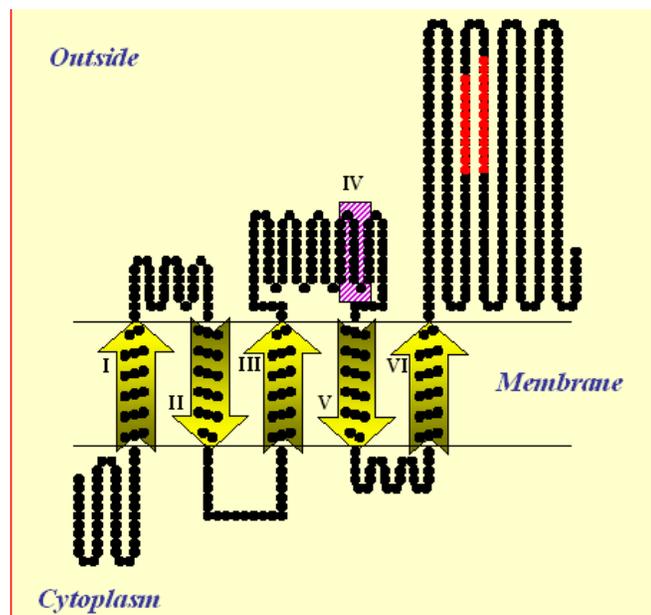


Fig. 44 Topologia modificata in seguito a delezione di un dominio transmembrana

La nuova topologia è stata validata dall'analisi citofluorimetrica utilizzando un anticorpo che riconosce la parte C-terminale della proteina. Questa validazione

dimostra che l'approccio è anche in grado di revisionare la predizione della topologia di una proteina.

In definitiva:

- una buona corrispondenza esiste tra i valori di FACS e quelli di spettrometria di massa
- il metodo permette di revisionare le predizioni di topologia di PSORT

Si può quindi concludere che questo metodo comporta l'identificazione di proteine esposte sulla superficie.

Bibliografia

- Areschoug T et al, (2004) *Vaccine*. **22 Suppl 1**:S9-S14.
- Belongia EA et al, (2003) *Clin Med Res*. **1(2)**:87-92.
- Bisno AL, et al, (2003). *Lancet Infect Dis*. **3(4)**:191-200.
- Byrd CM et al, (2002). *Vaccine*. **20(17-18)**:2197-2205.
- Chakravarti, D. N. et al, (2000), *Vaccine* **19**, 601-612.
- Carapetis JR et al, (2004). *Adv Exp Med Biol*. **549**:185-190.
- Carlsson F et al, (2003). *J Exp Med*. **198(7)**:1057-1068.
- Cheng Q et al, (2002). *Infect Immun*. **70(5)**:2408-2413.
- Courtney HS et al, (1999). *Mol Microbiol*. **32(1)**:89-98.
- Cunningham MW (2000). *Clin Microbil Rev*. **13(3)**:470-511.
- Dunn LA et al, (2002). *Vaccine*. **20(21-22)**:2635-2640.
- Ernst JD (2000). *Cell Microbiol*. **2(5)**:379-386.
- Fernie-King BA et al, (2004). *Immunology*. **111(4)**:444-452.
- Ferretti JJ et al, (2001). *Proc Natl Acad Sci USA*. **98(8)**:4658-4663.
- Fischetti VA (1989). *Clin Microbiol Rev*. **2(3)**:285-314.
- Fischetti VA et al (1990). *Mol Microbiol*. **4(9)**:1603-1605.
- Giannakis E, et al (2002). *J Immunol*. **168(9)**:4585-4592.
- Grandi G (2004). *Genomics, proteomics and vaccines*. Wiley.
- Hollingshead SK. et al, (1993). . *Mol Microbiol*. **8(4)**:707-717.
- Hu MC et al, (2002). *Infect Immun*. **70(4)**: 2171–2177.
- Jarva H et al, (2003). *Mol Immunol*. **40(2-4)**:95-107.
- Jawetz E et al, (1998). *Microbiologia medica*.
- Johnson IS (2005). *Nature*. **435(7044)**:881.
- Kotb M (1995). *Clin Microbiol Rev*. **8(3)**:411-426.
- Kohler G et al, (1975). *Nature*. **256(5517)**:495-497.
- Lancefield RC (1933). *J Exp Med*. **57**:571-595.

- Lei B, Smoot LM, et al, (2002). *Infect Immun.* **70(8)**:4494-500.
- Pizza M, et al (2000), *Science*; **287(5459)**:1816-20.
- Maione D, et al (2005, *Science.* **309(5731)**:148-150.
- Maignani V, et al (2002), *Expert Opin Biol Ther.* **2(8)**:895-905.
- Draveling C, et al (2001), *Protein Expression and Purification* **22**, 359-366.
- Gwyne P, et al (2002), Laboratory technology trends: lab automation and robotics.
- Lesley S.A, (2001), *Protein Expression and Purification* **22**, 159-164.
- Weinmann H et al, (2003), *Jala* **8**, 59-70.
- Noonan W. J et al, (2003), *Jala* **8**, 65-71.
- Rodriguez-Ortega M J et al, (2006) *Nature Biotechnology* **24**, 191-197.
- Lindhal G. et al, (2005), *Clin. Microbiol. Rev.* **18**, 102-127.
- Courtney H. S., et al, (2002), *Handbook of bacterial adhesion: Principles, Methods, Applications*, 553-579.
- Ton-That H., et al, (2004), *Biochim. Biophys. Acta* **1694**, 269-278.
- Guina T. et al, (2003), *J. Am. Soc. Mass. Spectrom.* **14**, 742-751.
- Marques M. A., et al, (1998) *Infect. Immun.* **66**, 2625-2631.
- Spence, J.M. et al, (2000), *Infect. Immun.* **68**, 5002-5010.
- Kurar E., et al (2000) *Vaccine* **15**, 1851-1857.
- Stalhammar-Carlemalm M, et al (1999), *Mol. Microbiol.* **33**, 208-219.
- McMillan DJ, et al (2004) *Vaccine*, **22**, 2783-2790.
- Cole JN, et al (2005) *Infect. Immun.* **73**, 3137-3146.
- Lancefield R.C. (1959), *J. Exp. Med.*, **110**, 271-292.
- Lancefield R. C. (1962), *J. Immunol.* **89**, 307-313.
- Harris T. et al (2003), *J. Clin. Invest.* **111**, 61-70.
- Carlsson F., et al (2006) *Nature*, **442(7105)**, 943-946.
- Tettelin et al., (2000) *Science.* **287(5459)**:1809-15[2].
- Mora M, et al (2005) *Proc Natl Acad Sci U S A.* **102(43)**:15 641.
- Valenzuela Pet al (1982) *Nature* **298(5872)**:347-50.

- Greco D, et al (1996) *N Engl J Med* **334(6)**:341-8.
- Pizza M, et al. (1989) *Science* **246(4929)**:497-500.
- Massignani V, et al (2003) *Mol Micr.* ;**50(3)**:1055-67.
- Ouskova G, et al (2004). *Glicobiology* **14(10)**:931-8.
- Bricker AL, et al (2005) *Infect immun.*, **73(10)**:6562-6.
- Rappuoli R., (2000) *Curr Opin Microbiol.*, **3(5)**:445-50.
- Gygi SP, et al. (1999), *Mol Cell Biol.* **19(3)**:1720-30.
- Nakai K, et al. (1991), *Proteins* **11(2)**:95-110.
- Nakai K. (2000), *Adv Prot. Chem.*, **54**:277-344.
- Grandi G. (2003) *Int J Parasitol* **33(5-6)**:615-20.
- Grandi G. (2001) *Trends Biotechnol.* **19(5)**:181-8.
- Montigiani et al, (2002) *Infect Immun.*;**70(1)**:368-79.
- Wizemann et al, (2001) *Infect Immun.*; **69(3)**:1593-8.
- Polly SM, et al (1975), *J Infect Dis.* **131(3)**:217-24.
- Giuliani MM, et al (2006) *Proc. Natl. Acad. Sci USA* **103(29)**:10834-9.
- Grifantini R, et al (2004) *Nature Biotechnol.* **20(9)**:914-21.
- Mora M, et al (2003) *Drug discover today* **8(10)**:459-64.
- Gorg A, et al (2000) *Electrophoresis*, **21(6)**:1037-53.
- Molloy P. et al, (2000), *Eur. J. Biochem.* **267**, 2871-2881.
- Rhomberg T., et al (2004), *Proteomics*, **4**, 3021-3033.
- Ferrari G, et al, (2006), *Proteomics*, **6**, 1856-66.
- Zuobi-Hasona K, et al (2005) **26**, 1200-5.