UNIVERSITÀ DEGLI STUDI DI PAVIA

DOTTORATO IN SCIENZE CHIMICHE E FARMACEUTICHE

XXX CICLO

Coordinatore: Chiar.mo Prof. Mauro Freccero

INNOVATIVE ANALYTICAL METHODOLOGIES FOR THE DEVELOPMENT OF NEW ANTITUBERCULAR AGENTS

Tutore

Chiar.ma Prof.ssa Enrica Calleri

Co-tutore

Chiar.ma Prof.ssa Caterina Temporini

Tesi di Dottorato di

FRANCESCA RINALDI

a.a. 2016-2017

CONTENTS

ABBREVIATIONS

AEC: Anion Exchange Chromatography

BCG: Bacillus Calmette-Guérin

DC: Dendritic Cell

DC-SIGN: Dendritic Cell-Specific Intracellular adhesion molecule 3-Grabbing Non-integrin

DLS: Dynamic Light Scattering

DR: Drug-Resistant

EK: EnteroKinase

EM: Electron Microscopy

ESI: ElectroSpray Ionization

FT-ICR: Fourier Transform Ion Cyclotron Resonance

HILIC: Hydrophilic Interaction Liquid Chromatography

HIV: Human Immunodeficiency Virus

*Hs*PNP: Human Purine Nucleoside Phosphorylase

HTS: High-Throughput Screening

IME: 2-IminoMethoxyEthyl

IMER: IMmobilized Enzyme Reactor

LAM: LipoArabinoMannan

LC: Liquid Chromatography

ManLAM: Mannose-capped LipoArabinoMannan

MALDI: Matrix-Assisted Laser Desorption/Ionization

MS: Mass Spectrometry

MDR: MultiDrug-Resistant

MMR: Macrophage Mannose Receptor

Mtb: *Mycobacterium tuberculosis*

*Mtb*PNP: *Mycobacterium tuberculosis* Purine Nucleoside Phosphorylase

PNP: Purine Nucleoside Phosphorylase

PGC: Porous Graphitized Carbon

RR: Rifampicin Resistant

SEC: Size Exclusion Chromatography

SPE: Solid Phase Extraction

SPR: Surface Plasmon Resonance

TB: TuBerculosis

TDR: Totally Drug-Resistant

TOF: Time-Of-Flight

WHO: World Health Organization

XDR: eXtensively Drug-Resistant

INTRODUCTION

1. Tuberculosis

Tuberculosis (TB) is one of the main causes of death in the world. Although the Stop TB Strategy claimed by World Health Organization (WHO) aimed at drastically decreasing the global TB epidemic by 2015, 10.4 million new cases and 1.4 million deaths were reported during that year. TB was indeed the deadliest infectious disease in 2015, even over human immunodeficiency virus (HIV) [1].

The infectious agent responsible for TB is the bacillus *Mycobacterium tuberculosis* (*Mtb*), which may be considered as the most successful intracellular bacterium for prevalence and distribution in the world. In fact, although Robert Koch identified *Mtb* in 1882 and different antitubercular treatments are currently available, the bacterium latently infected approximately one third of the global population and it still causes millions of deaths each year [2].

Since *Mtb* is an aerobic intracellular pathogen that mainly grows in tissues with high oxygen levels, TB usually affects the lungs, even if forms of extrapulmonary TB are also well-known. The main symptoms of pulmonary TB are chronic cough associated with hemoptysis, sputum production, appetite and weight loss, fever, night sweats. Extrapulmonary TB, which can affect 10-42% of patients, causes variable signs and symptoms and can hit any organ [3, 4].

Mtb is not classified as Gram positive or Gram negative due to its peculiar cell wall structure. The abundance of lipids in its wall hinders the permeability to basic dyes and the discolouration with acidified organic solvents. Therefore, it is defined as acid-fast [3].

TB transmission occurs principally by air, through *Mtb* diffusion from patients with active disease. Generally, only 5-15% of the people infected with *Mtb* will develop the disease throughout their lives, but the percentage grows considerably for HIV-positive individuals. Most immunocompetent individuals avoid active TB either removing the pathogen from the organism or keeping it in a latent state. People with latent TB do not show any damage or symptom because their immune system is able to manage *Mtb* replication [1, 3].

The first response after the inhalation of *Mtb* by the human host is from innate immunity. The bacterium initially interacts with alveolar macrophages, then various phagocytic cells including neutrophils, monocyte-derived macrophages, and dendritic cells (DCs) reach the infected lung and ingest bacilli. Recognition of *Mtb* by phagocytic cells is due to many receptors, among which macrophage mannose receptor (MMR, the first C-type lectin receptor linked to *Mtb* recognition) and DC-specific intracellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN) are believed to play

an important role in bacterial virulence. In fact, besides their role in phagocytosis, phagosome maturation and cytokine signaling, they recognize mannose-capped lipoarabinomannan (ManLAM). This lipoglycan is an essential component of *Mtb* cell wall and it is probably involved in the virulence of the pathogen because mannosylation of the LAM is typical for pathogenic mycobacteria. Conversely, environmental mycobacteria are characterized by a LAM capped with arabinose or inositol phosphate [5].

The early innate immune response is crucial to avoid a progressive infection and its main functions are the optimization of adaptive immune responses and the regulation of inflammation. However, without the action of adaptive immunity, it is not possible to control the infection in the long-term and to prevent the onset of active TB [5].

Adaptive anti-*Mtb* responses are characterized by a delayed start compared to those against other pathogens and by a continuous action to keep the bacillus in a latent state, contrasting and controlling *Mtb* replication. The response starts in the lymph nodes located in the upper part of the thorax, where dendritic cells, macrophages and other antigen-presenting cells coming from the infected lungs activate CD4+ and CD8+ T cells. Anti-*Mtb* specific T cells move to the infection site, where also neutrophils, monocytes/macrophages and B cells are activated, forming dynamic lymphoid structures typical for TB. These complexes of immune cells are known as granulomas and consist in compact and organized structures with a core of aggregate macrophages (**Figure 1**). Even if granulomas play a protective role for the host, also *Mtb* can exploit them for its survival, replication, diffusion and development of resistance mechanisms [2, 5, 6].

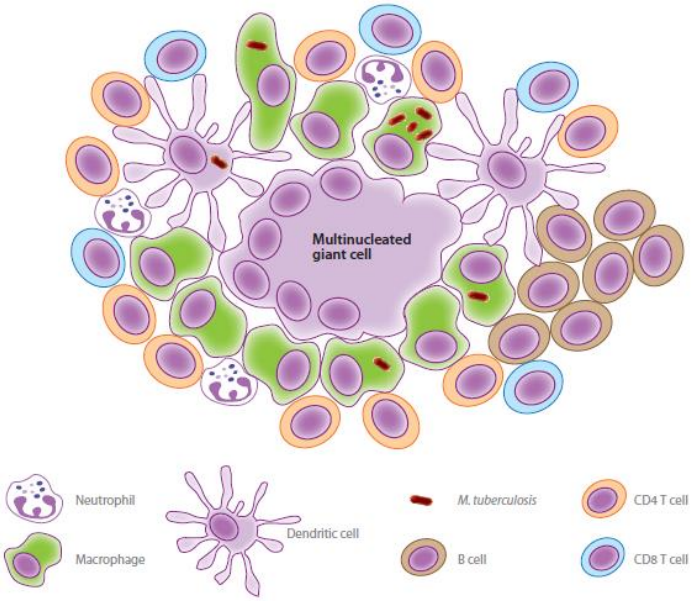


Figure 1. Structure of a representative human granuloma in response to *Mtb* infection [5].

Even if the functions of B cells in immune responses against TB are not completely elucidated, they seem to be important in both early and latent phases of the disease. B cells may act presenting mycobacterial antigens to T cells and producing cytokines and anti-*Mtb* specific antibodies. In addition to the cell-mediated immune response, the humoral response triggered by antibodies helps in the protection against the disease thanks to the recognition of various mycobacterial antigens [6].

Immune response is significantly different in HIV-positive patients, in which the cell-mediated response is poor. In these individuals, the production of specific CD4+ T cells and cytokines is reduced, with a consequent difficulty in the formation of functional granulomas. Both *Mtb* and HIV take advantage from these conditions, resulting in an increase in their replication and propagation [3].

1.1. Current status of the TB epidemic

TB deaths and incidence rates are progressively decreasing over time, but the disease diffusion is still high. In fact, although early diagnosis associated with a suitable therapy can be decisive in most cases, in 2015 TB was among the top 10 causes of human death [1].

Of the 10.4 million people infected with *Mtb* in 2015, around 11% were HIV-positive and therefore with a higher risk to develop active TB. Africa accounted for a higher percentage of patients coinfected with *Mtb* and HIV, especially in some southern regions in which it exceeded 50%. In 2015, TB caused the death of 0.4 million HIV-positive individuals, in addition to the 1.4 million among people without HIV [1].

In recent years, many efforts were made to reduce the burden of TB. The success of the applied strategies is proven by the substantial decrease of TB incidence (1.4% per year between 2000 and 2015), prevalence and mortality (42% and 47% from 1990 to 2015, respectively). Incidence defines the number of new and relapse cases in one year, prevalence is given by the number of TB cases at a fixed point in time and mortality is the number of deaths caused by the disease in a year [1].

Despite these improvements, WHO proposed new goals in order to end the global TB epidemics within 2035 (End TB Strategy and Sustainable Development Goals). Since India, Indonesia, China, Nigeria, Pakistan and South Africa are responsible for 60% of TB new cases (**Figure 2**), the achievement of WHO's goals relies on progresses in TB treatment in these six countries [1].

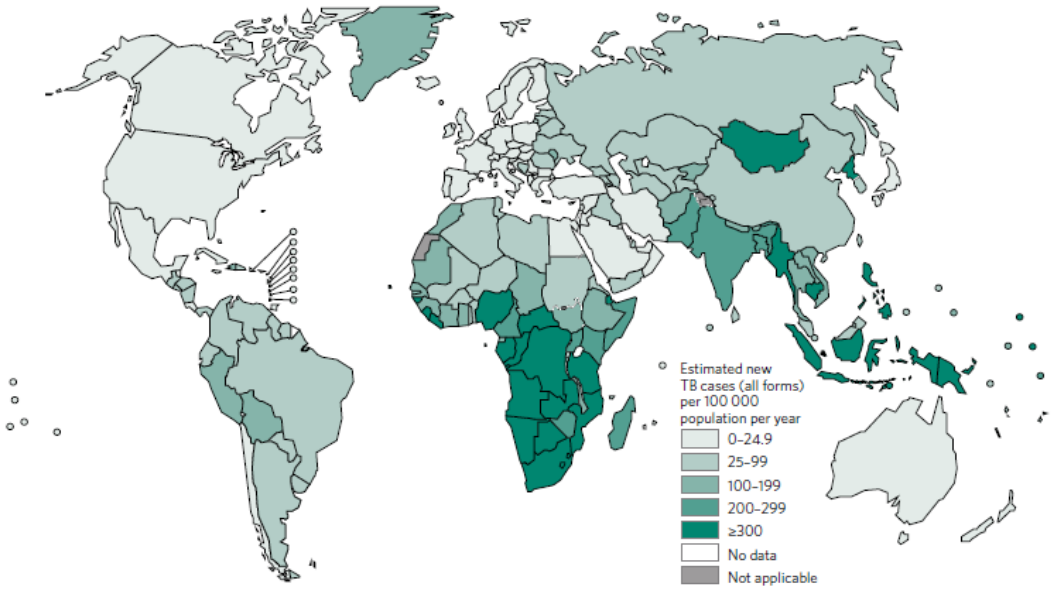


Figure 2. Estimated TB incidence rates in 2015 [1].

Since 1980s, the outbreak of drug-resistant (DR) TB is a growing public health concern in several countries. The emergence of resistance was due to spontaneous chromosomal mutations in *Mtb* strains, which gained a selective advantage thanks to the misuse of antitubercular drugs. At the beginning, *Mtb* became resistant to single drugs (for example to rifampicin), but in recent years more and more forms of multi (MDR), extensively (XDR) and totally drug-resistant (TDR) TB occurred due to the progressive accumulation of mutations causing resistance [4, 7].

In 2015, about 3.9% of new cases and 21% of resurgent cases of TB were caused by MDR or rifampicin resistant (RR) *Mtb* strains. Globally, MDR or RR-TB had an incidence of 580 000 and a mortality of 250 000 in 2015. China, India and the Russian Federation accounted for the 45% of the cases. An estimated 9.5% of MDR-TB is due to XDR strains of the bacillus [1].

1.2. Therapy

TB mortality in absence of any treatment is high. Tiemersma E.W. *et al.* [8] investigated the period before the introduction of antitubercular drugs and estimated death rates within 10 years from the onset of TB to be 70% for people with sputum smear-positive and 20% for individuals with culture-positive (but smear-negative) pulmonary TB. Patients are diagnosed as smear-positive if their sputum shows the presence of bacteria when observed by a microscope, as culture-positive if *Mtb* is found by culture methods (the reference standard to diagnose TB). The most infectious cases of TB are the smear-positive ones. In order to limit and prevent TB deaths and diffusion, an early and correct diagnosis followed by an adequate therapy is needed [1].

The first attempt to find a cure for TB was by Robert Koch at the end of 19th century. Unfortunately, the use of sterile filtrates from *Mtb* cultures did not confer any defense against infection in TB patients. In 1921, Albert Calmette and Camille Guérin succeeded in developing a vaccine from the attenuation of a laboratory strain of *Mycobacterium bovis*, which belongs to the same complex of *Mtb* and caused around 6% of European deaths from TB before the introduction of milk pasteurization practice. The stain grew for 13 years on culture and provided a reduction of *Mtb* virulence in animals. In addition, when tested on children, a decrease of 90% in death rates was observed. Currently, Bacillus Calmette-Guérin (BCG) is the only vaccine against TB for clinical use and it is the world's most widespread vaccine, with more than 3 billion vaccinated people [3, 9].

The first effective anti-TB drugs were developed in 1940s and the general knowledge of the bacillus and its specific antigens increased thanks to the *Mtb* genome sequencing in 1998. At present, a 6-month treatment with isoniazid, rifampicin, ethambutol and pyrazinamide (first-line drugs) is suggested for new cases of drug-susceptible TB (**Table 1**), with a success rate of 85% or more. RR and MDR cases of TB need a more expensive, more toxic and longer (at least 9-12 months) therapy. In general, anti-TB drug and preventive treatments need to confer protection for a long period because *Mtb* has a low replication speed (its reproduction lasts 15-20 h) and it often latently infects the host [1, 3].

Since HIV-associated and DR forms of TB require a specific attention in the clinical management, it is crucial to distinguish between the various types of the disease. Active TB diagnosis is usually carried out by sputum smear microscopy, rapid molecular tests and culture methods (that currently are the reference standard). Nowadays, the Xpert® MTB/RIF assay is the only rapid test recommended by WHO to diagnose TB. It is considerably more sensible and accurate compared to sputum smear microscopy and it provides results within 2 hours [1, 4].

The diagnosis of latent TB is different from the one of the active disease and it can occur by a tuberculin skin test or an interferon-gamma release assay. Treatment of latent infection is suggested in people with a higher risk to develop the active disease (among which patients with HIV coinfection). **Table 1** describes the suggested therapies for the various forms of the disease [4].

When *Mtb* infects a HIV-positive person, symptoms may be atypical and diagnosis difficult. For this reason, it is important to screen HIV-infected people for the presence of TB. The Xpert® MTB/RIF test provide an advantage in the identification of TB in these individuals, increasing the detection rate of 45% in comparison with sputum smear microscopy. Since HIV takes advantage from *Mtb* coinfection, a combination of antiretroviral and anti-TB therapies is suggested [3, 4].

DR forms of TB can be identified by different methods. The Xpert® MTB/RIF assay is more convenient than the standard culture methods because it simultaneously provides information about the presence/absence of TB and rifampicin resistance and it is considerably quicker. DR-TB treatment depends on various factors and usually combines first- and second-line drugs, often resulting in a high toxicity; some suggestions for MDR-TB are reported in **Table 1**. Handling of XDR-TB is very hard and it can cause up to 98% mortality in patients coinfected with HIV [1, 4].

Table 1. Recommended treatments for the various TB forms [4].

Type of Infection	Recommended Regimen	Comments
Active disease		
Newly diagnosed cases that are not multidrug-resistant	Isoniazid, rifampin, ethambutol, and pyrazinamide for 2 mo (intensive phase), followed by isoniazid and rifampin for 4 mo (continuation phase)	Pyridoxine supplementation recommended to prevent isoniazid-induced neuropathy
Multidrug-resistant disease	Four second-line antituberculosis drugs (as well as pyrazinamide), including a fluoroquinolone, a parenteral agent, ethionamide or prothionamide, and either cycloserine or para-aminosalicylic acid if cycloserine cannot be used	Initial treatment based on local disease patterns and pending drug-susceptibility results; later-generation fluoroquinolones (e.g., moxifloxacin or levofloxacin) preferred
Latent infection	Isoniazid at a dose of 300 mg daily for at least 6 mo and preferably for 9 mo	Recommended for 9 mo or more in HIV-infected persons; daily administration for 6 mo also an option but with lower efficacy; extension to 36 mo further reduces risk among HIV-positive patients in regions in which tuberculosis is endemic
	Isoniazid at a dose of 900 mg plus rifapentine at a dose of 900 mg weekly for 3 mo (directly observed therapy)	Studied with directly observed therapy in predominantly HIV-uninfected persons; higher completion rates and equal efficacy, as compared with isoniazid for 9 mo
	Rifampin at a dose of 600 mg daily for 4 mo	Shown to be effective in persons with silicosis
	Isoniazid at a dose of 300 mg plus rifampin at a dose of 600 mg daily for 3 mo	Effective alternative for HIV-infected persons
	Isoniazid at a dose of 900 mg plus rifampin at a dose of 600 mg twice weekly for 3 mo	Another effective alternative for HIV-infected persons

The treatment success rates reported in the *Global Tuberculosis Report 2016* [1] are 83% for drug-susceptible TB, 75% for people coinfected with HIV, 52% in case of MDR or RR-TB and only 28% in patients with XDR-TB. In addition, studies on the effectiveness of BCG pointed out that it prevents severe forms of the disease in children, but not in adults and that its efficacy substantially decrease in 10-15 years from the administration. Therefore, new effective agents for the prevention and cure of TB are needed. For this reason, one of the goals proposed by WHO in the context of the End TB Strategy is to increase research and innovation [1, 10].

Many efforts were made so far to develop new diagnostic tests, vaccines and drugs. In the diagnostic field, progress made in recent years includes rapid tests that can become reference standards, assays for the detection of resistance to first- and second-line drugs and technologies to predict the transition from latent to active TB. In 2015, 13 potential vaccines against TB, among which whole-cell derived

compounds, recombinant viral-vectored platforms, proteins combined with adjuvants and extracts from mycobacteria, were in clinical trials (**Figure 3**). These vaccines were designed to prevent infection, progression or reactivation of *Mtb*. In addition, 9 new drugs against TB were in Phase I, Phase II or Phase III trials in 2016 (**Figure 4**). The latest drugs and combination regimens were developed for drug-susceptible, MDR and latent forms of the disease [1].

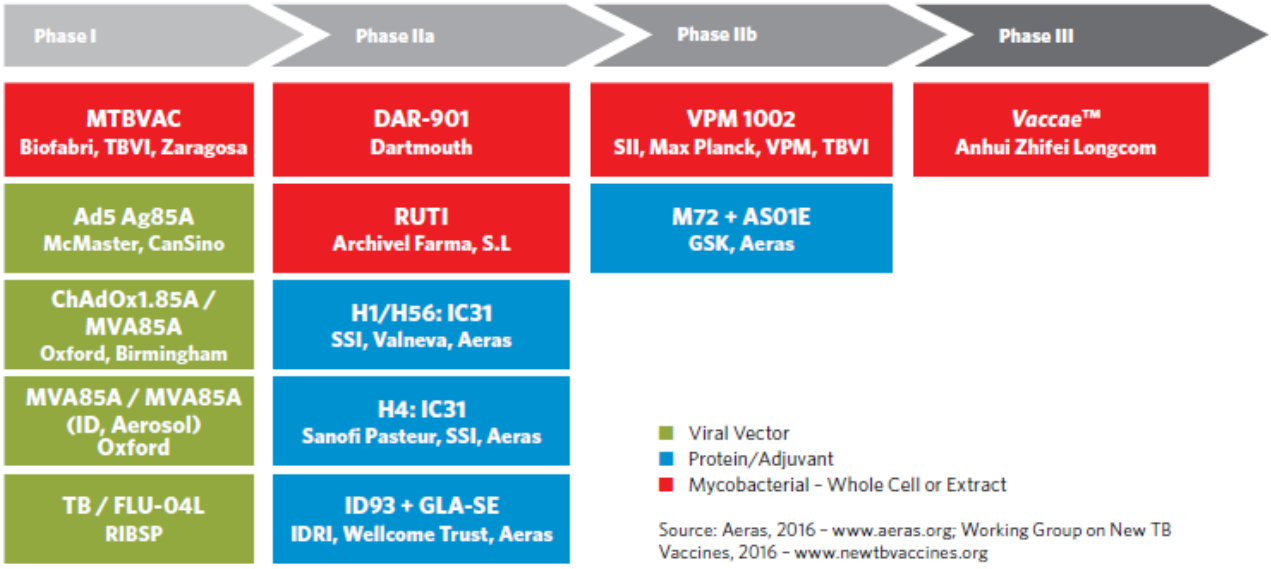


Figure 3. Status of potential TB vaccines under development in August 2015 [1].

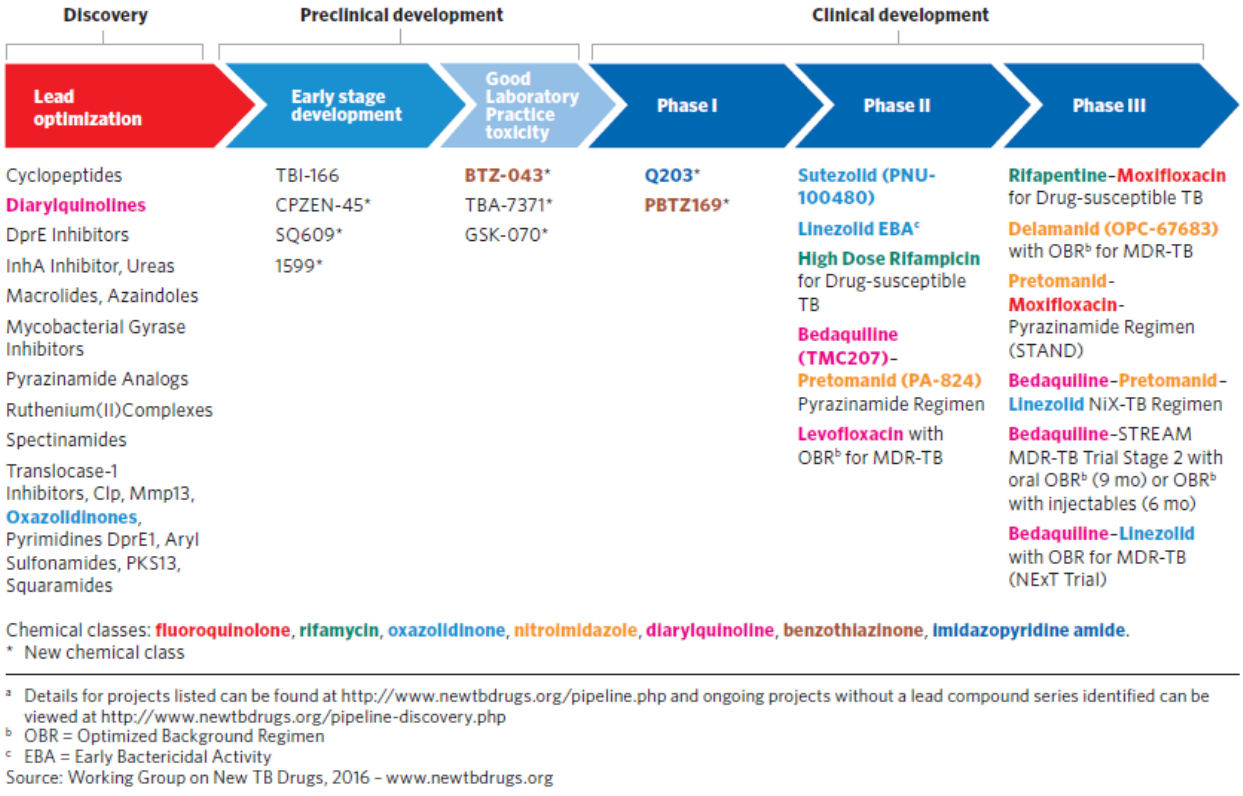


Figure 4. Status of new anti-TB drugs under development in August 2016 [1].

One of the main current strategies for the production of effective antitubercular vaccines is the use of antigenic proteins from *Mtb*, either combined with other antigens (fusion proteins) and adjuvants or expressed by viruses. 8 out of the 13 vaccines in clinical trials (**Figure 3**) are based on this approach, mostly exploiting protein antigens from ESAT-6 and Ag85 families for the stimulation of an immune response. The majority of new drugs against TB (**Figure 4**) acts blocking essential functions of the bacillus, like the synthesis of components needed for *Mtb* survival. In particular, many antimycobacterial drugs are enzyme inhibitors [1].

2. Background and aim of the work

In order to meet the need of new effective anti-TB agents, the present PhD project focuses on the investigation of two possible strategies to fight the disease: the design of a new glycoconjugate vaccine able to stimulate both humoral and T cell-mediated responses and the discovery of new antitubercular drugs characterized by an innovative mechanism of action (new molecular targets).

2.1. Glycoconjugate vaccines

The conjugation of carbohydrate antigens with a carrier protein is an effective strategy to enhance the immunogenicity of glycans that, unlike antigenic proteins, usually exhibit low affinity for their specific antibody. The efficacy of a carbohydrate-protein conjugate was proven for the first time in 1931, by the immunization of rabbits against Type III pneumococci [11]. Since that moment, various glycoconjugate vaccines have been developed, among which those against *Haemophilus influenzae* type b, *Streptococcus pneumoniae* and *Neisseria meningitides* are the most widespread [12, 13].

In general, bacterial polysaccharides stimulate only B cell immune responses. Nevertheless, saccharide-protein conjugates are recognized by the B cell receptor specific for polysaccharide antigens and they are internalized into the cells, where various proteases digest the protein carrier. The deriving peptide epitopes are exposed on the B cell surface and activate T cells, with a consequent stimulation of high-affinity IgG antibodies production. Thus, glycoconjugate vaccines take advantage of carbohydrate moieties including B cell epitopes and protein structures able to induce a T-dependent memory response, resulting in a long-term protection [12]. The use of antigenic carrier proteins from the same pathogen of polysaccharides has been demonstrated to increase the efficacy of the conjugate, resulting in an improved immunogenicity compared to the single constituents [14].

The production of glycoconjugates can be laborious and, if not controlled, can lead to non-homogeneous products with different features in terms of pharmacokinetics and immunogenicity or

with a decreased efficacy due to the masking of protein epitopes, hindering their clinical use. Therefore, well-defined glycoprotein therapeutics are needed, in which sugar moieties, glycosylation sites and glycan loading are uniform and deeply characterized. Innovative strategies for oligosaccharide synthesis and site-specific chemical coupling to proteins can be exploited to obtain glycovaccines with clear structure-activity relationships, thus enhancing their safety and efficacy [12, 15, 16].

The characterization of *neo*-glycoconjugates needs to be performed at different levels, both structural and functional (**Figure 5**). At the structural level, intact mass measurements allow to assess glycoprotein identity, purity, integrity and glycosylation profile in terms of yield and glycoform composition. Mass spectrometry with soft ionization techniques like matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) is the method of choice for intact protein/glycoprotein analysis, generally using time-of-flight (TOF), ion trap or Fourier transform ion cyclotron resonance (FT-ICR) analyzers. The coupling of liquid chromatography (LC) to mass spectrometry (MS) can provide simpler spectra useful in case of complex samples. The characterization of intact proteins/glycoproteins in terms of presence of size variants or aggregates and amount of glycans is an important structural investigation for biopharmaceuticals and it is usually performed by size exclusion chromatography (SEC), dynamic light scattering (DLS) or electron microscopy (EM). A detailed and updated study on these analytical approaches has been carried out in the context of the present PhD thesis (*Chapter 7*, pp. 130-151). The study at the peptide/glycopeptide level can elucidate the position of glycosylation sites, confirm protein sequences, localize eventual modifications and provide information on the glycan moieties. The (glyco)peptides are usually obtained through proteolytic digestion and analyzed by LC-MS. In order to trap and then to separate the glycosylated peptides, porous graphitized carbon (PGC) columns, hydrophilic interaction liquid chromatography (HILIC) and anion exchange chromatography (AEC) are widely used, generally followed by fragmentation in the mass analyzer through MS^n [15, 17-19].

The coupling procedure should not alter the tertiary structure or induce degradation of the carrier protein, as well as mask antigenic amino acids [20]. For these reason, a defined correlation between structure and function is required. To investigate protein antigenic regions and direct the glycosylation to non-antigenic amino acids, thus preserving the immunogenicity of the construct, valuable information can be derived from epitope identification. MS-based epitope determination approaches are among the most common and involve antigen digestion by a proteolytic enzyme, incubation with the ligand, a series of washing steps to remove unbound peptides and MS analysis of the elution fractions to identify the peptides involved in the binding [21]. Moreover, surface plasmon

resonance (SPR) biosensor technology has proven to be a precious technique for vaccine characterization, as it quantifies binding affinities and provides reliable kinetic data to evaluate specificity and mechanism of action. Noteworthy, SPR can be used to monitor immune responses directly in serum samples [22].

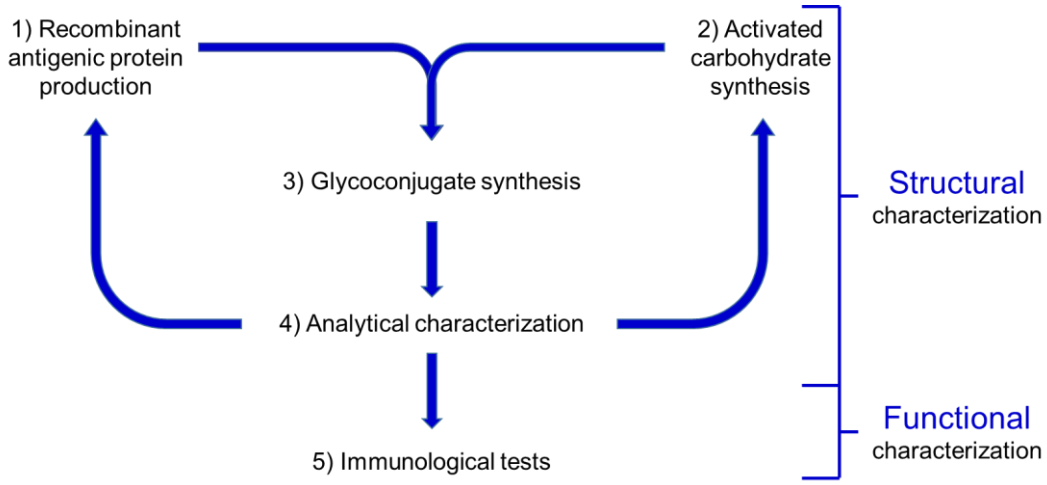


Figure 5. Phases of glycoconjugate vaccines preparation [23].

2.1.1. A new glycoconjugate vaccine against TB

A glycovaccine against TB can be obtained by conjugation of antigenic proteins, which are expressed by *Mtb* and are known to induce a strong immune reaction, with sugars able to act as LAM mimics, since LAM is the most important saccharidic surface antigen of *Mtb* [24].

In this study, TB10.4 and Ag85B were selected as carrier proteins since they are two of the most potent antigen species expressed by *Mtb* yet identified [10]. TB10.4 is a low molecular mass protein (11076.4 Da, **Figure 6**) belonging to the ESAT-6 family that is highly conserved in different *Mtb* clinical isolates of different geographical origin. Even if its biological function is still unknown, it was found to be strongly recognized by T cells in cultures of peripheral blood mononuclear cells isolated from patients with active TB and individuals vaccinated with BCG and to exert a protective effect against *Mtb* in mice immunized with this antigen [25, 26]. Ag85B is a mycolyl transferase with a medium molecular mass (31346.0 Da, **Figure 6**) and it is part of the Ag85 complex, consisting of the most abundantly secreted proteins by *Mtb*. Among them, Ag85B is one of the most potent antigen species, which has proven its anti-TB protective efficacy in murine and guinea pigs as well as other animal models and has been shown to induce both cellular and humoral immune responses [26-28]. A reliable method for the production and purification of the two proteins was developed, as described in a paper previously published by our research group [10].

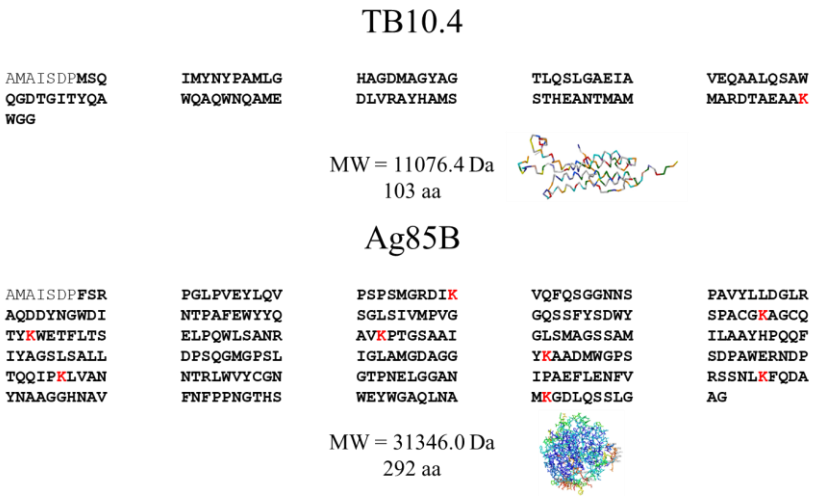


Figure 6. Amino acidic sequence, molecular weight (MW) and three-dimensional structure of TB10.4 and Ag85B recombinant proteins evaluated in the present study. The amino acids not in bold (AMAISDP) represent a pre-sequence originating from the cloning procedure, those in bold represent recombinant protein sequences encoded by the synthetic genes. Lysine residues (involved in the conjugation reaction) are in red.

The choice of carbohydrate antigens (saccharides of mannose) was based on literature data on *Mtb* reporting the presence of saccharidic surface structures able to stimulate the humoral response. Mannose was selected because it is the linker monosaccharide in LAM, a lipoglycan constituting an essential component of the *Mtb* cell wall (**Figure 7**). Studies on immune responses in lungs of patients with active TB showed the production of specific anti-LAM antibodies [29]. In addition, ManLAM probably represents a virulence factor, since mono-, di- or tri-mannosylation is characteristic for LAM of pathogenic mycobacteria [5, 30]. Our research group described the procedure for the synthesis of the sugars used in this project in [31].

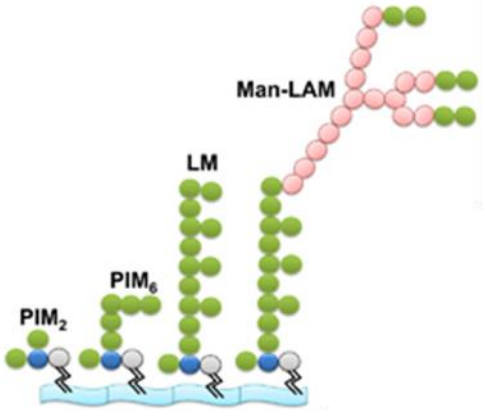
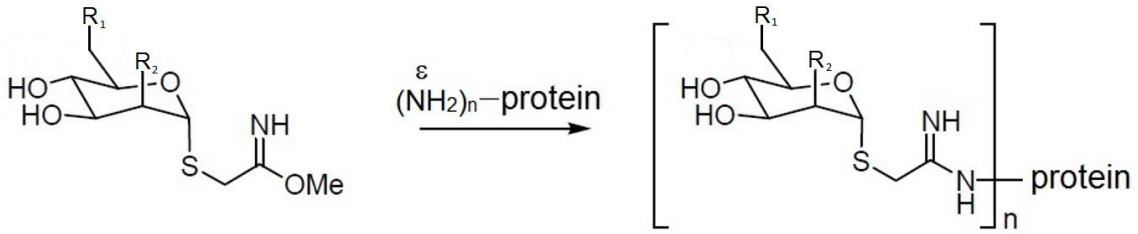


Figure 7. *Mtb* cell wall associated LAM and related glycoconjugates: lipomannan (LM) and phosphatidyl-*myo*-inositol mannosides (PIMs). Mannose units are in green, inositol in blue, phosphate in gray, arabinose in pink [30].

Among the various saccharide-protein conjugation methods, those targeting defined amino acids (like lysine, cysteine or aromatic amino acids) are reported to be the most effective [20]. In particular, our

group demonstrated the high selectivity of 2-iminomethoxyethyl (IME) thioglycosides towards lysine residues and optimized the reaction conditions to obtain pure and homogeneous *neo*-glycoconjugates [15]. **Figure 8** shows the reaction procedure with the optimized experimental conditions.



Experimental conditions: 100 mM Na tetraborate buffer pH 9.5, 5.5 mg/mL protein concentration, 1 mM benzamidine hydrochloride, 200/1 molar ratio sugar/protein, 37 °C, 24 h

Figure 8. Procedure followed for the synthesis of *neo*-glycoconjugates. Man-IME: R_1, R_2 = OH; Man(1→6)Man-IME: R_1 = α-mannose, R_2 = OH; Man(1→6)Man(1→6)Man-IME: R_1 = mannose(1→6)mannose, R_2 = OH; Ara(1→6)Man-IME: R_1 = α-arabinose, R_2 = OH; Ara(1→2)Man-IME: R_1 = OH, R_2 = α-arabinose.

Starting from this background, the aim of the present PhD project was the analytical characterization of the compounds in the various phases of the production process (**Figure 5**) to improve the rational design of the vaccine candidates. Different and complementary analytical techniques were employed to deeply investigate the products.

At first, the stability of recombinant proteins and derived *neo*-glycoconjugates was evaluated. In the latest years, HILIC has clearly emerged as a powerful technique to analyze intact glycoproteins [32-35]. Therefore, this chromatographic approach has been selected and applied to monitor compounds stability. *Chapter 1* of the experimental part of the present PhD thesis (pp. 25-36) describes the development and validation of a simple and rapid HILIC-UV method, which proved its suitability for the estimation of protein degradation under critical conditions (*i.e.* freeze-thaw cycles) and for the monitoring of the coupling reaction with saccharidic moieties, without the need of sample preparation [23]. To identify the degradation products observed in HILIC-UV and to better assess the performances of HILIC in the separation of glycoproteins, the coupling of this separation technique to high resolution TOF-MS and the use of three different columns were examined (*Chapter 2*, pp. 37-49) [36].

The two HILIC methods revealed the occurrence of degradation phenomena for different lots of the tested recombinant proteins, which were correlated to the presence of residual traces of enterokinase (EK) from the production process (EK is a specific protease commonly used in solution to release mature proteins from their chimeric constructs). Based on the experience of our research group in the

immobilization of enzymes [37-40] and on the presence in literature of papers reporting the successful immobilization of EK on different supports [41-43], the synthesis of an EK-immobilized enzyme reactor (IMER) appeared as an interesting approach to hydrolyze the fusion proteins in flow, thus preventing product contamination. Various IMERs were prepared and tested considering different immobilization chemistries and materials, as described in *Chapter 3* (pp. 50-72).

The *ex vivo* evaluation of the immunogenic activity of the different *neo*-glycoconjugates compared to non-glycosylated antigens [44] prompted us to consider new strategies to improve compounds effectiveness. The immunogenicity of glycoconjugate vaccines should be ensured by both antigenic protein/peptide and carbohydrate contribution, through the recognition of B and T cell epitopes [12]. Therefore, both moieties were investigated. As regards to the carrier protein, different experimental and computational methods suggested that Lys30 and Lys282 (the two most reactive amino acids of Ag85B involved in the glycosylation) are antigenic [44]. Conservative mutations of these two lysines to arginine residues were planned to prevent their glycosylation, thus preserving the immunogenicity of the constructs. The contribution of glycans was investigated by the conjugation of mono-, di- and tri-saccharides of mannose to Ag85B protein and its mutants.

A structural characterization of the various compounds was carried out by different analytical techniques (*Chapter 4*, pp. 73-93). Protein identity and purity, sugar activation yields and glycosylation profiles were defined by intact mass measurements in ESI-linear ion trap MS. A capillary reverse phase LC-UV-MS/MS system was exploited to confirm protein sequences, after proteolytic digestion by chymotrypsin. To characterize glycosylation sites, the *neo*-glycoconjugates were hydrolyzed by chymotrypsin. The released glycopeptides were selectively enriched by on-line solid phase extraction (SPE) and subsequently separated and identified by HILIC-MS/MS.

A correlation between structure and function of the obtained products was then performed to support the mutagenesis approach (*Chapter 5*, pp. 94-115). The areas of Ag85B protein involved in the interaction with antibodies from different sources (commercial monoclonal antibody from mouse, antibodies from sera of TB-patients, BCG-vaccinated people and a healthy control) were investigated using the proteolytic epitope extraction-MS method. Affinities of Ag85B protein, its mutants and its glycosylated forms for the commercial monoclonal antibody were then quantified by SPR analysis. This study revealed the suitability of the developed SPR method as an alternative screening assay for the immunological evaluation of antigenic proteins from *Mtb*.

2.2. Enzymes as drug targets

Among the available drug targets, enzymes hold a leading position due to some advantageous features. In fact, they play a fundamental role in all life processes, their activity is generally altered in pathological conditions and their peculiar structure makes them "druggable", since enzymes possess active sites particularly suited for high-affinity binding of small molecules [45, 46].

Nowadays, enzyme inhibitors constitute almost half of the drugs in clinical use and, noteworthy, they are usually administered by mouth. Therapeutic enzyme inhibitors have different mechanisms of action (drug-enzyme interactions can be reversible or irreversible, inhibitors can bind to the ground or the transition state of the enzyme, the inhibition can be competitive or noncompetitive) and are available for almost all categories of human pathologies (among which infectious diseases, cancer, cardiovascular problems, metabolic and neurological disorders) [45, 47].

In antibacterial drug discovery, the effectiveness of enzyme inhibition as a therapeutic strategy was proven in 1977 by Cohen S.S. on *Herpesvirus*, *Influenza A virus* and Protozoa [48]. Enzyme targets for the development of antibacterial agents should possess the following features: essential role in the life of the organism to design a drug able to inhibit bacterial growth or cause its death; common structure in different bacterial species to obtain a broad-spectrum antimicrobial agent; structure different enough from the human host to avoid toxic effects and finally presence of binding sites for small molecules [49].

Thanks to the increasing knowledge about bacteria molecular biology, target-directed screening procedures are becoming more and more common. In this panorama, the rational design of enzyme inhibitors usually starts from a high-throughput screening (HTS) of chemical libraries consisting of potential drug candidates. Common HTS methods for enzyme inhibitors research include radiometric, spectrophotometric or chromatographic assays. An activity assay specific for the target enzyme is usually developed to initially compare the percentage of inhibition induced by different compounds, using a fixed concentration of enzyme, substrate and tested inhibitor. After this screening step, IC_{50} (inhibitor concentration that reduces the enzymatic activity to 50%) and K_i (enzyme-inhibitor dissociation constant) are determined for the most interesting compounds in order to quantify the inhibitor potency and the affinity for its target. The activity assay in this case is performed in presence of increasing amounts of inhibitors [45, 49, 50].

2.2.1. *Mtb*PNP as an innovative target for the development of antimycobacterial drugs

In *Mtb*, the purine salvage pathway appears to be an interesting source of enzyme targets for the design of antitubercular drugs. This pathway is involved in the synthesis of nucleic acids, since its role is the conversion of adenine, guanine, and hypoxanthine nucleobases in their corresponding nucleotides (AMP, GMP, and IMP). Probably, the bacillus switches from the *de novo* to the salvage pathway especially in conditions of energy lack and during phases of rapid multiplication [50, 51].

The enzymes implicated in nucleotide metabolism usually differ from those of the human host, thus constituting good targets for the development of selective inhibitors. In fact, in 2003 Long M.C. and colleagues [52] proved the effectiveness of the nucleoside analogue 2-methyladenosine targeting adenosine kinase, an enzyme of the purine salvage pathway, as a selective antimycobacterial agent. Therefore, new drugs inhibiting enzymes in this pathway might satisfy the need of agents active against resistant TB strains and latent disease [51].

Purine nucleoside phosphorylase (PNP, EC 2.4.2.1) attracts particular interest since evidences suggest the presence of differences in structure and transition state between mycobacterial (*Mtb*PNP) and human (*Hs*PNP) enzymes. Both *Mtb*PNP and *Hs*PNP are homotrimeric enzymes that catalyze the reversible cleavage of the glycosidic bond of purine (deoxy)ribonucleosides in presence of inorganic phosphate as a second substrate to generate the nucleobase and α-D-(deoxy)ribose-1-phosphate (**Figure 9**). However, from the comparison of the two three-dimensional structures, Tyr188 emerged as the only residue involved in the binding with the ligand not conserved in human (in which Phe200 takes its place). In addition, Phe153 in *Mtb*PNP seems to interact with the 5' of the substrate sugar group, while the corresponding Phe159 in human behaves like a lid for the binding site of the adjacent subunit [50, 51, 53, 54].

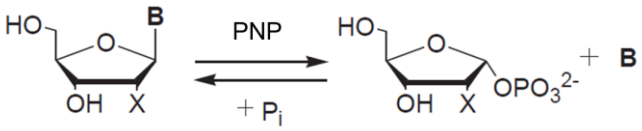


Figure 9. Reaction catalyzed by PNP. X = OH, H; B = purine base; Pi = inorganic phosphate [50].

In the context of the present PhD project, an enzymatic activity assay towards mycobacterial and human PNPs was developed in order to identify selective inhibitors as potential antitubercular lead-candidates. Starting from the promising results obtained by our research group [50], the conditions for the assay were adapted to the available laboratory equipment. Enzymatic reactions were conducted in batch and activity was assessed by monitoring the phosphorolysis of inosine to hypoxanthine. To this aim, a new HILIC-UV method was developed. After the definition of *Mtb*PNP

and *Hs*PNP kinetics, the assay was applied to the screening of 6- and 8-substituted nucleosides as potential inhibitors. The enzymatic assay allowed also to characterize the most interesting compounds (selected from the screening) in terms of inhibitory potency and affinity for their target, as described in *Chapter 6* (pp. 116-129) of the present PhD thesis (Experimental part).

References

[1] World Health Organization, *Global tuberculosis report 2016*, Geneva, Switzerland, WHO Press, **2016**.

[2] Bozzano F. et al., *Immunology of tuberculosis*, Mediterr J Hematol Infect Dis. **2014**, 6(1), e2014027.

[3] Lawn S.D., Zumla A.I., *Tuberculosis*, Lancet. **2011**, 378, 57-72.

[4] Zumla A. et al., *Tuberculosis*, N Engl J Med. **2013**, 368(8), 745-55.

[5] Philips J.A., Ernst J.D., *Tuberculosis pathogenesis and immunity*, Annu Rev Pathol. **2012**, 7, 353-384.

[6] Rao M. et al., *B in TB: B cells as mediators of clinically relevant immune responses in tuberculosis*, Clin Infect Dis. **2015**, 61(Suppl 3), S225–S234.

[7] Smith T. et al., *Molecular biology of drug resistance in Mycobacterium tuberculosis*, Curr Top Microbiol Immunol. **2013**, 374, 53-80.

[8] Tiemersma E.W. et al., *Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review*, PLoS One. **2011**, 6(4), e17601.

[9] Gupta U.D. et al., *Current status of TB vaccines*, Vaccine. 2007, 25(19), 3742-3751.

[10] Piubelli L. et al., *Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins*, Microb Cell Fact. **2013**, 12, 115.

[11] Avery O.T., Goebel W.F., *Chemo-immunological studies on conjugated carbohydrate-proteins*, J Exp Med. **1931**, 54(3), 437-447.

[12] Adamo R. et al., *Synthetically defined glycoprotein vaccines: current status and future directions*, Chem Sci. **2013**, 4(8), 2995-3008.

[13] Vella M., Pace D., *Glycoconjugate vaccines: an update*, Expert Opin Biol Ther. **2015**, 15(4), 529-546.

[14] Campodónico V.L. et al., *Efficacy of a conjugate vaccine containing polymannuronic acid and flagellin against experimental Pseudomonas aeruginosa lung infection in mice*, Infect Immun. **2011**, 79(8), 3455-3464.

[15] Temporini C. et al., *Liquid chromatography-mass spectrometry structural characterization of neo glycoproteins aiding the rational design and synthesis of a novel glycovaccine for protection against tuberculosis*, J Chromatogr A. **2014**, 1367, 57-67.

[16] Grayson E.J. et al., *A coordinated synthesis and conjugation strategy for the preparation of homogeneous glycoconjugate vaccine candidates*, Angew Chem Int Ed Engl. **2011**, 50(18), 4127-4132.

[17] Staub A. et al., *Intact protein analysis in the biopharmaceutical field*, J Pharm Biomed Anal. **2011**, 55(4), 810-822.

[18] Sandra K. et al., *Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization*, J Chromatogr A. **2014**, 1335, 81-103.

[19] Brusotti G. et al., *Advances on size exclusion chromatography and applications on the analysis of protein biopharmaceuticals and protein aggregates: a mini review*, Chromatographia. **2018**, 81(1), 3-23.

[20] Monsigny M. et al., *Neoglycoproteins*, Comprehensive Glycoscience: From Chemistry to Systems Biology, Elsevier **2007**, 3.23, 477-521.

[21] Opuni K.F.M. et al., *Mass spectrometric epitope mapping*, Mass Spectrom Rev. **2016**, doi: 10.1002/mas.21516.

[22] Hearty S. et al., *Surface plasmon resonance for vaccine design and efficacy studies: recent applications and future trends*, Expert Rev Vaccines. **2010**, 9(6), 645-664.

[23] Rinaldi F. et al., *Application of a rapid HILIC-UV method for synthesis optimization and stability studies of immunogenic neo-glycoconjugates*, J Pharm Biomed Anal. **2017**, 144, 252-262.

[24] Chan C.E. et al., *The diagnostic targeting of a carbohydrate virulence factor from M.Tuberculosis*, Sci Rep. **2015**, 5, 10281.

[25] Skjøt R.L.V. et al., *Comparative evaluation of low-molecular-mass proteins from Mycobacterium tuberculosis identifies members of the ESAT-6 family as immunodominant T-cell antigens*, Infect. Immun. **2000**, 68, 214-220.

[26] Dietrich J. et al., *Exchanging ESAT6 with TB10.4 in an Ag85B fusion molecule-based tuberculosis subunit vaccine: efficient protection and ESAT6-based sensitive monitoring of vaccine efficacy*, J Immunol. **2005**, 174(10), 6332-6339.

[27] Wiker H.G., Harboe M., *The antigen 85 complex: a major secretion product of Mycobacterium tuberculosis*, Microbiol Rev. **1992**, 56(4), 648-661.

[28] Kadir N.A. et al., *Cellular and humoral immunogenicity of recombinant Mycobacterium smegmatis expressing Ag85B epitopes in mice*, Int J Mycobacteriol. **2016**, 5(1), 7-13.

[29] Demkow U. et al., *Humoral immune response against mycobacterial antigens in bronchoalveolar fluid from tuberculosis patients*, J Physiol Pharmacol. **2005**, 56 (Suppl 4), 79-84.

[30] Mishra A.K. et al., *Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in Mycobacterium tuberculosis physiology and host-pathogen interaction*, FEMS Microbiol Rev. **2011**, 35(6), 1126-1157.

[31] Bavaro T. et al., *Chemoenzymatic synthesis of neoglycoproteins driven by the assessment of protein surface reactivity*, RSC Advances **2014**, 4, 56455-56465.

[32] Pedrali A. et al., *Characterization of intact neo-glycoproteins by hydrophilic interaction liquid chromatography*, Molecules. **2014**, 19(7), 9070-9088.

[33] Zhang Z. et al., *Polyacrylamide brush layer for hydrophilic interaction liquid chromatography of intact glycoproteins*, J Chromatogr A. **2013**, 1301, 156-161.

[34] Tetaz T. et al., *Hydrophilic interaction chromatography of intact, soluble proteins*, J Chromatogr A. **2011**, 1218(35), 5892-5896.

[35] Periat A. et al., *Potential of hydrophilic interaction chromatography for the analytical characterization of protein biopharmaceuticals*, J Chromatogr A. **2016**, 1448, 81-92.

[36] Tengattini S. et al., *Hydrophilic interaction liquid chromatography-mass spectrometry as a new tool for the characterization of intact semi-synthetic glycoproteins*, Anal Chim Acta. **2017**, 981, 94-105.

[37] Calleri E. et al., *Development of a bioreactor based on trypsin immobilized on monolithic support for the on-line digestion and identification of proteins*, J Chromatogr A. **2004**, 1045(1-2), 99-109.

[38] Temporini C. et al., *Optimization of a trypsin-bioreactor coupled with high-performance liquid chromatography-electrospray ionization tandem mass spectrometry for quality control of biotechnological drugs*, J Chromatogr A. **2006**, 1120(1-2), 121-131.

[39] Temporini C. et al., *Pronase-immobilized enzyme reactor: an approach for automation in glycoprotein analysis by LC/LC-ESI/MS^n*, Anal Chem. **2007**, 79(1), 355-363.

[40] Calleri E. et al., *Immobilized purine nucleoside phosphorylase from Aeromonas hydrophila as an on-line enzyme reactor for biocatalytic applications*, J Chromatogr B Analyt Technol Biomed Life Sci. **2014**, 968, 79-86.

[41] Suh C.W. et al., *Covalent immobilization and solid-phase refolding of enterokinase for fusion protein cleavage*, Process Biochem. **2005**, 40(5), 1755–1762.

[42] Kubitzki T. et al., *Application of immobilized bovine enterokinase in repetitive fusion protein cleavage for the production of mucin 1*, Biotechnol J. **2009**, 4(11), 1610-1618.

[43] Santana S.D. et al., *Immobilization of enterokinase on magnetic supports for the cleavage of fusion proteins*, J Biotechnol. **2012**, 161(3), 378-382.

[44] Bavaro T. et al., *Glycosylation of recombinant antigenic proteins from Mycobacterium tuberculosis: in silico prediction of protein epitopes and ex vivo biological evaluation of new semi-synthetic glycoconjugates*, Molecules. **2017**, 22(7), 1081.

[45] Copeland R.A., *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*, John Wiley & Sons, Inc. **2013**

[46] Copeland R.A. et al., *Targeting enzyme inhibitors in drug discovery*, Expert Opin Ther Targets. **2007**, 11(7), 967-978.

[47] Robertson J.G., *Mechanistic basis of enzyme-targeted drugs*, Biochemistry. **2005**, 44(15), 5561-5571.

[48] Cohen S.S., *A strategy for the chemotherapy of infectious disease*, Science. **1977**, 197(4302), 431-432.

[49] Silver L.L., *Challenges of Antibacterial Discovery*, Clin Microbiol Rev. **2011**, 24(1), 71-109.

[50] Cattaneo G. et al., *Development, validation and application of a 96-well enzymatic assay based on LC-ESI-MS/MS quantification for the screening of selective inhibitors against Mycobacterium tuberculosis purine nucleoside phosphorylase*, Anal Chim Acta. **2016**, 943, 89-97.

[51] Ducati R.G. et al., *Purine salvage pathway in Mycobacterium tuberculosis*, Curr Med Chem. **2011**, 18(9), 1258-1275.

[52] Long M.C. et al., *Identification and characterization of a unique adenosine kinase from Mycobacterium tuberculosis*, J Bacteriol. **2003**, 185(22), 6548-6555.

[53] Shi W. et al., *Structures of purine nucleoside phosphorylase from Mycobacterium tuberculosis in complexes with immucillin-H and its pieces*, Biochemistry. **2001**, 40(28), 8204-8215.

[54] Caceres R.A. et al., *Crystal structure and molecular dynamics studies of purine nucleoside phosphorylase from Mycobacterium tuberculosis associated with acyclovir*, Biochimie. **2012**, 94(1), 155-165.

EXPERIMENTAL PART

1. *Application of a rapid HILIC-UV method for synthesis optimization and stability studies of immunogenic neo-glycoconjugates*

Francesca Rinaldi, Sara Tengattini, Enrica Calleri, Teodora Bavaro, Luciano Piubelli, Loredano Pollegioni, Gabriella Massolini, Caterina Temporini

*Journal of Pharmaceutical and Biomedical Analysis 2017, 144, 252-262**

Journal of Pharmaceutical and Biomedical Analysis 144 (2017) 252–262

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Application of a rapid HILIC-UV method for synthesis optimization and stability studies of immunogenic *neo*-glycoconjugates

F. Rinaldi[a], S. Tengattini[a], E. Calleri[a], T. Bavaro[a], L. Piubelli[b,c], L. Pollegioni[b,c], G. Massolini[a], C. Temporini[a,*]

[a] Department of Drug Sciences, University of Pavia, Via Taramelli 12, 27100 Pavia, Italy
[b] Department of Biotechnology and Life Sciences, University of Insubria, via J.H. Dunant 3, I-21100 Varese, Italy
[c] The Protein Factory Interuniversity Centre, Politecnico of Milano and University of Insubria, via Mancinelli 7, I-20131, Milano, Italy

ARTICLE INFO

Article history:
Received 28 October 2016
Received in revised form 1 March 2017
Accepted 26 March 2017
Available online 28 March 2017

Keywords:
Neo-glycoproteins
HILIC-UV
Protein stability
Intact protein analysis

ABSTRACT

Proteins and glycoproteins with therapeutic activity are susceptible to environmental factors, which can cause their degradation and the loss of their activity. Thus, the maintenance of their stability during the production process is a critical factor. In this work, a simple and rapid hydrophilic interaction liquid chromatography HILIC-UV method was validated in terms of accuracy, precision, linearity, LOD, LOQ and specificity and applied to the investigation of the stability of intact proteins and their *neo*-glycoconjugates with antigenic activity against tuberculosis. The method proved to be suitable for the estimation of the degradation of the proteins under critical conditions (i.e. freeze-thaw cycles) and for the monitoring of their coupling reaction with saccharidic moieties, without the need of sample preparation. In addition, the chromatographic analysis allowed to calculate the yields of the protein glycosylation reaction.

1. Introduction

Biopharmaceuticals are particular categories of pharmaceutical substances for which the active ingredients are generally proteins (monoclonal antibodies and recombinant versions of endogenous proteins), polypeptides and oligonucleotides. The market of these products has grown rapidly in the last years and biopharmaceutical industry needs to increase the speed of product development, but still ensuring structural integrity and safety of novel biologic molecules [1]. In fact, in comparison to conventional small-molecule drugs, the chemical nature of proteins makes them less stable during production and storage. Biopharmaceuticals can be exposed to a number of stress conditions such as critical reaction conditions (i.e. medium and ionic content), temperature changes, freeze-thawing and light exposure and these harsh environmental factors can affect the native structure of a protein resulting in a possible modification of physico-chemical properties and therapeutic activity [2–4].

The labile stability of proteins implicates the need of user-friendly and rapid analytical methods to assess protein solubility, integrity and stability during all the steps of research and development studies, starting from the protein production, generally by recombinant DNA technology, to the biological activity evaluation [2].

Among biopharmaceuticals, the development of glycoconjugate vaccines, in which oligosaccharides are conjugated to carrier proteins (*neo*-glycoproteins), is one of the safest and more reliable strategies to prevent and control deadly infectious diseases [5]. This approach has been recently considered by our research group for the development of novel and more effective vaccines against tuberculosis (TB) [6], still one of the world's deadliest human diseases [7].

Potential TB vaccines have been synthesized by the conjugation of two antigenic *Mycobacterium tuberculosis* (MTB) proteins, namely TB10.4 and Ag85B [8], with 2-iminomethoxyethyl (IME) activated saccharides [9]. TB10.4 is a low molecular mass protein (11076.4 Da) belonging to the ESAT-6 family that is highly conserved in different MTB clinical isolates and that is strongly recognized by T cells [10]. Ag85B possesses a medium molecular mass (31346.0 Da) and it is part of the Ag85 complex, the most abundantly secreted proteins by MTB. Among them, Ag85B is one of the most potent antigen species yet identified, which is able to induce both humoral and cell-mediated immune response in MTB-infected subjects [11].

For the *neo*-glycoconjugates preparation, the two antigenic proteins were glycosylated using simple saccharidic moieties activated

* Corresponding author.
E-mail address: caterina.temporini@unipv.it (C. Temporini).

http://dx.doi.org/10.1016/j.jpba.2017.03.052
0731-7085/

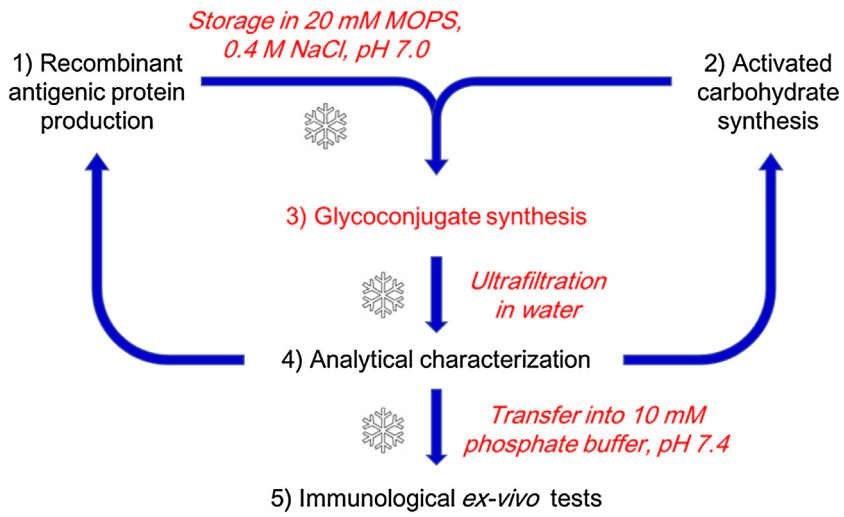


Fig. 1. Phases of the glycoconjugate vaccine preparation. The critical steps and conditions are reported in red. ❄ Symbol indicates where freeze-thaw cycles are considered. MOPS = 3-(N-Morpholino) PropaneSulfonic acid.

with the IME reagent [9], which selectively reacts with the ε-amino group of lysine residues [6]. TB10.4 includes only one lysine residue, but under strict reaction conditions generates two glycoforms, with a second saccharidic unit bound to the N-terminal amino acid [6]. Ag85B possesses 8 lysine residues, thus its glycosylation may produce glycoforms containing up to 8 sugar moieties [12].

In this project, the workflow followed to obtain glycoconiugates for biological evaluation is reported in Fig. 1. The critical steps and conditions requiring an analytical control are in red. In this regard a rapid, efficient and informative analytical tool, able to support the synthetic steps without laborious sample pre-treatment and to assess protein solubility and stability both during the synthesis process and storage conditions is required.

HPLC based methods in combination with UV and MS detection are the simplest and most common approaches currently used in biopharmaceutical analysis. Most of the protein chemical modification and degradation events can be evaluated and monitored by reverse phase chromatography (RP-LC) coupled to mass spectrometry (MS) [3,4]. However, in the case of chemical glycosylation RP-LC often lacks in selectivity to discriminate non-glycosylated proteins from glycosylated ones, and between different degrees of glycosylation. To date, some pioneering publications have demonstrated that analysis of intact natural and semi-synthetic glycoproteins can be successfully performed by hydrophilic interaction liquid chromatography (HILIC) [13–16]. Starting from a previously described HILIC-UV method developed on a model protein [13], we here describe its optimization in order to obtain a user-friendly and rapid tool for solubility and stability evaluation of non-conjugated TB10.4 and Ag85B under freeze-thaw stress. The method was validated and applied to the monitoring of the glycosylation process and glycoprotein stability studies.

2. Materials and methods

2.1. Reagents and chemicals

Sodium tetraborate (99%), benzamidine hydrochloride hydrate (98%), $HClO_4$ (70%), glycerol solution (86–89%), potassium phosphate dibasic (≥98%), trifluoroacetic acid (TFA, ≥99%), acetic acid (≥99.85%), NaCl, 3-(N-Morpholino) PropaneSulfonic acid (MOPS) and ribonuclease A (RNase A) from bovine pancreas were purchased from Sigma-Aldrich (Milan, Italy). Hydrochloric acid (37%) was from Carlo Erba Reagenti (Milan, Italy). Water was obtained from a Direct-QTM system Millipore (Millipore, Milan, Italy). Acetonitrile (ACN) and methanol HPLC grade were purchased by Sigma-Aldrich. 2-Iminomethoxyethyl (IME) thioglycosides, mannose-IME (Man-IME) and mannose(1 → 6)mannose-IME (ManMan-IME), were prepared according to the previously reported method [9]. Recombinant TB10.4 and Ag85B were prepared according to Piubelli et al. [8]. Identity and purity of the two proteins were confirmed by direct infusion (DI) of the samples in electrospray ionization mass spectrometry (ESI–MS) [8]. The deconvoluted mass spectra for the two antigenic proteins are reported in Fig. 2.

2.2. Glycosylation procedure

The glycosylation of TB10.4 and Ag85B was performed following a previously optimized protocol [6]. The antigenic proteins were produced by fermentation in *E. coli* [8] and stored in 20 mM MOPS, 0.4 M NaCl, pH 7.0 at −80 °C. Proteins were conjugated with the saccharidic moieties (synthesized and activated as in [9]) in sodium tetraborate buffer, 100 mM, pH 9.5.

Briefly, proteins were transferred in the conjugation buffer (final concentration 5.5 mg/mL) and then the solution was mixed with IME-glycoside in a glycoside/protein molar ratio of 200/1. Benzamidine hydrochloride (1 mM) was added to the reaction mixture to prevent the digestion by residual enterokinase derived from the protein production process [8]. The reaction mixture was vortexed for 1 min and incubated at 37 °C under continuous stirring. The obtained products underwent an ultrafiltration step in water before their structural characterization by ESI–MS. Finally, they were transferred into 10 mM phosphate buffer, pH 7.4 for the biological tests.

2.3. Sample preparation

Protein solutions (50 μg/mL) for system development (paragraph 3.1 and 3.2) were prepared in a water/methanol mixture (10/90 v/v) containing 10 mM $HClO_4$.

Stock solutions of TB10.4 and Ag85B (approximately 500 μg/mL) for stability studies (paragraph 3.3) were prepared in different conditions by ultrafiltration for buffer exchange [6]:

i) 20 mM MOPS, 0.4 M NaCl, pH 7.0 (post-production phase);
ii) water (post-glycosylation phase);
iii) 10 mM phosphate buffer, pH 7.4 (biological test phase).

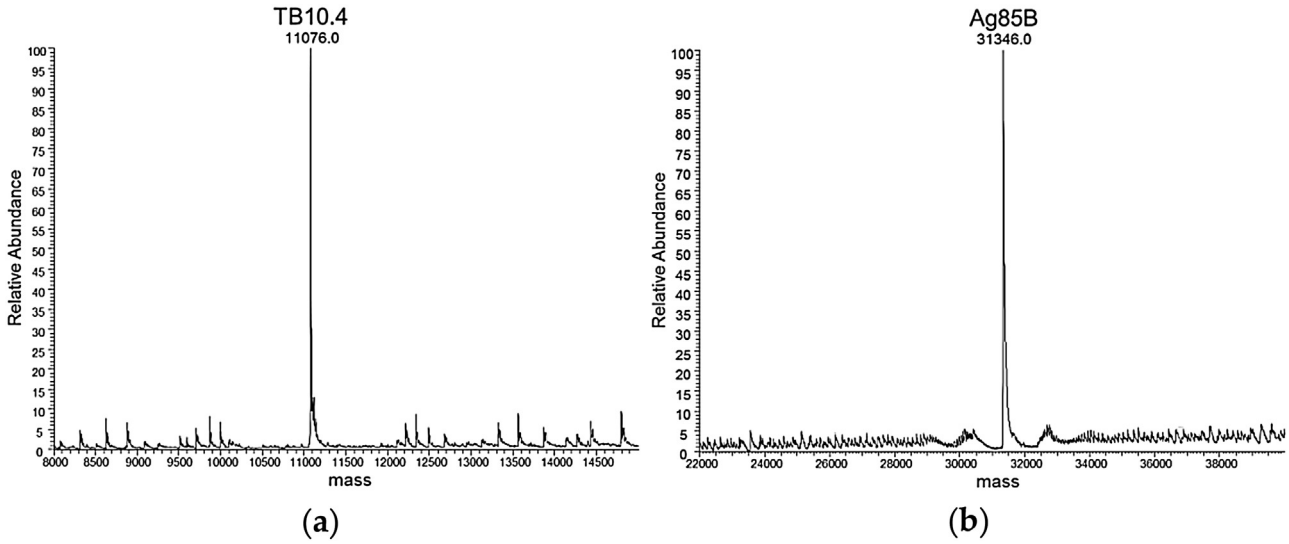


Fig. 2. Deconvoluted spectra of the TB10.4 (a) and Ag85B (b) samples used in this study.

The solutions were then diluted with methanol containing 10 mM $HClO_4$ to obtain working solutions at 50 μg/mL.

Stability studies were carried out either in the presence or absence of 10% glycerol.

For the stability study during the glycosylation procedure, 5 mg/mL or 1 mg/mL solutions of TB10.4 and 5.5 mg/mL solution of Ag85B in 100 mM sodium tetraborate buffer, pH 9.5, 1 mM benzamidine hydrochloride were used. *Neo*-glycoproteins were analyzed without purification by neutralizing the pH with HCl (37%) and diluting samples to a final concentration of 50 μg/mL in methanol/water (90/10 v/v) containing 10 mM $HClO_4$.

2.4. Instrumentation and HILIC chromatographic conditions

Chromatographic separations of intact proteins were performed on an Agilent HPLC series 1100 system (Santa Clara, CA, USA), equipped with mobile-phase online degasser, quaternary pump, autosampler, column thermostated compartment and diode array detector. For data acquisition and analysis, the HP ChemStation for LC software version Rev. A.06.03 [509] was used.

Protein and *neo*-glycoprotein separations were performed using a TSKgel Amide-80 (2 × 150 mm, 3 μm, 80 Å) column from Tosoh Bioscience (Montgomeryville, PA, USA). The mobile phase was composed of ACN (solvent A) and water (solvent B) both containing 10 mM $HClO_4$. Chromatographic conditions consisted of a linear gradient from 20% to 30% B in 20 min followed by isocratic elution at 30% B for 10 min. Column equilibration time was 70 min. Equilibration time between runs was set at 23 min. The column temperature was maintained at 50 °C, the injection volume was 10 μL and elution was carried out at constant flow of 200 μL/min. In all experiments, proteins were detected at 210 nm.

Intact MS experiments were carried out on a Linear Trap Quadrupole (LTQ) mass spectrometer with ESI source (Thermo Finnigan, San Jose, CA, USA), as described in [8]. The system was controlled by Xcalibur software 1.4 (Thermo Finnigan). The purified samples were directly introduced into the mass spectrometer with a 100 μL syringe moved at 10 μL/min by a syringe pump (Thermo Fisher Scientific, San Jose, CA, USA), using a concentration of 300 μg/mL in water/ACN (50/50 v/v) + 1% acetic acid. Full scan intact MS experiments were carried out under the following instrumental conditions: positive ion mode; mass range, 700–2000 *m/z*; source voltage, 4.5 kV; capillary voltage, 35 V; sheat gas, 15 (arbitrary units); auxiliary gas, 2 (arbitrary units); capillary temperature, 220 °C; tube lens voltage, 140 V. Multiple-charged protein ion signals were deconvoluted by using Bioworks Browser (Thermo Electron, revision 3.1) and the percentage abundance of

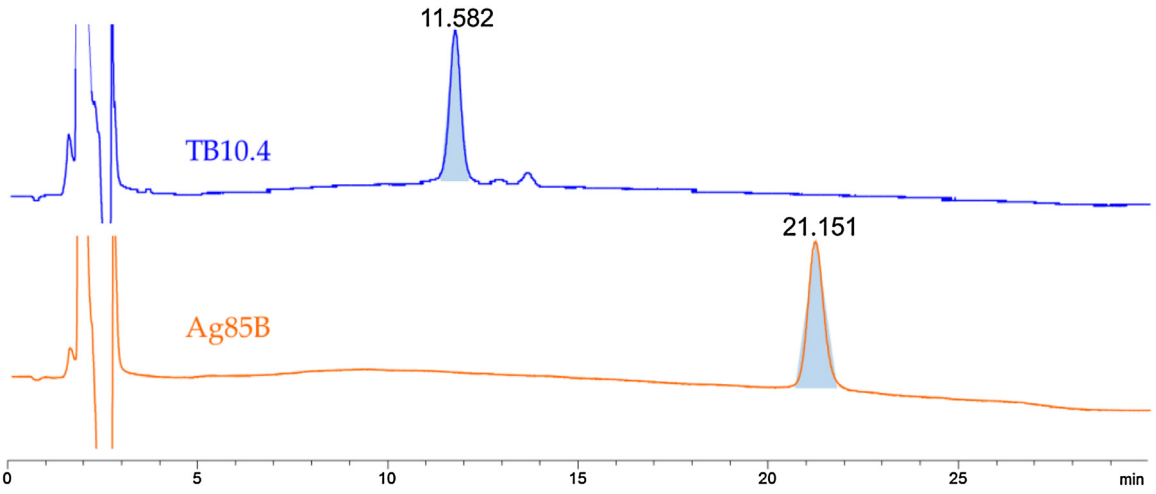


Fig. 3. Representative chromatograms of the antigenic proteins TB10.4 and Ag85B obtained applying the optimized conditions (50 μg/mL).

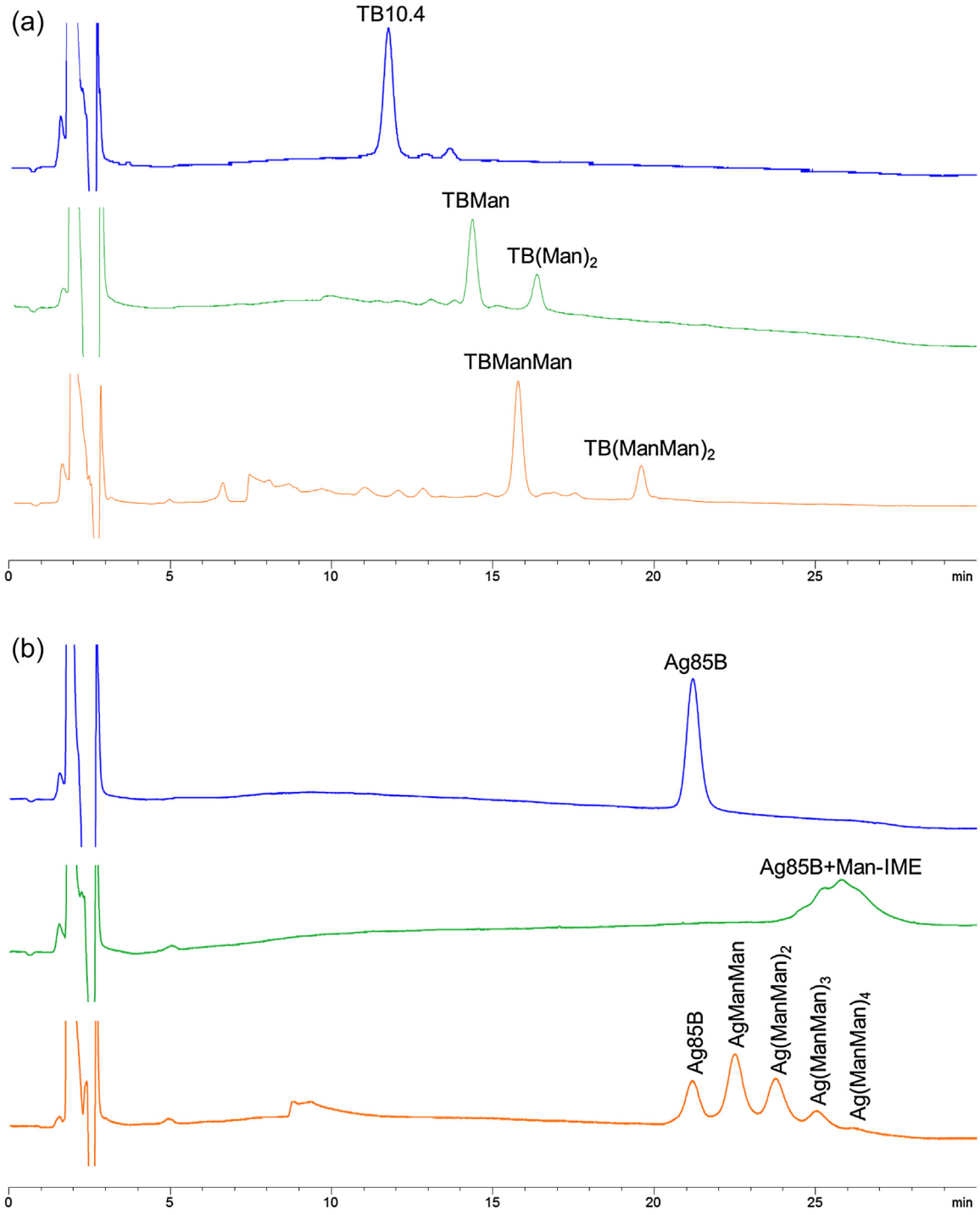


Fig. 4. Representative chromatograms of TB10.4 (a) and Ag85B (b) proteins compared with their glycosylated forms. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

different glycoforms calculated by the relative abundance of corresponding peaks in the deconvoluted spectra.

Theoretical masses and grand average of hydropathicity (GRAVY) values were deduced from the amino acid sequence using the "*ProtParam*" tool on ExPASy (http://web.expasy.org/protparam/).

2.5. HILIC-UV method validation

The parameters addressed to assess the performance of the analytical procedure were specificity, precision, linearity, accuracy, limit of detection (LOD) and limit of quantitation (LOQ).

The specificity of the method was intended as its ability to discriminate between non-glycosylated and glycosylated proteins. Chromatograms confirming the separation of those species are

Table 1
Recovery data obtained for RNase A commercial protein. Samples corresponding to 75, 100 and 125% of the test concentration were analyzed. SD: standard deviation. RSD: relative standard deviation.

Sample	Conc. (μg/mL)	Area (mAu × s)	Calculated conc. (μg/mL)	recovery %
		1835.5	39.4	104.99%
75%	37.5	1820.8	39.1	104.14%
		1695.9	36.3	96.84%
		2370.3	51.1	102.17%
100%	50	2294.4	49.4	98.84%
		2285.3	49.2	98.44%
		2946.4	63.7	101.91%
125%	62.5	3026.1	65.4	104.71%
		2827.2	61.1	97.74%
			% average recovery	101.09%
			SD	0.03
			RSD	3.14%

reported, and identity was assumed by chromatographic behaviour and DI-ESI–MS spectra of the same samples.

Precision was considered at two levels: repeatability and intermediate precision (inter-day). Repeatability was assessed using 21 determinations for RNase A (7 concentrations/3 replicates each) and 15 determinations for TB10.4 and Ag85B proteins (5 concentrations/3 replicates each). To determine the inter-day precision, 6 analyses of 50 μg/mL protein solutions were carried out in two different days (3 replicates for each day). Relative standard deviation (RSD) was calculated for each concentration level and the mean RSD was reported.

The linearity was evaluated analyzing the standard protein RNase A, followed by the antigenic proteins TB10.4 and Ag85B. For each protein, 5-point calibration curves were obtained by the least squares method. A concentration range of 12.5–100 μg/mL was investigated and three independent determinations performed at each concentration.

Accuracy was calculated as percent recovery of samples with a known added amount of analyte. The reference protein RNase A was analyzed at three different concentrations (37.5, 50 and 62.5 μg/mL), while TB10.4 and Ag85 B recovery was calculated only at the test concentration (50 μg/mL). Three replicates were performed for each concentration level.

LOD and LOQ were defined as follows:

$$\mathrm{LOD} = C_S \frac{3}{S/N}$$

$$\mathrm{LOQ} = C_S \frac{10}{S/N}$$

where C_s is the concentration of the injected analyte and S/N is the signal-to-noise ratio.

The reported LOQ and LOD were calculated as the mean of the values obtained for each determination.

Table 2
Recovery data obtained for TB10.4 recombinant protein. The measurements were performed at the test concentration.

Sample	Conc. (μg/mL)	Area (mAu × s)	Calculated conc. (μg/mL)	recovery %
		2705.1	50.7	101.43%
100%	50	2658.4	49.8	99.68%
		2637.6	49.5	98.91%
			% average recovery	100.01%
			SD	0.01
			RSD	1.29%

Table 3
Recovery data obtained for Ag85B recombinant protein. The measurements were performed at the test concentration.

Sample	Conc. (μg/mL)	Area (mAu × s)	Calculated conc. (μg/mL)	recovery %
		4664.6	52.0	103.91%
100%	50	4629.7	51.6	103.14%
		4719.7	52.6	105.13%
			% average recovery	104.06%
			SD	0.01
			RSD	0.97%

3. Results and discussion

3.1. Optimization of HILIC-UV method and immunogenic protein analysis

In a previous paper, an HILIC-UV method was reported for the analysis of intact *neo*-glycoproteins prepared starting from RNase A (124 amino acids, 13681 Da) as a model protein [13]. The method entailed the use of a carbamoyl TSKgel Amide-80 column and mobile phase composed of ACN (A) and water (B) both containing 10 mM $HClO_4$. Based on the good results achieved, we decided to start with the same analytical conditions for the analysis of the two MTB protein antigens (TB10.4 and Ag85B) and their glycoconjugates, and the possible impurities and degradation products present in the samples.

Under the chromatographic conditions described by Pedrali et al. (linear gradient from 25% to 35% B in 20 min followed by isocratic elution for 10 min) [13], intact TB10.4 and Ag85B proteins were poorly retained on HILIC column. A lower percentage of aqueous mobile phase (20% B) was therefore used to increase their retention. Under these chromatographic conditions, TB10.4 elutes at 11.60 ± 0.08 min (n = 3) and Ag85B elutes at 21.15 ± 0.09 min (n = 3). The elution order is in agreement with the grand average of hydropathicity (GRAVY) values of the two proteins (−0.208 and −0.279 for TB10.4 and Ag85B, respectively).

It is well known that in HILIC separations the composition of the sample solvent significantly affects peak shape and efficiency [17,18]. Meanwhile, protein solubility problems can arise due to the high organic solvent content required by HILIC at the beginning of the analyses. Thus, the sample preparation procedure was investigated in order to identify the injection solvent that allowed to achieve the greater chromatographic efficiency and solubility of the studied proteins.

This study was carried out using TB10.4 as a model protein. Starting from the solvent composition used in [13], namely ACN/water (50/50 v/v), different solvents such as methanol, ACN and their mixture with water and acids (TFA and $HClO_4$) were considered to prepare solutions with the final concentration of 50 μg/mL. The best results were obtained with a mixture of methanol and water (90/10 v/v) containing 10 mM $HClO_4$. Compared to the initial conditions, this solvent mixture approximately duplicated the peak area, reflecting an increased solubility both for TB10.4 and Ag85B.

In Fig. 3 the chromatographic traces of TB10.4 and Ag85B in the optimized solvent conditions are reported.

3.2. HILIC-UV method validation

3.2.1. Specificity

The specificity was assessed as the ability of the method to discriminate between non-glycosylated and glycosylated proteins. A good separation between the antigenic proteins [blue lines in Fig. 4 panel a) and b)] and their glycoconjugates was observed using both Man-IME [green lines in Fig. 4 panel a) and b)] and ManMan-IME

Table 4
Average area, standard deviation (SD) and relative standard deviation (RSD) of TB10.4 and Ag85B peaks (500 μg/mL stock solution; injected samples were diluted 1:10) in 10 mM phosphate buffer, pH 7.4, after 2 freeze-thaw cycles in absence or presence of 10% glycerol.

	TB10.4					Ag85B				
	Time	Average area	SD (n = 3)	RSD (%)	Recovery (%)	Time	Average area	SD (n = 3)	RSD (%)	Recovery (%)
No glycerol	0	1319.6	22.7	1.7	100.0	0	3536.3	117.1	3.3	100.0
	24	1240.1	n.d.	n.d.	94.0	72	3457.0	100.6	2.9	97.8
	96	1334.2	n.d.	n.d.	101.1	96	3481.3	72.2	2.1	98.4
10% glycerol	0	1300.2	21.3	1.6	100.0	0	3450.6	280.6	8.1	100.0
	24	1205.2	n.d.	n.d.	92.7	72	3033.7	37.5	1.2	87.9
	96	1231.8	n.d.	n.d.	94.7	96	2915.4	104.8	3.6	84.5

n.d., not determined

[orange lines in Fig. 4 panel a) and b)] as saccharidic moieties. Unfortunately, the method did not allow to discriminate between the different glycoforms of Ag85B glycosylated with Man-IME [green line, Fig. 4 panel b)], but it was possible to completely resolve the non-glycosylated protein from the glycoconjugates.

3.2.2. Precision

The repeatability, calculated as relative standard deviation (RSD), was evaluated by triplicate injections of the standard RNase A at 7 different concentrations (ranging from 12.5 to 100 μg/mL). A mean RSD of 3.11% for areas and of 0.88% for retention times was obtained.

The repeatability for TB10.4 and Ag85B samples was calculated on three determinations at 5 different concentration levels from 12.5 to 100 μg/mL. Since in a preliminary phase Ag85B showed a high variability (RSD >10%), we hypothesized that either protein solubility and/or solution homogeneity problems might occur. To circumvent these problems, the sample preparation procedure was optimized by including a 5 s vortex mixing to both the stock solution and the prepared dilutions before each LC-UV analysis.

A very good repeatability was obtained in terms of area (average RSD = 1.93% for TB10.4 and 2.34% for Ag85B) and retention time (average RSD = 0.44% for TB10.4 and 1.41% for Ag85B). These results make the chromatographic method suitable for the qualitative evaluation of protein identity (retention time) and for a semiquantitative evaluation of protein concentration (peak areas).

Since this work aims to evaluate the protein stability during storage, the assessment of the inter-day variability is required. Thus, the intermediate precision was established by analyzing 50 μg/mL protein solutions in two different days. Three replicates were performed on each day. The good precision of the method is demonstrated by the RSD values of peak areas: 1.93% for RNase A, 1.74% for TB10.4 and 2.03% for Ag85B.

3.2.3. Linearity

The linearity of the optimized LC-UV method was first established for the commercial protein RNase A. The investigated concentration range was 12.5–100 μg/mL, corresponding to 25–200% of the test concentration. 5 levels were considered (12.5, 25, 50, 75 and 100 μg/mL) and three independent determinations were performed for each point. The response was linear and the equation of the calibration curve was y = 45.671 x + 37.295 (R^2 = 0.9998).

The linearity was then assessed for TB10.4 and Ag85B antigens, prepared as recombinant proteins in "The Protein Factory" Research Center (Politecnico di Milano and Università degli Studi dell'Insubria). The concentration of the standard protein solution was determined by Bradford assay. Five-point calibration curves were obtained for the two recombinant proteins in the same concentration range used for the standard protein RNase A. Three replicates were performed for each point. Calibration equations of y = 53.552x – 10.716 (R^2 = 0.9998) and y = 90.322x – 28.021 (R^2 = 0.9975) were obtained for TB10.4 and Ag85B, respectively.

Correlation coefficient (R^2) was found to be greater than 0.99 for all the proteins, thus the developed method showed very good linearity.

3.2.4. Accuracy

Accuracy was determined as percent recovery of analyte in the sample. As an initial stage, the recovery of the reference protein RNase A was evaluated on three concentration levels (corresponding to 75, 100 e 125% of the test concentration). Three replicates were analyzed for each concentration. Recovery ranged from 96.84 to 104.99%, with an average RSD of 3.14%. The complete data are presented in Table 1.

The percent recovery of TB10.4 and Ag85B samples analyzed at the test concentration was also defined. Recovery values are presented in Tables 2 (TB10.4) and 3 (Ag85B).

The mean recovery percentages ranged from 100.01 to 104.06% for the three proteins, with average RSD from of 3.14% to 0,97%, corroborating the accuracy of the method.

3.2.5. Limit of detection (LOD) and limit of quantitation (LOQ)

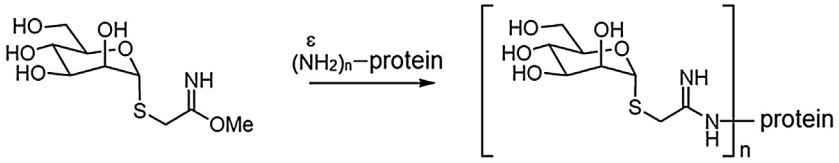
LOD and LOQ were determined as described *in paragraph 2.5*. The detection limit was 3 μg/mL for RNase A, 2 μg/mL for TB10.4 and 2 μg/mL for Ag85B. The quantitation limits for RNase A, TB10.4 and Ag85B were 11, 7 and 7 μg/mL, respectively.

3.3. Stability evaluation of immunogenic proteins

During the synthetic process and immunological evaluation, the protein antigens are stored in different buffers and can be submitted to one or more freeze-thaw cycles (Fig. 1). It is well known that freeze-thawing combines a number of stressing factors that can potentially cause protein precipitation and/or aggregation [19,20]. The stability of TB10.4 and Ag85B solutions was thus investigated after subjecting them to 2 freeze-thaw cycles in different buffer solutions. Noteworthy, experiments were designed on unconjugated proteins since it is well known that glycosylation can increase the overall stability of proteins [21].

Table 5
Recovery (expressed as percentage) over time of TB10.4 (5 or 1 mg/mL solutions in glycosylation buffer).

Time (h)	Recovery%	
	(5 mg/mL)	(1 mg/mL)
0	100	100
1	99	86
5	72	86
6	56	86
21	11	11
22	8	9
24	0	9



Scheme 1. Synthesis of *neo*-glycoproteins by coupling of IME-activated mannose (Man-IME) with antigenic proteins.

In this study, the considered buffers were: i) 20 mM MOPS, 0.4 M NaCl, pH 7.0 used as protein storage buffer [22]; iii) 10 mM phosphate buffer, pH 7.4, used to solubilize the *neo*-glycoproteins during *ex-vivo* immunological assays. In addition, as protein analysis and structural investigation (e.g. by DI-MS [8] and capillary electrophoresis (CE) [22]) can require the buffer removal, protein solutions in plain water were also considered (ii). The stability of TB10.4 and Ag85B added with 10% glycerol was also investigated as glycerol is routinely employed for the stabilization of aqueous solutions of proteins [23].

Protein solutions were analyzed at fixed time intervals (0, 24 and 96 h for TB10.4; 0, 72 and 96 h for Ag85B). The stability was quantified by recovery values calculated as area ratio percentage in respect to time zero value. 90% recovery was considered the critical threshold. The semiquantitative analysis indicates that the two proteins behave similarly regardless of the buffer and storage conditions used. Overall, in all tested conditions, and for the two antigenic proteins, stability was maintained for 96 h and after 2 freeze-thaw cycles.

Interestingly, only in the case of Ag85B, the addition of 10% glycerol causes a progressive reduction of peak areas and consequent recovery, suggesting that glycerol does not exert its positive effect on the solubility of Ag85B under the explored conditions. Although unexpected, it is reported that the effectiveness of glycerol in preserving protein solubility varies widely among different proteins, even if the reasons are not completely understood [23].

As an example, Table 4 reports the semiquantitative data obtained for TB10.4 and Ag85B in phosphate buffer with and without glycerol.

3.4. Immunogenic protein stability under conjugation conditions

The development of a *neo*-glycovaccine includes a chemical glycosylation step. In our study, TB10.4 and Ag85B were glycosylated with IME-activated saccharides. The IME group selectively reacts with the ε-amino function of the protein lysine residues (Scheme 1.). The glycosylation conditions are reported in paragraph 2.2. The moderately high protein concentration (5.5 mg/mL) and buffer pH (9.5), might affect the integrity of the antigens during the conjugation process. Therefore protein stability during the glycosylation phase was investigated.

3.4.1. TB10.4

The chromatographic trace of the glycosylation reaction of TB10.4 with Man-IME at 24 h was compared to chromatographic profile of the same unmodified protein. A significant reduction of the peak area was observed suggesting that precipitation/enzymatic degradation events might occur. To elucidate the cause of this phenomenon, the behaviour of wild-type TB10.4 was investigated throughout a simulation of the glycosylation reaction carried out in absence of activated glycans (i.e. in 100 mM sodium tetraborate buffer, pH 9.5, 1 mM benzamidine hydrochloride, 37 °C,

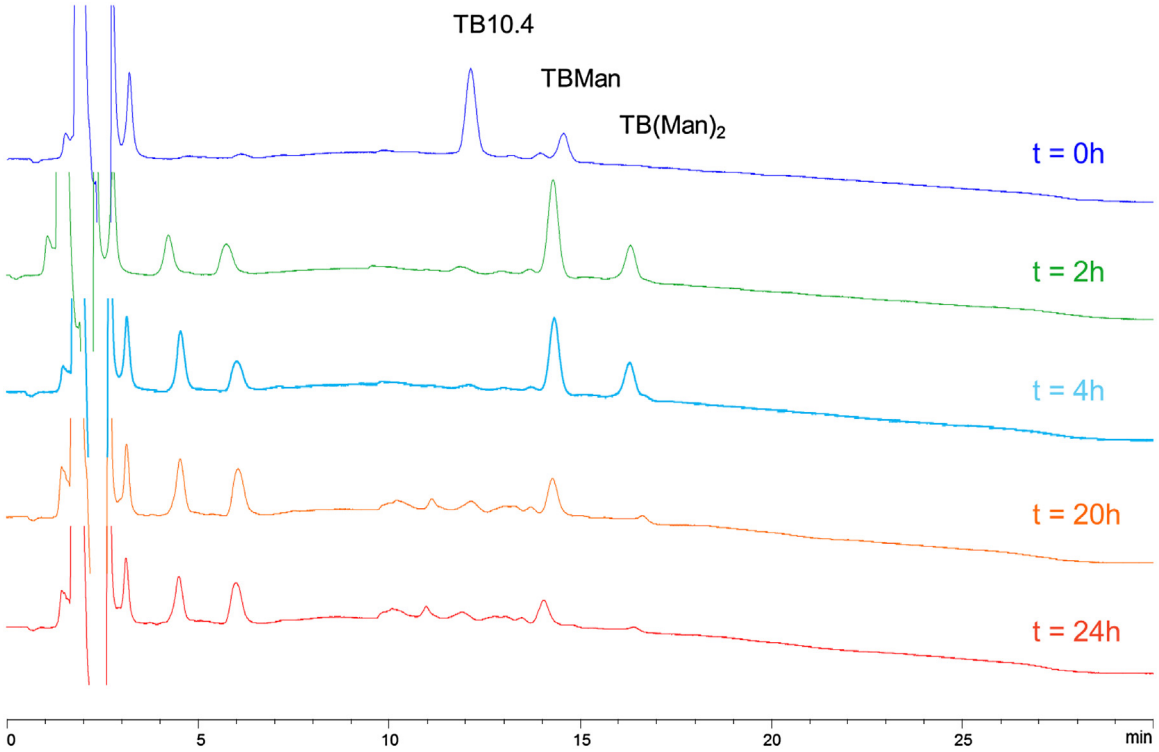


Fig. 5. Representative chromatograms of the antigenic protein TB10.4 conjugated with Man-IME (5.5 mg/mL in 100 mM sodium tetraborate buffer, pH 9.5, 1 mM benzamidine hydrochloride, glycosidic reagent/protein molar ratio of 200/1). The chromatograms were obtained at 0, 2, 4, 20 and 24 h.

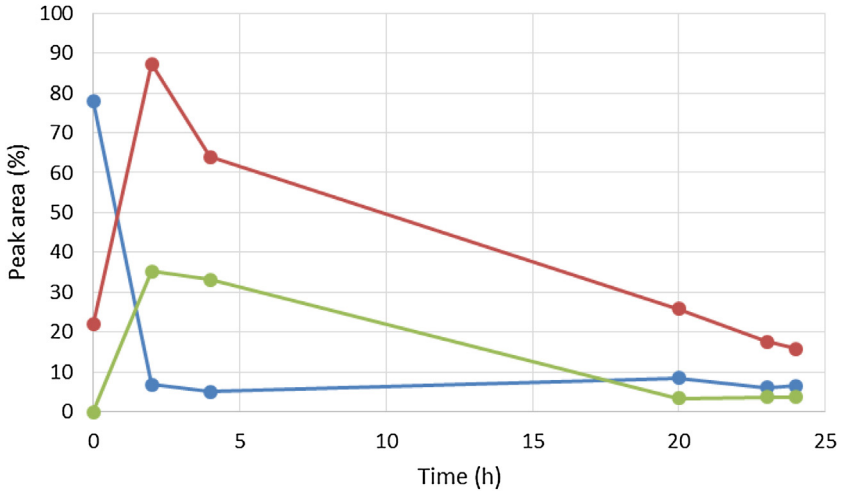


Fig. 6. Trend of the TB10.4 (blue), TB10.4-Man (red) and TB10.4-(Man)2 (green) recovered peak areas during the glycosylation reaction carried out under standard [6] conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

24 h). This study was performed both at 5 mg/mL and at 1 mg/mL to assess the influence of the protein concentration. Aliquots of TB10.4 solution were analyzed at fixed intervals for qualitative and semi-quantitative evaluation. The obtained data are reported in Table 5.

The analyses showed a gradual reduction of TB10.4 peak area during reaction time. This phenomenon seems to be accelerated in the most concentrated protein solution, even if an almost complete disappearance of the protein at the end of the reaction time (24 h) was observed for both the analyzed samples (Table 5).

In order to understand if the glycoconjugates could undergo the same gradual degradation of the protein, the same TB10.4 lot was subjected to chemical conjugation with Man-IME. Aliquots of the reaction mixture were therefore analyzed at the fixed time intervals. The chromatographic traces are reported in Fig. 5 and the recovered peak areas are represented in Fig. 6. The elution profile of the t0 sample (Fig. 4) clearly shows the presence of 2 main peaks. Considering their Rt, they have been identified as unconjugated TB10.4 (11.989 min) and mono-mannose derivative (14.393 min), suggesting that the kinetics of the conjugation reaction is extremely rapid. After 2 h, unmodified TB10.4 was minor than 10%, while a third peak appeared at 16.337 min. The DI-ESI–MS analysis of the same 2 h-sample (Fig. 7a) showed the presence of the expected mono-glycosylated isoform (11311.0 Da) and the occurrence of a di-glycosylated isoform (11547.0 Da). It was already demonstrated that even if TB10.4 possesses only one lysine residue, under harsh reaction conditions a second mannose unit binds at the N-terminal amino acid [6].

As the retention in HILIC is mainly controlled by the number of the glycan molecules, it can be reasonably speculated that the elution order reflects the glycosylation degree. Hence the developed HILIC-UV method enables the complete resolution of unmodified, mono- and di-glycosylated TB10.4, allowing the direct and rapid estimation of the relative abundance of each species without any sample preparation.

Following the chromatographic monitoring up to 24 h, also the glycoconjugates undergo a gradual degradation as already observed for the unmodified protein. Therefore, the reaction time was shortened to 2 h in order to reach the maximum yield and to preserve the protein integrity.

The new reaction protocol (2 h of reaction) was applied to the conjugation of TB10.4 with ManMan-IME: Fig. 8 reports the chromatographic traces of intact TB10.4 and conjugated TB10.4. Also in this experiment, direct MS infusion analysis was carried out (Fig. 7b): two glycoforms were detected in agreement with the

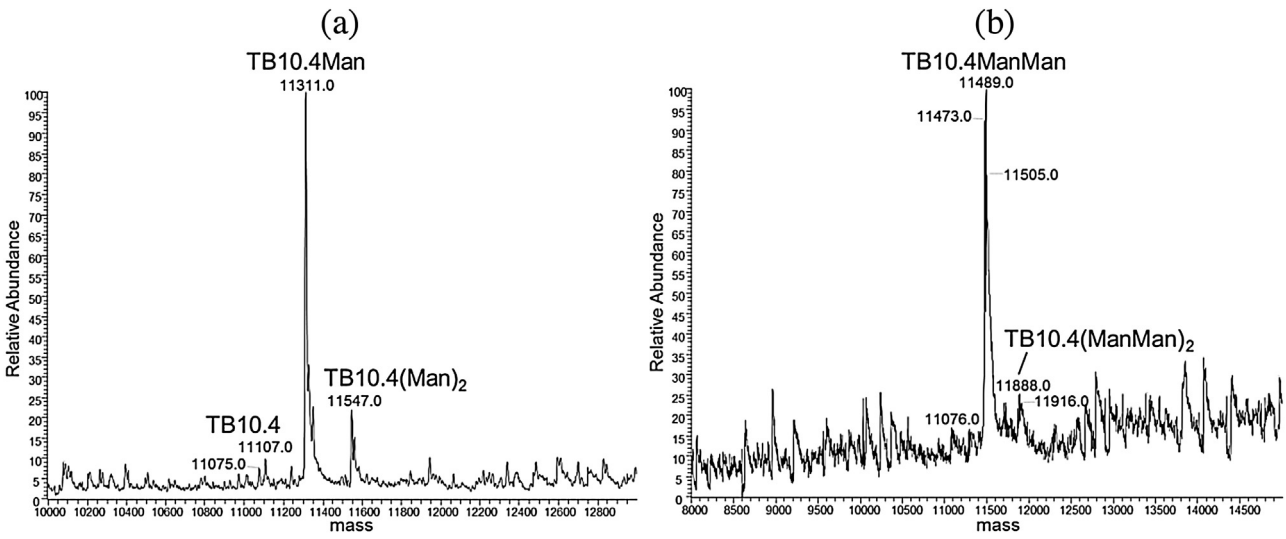


Fig. 7. Mass spectrum of the TB10.4 + Man-IME (a) and TB10.4 + ManMan-IME (b) samples.

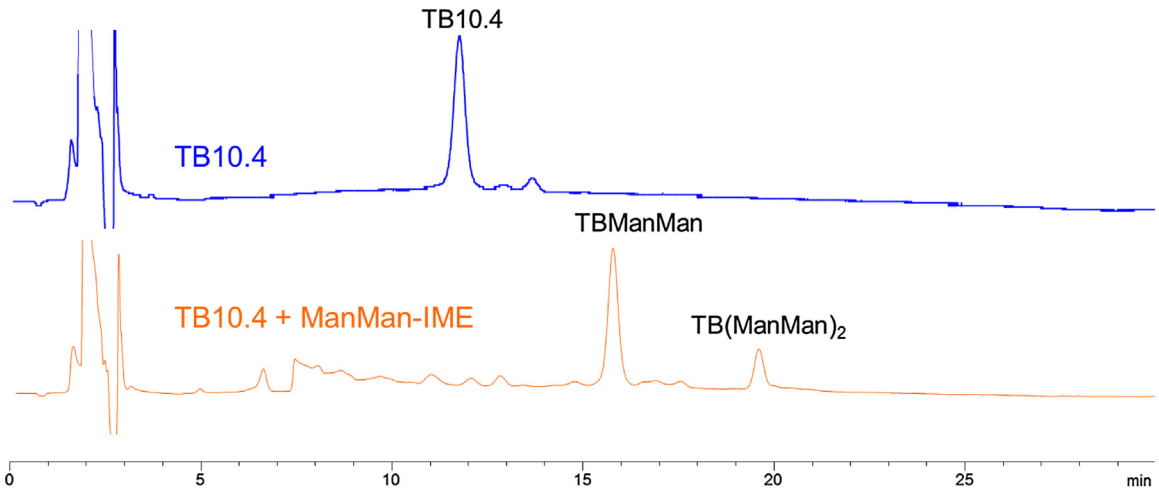


Fig. 8. Chromatographic UV traces of TB10.4 conjugated with ManMan-IME using the 2 h-procedure compared to the intact protein.

HILIC-trace. The identity of the peaks was also inferred considering that the polar groups, such as the saccharide units, promote the retention in HILIC.

3.4.2. Ag85B

A simulation of the glycosylation reaction in absence of activated glycans was also performed on Ag85B. The protein stability was monitored at fixed times. Differently from TB10.4, Ag85B was stable during the 24 h required for the conjugation reaction, with an average area of 4314 ± 105 and a relative standard deviation of 2.43% calculated on the chromatograms reported in Fig. 9.

Once assessed protein stability in the conjugation conditions, Ag85B was glycosylated using Man-IME. Ag85B has 8 lysine residues in its sequence and thus its glycosylation results in the formation of glycoforms differing in the number of incorporated saccharides. The analysis of the reaction mixture at 24 h is reported in Fig. 10. No degradation products were observed and the HILIC method completely resolved the non-glycosylated protein from the glycoconjugates allowing to define a glycosylation yield of 100% (green trace). Unfortunately, the method selectivity did not allow to discriminate between the different glycoforms. However, the elongation of the saccharidic chain used in the conjugation process (ManMan-IME) results in a successful separation of the glycoforms (Fig. 10) due to the greater contribution of the hydrophilic moiety, in agreement with the retention mechanism of amide HILIC stationary phase. Also for these experiments the DI-ESI–MS spectra were recorded (Fig. 11), the presence of Ag85B-Man (green trace in Fig. 10) and Ag85B-ManMan glycoforms (orange trace in Fig. 10) was detected in the respective samples, thus confirming glycoconjugate composition and identity.

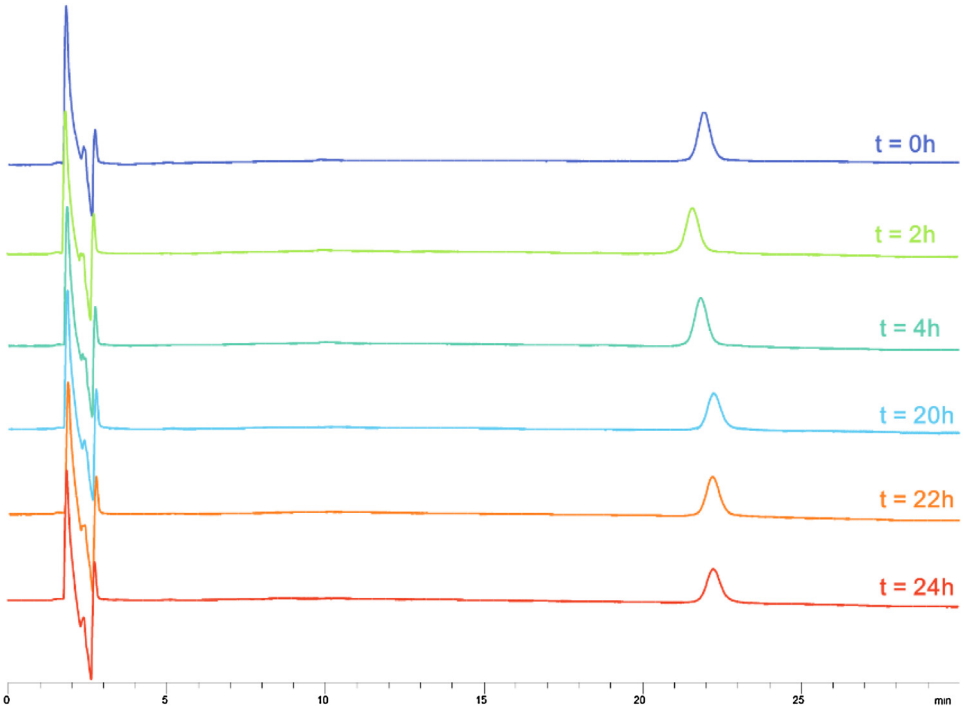


Fig. 9. Chromatographic traces of Ag85B peaks obtained by analyses in HPLC of appropriate dilutions of the 5.5 mg/mL stock solution in 100 mM sodium tetraborate buffer, pH 9.5, 1 mM benzamidine hydrochloride. The stability of the protein was investigated at 0, 2, 4, 20, 22 and 24 h.

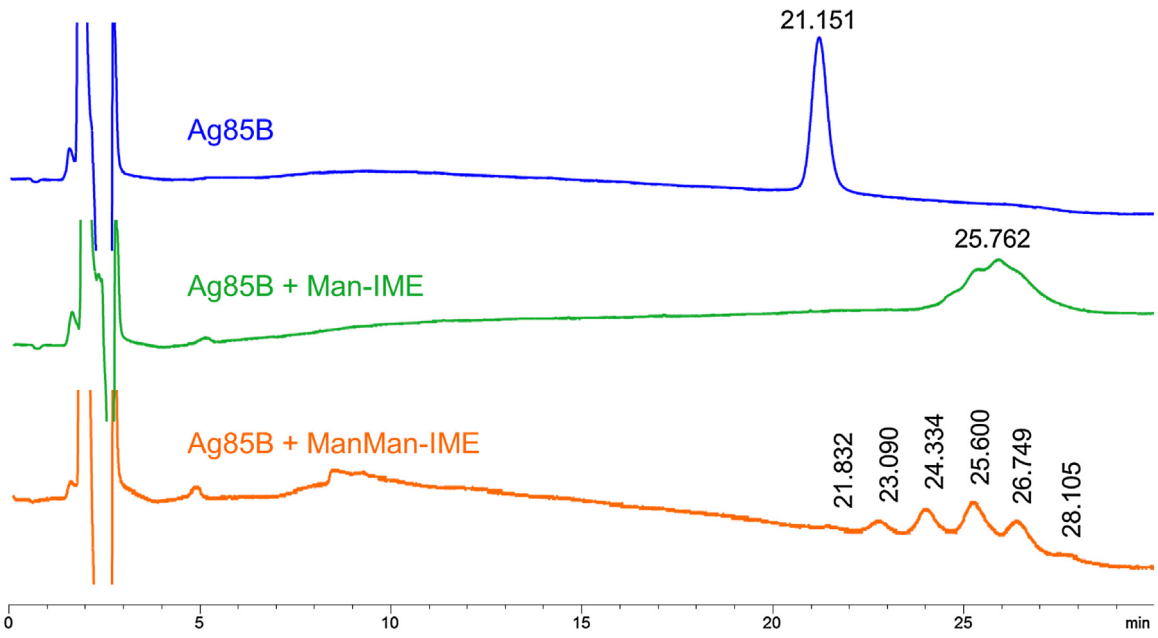


Fig. 10. Chromatographic traces of Ag85B (50 μg/mL) glycosylated with Man-IME and ManMan-IME compared to the intact protein.

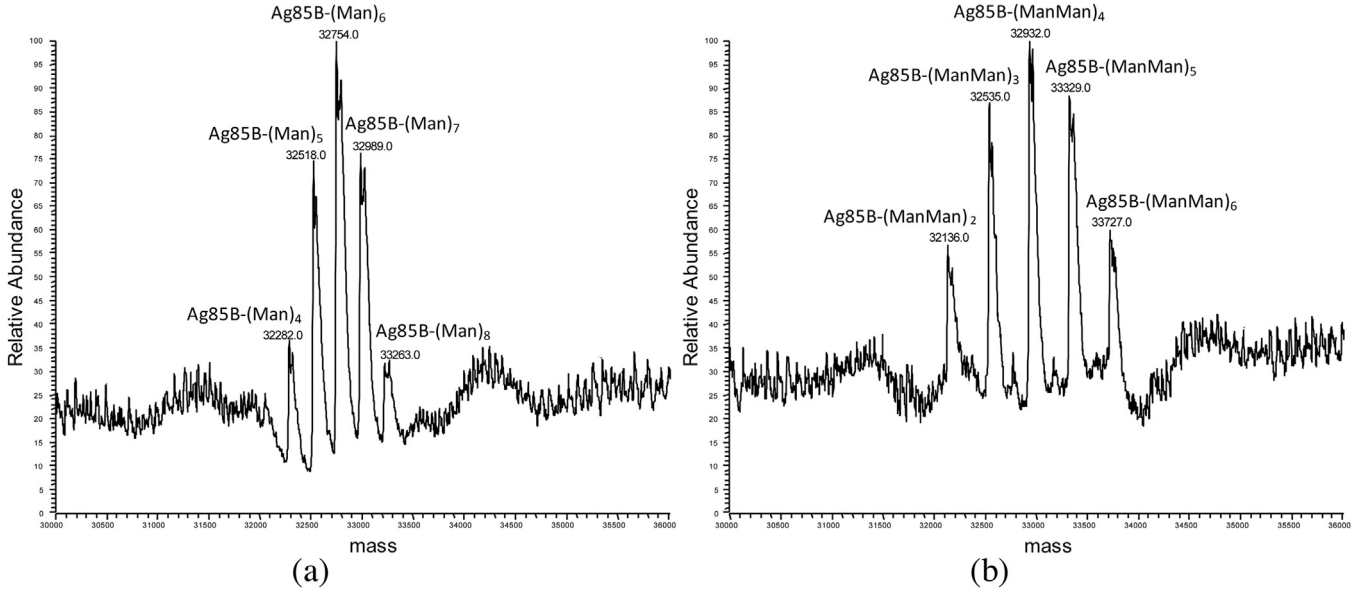


Fig. 11. Deconvoluted mass spectrum of the Ag85B + Man-IME (a) and Ag85B + ManMan-IME (b) samples.

4. Conclusions

In this work, a rapid HILIC-UV method was developed for the analysis of intact proteins and *neo*-glycoproteins with antigenic activity against tuberculosis. The method allowed to rapidly estimate the protein stability under critical conditions and during the coupling reaction, without the need of any purification step. In particular, the obtained results led to the optimization of the glycosylation protocol thus preventing protein degradation. The method represents a valid, informative and rapid tool able to monitor protein content during storage conditions and to support the *neo*-glycoprotein synthesis by monitoring the stability of the native protein and its glycoforms. The present study represents a further application of the potential of HILIC chromatography for the separation of intact recombinant proteins, with the possibility to reach an adequate selectivity towards glycosylated derivatives. These promising results could open the possibility to combine the HILIC method to MS for direct structural characterization of each glycoform.

Acknowledgments

This work was funded by Regione Lombardia, Italy (VATUB project, Project Framework agreement Lombardy Region Universities-DGR 9139) and by Fondazione Banca del Monte di Lombardia (Italy) FBML. LuPi and LP were supported from Fondo di Ateneo per la Ricerca.

References

[1] ICH Topic Q 5C, Quality of Biotechnological Products: Stability Testing of Biotechnological/Biological Products, 1995. http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html.

[2] M.L. Brader, T. Estey, S. Bai, R.W. Alston, K.K. Lucas, S. Lantz, P. Landsman, K.M. Maloney, Examination of thermal unfolding and aggregation profiles of a series of developable therapeutic monoclonal antibodies, Mol. Pharm. 12 (4) (2015) 1005–1017.

[3] B. Bobaly, E. Sipkò, J. Fekete, Challenges in liquid chromatographic characterizations of proteins, J. Chromatogr. B 1032 (2016) 3–22.

[4] M.K. Parr, O. Montacir, H. Montacir, Physicochemical characterization of biopharmaceuticals, J. Pharm. Biomed. Anal. 130 (2016) 366–389.

[5] R. Adamo, A. Nilo, B. Castagner, O. Boutureira, F. Berti, G.J.L. Bernardes, Synthetically defined glycoprotein vaccines: current status and future directions, Chem. Sci. 4 (2013) 2995–3008.

[6] C. Temporini, T. Bavaro, S. Tengattini, I. Serra, G. Marrubini, E. Calleri, F. Fasanella, L. Piubelli, F. Marinelli, L. Pollegioni, G. Speranza, G. Massolini, M. Terreni, Liquid chromatography-mass spectrometry structural characterization of *neo* glycoproteins aiding the rational design and synthesis of a novel glycovaccine for protection against tuberculosis, J. Chromatogr. A 1367 (2014) 57–67.

[7] World Health Organization, Global tuberculosis report 2015. Available online: http://www.who.int/tb/publications/global_report/en.

[8] L. Piubelli, M. Campa, C. Temporini, E. Binda, F. Mangione, M. Amicosante, M. Terreni, F. Marinelli, L. Pollegioni, Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins, Microb. Cell Fact. 12 (2013) 115.

[9] T. Bavaro, M. Filice, C. Temporini, S. Tengattini, I. Serra, C.F. Morelli, G. Massolini, M. Terreni, Chemoenzymatic synthesis of neoglycoproteins driven by the assessment of protein surface reactivity, RSC Adv. 4 (2014) 56455–56465.

[10] R.L.V. Skjøt, T. Oettinger, I. Rosenkrands, P. Ravn, I. Brock, S. Jacobsen, P. Andersen, Comparative evaluation of low-molecular-mass proteins from mycobacterium tuberculosis identifies members of the ESAT-6 family as immunodominant T-cell antigens, Infect. Immun. 68 (2000) 214–220.

[11] J. Dietrich, C. Aagaard, R. Leah, A.W. Olsen, A. Stryhn, T.M. Doherty, P. Andersen, Exchanging ESAT6 with TB10.4 in an Ag85B fusion molecule-based tuberculosis subunit vaccine: efficient protection and ESAT6-based sensitive monitoring of vaccine efficacy, J. Immunol. 174 (2005) 6332–6339.

[12] S. Tengattini, E. Domínguez-Vega, C. Temporini, M. Terreni, G.W. Somsen, Monitoring antigenic protein integrity during glycoconjugate vaccine synthesis using capillary electrophoresis-mass spectrometry, Anal. Bioanal. Chem. 408 (2016) 6123–6132.

[13] A. Pedrali, S. Tengattini, G. Marrubini, T. Bavaro, P. Hemström, G. Massolini, M. Terreni, C. Temporini, Characterization of intact *neo*-glycoproteins by hydrophilic interaction liquid chromatography, Molecules 19 (2014) 9070–9088.

[14] Z. Zhang, Z. Wu, M.J. Wirth, Polyacrylamide brush layer for hydrophilic interaction liquid chromatography of intact glycoproteins, J. Chromatogr. A 1301 (2013) 156–161.

[15] T. Tetaz, S. Detzner, A. Friedlein, B. Molitor, J.-L. Mary, Hydrophilic interaction chromatography of intact, soluble proteins, J. Chromatogr. A 1218 (2011) 5892–5896.

[16] A. Periat, S. Fekete, A. Cusumano, J.-L. Veuthey, A. Beck, M. Lauber, D. Guillarme, Potential of hydrophilic interaction chromatography for the analytical characterization of protein biopharmaceuticals, J. Chromatogr. A 1448 (2016) 81–92.

[17] J.C. Heaton, D.V. McCalley, Some factors that can lead to poor peak shape in hydrophilic interaction chromatography, and possibilities for their remediation, J. Chromatogr. A 1427 (2016) 37–44.

[18] J. Ruta, S. Rudaz, D.V. McCalley, J.L. Veuthey, D. Guillarme, A systematic investigation of the effect of sample diluent on peak shape in hydrophilic interaction liquid chromatography, J. Chromatogr. A 1217 (52) (2010) 8230–8240.

[19] R. Chaudhuri, Y. Cheng, C.R. Middaugh, D.B. Volkin, High-throughput biophysical analysis of protein therapeutics to examine interrelationships between aggregate formation and conformational stability, AAPS J. 16 (1) (2014) 48–64.

[20] A. Zhang, S.K. Singh, M.R. Shirts, S. Kumar, E.J. Fernandez, Distinct aggregation mechanisms of monoclonal antibody under thermal and freeze-thaw stresses revealed by hydrogen exchange, Pharm. Res. 29 (1) (2012) 236–250.

[21] R.J. Solá, K. Griebenow, Effects of glycosylation on the stability of protein pharmaceuticals, J. Pharm. Sci. 98 (4) (2009) 1223–11245.

[22] B.S. Gupta, M. Taha, M.J. Lee, Buffers more than buffering agent: introducing a new class of stabilizers for the protein BSA, Phys. Chem. Chem. Phys. 17 (2) (2015) 1114–1133.

[23] V. Vagenende, M.G. Yap, B.L. Trout, Mechanisms of protein stabilization and prevention of protein aggregation by glycerol, Biochemistry 48 (46) (2009) 11084–11096.

2. *Hydrophilic interaction liquid chromatography-mass spectrometry as a new tool for the characterization of intact semi-synthetic glycoproteins*

Sara Tengattini, Elena Domínguez-Vega, Caterina Temporini, Teodora Bavaro, Francesca Rinaldi, Luciano Piubelli, Loredano Pollegioni, Gabriella Massolini, Govert Willem Somsen

*Analytica Chimica Acta 2017, 981, 94-105**

Analytica Chimica Acta 981 (2017) 94–105

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Hydrophilic interaction liquid chromatography-mass spectrometry as a new tool for the characterization of intact semi-synthetic glycoproteins

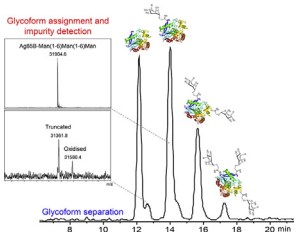
Sara Tengattini [a, b, 1], Elena Domínguez-Vega [b, 1], Caterina Temporini [a, *], Teodora Bavaro [a], Francesca Rinaldi [a], Luciano Piubelli [c, d], Loredano Pollegioni [c, d], Gabriella Massolini [a], Govert W. Somsen [b]

[a] Department of Drug Sciences, University of Pavia, Via Taramelli 12, I-27100, Pavia, Italy
[b] Division of BioAnalytical Chemistry, Department of Chemistry and Pharmaceutical Sciences, Vrije Universiteit Amsterdam, de Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands
[c] Department of Biotechnology and Life Sciences, University of Insubria, Via J.H. Dunant 3, I-21100, Varese, Italy
[d] The Protein Factory Interuniversity Centre, Politecnico of Milano, University of Insubria, Via Mancinelli 7, I-20131, Milano, Italy

HIGHLIGHTS

- New potential glycoconjugate vaccines are characterized by HILIC-UV and HILIC-MS.
- Amide HILIC stationary phases can provide efficient separation of intact protein glycoforms.
- HILIC-MS allows reliable glycoform assignment of intact semi-synthetic glycoproteins.
- Intact glycoform fragmentation in HILIC-MS/MS detects N-terminal glycosylation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 6 February 2017
Received in revised form
8 May 2017
Accepted 14 May 2017
Available online 2 June 2017

Keywords:
Hydrophilic interaction liquid chromatography
Neo-glycoproteins
Intact protein analysis
Mass spectrometry

ABSTRACT

Improved methods for detailed characterization of complex glycoproteins are required in the growing sector of biopharmaceuticals. Hydrophilic interaction liquid chromatography (HILIC) coupled to high resolution (HR) time-of-flight mass spectrometric (TOF-MS) detection was examined for the characterization of intact *neo*-glycoproteins prepared by chemical conjugation of synthetic saccharides to the lysine residues of selected recombinant proteins. The separation performances of three different amide HILIC columns (TSKgel Amide-80, XBridge BEH and AdvanceBio Glycan Mapping) were tested. Water-acetonitrile gradients and volatile eluent additives have been explored. Addition of 0.05% (v/v) trifluoroacetic acid to the mobile phase appeared to be essential for achieving optimum resolution of intact glycoforms and minimal ion suppression effects. Gradient elution conditions were optimized for each protein on every column. HILIC stationary phases were evaluated for the analysis of highly heterogeneous semi-synthetic derivatives of the same protein (ribonuclease A), and in the enhanced characterization of TB10.4 and Ag85B glycoconjugates, selected antigens from *Mycobacterium tuberculosis* (MTB). HILIC-MS results indicated that the HILIC selectivity is predominantly governed by size of the conjugated glycans and number of glycans attached, providing efficient glycoform separation. Moreover, HILIC separation prior to HRMS detection allowed assignment of several product impurities. Additional top-

* Corresponding author.
E-mail address: caterina.temporini@unipv.it (C. Temporini).
[1] Contributed equally.

http://dx.doi.org/10.1016/j.aca.2017.05.020

down MS/MS experiments confirmed conjugation at the N-terminus of TB10.4 next to its lysine residue. Overall, the obtained results demonstrate that amide-stationary-phase based HILIC coupled to MS is highly useful for the characterization of intact *neo*-glycoproteins allowing assessment of the number, identity and relative abundance of glycoforms present in the semi-synthetic products.

1. Introduction

Biopharmaceutical products and drugs of biological origin represent an important class of therapeutics. In particular, glycoproteins are among the most commercially successful biological drugs due to their numerous applications and indications ranging from cancer to inflammation to infectious diseases (glycovaccines). It is well known that glycans attached to a protein can influence the therapeutic efficacy of protein drug and the protein physicochemical and pharmacological properties. Thus, there is a need for the production of well-defined glycoproteins and elegant synthetic strategies have been explored for the preparation of homogeneous glycoproteins with particular emphasis on glycovaccines.

An interesting approach, often used for the production of glycoconjugate vaccines, entails the extraction or the synthesis of saccharides (antigens) and their covalent linkage to a carrier protein produced by recombinant DNA technology. The conjugation is commonly achieved by chemical activation of the glycan portion allowing targeting of nucleophilic functional groups of specific amino acid residues (lysines, aspartic/glutamic acids or cysteines) in the protein [1,2].

The characterization of *neo*-glycoprotein products, obtained as described above, is quite challenging. In fact, in addition to the inevitable heterogeneity of recombinant proteins due to the biotechnological manufacturing, the heterogeneity related to the carbohydrate component (structure, number, and position) has to be considered. Similarly to native glycoproteins, also for semi-synthetic glycosylated proteins, the inherent structural complexity (i.e. sequence mutation, oxidation, deamidation, N- and C-terminal alteration, glycoform variety due to multiple conjugation sites) necessitates the development of analytical strategies for their characterization [3].

Liquid chromatography (LC) coupled to electrospray ionization mass spectrometry (ESI-MS) is widely used for glycoprotein characterization by bottom-up approaches, in which the glycoproteins are first digested into peptides, providing structural information and allowing localization of glycosylation sites [4,5]. Nevertheless, quality control of the intact glycoproteins is of great added value in biopharmaceutical analysis, as it provides specific information on the presence of proteoforms as well as on the exact combinations of multiple modifications, which cannot be revealed by merely bottom-up approaches. Moreover, intact protein analysis is relatively fast, requires minimum sample treatment and prevents undesired modifications induced by enzymatic treatments [6,7].

Hydrophilic interaction liquid chromatography (HILIC) has demonstrated to be a highly useful technique for analysis of amino acids, released glycans and glycopeptides [8]. In contrast to reversed-phase (RP) LC, HILIC has the ability to retain and resolve (highly) polar compounds, based on a complex retention mechanism, involving hydrophilic partitioning and polar interactions. As a consequence, glycans and glycopeptides show stronger HILIC retention with an increasing number and size of sugar units [9,10].

Only few studies have been published on the use of HILIC methods for the characterization of intact proteins [8,11,12] and for the determination of their glycoform composition [11–14], indicating HILIC to be complementary and orthogonal to RPLC. So far, the application of HILIC for the separation of native glycoforms of intact proteins has been limited to the model glycoprotein ribonuclease B (RNase B), containing only one glycosylation site, and monoclonal antibodies [11,12]. Recently, we have reported a preliminary work where HILIC-UV was used for the monitoring of the glycosylation reaction of a model protein ribonuclease A (RNase A), with different glycans [15]. The developed analytical method was not compatible with MS detection, however, it was suitable for the analysis of intact *neo*-glycoproteins, allowing discrimination of non-glycosylated protein from glycoproteins and separation of different glycoforms, including isomers, as demonstrated by off-line MS analysis [15].

On-line MS detection after HILIC separation of glycoforms, would offer many advantages, such as the assignment of glycoform structures and other potential protein modifications and degradations. Moreover, following a top-down approach, HILIC-MS/MS could provide information on N-terminal sequence, including glycosylation occurrence at the α-amino group. Nevertheless, on-line HILIC-MS of intact glycoproteins has been reported only for RNase B and for IdeS-digested and reduced monoclonal antibodies [11,12,14].

In order to expand the applicability of HILIC with UV and HRMS detection to the characterization of intact *neo*-glycoproteins, we studied the chromatographic performances of three different amide HILIC columns (TSKgel Amide-80, XBridge BEH and AdvanceBio Glycan Mapping). *Neo*-glycoproteins, obtained by chemical conjugation of synthetic mono-, di- and tri-mannose to lysine residues of RNase A and of the two recombinant tuberculosis antigens TB10.4 and Ag85B [16], were considered as representative real-life proteins in this work. TB10.4 and Ag85B have been evaluated by our research group for the development of new effective vaccines against tuberculosis [3,17]. Their semi-synthetic glycoconjugates comprise a large number of glycoforms with multiple conjugation sites, each carrying short carbohydrate chains. This intrinsic heterogeneity combined with the relatively small hydrophilic contribution of every individual carbohydrate chain, makes chromatographic resolution and characterization of these *neo*-glycoproteins an interesting case for the application of the three amide HILIC columns. MS-compatible mobile phase composition, gradients and column temperature were studied in order to achieve good separation of the semi-synthethic glycoprotein products. Moreover, the HILIC-MS results obtained aid in understanding structure-retention relationships of glycoproteins. The potential of the developed HILIC-MS(/MS) methods for the detection of glycoconjugate product impurities and for the confirmation of the glycan position in conjugated TB10.4 was also examined.

2. Experimental

2.1. Reagents

RNase A and RNase B from bovine pancreas were purchased from Sigma-Aldrich (St. Louis, MO, USA) and were used without further purification. Formic acid (FA), ammonium hydroxide (30%),

sodium tetraborate, trifluoroacetic acid (TFA), phosphate-buffered saline (PBS), benzamidine chloride were from Sigma-Aldrich (St. Louis, MO, USA). Water was obtained from a Direct-QTM Millipore system Millipore (Millipore, Billerica, MA, USA). Acetonitrile (ACN) MS grade was purchased from Sigma-Aldrich (St. Louis, MO, USA).

In this study three IME-thioglycosides were used for protein glycosylation: Mannose-IME (Man-IME), Mannose(1–6)Mannose-IME (Man(1–6)Man-IME) and Mannose(1–6)Mannose(1–6)Mannose-IME (Man(1–6)Man(1–6)Man-IME). IME-thioglycosides were *ad hoc* synthetized according to the procedure previously reported [18] (see Supplementary Material). The TB10.4 and Ag85B immunogenic proteins were obtained as recombinant forms in *Escherichia coli* as reported in Piubelli et al. [16] and finally collected in 20 mM 3-(N-morpholino)propanesulfonic acid, 0.4 M NaCl, pH 7.0 at different concentrations.

2.2. Protein glycosylation

2.2.1. RNase A glycosylation procedure

According to the glycosylation protocol previously reported [15], the reaction was carried out in sodium tetraborate buffer, 100 mM, pH 9.5. RNase A was dissolved in the buffer to reach a final concentration of 1.7 mg mL^{-1} and then the solution was mixed with IME-glycoside to a glycoside/protein molar ratio of 200/1. The reaction mixture was vortexed for 1 min and incubated for 24 h at 25 °C under continuous stirring. Based on thioglycosides used, the prepared glycoconjugates were named RNase-Man and RNase-Man(1–6)Man.

2.2.2. TB10.4 and Ag85B glycosylation procedure

The conditions used for the antigenic protein glycosylation required a protein concentration of 5.5 mg mL^{-1} and a glycoside/protein molar ratio of 200/1 in the same reaction buffer as above. Benzamidine chloride (1 mM) was added to prevent protein digestion by traces of enterokinase derived from protein production. Five *neo*-glycoproteins were synthesized, namely TB10.4-Man, TB10.4-Man(1–6)Man, Ag85B-Man, Ag85B-Man(1–6)Man and Ag85B-Man(1–6)Man(1–6)Man.

2.3. Sample preparation

Standard stock solutions of RNase B were prepared in pure water at a concentration of 2 mg mL^{-1} and then diluted to obtain working solutions at 0.5 mg mL^{-1} in ACN-water (50:50, v/v). *Neo*-glycoproteins, prepared from RNase A and antigenic proteins, were purified in order to remove reagents and salts. The reaction mixture was submitted to seven 20 min steps of ultrafiltration at 13000 g and 4 °C using centrifuge 5804-R (Eppendorf s.r.l., Milan, Italy) and Millipore's Amicon® Ultra filters with a nominal molecular weight limit of 3 or 10 kDa and load capacity of 500 μL. Proteins and *neo*-glycoproteins were finally collected and stored in PBS at concentration of 2 mg mL^{-1}. Prior to HILIC-UV or HILIC-MS analysis, samples were diluted to obtaine a final concentration of 0.5 mg mL^{-1} in ACN-water-PBS (50:25:25, v/v/v).

2.4. Chromatographic equipment and conditions

Chromatographic separations of intact (*neo*-)glycoproteins were performed on an Agilent HPLC series 1200 system, equipped with a mobile-phase online degasser, quaternary pump, autosampler, column thermostated compartment, and diode array detector. For data acquisition and analysis, ChemStation software version Rev. B.04.01 was used. The HILIC columns studied were (1) TSKgel Amide-80 (150 × 2.0 mm, 3 μm, 80 Å) from Tosoh Bioscience (Montgomeryville, PA, USA), (2) XBridge BEH Amide (150 × 3.0 mm, 2.5 μm, 130 Å) from Waters (Milford, MA, USA), and (3) AdvanceBio Glycan Mapping (150 × 2.1 mm, 2.7 μm) from Agilent Technologies (Palo Alto, CA, USA). The injection volume was fixed at 2 μL, the column temperature at 50 °C and the flow rate was set at 0.3 mL min^{-1} for the TSKgel Amide-80 column, and at 0.5 mL min^{-1} for the other two HILIC columns.

The final mobile phases were composed of ACN (A) and water (B) both containing 0.05% (v/v) TFA. Gradient elution conditions were optimized for each protein on every column. For RNase B and RNase glycoconjugates: with TSKgel Amide-80 from 32% to 42% B in 20 min followed by isocratic elution at 42% B for 10 min; with XBridge BEH Amide and AdvanceBio Glycan Mapping from 28% to 38% B in 20 min followed by isocratic elution at 38% B for 10 min. For TB10.4 glycoconjugates: with TSKgel Amide-80 from 21% to 31% B in 20 min followed by isocratic elution at 31% B for 10 min; with XBridge BEH Amide and AdvanceBio Glycan Mapping from 17% to 27% B in 20 min followed by isocratic elution at 27% B for 10 min. For Ag85B glycoconjugates: with TSKgel Amide-80 from 26% to 36% B in 20 min followed by isocratic elution at 36% B for 10 min; with XBridge BEH Amide and AdvanceBio Glycan Mapping from 22% to 32% B in 20 min followed by isocratic elution at 32% B for 10 min. UV absorbance was monitored at 214 nm.

2.5. MS detection

MS detection was performed using a maXis HD ultra-high resolution quadrupole time-of-flight (QTOF) mass spectrometer (Bruker Daltonics, Bremen, Germany) with an ESI source. The mass spectrometer was operated in positive-ion mode with an electrospray voltage of 4.5 kV. The nebulizer and drying gas conditions were 1.0 bar and 8.0 $Lmin^{-1}$ nitrogen at 200 °C, respectively. Quadrupole ion and collision cell energy were 3.0 and 15.0 eV, respectively. Transfer and pre plus storage times were set at 190.0 and 20.0 μs, respectively. In source collision induced dissociation (ISCID) was set at 120 eV in order to dissociate protein-TFA adducts formed during ESI. The monitored mass range was 250–4000 m/z. Extracted ion chromatograms were obtained with an extraction window of ±0.5 m/z and using the smooth option of the software (Gaussian at 1 point). For fragmentation experiments, the most abundant ion of each protein was fragmented by collision induced dissociation (CID) using multiple reaction monitoring mode. The collision energy was separately optimized for each species and the final applied values were in the range 90–110 eV.

3. Results and discussion

In this work, HILIC-UV-MS was evaluated for the characterization of *neo*-glycoproteins. Three different commercial HILIC columns were investigated for their capacity to reduce heterogeneity complexity by resolving intact glycoforms as well as specific proteoforms to facilitate their assignment. Initial chromatographic optimization was carried out by HILIC-UV using MS-compatible conditions and RNase B as model glycoprotein. Then, HILIC-MS methods were applied to the analysis of semi-synthetic glycoproteins prepared starting from RNase A and from two recombinant tuberculosis antigen proteins, TB10.4 and Ag85B.

3.1. Development of a HILIC method for intact glycoform resolution

Previous studies of our research group have shown that a carbamoyl-silica-based stationary phase (TSKgel Amide-80) has good potential for the glycoform profiling of intact glycoproteins and *neo*-glycoproteins [15]. Encouraged by these interesting results, in the present study we also considered two other amide HILIC columns (XBridge BEH Amide and AdvanceBio Glycan

Mapping) for glycoprotein analysis. The performance of the three HILIC columns was studied under MS-compatible conditions using the glycoprotein RNase B (MW 15 kDa) as test compound. RNase B is N-glycosylated at Asn 34 giving rise to five glycoforms differing in the number of mannose residues (5–9). Mobile phases consisting of ACN and water with different concentrations of FA, ammonium formate and TFA were tested. The injection volume was fixed at 2 μL in order to avoid the deterioration of peak shape due to the relatively high percentage of water (50% v/v) in the diluents necessary for full protein solubilization.

The eluent composition and gradient conditions were adjusted to obtain an appropriate elution window for RNase B. However, with the addition of 0.1% (v/v) FA or 10 mM ammonium formate (adjusted at pH 3.0) no separation among the five glycoforms was observed with each of the tested HILIC columns. Improved glycoform separation was obtained by increasing the percentage FA to 0.5% (v/v) or ammonium formate concentration to 100 mM (when allowed by column specifications), and by rising the column temperature from 25 to 50 °C. However, the obtained glycoform resolution was still unsatisfactory. Temperatures above 50 °C were tested for the XBridge BEH column (as permitted according to column specifications), but not significant increase in separation quality was observed.

Using a column temperature of 50 °C, the effect of TFA as alternative eluent additive [11,14] on the three HILIC columns was studied. The addition of 0.025% (v/v) of TFA to the mobile phase resulted in an increased resolution of the five glycoforms of RNase B. The concentration of TFA was next evaluated in the range 0.025%–0.1%. Higher percentages were not considered, as TFA is known to suppress ESI [19–21]. For all three HILIC columns, the overall peak profile of the RNase B glycoforms was hardly affected by increasing the concentration of TFA in the mobile phase. For instance, for the XBridge BEH column using a gradient of 30%–40% water in 10 min followed by isocratic elution at 40% water for 10 min, the average distance between the apices of the consecutive glycoform peaks was 0.43 ± 0.06 min, 0.49 ± 0.06 min and 0.49 ± 0.05 min at 0.025%, 0.05% and 0.1% TFA, respectively. However, more narrow peaks were observed with rising TFA concentration, leading to an overall better glycoform resolution. Indeed, with the same column, the average resolution for consecutive peaks was 0.95, 1.19 and 1.25 using, respectively, 0.025%, 0.05% and 0.1% TFA (Fig. S1 in Supplementary data). From these experiments it is clear that TFA must be used to achieve satisfactory separation of the glycoforms. Thus, for further experiments we selected a TFA concentration of 0.05% to achieve good glycoform resolution while preventing excessive ion suppression in MS detection.

Fig. 1 shows the HILIC-UV chromatograms for RNase B obtained under optimized conditions and Table 1 lists the corresponding Rs values for the separated glycoforms. A significantly lower resolution of the RNase B glycoforms was observed for the TSKgel Amide-80 column (average Rs 1.0 ± 0.1). Considering the similar peak spacing provided by the three stationary phases, the reduced separation power of this column is a result of the protein peak broadening, most probably due to the small pore size (80 Å) of the stationary phase. The XBridge BEH Amide and AdvanceBio Glycan Mapping materials provided baseline glycoform resolution with average Rs values of 1.5 ± 0.2. The AdvanceBio Glycan Mapping column exhibited the shortest retention times and a somewhat better resolution for the low abundant glycoforms.

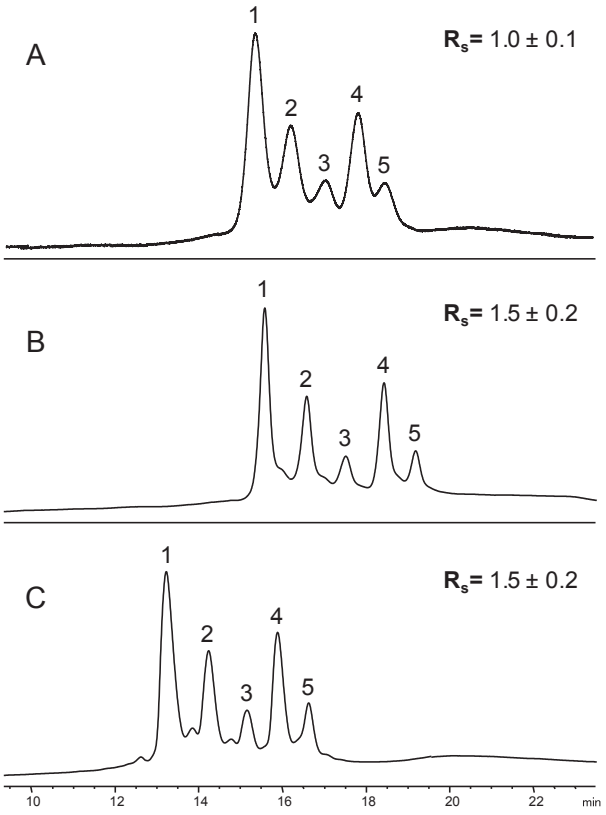


Fig. 1. HILIC-UV chromatograms obtained for RNase B (0.5 mg mL^{-1} in ACN-water (50:50, v/v)). Column, (A) TSKgel Amide-80, (B) Waters XBridge BEH Amide, and (C) Agilent AdvanceBio Glycan Mapping. Mobile phase comprised of ACN (solvent A) and water (solvent B) both containing 0.05% (v/v) TFA; injection volume, 2 μL; column temperature, 50 °C. Flow rate, (A) 0.3 mL min^{-1} and (B,C) 0.5 ml/min. Gradient, (A) from 32% to 42% B in 20 min followed by isocratic elution at 42% B for 10 min, and (B,C) from 28% to 38% B in 20 min followed by isocratic elution at 38% B for 10 min.

3.2. HILIC-UV-MS of neo-glycoproteins

3.2.1. RNase glycoconjugates

In order to investigate the suitability of the developed HILIC methods for the characterization of semi-synthetic glycoproteins, glycoconjugates of RNase A were prepared and analysed. These conjugates were obtained by coupling a mono- or disaccharide activated with IME to the ε-amino group of lysine residues of RNase A (see preparation procedure in Section 2.2.1). RNase A (124 amino acids, MW 13,681 Da) contains ten lysines and, consequently, its glycosylation leads to the formation of a mixture of semi-synthetic glycoforms with different saccharide loading and positioning. These heterogeneous glycoconjugate products allow the assessment of the ability of the HILIC columns to discriminate among the different glycoforms. The *neo*-glycoproteins were dissolved in ACN-water (50:50, v/v) and analysed by HILIC-UV-MS using the three considered HILIC columns. MS detection was achieved using an ESI sprayer and a high resolution time-of-flight (HR TOF) mass spectrometer operated in the positive-ion mode.

Fig. 2 reports the UV chromatograms obtained for RNase A conjugated with Man-IME (upper panel) and Man(1–6)Man-IME (lower panel), showing clusters of partially resolved glycoform peaks. The median retention times and the width of the clusters are significantly larger for the disaccharide conjugate products, indicating that the HILIC retention of the intact glycoproteins increases with increasing glycan size. The HILIC-UV chromatograms obtained with the TSKgel Amide-80 column show less resolved glycoform profiles than the chromatograms acquired with the other two HILIC columns. Fig. 3 shows the extracted-ion chromatograms (EICs) obtained for the conjugated RNase A products using the AdvanceBio Glycan Mapping column and depicting the intensity for the most abundant ion (charge state 6+) of each glycoform. The

Table 1
Retention times (RT), and resolutions (Rs) obtained for successively eluting glycoforms during HILIC-UV of RNase B on three columns. SD, standard deviation.

HILIC column	mean RT (min)	Rs (2,1)	Rs (3,2)	Rs (4,3)	Rs (5,4)	Average Rs ± SD
TSK-gel Amide 80	16.84	1.1	1.0	1.0	0.9	1.0 ± 0.1
XBridge BEH Amide	17.45	1.9	1.4	1.4	1.3	1.5 ± 0.2
AdvanceBio Glycan Mapping	15.02	1.7	1.6	1.3	1.4	1.5 ± 0.2

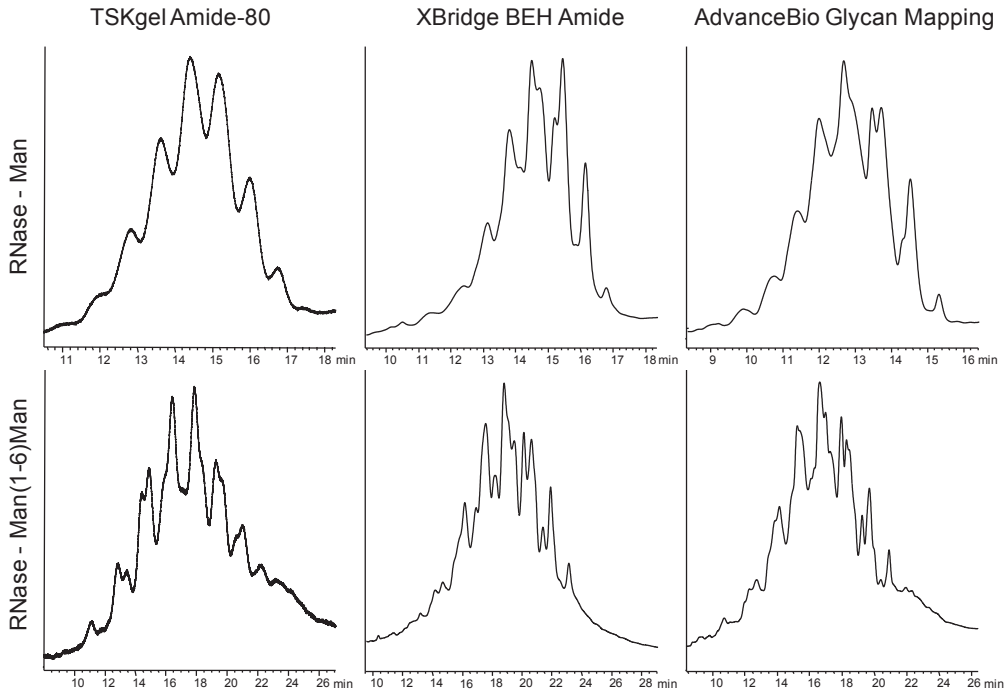


Fig. 2. HILIC-UV chromatograms obtained for RNase A (0.5 mg mL^{-1}) conjugated with Man-IME (upper panel) and Man(1–6)Man-IME (lower panel) obtained using the (left) TSKgel Amide-80, (middle) XBridge BEH Amide, and (right) AdvanceBio Glycan Mapping columns. Further conditions, see Fig. 1.

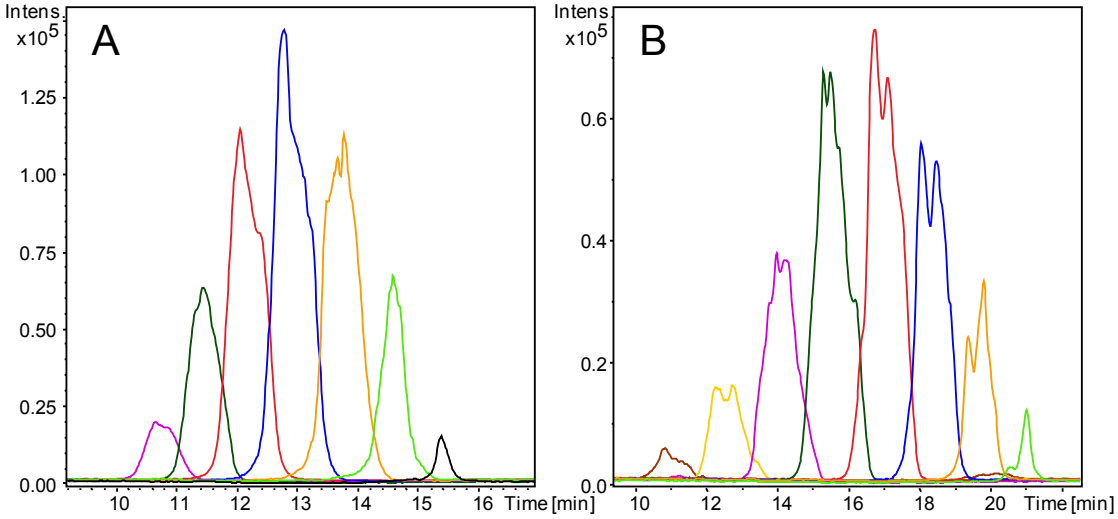


Fig. 3. Extracted-ion chromatograms obtained during HILIC-UV-MS of RNase A conjugated with (A) Man and (B) Man(1–6)Man using the AdvanceBio Glycan Mapping column. The ion intensity of the 6 + charge state of each glycoform is depicted. Number of conjugated saccharides, 2 (dark red), 3 (yellow), 4 (pink), 5 (dark green), 6 (red), 7 (blue), 8 (orange), 9 (light green) and 10 (black). Further conditions, see Fig. 1 and Experimental Section. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

acquired mass data revealed that seven glycoforms carrying 4–10 saccharide units, were detected for the Man conjugates, whereas eight glycoforms carrying 2–9 saccharide units were detected for the Man(1–6)Man conjugates. The results nicely show that the HILIC retention of the intact protein glycoforms increases proportionally with the number of conjugated saccharides, yielding a

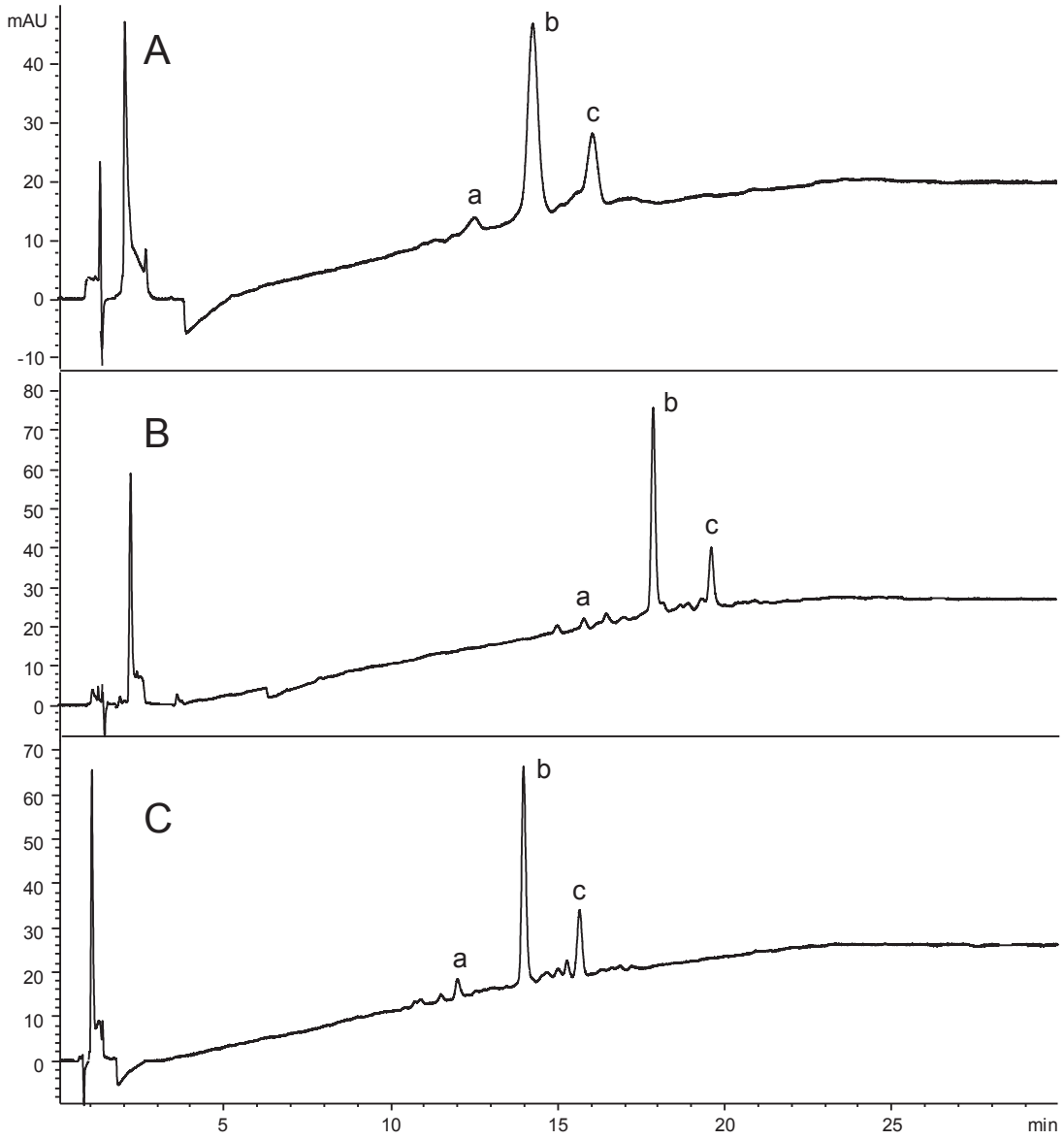


Fig. 4. HILIC-UV chromatograms obtained for TB10.4 (0.5 mg mL^{-1}) conjugated with Man-IME. Column, (A) TSKgel Amide-80, (B) XBridge BEH Amide, and (C) AdvanceBio Glycan Mapping. Sample solvent, ACN-water (50:50, v/v); gradient, (A) from 21% to 31% B in 20 min followed by isocratic elution at 31% B for 10 min, and (B,C) from 17% to 27% B in 20 min followed by isocratic elution at 27% B for 10 min. Peak assignment, (a) TB10.4, (b) TB10.4-Man, (c) TB10.4-$(Man)_2$. Further conditions, see Experimental Section.

Table 2
Resolutions (Rs) obtained during HILIC-UV of TB10.4 conjugated with Man-IME and Man(1–6)Man-IME on three columns. Components, (a) non-glycosylated TB10.4, (b) monoglycosylated TB10.4, and (c) diglycosylated TB10.4.

HILIC column	TB10.4-Man			TB10.4-Man(1–6)Man		
	RT a; b; c (min)	Rs (a,b)	Rs (b,c)	RT a; b; c (min)	Rs (a,b)	Rs (b,c)
TSK-gel Amide 80	12.51; 14.33; 16.09	3.0	3.0	12.49; 16.43; 19.85	5.8	6.13
XBridge BEH Amide	16.04; 18.11; 19.82	8.1	6.2	16.01; 20.55; 24.23	17.3	12.7
AdvanceBio Glycan Mapping	12.02; 13.97; 15.65	7.4	7.0	11.99; 15.99; 19.30	15.6	13.7

resolution which is based on the glycan load. Moreover, the increase of the HILIC retention *per* glycan is clearly larger for Man(1–6)Man (disaccharide) than for Man (monosaccharide). The observed glycoform peaks (Fig. 3) are relatively broad, occasionally showing fine structure, particularly for the Man(1–6)Man conjugates. These wide, split peaks most probably are caused by the partial HILIC separation of positional isomers of *neo*-glycoproteins with the same number of conjugated saccharides, but different binding location on the protein backbone.

3.2.2. TB10.4 glycoconjugates

The tuberculosis antigen TB10.4 (103 amino acids, MW 11,076 Da) comprises one lysine residue in its amino acid sequence. The protein was glycosylated with Man-IME and Man(1–6)Man-IME using a previously optimized procedure [3], and the reaction products were purified and analysed. The HILIC gradient conditions

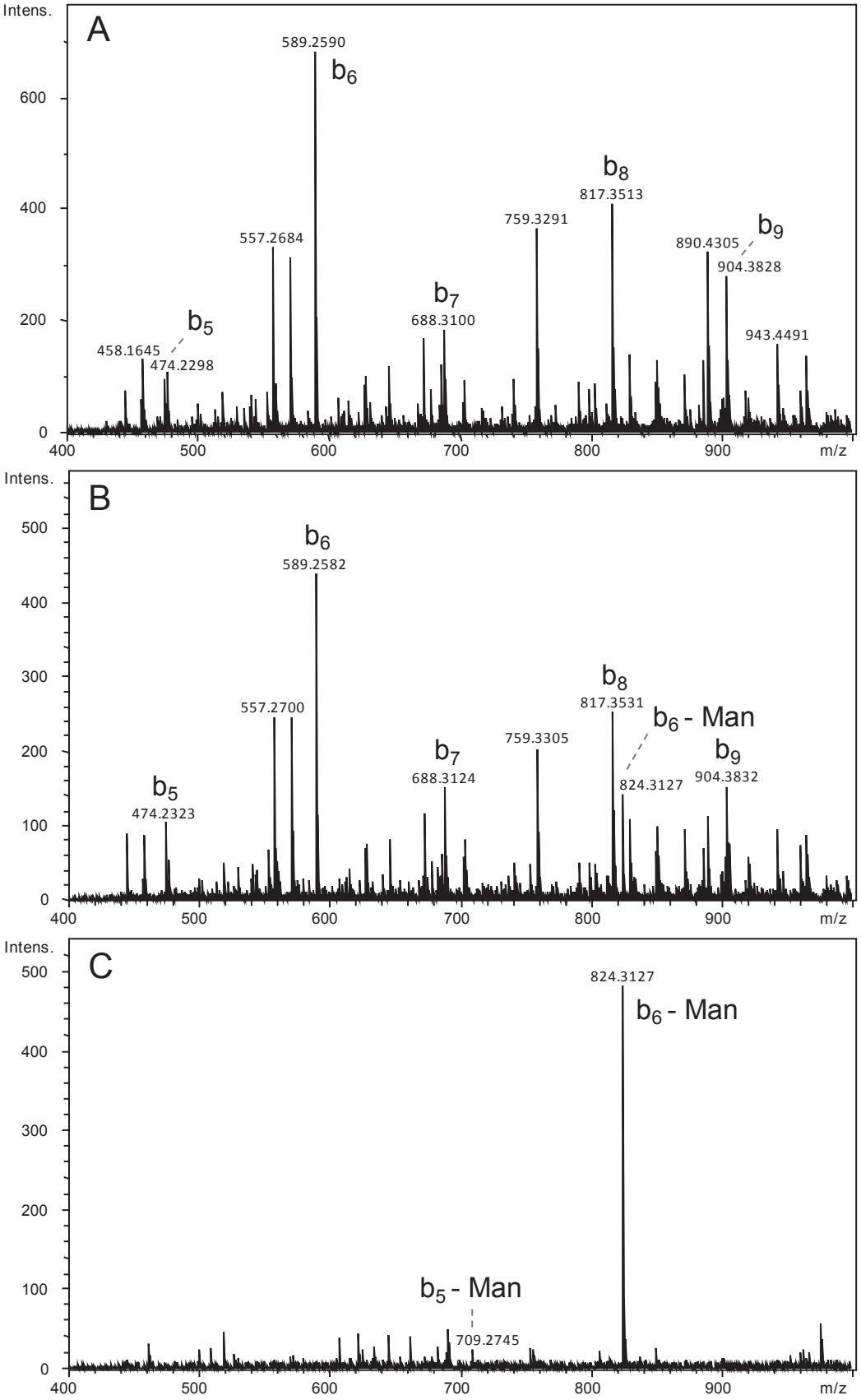


Fig. 5. MS/MS spectra (400–800 *m/z*) of (A) TB10.4, (B) TB10.4-Man and (C) TB10.4-$(Man)_2$ obtained during HILIC-MS/MS of TB10.4 (0.5 mg mL^{-1}) conjugated with Man-IME. Parent ion, 6 + charge state of each protein species; column, AdvanceBio Glycan Mapping. Further conditions, see Fig. 4 and Experimental Section.

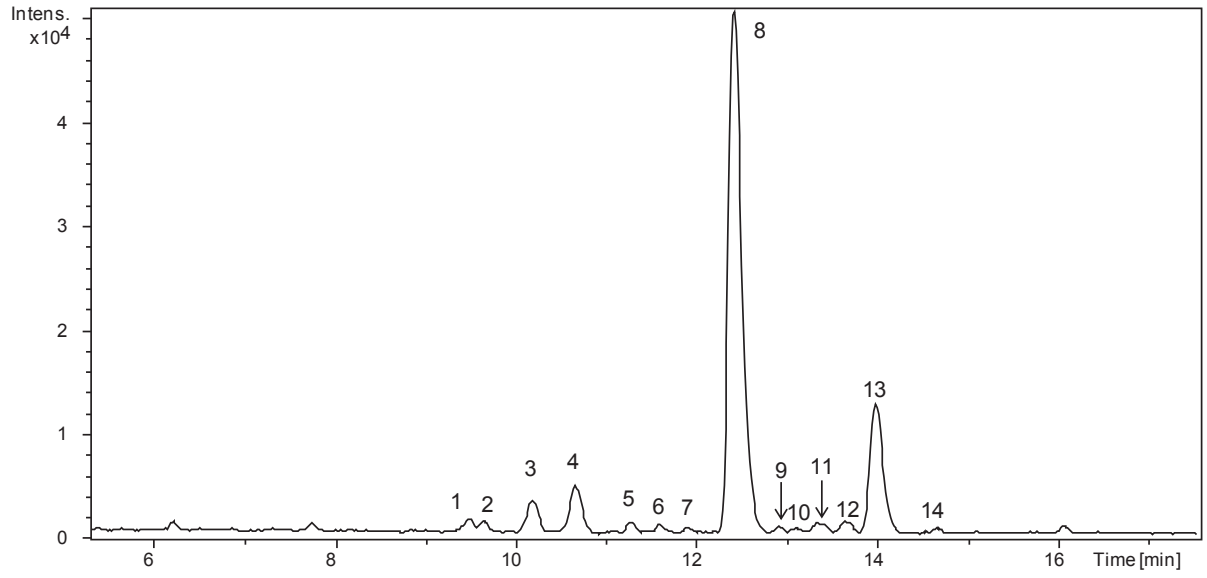


Fig. 6. HILIC-MS of TB10.4 (0.5 mg mL^{-1}) conjugated with Man-IME. Column, AdvanceBio Glycan Mapping; gradient, from 17% to 27% B in 20 min followed by isocratic elution at 27% B for 10 min. Further conditions, see Experimental Section.

Table 3
Species observed during HILIC-MS analysis of TB10.4 conjugated with Man-IME.

Peak	RT (min)	Experimental mass (Da)	Assignment	Theoretical mass (Da)
1	9.5	9933.4	A1-A92	9933.1
		9862.3	A1-M91	9862.0
2	9.7	10088.6	A1-R93	10088.8
3	10.2	9730.5	A1-M90	9730.8
4	10.7	9528.4	A1-M88	9528.5
		9599.5	A1-A89	9599.6
		10505.8	A1-E97	10505.6
		10646.9	A1-A99	10647.8
		11076.1	**TB10.4**	11076.3
5	11.2	7651.4	G36-G103	7651.1
		7979.6	Q33-G103	7979.5
		10097.4	A1-A91-Man	10097.1
		10168.5	A1-A92-Man	10168.2
6	11.6	6469.9	S48-G103-Man	6469.9
		6711.0	L46-G103-Man	6711.2
		9426.3	M18-G103-Man	9426.2
		8916.1	G23-G103-Man	8916.4
7	11.9	9965.5	A1-M90-Man	9965.8
8	12.4	11311.2	**TB10.4-Man**	11311.3
9	12.9	11011.0	A1-K100-Man	11011.0
10	13.1	10494.7	S9-G103-Man	10494.4
11	13.3	11327.2	Oxidised TB10.4-Man	11327.3
12	13.6	11327.2	Oxidised TB10.4-Man	11327.3
		9061.8	Q10-M90-Man	9061.8
13	14.0	11546.2	**TB10.4-Man_2**	11546.5
14	14.7	11562.3	Oxidised TB10.4-Man_2	11562.5

The main products TB10.4 and its glycoconjugates are in bold.

Table 4
Retention times (RT) and resolutions (Rs) obtained for successively eluting glycoforms during HILIC-UV on three columns of A85B conjugated with Man-IME. ND, not determinable.

HILIC column	mean RT (min)	Rs (2,1)	Rs (3,2)	Rs (4,3)	Rs (5,4)	Rs (6,5)	Average $\alpha \pm$ SD Rs $\pm$ SD
TSK-gel Amide 80	17.43	ND	ND	ND	ND	ND	ND
Waters XBridge BEH Amide	15.93	0.9	1.0	1.0	1.0	0.9	1.0 ± 0.1
Agilent AdvanceBio Glycan Mapping	15.01	1.1	1.1	1.0	1.2	1.2	1.1 ± 0.1

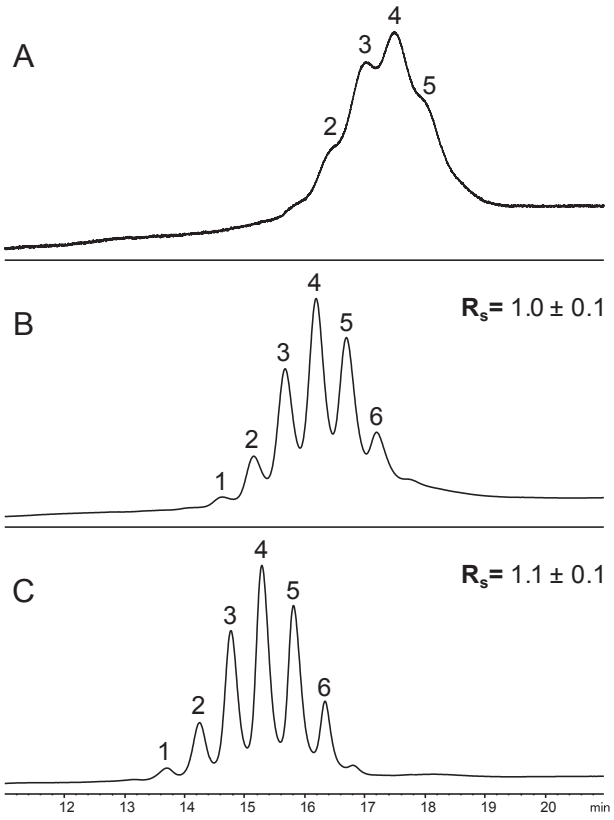


Fig. 7. HILIC-UV chromatograms obtained for Ag85B (0.5 mg mL^{-1}) conjugated with Man-IME. Column, (A) TSKgel Amide-80, (B) XBridge BEH Amide, and (C) AdvanceBio Glycan Mapping. Sample solvent, ACN-water (50:50, v/v); gradient, (A) from 26% to 36% B in 20 min followed by isocratic elution at 36% B for 10 min, and (B,C) from 22% to 32% B in 20 min followed by isocratic elution at 32% B for 10 min. Further conditions, see Fig. 1.

were slightly adjusted in order to achieve appreciable retention and good glycoform separation. The starting water percentage was decreased from 32% to 21% for the TSKgel Amide-80 column, and from 28% to 17% for the XBridge BEH Amide and AdvanceBio Glycan Mapping columns, while maintaining the same gradient slope (0.5% min^{-1}) and time (20 min) as for the RNase B and A samples. Fig. 4 shows the HILIC-UV chromatograms for TB10.4 conjugated with Man-IME using the various columns. The resolution for the main peaks observed for the TB10.4 samples after conjugation with Man-IME and Man(1–6)Man-IME, are reported in Table 2. Each HILIC column separated the TB10.4 variants (a, b, and c in Fig. 4A–C). However, due to narrower peaks obtained with the XBridge BEH Amide and AdvanceBio Glycan Map columns, obtained resolution values for these columns were significantly higher than for the TSKgel Amide-80. The better separation also revealed the presence of several minor sample components (Fig. 4B and C; see below for assignment). Comparison of Rs values obtained for the Man-IME and Man(1–6)Man-IME conjugates (Table 2) shows that doubling the size of the conjugated saccharide significantly enhances the HILIC resolution (factor 2.0). Clearly, the relative contribution of the size of the glycan moiety on the HILIC retention of the intact protein species is very large.

HILIC-MS allowed assignment of the peaks a, b and c to the non-glycosylated, monoglycosylated, and diglycosylated TB10.4 species, respectively. The presence of the latter can be explained by simultaneous conjugation of the amino group of the lysine residue and the N-terminal alanine of TB10.4 [3,22]. Glycosylation of the N-terminus of TB10.4 conjugated with Man-IME was confirmed by employing the MS/MS capabilities of the QTOF mass spectrometer after HILIC separation. The most abundant ion (charge state, 6+) obtained for TB10.4 (peak a) and mannosylated TB10.4 (peaks b and c) were fragmented by CID and the product-ion spectra were recorded (Fig. 5). The b_6 fragment was used as diagnostic ion for the assignment of N-terminal glycosylation, as it was observed with good intensity for all fragmented species. The spectrum of TB10.4 (Fig. 5A) shows an ion with *m/z* of 589.26 which corresponds to the expected value for the b_6 ion of TB10.4 that has a non-glycosylated N-terminus. The product-ion spectrum of TB10.4-$(Man)_2$ (Fig. 5C) shows a clear mannosylated b_6 ion (*m/z* 824.31) indicating glycosylation at the N-terminus. Interestingly, the product-ion spectrum of TB10.4-(Man) (Fig. 5B), shows b_6 ions (*m/z* 589.26 and 824.31) corresponding, respectively, to the non-glycosylated and glycosylated N-terminus of TB10.4. This indicates that the main mannosylated TB10.4 product actually is comprised of two isomeric glycoforms, which are not separated under the applied HILIC conditions, and which are conjugated either at the lysine or at the N-terminal alanine. Notably, this kind of structural information cannot be obtained by LC-MS using a bottom-up procedure.

Using HILIC-MS employing the AdvanceBio Glycan Mapping column, several minor components were observed next to the three TB10.4 species (Fig. 6), which were formed during the glycosylation process. Deconvolution of the mass spectra obtained in the apices of the peaks revealed the presence of several degradation products formed during the glycosylation process. A large part of these products could be assigned to truncated protein forms that lost a number of amino acids from the C- or N-terminus side (Table 3). In addition, some oxidation products were observed showing larger retention than the native proteins. Oxidation is a common protein degradation occurring under the glycosylation conditions [22]. Notably, the HILIC column is capable to fully resolve the relatively small difference in protein structure induced by a single oxidation. The two peaks observed for oxidised TB10.4-Man most probably represent positional isomers.

3.2.3. *Ag85B glycoconjugates*

The antigenic protein Ag85B (292 amino acids, MW 31,346 Da) presents eight lysine residues as potential glycosylation sites. *Neo*-glycoproteins were produced by conjugation with Man-IME, Man(1–6)Man-IME and Man(1–6)Man(1–6)Man-IME. For larger proteins, solubility in the ACN-rich eluent and irreversible adsorption to HILIC materials might become an issue [23]. However, using gradient elution with a starting water percentage of 26% for the TSKgel Amide-80 column, and 22% for the XBridge BEH Amide and AdvanceBio Glycan Mapping columns, symmetric peaks and repeatable retention times and peak areas were obtained in the HILIC-UV traces of non-glycosylated Ag85B. For all columns the RSD values (n = 5) for retention time and peak area were below 0.4% and 2.5%, respectively, indicating proper performance.

Fig. 7 shows the HILIC separations of Ag85B glycoforms obtained after Man-IME conjugation using the three columns. The resolutions obtained for the Ag85B *neo*-glycoproteins differed considerably (Table 4). Using the AdvanceBio Glycan Mapping column (Fig. 7C) an almost baseline separation of six glycoforms was achieved. The same number of peaks was detected with the XBridge BEH Amide column (Fig. 7B), but the glycoforms partially co-eluted. The profile of overlapping peaks obtained with the TSKgel Amide-80 column (Fig. 7A) did not allow calculation of peak resolutions. The relatively small pore size of the TSKgel Amide-80 particles clearly compromises the separation performance for proteins larger than 15 kDa.

In order to achieve glycoform assignment, Ag85B conjugated with the mono-, di- and tri-mannose saccharides were analysed by

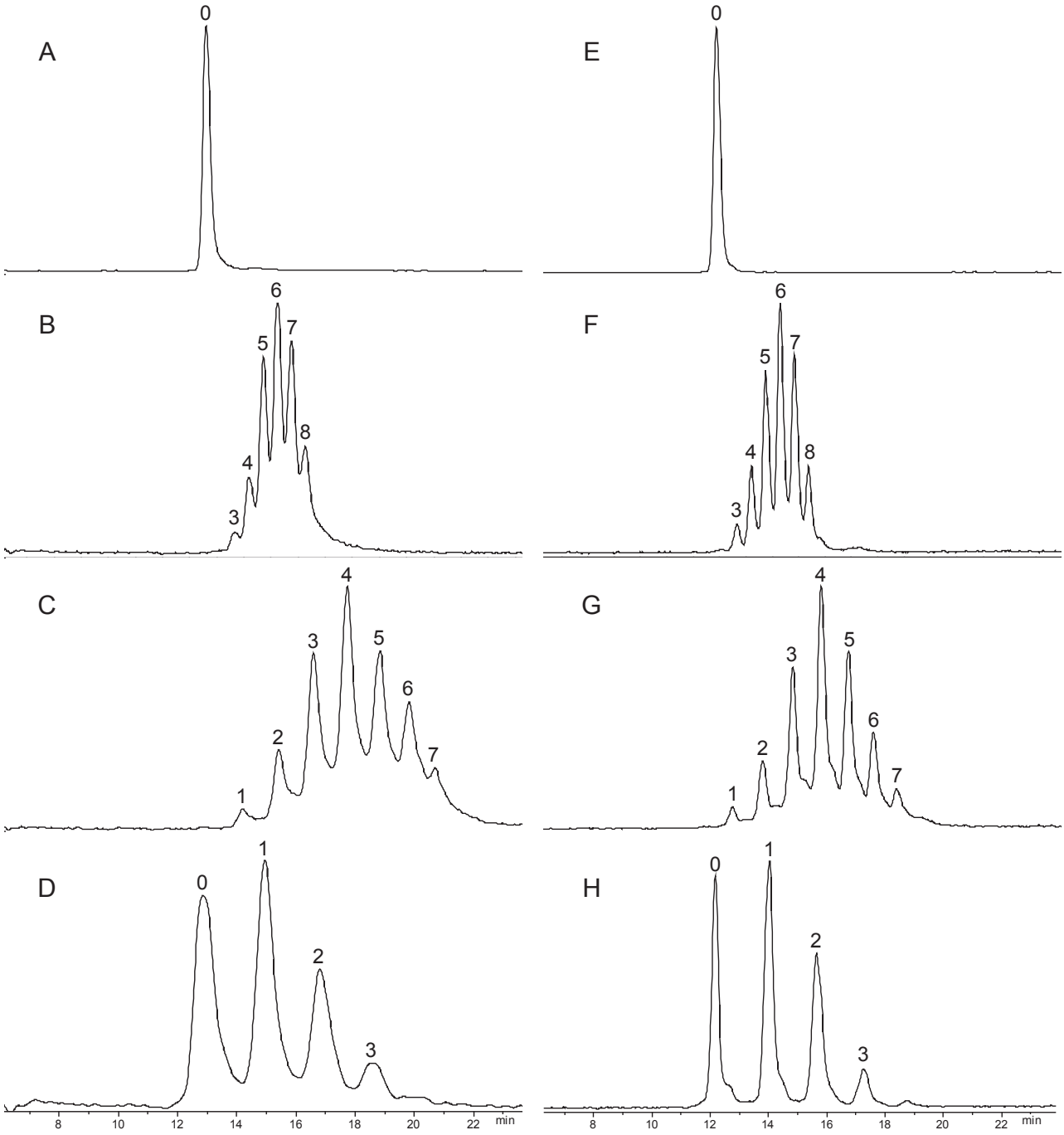


Fig. 8. HILIC-MS chromatograms obtained for Ag85B (A,E) conjugated with Man-IME (B,F), Man(1–6)Man-IME (C,G) and Man(1–6)Man(1–6)Man-IME (D,H). Column, (A–D) XBridge BEH Amide, and (E–H) AdvanceBio Glycan Mapping. Numbers indicate number of conjugated glycans. Further conditions, see Fig. 7 and Experimental Section.

HILIC-MS employing the XBridge BEH Amide and AdvanceBio Glycan Mapping columns (Fig. 8). MS detection enabled the unambiguous determination of the number of saccharide units conjugated. For Ag85B reacted with Man-IME, protein species bearing 3 to 8 glycans could be detected (Fig. 8B and F), indicating all Ag85B molecules had been conjugated. Using Man(1–6)Man-IME (Fig. 8C and G) Ag58B quantitatively reacted with the sugar, but the average degree of substitution (1–7 glycans attached) was lower. The sample coming from Ag85B conjugation with Man(1–6)Man(1–6) Man-IME (Fig. 8D and H) contained a significant amount of unmodified antigen and only mono-, di-, and tri-glycosylated protein species. Evidently, the degree of substitution decreases with the size of the attached glycan, which is in accordance with the relative chemical reactivity of the activated sugars. From the overall results provided in Fig. 8, it is very clear that the number and size of the glycans attached to the protein strongly add to the HILIC retention and resolution of the intact species.

With the AdvanceBio Glycan Mapping column highest glycoform resolutions were obtained for the Ag58B *neo*-glycoproteins (Fig. 8E–H). Interestingly, this good performance allowed detection of an extra low abundant component in the Ag58B sample conjugated with Man(1–6)Man(1–6)Man-IME (Fig. 9A) which eluted

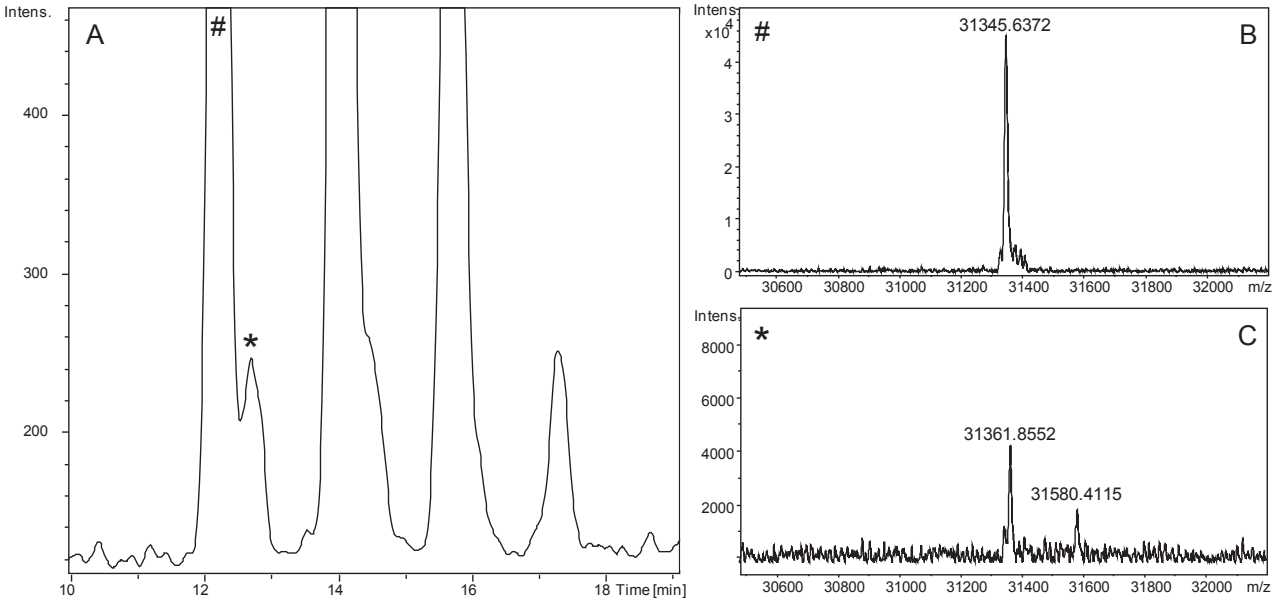


Fig. 9. (A) Zoom of HILIC-MS base peak chromatogram of Ag85B conjugated with Man(1–6)Man(1–6)Man-IME. (B) Deconvoluted mass spectrum obtained at the apex of the minor peak eluting at 12.7 min. Conditions, see Fig. 8.

just after the native Ag85B protein. The deconvoluted spectrum recorded at the apex of the extra peak (Fig. 9C) reveals two masses (31,361.8 and 31,580.4 Da) which are higher than the molecular mass of Ag85B (31,345 Da). The first mass might be due to mono-oxidised Ag85B. The second mass indicates the presence of Ag85B conjugated with Man-IME. The Man(1–6)Man(1–6)Man-IME reagent is not totally pure and may contain traces of Man-IME, leading to formation of minor amounts of Ag85B-Man conjugates. Reliable MS detection of these minor sample components among excess of other protein species is only possible due to the pre-separation provided by the HILIC selectivity.

4. Conclusion

HILIC-UV-MS(/MS) methods for the characterization of intact *neo*-glycoproteins were developed. Under optimized separation conditions, resolution and assignment of proteoforms and individual glycoforms of semi-synthetic intact glycoproteins were achieved. HILIC retention and selectivity appeared to be dominated by the size and number of the saccharides conjugated to the carrier protein. Three tested amide HILIC stationary phases showed similar selectivity for the glycosylated proteins. However, the XBridge BEH Amide and AdvanceBio Glycan Mapping columns showed better separation performance than the TSKgel Amide-80 column, leading to better glycoform resolution. For the larger glycoconjugated antigen Ag85B (MW, 31.3 kDa), optimum resolution was achieved with the AdvanceBio Glycan Mapping material, most likely due to the superficially porous nature of its particles. The HILIC-MS selectivity also permitted to detect and tentatively assign minor protein impurities and degradation products. Furthermore, MS/MS detection involving CID of glycoconjugates provided information on glycan position, confirming N-terminal glycosylation of the TB10.4 antigen.

Overall, the HILIC-UV-MS results obtained for the different glycosylated proteins suggest that with proper selection of the HILIC stationary phase and optimization of MS-compatible chromatographic conditions, it is possible to achieve very good glycoconjugate resolution. Moreover, intact glycoform identities of semi-synthethic glycoprotein products can be assessed without the need for sample pre-treatment.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.aca.2017.05.020.

References

[1] R. Adamo, A. Nilo, B. Castagner, O. Boutureira, F. Berti, G.J.L. Bernardes, Synthetically defined glycoprotein vaccines: current status and future directions, Chem. Sci. 4 (2013) 2995–3008.
[2] M. Dalziel, M. Crispin, C.N. Scanlan, N. Zitzmann, R.A. Dwek, Emerging principles for the therapeutic exploitation of glycosylation, Science 343 (2014) 1235681–1235688.
[3] C. Temporini, T. Bavaro, S. Tengattini, I. Serra, G. Marrubini, E. Calleri, F. Fasanella, L. Piubelli, F. Marinelli, L. Pollegioni, G. Speranza, G. Massolini, M. Terreni, Liquid chromatography–mass spectrometry structural characterization of *neo*-glycoproteins aiding the rational design and synthesis of a novel glycovaccine for protection against tuberculosis, J. Chromatogr. A 1367 (2014) 57–67.
[4] K. Mariño, J. Bones, J.J. Kattla, P.M. Rudd, A systematic approach to protein glycosylation analysis: a path through the maze, Nat. Chem. Bio 6 (2010) 713–723.
[5] M. Thaysen-Andersen, N.H. Packer, Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome, Biochim. Biophys. Acta 1844 (2014) 1437–1452.
[6] K. Sandra, I. Vandenheede, P. Sandra, Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization, J. Chromatogr. A 1335 (2014) 81–103.
[7] A. Staub, D. Guillarme, J. Schappler, J.-L. Veuthey, S. Rudaz, Intact protein analysis in the biopharmaceutical field, J. Pharm. Biomed. Anal. 55 (2011) 810–822.
[8] A. Periat, I.S. Krull, D. Guillarme, Applications of hydrophilic interaction chromatography to amino acids, peptides, and proteins, J. Sep. Sci. 38 (2015) 357–367.
[9] J. Bones, S. Mittermayr, N. O'Donoghue, A. Guttman, P.M. Rudd, Ultra performance liquid chromatographic profiling of serum N-Glycans for fast and efficient identification of cancer associated alterations in glycosylation, Anal. Chem. 82 (2010) 10208–10215.
[10] Y. Takegawa, K. Deguchi, T. Keira, H. Ito, H. Nakagawa, S.I. Nishimura, Separation of isomeric 2-aminopyridine derivatized N-glycans and N-glycopeptides of human serum immunoglobulin G by using a zwitterionic type of hydrophilic-interaction chromatography, J. Chromatogr. A 1113 (2006) 177–181.
[11] A. Periat, S. Fekete, A. Cusumano, J.-L. Veuthey, A. Beck, M. Lauber,

D. Guillarme, Potential of hydrophilic interaction chromatography for the analytical characterization of protein biopharmaceuticals, J. Chromatogr. A 1448 (2016) 81–92.
[12] V. D'Atri, S. Fekete, A. Beck, M. Lauber, D. Guillarme, Hydrophilic interaction chromatography hyphenated with mass spectrometry: a powerful analytical tool for the comparison of originator and biosimilar therapeutic monoclonal antibodies at the middle-up level of analysis, Anal. Chem. 89 (2017) 2086–2092.
[13] T. Tetaz, S. Detzner, A. Friedlein, B. Molitor, J.L. Mary, Hydrophilic interaction chromatography of intact, soluble proteins, J. Chromatogr. A 1218 (2011) 5892–5896.
[14] Z. Zhang, Z. Wu, M.J. Wirth, Polyacrylamide brush layer for hydrophilic interaction liquid chromatography of intact glycoproteins, J. Chromatogr. A 1301 (2013) 156–161.
[15] A. Pedrali, S. Tengattini, G. Marrubini, T. Bavaro, P. Hemström, G. Massolini, M. Terreni, C. Temporini, Characterization of intact neo-glycoproteins by hydrophilic interaction liquid chromatography, Molecules 19 (2014) 9070–9088.
[16] L. Piubelli, M. Campa, C. Temporini, E. Binda, F. Mangione, M. Amicosante, M. Terreni, F. Marinelli, L. Pollegioni, Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins, Microb. Cell Fact. 12 (2013) 115.
[17] G. Kallenius, A. Pawlowski, B. Hamasur, S.B. Svenson, Mycobacterial glycoconjugates as vaccine candidates against tuberculosis, Trends Microbiol. 16 (2008) 456–462.
[18] T. Bavaro, M. Filice, C. Temporini, S. Tengattini, I. Serra, C.F. Morelli, G. Massolini, M. Terreni, Chemoenzymatic synthesis of *neo*glycoproteins driven by the assessment of protein surface reactivity, RSC Adv. 4 (2014) 56455–56465.
[19] W.Z. Shou, W. Naidong, Simple means to alleviate sensitivity loss by trifluoroacetic acid (TFA) mobile phases in the hydrophilic interaction chromatography–electrospray tandem mass spectrometric (HILIC–ESI/MS/MS) bioanalysis of basic compounds, J. Chromatogr. B 825 (2005) 186–192.
[20] R. Kostiainen, T.J. Kauppila, Effect of eluent on the ionization process in liquid chromatography–mass spectrometry, J. Chromatogr. A 1216 (2009) 685–699.
[21] Y. Chen, A.R. Mehok, C.T. Mant, R.S. Hodges, Optimum concentration of trifluoroacetic acid for reversed-phase liquid chromatography of peptides revisited, J. Chromatogr. A 1043 (2004) 9–18.
[22] S. Tengattini, E. Domínguez-Vega, L. Piubelli, C. Temporini, M. Terreni, G.W. Somsen, Monitoring antigenic protein integrity during glycoconjugate vaccine synthesis using capillary electrophoresis-mass spectrometry, Anal. Bioanal. Chem. 408 (2016) 6123–6132.

3. *Enterokinase monolithic bioreactor as an efficient tool for biopharmaceuticals preparation: on-line cleavage of fusion proteins and analytical characterization of released products*

Sara Tengattini, Francesca Rinaldi, Luciano Piubelli, Tom Kupfer, Benjamin Peters, Teodora Bavaro, Enrica Calleri, Gabriella Massolini, Caterina Temporini

Draft

Enterokinase monolithic bioreactor as an efficient tool for biopharmaceuticals preparation: on-line cleavage of fusion proteins and analytical characterization of released products

Sara Tengattini[a], Francesca Rinaldi[a], Luciano Piubelli[b], Tom Kupfer[c], Benjamin Peters[c], Teodora Bavaro[a], Enrica Calleri[a], Gabriella Massolini[a], Caterina Temporini[a]

[a] *Department of Drug Sciences, University of Pavia, Viale Taramelli 12, 27100 Pavia, Italy*
[b] *Department of Biotechnology and Life Sciences, University of Insubria, Via Dunant 3, 21100 Varese, Italy and The Protein Factory Research Centre, Politecnico of Milano and University of Insubria, Via Mancinelli 7, 20131 Milano, Italy*
[c] *Merck KGaA, Frankfurter Straße 250, 64293 Darmstadt, Germany*

Abstract

One of the most popular enzymes used for the *in vitro* cleavage of fusion proteins is enterokinase (EK, EC 3.4.21.9). EK cleaves with high specificity after the sequence Asp_4-Lys (DDDDK), which allows the fusion protein to preserve its native amino acid terminus without any additional unwanted cleavage residue from the recognition sequence. However, the complete removal of EK after protein cleavage is a critical step to ensure protein identity and stability.

As enzyme immobilization increases stability and reusability of the biocatalyst while reducing operating costs and sample contamination, in this work we report the covalent immobilization of recombinant EK (rEK) on monolithic chromatographic supports with different binding chemistries for the development of a rEK-chromatographic-bioreactor. An on-line assay for the determination of the activity of immobilized rEK was set up using a synthetic substrate (Gly-Asp_4-Lys-β-naphthylamide, GD_4K-NA). The assay was used to study the improvement of the operational conditions (temperature and flow rate) on hydrolytic activity of the bioreactor. The immobilization yields, as well as the cleavage activity of immobilized rEK on GD_4K-NA, were highly satisfactory when the immobilized enzyme reactor was used in recirculation. The ability of the immobilized rEK to cleave fusion proteins was tested by recirculation of thioredoxin (Trx)-TB10.4 and Trx-Ag85B His-tagged proteins yielding the mature antigens TB10.4 and Ag85B, to be used in the preparation of potential novel glycovaccines against tuberculosis. The prepared rEK-based immobilized enzyme reactors proved to efficiently cleave the considered fusion proteins even if the cleavage specificity at

the canonical site was not fully achieved. The immobilized rEK showed very good stability and reusability.

Abbreviations

2NA: 2-naphthylamide
ACN: Acetonitrile
EK: Enterokinase
GD_4K-NA: Gly-Asp_4-Lys-β-naphthylamide
His-tag: Poly-histidine tag
HPLC: High Performance Liquid Chromatography
IMER: Immobilized enzyme reactor
LC: Liquid Chromatography
MS: Mass Spectrometry
rEK: recombinant enterokinase
RP: Reverse Phase
SDS-PAGE: Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis
TFA: Trifluoroacetic acid
Tris: Tris(hydroxymethyl)aminomethane
Trx: Thioredoxin

1. Introduction

Proteins have become attractive for therapeutic applications, as witnessed by the high number of approved protein therapeutics. Most of them are produced by biotechnological processes in host organisms, involving recombinant DNA technology that allows therapeutic proteins to be produced on a large scale. The successful production of recombinant proteins requires an efficient expression, a good solubility of the protein and its easy purification. Fusion protein technology is frequently used to attach a "tag", which facilitates the purification and identification of the protein of interest. In addition, the incorporation of fusion tags can increase expression yields and improve protein solubility and folding [1]. Fusion proteins have been used for the production of recombinant proteins of therapeutic interest such as antibodies, coagulation factors, growth hormones and vaccines [2, 3].

Numerous tags have been developed for recombinant protein production and purification. The most frequently used include FLAG-tag, histidine affinity tag (HAT-tag), poly-histidine tag (His-tag), glutathione S-transferase (GST), maltose-binding protein (MBP), N-utilization substance protein A (NusA), S-tag, streptavidin-binding peptide (SBP)-tag, Strep-tag, and thioredoxin (Trx). Some tags (such as MBP, GST and Trx) are useful for increasing solubility and folding, others are needed for purification (*e.g.* His-tag) [4-6].

Once the fusion protein is expressed and isolated, it is often necessary to remove the tag as the presence of the fusion partner may affect the structure or the biological function of the protein of interest. Tag cleavage can be achieved through chemical (*e.g.* cyanogen bromide or hydroxylamine) or enzymatic treatments. Enzymatic cleavage is often preferred in research laboratories due to the higher selectivity and the milder reaction conditions, while the industrial application of the biocatalytic approach is limited by the relatively high amounts of enzyme requested and its high cost [7]. After the production of mature (*i.e.* tagless) protein, the cleaving enzyme needs to be removed from the purified protein, to avoid protease contamination and instability of the final protein preparation.

The most used proteolytic enzymes include enterokinase (EK), factor Xa, small ubiquitin related modifier (SUMO) protease, tobacco etch virus (TEV) protease, thrombin and human rhinovirus 3C protease (HRV 3C) [8].

Enterokinase (synonym: enteropeptidase, EC 3.4.21.9) is a heterodimeric serine protease found in the duodenum that activates trypsinogen by highly specific cleavage of the trypsinogen activation peptide following the sequence Asp_4-Lys [9]. Its high site-specificity makes this enzyme of great biotechnological interest because the sequence Asp_4-Lys can be placed at the native amino acid *N*-terminus without any additional unwanted cleavage residue from the recognition sequence, yielding the exact native protein upon EK cleavage of the fusion protein.

In research laboratories, recombinant enterokinase (rEK) is often used. rEK is a highly purified preparation of the catalytic subunit of EK able to recognize the amino acid sequence Asp_4-Lys as the native enzyme. It has been demonstrated that rEK exhibits superior cleavage rates of fusion proteins containing the recognition sequence when compared to native enzyme [10-12]. The immobilization of rEK might represent an attractive option in order to improve its activity, stability and reusability and to obtain therapeutic proteins at low manufacturing costs.

rEK has been immobilized on porous materials [13] and on commercial paramagnetic microspheres [14, 15]. More recently, its immobilization was performed onto iron oxide magnetic nanoparticles coated with biopolymers. Two different chemistries have been explored for the covalent coupling of rEK, namely carbodiimide (EDC coupling) and maleimide (Sulfo coupling) activation [16]. Although the reusability of the biocatalyst was assessed for several hydrolysis cycles in all the studies, the activity on control fusion proteins was generally demonstrated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). This technique can give information on the cleavage of fusion proteins but no details on site-specificity and on the quality of the released protein

can be obtained. To our best knowledge, no immobilized enzyme reactors (IMERs) based on rEK to be used on a dynamic system (*i.e.* liquid chromatography) for the cleavage of fusion tag proteins are reported in literature. The application of EK-IMERs in tag removal can reduce enzyme consumption and the use of an in-flow reaction system, in principle, can allow the production of large amounts of therapeutic proteins. Moreover, the use of stable covalently bound EK would decrease the risk of residual enzyme traces in the released protein solution, thus improving product stability.

In this work, we describe the development of a rEK-based silica monolith chromatographic reactor. The amount, as well as the activity of the immobilized enzyme were determined. The kinetic characterization of the IMER was carried out on the synthetic peptide Gly-Asp_4-Lys-β-naphthylamide (GD_4K-NA). The experimental parameters (flow rate and temperature) were studied to maximize the reaction yields. After evaluating the stability and reaction characteristics of the immobilized enzyme, we compared its cleavage performance with the free enzyme in both single injection and recirculating reaction modes. Finally, the specificity of the cleavage site was verified by liquid chromatography-mass spectrometry (LC-MS) and compared with the results obtained with the free enzyme using two model fusion proteins, namely Trx-Ag85B and Trx-TB10.4, yielding the mature proteins Ag85B and TB10.4 used as protein scaffolds in the preparation of a novel glycovaccine against tuberculosis [17-22]. To improve the cleavage site-specificity of the immobilized rEK, the effect of immobilization chemistry (*via* epoxy and propylamine supports), support features (normal and wide pore monolithic materials), as well as the type of enzyme were considered.

2. Materials and methods

2.1. Reagents and chemicals

All chemicals were of analytical grade. Gly-Asp_4-Lys-β-naphthylamide (GD_4K-NA), gradient grade acetonitrile (ACN) for HPLC, urea, sodium periodate, methanol, glutaraldehyde and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Imidazole, sodium azide, ammonium sulphate, sodium dihydrogen phosphate and tris(hydroxymethyl)aminomethane (Tris) were from Merck KGaA (Darmstadt, Germany). Hydrochloric acid, sulfuric acid, ortophosphoric acid, calcium chloride, sodium chloride, sodium hydroxide, potassium phosphate and monoethanolamine were purchased from Carlo Erba Reagents (Cornaredo, Italy). Water was obtained from a Milli-Q® Integral system (Merck KGaA, Darmstadt, Germany). Sodium cyanoborohydride was from Alfa Aesar (Haverhill, MA, USA).

Syringe filters were purchased from AISIMO corporation CO., LTD (London, England).

Trx-TB10.4 and Trx-Ag85B fusion proteins and TB10.4 and Ag85B mature proteins were obtained as recombinant forms in *Escherichia coli* as reported by Piubelli *et al.* [17] and finally collected in 20 mM 3-(N-morpholino) propanesulfonic acid (MOPS), 0.4 M NaCl, pH 7.0.

2.2. Enterokinase

Recombinant bovine enterokinase (rEK) was purchased from Sino Biological Inc. (Beijing, China) (purity > 95%) and from Merck KGaA (Darmstadt, Germany) (purity > 99%). The bovine rEK consists of 234 amino acids and has a calculated molecular mass of 26.1 kDa. As a result of glycosylation, the recombinant protein migrates as an approximately 45 kDa protein in SDS-PAGE under reducing conditions.

rEK from Sino Biological was purchased as a powder (100 U) and was resuspended in sterile water to obtain a stock solution of 100 U/mL in 20 mM Tris-HCl, 200 mM NaCl, 10% glycerol, 5% mannitol, 5% trehalose, pH 7.2. rEK from Merck was purchased as solution in 20 mM Tris-HCl, 200 mM NaCl, 2 mM $CaCl_2$, 50% glycerol, pH 7.4.

When not specified, experiments were carried out using rEK from Sino Biological.

2.3. In solution enzymatic assays

2.3.1. Assay using GD_4K-NA substrate

To evaluate in solution rEK activity, the synthetic substrate GD_4K-NA was used. The digestion was conducted in 20 mM Tris-HCl, 50 mM NaCl, 2 mM $CaCl_2$, pH 7.4 (digestion buffer) and at 30 °C under continuous stirring, using a substrate concentration of 500 µg/mL and a rEK/substrate ratio of 1 U/1 mg. First, to define the reaction kinetic, digestion was monitored until complete substrate conversion was observed (6h). To determine the rEK immobilized units, the assay was carried out on both pre- and post-immobilization solutions in the linear velocity range (30 min).

For all the samples, digestion was stopped by addition of 100 mM HCl to pH 3. For quantification, the resulting solutions (50 µg/mL) were analyzed by high performance liquid chromatography (HPLC)-UV (see paragraph 2.5.1).

2.3.2. Assay using fusion proteins

To evaluate in solution rEK activity, the digestions of the recombinant fusion proteins Trx-TB10.4 and Trx-Ag85B were also performed. Digestions were conducted in 20 mM Tris-HCl, 50 mM NaCl,

2 mM $CaCl_2$, pH 7.4, using a fusion protein concentration of 600 µg/mL and a rEK/protein ratio of 1 U/2 mg.

The solutions were incubated for 2 h at room temperature under continuous stirring and stopped by storing at -20 °C. After dilution to 60 µg/mL in the digestion buffer/ACN, 60/40 (v/v) the rEK reaction solutions were injected in HPLC-UV and HPLC-MS (see paragraph 2.5.2) in order to define rEK activity and cleavage specificity.

2.4. Enzyme immobilization

2.4.1. Immobilization on epoxy silica

rEK was immobilized on an epoxy-modified silica Chromolith® Flash (4.6 × 25 mm) monolithic support (Merck KGaA, Darmstadt, Germany) following a procedure developed by our group [23]. Briefly, 10 U of rEK were dissolved in 5 mL of grafting solution (50 mM phosphate buffer 1.9 M in ammonium sulphate, pH 8.0) and filtered by 0.45 µm filters. Before immobilization, the monolithic column was equilibrated with 30 mL of grafting solution. The rEK solution was then pumped for 4 h through the support at 500 µL/min, flushing and back-flushing every 15 min. After immobilization, the column was washed with 40 mL of 10 mM phosphate buffer, pH 6.0, at 50 µL/min, then the unreacted epoxide groups were blocked with 60 mL of a 1 M monoethanolamine solution (flow rate 500 µL/min). Finally, the obtained bioreactor was equilibrated with 50 mM NaCl, 20 mM Tris-HCl, 2 mM $CaCl_2$, pH 7.4.

All the flushing steps were performed using a quaternary pump from Agilent series 1100 system (Santa Clara, CA, USA).

To determine the immobilization yields, the rEK solutions were submitted to the enzyme activity assay described in paragraph 2.3.1, before and after the immobilization procedure.

2.4.2. Immobilization via Schiff base mechanism

For the immobilization *via* Schiff base mechanism, the enzyme was immobilized on a Chromolith® (4.6 × 25 mm) epoxy support (Merck KGaA, Darmstadt, Germany) following the protocol suggested by Merck (protocol A) [24] and on a propylamine Chromolith® support (4.6 × 25 mm) (Merck KGaA, Darmstadt, Germany) using a method described by our research group (protocol B) [25].

For these columns too, all the flushing steps were performed using a quaternary pump from Agilent series 1100 system (Santa Clara, CA, USA) and the rEK solutions before and after the immobilization

procedures were submitted to the enzyme activity assay described in paragraph 2.3.1 to define the immobilization yields.

Protocol A: 10 U of rEK were dissolved in 20 mL of 50 mM phosphate buffer 1.9 M in ammonium sulphate, pH 8.0 in presence of 5 mM $NaBH_3CN$. After washing the column with 9 mL water, it was equilibrated using 30 mL of 2% H_2SO_4 in water to hydrolyze epoxide groups (flow rate 2 mL/min). The hydrolysis was followed by another washing step in water and an oxidation of diol functions by flowing 100 mM $NaIO_4$ in H_2O/MeOH 4/1 (v/v) at 2 mL/min for 21 min. The activated support was washed again with water and the rEK solution was continuously flowed in the column for 16 h at 200 µL/min. After immobilization, the unreacted aldehyde functions were reduced by flowing 30 mL of 20 mM $NaBH_3CN$ in 50 mM NaH_2PO_4, pH 3 (flow rate 2 mL/min). At the end, the obtained bioreactor was equilibrated with 50 mM NaCl, 20 mM Tris-HCl, 2 mM $CaCl_2$, pH 7.4.

Protocol B: the monolithic column was inserted into the HPLC system and conditioned with 50 mM phosphate buffer, pH 7.5 at 500 µL/min for 20 min. Subsequently, the support was activated with a 10% (v/v) glutaraldehyde solution in the same buffer at 500 µL/min for 5 h. The activated support was then washed with 100 mM phosphate buffer, pH 7.5 at 500 µL/min for 10 min to remove the excess of glutaraldehyde. Immediately after, 20 mL of the enzyme solution (0.5 U/mL) in 100 mM phosphate buffer, pH 7.5 were recirculated through the column at 25 °C overnight. The imines resulting from the reaction of glutaraldehyde with the enzyme amino groups were reduced by pumping a 100 mM $NaBH_3CN$ solution in the same buffer at 200 µL/min for 1.5 h. The column was washed with the buffer for 2 h with a 100 mM monoethanolamine solution in phosphate buffer (100 mM, pH 7.5). Finally, the column was washed again with phosphate buffer (100 mM, pH 7.5).

2.4.3. rEK IMER activity assay

The activity of immobilized rEK was assessed using GD_4K-NA, a synthetic substrate widely used for the determination of EK activity [26]. A solution of GD_4K-NA (100 µg/mL) in digestion buffer was continuously recirculated through the bioreactor using a HPLC pump from Agilent series 1100 system (Santa Clara, CA, USA). The flow rate was set at 400 µL/min and the temperature at 30 °C. At scheduled times, 20 µL of substrate solution were collected and injected without dilution in HPLC-UV or HPLC-MS. The amount of substrate was derived by the calibration curve and the conversion percentage was calculated.

2.5. Instrumentation and analytical methods

LC-UV separations were performed on an Agilent HPLC series 1100 system (Santa Clara, CA, USA), equipped with mobile phase online degasser, quaternary pump, autosampler, column thermostated compartment and diode array detector.

For LC-MS experiments, a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA, USA) controlled by Chromeleon software (version 6.8) was connected to a linear trap quadrupole (LTQ) MS with an ESI source (Thermo Fisher Scientific, Waltham, MA, USA), controlled by Xcalibur software 1.4.

2.5.1. Analysis of GD_4K-NA and 2-naphthylamide

GD_4K-NA and 2-naphthylamide (2NA) analyses were performed using a XTerra MS C18 column (5 μm, 125 Å, 2.1 × 100 mm) from Waters (Milford, MA, USA). The mobile phase was composed of water (A) and ACN (B), both containing 0.1% TFA. Chromatographic conditions consisted of a linear gradient from 10% to 60% B in 20 min. The column temperature was maintained at 30 °C, the injection volume was 5 μL and elution was carried out at constant flow of 150 μL/min. UV adsorption was monitored at 245 nm, the λ of maximum adsorption of GD_4K-NA as determined by acquisition of its UV adsorption spectrum.

For GD_4K-NA calibration curve, five different concentrations were considered (1, 20, 50, 100, 200 μg/mL) and three independent determinations were performed for each point. The response was linear and the equation of calibration curve was $y = 82.90(8)x + 7(8)$, $R^2=1.000$.

GD_4K-NA and 2NA identities were confirmed by LC-MS analysis. Full scan MS experiments were carried out under the following instrumental conditions: positive ion mode; mass range, 200-2000 m/z; source voltage, 4.5 kV; capillary voltage, 31 V; sheat gas, 15 (arbitrary units); auxiliary gas, 2 (arbitrary units); capillary temperature, 250 °C; tube lens voltage, 95 V.

2.5.2. Analysis of fusion proteins and of their cleavage products

The separation of TB10.4, Ag85B and their fusion proteins was performed using a Symmetry C18 column (3.5 μm, 300 Å, 2.1 × 100 mm) from Waters (Milford, MA, USA). The mobile phase was composed of water (A) and ACN (B), both containing 0.1% TFA. Chromatographic conditions consisted of a linear gradient from 40% to 70% B in 10 min, followed by a linear gradient from 70% to 80% B in 3 min. The column temperature was maintained at 45 °C, the injection volume was 20

µL and elution was carried out at constant flow of 300 µL/min. In all experiments, proteins were detected at 214 nm.

UV calibration curves at five different concentrations were built in the range 1-200 µg/mL for TB10.4, Trx-TB10.4, Ag85B and Trx-Ag85B. Good linearity was assessed for all the four analytes (TB10.4: y = 77.3(6)x - 98(67), R^2=1.000; Trx-TB10.4: y = 73(2)x - 46(31), R^2=0.997; Ag85B: y = 50(3)x - 87(120), R^2=0.988; Trx-Ag85B: y = 19(1)x - 46(87), R^2=0.982).

Full scan intact MS experiments were carried out under the following instrumental conditions: positive ion mode; mass range, 700-2000 m/z; source voltage, 4.5 kV; capillary voltage, 35 V; sheat gas, 15 (arbitrary units); auxiliary gas, 2 (arbitrary units); capillary temperature, 220 °C; tube lens voltage, 140 V. Multiple-charged protein ion signals were deconvoluted by using Bioworks Browser (Thermo Fisher Scientific, revision 3.1) and the percent abundance of different species calculated by the relative abundance of corresponding peaks in the deconvoluted spectra. The accuracy of mass determination was calculated comparing the experimental value with the one calculated from the amino acid sequence by "Peptide Mass Calculator" on IonSource Mass Spectrometry Educational Resource (*www.ionsource.com*).

2.6. Cleavage of fusion proteins with immobilized rEK by recirculation through the IMER

Solutions of Trx-TB10.4 and Trx-Ag85B (100 µg/mL) in buffer were continuously recirculated through the bioreactor at 100 µL/min and 30 °C. At scheduled times, 60 µL of substrate solution were collected and injected in HPLC-UV or HPLC-MS. The TB10.4 and Ag85B products were quantified as described in paragraph 2.5.2 and the conversion percentage were calculated.

2.7. Storage stability

When not in use, the bioreactor was stored at 4 °C in a 0.01% (w/v) solution of sodium azide.

The storage stability of the rEK-IMER was determined by analysis of the synthetic peptide as described in paragraph 2.4.3.

3. Results and discussion

The aim of this study was the development of an active and stable monolithic silica chromatographic bioreactor based on rEK to be used in a chromatographic system for the cleavage of fusion proteins to produce mature recombinant therapeutic proteins.

3.1. Development and characterization of rEK-IMER

In the last 15 years, the application of monolithic supports for bioconversion processes has rapidly expanded, especially in the field of protein digestion. The development of flow-through IMERs is expected to consider the selection of appropriate solid matrix and immobilization approach to obtain an active and useful bioreactor [27].

Silica monolithic supports (Chromolith®) activated with epoxide functionalities were selected for the covalent immobilization of rEK in a one-step reaction, adapting an *in situ* procedure previously described by our group [23]. Enzyme immobilization *via* epoxy groups is a very convenient procedure when preparing bioreactors that can be integrated in HPLC systems, due to the highly stable multipoint covalent attachment of the enzyme on the chromatographic support [28-30]. Since the characterization in terms of amount of immobilized enzyme is very important, the cleavage activity of rEK was tested before and after the immobilization process using the standard substrate GD_4K-NA in order to estimate the rEK immobilization yield. A reverse phase (RP)-LC-UV method was developed to separate the product, 2NA, from the substrate (method details are reported in paragraph 2.5.1) and the identity of eluting peaks was confirmed by ESI-MS coupling. For quantitative measurements, a calibration curve of standard substrate GD_4K-NA was built in the range 1-200 µg/mL ($y = 82.90(8)x + 7(8)$, $R^2=1.000$). The cleavage yields were calculated by quantitative determinations of residual substrate.

The residual activity of post-immobilization rEK solution was determined under saturating conditions, and was found to be 12%, corresponding to 8.8 ± 1.0 (n=3) immobilized enzyme units (U).

It is well known that the immobilization procedure can change the enzyme activity. Thus, the cleavage performance of the rEK bioreactor was tested using the same standard substrate, GD_4K-NA. At first, a GD_4K-NA solution (20 µL at 1 mg/mL) was injected into the bioreactor at a flow rate of 100 µL/min. Under the explored experimental conditions, with an enzyme/substrate contact time of 6 min, no cleavage products were observed in the collected fractions. We ascribed this negative result to the short contact time between the immobilized enzyme and the substrate. Therefore, we considered a

different set up and the substrate solution (1.5 mL of GD_4K-NA at 100 µg/mL concentration) was continuously recirculated through the IMER with an incubation time of 24 h. The concentration of GD_4K-NA was monitored over time.

Since reaction conditions, such as temperature and contact time, can affect cleavage rate, two different temperatures and flow rates were studied in order to maximize the conversion yield. As reported in **Figure 1**, at 30 °C and with a flow rate of 400 µL/min, a complete cleavage of the GD_4K-NA solution (100% of conversion yield) was achieved in 24 hours. These experiments demonstrated that immobilized rEK maintains its activity in the immobilized form. Additionally, the specificity for the Asp_4-Lys sequence in this substrate conversion was also confirmed. The optimal experimental conditions were used in all the prepared rEK-bioreactors with GD_4K-NA as substrate.

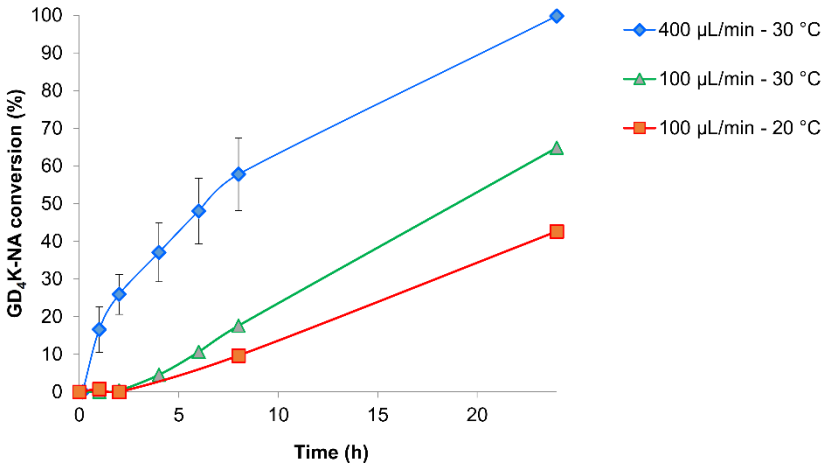


Figure 1. Conversion yields obtained for GD_4K-NA substrate solutions recirculated through the IMER in different conditions.

One of the purposes of the immobilization strategy is to increase the enzyme stability for its prolonged utilization. The stability of the rEK-IMER was monitored using the same recirculation set up to quantify the enzymatic activity over a defined time frame (3, 6 and 10 months). A reduction of the enzyme activity (33%) was observed after the first 3 months, but the IMER preserved 64% of its activity after 10 months and 155 working hours (**Figure 2**). These stability and reuse data are highly significant also considering that the soluble rEK diluted in storage buffer and preserved at -20 °C, as recommended by the supplier, maintains in the same period only 14% of its activity.

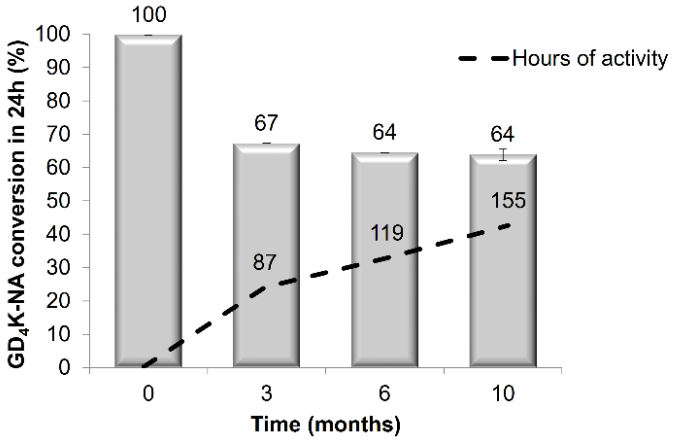


Figure 2. Stability evaluation of rEK IMER.

3.2. Cleavage reaction of fusion proteins

In order to investigate the catalytic activity of both free and immobilized rEK, the cleavage reaction of two recombinant proteins (Trx-TB10.4 and Trx-Ag85B) was considered. The two fusion proteins contain the same fusion partner composed by Trx sequence, to enhance protein solubility, a His_6-tag to allow protein purification, and a linker containing the specific cleavage sequence for EK (**Figure 3**). The cleavage of these constructs yields the two mature proteins, TB10.4 (11077 kDa) and Ag85B (31346 kDa) selected for the preparation of potential novel glycovaccines against tuberculosis [20].

Trx-TB10.4 (261 aa)

```
        10         20         30         40         50         60         70
MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QGKLTVAKLN IDQNPGTAPK
        80         90        100        110        120        130        140
YGIRGIPTLL LFKNGEVAAT KVGALSKGQL KEFLDANLAG SGSGHMHHHH HHSSGLVPRG SGMKETAAAK
       150      158          10         20         30         40         50
FERQHMDSPD LGTDDDDK   AMAISDPMSQ IMYNYPAMLG HAGDMAGYAG TLQSLGAEIA VEQAALQSAW
        60         70         80         90        100
QGDTGITYQA WQAQWNQAME DLVRAYHAMS STHEANTMAM MARDTAEAAK WGG
```

Trx-Ag85B (450 aa)

```
        10         20         30         40         50         60         70
MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QGKLTVAKLN IDQNPGTAPK
        80         90        100        110        120        130        140
YGIRGIPTLL LFKNGEVAAT KVGALSKGQL KEFLDANLAG SGSGHMHHHH HHSSGLVPRG SGMKETAAAK
       150      158          10         20         30         40         50
FERQHMDSPD LGTDDDDK   AMAISDPFSR PGLPVEYLQV PSPSMGRDIK VQFQSGGNNS PAVYLLDGLR
        60         70         80         90        100        110        120
AQDDYNGWDI NTPAFEWYYQ SGLSIVMPVG GQSSFYSDWY SPACGKAGCQ TYKWETFLTS ELPQWLSANR
       130        140        150        160        170        180        190
AVKPTGSAAI GLSMAGSSAM ILAAYHPQQF IYAGSLSALL DPSQGMGPSL IGLAMGDAGG YKAADMWGPS
       200        210        220        230        240        250        260
SDPAWERNDP TQQIPKLVAN NTRLWVYCGN GTPNELGGAN IPAEFLENFV RSSNLKFQDA YNAAGGHNAV
       270        280        290
FNFPPNGTHS WEYWGAQLNA MKGDLQSSLG AG
```

Figure 3. Amino acid sequences of Trx-TB10.4 and Trx-Ag85B. The fusion partner sequence is reported in black and the mature protein sequences are reported in red. The cleavage site (DDDDK) and the His_6-tag present in both fusion proteins are evidenced in bold.

To monitor the cleavage activity and specificity, a RP-HPLC-UV-MS method was developed to separate mature TB10.4 and Ag85B from their substrates. A 300 Å wide pore C18 column was selected to achieve proper chromatographic efficiency and mobile phases composed by water (A) and ACN (B) both containing 0.1% TFA were used to allow direct ESI-MS coupling. For both the considered analyte pairs, a linear gradient from 40% to 70% B in 10 min allowed the mature protein to elute well-resolved from the substrate peak. Individual calibration curves were built for Trx-Ag85B / Ag85B, and Trx-TB10.4 / TB10.4 pairs. Linearity and good correlation values were found for all the analytes.

3.2.1. In solution cleavage

Trx-TB10.4 and Trx-Ag85B were cleaved in solution according to the protocol reported by the rEK supplier. It is indeed extensively documented that rEK sometimes lacks of specificity in fusion protein digestion [31, 32]. The UV traces of both the digestion mixtures analyzed after two hours of incubation (**Figure 4c and 4f**) show the presence of a peak at the expected retention time for the mature protein. The deconvoluted MS spectrum from the analysis of the Trx-TB10.4 digestion mixture demonstrates the coelution of two species with mass of 11076 Da and 10776 Da and relative abundance of 85.5% and 11.5%, respectively. The most intense mass peak corresponds to the expected MW for mature TB10.4 (11077 Da). The less abundant one can be assigned to the truncated species A_1-K_{100} (10776 Da) generated by the non-specific rEK cleavage after the lysine residue in position 100. Differently, in the digestion of Trx-Ag85B, only one species with the expected mass of mature protein (31346 Da) was detected in Ag85B peak, and non-specific cleavages were not observed.

LC-MS analyses revealed also that in both Trx-TB10.4 and Trx-Ag85B digestions, the fusion partner Trx-His_6-tag was not resolved from the substrate peak. Modification in gradient slope did not allow to achieve the desired resolution. When necessary, for semi-quantitative measurements, the relative abundance of coeluting species was assessed by MS spectra assuming the same ionization efficiency and UV absorbance of the two species.

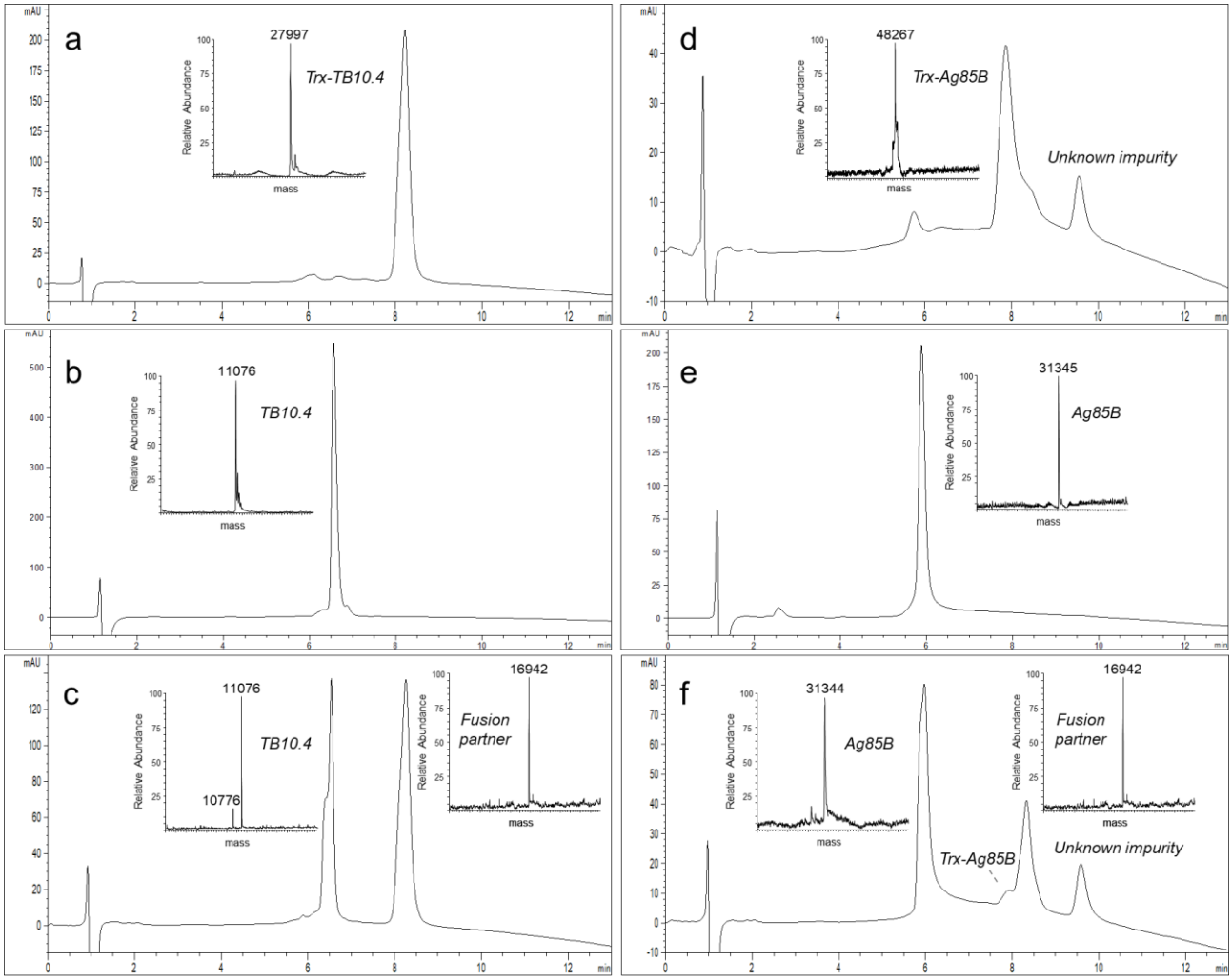


Figure 4. LC-UV traces of Trx-TB10.4 (a); TB10.4 (b); in solution digested Trx-TB10.4 (c); Trx-Ag85B (d); Ag85B (e); in solution digested Trx-Ag85B (f).

3.2.2. rEK-IMER activity evaluation

Reactions of the two fusion recombinant proteins with the rEK-IMER were carried out, cleavage yields were quantitatively evaluated by LC-UV and the cleavage specificity assessed by LC-MS.

A Trx-TB10.4 solution (1 mL of 100 µg/mL) was continuously flowed into the bioreactor in the same conditions used for the standard substrate GD_4K-NA (400 µL/min; 30 °C), monitoring the solution composition over time. A fast reduction of substrate amount was observed, which was not balanced by a coherent formation of digestion products. This behaviour was ascribed to the adsorption of the fusion protein on the support material due to non-specific interactions, possibly involving the poly-histidine tag. To contrast this phenomenon, two buffer modifiers (ACN and imidazole) were tested. Additives were used at the highest concentrations reported in the literature on soluble rEK, which were referred as not detrimental for the enzyme activity. With the addition of 10% ACN, the total recovered area (substrate and product) increase was not satisfactory. Differently, by adding 50 mM

imidazole to the buffer, the total expected area of both substrate and product was found after 60 min of reaction in recirculation conditions. These data confirm that the adsorption of the fusion protein on the monolithic material is mainly mediated by the presence of the His_6-tag sequence, being this interaction reversed by the presence of imidazole, with a competitive mechanism.

Due to the reduction of adsorption phenomena, digestion yield increased to 38% in one-hour cleavage, and imidazole was thus selected as buffer additive in all the subsequent experiments.

Since conversion percentage was still not satisfactory, we investigated a lower flow rate considering that the best enzyme/substrate contact time is strictly dependent on the nature and dimension of proteins, making the flow rate selected for a small dimension substrate as GD_4K-NA not necessarily optimal for higher dimension substrates, such as fusion proteins.

Trx-TB10.4 in the digestion buffer added with 50 mM imidazole was recirculated into the IMER at a flow rate of 100 µL/min. In these conditions, a satisfactory conversion yield (97%) was achieved in only 2 hours.

Trx-Ag85B was subjected to hydrolysis under the same experimental conditions used for Trx-TB10.4 and reaction degree was estimated after 2 hours. Even with this larger substrate, an almost quantitative conversion (97%) was achieved.

It is well known that the covalent immobilization of a protein on a solid matrix might induce some conformational changes as a consequence of the steric restrictions generated by the newly introduced covalent bonds between the protein surface amino acids and the functional groups of the support. When the macromolecule is an enzyme, such conformational changes might be detrimental for the catalytic properties or, even worse, affect the site-specificity of the catalytic pocket.

In the case of immobilized EK, the applications reported in literature for hydrolysis of fusion proteins monitored the success of the reaction only by SDS-PAGE [13-16]. Although proving the maintenance of a hydrolytic activity of EK after immobilization on the different supports, the site-specificity was not demonstrated. Indeed, products derived by adventitious cleavages might show small differences in molecular weight (*i.e.* few amino acid residues) that cannot be appreciated by SDS-PAGE.

In this work, all the digestion mixtures were analyzed by LC-MS in order to assess the identity of the released products. The LC-UV analysis of Trx-TB10.4 solutions containing 50 mM imidazole, recirculated at 100 µL/min for 2 h in the rEK-IMER revealed digestion products at a different retention time compared to the one expected for mature TB10.4, suggesting that different protein products might be obtained due to non-specific cleavage of the substrate. The average deconvoluted

spectrum obtained for this peak is reported in **Figure 5a**. The examination of the detected masses allowed to attribute these species to non-specific cleavages in proximity or inside the six histidine sequence of the fusion protein.

Under the same experimental conditions, also Trx-Ag85B was digested and site-specificity evaluated by LC-MS. As shown in **Figure 5b**, the UV trace revealed the presence of hydrolyzed products at the expected retention time for mature Ag85B. The deconvoluted spectrum shows the presence of the correct Ag85B product with a relative abundance of 11.5% (**Figure 5b**). The remaining components were identified as different proteoforms resulting from non-specific cleavages occurring in proximity of the specific site (**Figure 6**).

Given these results, rEK specificity seems to be influenced by the immobilization process. While with Trx-TB10.4 a complete loss of cleavage specificity was observed, in the case of Trx-Ag85B the obtained product is represented by a mixture of closely related proteoforms of the desired mature Ag85B.

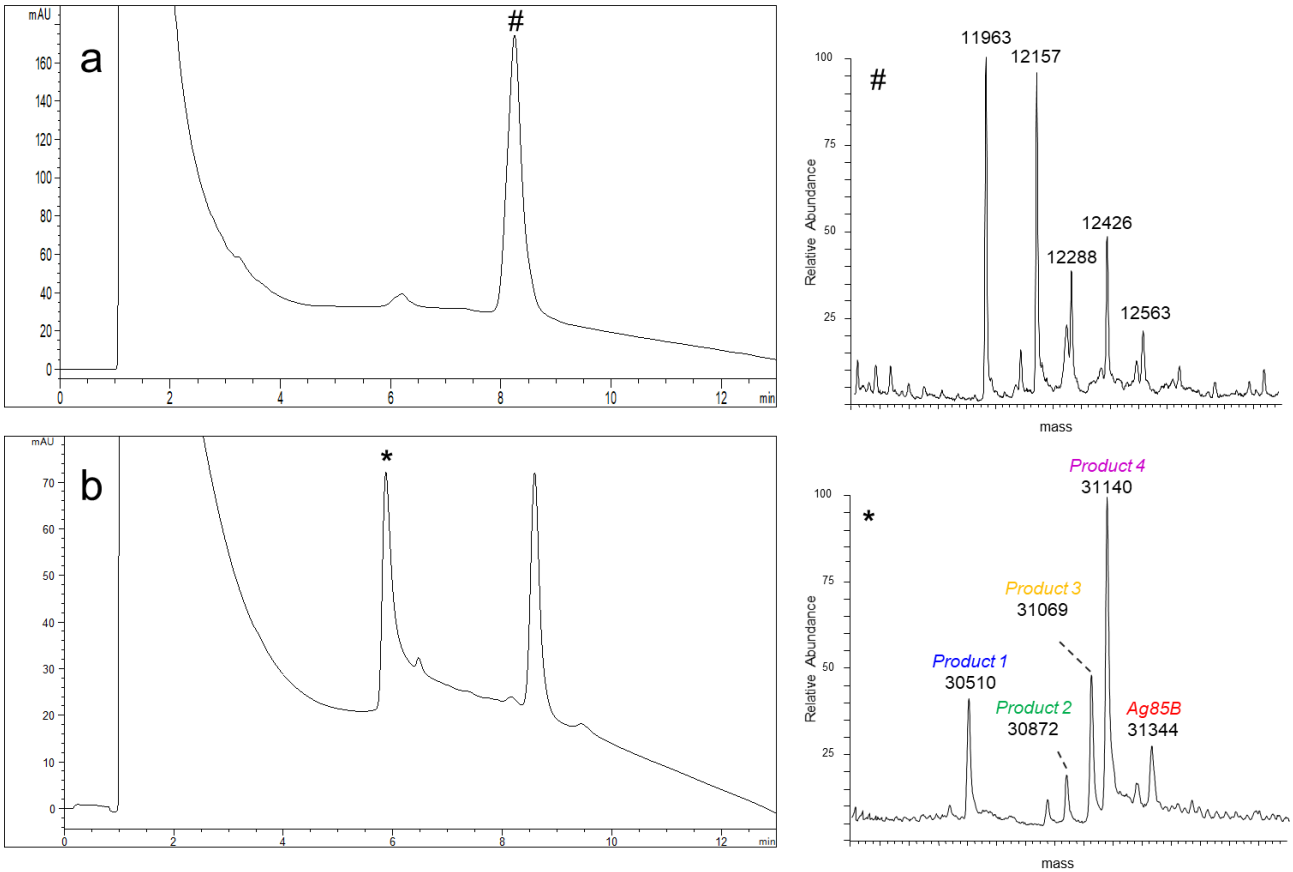


Figure 5. LC-UV traces of in-flow digested Trx-TB10.4 (a) and in-flow digested Trx-Ag85B (b). The deconvoluted spectra obtained at the apex of digestion product peaks are reported.

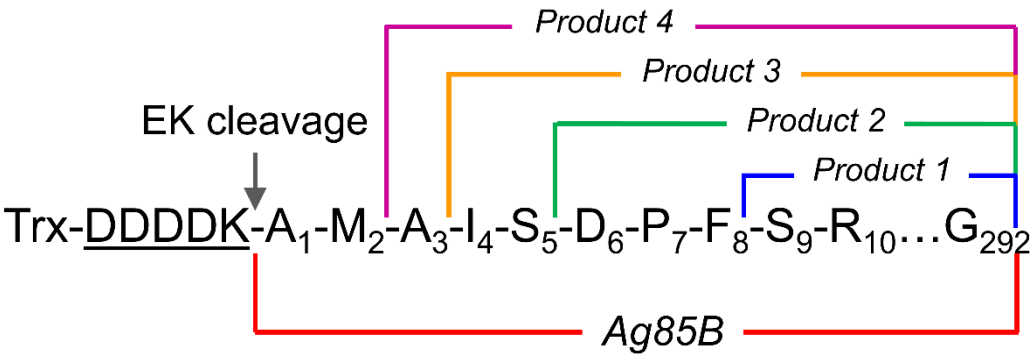


Figure 6. Identity assignment of components revealed in the in-flow digestion of Trx-Ag85B.

3.3. Strategies to improve the cleavage efficiency and specificity of rEK-IMER

Several attempts have been made to improve site-specificity of immobilized rEK, based on the current literature on IMERs [29, 33]. We first hypothesized that the presence of imidazole in the buffer could negatively affect enzyme specificity; to rule out this possibility the digestion products obtained with unmodified buffer were also analyzed in LC-MS, but no significant differences were observed. A second hypothesis was that the non-specific cleavage could be related to inaccessibility of the rEK to the canonical sequence during in-flow proteolysis. Since literature documents the positive effect that the presence of urea can provide on rEK specificity [12], digestion products obtained by adding to 2 M urea to fusion protein solutions were analyzed by HPLC-MS, but the same species were revealed in the deconvoluted spectrum of product peaks.

A further attempt to improve site-specificity of the rEK cleavage towards protein substrates consisted in the evaluation of wide pore monolithic silica for enzyme immobilization. New generation epoxy Chromolith® supports with 300 Å pores were obtained by Merck as research samples and were evaluated for rEK immobilization. The same immobilization protocol with 10 U of enzyme was applied (paragraph 2.4.1). A high immobilization yield was obtained (95%, corresponding to 9.5 U) with the preservation of the catalytic activity. However, the site-specificity in fusion protein digestion did not benefit of the higher support permeability under the considered conditions, since the digestion of both Trx-TB10.4 and Trx-Ag85B resulted in the same proteoforms, differing only in their relative abundance.

We also considered alternative immobilization protocols on monoliths with different functionalities. The nature of chemical bonds and support-enzyme spacers (*i.e.* length, hydrophobicity) can play a crucial role in preserving the optimal conformation of rEK upon its covalent immobilization. Longer spacers are expected to allow a higher conformational flexibility of the protein. This might be an advantage for enzymes undergoing significant conformational changes as they interact with the

substrate. On the other hand, shorter spacers can confer higher thermal stability since they restrict the enzyme mobility and prevent unfolding [34].

Two alternative immobilization protocols were exploited. 10U of rEK were immobilized on an epoxy support hydrolyzed to diol and oxidized to aldehyde for the covalent Schiff bases formation [22, 27], and on a propylamine Chromolith® support, functionalized with glutaraldehyde (longer spacer) [22]. Although excellent immobilization yields were obtained (100% in both cases), reduced activities on both standard peptide and fusion protein substrates were obtained, with no improvements in site-specificity in the cleavage of the Trx-His_6-tag proteins.

The type of enzyme (*i.e.* higher purity) used for the immobilization was also examined in order to obtain a site-specific bioreactor. 10 U of rEK from Merck (purity > 99%) were used to prepare a new rEK-IMER on a Chromolith® epoxy support. A 76% immobilization yield was obtained, and the bioreactor was applied to in-flow digestion of Trx-TB10.4 and Trx-Ag85B. In Trx-TB10.4 hydrolysis, the same site-specificity as the previously prepared bioreactors was observed, even if a more homogeneous product (mass of 12287 Da) was obtained. Interestingly, in the digestion of Trx-Ag85B, the use of this enzyme resulted in the formation of a single product, corresponding to the truncated form R_{10}-G_{292} (**Figure 7a**). It can be observed that this proteoform is similar to those obtained with immobilized rEK from Sino Biological, resulting from the non-specific cleavage after the serine residue in position 9. In addition to a higher product homogeneity, the quality of the released mature protein is highly comparable to the one obtained in solution. Indeed, the same rEK was also tested in solution on Trx-Ag85B, and the R_{10}-G_{292} proteoform was observed as one of the products together with the expected mature Ag85B (**Figure 7b**).

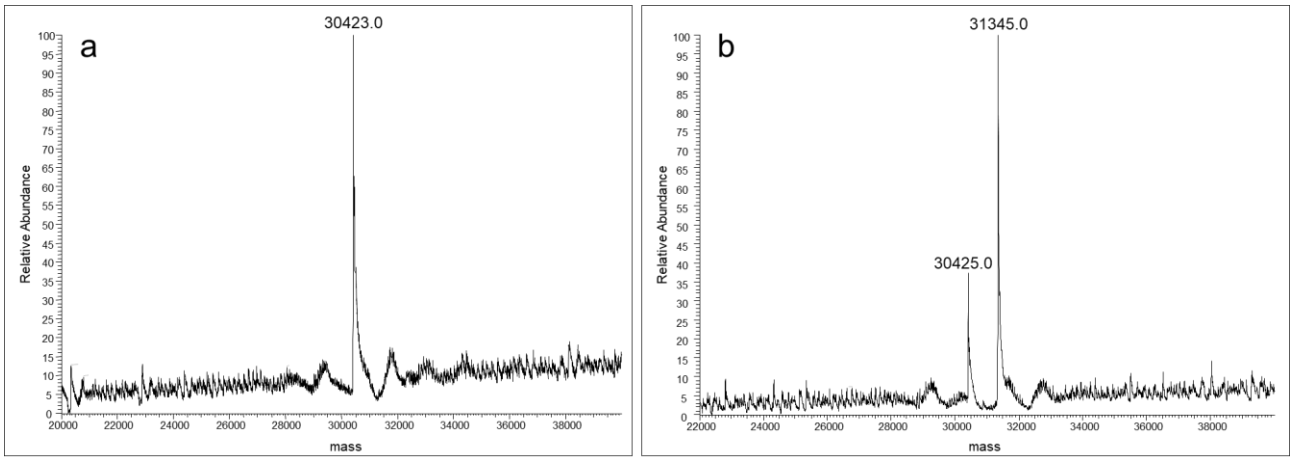


Figure 7. ESI-MS deconvoluted spectra obtained for the products generated from digestion of Trx-Ag85B carried out with immobilized (a) and in solution (b) rEK from Merck.

4. Conclusions

In this paper, a new rEK-IMER based on monolithic material has been developed. The rEK-IMER was designed to cleave fusion proteins avoiding the presence of residual traces of this protease in the final protein preparations.

The hydrolytic performances of the immobilized enzyme were first tested and optimized using the model substrate GD_4K-NA. Results demonstrated that the enzyme can be successfully immobilized on the monolithic chromatographic support with different chemistries, and the best cleavage activity was obtained with the enzyme immobilized on epoxy monolithic material. The rEK immobilization resulted in a significant increase in enzyme stability (more than 155 hours of activity in 10 months) compared to the free enzyme. rEK-IMERs were applied to the in-flow hydrolysis of two fusion proteins selected as models. A detailed study was carried out to minimize the non-specific binding of the substrate to the chromatographic support, and maximize the conversion yield. A cleavage yield of 97% was achieved in only 2 hours for both Trx-TB10.4 and Trx-Ag85B. Overall, the hydrolytic activity and reusability/stability data of immobilized rEK are highly satisfactory also in comparison to those reported in literature for both soluble and immobilized EK.

The site-specificity issue was deeply investigated by exploring several parameters during IMER synthesis (*i.e.* binding chemistry, support pore size, enzyme purity) and by defining the exact product identity by LC-MS. The obtained results indicate that the specificity of the rEK-IMER differs from the one of rEK in solution, and that neither the support permeability nor the explored immobilization chemistries can minimize this gap. Enzyme purity, on the other hand, can improve product quality. Enzyme with a lower purity produces less homogeneous products, while rEK with higher purity degree results in a single product proteoform. However, the experimental evidences suggest that the most critical aspect in the success of the on-line digestion is the nature of substrate construct, which seems to influence the quality of the obtained product stronger than the intrinsic enzyme site-specificity. It might be hypothesised the occurrence of additional factors possibly affecting the accessibility of the cleavage site during digestion carried out on rEK-IMER, which will be further investigated.

References

[1] LaVallie E.R. et al., *Thioredoxin and related proteins as multifunctional fusion tags for soluble expression in E. coli*, Methods Mol Biol. **2003**, 205, 119-140.

[2] Schmidt S.R., *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges*, first edition, Hoboken, New Jersey, Wiley **2013**.

[3] Schmidt S.R., *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges*, MAbs. **2015**, 7(3), 456-460.

[4] Young C.L. et al., *Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications*, Biotechnol J. **2012**, 7(5), 620-634.

[5] Arnau J. et al., *Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins*, Protein Expr Purif. **2006**, 48(1), 1-13.

[6] Bornhorst A., Falke J.J., *Purification of Proteins Using Polyhistidine Affinity Tags*, Methods Enzymol. **2000**, 326, 245-254.

[7] Waugh D.S., *An overview of enzymatic reagents for the removal of affinity tags*, Protein Expr Purif. **2011**, 80(2), 283-229.

[8] Gasparian M.E. et al., *Strategy for improvement of enteropeptidase efficiency in tag removal processes*, Protein Expr Purif. **2011**, 79(2), 191-196.

[9] Bricteux-Gregoire S. et al., *Phylogeny of trypsinogen activation peptides*, Comp Biochem Physiol B. **1972**, 42(1), 23-39.

[10] Light A., Fonseca P., *The preparation and properties of the catalytic subunit of bovine enterokinase*, J Biol Chem. **1984**, 259(21), 13195-13198.

[11] Collins-Racie L. A. et al., *Production of recombinant bovine enterokinase catalytic subunit in Escherichia coli using the novel secretory fusion partner DsbA*, Biotechnology (N Y). **1995**, 13(9), 982-987.

[12] Shahravan S.H. et al., *Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag*, Protein Expr Purif. **2008**, 59(2), 314-319.

[13] Suh C.W. et al., *Covalent immobilization and solid-phase refolding of enterokinase for fusion protein cleavage*, Process Biochem. **2005**, 40(5), 1755-1762.

[14] Kubitzki T. et al., *Application of immobilized bovine enterokinase in repetitive fusion protein cleavage for the production of mucin 1*, Biotechnol J. **2009**, 4(11), 1610-1618.

[15] Kubitzki T. et al., *Immobilisation of bovine enterokinase and application of the immobilised enzyme in fusion protein cleavage*, Bioprocess Biosyst Eng. **2008**, 31(3), 173-182.

[16] Santana S.D. et al., *Immobilization of enterokinase on magnetic supports for the cleavage of fusion proteins*, J Biotechnol. **2012**, 161(3), 378-382.

[17] Piubelli L. et al., *Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins*, Microb Cell Fact. **2013**, 12, 115.

[18] Temporini C. et al., *Liquid chromatography-mass spectrometry structural characterization of neo glycoproteins aiding the rational design and synthesis of a novel glycovaccine for protection against tuberculosis*, J Chromatogr A. **2014**, 1367, 57-67.

[19] Tengattini S. et al., *Monitoring antigenic protein integrity during glycoconjugate vaccine synthesis using capillary electrophoresis-mass spectrometry*, Anal Bioanal Chem. **2016**, 408(22), 6123-6132.

[20] Bavaro T. et al., *Glycosylation of recombinant antigenic proteins from Mycobacterium tuberculosis: in silico prediction of protein epitopes and ex vivo biological evaluation of new semi-synthetic glycoconjugates*, Molecules. **2017**, 22(7), 1081.

[21] Tengattini S. et al., *Hydrophilic interaction liquid chromatography-mass spectrometry as a new tool for the characterization of intact semi-synthetic glycoproteins*, Anal Chim Acta. **2017**, 981, 94-105.

[22] Rinaldi F. et al., *Application of a rapid HILIC-UV method for synthesis optimization and stability studies of immunogenic neo-glycoconjugates*, J Pharm Biomed Anal. **2017**, 144, 252-262.

[23] Temporini C. et al., *Optimization of a trypsin-bioreactor coupled with high-performance liquid chromatography-electrospray ionization tandem mass spectrometry for quality control of biotechnological drugs*, J Chromatogr A. **2006**, 1120(1-2), 121-131.

[24] Chromolith® WP 300 Epoxy HPLC columns, *Immobilization via Schiff base mechanism*, Merck KGaA, Darmstadt, Germany.

[25] Cattaneo G. et al., *Synthesis of adenine nucleosides by transglycosylation using two sequential nucleoside phosphorylase-based bioreactors coupled on-line to an HPLC system for reaction monitoring*, ChemCatChem. **2017**, 9(24), 4614-4620.

[26] Antonowicz I. et al., *The application of a new synthetic substrate to the determination of enteropeptidase in rat small intestine and human intestinal biopsies*, Clin Chim Acta. **1980**, 101(1), 69-76.

[27] Vlakh E.G., Tennikova T.B., *Flow-through immobilized enzyme reactors based on monoliths: I. Preparation of heterogeneous biocatalysts*, J Sep Sci. **2013**, 36(1), 110-127.

[28] Mateo C. et al., *Multifunctional epoxy supports: a new tool to improve the covalent immobilization of proteins. The promotion of physical adsorptions of proteins on the supports before their covalent linkage*, Biomacromolecules. **2000**, 1(4), 739-745.

[29] Barbosa O. et al., *Heterofunctional supports in enzyme immobilization: from traditional immobilization protocols to opportunities in tuning enzyme properties*, Biomacromolecules. **2013**, 14(8), 2433-2462.

[30] Mateo C. et al., *Improvement of enzyme activity, stability and selectivity via immobilization techniques*, Enzyme Microb Technol. **2007**, 40(6), 1451-1463.

[31] Liew O.W. et al., *Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites*, Protein Expr Purif. **2005**, 41(2), 332-340.

[32] Choi S.I. et al., *Recombinant enterokinase light chain with affinity tag: expression from Saccharomyces cerevisiae and its utilities in fusion protein technology*, Biotechnol Bioeng. **2001**, 75(6), 718-724.

[33] Hartmann M., Kostrov X., *Immobilization of enzymes on porous silicas - benefits and challenges*, Chem Soc Rev. **2013**, 42(15), 6277-6289.

[34] Hanefeld U. et al., *Understanding enzyme immobilisation*, Chem Soc Rev. **2009**, 38(2), 453-468.

4. *Rational design, preparation and structural characterization of a potential antitubercular glycoconjugate vaccine with improved immunogenicity*

Francesca Rinaldi, Sara Tengattini, Teodora Bavaro, Luciano Piubelli, Gabriella Massolini, Enrica Calleri, Loredano Pollegioni, Marco Terreni, Caterina Temporini

Draft

Rational design, preparation and structural characterization of a potential antitubercular glycoconjugate vaccine with improved immunogenicity

Francesca Rinaldi[a], Sara Tengattini[a], Teodora Bavaro[a], Luciano Piubelli[b], Gabriella Massolini[a], Enrica Calleri[a], Loredano Pollegioni[b], Marco Terreni[a], Caterina Temporini[a]

[a] *Department of Drug Sciences, University of Pavia, Viale Taramelli 12, 27100 Pavia, Italy*
[b] *Department of Biotechnology and Life Sciences, University of Insubria, Via Dunant 3, 21100 Varese, Italy and The Protein Factory Research Centre, Politecnico of Milano and University of Insubria, Via Mancinelli 7, 20131 Milano, Italy*

Abstract

Tuberculosis is the deadliest infectious disease in the world. The variable efficacy of the current treatments highlights the need for more effective agents against this disease. Our research group is actively working on the investigation of a potential antitubercular glycoconjugate vaccine starting from Ag85B, a potent protein antigen from *Mycobacterium tuberculosis*. After a preliminary assessment of its biological activity, structural modifications were rationally designed to improve its immunogenicity. This paper describes the characterization of the effects of the proposed variations on *neo*-glycoconjugate structure by intact mass spectrometry (MS) measurements and glycopeptides solid phase extraction-hydrophilic interaction liquid chromatography-MS/MS analyses.

Abbreviations

AraMan: Arabinose(1→6)Mannose
ACN: Acetonitrile
CID: Collision Induced Dissociation
DTT: Dithiothreitol
ESI: Electrospray ionization
HILIC: Hydrophilic Interaction Liquid Chromatography
IME: 2-Iminomethoxyethyl
LAM: Lipoarabinomannan
LC: Liquid Chromatography
LTQ: Linear Trap Quadrupole
Man: Mannose
ManMan: Mannose(1→6)Mannose
ManManMan: Mannose(1→6)Mannose(1→6)Mannose
MOPS: 3-(N-morpholino) propanesulfonic acid
MS: Mass Spectrometry

Mtb: *Mycobacterium tuberculosis*
MW: Molecular Weight
PGC: Porous Graphitized Carbon
rEK: Recombinant enterokinase
RPLC: Reversed Phase Liquid Chromatography
SDS-PAGE: Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis
SPE: Solid Phase Extraction
TFA: Trifluoroacetic acid
TB: Tuberculosis
Trx: Thioredoxin

1. Introduction

The use of glycoconjugate vaccines is a promising strategy to fight various pathologies, among which different infectious diseases and forms of cancer [1]. The efficacy of a carbohydrate-protein conjugate was proven for the first time in 1931, by the immunization of rabbits against Type III pneumococci [2]. Since that moment, several vaccines of this kind have been developed and those against *Haemophilus influenzae* type b, *Streptococcus pneumoniae* and *Neisseria meningitides* are the most widespread [1, 3, 4].

In general, a glycoconjugate vaccine is composed of a carbohydrate moiety able to stimulate the B cell immune response and a carrier protein or peptide inducing the T cell mediated immunity. The conjugation of glycans derived from bacterial surface polysaccharides to immunogenic carrier proteins, usually from the same pathogen, is today recognized as a valid approach to produce high-effective glycoconjugates able to induce a long-term protection against encapsulated bacteria [1, 5]. Typically, saccharide-protein conjugates are recognized by the B cell receptor specific for polysaccharide antigens and are internalized into the cells, where various proteases digest the protein carrier. The deriving peptide epitopes are exposed on the B cell surface and activate T cells, with a consequent stimulation of high-affinity IgG antibodies production. Thus, both antigenic protein/peptide and carbohydrate components contribute to vaccine immunogenicity [3].

The production of glycoconjugates can be laborious and, if not controlled, can lead to non-homogeneous products with different features in terms of pharmacokinetics and immunogenicity, hindering their clinical use. Therefore, glycoprotein therapeutics with clear structure-activity relationships are required [3, 4]. The characterization of *neo*-glycoconjugates needs to be performed at different levels, both structural and functional. The most important structural features to investigate are: identity of proteins, sugars and conjugates, purity, integrity and glycosylation profile (yield, glycoform composition and glycosylation sites). This is usually achieved by a combination of analytical techniques such as liquid chromatography (LC) and mass spectrometry (MS) in different

modalities and operating conditions [6-9]. The functional characterization of vaccines generally examines their mechanism of action, immunogenicity, immunostimulation, efficacy and safety by *in vitro*, *ex vivo* or *in vivo* assays [10].

This work is part of a wider project aimed at studying a new potential glycoconjugate vaccine against tuberculosis (TB), one of the main causes of death in the world [11]. The vaccine under investigation is composed of antigenic proteins from *Mycobacterium tuberculosis* (*Mtb*), namely TB10.4 and Ag85B, and saccharides of mannose. In previous phases of the project, a method for the production and purification of the two proteins [12], a procedure for the synthesis and activation of the saccharidic moieties [13] and an optimized conjugation reaction to obtain pure and homogeneous *neo*-glycoconjugates [6] were developed. However, the *ex vivo* immunological evaluation of the compounds revealed a reduction in the effectiveness of Ag85B glycoconjugates compared to the non-glycosylated antigen [14]. The reason of this decrease was attributed to the putative antigenicity of Lys30 and Lys282 (the two most reactive amino acids of Ag85B involved in the glycosylation), as predicted using a combination of computational methods [14].

This paper describes the examination of strategies for the rational design of glycoconjugates with an improved immunogenicity, considering both protein and saccharidic moieties. As regards to the carrier protein, conservative mutations of Lys30 and Lys282 to arginine residues were planned to direct the glycosylation towards non-antigenic amino acids, thus preserving immunogenicity of the constructs. The contribution of glycans was investigated by the conjugation of mono-, di- and tri-saccharides of mannose to Ag85B protein and its mutants. The elongation of the chain was evaluated in order to move towards polysaccharides mimicking lipoarabinomannan (LAM), an immuno-modulating lipoglycan constituting an essential component of the mycobacterial cell wall that has been shown to contribute to the pathogenesis of *Mtb* infections [15-17].

In this work, a structural characterization of the various compounds was carried out by different analytical techniques. Protein identity and purity, sugar activation yields and glycosylation profiles were defined by intact mass measurements in electrospray ionization (ESI)-linear ion trap MS. A capillary reverse phase (RP) LC-UV-MS/MS system was exploited to confirm protein sequences after proteolytic digestion by chymotrypsin. To characterize glycosylation sites in *neo*-glycoconjugates, chymotryptic glycopeptides were selectively enriched by on-line solid phase extraction (SPE) and subsequently separated and identified by hydrophilic interaction liquid chromatography (HILIC)-MS/MS.

2. Materials and methods

2.1. Reagents and chemicals

Acetic acid, acetonitrile (ACN), ammonium bicarbonate, benzamidine hydrochloride, α-chymotrypsin from bovine pancreas, dithiothreitol (DTT), methanol, NaCl, 3-(N-morpholino) propanesulfonic acid (MOPS), sodium tetraborate and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). All the employed reagents were of analytical grade.

Recombinant bovine enterokinase (rEK, purity > 99%) and the Cleavage Capture Kit for rEK removal were purchased from Merck KGaA (Darmstadt, Germany). One rEK unit is defined as the amount of enzyme that cleaves > 95% of 50 μg of a control protein (supplied by the producer) in 16 hours at room temperature and in 20 mM TrisHCl, 50 mM NaCl, 2 mM $CaCl_2$, pH 7.4.

Deionized water was obtained from a Milli-Q® Integral purification system (Merck KGaA, Darmstadt, Germany).

Sample desalting required the use of Amicon® Ultra-0.5 mL centrifugal filters (Merck KGaA, Darmstadt, Germany).

2.2. Production of recombinant Ag85B proteins (wild-type and variants)

The wild-type Ag85B antigen from *Mtb* was produced in a recombinant form in *Escherichia coli* BL21(DE3) as a chimeric protein bearing thioredoxin (Trx), a His_6-tag and a sequence recognized by enterokinase at its N-terminus. The mature protein was obtained through the cleavage of the fusion protein by rEK, which was subsequently removed from the solution by affinity-based capture on EKapture™ Agarose (Merck KGaA, Darmstadt, Germany). Ag85B and Trx derived from the cleavage were separated by metal-chelating chromatography (HiTrap Chelating, GE Healthcare, Piscataway, NJ, USA). The complete production and purification process is described in [12].

Single (K30R and K282R) and double (K30R/K282R) point variants of Ag85B were produced using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). Site-directed mutagenesis was carried out *via* a PCR reaction on wild-type (for K30R and K282R) or Ag85B-K30R (for K30R/K282R) cDNA cloned into pET32b expression plasmids. For each reaction, two complementary oligonucleotide primers containing the desired mutation, flanked by unmodified nucleotide sequence, were used (mutated bases are underlined):

1.1) Ag85B-K30R: 5' GGGTCGTGATATTA<u>G</u>AGTTCAGTTTCAGAGCGG 3';
1.2) Ag85B-K30R: 3' CCCAGCACTATAAT<u>C</u>TCAAGTCAAAGTCTCGCC 5';

2.1) Ag85B-K282R: 5' GCTGAATGCAATGAGAGGTGATCTGCAGAGC 3';
2.2) Ag85B-K282R: 3' CGACTTACGTTACTCTCCACTAGACGTCTCG 5'.

For the K30R/K282R double variant, the oligonucleotides for the production of Ag85B-K282R (2.1 and 2.2) were employed.

The presence of the desired mutations was confirmed by automated DNA sequencing.

The Trx-Ag85B variant proteins were produced as reported for the wild-type counterpart [12]. Briefly, *E. coli* BL21(DE3) cells transformed with the pET32b-Ag85B plasmids encoding for the different protein variants were grown in SB medium at 37 °C, added of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600nm}=2 and grown at 18 °C for additional 16 hours (overnight). Using these conditions, ~20% of the single point chimeric proteins were soluble *vs.* ~10% for the double variant. Trx-Ag85B variants were isolated by a single chromatographic step on a nickel-affinity column (HiTrap Chelating, GE Healthcare, Piscataway, NJ, USA). Maturation of the fusion proteins was obtained using 3 U of rEK per mg of Trx-Ag85B, at 20 °C for 16 hours, followed by a further HiTrap Chelating chromatography. The overall yields were: Ag85B-K30R, 1.5 mg pure protein/L fermentation broth; Ag85B-K282R, 1.25 mg pure protein/L fermentation broth; Ag85B-K30R/K282R, 0.9 mg pure protein/L fermentation broth. The proteins were finally collected in 20 mM MOPS, 0.4 M NaCl, pH 7.0 at different concentrations. Purity and conformation of the recombinant proteins used in this study were assessed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), MS and circular dicroism (CD), as described in [12, 18].

2.3. Synthesis of mannose-based saccharides

2-iminomethoxyethyl (IME) activated mono-, di- and tri-saccharides of mannose were synthesized following the procedure reported in [13]. Mannose-IME (Man-IME), mannose(1→6)mannose-IME (ManMan-IME), arabinose(1→6)mannose-IME (AraMan-IME) and mannose(1→6)mannose(1→6) mannose-IME (ManManMan-IME) carbohydrates were examined in this study. Products were characterized by nuclear magnetic resonance (NMR) [13] and MS.

2.4. Production of neo-glycoconjugates

Glycosylation of Ag85B and its variants was performed as previously described in [6] and [19], with minimal adjustments. Proteins were dissolved in 100 mM sodium tetraborate buffer, pH 9.5 to reach a final concentration of 5.5 mg/mL and the solution was mixed with IME-glycosides to a glycoside/protein molar ratio of 200/1. Benzamidine hydrochloride was added to a final concentration

of 1 mM to prevent the digestion by residual rEK derived from the protein production process. Reaction mixtures were vortexed for 1 min and incubated for 24 h at 25 °C under continuous stirring.

2.5. Intact MS measurements

Proteins, activated saccharides and glycoconjugates were analyzed in a linear trap quadrupole (LTQ)-MS with an ESI source controlled by X-calibur software 2.0.7 (Thermo Fisher Scientific, Waltham, MA, USA).

Before direct infusion in ESI-MS, protein and glycoprotein solutions were desalted by ultrafiltration using Amicon® Ultra centrifugal filters with a nominal molecular weight (MW) cut-off of 10 kDa and a 500 µL volume. Seven washing steps were performed at 13000 rcf and 4 °C for 10 minutes to transfer the (glyco)proteins in deionized water.

(Glyco)protein samples were prepared in water/ACN/acetic acid (49/50/1 v/v/v) to a final concentration of 300 µg/mL, while lyophilized saccharides were dissolved in methanol. Samples were introduced into the mass spectrometer with a syringe pump at a flow rate of 10 µL/min and full scan intact MS experiments were carried out under the following instrumental conditions: positive ion mode, mass range 700-2000 m/z or 150-1000 m/z, source voltage 4.5 or 4.6 kV, capillary voltage 35 or 36 V, sheath gas flow rate 15 or 6 (arbitrary units), auxiliary gas flow rate 2 or 1 (arbitrary units), capillary temperature 220 or 250 °C, tube lens voltage 140 or 80 V for (glyco)proteins and carbohydrates, respectively. Deconvolution of the (glyco)protein spectra was performed using Bioworks Browser (Thermo Fisher Scientific, revision 3.1) and the abundance of the different species was defined by the relative intensity of the corresponding peaks in the deconvoluted spectra. The accuracy of mass determination was calculated comparing the experimental value with the one calculated from the amino acid sequence by "Peptide Mass Calculator" on IonSource Mass Spectrometry Educational Resource (http://www.ionsource.com/). Identification of protein truncated forms was carried out by MS-NonSpecific database search program of ProteinProspector v 5.20.0 (http://prospector.ucsf.edu).

2.6. Chymotryptic digestion of (glyco)proteins

All the (glyco)proteins were enzymatically digested with chymotrypsin prior to LC-MS analyses. Briefly, 100 µL of 100 µM (glyco)protein solution in water were added with 90 µL of 100 mM ammonium bicarbonate, pH 8.5, and 10 µL of 100 mM DTT solution in 100 mM ammonium bicarbonate, pH 8.5. The solution was heated at 60 °C for 30 minutes for disulfide bridge reduction and then added of chymotrypsin to a final (glyco)protein/enzyme ratio of 50/1 (w/w). The reaction

mixture was incubated for 3 hour at 37 °C under continuous stirring and the reaction was stopped by adding 2.5% TFA (v/v).

2.7. Capillary RPLC-UV-MS peptide mapping

Peptide solutions deriving from the chymotriptic digestion of non-glycosylated proteins were filtered and diluted in water/ACN (95/5 v/v) to a final concentration of 8 µM before LC-UV analyses.

The separation of peptides was performed on a Dionex UltiMate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA, USA) equipped with an UltiMate 3000 Autosampler and an UltiMate 3000 Variable Wavelength Detector and controlled by Chromeleon software (version 6.8). The analytical column was a C18 Acclaim® PepMap RSLC (2 µm, 100 Å, 300 µm x 150 mm) from Dionex (Thermo Fisher Scientific, Waltham, MA, USA). The mobile phases consisted of water + 0.05% TFA (v/v) (solvent A) and water/ACN 20/80 + 0.04% TFA (all v/v) (solvent B). The gradient was from 4 to 55% B in 30 min with a flow rate of 4 µL/min.

UV detection was performed at a wavelength of 214 nm.

MS detection was carried out on a LTQ-MS with an ESI source controlled by X-calibur software 2.0.7 (Thermo Fisher Scientific, Waltham, MA, USA), under the following instrumental conditions: positive ion mode, source voltage 4.5 kV, capillary voltage 31 V, sheath gas flow rate 40 (arbitrary units), auxiliary gas flow rate 10 (arbitrary units), capillary temperature 250 °C, tube lens voltage 95 V. Full scan mass range was set up from 300 to 2000 Da. MS^2 and MS^3 spectra were obtained by collision induced dissociation (CID) with normalized collision energy of 35.0.

Data processing was performed using Bioworks Browser (Thermo Fisher Scientific, revision 3.1) by comparing experimental data with protein sequences in FASTA. Peptides selection was achieved using "Xcorr *vs.* Charge State filter" (+1: 1.50, +2: 2.00, +3: 2.50). The score provided by the software measures the quality of the match (higher scores correspond to better matches).

2.8. On-line SPE-HILIC-MS analysis of neo-glycopeptides

The *neo*-glycoprotein digestion mixtures were analyzed according to a method previously described in [6]. Briefly, glycopeptides were on-line extracted on a porous graphitized carbon (PGC) trap column Hypersil Hypercarb (4.6 x 50 mm) from Alltech Associates (Deerfield, IL, USA) by applying the following conditions: 10 min of desorption from the trap column with 80% solvent A (ACN + 0.05% v/v TFA) and 20% solvent B (water + 0.05% v/v TFA) at a flow rate of 100 µL/min. Glycopeptides separation was performed on a Dionex UltiMate 3000 HPLC system (Thermo Fisher

Scientific, Waltham, MA, USA) controlled by Chromeleon software, using a TSK-gel Amide-80 column (3 μm, 80 Å, 2 x 150 mm) from Tosoh Bioscience LLC (Montgomeryville, PA, USA). The gradient for chromatographic analyses was from 30% to 57% B in 11 min with a flow rate of 200 μL/min. The analytes were revealed by an ESI-LTQ-MS (Thermo Fisher Scientific, Waltham, MA, USA) controlled by X-calibur software. Mass spectra were generated in positive ion mode under the following instrumental conditions: source voltage 4.0 kV, capillary voltage 46 V, sheath gas flow rate 40 (arbitrary units), auxiliary gas flow rate 10 (arbitrary units), capillary temperature 250 °C and tube lens voltage 105 V. MS^2 and MS^3 spectra were obtained by CID with normalized collision energy of 35.0. Glycopeptides were identified on the basis of MS^2 and MS^3 spectra by Bioworks Browser (Thermo Fisher Scientific, revision 3.1). Only identifications with a Xcorr higher than 1 were considered and, to avoid false positives, MS^2 and MS^3 spectra of all the species recognized as glycopeptides were manually evaluated.

3. Results and discussion

3.1. Rational design and analytical characterization of Ag85B variants

In a previous phase of this project [14], the reduction in the immunogenic activity observed for mannosylated Ag85B compared to the non-glycosylated protein was ascribed to the putative antigenicity of Lys30 and Lys282. The characterization of Ag85B glycosylation sites and the use of a combination of *in silico* prediction methods pointed out that these two lysines are the most reactive residues and are involved in the formation of protein epitopes [14]. On the light of these data, in order to produce *neo*-glycoconjugates with higher immunogenicity, avoiding the glycosylation of antigenic amino acids was considered a mandatory issue to preserve epitope formation. Thus, conservative mutations (K30R, K282R and K30R/K282R) of Ag85B protein were planned. The substitution of lysine with arginine residues was considered to be chemically conservative (pKa of the ε amino-group of Lys = 10.54; pKa of the guanidine-side group of Arg = 12.48).

Before starting the production of the selected Ag85B variants, the effect of the planned modifications on protein antigenicity was evaluated *in silico* by analyzing the three-dimensional structure of the query protein (PDB 1F0N for Ag85B, mutations were manually inserted in the sequence) using two prediction systems already exploited in [14]:

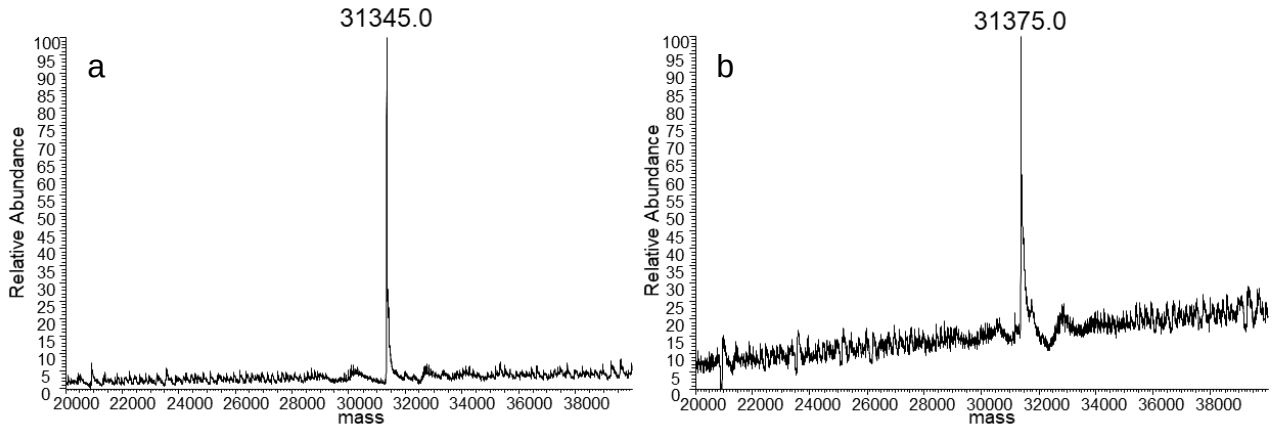
1) SEPPA, Spatial Epitope Prediction of Protein Antigens server [20] (http://lifecenter.sgst.cn/seppa/), a tool for conformational B cell epitope prediction that

attributes to each residue a score according to its neighborhood residues. Higher scores correspond to a higher probability for the residue to be part of an epitope (the default threshold value is set at 1.80);

2) EPCES, prediction of antigenic Epitopes on Protein surfaces by ConsEnsus Scoring server [21] (http://sysbio.unl.edu/EPCES/), which employs six different scoring functions (residue epitope propensity, conservation score, side-chain energy score, contact number, surface planarity score, and secondary structure composition).

Outputs indicate that K30R and K282R substitutions do not alter the original antigenicity of Ag85B. K30 residue was identified as antigenic only by SEPPA method, giving an antigenic prediction score of 1.83 that was not altered introducing K30R mutation (1.83). Similarly, K282 and R282 were predicted as antigenic residues by EPCES and gave comparable scores using SEPPA method (1.59 for lysine and 1.63 for arginine).

Two single point variants (K30R and K282R) and a double variant (K30R/K282R) of Ag85B protein were thus produced by site-directed mutagenesis and overexpressed in *E. coli* under the same conditions employed for the wild-type counterpart [12]. SDS-PAGE and amino acid sequencing were used to preliminary assess purity and identity of the products (data not shown). These two parameters were characterized more in detail by intact MS measurement of Ag85B mutants compared to the wild-type protein. Deconvoluted spectra are reported in **Figure 1 (a-d)**. For wild-type Ag85B and the two single point variants, the experimental mass corresponded to the expected MW of 31346.0 and 31374.0 Da, respectively (**Figure 1a, b** and **c**). The mass spectrum of the double mutant Ag85B-K30R/K282R (**Figure 1d**) showed a main peak at 31401.0 Da, which well agrees with the expected MW of 31402.0 Da. However, a second species was observed at 30480.0 Da with a relative abundance of around 20%. Based on its mass, this species was identified as the truncated by-product R_{10}-G_{292}, previously detected as impurity also in different batches of wild-type Ag85B.



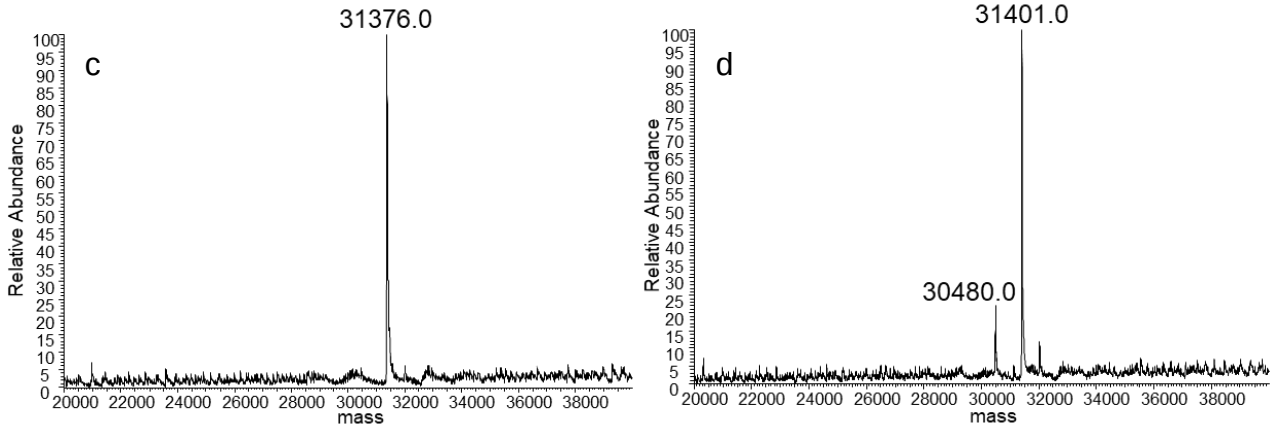


Figure 1. Deconvoluted ESI-MS spectra of purified Ag85B (a), Ag85B-K30R (b), Ag85B-K282R (c) and Ag85B-K30R/K282R (d).

In order to confirm protein sequences, the presence and position of the desired mutations, Ag85B and its mutants were digested with chymotrypsin and the resulting peptide mixtures were analyzed by capillary RPLC-UV-MS/MS. Results allowed to verify the identity of all the proteins, being the sequence coverage (% aa) over 75% in all cases (**Table 1**). Peptides containing each expected amino acid substitution were found in the corresponding variant sample, confirming the presence of the introduced mutations in the correct sequence position (**Table 2**). The extracted ion chromatograms of the m/z values corresponding to peptides comprising K30, K282, R30 and R282 provided an additional confirmation of the presence/absence of the mutations in all the analyzed proteins (data not shown).

Table 1. Summary results of the peptide mapping of Ag85B and its variants. Score values are calculated by the software and indicate the quality of the match between the experimental MS/MS data and the theoretical data.

Protein	Number of identified peptides	Protein sequence coverage (% aa)	Score
Ag85B	44	79.11%	611
Ag85B-K30R	49	75.00%	601
Ag85B-K282R	68	97.26%	1030
Ag85B-K30R/K282R	40	86.64%	540

Table 2. Identified peptides in the mappings of Ag85B and its mutants. Peptides including R30 are underlined in green, those with R282 are in light blue, while peptides comprising the unmodified K30 and K282 residues are in blue.

Position	$[M+H]^+$	Sequence	Ag85B	Ag85B-K30R	Ag85B-K282R	Ag85B-K30R/K282R
[1-8]	851.99	-.AMAISDPF.S	✓	✓	✓	✓
[1-17]	1851.12	-.AMAISDPFSRPGLPVEY.L			✓	✓
[3-8]	649.72	M.AISDPF.S	✓	✓		
[3-17]	1648.84	M.AISDPFSRPGLPVEY.L			✓	✓
[9-17]	1018.15	F.SRPGLPVEY.L	✓	✓	✓	✓
[18-24]	727.83	Y.LQVPSPS.M		✓		
[18-30]	1456.70	Y.LQVPSPSMGRDIR.V		✓		
[18-33]	1803.12	Y.LQVPSPSMGRDIKVQF.Q	✓		✓	
[18-33]	1831.13	Y.LQVPSPSMGRDIRVQF.Q		✓		✓
[18-44]	2878.21	Y.LQVPSPSMGRDIKVQFQSGGNNSPAVY.L			✓	
[19-33]	1689.96	L.QVPSPSMGRDIKVQF.Q	✓		✓	
[19-44]	2765.06	L.QVPSPSMGRDIKVQFQSGGNNSPAVY.L			✓	
[20-33]	1561.83	Q.VPSPSMGRDIKVQF.Q	✓		✓	
[25-33]	1094.31	S.MGRDIKVQF.Q	✓			
[34-44]	1094.12	F.QSGGNNSPAVY.L		✓		
[44-58]	1799.92	V.YLLDGLRAQDDYNGW.D		✓		
[45-55]	1279.38	Y.LLDGLRAQDDY.N		✓	✓	
[45-58]	1636.75	Y.LLDGLRAQDDYNGW.D	✓	✓	✓	✓
[46-58]	1523.59	L.LDGLRAQDDYNGW.D	✓	✓	✓	✓
[45-68]	2874.07	Y.LLDGLRAQDDYNGWDINTPAFEWY.Y			✓	
[46-65]	2282.41	L.LDGLRAQDDYNGWDINTPAF.E			✓	
[47-58]	1410.43	L.DGLRAQDDYNGW.D			✓	
[50-58]	1125.13	L.RAQDDYNGW.D			✓	
[56-68]	1613.71	Y.NGWDINTPAFEWY.Y			✓	
[58-68]	1256.35	W.DINTPAFEWY.Y				✓
[59-67]	1093.17	W.DINTPAFEW.Y	✓	✓	✓	✓
[59-68]	1256.35	W.DINTPAFEWY.Y	✓	✓	✓	
[59-73]	1804.94	W.DINTPAFEWYYQSGL.S			✓	
[68-86]	2084.34	W.YYQSGLSIVMPVGGQSSFY.S			✓	
[69-85]	1757.99	Y.YQSGLSIVMPVGGQSSF.Y			✓	✓
[69-86]	1921.16	Y.YQSGLSIVMPVGGQSSFY.S			✓	✓
[74-85]	1209.40	L.SIVMPVGGQSSF.Y		✓		
[74-86]	1372.57	L.SIVMPVGGQSSFY.S	✓	✓	✓	✓
[74-90]	1924.12	L.SIVMPVGGQSSFYSDWY.S			✓	
[78-86]	942.01	M.PVGGQSSFY.S			✓	✓
[86-102]	1901.07	F.YSDWYSPACGKAGCQTY.K			✓	
[87-102]	1737.90	Y.SDWYSPACGKAGCQTY.K	✓	✓	✓	✓
[90-102]	1349.52	W.YSPACGKAGCQTY.K			✓	
[91-102]	1186.35	Y.SPACGKAGCQTY.K		✓	✓	
[92-102]	1099.27	S.PACGKAGCQTY.K			✓	
[103-107]	710.80	Y.KWETF.L		✓		✓
[103-115]	1665.87	Y.KWETFLTSELPQW.L			✓	
[103-116]	1779.03	Y.KWETFLTSELPQWL.S	✓		✓	✓
[104-107]	710.80	Y.KWETF.L	✓			
[108-116]	1087.25	F.LTSELPQWL.S	✓	✓		
[117-132]	1513.73	L.SANRAVKPTGSAAIGL.S	✓	✓	✓	✓
[120-130]	1071.26	N.RAVKPTGSAAI.G		✓		
[120-132]	1241.47	N.RAVKPTGSAAIGL.S	✓	✓	✓	✓
[121-132]	1085.28	R.AVKPTGSAAIGL.S		✓		
[122-132]	1014.20	A.VKPTGSAAIGL.S		✓		
[141-150]	1188.36	M.ILAAYHPQQF.I	✓		✓	✓
[151-156]	623.72	F.IYAGSL.S		✓		
[143-150]	962.04	L.AAYHPQQF.I	✓	✓	✓	✓
[151-173]	2261.62	F.IYAGSLSALLDPSQGMGPSLIGL.A			✓	✓
[151-181]	2984.40	F.IYAGSLSALLDPSQGMGPSLIGLAMGDAGGY.K			✓	
[153-173]	1985.29	Y.AGSLSALLDPSQGMGPSLIGL.A			✓	✓
[153-181]	2708.06	Y.AGSLSALLDPSQGMGPSLIGLAMGDAGGY.K			✓	
[153-184]	2978.39	Y.AGSLSALLDPSQGMGPSLIGLAMGDAGGYKAA.D			✓	
[157-173]	1656.93	L.SALLDPSQGMGPSLIGL.A	✓	✓	✓	✓
[157-181]	2379.70	L.SALLDPSQGMGPSLIGLAMGDAGGY.K			✓	✓
[158-173]	1569.85	S.ALLDPSQGMGPSLIGL.A		✓		
[160-173]	1385.61	L.LDPSQGMGPSLIGL.A	✓	✓		
[160-181]	2108.38	L.LDPSQGMGPSLIGLAMGDAGGY.K			✓	
[161-173]	1272.45	L.DPSQGMGPSLIGL.A		✓		
[161-181]	1995.22	L.DPSQGMGPSLIGLAMGDAGGY.K				
[182-187]	721.85	Y.KAADMW.G	✓	✓		
[182-206]	2827.08	Y.KAADMWGPSSDPAWERNDPTQQIPK.L	✓	✓	✓	✓
[182-207]	2940.24	Y.KAADMWGPSSDPAWERNDPTQQIPKL.V	✓	✓		✓
[182-210]	3224.55	Y.KAADMWGPSSDPAWERNDPTQQIPKLVAN.N	✓		✓	
[182-211]	3338.66	Y.KAADMWGPSSDPAWERNDPTQQIPKLVANN.T	✓			✓
[182-212]	3439.76	Y.KAADMWGPSSDPAWERNDPTQQIPKLVANNT.R	✓		✓	✓

[182-215]	3895.32	Y.KAADMWGPSSDPAWERNDPTQQIPKLVANNTRLW.V			✓	✓
[188-206]	2124.26	W.GPSSDPAWERNDPTQQIPK.L	✓	✓		
[188-207]	2237.42	W.GPSSDPAWERNDPTQQIPKL.V	✓	✓		
[188-208]	2336.55	W.GPSSDPAWERNDPTQQIPKLV.A	✓	✓		
[188-210]	2521.73	W.GPSSDPAWERNDPTQQIPKLVAN.N	✓	✓		
[188-211]	2635.83	W.GPSSDPAWERNDPTQQIPKLVANN.T	✓	✓		
[188-212]	2736.94	W.GPSSDPAWERNDPTQQIPKLVANNT.R	✓	✓		✓
[188-215]	3192.49	W.GPSSDPAWERNDPTQQIPKLVANNTRLW.V	✓		✓	✓
[196-206]	1326.44	W.ERNDPTQQIPK.L		✓		
[196-215]	2394.68	W.ERNDPTQQIPKLVANNTRLW.V			✓	✓
[207-212]	631.70	K.LVANNT.R		✓		
[207-215]	1087.26	K.LVANNTRLW.V	✓		✓	
[216-235]	2024.20	W.VYCGNGTPNELGGANIPAEF.L			✓	
[216-239]	2527.75	W.VYCGNGTPNELGGANIPAEFLENF.V			✓	
[216-245]	3184.49	W.VYCGNGTPNELGGANIPAEFLENFVRSSNL.K			✓	
[218-235]	1761.90	Y.CGNGTPNELGGANIPAEF.L			✓	
[218-239]	2265.45	Y.CGNGTPNELGGANIPAEFLENF.V			✓	
[236-245]	1179.31	F.LENFVRSSNL.K	✓		✓	✓
[238-245]	937.04	E.NFVRSSNL.K	✓			
[246-251]	771.84	L.KFQDAY.N		✓		
[246-261]	1710.83	L.KFQDAYNAAGGHNAVF.N	✓		✓	✓
[252-261]	958.01	Y.NAAGGHNAVF.N	✓	✓	✓	✓
[262-274]	1635.72	F.NFPPNGTHSWEYW.G			✓	
[264-271]	895.94	F.PPNGTHSW.E	✓		✓	
[272-278]	866.94	W.EYWGAQL.N	✓		✓	
[275-289]	1561.75	W.GAQLNAMRGDLQSSL.G			✓	
[275-292]	1746.93	W.GAQLNAMRGDLQSSLGAG.-			✓	✓
[279-289]	1192.33	L.NAMRGDLQSSL.G			✓	✓
[279-291]	1164.32	L.NAMKGDLQSSL.G		✓		
[279-292]	1349.50	L.NAMKGDLQSSLGAG.-	✓	✓		
[279-292]	1377.51	L.NAMRGDLQSSLGAG.-			✓	✓
[280-291]	1050.21	N.AMKGDLQSSL.G		✓		
[280-292]	1235.39	N.AMKGDLQSSLGAG.-	✓	✓		
[280-292]	1263.41	N.AMRGDLQSSLGAG.-			✓	✓
[282-292]	1033.12	M.KGDLQSSLGAG.-		✓		

3.2. Synthesis of activated saccharides assisted by ESI-MS analysis

Another strategy applied to improve the glycoconjugate vaccine effectiveness was the amplification of glycans contribution to the overall immunogenicity by the elongation of their chain. The final aim is to synthesize more complex polysaccharides able to mimic LAM, an important saccharidic superficial antigen of *Mtb* [15-17].

The effect of chain extension was first evaluated comparing mono-, di- and tri-saccharides of mannose (**Figure 2**), which were synthesized as described in [13] and activated with the IME functional group. This kind of activation was chosen after the demonstration of its high selectivity towards lysine residues of the carrier protein [6].

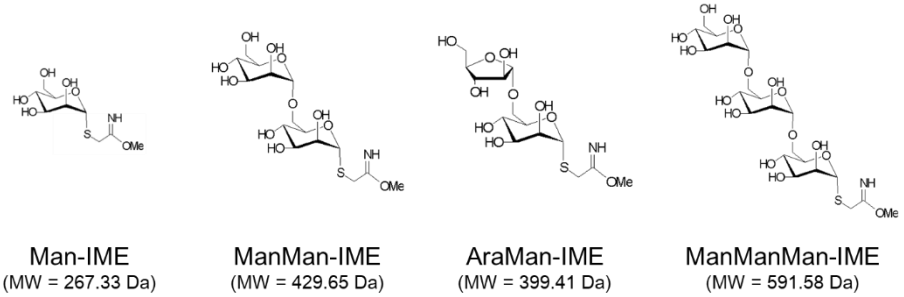


Figure 2. IME-activated saccharides considered in the study.

Direct infusions of the various carbohydrates in ESI-MS allowed to define their purity and activation yields before the conjugation with Ag85B and its variants. The purity of the activated products was found to decrease with the elongation of the chain (**Figure 3**, **a-d**), as previously reported in literature [22].

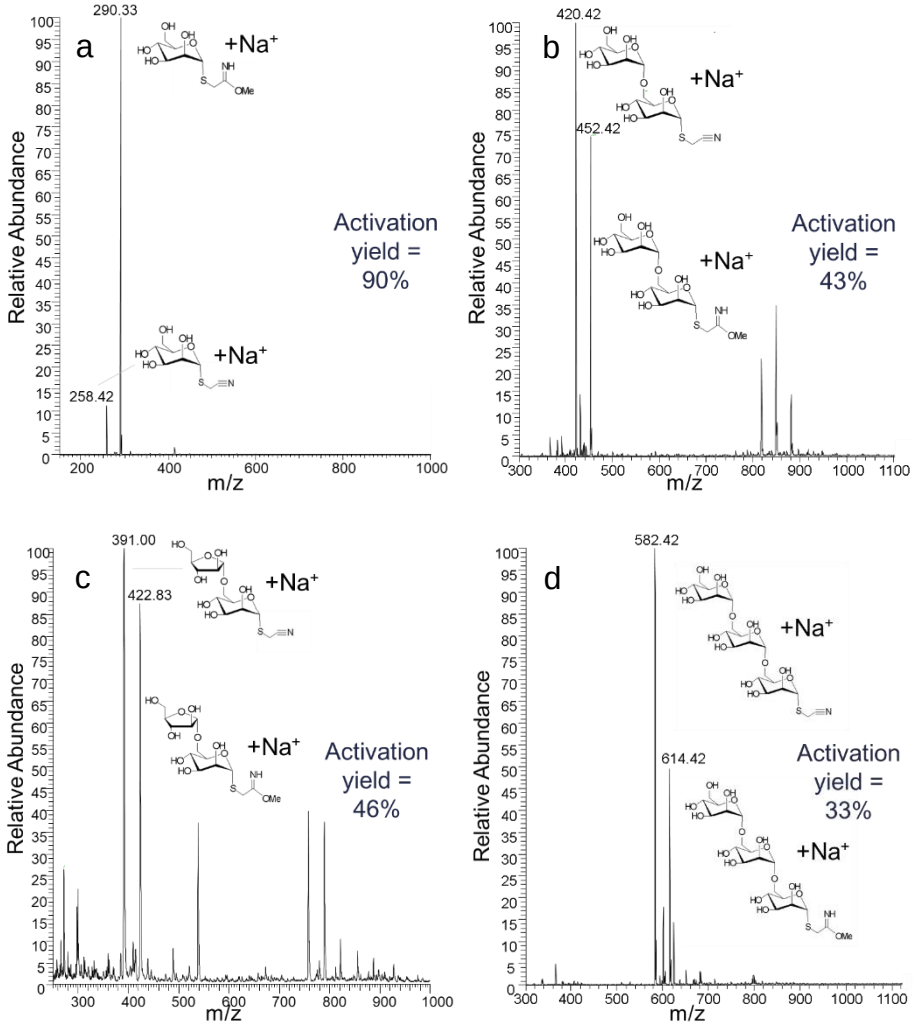


Figure 3. ESI-MS spectra of Man-IME (a), ManMan-IME (b), AraMan-IME (c) and ManManMan-IME (d) glycosides employed in this study.

3.3. Production and characterization of neo-glycoconjugates

Ag85B and its single (K30R and K282R) and double (K30R/K282R) variants were glycosylated with all the carbohydrate moieties under previously optimized experimental conditions [6] with minimal adjustments (see paragraph 2.4), in order to assess lysine and glycan reactivity and to characterize the obtained *neo*-glycoconjugates before their *in vitro* and *ex vivo* immunological evaluation.

Initially, lysine reactivity was evaluated using the conjugation with Man-IME as a model. Intact MS measurements of the purified reaction mixtures revealed reaction yields (Y) of 100% for all the

coupling reactions, as unmodified proteins were not detected. The % relative abundance of each glycoform was calculated as the ratio between the relative abundance indicated in the deconvoluted spectrum and the total ion intensities of the pattern (**Table 3**). The maximum number of incorporated mannose units corresponded to the total number of lysine residues in each protein sequence (8 for Ag85B, 7 for K30R and K282R single mutants and 6 for K30R/K282R double variant), indicating that all lysine residues are involved in the glycosylation. Considering the average number of bound glycosides as an indicator of the protein surface reactivity, it can be deduced that K30 is more reactive than K282, as the absence of this residue led to a more significant reduction of bound mannose unit loading (from 5.9 to 5.0 unit *per* protein) than mutation in position 282 (from 5.9 to 5.2 unit *per* protein).

Table 3. Glycoform distribution (% relative abundance), glycosylation yield (%) and Man/protein ratio (mol/mol) calculated after direct infusion of each purified reaction mixture in ESI-MS. Mean values for duplicate analysis are reported.

Protein	**Number of incorporated Man units (%)**									**Y (%)**	**Man/protein (mol/mol)**
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**		
Ag85B	-	-	-	1.2	9.7	25.7	34.9	21.9	6.7	100.0	5.9
Ag85B-K30R	-	-	-	9.4	24.5	34.4	24.5	7.3	-	100.0	5.0
Ag85B-K282R	-	-	-	5.6	20.3	35.4	29.1	9.7	-	100.0	5.2
Ag85B-K30R/K282R	-	-	9.2	26.9	37.4	22.0	4.4	-	-	100.0	3.8

A detailed characterization of glycosylation sites was performed through on-line SPE-HILIC-MS/MS analyses following the chymotriptic digestion of the various *neo*-glycoconjugates. The relative abundance of each glycosylation site was calculated as % area of all the relative glycopeptides (**Figure 4**).

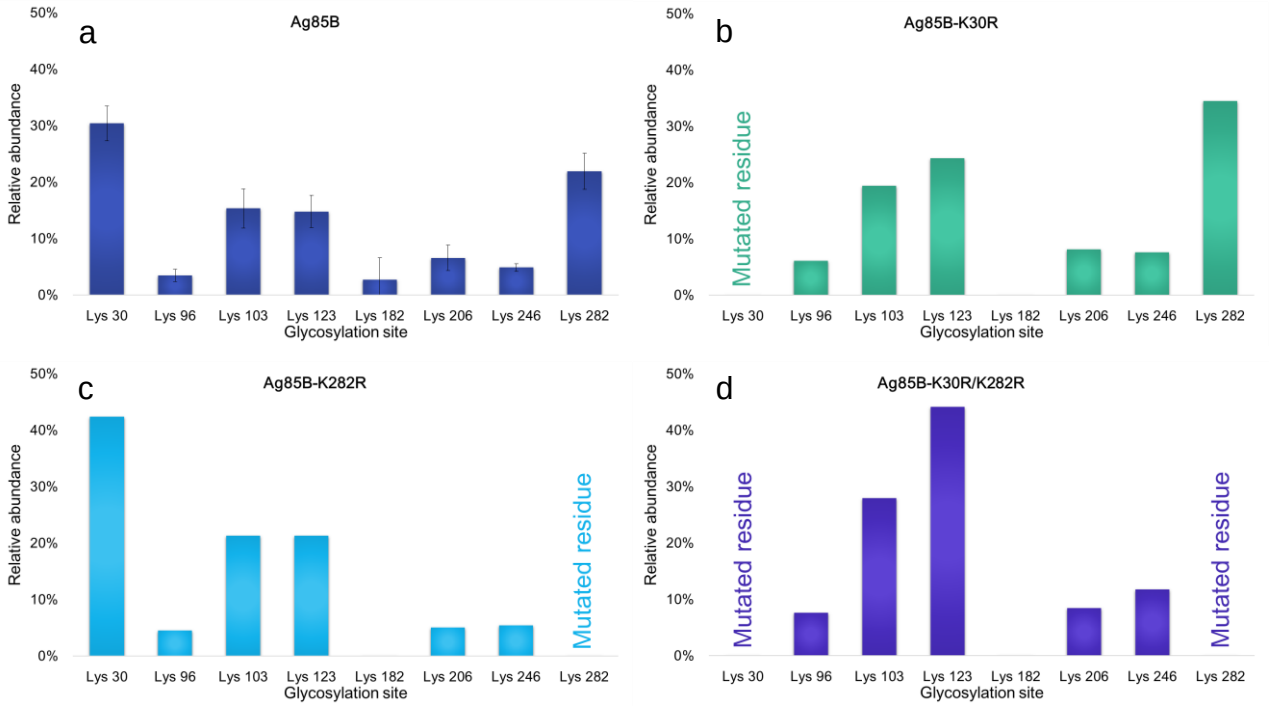


Figure 4. Lysine reactivity (% relative abundance) in Ag85B (a), Ag85B-K30R (b), Ag85B-K282R (c) and Ag85B-K30R/K282R (d). For wild-type Ag85B, mean values for duplicate analyses are shown (bars indicate variability).

In accordance with previously reported data [14], the main glycosylated lysines in wild-type Ag85B were K30 and K282. No glycosylated peptides corresponding to the mutated lysines were detected in the other samples, confirming that mutations were properly introduced and effectively protected the residues from the conjugation reaction. Amino acid reactivity reflected the absence of the silenced lysines, with a resulting predominant glycosylation of K282 in Ag85B-K30R and of K30 in Ag85B-K282R. In the variant Ag85B containing both amino acid substitutions, the major glycosylated site was K123, which was thus recognized as the most reactive lysine after K30 and K282. Moreover, the relative abundance order was generally respected in the four samples (**Figure 4**). Minimal variations in K206-K246 and K123-K103 relative orders were observed, since abundances of the sites of each couple were very close.

The absolute abundances of each glycosylation site, calculated as the sum of mass peak areas of all the relative glycopeptides, are comparable between wild-type and mutated proteins (**Figure 5**) although the relative abundance of each occupied site is significantly different between the various samples (**Figure 4**). This finding suggests that the chemical reactivity of the individual lysines remains unaltered even in the absence of the most reactive groups and consequently the coupling reaction does not possess a competitive behaviour. Thus, the substitution of the most reactive lysines does not induce alterations in the glycosylation of less reactive ones, further supporting the success of the mutagenesis approach.

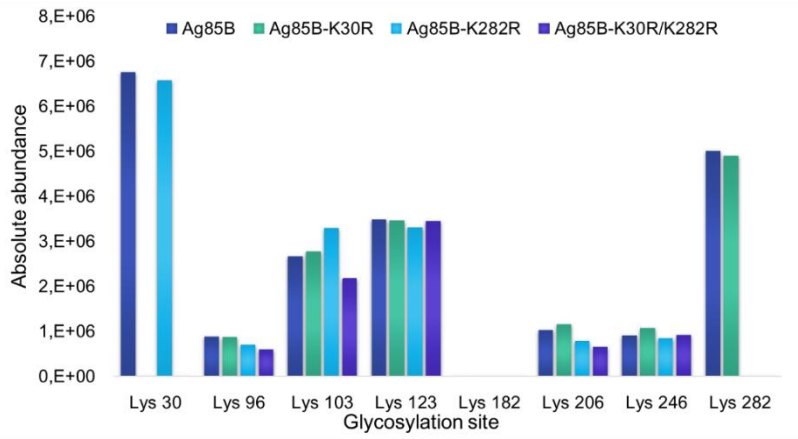


Figure 5. Comparison of glycosylated lysines absolute abundances (mass peak areas) between Ag85B and its mutants.

After the deep investigation of amino acid reactivity, the effects of the glycan type and length on the conjugation reaction (yield and position) were assessed. Since data suggested the comparability between the surface reactivity of Ag85B and its variants, the characterization of glycosylation sites using the various carbohydrates (Man-IME, ManMan-IME, AraMan-IME and ManManMan-IME) was carried out on wild-type Ag85B as a model. Conjugation reactions were performed considering the purity of each activated saccharide and following the same procedure used for Man-IME. Results of the glycopeptide mapping by SPE-HILIC-MS/MS (**Figure 6**) highlight that lysines 30, 282, 103 and 123 are the most reactive with all the tested glycans. Data show that the binding to Lys206 decreases with the elongation of the sugar chain. Taking into consideration the three-dimensional structure of Ag85B protein (PDB 1F0N), we propose that the biggest glycan moieties amplify the impact of the poor accessibility of K206 side chain, mainly due to steric effects.

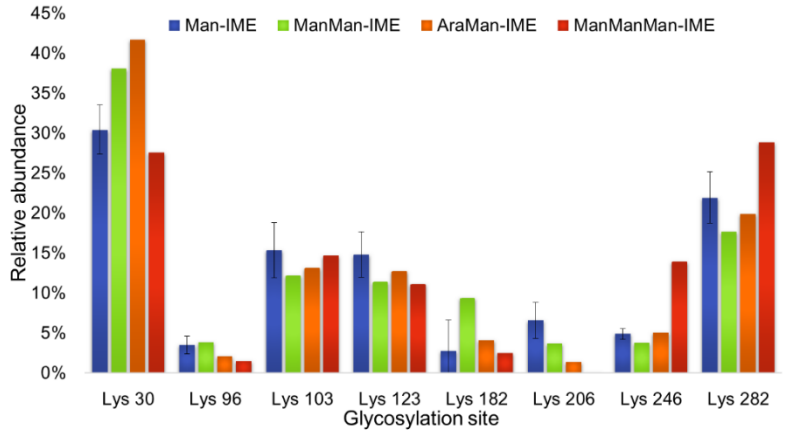


Figure 6. Lysine reactivity (% relative abundance) in Ag85B protein conjugated with Man-IME (blue), ManMan-IME (green), AraMan-IME (orange) and ManManMan-IME (red).

Glycan reactivity was then derived after the analysis of intact *neo*-glycoconjugates in ESI-MS. For the wild-type protein, the conjugation with different saccharidic moieties resulted in a change of the glycoform profile. Even if glycopeptides comprising each lysine residue were found in all the HILIC-

MS/MS analyses, intact MS measurements revealed a reduction of the overall glycan loading with the extension of the chain (**Table 4**). This observation can be ascribed to a steric hindrance effect.

Table 4. Glycoform distribution (% relative abundance), glycosylation yield (%) and glycoside/protein ratio (mol/mol) calculated for wild-type Ag85B conjugated with various saccharidic moieties.

Glycoconjugate	Number of incorporated glycosides (%)									Y (%)	Glycoside/protein (mol/mol)
	0	1	2	3	4	5	6	7	8		
Ag85B + Man-IME	-	-	-	1.2	9.7	25.7	34.9	21.9	6.7	100.0	5.9
Ag85B + ManMan-IME	-	-	11.7	25.1	31.4	21.9	9.8	-	-	100.0	3.9
Ag85B + AraMan-IME	-	4.8	19.8	25.4	22.5	18.5	9.0	-	-	100.0	3.6
Ag85B + ManManMan-IME	-	8.2	13.9	28.5	29.1	20.3	-	-	-	100.0	3.4

In order to assess if protein variants were equally influenced in the change of the glycosylation profile, Ag85B variants were conjugated with ManMan-IME and analyzed in ESI-MS by direct infusion. In glycan-mutant glycoconjugates, the differences between mono- and di-saccharides are more evident, probably amplified by the silencing of the two most reactive residues (**Table 5**). The reduction of the glycoside/protein ratio between Man-IME and ManMan-IME was indeed 34% for the wild-type protein, 38% for K30R mutant, 39% for Ag85B-K282R and 47% for the double variant.

Table 5. Glycoform distribution (% relative abundance), glycosylation yield (%) and glycoside/protein ratio (mol/mol) calculated for Ag85B variants conjugated with Man-IME and ManMan-IME.

Glycoconjugate	Number of incorporated glycosides (%)									Y (%)	Glycoside/ protein (mol/mol)
	0	1	2	3	4	5	6	7	8		
Ag85B-K30R + Man-IME	-	-	-	9.4	24.5	34.4	24.5	7.3	-	100.0	5.0
Ag85B-K30R + ManMan-IME	-	9.6	20.4	30.0	25.2	14.8	-	-	-	100.0	3.1
Ag85B-K282R + Man-IME	-	-	-	5.6	20.3	35.4	29.1	9.7	-	100.0	5.2
Ag85B-K282R + ManMan-IME	-	9.2	18.0	28.7	28.4	15.7	-	-	-	100.0	3.2
Ag85B-K30R/K282R + Man-IME	-	-	9.2	26.9	37.4	22.0	4.4	-	-	100.0	3.8
Ag85B-K30R/K282R + ManMan-IME	8.5	25.5	32.8	22.3	11.0	-	-	-	-	91.5	2.0

4. Conclusions

In this work, two strategies to improve the immunogenicity of a potential glycoconjugate vaccine against tuberculosis have been proposed: the mutation of epitope amino acids to prevent their masking by glycosylation and the use of more complex carbohydrate moieties to mimic LAM (an important saccharidic superficial antigen of *Mtb*).

After a first characterization of all the compounds, the structural effects of the introduced variations were deeply investigated by defining the glycosylation profile and the protein surface reactivity for each coupling reaction through MS and LC-MS/MS techniques.

Data allowed to verify the success of the mutagenesis, since amino acid substitutions were identified in the predicted position and efficiently avoided the glycosylation of these residues. Moreover, the introduced mutations did not alter the glycosylation degree of the remaining lysine residues.

The comparison of results obtained using mono-, di- and tri-saccharides of mannose for the conjugation suggested that protein surface reactivity is not influenced by the change in the sugar moiety. However, the glycoform profiles moved towards lower values, with less glycosides incorporated in the glycoconjugates. Coherently, this effect was more evident for Ag85B single and double point variants.

The structural characterization of the products and the confirmation of the occurrence of the planned modifications lay the basis for *in vitro* and *ex vivo* immunological assays. In this way, the effectiveness of the two strategies will be assessed.

Acknowledgments

This work was funded by Fondazione Banca del Monte di Lombardia (FBML, Italy). Luciano Piubelli and Loredano Pollegioni were supported from Fondo di Ateneo per la Ricerca (University of Insubria).

References

[1] Astronomo R.D., Burton D.R., *Carbohydrate vaccines: developing sweet solutions to sticky situations?*, Nat Rev Drug Discov. **2010**, 9(4), 308-324.

[2] Avery O.T., Goebel W.F., *Chemo-immunological studies on conjugated carbohydrate-proteins*, J Exp Med. **1931**, 54(3), 437-447.

[3] Adamo R. et al., *Synthetically defined glycoprotein vaccines: current status and future directions*, Chem Sci. **2013**, 4(8), 2995-3008.

[4] Vella M., Pace D., *Glycoconjugate vaccines: an update*, Expert Opin Biol Ther. **2015**, 15(4), 529-546.

[5] Campodónico V.L. et al., *Efficacy of a conjugate vaccine containing polymannuronic acid and flagellin against experimental Pseudomonas aeruginosa lung infection in mice*, Infect Immun. **2011**, 79(8), 3455-3464.

[6] Temporini C. et al., *Liquid chromatography-mass spectrometry structural characterization of neo glycoproteins aiding the rational design and synthesis of a novel glycovaccine for protection against tuberculosis*, J Chromatogr A. **2014**, 1367, 57-67.

[7] Staub A. et al., *Intact protein analysis in the biopharmaceutical field*, J Pharm Biomed Anal. **2011**, 55(4), 810-822.

[8] Sandra K. et al., *Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization*, J Chromatogr A. **2014**, 1335, 81-103.

[9] Brusotti G. et al., *Advances on size exclusion chromatography and applications on the analysis of protein biopharmaceuticals and protein aggregates: a mini review*, Chromatographia. **2018**, 81(1), 3-23.

[10] National Research Council Canada, *Evaluating the behaviour of complex molecules - Functional Characterization*, **2016**

[11] World Health Organization, *Global tuberculosis report 2016*, Geneva, Switzerland, WHO Press, **2016**.

[12] Piubelli L. et al., *Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins*, Microb Cell Fact. **2013**, 12, 115.

[13] Bavaro T. et al., *Chemoenzymatic synthesis of neoglycoproteins driven by the assessment of protein surface reactivity*, RSC Advances **2014**, 4, 56455-56465.

[14] Bavaro T. et al., *Glycosylation of recombinant antigenic proteins from Mycobacterium tuberculosis: in silico prediction of protein epitopes and ex vivo biological evaluation of new semi-synthetic glycoconjugates*, Molecules. **2017**, 22(7), 1081.

[15] Källenius G. et al., *Mycobacterial glycoconjugates as vaccine candidates against tuberculosis*, Trends Microbiol. **2008**, 16(10), 456-462.

[16] Demkow U. et al., *Humoral immune response against mycobacterial antigens in bronchoalveolar fluid from tuberculosis patients*, J Physiol Pharmacol. **2005**, 56 (Suppl 4), 79-84.

[17] Mishra A.K. et al., *Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in Mycobacterium tuberculosis physiology and host-pathogen interaction*, FEMS Microbiol Rev. **2011**, 35(6), 1126-1157.

[18] Caldinelli L. et al., *Dissecting the structural determinants of the stability of cholesterol oxidase containing covalently bound flavin*, J Biol Chem. **2005**, 280(24), 22572-22581.

[19] Rinaldi F. et al., *Application of a rapid HILIC-UV method for synthesis optimization and stability studies of immunogenic neo-glycoconjugates*, J Pharm Biomed Anal. **2017**, 144, 252-262.

[20] Sun J. et al., *SEPPA: a computational server for spatial epitope prediction of protein antigens*, Nucleic Acids Res. **2009**, 37, W612-W616.

[21] Liang S. et al., *Prediction of antigenic epitopes on protein surfaces by consensus scoring*, BMC Bioinformatics. **2009**, 10, 302.

[22] Davis B.G. et al., *Imaging agent*, International Patent, WO 2007/020450 A2, 22 February **2007**.

5. *Epitope identification and affinity determination of Mycobacterium tuberculosis Ag85B antigen towards anti-Ag85 antibodies from different sources*

Francesca Rinaldi, Loredana Lupu, Hendrik Rusche, Zdeněk Kukačka, Sara Tengattini, Roberta Bernardini, Teodora Bavaro, Luciano Piubelli, Stefan Maeser, Massimo Amicosante, Loredano Pollegioni, Gabriella Massolini, Enrica Calleri, Caterina Temporini, Michael Przybylski

Draft

Epitope identification and affinity determination of *Mycobacterium tuberculosis* Ag85B antigen towards anti-Ag85 antibodies from different sources

Francesca Rinaldi[a,b], Loredana Lupu[b,*], Hendrik Rusche[b,*], Zdeněk Kukačka[b,*], Sara Tengattini[a], Roberta Bernardini[c], Teodora Bavaro[a], Luciano Piubelli[d], Stefan Maeser[b], Massimo Amicosante[c], Loredano Pollegioni[d], Gabriella Massolini[a], Enrica Calleri[a], Michael Przybylski[b], Caterina Temporini[a]

[a] *Department of Drug Sciences, University of Pavia, Viale Taramelli 12, 27100 Pavia, Italy*

[b] *Steinbeis Centre for Biopolymer Analysis and Biomedical Mass Spectrometry, Marktstraße 29, 65428 Rüsselsheim am Main, Germany*

[c] *Department of Biomedicine and Prevention and Animal Technology Station, University of Rome "Tor Vergata", Via Montpellier 1, 00133 Rome, Italy*

[d] *Department of Biotechnology and Life Sciences, University of Insubria, Via Dunant 3, 21100 Varese, Italy and The Protein Factory Research Centre, Politecnico of Milano and University of Insubria, Via Mancinelli 7, 20131 Milano, Italy*

* *Contributed equally*

Abstract

Tuberculosis (TB) is the first cause of death from an infectious disease worldwide. Nowadays, only one anti-TB vaccine is available for clinical use, but various studies highlighted that its efficacy is not always achieved. The aim of this work is to provide a basis for the rational design of a *neo*-glycoconjugate vaccine against TB. After initial structural characterization of recombinant antigenic proteins from *Mycobacterium tuberculosis* Ag85B (wild-type, variants and semi-synthetic glycoconjugates), identification of antibody epitopes by proteolytic affinity-mass spectrometry and surface plasmon resonance (SPR) biosensor analysis were used to qualitatively and quantitatively characterize interactions of the antigens with antibodies from different sources. A commercial monoclonal antibody and polyclonal antibodies from patients with active TB, vaccinated individuals and a healthy control were employed to analyze antigen-antibody interactions. The combination approach provided the identification of different protein epitope regions involved in the interaction with specific antibodies, and a quantitative comparison between the affinities of the selected antigens.

Abbreviations

AraMan: Arabinose(1 → 6)Mannose
BCG: Bacillus Calmette-Guérin
DHB: Dihydroxybenzoic acid
DTT: Dithiothreitol
EDC: N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride
HILIC: Hydrophilic Interaction Liquid Chromatography
IME: 2-Iminomethoxyethyl
LAM: Lipoarabinomannan
mAb: monoclonal antibody
MALDI-TOF: Matrix-Assisted Laser Desorption/Ionization-Time Of Flight
ManMan: Mannose(1 → 6)Mannose
MS: Mass Spectrometry
Mtb: *Mycobacterium tuberculosis*
NHS: N-hydroxysuccinimide
PBS: Phosphate Buffered Saline
SAM: Self-Assembled Monolayer
SDHB: Super DHB
SPR: Surface Plasmon Resonance
TB: Tuberculosis
TFA: Trifluoroacetic acid

1. Introduction

Tuberculosis (TB) still constitutes one of the main causes of death in the world and in 2015 it ranked above HIV/AIDS as the most lethal infectious disease. TB is caused by the bacillus *Mycobacterium tuberculosis* (*Mtb*), from which various drug-resistant strains emerged and spread. The only antitubercular vaccine currently in clinical use is Bacillus Calmette-Guérin (BCG), developed in 1921 from the attenuation of a laboratory strain of *Mycobacterium bovis*. However, the variable protective efficacy observed for BCG and the growing public health concern for drug-resistant forms of TB pointed out the need for new effective agents against this disease [1].

The development of a new vaccine requires a well-defined correlation between design and protection. Two key prerequisites for the rational design of effective prophylactic agents capable to elicit a long-term protection are the comprehension of epitope/antibody interactions and the quantitative evaluation of vaccine-induced immune responses [2, 3]. To identify antigenic regions of molecules coming from microorganisms, which might be valuable vaccine candidates, B cell epitope determination is a promising approach, especially for the discovery of assembled discontinuous epitopes [2]. Mass spectrometry (MS)-based epitope identification approaches are among the most powerful; a strategy widely used (proteolytic epitope extraction) includes antigen digestion by a proteolytic enzyme, incubation with the binder antibody, washing steps to remove unbound peptides

and MS analysis of the elution fractions to identify the epitope peptides involved in the binding [4-6]. Surface plasmon resonance (SPR) biosensor technology has proven to be a precious technique for vaccine characterization, as it quantifies binding affinities and provides reliable kinetic data to evaluate specificity and mechanism of action [7]. Noteworthy, affinity-MS epitope determination and SPR approaches can be used as a combination to investigate immune responses directly in serum samples, representative of the total human antibody set [4, 7].

In this work, several analytical techniques were employed for the characterization of a new potential glycoconjugate vaccine against TB. Such a glycovaccine can be obtained by conjugation of antigenic proteins, which are expressed by *Mtb* and are known to induce a strong immune reaction, with carbohydrate moieties capable to mimic lipoarabinomannan (LAM), which is the most important saccharidic superficial antigen of *Mtb*. The *neo*-glycoconjugate might induce both humoral and T cell-mediated immune responses, resulting in an improved antigenicity compared to the single constituents [8]. Ag85B was selected as carrier protein since it is one of the most potent antigen species expressed by *Mtb*, has proven its anti-TB protective efficacy in murine, guinea pigs and other animal models and has shown good antibody and T cell responses in patients with active TB as well as BCG-vaccinated individuals [9-12]. The choice of carbohydrate antigens as simplified models (saccharides of mannose [8, 12-15]) was based on literature data on *Mtb* reporting the production of specific anti-LAM antibodies in lungs of patients with active TB [16] and the characteristic mono-, di- or tri-mannosylation of LAM of pathogenic mycobacteria [17, 18].

In a previous part of this project [12], different experimental and computational methods suggested that Lys30 and Lys282 residues (the two most reactive amino acids of Ag85B involved in the glycosylation) are antigenic. Mutations of these two lysines to arginine residues were designed to direct the glycosylation towards non-antigenic amino acids, thus preserving immunogenicity of the constructs (see *Chapter 4* of the experimental part of the present PhD thesis, pp. 73-93).

The areas of Ag85B protein involved in the interaction with antibodies from different sources (commercial monoclonal antibody from mouse, total antibodies from sera of active TB-patients, BCG-vaccinated subjects and a healthy control) were investigated using the proteolytic epitope extraction-MS method. In addition, affinities of Ag85B protein, its mutants and its glycosylated forms for the commercial monoclonal antibody (mAb) were compared by SPR analysis to support the mutagenesis approach.

2. Materials and methods

2.1. Reagents and chemicals

Mycobacterium tuberculosis Ag85 monoclonal antibody was purchased from Thermo Fisher Scientific (Darmstadt, Germany). Total antibodies from sera of patients with active TB, individuals vaccinated against TB (BCG vaccination) and a healthy control were provided from the University of Rome "Tor Vergata" (Department of Biomedicine and Prevention and Animal Technology Station, Rome, Italy). Ammonium sulfate precipitation was used to enrich and concentrate total immunoglobulins from sera and antibodies were then transferred in a phosphate buffered saline (PBS) solution. All sera were tested for antibody reactivity to the protein antigens by ELISA assays, following the same procedure described in [9].

Ag85B recombinant protein was produced in The Protein Factory Research Centre (Politecnico of Milano and University of Insubria, Milan, Italy) as described in [9]. K30R, K282R and K30R/K282R variants of Ag85B protein were also prepared by The Protein Factory Research Centre (see *Chapter 4* of the experimental part of the present PhD thesis, pp. 73-93). For SPR measurements, Ag85B protein was conjugated with Mannose(1→6)Mannose-IME (ManMan-IME) and Arabinose(1→6)Mannose-IME (AraMan-IME) carbohydrates synthesized and activated as in [13]. Glycosylation was performed as previously described [8, 14].

16-Mercaptohexadecanoic acid, α-chymotrypsin from bovine pancreas, $CaCl_2$, ethanolamine, iodoacetamide, N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC), Na_2HPO_4, NaN_3, N-hydroxysuccinimide (NHS), sodium acetate and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile and NaCl were from Carl Roth GmbH (Karlsruhe, Germany), dithiothreitol (DTT) from Roche (Mannheim, Germany), urea from VWR (Radnor, PA, USA).

Water was obtained from a Direct-Q® 3 UV system (Merck KGaA, Darmstadt, Germany).

2,5-dihydroxybenzoic acid (2,5-DHB) and Super DHB (SDHB) were purchased from Bruker Daltonics (Billerica, MA, USA).

POROS™ A material and Zeba™ spin desalting columns were from Thermo Fisher Scientific (Darmstadt, Germany).

Amicon® Ultra-0.5 mL centrifugal filters, ZipTip® pipette tips and ammonium bicarbonate were from Merck KGaA (Darmstadt, Germany).

2.2. Proteolytic extraction-mass spectrometry

For proteolytic extraction MS experiments, Ag85B protein and its mutants were digested by chymotrypsin prior to incubation with immobilized antibodies.

1 mg/mL protein solutions in 100 mM ammonium bicarbonate buffer, pH 8.5 were incubated at 95 °C for 20 minutes. After cooling to room temperature, urea and DTT were added to a final concentration of 6 M and 5 mM, respectively. Solutions were incubated at 60 °C for 30 minutes to reduce disulfide bonds, cooled to room temperature and spun to collect condensation. The addition of iodoacetamide to 14 mM final concentration was followed by incubation at room temperature in the dark for 30 minutes to alkylate cysteines. Protein mixtures were added with DTT to an additional 5 mM concentration and incubated at room temperature in the dark for 15 minutes to quench unreacted iodoacetamide. Solutions were diluted 1:5 in 100 mM ammonium bicarbonate buffer, pH 8.5 to reduce the concentration of urea to less than 2 M, followed by the addition of $CaCl_2$ to 1 mM concentration. Chymotrypsin was added to 1/50 enzyme/substrate ratio (w/w) and the reaction mixture was incubated overnight at 37 °C. Reactions were stopped by separation of the peptide mixture from the enzyme using 10 kDa Amicon® Ultra centrifugal filters (centrifugation at 13000 rpm for 5 minutes, 2 steps). Peptides were transferred in 10 mM Na_2HPO_4, pH 7 in presence of 0.02% (w/v) NaN_3 by Zeba™ spin desalting columns and concentrated to 2 mg/mL.

A Protein A affinity column was prepared by loading POROS™ A medium on a 100 µL micro-column. The column was equilibrated with the running buffer (10 mM Na_2HPO_4, pH 7 added with 0.02% w/v NaN_3) for 10 minutes with a constant flow rate of 1 mL/min. For the immobilization, 50 µL of each antibody solution (commercial mAb, total antibodies from sera of active-TB patients, vaccinated subjects and a healthy control) in phosphate buffer were injected onto the Protein A affinity column at a 500 µL/min constant flow rate. The column was washed for 10 minutes at 500 µL/min to remove any species not captured from the Protein A.

For epitope extraction, 50 µL of the previously prepared Ag85B (or Ag85B variants) proteolytic peptide mixtures were loaded on the column in the running buffer at a flow rate of 20 µL/min. The flow through was monitored at 220 and 280 nm using a UV-Vis detector (Dionex, Thermo Fisher Scientific, Darmstadt, Germany) and the column was washed until absorption reached again the baseline. Elution was performed by switching the running buffer to 150 mM NaCl, pH 2.5 with 0.02% (w/v) NaN_3 and the flow rate to 500 µL/min. Elution fractions were collected and peptides were separated from antibodies using 10 kDa Amicon® Ultra centrifugal filters. After 2 centrifugation steps at 13000 rpm for 5 minutes, 0.1% TFA in water (v/v) was added to the filters, which were centrifuged

for additional 10 minutes at 13000 rpm. The peptide solutions passed through the filters were concentrated, desalted using the ZipTip® procedure [19] and analyzed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS.

2.3. Mass spectrometric analysis

Mass spectrometric analyses were performed using an Autoflex III Smartbeam (Bruker Daltonics, Billerica, MA, USA) MALDI-TOF system equipped with a Nd:YAG laser source (4000 resolving power at 3200 Da in linear mode and 11000 resolving power below 3500 Da in reflectron mode).

For intact protein determinations, 0.5 µL of each sample solution were spotted on a Ground Steel MALDI Target Plate (Bruker Daltonics, Billerica, MA, USA), mixed with 0.5 µL of Super DHB matrix solution (50 mg/mL SDHB in 50/50 v/v acetonitrile / 0.1% TFA in water) and allowed to dry at room temperature. Spectra were acquired using the following instrumental conditions: linear mode; mass range 10000-35000 Da for proteins and glycoconjugates, 20000-100000 and 100000-200000 Da for antibodies; detector gain 10 (proteins and glycoconjugates) or 100 V (antibodies); sample rate 2.00 GS/s; electronic gain 50 mV.

For peptide mapping and epitope identification, 0.5 µL of 2,5-DHB matrix solution (10 mg/mL 2,5-DHB in 30/70 v/v acetonitrile / 0.1% TFA in water) were spotted on an AnchorChip MALDI Target Plate (Bruker Daltonics, Billerica, MA, USA). 1 µL of sample solution (after desalting by ZipTip®) was deposited onto each matrix spot and allowed to dry. Spectra were acquired using the following instrumental conditions: linear mode; mass range 240-8000 Da; detector gain 5.5 V; sample rate 2.00 GS/s; electronic gain 50 mV.

Data processing was performed using mMass open source software (version 5.5) [20].

2.4. SPR biosensor determinations

Determination of affinity constants was carried out on a SR7500DC SPR biosensor (Ametek-Reichert Technologies, Depew, NY). *Mycobacterium tuberculosis* Ag85 mAb (Thermo Fisher Scientific, Darmstadt, Germany) was immobilized on two different sensor slides: a carboxymethyl dextran hydrogel surface sensor chip and a gold surface sensor chip coated with a self-assembled monolayer (SAM) of 16-mercaptohexadecanoic acid.

Carboxyl groups on the chip surface were activated by injecting 250 µL of a freshly prepared solution of 200 mM EDC and 100 mM NHS at a constant flow rate of 25 µL/min. After washing the chip surface with 1 mL of 10 mM PBS (10 mM Na_2HPO_4, 30 mM NaCl), pH 7.5, 250 µL of a 35 µg/mL

antibody solution (in 10 mM sodium acetate buffer, pH 5.2) were injected at 25 µL/min for the immobilization. Unreacted NHS groups were blocked with 1 M ethanolamine, pH 8.5. The antibody was immobilized on left channel and right channel was used as a reference.

Affinity determinations were performed at 22 °C (SAM-coated chip) or 27 °C (dextran chip) in 10 mM PBS binding buffer, pH 7.5 at a flow rate of 25 µL/min, by injecting 100 µL of each sample. A dilution series from 0.2 to 6.4 µM (6.25-200 µg/mL) was analyzed for each protein. Each sample injection was followed by a PBS injection (blank). Affinity K_D constants and associated errors were calculated using the TraceDrawer software (Ridgeview Instruments AB, Vange, Sweden) according to a 1:1 binding model.

3. Results and discussion

3.1. Preliminary characterization of the samples

All proteins (wild-type Ag85B, Ag85B-K30R, Ag85B-K282R, Ag85B-K30R/K282R), wild-type Ag85B glycoconjugates and antibodies used in this study were analyzed by MALDI-TOF MS.

Following determination of intact proteins providing a first identity characterization, peptide mapping analysis was carried out on Ag85B recombinant protein and its variants to confirm protein sequences and identify and locate mutations. These analyses were performed by MALDI-TOF MS after a chymotryptic digestion. Wild-type and mutated Ag85B were completely digested and provided a high sequence coverage (>95%) to ascertain the identity of all proteins. As an example, **Figure 1, Table 1** and **Figure 2** report the results of peptide mapping for the chymotryptic digest of Ag85B-K30R mutant.

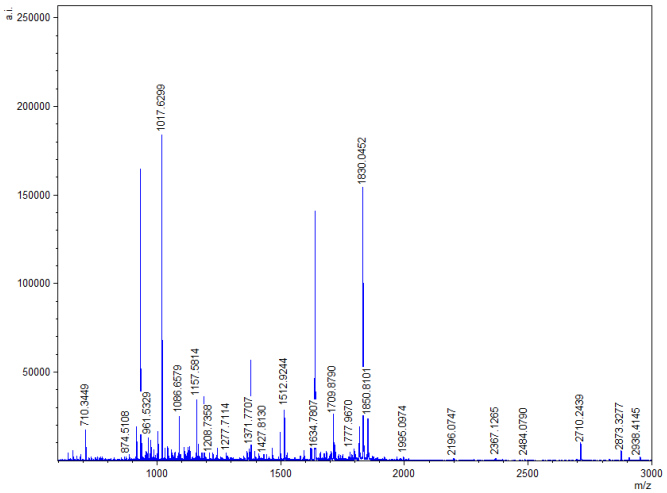


Figure 1. MALDI-TOF MS spectrum of Ag85B-K30R proteolytic mixture.

Table 1. List of the peptides identified from the peptide mapping of Ag85B-K30R protein. Peptides including R30 mutation are in orange.

Slice	Mis	m/z	z	Sequence
[1-17]	1	1849.9153	1	.AMAISDPFSRPGLPVEY.l
[9-17]	0	1017.5364	1	f.SRPGLPVEY.l
[9-18]	1	1130.6204	1	f.SRPGLPVEYL.q
[18-33]	1	1829.9691	1	y.LQVPSPSMGRDIRVQF.q
[18-44]	2	2904.4421	1	y.LQVPSPSMGRDIRVQFQSGGNNSPAVY.l
[19-33]	0	1716.8850	1	l.QVPSPSMGRDIRVQF.q
[19-45]	2	2904.4421	1	l.QVPSPSMGRDIRVQFQSGGNNSPAVYL.l
[34-44]	0	1093.4909	1	f.QSGGNNSPAVY.l
[34-58]	5	2710.2492	1	f.QSGGNNSPAVYLLDGLRAQDDYNGW.d
[45-55]	3	1278.6325	1	y.LLDGLRAQDDY.n
[45-58]	4	1635.7762	1	y.LLDGLRAQDDYNGW.d
[45-68]	7	2872.3140	1	y.LLDGLRAQDDYNGWDINTPAFEWY.y
[46-58]	3	1522.6921	1	l.LDGLRAQDDYNGW.d
[47-58]	2	1409.6080	1	l.DGLRAQDDYNGW.d
[47-67]	4	2483.0898	1	l.DGLRAQDDYNGWDINTPAFEW.y
[50-58]	1	1124.4756	1	l.RAQDDYNGW.d
[50-67]	3	2197.9574	1	l.RAQDDYNGWDINTPAFEW.y
[50-68]	4	2361.0207	1	l.RAQDDYNGWDINTPAFEWY.y
[59-67]	1	1092.4997	1	w.DINTPAFEW.y
[59-68]	2	1255.5630	1	w.DINTPAFEWY.y
[74-85]	0	1208.5980	1	l.SIVMPVGGQSSF.y
[74-86]	1	1371.6613	1	l.SIVMPVGGQSSFY.s
[87-102]	2	1736.7043	1	y.SDWYSPACGKAGCQTY.k
[91-102]	0	1185.5027	1	y.SPACGKAGCQTY.k
[103-107]	1	710.3508	1	y.KWETF.l
[103-116]	4	1777.9159	1	y.KWETFLTSELPQWL.s
[105-116]	3	1463.7417	1	w.ETFLTSELPQWL.s
[108-116]	2	1086.5830	1	f.LTSELPQWL.s
[117-132]	0	1512.8493	1	l.SANRAVKPTGSAAIGL.s
[133-142]	0	967.4587	1	l.SMAGSSAMIL.a
[143-150]	1	961.4526	1	l.AAYHPQQF.i
[143-152]	2	1237.6000	1	l.AAYHPQQFIY.a
[146-156]	2	1260.6371	1	y.HPQQFIYAGSL.s
[153-170]	3	1700.8524	1	y.AGSLSALLDPSQGMGPSL.i
[153-173]	4	1984.0420	1	y.AGSLSALLDPSQGMGPSLIGL.a
[157-173]	3	1655.8673	1	l.SALLDPSQGMGPSLIGL.a
[161-181]	2	1993.8994	1	l.DPSQGMGPSLIGLAMGDAGGY.k
[161-187]	3	2696.2153	1	l.DPSQGMGPSLIGLAMGDAGGYKAADMW.g
[182-207]	2	2938.4152	1	y.KAADMWGPSSDPAWERNDPTQQIPKL.v
[196-215]	2	2393.2684	1	w.ERNDPTQQIPKLVANNTRLW.v
[215-226]	2	1352.5940	1	l.WVYCGNGTPNEL.g
[216-235]	2	2022.9226	1	w.VYCGNGTPNELGGANIPAEF.l
[218-236]	2	1873.8749	1	y.CGNGTPNELGGANIPAEFL.e
[227-235]	0	875.4258	1	l.GGANIPAEF.l
[236-245]	2	1178.6164	1	f.LENFVRSSNL.k
[240-247]	1	950.5418	1	f.VRSSNLKF.q
[240-251]	2	1427.7278	1	f.VRSSNLKFQDAY.n
[240-261]	3	2366.1636	1	f.VRSSNLKFQDAYNAAGGHNAVF.n
[246-261]	2	1709.8030	1	l.KFQDAYNAAGGHNAVF.n
[252-261]	0	957.4537	1	y.NAAGGHNAVF.n
[262-271]	0	1156.5170	1	f.NFPPNGTHSW.e
[262-274]	2	1634.7023	1	f.NFPPNGTHSWEYW.g
[262-278]	3	2003.9035	1	f.NFPPNGTHSWEYWGAQL.n
[275-292]	3	1717.8538	1	w.GAQLNAMKGDLQSSLGAG.
[279-289]	1	1163.5725	1	l.NAMKGDLQSSL.g
[279-292]	2	1348.6525	1	l.NAMKGDLQSSLGAG.

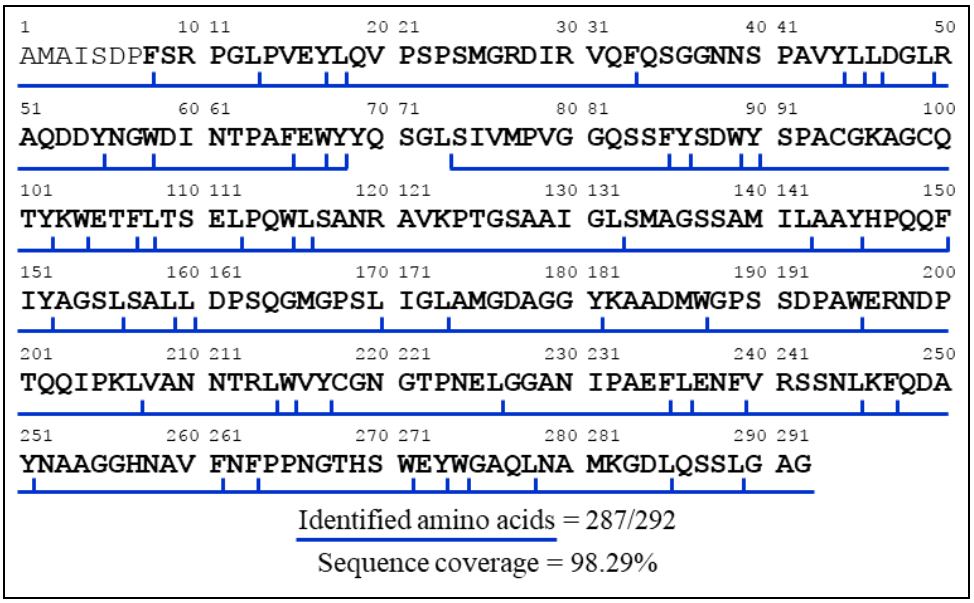
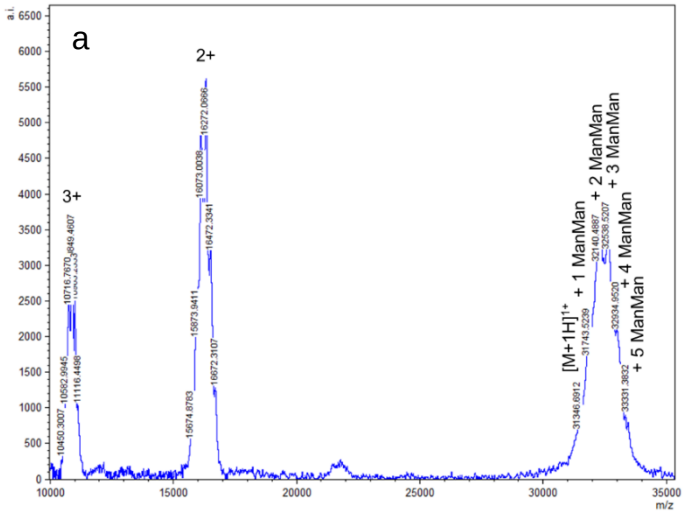


Figure 2. Sequence coverage obtained by the peptide mapping of Ag85B-K30R chymotryptic digest. Vertical lines indicate cleavage sites.

Ag85B glycoconjugates were also prepared and characterized. Carbohydrates used for the conjugation were activated with the 2-iminomethoxyethyl (IME) functional group [13], which selectively reacts with the ε-amino group of lysine residues [8]. The conjugation method may involve all surface lysine residues (8 in Ag85B), yielding to a mixture of glycoproteins with variable saccharide loading number and positioning. Intact protein determinations allowed to confirm the effectiveness of the glycosylation procedure. Ag85B-ManMan glycoconjugates investigated in this work present up to 5 glycan moieties *per* protein (**Figure 3a**), while conjugation with AraMan-IME was less efficient, with a maximum of 4 disaccharides bound to the protein (**Figure 3b**). The glycosylation profile was also characterized by a previously developed hydrophilic interaction liquid chromatography (HILIC)-UV method [14], which allowed to assess yield and glycoform composition from each coupling reaction (**Table 2**).



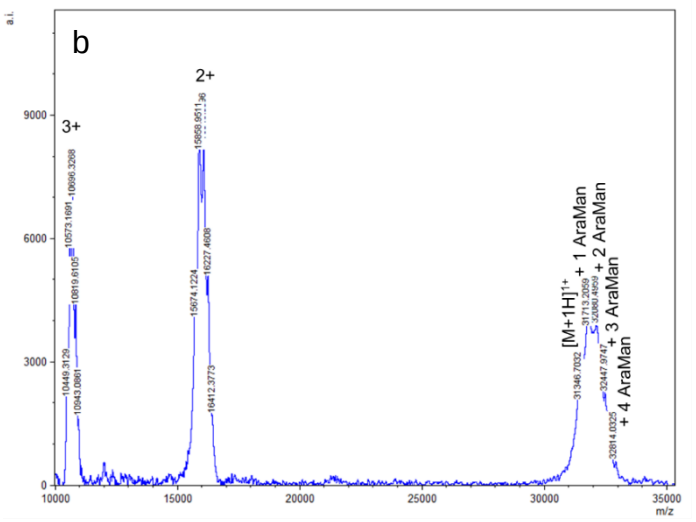


Figure 3. MALDI-TOF MS spectra of Ag85B protein conjugated with ManMan-IME(a) and AraMan-IME(b). Mono-, bi- and tri-charged ions of the glycoproteins are shown.

Table 2. Glycoform distribution (%), glycosylation yield (Y, %) and glycoside/protein ratio (mol/mol) calculated from peak areas obtained by HILIC-UV analysis of Ag85B glycoconjugates.

Glycoconjugate	Number of incorporated glycosides (%)									Y (%)	Glycoside/protein (mol/mol)
	0	1	2	3	4	5	6	7	8		
Ag85B + ManMan-IME	2.6	10.8	27.2	32.6	18.7	8.1	-	-	-	97.4	2.8
Ag85B + AraMan-IME	9.2	25.1	32.1	24.1	9.5	-	-	-	-	90.8	2.0

MS analyses of all antibodies used in this study were also performed. The antibody purchased from Thermo Fisher is a monoclonal IgG1 produced by immunization of mice with bacterial press extract from *Mtb* H37Rv (the most studied TB strain) and detects the *Mtb* complex of mycolytransferases composed of Ag85A, B and C proteins. MS analysis of the intact antibody provided a single main species with a molecular weight around 146 kDa (**Figure 4a**), confirming its monoclonality.

As expected, mass spectra of antibodies obtained from sera samples showed the presence of different populations (**Figure 4b**), being representative for the total immunoglobulin content (comprising non-specific and specific polyclonal antibodies) of human sera.

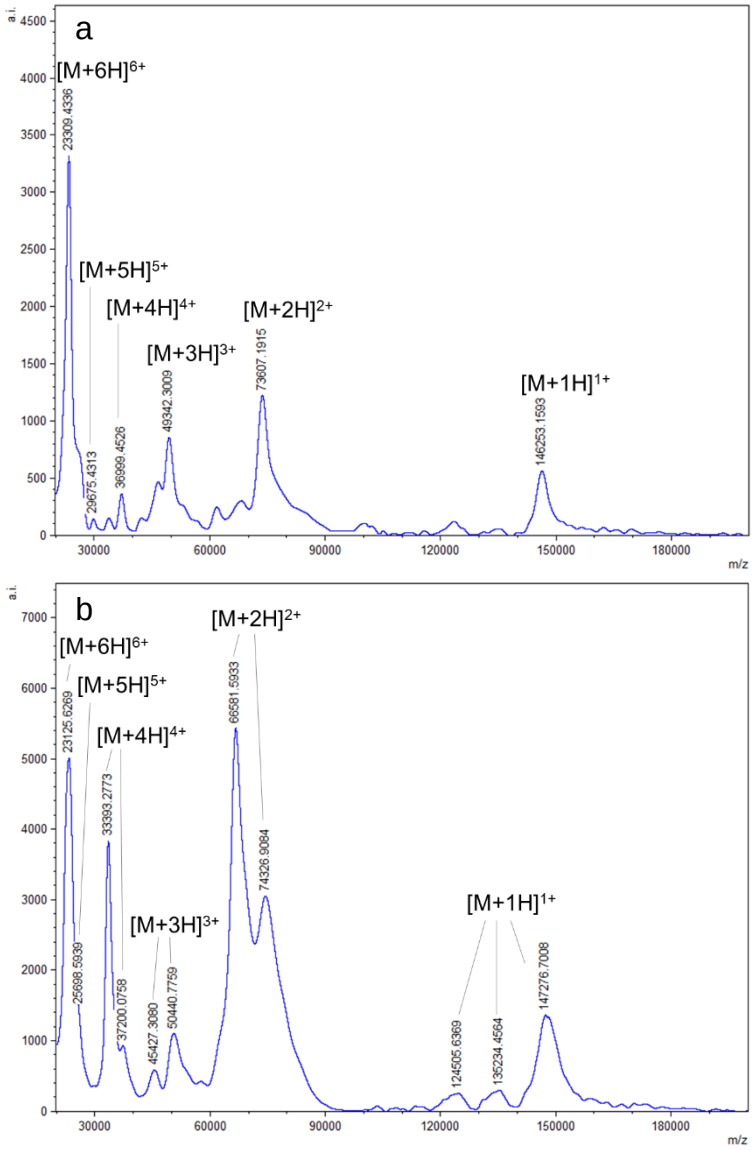


Figure 4. MALDI-TOF MS spectra of *Mtb* Ag85 mAb from Thermo Fisher (a) and polyclonal antibodies from serum of a BCG-vaccinated individual (b).

3.2. Mass spectrometric identification of epitope peptides

In a previous study of our group [12], the assessment of Ag85B glycosylation sites, the *ex vivo* evaluation of *neo*-glycoproteins immunogenic activity and the prediction of protein B and T cell epitopes by a combination of *in silico* systems suggested that the two main glycosylated sites (Lys30 and Lys282) are also antigenic. Thus, Ag85B variants at 30 and 282 residues were designed in order to prevent the glycosylation in these positions and to preserve the immunogenicity of the glycoconjugates (see *Chapter 4* of the experimental part of the present PhD thesis, pp. 73-93).

To support the rational design of an effective vaccine against TB, the antibody epitope of Ag85B antigen was determined using proteolytic extraction-MS.

The use of clinical samples and in particular of sera for epitope determination is representative of the overall human antibody set [4]. Therefore, antibodies from sera of patients with active TB, BCG-vaccinated subjects and a healthy control were employed in this study. These samples consisted of the total immunoglobulins from human sera, including both specific and non-specific antibodies for Ag85B antigen. Anti-Ag85B specific antibodies are produced in patients and vaccinated subjects, but are not present in healthy controls. These specific immunoglobulins represent a minority component of the total antibody content, generally up to 5 % in patients with active TB infection. For this reason, a key point during the set-up of the immobilization procedure was to find a suitable binding chemistry able to orient the antibodies with the variable regions exposed towards the mobile phase and to maximize the specific interaction of the analyte.

In addition to clinical samples, a commercial monoclonal antibody against Ag85B was studied to experimentally determine the binding areas of the same antigens to the mAb and assess whether it might constitute a simplified model to study the interaction with recombinant proteins from *Mtb*.

After testing different immobilization chemistries (cyanogen bromide method using CNBr-activated Sepharose 4B and amine coupling through EDC/NHS activation, data not shown), a POROS™ Protein A affinity column was selected. Protein A is naturally produced by *Staphylococcus aureus* and specifically binds the heavy chain constant region (Fc) of IgGs in the CH_2-CH_3 portion, thus allowing the interaction between the antibody variable region and its antigens.

For proteolytic extraction experiments, digestion of Ag85B protein with trypsin resulted in a poor sequence coverage. This might be explained by the globular folded structure of the protein hindering the enzyme to reach specific cleavage sites. For this reason, chymotryptic digests were used even if the presence of redundant peptides, some of them containing missed cleavages, can render the data interpretation more laborious.

The chymotryptic digests of Ag85B were incubated with the immobilized antibodies. After a washing step to remove unbound peptides, the elution fractions were collected and analyzed by MALDI-TOF MS to identify the peptides involved in the interaction. **Table 3** include all peptides identified in the elution fractions collected from each column (prepared with the commercial mAb or with antibodies from individual serum samples). Interestingly, peptide maps from the same areas of the antigen protein (in red in **Figures 5** and **6**) were found in the elution fractions from all the columns with antibodies from mouse (commercial mAb), patients and vaccinated subjects. Areas in blue in **Figures 5** and **6** represent recurring peptides, found in at least two elution fractions. The results showed that the detected common peptides belong to the same side of the protein, suggesting the presence of an

assembled epitope in this section. One important output of these data is that the commercial mAb binds to the same regions of the protein, thus validating the model. As expected, the peptides were not present in the elution from the healthy control column (three replicates). The specificity of binding peptides for anti-Ag85B antibodies was also verified by performing the proteolytic extraction experiment on the unmodified Protein A column (two replicates).

These findings suggest a common assembled epitope comprising different sequences recognized by anti-Ag85B antibodies from different sources.

Table 3. List of the identified peptides in the elution fractions of Ag85B proteolytic mixtures loaded on affinity columns with immobilized antibodies from mouse (commercial mAb), TB-patients and BCG-vaccinated people.

Slice	Mis	m/z	z	Sequence	Commercial mAb	Patient 1	Patient 2	Patient 3	Vaccinated 1	Vaccinated 2
[1-17]	1	1849.9153	1	.AMAISDPFSRPGLPVEY.l						✓
[9-17]	0	1017.5364	1	f.SRPGLPVEY.l	✓	✓	✓	✓	✓	✓
[45-58]	4	1635.7762	1	y.LLDGLRAQDDYNGW.d		✓	✓	✓		✓
[46-58]	3	1522.6921	1	l.LDGLRAQDDYNGW.d	✓	✓				
[47-55]	1	1052.4643	1	l.DGLRAQDDY.n	✓				✓	
[47-58]	2	1409.6080	1	l.DGLRAQDDYNGW.d	✓	✓			✓	✓
[50-58]	1	1124.4756	1	l.RAQDDYNGW.d	✓					
[59-67]	1	1092.4997	1	w.DINTPAFEW.y						✓
[74-86]	1	1371.6613	1	l.SIVMPVGGQSSFY.s						✓
[103-107]	1	710.3508	1	y.KWETF.l					✓	
[105-116]	3	1463.7417	1	w.ETFLTSELPQWL.s						✓
[133-150]	2	1909.8935	1	l.SMAGSSAMILAAYHPQQF.i			✓			
[143-150]	1	961.4526	1	l.AAYHPQQF.i						✓
[196-207]	0	1438.7649	1	w.ERNDPTQQIPKL.v					✓	
[208-217]	2	1235.6531	1	l.VANNTRLWVY.c	✓					
[215-226]	2	1352.5940	1	l.WVYCGNGTPNEL.g	✓				✓	
[216-226]	1	1166.5146	1	w.VYCGNGTPNEL.g			✓			
[216-235]	2	2022.9226	1	w.VYCGNGTPNELGGANIPAEF.l			✓			
[236-245]	2	1178.6164	1	f.LENFVRSSNL.k		✓		✓		✓
[237-245]	1	1065.5323	1	l.ENFVRSSNL.k				✓		
[240-251]	2	1427.7278	1	f.VRSSNLKFQDAY.n			✓			
[246-261]	2	1709.8030	1	l.KFQDAYNAAGGHNAVF.n			✓	✓		✓
[252-261]	0	957.4537	1	y.NAAGGHNAVF.n	✓	✓			✓	✓
[262-271]	0	1156.5170	1	f.NFPPNGTHSW.e	✓	✓		✓		✓
[262-273]	1	1448.6230	1	f.NFPPNGTHSWEY.w	✓		✓		✓	✓
[272-278]	2	866.4043	1	w.EYWGAQL.n						✓
[272-289]	4	2010.9590	1	w.EYWGAQLNAMKGDLQSSL.g			✓			
[279-292]	2	1348.6525	1	l.NAMKGDLQSSLGAG.	✓			✓		✓

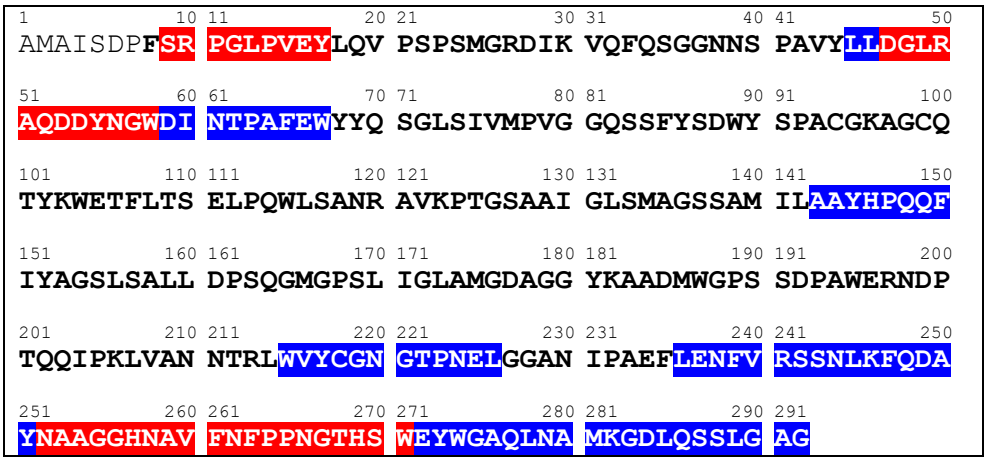


Figure 5. Epitope identification from the elution fractions collected after incubation of chymotryptic digests of Ag85B with immobilized antibodies from different sources. Amino acids underlined in red were identified in all the elution fractions (except from those collected from the control columns), while areas in blue represent recurring peptides.

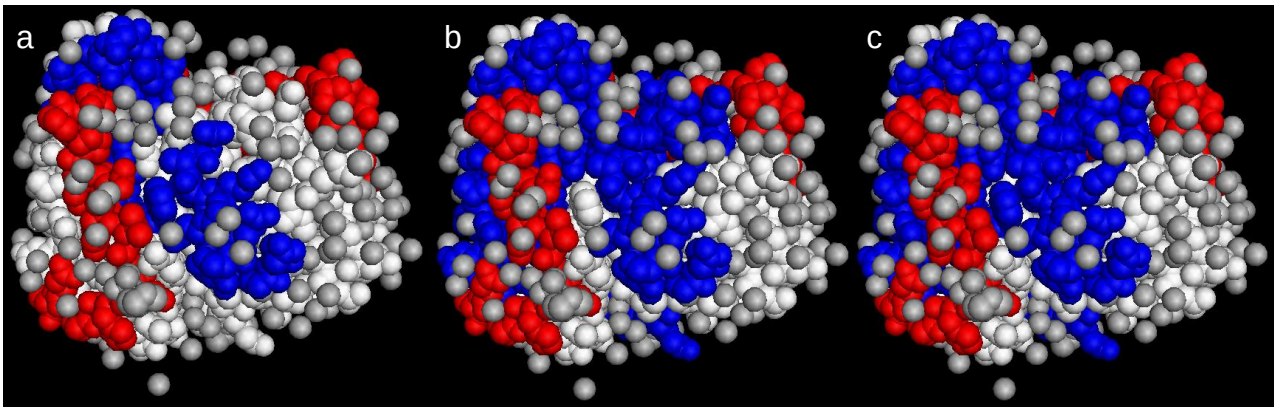


Figure 6. Areas of Ag85B protein interacting with commercial mAb (a), TB-patients polyclonal antibodies (b) and vaccinated people polyclonal antibodies (c). Amino acids in red were identified in all the elution fractions collected from the different affinity columns with immobilized antibodies, areas in blue in at least two elution fractions.

The repeatability of the epitope determination was assessed by duplicate experiments of the entire epitope extraction procedure, starting from the immobilization of a new aliquot of antibodies from the same source, for one patient (Patient 3) and one vaccinated individual (Vaccinated 2). The results (**Figure 7**, **Tables 4** and **5**) show that the majority of identified protein areas are found in both analyses, despite the use of different proteolytic mixtures and the preparation of different affinity columns.

A further step was the confirmation that the conservative mutations introduced in Ag85B protein (K30R and K282R) do not alter its antibody binding. To this aim, digestion mixtures of Ag85B variants were used for epitope extraction. Antibodies from a vaccinated individual (Vaccinated 2) were immobilized and a mixture of Ag85B-K30R and Ag85B-K282R chymotryptic digests was incubated with the affinity column. In the elution, peptides including both unmodified (Lys) and mutated (Arg) 282 residue were identified (**Table 5**). Also antibodies from Patient 3 were incubated

with Ag85B-K282R chymotryptic digest (**Table 4**). Peptides comprising Arg282 were found in the elution fraction, supporting the evidence that mutation of this residue do not prevent antigen-antibody interactions.

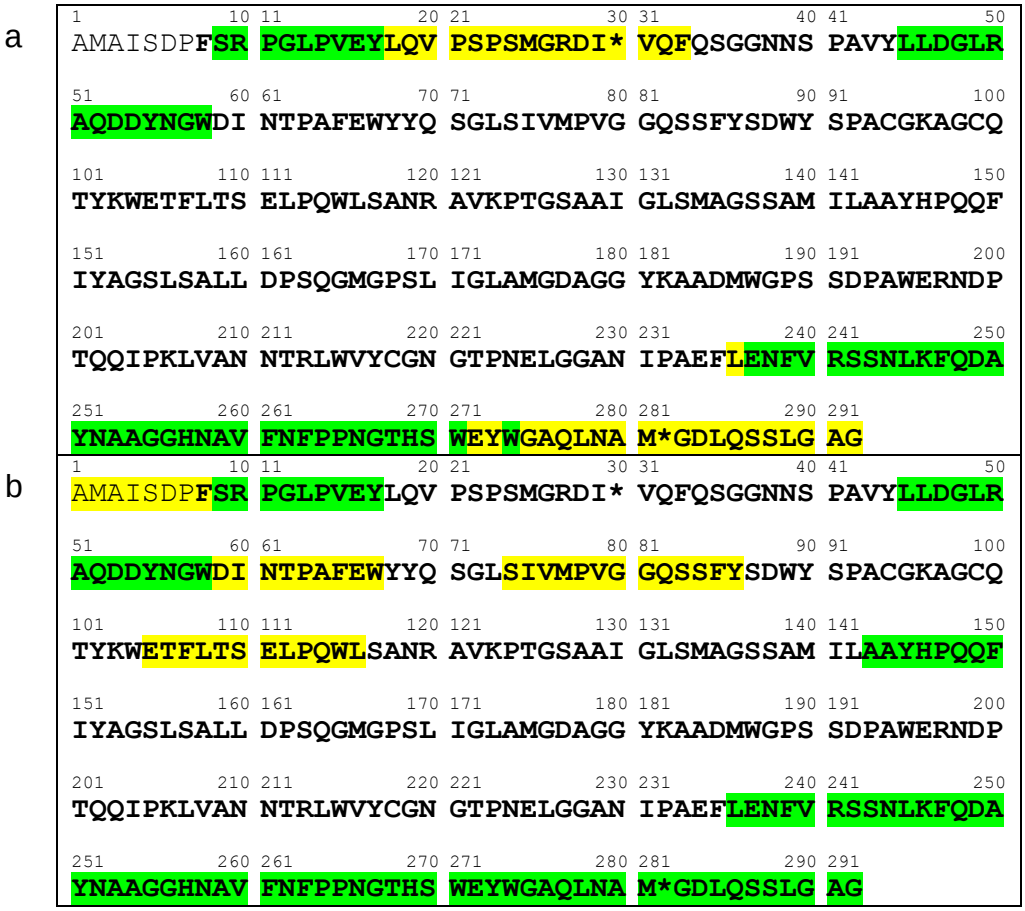


Figure 7. Identified sequences in the elution fractions from two analyses performed loading different Ag85B (or Ag85B variants) digestion mixtures on affinity columns prepared immobilizing two different aliquots of antibodies from the same patient (a) or vaccinated individual (b). Amino acids underlined in green were identified in both the elution fractions, while areas in yellow only in one analysis. * indicates K or R residues, depending on the use of wild-type or mutant Ag85B for the epitope extraction experiments.

Table 4. List of the peptides identified from the epitope determination of Ag85B and mutants peptide mixtures incubated with antibodies from Patient 3 (two replicates). Peptides including R282 mutation are in orange.

Slice	Mis	m/z	z	Sequence	Replicate 1	Replicate 2
[9-17]	0	1017.5364	1	f.SRPGLPVEY.l	✓	✓
[18-33]	1	1829.9691	1	y.LQVPSPSMGRDIRVQF.q		✓
[45-58]	4	1635.7762	1	y.LLDGLRAQDDYNGW.d	✓	✓
[236-245]	2	1178.6164	1	f.LENFVRSSNL.k	✓	
[237-245]	1	1065.5323	1	l.ENFVRSSNL.k		✓
[246-261]	2	1709.8030	1	l.KFQDAYNAAGGHNAVF.n	✓	✓
[262-271]	0	1156.5170	1	f.NFPPNGTHSW.e	✓	✓
[262-274]	2	1634.7023	1	f.NFPPNGTHSWEYW.g		✓
[274-289]	3	1746.8592	1	y.WGAQLNAMRGDLQSSL.g	✓	
[275-292]	3	1745.8599	1	w.GAQLNAMRGDLQSSLGAG.	✓	
[279-292]	2	1376.6587	1	l.NAMRGDLQSSLGAG.	✓	

Table 5. List of the peptides identified from the epitope determination of Ag85B and mutants peptide mixtures incubated with antibodies from Vaccinated 2 (two replicates). Peptides including K282 are underlined in light blue, those with R282 mutation are in orange.

Slice	Mis	m/z	z	Sequence	Replicate 1	Replicate 2
[1-17]	1	1849.9153	1	.AMAISDPFSRPGLPVEY.l		✓
[9-17]	0	1017.5364	1	f.SRPGLPVEY.l	✓	✓
[45-58]	4	1635.7762	1	y.LLDGLRAQDDYNGW.d	✓	✓
[47-58]	2	1409.6080	1	l.DGLRAQDDYNGW.d	✓	
[59-67]	1	1092.4997	1	w.DINTPAFEW.y		✓
[74-86]	1	1371.6613	1	l.SIVMPVGGQSSFY.s	✓	
[105-116]	3	1463.7417	1	w.ETFLTSELPQWL.s		✓
[143-150]	1	961.4526	1	l.AAYHPQQF.i	✓	✓
[236-245]	2	1178.6164	1	f.LENFVRSSNL.k	✓	✓
[246-261]	2	1709.8030	1	l.KFQDAYNAAGGHNAVF.n	✓	✓
[252-261]	0	957.4537	1	y.NAAGGHNAVF.n	✓	✓
[262-271]	0	1156.5170	1	f.NFPPNGTHSW.e	✓	✓
[262-273]	1	1448.6230	1	f.NFPPNGTHSWEY.w	✓	
[262-274]	2	1634.7023	1	f.NFPPNGTHSWEYW.g		✓
[272-278]	2	866.4043	1	w.EYWGAQL.n	✓	
[274-289]	3	1746.8592	1	y.WGAQLNAMRGDLQSSL.g		✓
[275-292]	3	1745.8599	1	w.GAQLNAMRGDLQSSLGAG.		✓
[279-292]	2	1376.6587	1	l.NAMRGDLQSSLGAG.	✓	✓
[279-292]	2	1348.6525	1	l.NAMKGDLQSSLGAG.	✓	

A literature search was performed to compare our data with previously published results. Although different T cell epitope determinations (reviewed in [21]) were carried out for *Mtb* Ag85B, few experimental data on B cell epitopes are available for this antigen. To our knowledge, this is the first report describing the extraction of epitope peptides from proteolytic mixtures of Ag85B incubated with immobilized antibodies. All the examined works about the interactions between antibodies and Ag85B peptides [11, 22-24] were carried out with synthetic peptides, either selected on the basis of *in silico* predictions of B epitopes [11, 22] or synthesized to cover the entire Ag85B protein sequence [23, 24]. An intriguing result from the literature search is that Kadir *et al.* [11] recently ascertained the involvement of the T cell epitopes L_{108}TSELPQWLSANRAV$_{122}$, S_{133}MAGSSAMILAAYHP$_{147}$ and T_{268}HSWEYWGAQLNAMK$_{282}$ (identified in [25]) also in the humoral response. The sequences A_{143}AYHP$_{147}$ (part of the second epitope peptide in [11]) and T_{268}HSWEYWGAQLNAMK$_{282}$ (third epitope) were in agreement with the epitope determination in our present work. In [22], peptides based on the sequences P_{14}VEYLQVPS$_{22}$ and G_{256}HNAVFN$_{262}$ of Ag85 proteins were demonstrated to be immunoreactive by ELISA. Both sequences (except for L_{18}QVPS$_{22}$ amino acids) were identified in all elution fractions examined in the present study. Some epitope regions in our work (L_{45}-G_{57} and W_{215}-L_{226}) were also recognized by antibodies from TB patients in [23]. The peptide P_{265}NGTHSWEYWGAQ$_{277}$, included in the areas of Ag85B protein found in our epitope

determination, showed a significantly higher humoral response in TB patients than in healthy individuals in a previous microarray experiment on human sera [24]. Finally, most epitope regions identified in the present work were found in the latest B cell epitope predictions of Ag85B [26] and Ag85A [27], a protein belonging to the same complex of Ag85B that shows high sequence homology with this antigen (about 77% shared amino acids).

3.3. Determination of affinity binding constants

The validation of the simplified monoclonal model by the epitope determination prompted us to use the commercial *Mtb* Ag85 antibody to assess the affinity binding constants of Ag85B, its variants and its glycosylated forms by SPR analysis. The same study was not possible using antibodies from sera due to the low content of anti-Ag85B specific antibodies in these samples, which did not allow to appreciate any specific interaction (data not shown).

At first, the effect of protein glycosylation was evaluated. The use of a carboxymethyl dextran chip, which provides a very stable and highly flexible surface favouring the easy coupling of biomolecules and the accessibility of binding sites, was not suitable for the determination of glycoprotein equilibrium dissociation constants (K_D). The glycosylated antigen showed a similar binding to the channels with and without (reference) the immobilized antibody, suggesting the occurrence of non-specific interactions between the carbohydrate moieties and the dextran surface. Therefore, a SAM-coated chip was employed as an alternative immobilization support. The two supports gave comparable maximal responses upon analyte binding (R_{max}), 15.3 ± 2.5 µRIU for the dextran chip and 18 ± 3.6 µRIU for the SAM-coated chip (mean values of three independent determinations). The response depends on the number of immobilized ligand molecules and on the size of ligand and analyte (which were identical for the two chips), providing a measure of the maximum analyte binding capacity of the surface. The obtained R_{max} values suggest that the two surfaces gave similar yields in terms of amount and proper orientation for the interaction of the immobilized antibodies. An additional confirmation of the comparable antibody immobilization on the two chips was obtained from the Ag85B K_D values, which were found to be consistent with each other (**Table 6**). The 1:1 binding model, indicating that one immobilized antibody binds one analyte from the injected solution, was chosen to fit the obtained sensorgrams. The use of the SAM-coated chip allowed to define the K_D for Ag85B glycosylated with ManMan-IME and AraMan-IME, resulting in a 3- or 9-fold decrease of the binding affinity relative to the wild-type protein (**Table 6**). This finding is in agreement with the hypothesis of the masking of some epitope amino acid residues by glycosylation.

The two main Ag85B glycosylation sites (Lys30 and Lys282) were predicted to be involved in the formation of various T cell epitopes [12], as demonstrated by experimental evidences [21, 25]. In addition, we showed the interaction of Lys282 with monoclonal and polyclonal anti-Ag85B antibodies. In order to provide a proof of principle of the mutagenesis approach to prevent the glycosylation of antigenic amino acids, the affinity of single point variants (K30R and K282R) and a double variant (K30R/K282R) of Ag85B protein was quantified. The measurements were performed on both SPR chips, giving comparable results (**Table 6**). In summary, these data appear to support the validity of the mutagenesis approach, since the introduced mutations do not considerably alter the binding affinity while glycosylation reduces affinity compared to Ag85B wild-type protein.

The injection of a 0.2-6.4 µM dilution series of a negative control protein (myoglobin) confirmed the specificity of the interaction of Ag85B, its mutants and its glycoconjugates with the immobilized antibodies.

Table 6. Equilibrium dissociation constants (K_D) and estimated errors calculated for Ag85B protein, its variants (Ag85B-K30R, Ag85B-K282R and Ag85B-K30R/K282R) and its glycosylated forms (Ag85B+ManMan and Ag85B+AraMan) interacting with *Mtb* Ag85 mAb.

Sample	K_D (µM) Dextran chip	K_D (µM) SAM-coated chip
Ag85B	0.64 ± 0.0091	0.98 ± 0.0580
Ag85B-K30R	1.55 ± 0.1420	1.84 ± 0.0045
Ag85B-K282R	1.17 ± 0.0356	1.36 ± 0.0108
Ag85B-K30R/K282R	not determined	1.04 ± 0.0047
Ag85B+ManMan	not determined	3.31 ± 0.0188
Ag85B+AraMan	not determined	9.17 ± 0.0162

4. Conclusions

In this work, the mass spectrometric epitope determination and SPR biosensor analysis allowed to investigate the interactions between a known antigenic protein from *Mtb* (Ag85B) and antibodies from different sources, both qualitatively and quantitatively.

It was possible to use human clinical samples to determine the Ag85B epitope. Noteworthy, it was not necessary to isolate Ag85B-specific antibodies since a simple ammonium sulfate precipitation of

the total immunoglobulin content of sera samples followed by antibody immobilization on a Protein A column was sufficient to isolate epitope peptides from Ag85B digestion mixtures.

The identification of the same binding areas of the protein interacting with a commercial monoclonal antibody and with antibodies from human sera allowed to validate the simplified monoclonal model, confirming its applicability in SPR experiments. Therefore, an alternative screening method for the immunological evaluation of antigenic proteins from *Mtb* has been developed. With this *in vitro* approach, it was possible to study the effect of the introduction of conservative mutations (K30R and K282R) in Ag85B protein.

Moreover, both the SPR and epitope identification results indicate that the introduced mutations do not alter the binding properties of Ag85B, while glycosylation reduces the binding affinity due to the masking of some epitope amino acids (like K282).

These findings provide a molecular basis for an improved rational design of a new potential glycoconjugate vaccine against TB.

Acknowledgments

The research visiting stay of Francesca Rinaldi in the Steinbeis Centre for Biopolymer Analysis and Biomedical Mass Spectrometry was supported by an increase of the PhD scholarship and a grant for international mobility from the University of Pavia.

Professor Massimo Amicosante made a substantial contribution, sharing his knowledge and providing valuable suggestions. His busy schedule did not prevent him to assist us with great humanity and professionalism.

References

[1] World Health Organization, *Global tuberculosis report 2016*, Geneva, Switzerland, WHO Press, **2016**.

[2] Ahmad T.A. et al., *B-cell epitope mapping for the design of vaccines and effective diagnostics*, Trials Vaccinol. **2016**, 5, 71-83.

[3] He L., Zhu J., *Computational tools for epitope vaccine design and evaluation*, Curr Opin Virol. **2015**, 11, 103-112.

[4] Opuni K.F.M. et al., *Mass spectrometric epitope mapping*, Mass Spectrom Rev. **2016**, doi: 10.1002/mas.21516.

[5] Suckau D. et al., *Molecular epitope identification by limited proteolysis of an immobilized antigen-antibody complex and mass spectrometric peptide mapping*, Proc Natl Acad Sci U S A. **1990**, 87(24), 9848-9852.

[6] Macht M. et al., *Mass spectrometric mapping of protein epitope structures of myocardial infarct markers myoglobin and troponin T*, Biochemistry. **1996**, 35(49), 15633-15639.

[7] Hearty S. et al., *Surface plasmon resonance for vaccine design and efficacy studies: recent applications and future trends*, Expert Rev Vaccines. **2010**, 9(6), 645-664.

[8] Temporini C. et al., *Liquid chromatography-mass spectrometry structural characterization of neo glycoproteins aiding the rational design and synthesis of a novel glycovaccine for protection against tuberculosis*, J Chromatogr A. **2014**, 1367, 57-67.

[9] Piubelli L. et al., *Optimizing Escherichia coli as a protein expression platform to produce Mycobacterium tuberculosis immunogenic proteins*, Microb Cell Fact. **2013**, 12, 115.

[10] Wiker H.G., Harboe M., *The antigen 85 complex: a major secretion product of Mycobacterium tuberculosis*, Microbiol Rev. **1992**, 56(4), 648-661.

[11] Kadir N.A. et al., *Cellular and humoral immunogenicity of recombinant Mycobacterium smegmatis expressing Ag85B epitopes in mice*, Int J Mycobacteriol. **2016**, 5(1), 7-13.

[12] Bavaro T. et al., *Glycosylation of recombinant antigenic proteins from Mycobacterium tuberculosis: in silico prediction of protein epitopes and ex vivo biological evaluation of new semi-synthetic glycoconjugates*, Molecules. **2017**, 22(7), 1081.

[13] Bavaro T. et al., *Chemoenzymatic synthesis of neoglycoproteins driven by the assessment of protein surface reactivity*, RSC Advances **2014**, 4, 56455-56465.

[14] Rinaldi F. et al., *Application of a rapid HILIC-UV method for synthesis optimization and stability studies of immunogenic neo-glycoconjugates*, J Pharm Biomed Anal. **2017**, 144, 252-262.

[15] Tengattini S. et al., *Hydrophilic interaction liquid chromatography-mass spectrometry as a new tool for the characterization of intact semi-synthetic glycoproteins*, Anal Chim Acta. **2017**, 981, 94-105.

[16] Demkow U. et al., *Humoral immune response against mycobacterial antigens in bronchoalveolar fluid from tuberculosis patients*, J Physiol Pharmacol. **2005**, 56 (Suppl 4), 79-84.

[17] Philips J.A., Ernst J.D., *Tuberculosis pathogenesis and immunity*, Annu Rev Pathol. **2012**, 7, 353-384.

[18] Mishra A.K. et al., *Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in Mycobacterium tuberculosis physiology and host-pathogen interaction*, FEMS Microbiol Rev. **2011**, 35(6), 1126-1157.

[19] https://www.danforthcenter.org/docs/default-source/CoreFacilities/pmsf/ziptip-protocol.pdf?sfvrsn=2

[20] Strohalm M. et al., *mMass data miner: an open source alternative for mass spectrometric data analysis*, Rapid Commun Mass Spectrom. **2008**, 22(6), 905-908.

[21] Huygen K., *The immunodominant T-cell epitopes of the mycolyl-transferases of the antigen 85 complex of M. tuberculosis*, Front Immunol. **2014**, 5, 321.

[22] Kashyap R.S. et al., *Diagnosis of tuberculosis infection based on synthetic peptides from Mycobacterium tuberculosis antigen 85 complex*, Clin Neurol Neurosurg. **2013**, 115(6), 678-683.

[23] Shen G. et al., *Peptide-based antibody detection for tuberculosis diagnosis*, Clin Vaccine Immunol. **2009**, 16(1), 49-54.

[24] Nahtman T. et al., *Validation of peptide epitope microarray experiments and extraction of quality data*, J Immunol Methods. **2007**, 328(1-2), 1-13.

[25] Lee B.Y., Horwitz M.A., *T-cell epitope mapping of the three most abundant extracellular proteins of Mycobacterium tuberculosis in outbred guinea pigs*, Infect Immun. **1999**, 67(5), 2665-2670.

[26] Zhang F. et al., *The prediction of T- and B-combined epitope of Ag85B antigen of Mycobacterium tuberculosis*, Int J Clin Exp Med. **2016**, 9(2), 1408-1421.

[27] Dewi I.P., *B-cell epitope prediction of Mycobacterium tuberculosis Ag85A antigen*, UNEJ e-Proceeding. **2017**, S.l., 108-111.

6. *Evaluation of Mycobacterium tuberculosis purine nucleoside phosphorylase as a new therapeutic target for the development of antimycobacterial agents*

Francesca Rinaldi, Enrica Calleri, Daniela Ubiali, Caterina Temporini, Marco Rabuffetti, Marcela Cristina de Moraes, Giovanna Speranza, Gabriella Massolini

Draft

Evaluation of *Mycobacterium tuberculosis* purine nucleoside phosphorylase as a new therapeutic target for the development of antimycobacterial agents

Francesca Rinaldi[a], Enrica Calleri[a], Daniela Ubiali[a], Caterina Temporini[a], Marco Rabuffetti[b], Marcela Cristina de Moraes[c], Giovanna Speranza[b], Gabriella Massolini[a]

[a] *Department of Drug Sciences, University of Pavia, Viale Taramelli 12, 27100 Pavia, Italy*
[b] *Department of Chemistry, University of Milan, Via Golgi 19, 20133 Milano, Italy*
[c] *Department of Organic Chemistry, Institute of Chemistry, Fluminense Federal University, Outeiro de São João Batista s/n, 24020-141 Niterói, Rio de Janeiro, Brazil*

Abstract

Purine nucleoside phosphorylase (PNP) from *Mycobacterium tuberculosis* is a promising target to design effective agents against resistant and latent tuberculosis forms, representing a growing issue for global health. The present paper describes the development of a new enzymatic assay towards mycobacterial and human PNPs to discover selective inhibitors. Enzymatic reactions were conducted in batch and activity was assessed monitoring the phosphorolysis of inosine to hypoxanthine by hydrophilic interaction liquid chromatography. A small library of 6- and 8-substituted nucleosides was screened and the inhibition potency of the most interesting compounds was quantified through K_i and IC_{50} determination. The selected inhibitors were also investigated by *in vitro* biological tests.

Abbreviations

6-ClIno: 6-chloroinosine
6-i-PrOxIno: 6-i-propoxyinosine
8-BrIno: 8-bromoinosine
8-i-PrNHIno: 8-i-propylaminoinosine
8-NHIno: 8-aminoinosine
8-NHMeIno: 8-methylaminoinosine
8-N,N-2MeNHIno: 8-N,N-dimethylaminoinosine
ACN: Acetonitrile
Acv: Acyclovir
dAdo: 2'-deoxyadenosine
DMSO: Dimethylsulfoxide
HILIC: Hydrophilic Interaction Liquid Chromatography
Hpx: Hypoxanthine
*Hs*PNP: Human Purine Nucleoside Phosphorylase
Ino: Inosine

Mtb: *Mycobacterium tuberculosis*

*Mtb*PNP: *Mycobacterium tuberculosis* Purine Nucleoside Phosphorylase

N-i-PrA: N-i-propyladenosine

N,N-2MeA: N,N-dimethyladenosine

PNP: Purine Nucleoside Phosphorylase

TB: Tuberculosis

1. Introduction

The outbreak of drug-resistant tuberculosis (TB) is a growing public health concern in several countries, with multi, extensively and totally drug-resistant *Mycobacterium tuberculosis* (*Mtb*) strains spreading worldwide [1-3]. The treatment success rates of these forms of TB are very low, especially for extensively drug-resistant cases (28% in 2013) [1]. Therefore, new strategies and new molecular targets to design effective agents against this infectious disease are needed.

Enzymes hold a leading position among the available drug targets, since they play a fundamental role in all life processes, their activity is generally altered in pathological conditions and their peculiar structure makes them “druggable” (they possess active sites particularly suited for high-affinity binding of small molecules) [4, 5].

In *Mtb*, the purine salvage pathway appears to be an interesting source of enzyme targets for the design of antitubercular drugs because it is involved in the synthesis of nucleic acids, essential for life. Moreover, enzymes implicated in nucleotide metabolism usually differ from those of the human host, thus favouring the development of selective inhibitors. Purine nucleoside phosphorylase (PNP) attracts particular interest since evidences suggest the presence of differences in structure and transition state between mycobacterial (*Mtb*PNP) and human (*Hs*PNP) enzymes. Both *Mtb*PNP and *Hs*PNP are homotrimeric enzymes that catalyze the reversible cleavage of the glycosidic bond of purine (deoxy)ribonucleosides in presence of inorganic phosphate as a second substrate to generate the nucleobase and α-D-(deoxy)ribose-1-phosphate. However, from the comparison of the two three-dimensional structures, Tyr188 emerged as the only residue involved in the binding with the ligand not conserved in human (in which Phe200 takes its place). In addition, Phe153 in *Mtb*PNP seems to interact with the 5' of the substrate sugar group, while the corresponding Phe159 in human behaves like a lid for the binding site of the adjacent subunit [6-10].

The rational design of enzyme inhibitors usually starts from a screening (through radiometric, spectrophotometric or chromatographic assays) of chemical libraries consisting of potential drug candidates. An activity assay specific for the target enzyme is developed to initially compare the

inhibition percentage induced by different compounds, using a fixed concentration of enzyme, substrate and tested inhibitor. After this screening step, IC_{50} (inhibitor concentration that reduces the enzymatic activity to 50%) and K_i (enzyme-inhibitor dissociation constant) are determined for the most interesting molecules in order to quantify the inhibitor potency and the affinity for its target. The activity assay in this case is performed in presence of increasing amounts of inhibitors [4, 6].

In this work, a new enzymatic assay towards mycobacterial and human PNPs was developed to identify new selective inhibitors as potential antitubercular lead-candidates. Enzymatic reactions were conducted in batch and activity was determined by monitoring the phosphorolysis of inosine (Ino, substrate) to hypoxanthine (Hpx, product). To this aim, a new hydrophilic interaction liquid chromatography (HILIC)-UV method was developed. The activity assay allowed to characterize the kinetics of both PNPs, to screen ten purine analogues as potential inhibitors and to quantitatively assess the inhibitory activity of two nucleosides of interest. The two compounds were also tested *in vitro* on *Mtb* cultures to confirm data obtained using the isolated target.

2. Materials and methods

2.1. Reagents and chemicals

Inosine (Ino), Acyclovir (Acv, internal standard of Ino) and 2'-deoxyadenosine (dAdo, internal standard of Hpx) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Hypoxanthine (Hpx) was from Alfa Aesar (Haverhill, MA, USA).

8-methylaminoinosine (8-NHMeIno), 6-chloroinosine (6-ClIno), 8-bromoinosine (8-BrIno), 8-aminoinosine (8-NHIno), N,N-dimethyladenosine (N,N-2MeA), 8-N,N-dimethylaminoinosine (8-N,N-2MeNHIno), 8-i-propylaminoinosine (8-i-PrNHIno), N-i-propyladenosine (N-i-PrA) and 6-i-propoxyinosine (6-i-PrOxIno) were synthesized and kindly provided by Prof. Giovanna Speranza (Department of Chemistry, University of Milan) [6].

The two recombinant enzymes *Hs*PNP and *Mtb*PNP were produced in the Fluminense Federal University (Department of Organic Chemistry). *Hs*PNP activity was 0.57 IU/mg (stock solution: 0.4 mg/mL), while *Mtb*PNP activity was 0.10 IU/mg (stock solution: 1.5 mg/mL), as previously described [6].

Water was obtained from a Milli-Q® Integral system (Merck KGaA, Darmstadt, Germany).

Acetonitrile (ACN) HPLC grade, ammonium phosphate monobasic (≥99%) and ammonium formate (≥99%) were purchased from Sigma-Aldrich (St. Louis, MO, USA), while dimethylsulfoxide (DMSO) was from Carlo Erba Reagents (Cornaredo, Italy).

2.2. Instrumentation and chromatographic conditions

Chromatographic analyses were performed on an Agilent HPLC series 1100 system (Santa Clara, CA, USA), equipped with mobile-phase online degasser, quaternary pump, autosampler, column thermostated compartment and diode array detector. For data acquisition and analysis, the HP ChemStation for LC software version Rev. A.06.03 [509] was used.

A TSKgel Amide-80 (3 μm, 80 Å, 2 × 150 mm) column from Tosoh Bioscience (Montgomeryville, PA, USA) was employed. The mobile phase was composed of a mixture of ACN and 15 mM ammonium formate 92/8 (v/v), pH 7.0. Chromatographic conditions consisted of an isocratic elution for 17 min, a temperature of 25 °C, an injection volume of 10 μL and a constant flow rate of 300 μL/min. In all experiments, compounds were detected at 254 nm.

The linearity of the method was assessed for Ino and Hpx in the concentration range 0.5-50 μM. Eight concentration levels were considered and three independent determinations were performed for each point. A fixed amount (5 μM) of the internal standards dAdo and Acv was added to each sample. Calibration curves were obtained by the least squares method plotting Hpx/dAdo and Ino/Acv peak area ratios (y-axis) against Hpx and Ino concentrations (x-axis), respectively.

2.3. Sample preparation

For the calibration curve determination, stock solutions of 10 mM Ino (in water), Hpx, dAdo and Acv (in DMSO) were prepared. Working solutions (0.025 mM, 0.25 mM and 1 mM for Ino and Hpx, 0.25 mM for dAdo and Acv) were obtained by dilution of the stock solutions in water. The working solutions of compounds and internal standards were then combined and made up to a 500 μL final volume with ACN in order to reach the required concentrations.

For the enzymatic assay, 10 mM Ino (in water) and inhibitor (in DMSO) stock solutions were diluted to different concentrations with 10 mM ammonium phosphate, pH 7.5 to reach a volume of 480 μL. The addition of 20 μL of enzyme solutions, prepared by diluting 1:100 the *Hs*PNP and *Mtb*PNP stock solutions (see paragraph 2.1), allowed to obtain the final volume (500 μL) for the reaction.

2.4. Enzymatic activity assay

The procedure followed for the enzymatic assays was adapted to the available laboratory equipment starting from the one described by Cattaneo and colleagues [6].

Briefly, microcentrifuge tubes containing 480 µL solutions of Ino and phosphate buffer for the kinetic assay or Ino, phosphate buffer and inhibitor for the inhibition assay, were conditioned for 5 min at 37 °C under continuous stirring. After, 20 µL enzyme solutions (see paragraph 2.3) were added to each tube (0.08 µg enzyme/microcentrifuge tube for *Hs*PNP, 0.3 µg enzyme/microcentrifuge tube for *Mtb*PNP).

The reaction was run for 5 min at 37 °C under continuous stirring and subsequently stopped by pipetting 25 µL of the reaction mixture in a new microcentrifuge tube containing 475 µL ACN and the internal standards (1.05 µM of dAdo and Acv).

The obtained solutions were mixed by pipetting 150 µL up and down for three times and the inactivated enzyme was precipitated by centrifugation (5000 rpm for 10 min).

Finally, the supernatant was analyzed by HILIC-UV.

2.4.1. Kinetic studies

Ino kinetic constants were defined by performing the enzymatic assay in presence of increasing Ino concentrations (seven concentrations in the range 0.01-1 mM, mean of three replicates) and a fixed amount of P_i (10 mM ammonium phosphate, pH 7.5).

To calculate reaction rates for *Hs*PNP and *Mtb*PNP, µmol of produced Hpx were derived taking into account the Hpx/dAdo calibration line and were divided by the reaction time (5 min).

The Michaelis-Menten trend was obtained by plotting the rate of the enzymatic reaction, expressed as µmol of produced Hpx/min, against the substrate (Ino) concentration expressed as mM. The kinetic parameters V_{max} (maximal reaction rate, expressed as µmol/min), K_M (Michaelis-Menten constant, expressed as mM) and K_{cat} (turnover number, expressed as 1/sec) were calculated by Prism® software (GraphPad Software, San Diego, CA, USA) using a non-linear regression curve fitting. To define V_{max} and K_M, the Michaelis-Menten model $Y = V_{max}*X/(K_M+X)$ was chosen. The enzyme kinetics equation $Y = E_t*K_{cat}*X/(K_M + X)$ was used to determine K_{cat}. The term E_t represents the concentration of enzyme catalytic sites and it was calculated considering that both *Hs*PNP (MW = 96354 Da) and *Mtb*PNP (MW = 82716 Da) are homotrimeric.

2.4.2. Inhibition studies

The enzymatic assay was applied to the screening of ten nucleoside analogues (Acv, 8-NHMeIno, 6-ClIno, 8-BrIno, 8-NHIno, N,N-2MeA, 8-N,N-2MeNHIno, 8-i-PrNHIno, N-i-PrA and 6-i-PrOxIno) as potential PNP inhibitors. The assay was carried out using a high inhibitor concentration (0.5 mM) and fixed concentrations of Ino (close to the K_M level: 0.1 mM for *Hs*PNP, 0.075 mM for *Mtb*PNP) and P_i (10 mM ammonium phosphate, pH 7.5). *Mtb*PNP and *Hs*PNP residual activities (%) were calculated comparing the produced Hpx in presence of each inhibitor to the reaction carried out in the same conditions, but in absence of inhibitor (100% residual activity). Each resulting percentage is the mean of three replicates.

Inhibition constants (K_i) were assessed in the same conditions of the screening assay, except for the amount of inhibitors. Five concentrations of each inhibitor (0.01-1 mM, three replicates) were incubated with the two PNPs and reactions without inhibitor were carried out in parallel. Residual activities (%) of *Hs*PNP and *Mtb*PNP were calculated considering the Hpx/dAdo calibration line, the dilution factor and the assay volume (500 µL) and were plotted against the inhibitor concentration (M) in a logarithmic scale. K_i and IC_{50} (inhibitor concentration that gives 50% of the maximal response) parameters were calculated by Prism® software (GraphPad Software, San Diego, CA, USA) according to the One site - Fit K_i (1) and One site - Fit $LogIC_{50}$ (2) equations, respectively:

(1) $LogIC_{50} = Log(10^{LogK_i} \cdot (1+[Ino]/ InoK_M))$

(2) $Y = (Top-Bottom)/(1+10^{(X-LogIC_{50})}) + Bottom$

where K_i and IC_{50} are expressed in M, [Ino] is the constant concentration of substrate (M), $InoK_M$ is the Michaelis-Menten constant of the substrate (M), top and bottom represent the plateaus reached on the y-axis (%).

2.5. In vitro biological evaluation

The *in vitro* biological evaluation of the inhibitory activity was performed in 96-well U-bottom polystyrene microplates. Isoniazid (control drug) and solutions of the compounds were prepared at 5 mg/mL concentration in neat DMSO. Solutions were diluted in Middlebrook 7H9 medium containing 10% ADC (albumin, dextrose and catalase) to a concentration of 100 µg/mL. Growth controls containing no antibiotic and sterility controls without inoculation were included.

The inhibitory activity was determined for *Mtb* H37Rv strain, which was grown in Middlebrook 7H9 containing 10% OADC (oleic acid, albumin, dextrose and catalase) and 0.05% Tween® 80. Cells were vortexed with sterile glass beads (4 mm) for 5 min to disrupt clamps and allowed to settle for

20 min. The supernatant was measured spectrophotometrically at a wavelength of 600 nm. The *Mtb* suspension was aliquoted and stored at -20 °C. Each suspension was appropriately diluted in Middlebrook 7H9 broth containing 10% ADC to achieve an optical density of 0.006 at 600 nm and 100 µL were added to each well of the plate except for sterility controls. A final concentration of 2.5% DMSO was maintained in each well. The plates were covered, sealed with parafilm, and incubated at 37 °C. After 7 days of incubation, 60 µL of 0.01% resazurin solution were added to each well and incubated for additional 48 hours at 37 °C [11]. A change in color from blue to pink indicates the growth of bacteria, therefore the inhibitory activity was evaluated by the prevention of the color change. Three tests were carried out independently and inhibitory values reported were observed in at least two assays.

3. Results and discussion

3.1. HILIC-UV method development

A new HILIC-UV method was developed to monitor the phosphorolysis of Ino to Hpx and consequently *Hs*PNP and *Mtb*PNP activity under different conditions.

The chromatographic conditions that allowed to properly separate and quantify Ino, Hpx and their internal standards (Acv and dAdo, respectively) required the use of a carbamoyl TSKgel Amide-80 column and a mobile phase composed of 92/8 (v/v) ACN/15 mM ammonium formate buffer, pH 7.0, as described in paragraph 2.2. **Figure 1** shows the chromatographic trace obtained applying the developed method.

The linearity of the HILIC method was established for both the enzymatic substrate (Ino) and product (Hpx) in the concentration range 0.5-50 µM. Calibration curves of y=0.1192x+0.0601 (R^2=0.9987) and y=0.1225x+0.0379 (R^2 = 0.9984) were obtained for Ino/Acv and Hpx/dAdo, respectively.

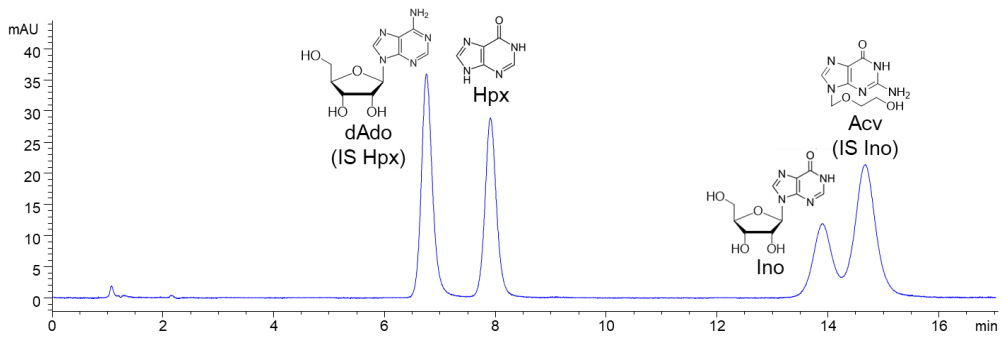


Figure 1. Representative chromatogram of Ino, Hpx and their internal standards (IS) obtained applying the conditions described in paragraph 2.2.

3.2. Enzymatic activity assay

Kinetic and inhibition enzymatic assays were performed following the same procedure, as described in paragraph 2.4.

The time for enzymatic reactions was fixed at 5 minutes, after which the enzyme was inactivated by dilution of the reaction mixture in ACN containing internal standards.

Before the quantification of the produced Hpx by HILIC-UV, the denatured enzyme was separated from the supernatant through centrifugation.

3.2.1. Kinetic studies

Before the investigation of potential enzyme inhibitors, it is necessary to characterize the target enzyme by defining its activity in absence of inhibitor. Usually, enzymatic activity assays quantify the product formation or the substrate consumption over time to define reaction rate [4]. The kinetic properties of the majority of enzymes can be described by the Michaelis-Menten equation:

$$v = \frac{V_{max}\,[S]}{K_M + [S]}$$

where v is the reaction rate, V_{max} is the maximum reaction rate, [S] represents the concentration of the enzymatic substrate and K_M or Michaelis-Menten constant is the substrate concentration that allows to reach half of the V_{max}. Another important kinetic parameter to assess enzyme activity is the turnover number K_{cat}, which expresses the number of substrate molecules converted into product per catalytic site per unit of time and is given by the ratio between enzyme concentration and V_{max} [4].

To investigate *Hs*PNP and *Mtb*PNP activities, the two enzymes were incubated with increasing amounts of Ino (0.01-1 mM, three replicates for each concentration). Since the reaction catalyzed by PNPs requires the presence of inorganic phosphate as a second substrate, the assay was performed in 10 mM ammonium phosphate buffer, pH 7.5, used as a source of P_i in saturating concentration.

The resulting kinetic constants are reported in **Table 1**, while Michaelis-Menten kinetics of the two enzymes are shown in **Figure 2**. The obtained values are consistent and of the same order of magnitude compared to those reported by Cattaneo *et al.* [6].

Table 1. Kinetic parameters calculated for *Hs*PNP and *Mtb*PNP using Ino as a substrate.

Enzyme	K_M (mM)	V_{max} (µmol/min)	K_{cat} (1/sec)
*Hs*PNP	0.1267 ± 0.0213	0.0078 ± 0.0004	52.52 ± 3.00
*Mtb*PNP	0.0752 ± 0.0129	0.0072 ± 0.0004	11.01 ± 0.56

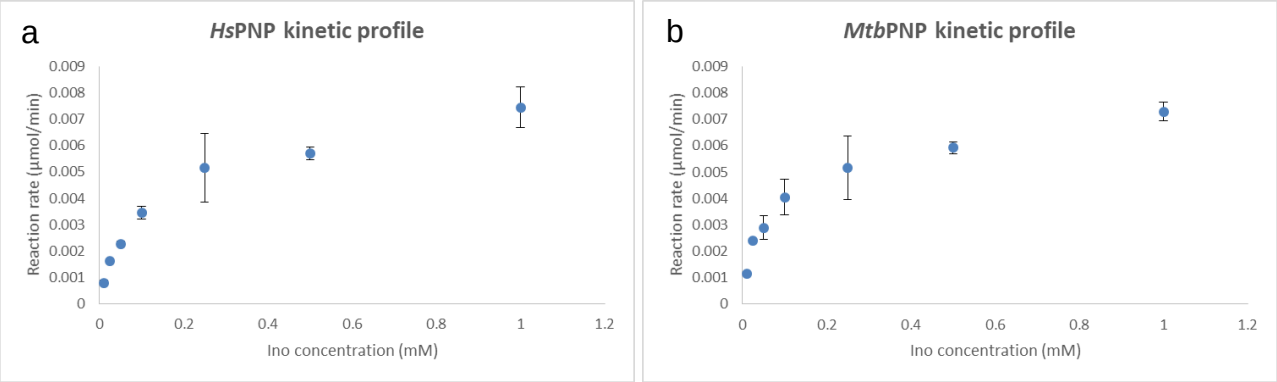


Figure 2. Kinetic profiles of *Hs*PNP (a) and *Mtb*PNP (b) obtained incubating the enzymes with increasing concentrations of Ino. Each point is the mean of three replicates, with error bars indicating ± standard deviation.

3.2.2. Inhibition studies

Following the results obtained in a previous work of our research group [6], some 6- and 8-substituted nucleosides (**Figure 3**) were selected and tested as potential PNP inhibitors.

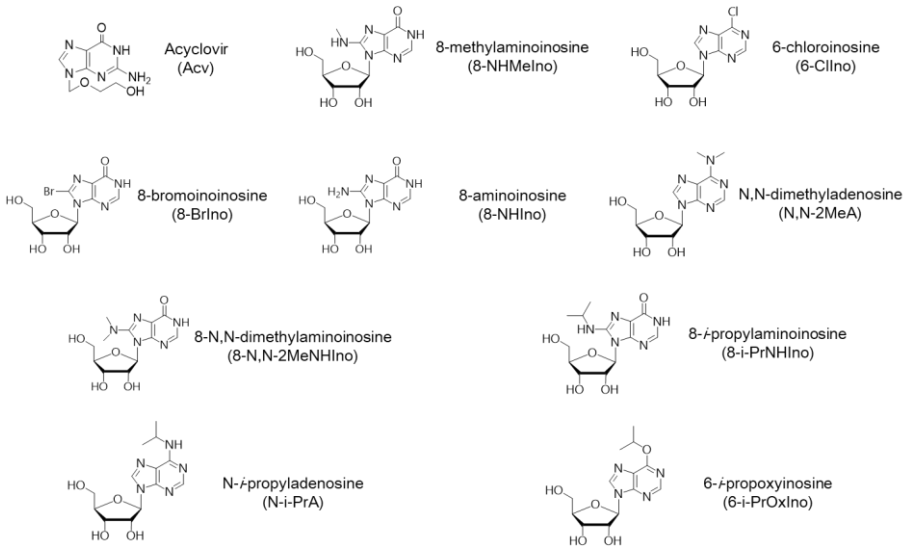


Figure 3. Structures of the nucleoside analogues tested as potential inhibitors of PNP enzymes.

The inhibition induced by the different potential lead-candidates was initially determined by performing a screening assay. Thus, the developed enzymatic activity assay was carried out at fixed concentrations of the two substrates (close to the K_M for Ino, saturating for P_i) and of the tested compounds (0.5 mM).

Since the aim of the present study is the identification of selective inhibitors towards *Mtb*PNP as a starting point to develop new antimycobacterial drugs, the screening assay was carried out in parallel on human and mycobacterial enzymes. The activity of the two enzymes was calculated in presence and in absence of each compound and inhibition percentages (expressed as enzyme residual activity) were compared.

The majority of the tested nucleosides did not show any selectivity for the PNP from *Mtb* (**Figure 4**), giving comparable residual activities of the two enzymes (6-ClIno, 8-BrIno, 8-i-PrNHIno, 6-i-PrOxIno) or showing selectivity for *Hs*PNP (Acv, 8-NHMeIno, 8-NHIno, N,N-2MeA, 8-N,N-2MeNHIno). The only compound that appeared to have a selective inhibitory activity for mycobacterial PNP was N-i-PrA, resulting in a residual activity of 73.40 ± 1.86% for *Mtb*PNP and 84.11 ± 2.67% for *Hs*PNP (mean of three independent determinations). N-i-PrA was therefore selected for further investigations on inhibitor potency and affinity for its target through K_i and IC_{50} determination. 8-NHIno was also considered a valuable candidate for additional studies since it exhibited a very high inhibitory activity, with 43.18 ± 2.09% *Mtb*PNP and 29.26 ± 1.75% *Hs*PNP residual activities.

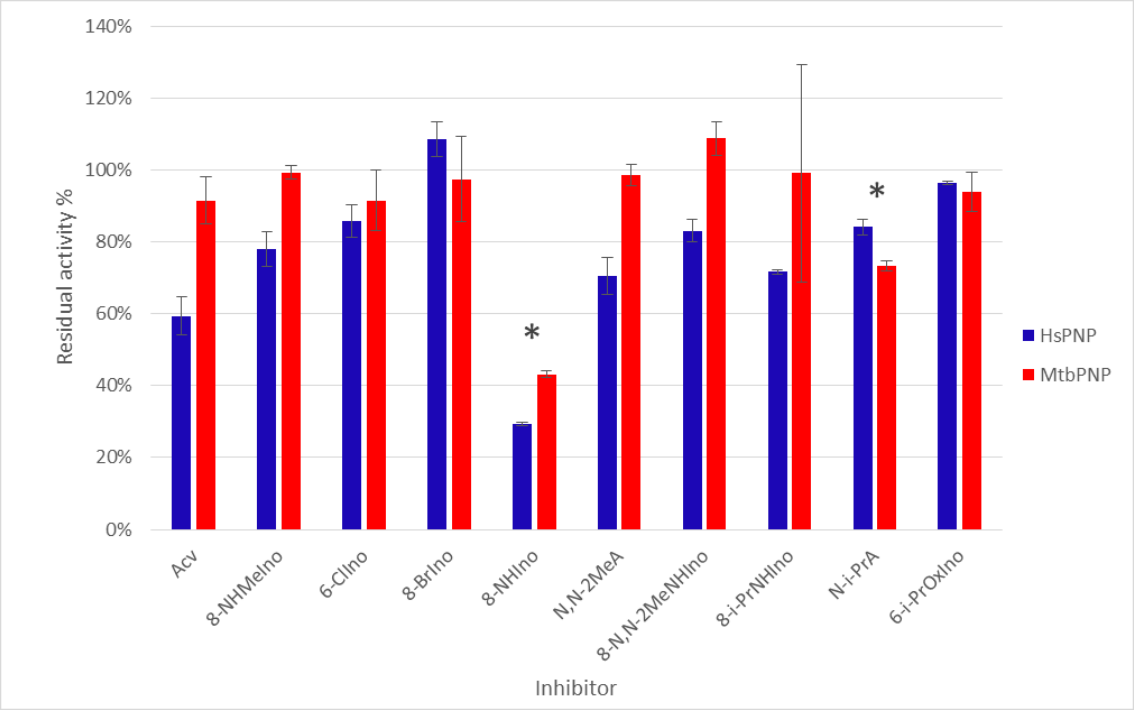


Figure 4. Residual activities of *Hs*PNP and *Mtb*PNP after incubation with 0.5 mM of each potential inhibitor (a residual activity of 100% corresponds to the produced Hpx in absence of the tested compound). Results derive from the mean of three independent determinations, with standard deviation indicated by error bars. Nucleosides selected for further investigations (K_i and IC_{50} determination) are identified by * symbol.

K_i measures the enzyme-inhibitor dissociation constant, while IC_{50} expresses the inhibitor concentration required to obtain a 50% reduction in enzyme activity. Both K_i and IC_{50} quantify inhibitor potency, but usually K_i is preferred because, unlike IC_{50}, it is independent of substrate identity and concentration [4].

For K_i and IC_{50} determination, the activity assay was performed in the same conditions used for the kinetic and screening assays, but with increasing inhibitor concentrations (0.01-1 mM, three

replicates). Residual activities of the two PNPs were calculated by comparing the results obtained in presence and in absence of each nucleoside. The two inhibition parameters were calculated by fitting the curve of inhibitor concentrations (Log(M)) *versus* residual activities (%) through Prism® software (paragraph 2.4.2).

Data (**Table 2**) confirmed the information deduced by the screening assay and provided a quantitative assessment of the inhibitory activity of 8-aminoinosine and N-i-propyladenosine. The first inhibitor showed a high potency, with K_i and IC_{50} in the micromolar range. Investigations on N-i-PrA proved its selectivity towards *Mtb*PNP. Even if this nucleoside showed a low inhibitor potency, giving values in the millimolar range for *Mtb*PNP, a significant difference in the activity towards the two enzymes was observed. In fact, its effectiveness in the inhibition of the human enzyme was so low in the studied range that it was not possible to calculate K_i and IC_{50}. In addition, the highest tested concentration (1 mM) caused a reduction of enzymatic activities to 50.66 ± 0.89% and 67.22 ± 0.95% for mycobacterial and human PNPs, respectively.

Table 2. Inhibition results towards *Hs*PNP and *Mtb*PNP. * it was not possible to calculate the value of the parameter due to the low potency of the compound in the studied range.

Inhibitor	Enzyme	K_i (mM)	Log(K_i) Log(M)	IC_{50} (mM)	Log(IC_{50}) Log(M)
8-NHIno	*Hs*PNP	0.0481	-4.32 ± 0.0539	0.0860	-4.06 ± 0.0539
	*Mtb*PNP	0.1319	-3.88 ± 0.0687	0.2635	-3.58 ± 0.0687
N-i-PrA	*Hs*PNP	not determined*	not determined*	not determined*	not determined*
	*Mtb*PNP	2.40	-2.62 ± 0.1693	4.79	-2.32 ± 0.1693

3.3. In vitro biological evaluation

Since 8-NHIno and N-i-PrA showed promising results on the inhibition of isolated *Mtb*PNP and *Hs*PNP, the biological activity of the two nucleosides was evaluated by *in vitro* assays on mycobacterial cultures.

Compounds were incubated with cultures of the most common strain of *Mtb* (H37Rv) for 7 days. The assays were carried out in triplicate and no significant inhibitory activity was observed in two non-consecutive days.

This negative result can be ascribed to the abundance of lipids in *Mtb* cell wall hindering the permeation of hydrophilic molecules. The difficult passage of the hydrophilic nucleosides prevents their interaction with the target, resulting in the survival of the bacillus. Drug delivery systems should be considered to allow the compounds to reach their target *in vitro* and *in vivo*.

4. Conclusions

An enzymatic activity assay towards *Mtb* and human purine nucleoside phosphorylases was developed and successfully applied to the screening of ten purine analogues as potential enzyme inhibitors, after a kinetic characterization of the two PNPs.

The aim of the work was the identification of molecules with a selective inhibitory activity towards *Mtb*PNP as powerful and innovative antitubercular lead-candidates. Only N-i-propyladenosine showed the desired selectivity and was therefore selected for a quantitative assessment of its inhibition potency. The developed enzymatic activity assay proved its suitability in the determination of inhibitor K_i and IC_{50}, confirming the observed selectivity. Further investigations were also carried out on 8-aminoinosine, which induced a strong reduction in the activity of both the tested PNPs.

The two nucleosides were subsequently tested by *in vitro* assays on mycobacterial cultures to confirm data obtained using the isolated target, but no significant reduction on cell viability was observed. This result was correlated to the lipophilicity of *Mtb* cell wall hampering the passage of the nucleosides and their interaction with the target enzyme. The use of suitable drug delivery systems might be useful to obtain the desired effect.

Molecular modelling studies were recently planned in order to understand on a molecular basis the selectivity and activity shown by the two candidates.

References

[1] World Health Organization, *Global tuberculosis report 2016*, Geneva, Switzerland, WHO Press, **2016**.

[2] Zumla A. et al., *Tuberculosis*, N Engl J Med. **2013**, 368(8), 745-55.

[3] Smith T. et al., *Molecular biology of drug resistance in Mycobacterium tuberculosis*, Curr Top Microbiol Immunol. **2013**, 374, 53-80.

[4] Copeland R.A., *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*, John Wiley & Sons, Inc. **2013**

[5] Copeland R.A. et al., *Targeting enzyme inhibitors in drug discovery*, Expert Opin Ther Targets. **2007**, 11(7), 967-978.

[6] Cattaneo G. et al., *Development, validation and application of a 96-well enzymatic assay based on LC-ESI-MS/MS quantification for the screening of selective inhibitors against Mycobacterium tuberculosis purine nucleoside phosphorylase*, Anal Chim Acta. **2016**, 943, 89-97.

[7] Ducati R.G. et al., *Purine Salvage Pathway in Mycobacterium tuberculosis*, Curr Med Chem. **2011**, 18(9), 1258-1275.

[8] Long M.C. et al., *Identification and characterization of a unique adenosine kinase from Mycobacterium tuberculosis*, J Bacteriol. **2003**, 185(22), 6548-6555.

[9] Shi W. et al., *Structures of purine nucleoside phosphorylase from Mycobacterium tuberculosis in complexes with immucillin-H and its pieces*, Biochemistry. **2001**, 40(28), 8204-8215.

[10] Caceres R.A. et al., *Crystal structure and molecular dynamics studies of purine nucleoside phosphorylase from Mycobacterium tuberculosis associated with acyclovir*, Biochimie. **2012**, 94(1), 155-165

[11] Palomino J.C. et al., *Resazurin microtiter assay plate: simple and inexpensive method for detection of drug resistance in Mycobacterium tuberculosis*, Antimicrob Agents Chemother. **2002**, 46(8), 2720-2722.

7. *Advances on size exclusion chromatography and applications on the analysis of protein biopharmaceuticals and protein aggregates: a mini review*

Gloria Brusotti, Enrica Calleri, Rafaella Colombo, Gabriella Massolini, Francesca Rinaldi, Caterina Temporini

*Chromatographia 2018, 81(1), 3-23**

Chromatographia (2018) 81:3–23
https://doi.org/10.1007/s10337-017-3380-5

REVIEW

Advances on Size Exclusion Chromatography and Applications on the Analysis of Protein Biopharmaceuticals and Protein Aggregates: A Mini Review

Gloria Brusotti[1] · Enrica Calleri[1] · Raffaella Colombo[1] · Gabriella Massolini[1] · Francesca Rinaldi[1] · Caterina Temporini[1]

Received: 8 May 2017 / Revised: 31 July 2017 / Accepted: 2 August 2017 / Published online: 8 August 2017

Abstract Biopharmaceutical drugs are large, complex protein molecules derived from living cells. The characterization of the protein-based products is a complicated and challenging task necessary to guarantee their efficacy and safety. A major concern in manufacturing protein biopharmaceuticals is their propensity to form aggregates. The separation of proteins by size-exclusion chromatography (SEC) is a well-established method for the characterization of macromolecules and their aggregates. The demand for increased resolution and faster analysis in SEC has led to the development of small, more rigid fully porous particles and superficially porous particles with different pore sizes. The coupling of SEC with three or four detectors has enhanced protein characterization by generating accurate molecular weights (absolute determination) and by giving insight into the degradation pathways and products associated with aggregation. The intent of this review is to report the continuous developments of SEC stationary phases and their applications on the characterization of intact proteins, their variants or aggregates and protein complexes. Thus, the most recent applications of SEC with the focus on proteins and protein aggregates are reported and discussed.

Keywords Size exclusion chromatography · Therapeutic proteins · Protein aggregates · Biopharmaceuticals

Published in the topical collection *Peptide and Protein Analysis* with Debby Mangelings and Gerhard K. E. Scriba as editors.

✉ Gabriella Massolini
gabriella.massolini@unipv.it

[1] Department of Drug Sciences, University of Pavia, Viale Taramelli 12, 27100 Pavia, Italy

Introduction

Pharmaceutical products obtained by biotechnological approaches constitute a wide group of drugs (monoclonal antibodies, vaccines, enzymes, cytokines etc.), endowed with high specificity and potency, used to treat important diseases such as cancer, inflammation, immune disorders, infections and cardiovascular diseases [1, 2]. In the last three decades the development of key technologies enabled recombinant DNA protein production to progress rapidly, leading to a drastic change in the progress of proteins as therapeutic agents. However, the large-scale production of protein-based drugs is a complex and demanding activity that includes inoculation expression (molecular biology), product fermentation (cell culture development), product recovery from cell culture and purification. During every step, the process control is crucial since even small changes can alter the native proteins structure. In addition, different environmental factors such as temperature, pH, salt type, salt concentration and mechanical stresses (filtration, agitation etc.) during the manufacturing process may trigger protein degradation and inactivation [3].

After the protein production, other steps such as formulation, packaging, shipping and storage may lead to various degradation processes and structure modifications in the lifecycle, possibly decreasing the potency of the protein-based drugs [4].

Protein degradation pathways can be divided in: (1) chemical modifications of protein, that include oxidation, deamidation, reduction, hydrolysis; and (2) physical modifications that involve unfolding, dissociation, denaturation, aggregation, and precipitation. In some cases, the two processes are synergistic and a chemical event may trigger a physical event [5, 6].

Biopharmaceutical drugs should be characterized at different levels, ranging from primary structure (i.e. amino

acid sequence and post-translational modification (PTM)) to conformation (secondary and tertiary structure) to correct assembly of multi-unit proteins (quaternary structure) in order to guarantee their efficacy and safety.

The characterization of the protein-based products is a complicated and challenging activity, due to the large size and complex structure of proteins, often characterized by micro heterogeneities commonly observed even in highly purified protein fractions [7–9].

A variety of established analytical techniques that can be successfully applied to the complete characterization of a given protein are already available [9–11]: (1) amino acid analysis or UV absorbance can be used for the quantitative determination of the protein concentration; (2) spectroscopic methods (X-ray, NMR, Circular Dichroism) are used to assess protein secondary and tertiary structure; (3) mass spectrometry is suitable for measuring the molecular weight of intact proteins by the "top down" approach, while the "bottom up" approach elucidates the protein sequence using N- and C-terminal sequencing [12, 13]. In addition, different liquid chromatographic methods (reversed phase (RP), size exclusion (SEC), ion exchange (IEC), hydrophilic-interaction chromatography (HILIC), hydrophobic interaction chromatography (HIC) etc. in 1 or 2 dimension systems) [14–16] and electrophoresis (gel-electrophoresis, capillary electrophoresis) coupled to different detection techniques (DAD, MS, Light scattering) may support a more detailed analysis of biopharmaceuticals as well as of product related impurities/degradants [17].

Among the different chromatographic techniques, size exclusion chromatography (SEC) can be considered the method of choice for characterizing proteins on the bases of molecular weight and/or molecular weight distribution [18, 19]. SEC is a non-destructive separation technique that can separate and quantify protein mixtures, and is therefore valuable for the quality control in recombinant protein production (i.e. to guide cell-line selection) and to direct the development of the purification process for biopharmaceuticals. SEC can also be used for the routine monitoring of aggregates in biopharmaceuticals as part of the manufacturing process development.

Indeed, one of the most common problem connected with physical modifications of protein-based biopharmaceuticals is the change in protein higher order structure leading to permanent (or transient) partial unfolding or to aggregation. In the last years a lot of attention has been paid to protein aggregation as it can affect protein activity and solubility, or even cause undesired immune reactions [20, 21]. Thus, it is possible to state that for therapeutic proteins aggregation is an important concern and there is a considerable interest for the monitoring of protein aggregation in term of its detection and characterization [22]. Aggregates have been defined as reversible/irreversible, soluble/insoluble, covalent/non-covalent, small/large self-associated species ranging in size from nanometers to hundreds of micrometers [23].

Regulatory guidelines consider aggregates as process or product related impurities and clearly state that aggregates must be resolved from the desired product and quantified. Considering the increasing popularity of biopharmaceutical drugs, there is a need of analytical methods for the characterization of proteins aggregation during the long, complex and costly development process.

Structural characterization of protein aggregates can be challenging due to the broad aggregate size ranges, their different physico-chemical properties and lifetime. In addition, during sample preparation and/or analysis it is possible to destroy or even create some aggregates.

A number of complementary and orthogonal techniques can be used for the profiling of protein aggregation (SDS-PAGE, capillary electrophoresis, light scattering based methods, analytical ultracentrifugation, flow-field-flow fractionation, high performance size exclusion chromatography) [24–27]. SEC is the favoured method for routine and validated analysis of aggregates. The separation of aggregates with SEC can be performed in non-denaturing elution conditions and the method provide sensitive, reproducible and fast information on presence of aggregates. However, SEC requires careful method development to ensure accuracy of aggregate forms present in a sample due to filtration or non-specific binding of the high molecular weight species by the column. Pros and Cons of SEC are reported in Table 1.

This review is not aimed at providing a comprehensive literature review in the area of SEC, as recent exhaustive reviews have been published in the last years [19, 28, 29] and the readers are encouraged to consult the references especially for going into the details on the theory of SEC. This review paper focuses on the new developments in the area of SEC stationary phases and on the more used and new detectors. The recent applications of SEC (covering the last 3 years) in the analysis of proteins and protein aggregates are also reported to give to the reader an updated overview on innovative approaches in the field of biopharmaceutical.

Size Exclusion Mechanism

Size exclusion chromatography separates molecules according to their physical size in solution (hydrodynamic radius) by interacting with a porous structure.

The elution order typically follows molecular weight and molecules with the highest molecular weight elute first because they are excluded from the pores. Smaller molecules, which are able to access pores within the resin particles, permeate a larger accessible volume within the column and are eluted later. Since protein aggregates are much larger

Table 1 Pros and cons of SEC

Pros of SEC	Cons of SEC
Simplicity of the physical separation mechanism (molecular size and shape)	Stationary phase can act as a filter
Easy to use to validate	Limited dynamic range
High speed and precision in separation	Nonspecific interactions of the biopharmaceutical and its aggregates with the chromatographic material, column hardware, especially the column frits
Good variety of column from different manufacturers	Inaccuracy due to alteration of size distribution, in particular for reversible aggregates
Low levels of aggregation can be measured with very small amounts of material	Few columns type for separating aggregates larger than a few MDa Complex MS-coupling

(2× larger or more) than the native protein, they are expected to elute earlier.

To achieve sufficient differentiation in residence time in the column, the SEC media must be porous and have a high pore volume. Unlike other modes of chromatography, SEC relies on the absence of any interaction between the analyte and the stationary phase packed in the column thus separation can be achieved in isocratic conditions.

A complete guide to the theories, methods with detailed descriptions of techniques, and applications of size-exclusion chromatography can be found in the second Edition of Modern Size-Exclusion Chromatography [30, 31].

SEC is an ideal solution for separating and analysing intact proteins from contaminants that can include aggregates, excipients, cell debris, and other impurities arising from degradation with minimal sample preparation. In addition, SEC can be used for the estimation of molecular mass using calibration standards or absolute methods (light scattering) or for sample clean up. We are not discussing the development of SEC method (mobile phase selection, flow-rate, temperature, calibration methods etc.), instead we will briefly describe the technological evolution of SEC stationary phases together with some new advances on instrumentation.

SEC Stationary Phases

The first material used for the separation of proteins and peptides was starch; the low mechanical strength and the poorly defined structure of this material lead to the preparation of new supports based on cross-linked dextran gels commercialized with the brand name Sephadex (Pharmacia) or polyacrylamide gels (Bio-Rad) with controlled porosity. However, these supports were limited in mechanical performance. In the same years, Moore at Dow Chemical Co., developed the synthesis of rigid cross-linked polystyrene–divinylbenzene (PS–DVB) gel to be used as column packing (Styragel).

Developments in modern high performance SEC rely on small, more rigid porous particle that can be divided in organic (polymeric) or inorganic (silica with or without surface modification) materials. A complete list of the commercialized SEC stationary phases has been reported by Saunders and Barth together with a systematic and exhaustive discussion of the state of the art of SEC columns [32].

The SEC column can be optimized in terms of surface chemistry (non-specific adsorption), specific resolution (pore size and volume), efficiency (particle size) and column stability (mechanical, chemical and thermal). A summary of the most important characteristics of SEC stationary and mobile phases for better separations is given in Table 2. Users can consider these properties for special applications such as proteins and aggregates analysis.

In general, silica material has the advantage of being rigid, which is beneficial for high pressure separation and it can be used with a much wider variety of mobile phases than polymer-based stationary phases. An important aspect to be considered is the surface chemistry of SEC materials: an ideal support should be inert in order to obtain minimal adsorption of proteins and protein aggregates leading to more accurate quantitation of molecular weights. In addition, for higher molecular weight analytes, stationary phases with higher pore size and greater pore volume can provide improved separations, thus, useful when dealing with multimeric protein aggregates or conjugated proteins.

The most common stationary phases consist of silica with chemically bonded 1,2-propanediol functional groups. This hydrophilic stationary phase, where the acidic silanol groups are derivatized, is neutral and it is a good choice for the separation of biopolymers and synthetic water-soluble polymers. At present diol phase is the most used silica-surface for SEC of proteins and their aggregates. These materials minimize any secondary interaction between the stationary phase and the protein analytes, thus leading to separation based primarily on the ability of the proteins to permeate the porous silica particle.

However, 1,2-propanediol phase is not a good choice for basic polymers since it will ion exchange onto residual

Table 2 Parameters of SEC stationary and mobile phases and their effect on the chromatographic performances

Property	Effect
Chemical composition of packing	Reduced non-specific adsorption Improved MW accuracy
Particle size and shape	Improved efficiency and resolution Reduced analysis time Increased shear-degradation of samples Increased formation of aggregates
Particle mechanical rigidity	Improved resistance at high pressures Improved suitability for various mobile phases
Pore size and volume	Improved efficiency, resolution, MW separation range Reduced analysis time
Interstitial volume	Improved exclusion limits
Column length and i.d. (not discussed)	Improved resolution
Flow rate Injected sample concentration	Reduced shear-degradation of samples
Mobile phase ionic strength and pH	Reduced non-specific adsorption

acidic silanol groups or even diol-modified silica. Differently, anionic polymers will be excluded from negatively charged pores. The mobile phase parameters such as ionic strength and pH can influence the adsorption of the proteins to the stationary phase. Working at pH values near the isoelectric point (pI) of the protein can minimise the adsorption as well as the use of a significant amount of electrolyte added to the mobile phase.

Silica is also a useful packing material for the analysis of nonionic polymers with organic mobile phases (polar or nonpolar), also for high-temperature applications. On the contrary silica is not recommended for aqueous mobile phases because of the solubility of silica in aqueous buffers, especially at elevated temperatures. Furthermore, silica packings can undergoes hydrolysis.

The newest type of silica-related packing is an ethylene-bridged hybrid inorganic–organic (BEH) support commercialized by Waters Co. [33]. This packing is a mixed composition of silica and organosiloxanes which form poly-ethoxyoligosilane polymers. Compared to 100% silica packings, BEH particles have improved chemical stability, reduced silanol reactivity and acidity. In fact BEH column exhibits less adsorption of the basic protein lysozyme (pI ~ 11) at pH 7, this is a classical test for acidic silanols, when compared to a commercially available silica SEC column. BEH material has been prepared with controlled different pore sizes (125–900 Å) and pore volume (0.3–1.6 cm^3/g) without losing the mechanical rigidity [19, 34].

The polymeric materials for the separation of nonpolar (hydrophobic) polymers is cross-linked poly-styrene-co-divinylbenzene (or polydivinyl benzene), introduced in 1964 by Moore. In this material the degree of cross-linking, which in turn influences the pore size, can be adjusted by carefully controlling the polymerization conditions. Different hydrophilic cross-linked packings developed for the SEC of biopolymers and synthetic water-soluble polymers are now commercially available.

Most of these packings are hydroxylated derivatives of cross-linked polymethacrylates. Unusual polymeric packings for aqueous SEC include sulfonated cross-linked polystyrene, polydivinylbenzene derivatized with glucose or anion-exchange groups, a polyamide polymer, and Novarose, a high-performance crossed-linked agarose.

Fully porous 5–10 μm particles (FPP) have been exclusively used until 2000. For FPP the interstitial volume is much larger than for non-porous particle thus increasing the accessible molar mass (M) of the biopolymers. Modern supports have reduced average particle diameter to sub-2 μm to enhance column efficiency and to gain faster analysis [35]. In general, resolution between the peaks of aggregates and native protein is improved using smaller particle size. However, reducing the particle size into the 2 μm range increases the risk of shear-degradation of biopolymers and their aggregates. In particular, the creation of low-molecular weight (MW) fragments from high MW biopolymers or the breaking of aggregates can be observed when large shear force is applied. Shear stresses generated at high flow rates may also cause deformation of flexible proteins thus their Stokes radii do not correlate exactly with molecular weight. In addition, the sintered frit porosity and inner diameter of capillary connections must be sufficiently large to prevent polymer shear effects. On the other site the reduction of the particle size increases the risk of creating on-column aggregates, when analysing sensitive proteins, due to the high pressure conditions.

These limitations can be overcome by an accurate dimensional comparison of the pores, the interstitial volume and the size of the aggregates or proteins. The reduction of flow rate as well as minimal injected sample concentration also can help to eliminate biopolymer shear degradation.

Meanwhile, highly efficient core–shell packing materials were developed. These new materials show a useful opportunity for fast and efficient separations of proteins. Superficially porous particles (SPPs) (core–shell particles) have been developed for SEC chromatography. Although these materials have reduced total pore volume (40% less pore volume), SPPs have kinetic advantages that compensate for the loss in pore volume compared to FPPs.

Standard SPPs have pore diameters of about 100 Å designed for separating small molecules. However, SPPs with 160–200 Å pore diameter have also been made available for protein separation [36, 37]. Generally, the core–shell particles work over a smaller range of molecular weights than the porous particles. Very recently the feasibility of using core–shell particles in SEC has been investigated and core–shell materials were compared with conventional packing materials with similar particle diameters and pore sizes using equal flow rates [38]. Polystyrene molecular weight standards were used and the analyses were carried out on a standard HPLC equipment with tetrahydrofuran as mobile phase. The parameters used for comparison include the elution range (smaller with SPPs), efficiency (larger for SPPs), the rate production of theoretical plates (larger for SPPs) and the specific resolution (larger with FPPs). The Authors conclude that SEC using SPPs offers similar peak capacities to that using FPPs but with shorter time analysis.

Figure 1 reports chromatograms of the solutes run with SPPs and FPPs. The peaks are overlaid for specific molecular weight analytes run on 160 Å SPP and 200 Å FPP pore size materials (Fig. 1a, c respectively). The elution curves are overlaid for the larger pore (1000 Å) SPP and FPP materials (Fig. 1b, d respectively).

SPPs packing material was also found particularly useful in fast SEC high-throughput applications, achieving high SEC resolution within very short analysis times [39]. The experiments were carried out with a standard HPLC. The reduced pore volume in comparison to fully porous particles is compensated by an increase in efficiency. Moreover, while conventional SEC columns perform more evenly across a broader range of analyte molar masses, the investigated core–shell columns exhibited excellent performance across more limited ranges, depending on the pore size. Analysis times of 0.4 min or less were obtained. The Authors suggest that these results are very promising for application as a second-dimension separation in an LC × LC separation system.

A recent paper describes the development and properties of superficially porous silica particles (fused core silica particles) with very large pores (1000 Å) specifically developed for separating very large biomolecules and polymers without restricted diffusion. Separations of DNA fragments, monoclonal antibodies, large proteins and large polystyrene standards were used to illustrate the utility of these particles in minimizing mass transport for efficient, high-resolution applications [40]. The separation of specific polystyrene standards on a 400 Å SPP column and 1000 Å SPP column is shown in Fig. 2. The largest MW standard is excluded by both pore systems. The 1000 Å pore material shows high resolution in a short amount of time. Separations of large biomolecules confirm the advantages of particles with very large pores in minimizing mass transport effects for better resolution.

Ultra-High Performance Size-Exclusion Chromatography (UHPSEC)

The introduction of small particle size packing materials (2 µm and less) improved dramatically the resolution, sensitivity and throughput provided by liquid chromatography [41]. In 2004, Waters presented the first LC equipment able to work with reduced sub-2 µm particles, the equipment was named ultra-performance or ultra-high performance liquid chromatography (UHPLC).

Modern SEC has also benefited from the improvement of stationary phase materials using low dispersion UHPLC instrumentation [29, 42, 43].

Since the introduction of SEC, columns with 300 × 6–8 mm ID packed with 5–10 µm particles have been typically used. With the advent of UHPSEC, columns packed with 1.7, 1.8 or 2.0 µm particles (pore sizes of 125–900 Å) in 150 × 4.6 mm ID dimensions offer new opportunities to improve both the resolution and throughput. It has been shown that UPSEC can shorten the analysis time typically into 3–5 min range [44]. However, the large amount of heat generated by friction in UHPSEC columns may affect analytes with temperature-dependent properties.

Indeed, the speed of analysis is critical when screening multiple samples, such as when looking at cell culture optimization or process variables on the level of aggregation. However, when SEC is applied for the analysis and quantitation of monomers, dimers, and multimers of proteins, there is a risk of shear forces causing changes to the sample composition or the column acting as a filter and trapping the multimers.

In a recent article, practical possibilities and limitations of recently introduced SEC columns of 150 × 4.6 mm ID, with sub–3 µm packing materials for the separation of model proteins biopharmaceutical were evaluated in a UPLC system. Columns packed with 1.7–2.0 µm particles improved the plate count in agreement with theoretical expectations. In addition, possible secondary hydrophobic and/or electrostatic interactions between the SEC stationary phases and biopharmaceutical proteins were systematically studied. Significant differences in nonspecific interactions were observed, with hydrophobic interactions generally exerting more influence than electrostatic interactions [45].

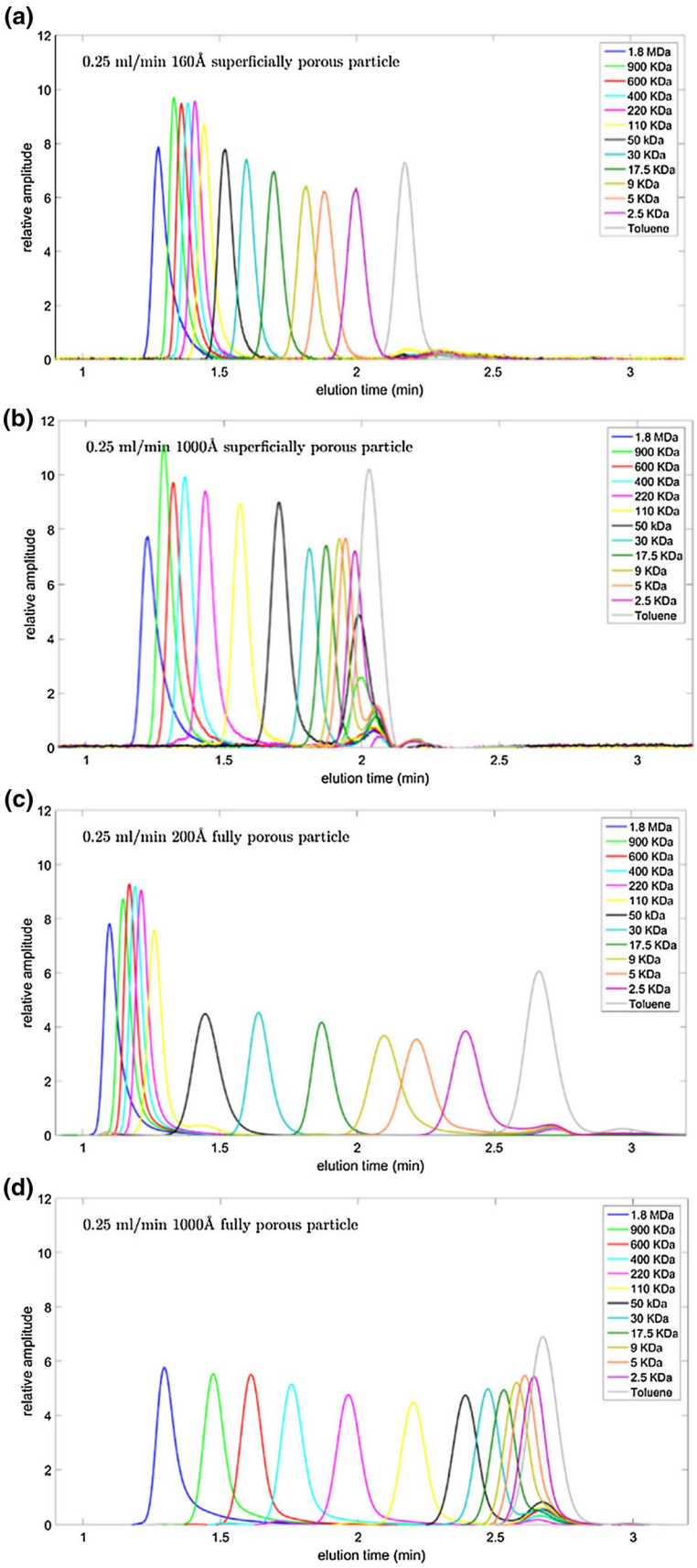


Fig. 1 Chromatographic elution of polystyrene solutes and toluene superimposed on the time axis. **a** SPPs with pore size of 160 Å, **b** SPPs with pore size of 1000 Å, **c** FPPs with pore size 200 Å, **d** FPPs with pore size of 1000 Å. Flow rate of 0.25 mL/min [38]

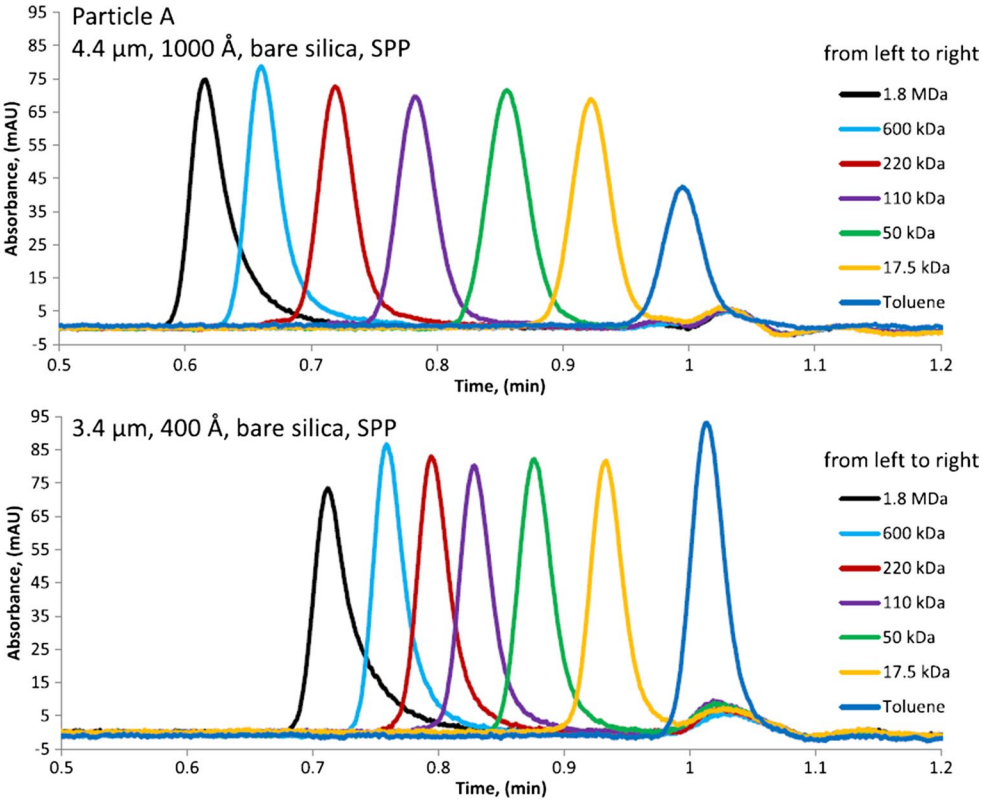


Fig. 2 Separation of polystyrene standards with the same molecular weight on (*top*) 1000 Å silica particles and (*bottom*) 400 Å silica particles [40]

An interesting study has been performed to evaluate the loss of efficiency observed with three different instruments (regular HPLC, non-optimized UHPLC and fully optimized UHPLC) [46]. It was demonstrated that the new 150 × 4.6 mm ID, sub–3 µm SEC columns cannot be employed on a regular HPLC instrument, since the efficiency loss was approximately 60–85%, when analysing mAb sample. The authors conclude that the apparent efficiency of SEC columns packed with sub–2 µm particles can indeed be hampered when using inappropriate system. Considering the extra-column band broadening contribution of current UHPLC instruments, a further decrease of SEC column dimension is therefore not desired.

SEC Methods and Detection Modes in the Characterization of Biopharmaceuticals

Protein aggregation can affect both the economic viability of a biopharmaceutical product and its efficacy. In fact, the presence of aggregates is predictive for immunogenicity of therapeutic proteins. However, it is not clear what type and size of aggregates are inducing antibodies. Thus, optimal production, purification and formulation is necessary to prevent an immunogenic response.

The definition of the presence of size variants and aggregation in biopharmaceuticals is a complex analytical issue which requires the combination of different techniques to gain comprehensive information about the formation of protein aggregates and their behaviour [47]. SEC, dynamic light scattering (DLS) and electron microscopy (EM) represent the more established techniques for the characterization of proteins and their aggregates [26]. Only recently, a new technique, Water Proton NMR (wNMR), has been proposed to monitor a wide size range of aggregates in biopharmaceuticals. wNMR seems to overcome the size limits of SEC in detecting large micron-sized or insoluble aggregates, even if further investigations are needed [48]. Nowadays SEC remains one of the most conventional and frequently employed method to measure protein mass, size, homogeneity and aggregation during the development, the quality control and the stability of biopharmaceuticals. This technique is suitable for the analysis of aggregates in the size range of 1–50 nm and it is considered robust and accurate, when the method induces no changes to the aggregation state and nonspecific interactions with the column packing media are eliminated.

SEC analysis is a two-stage process in which protein samples are firstly separated on the bases of their hydrodynamic volumes and then revealed using different detectors: UV, refractive index (RI), MS, viscometer, static and dynamic light scattering (SLS and DLS). SEC in combination with UV detection measures the concentration of chromophore-containing species to reveal the percentage of aggregation.

Although sensitivity is not very high, SEC/UV is still the most used approach for quantitative determinations.

High performance size exclusion chromatography (HP-SEC) with LS detection, in one or more of its formats (static, dynamic, triple angle, multiangle, right angle, and so forth) allows to obtain moderate accurate molecular mass results [49, 50]. These detectors are very useful during the protein-based product development for demonstrating the presence or absence of protein complexes, as well as unwanted aggregates of the monomeric species.

To achieve more precise and accurate data, MS-based methods can be used to gain insight in identity, MW, modifications type and site [51]. Furthermore, the additional dimension offered by MS analysis allows the identification of individual components in aggregate peak. However, interfacing SEC with MS may be difficult due to mobile phases with high concentration of non-volatile salts. In the last years, many efforts have been paid to the search for near-native elution conditions to promote the online combination of SEC and MS, gaining implemented information on the separated intact proteins, their variants or aggregates and protein complexes by their MW determination [51].

MS or LS detection should be considered as complementary techniques; MS is more suitable to acquire very specific details about a limited number of individual species of not too high MW, while LS is well suited for polydisperse high M samples. Another advantage is that LS is very sensitive to high molar masses, even at low concentrations. LS can often detect higher aggregates with higher sensitivity than any other detector.

In general, SEC/UV and SEC/MS are the most used tools to have information on protein purity, impurities and also on fragmentation and aggregation. The combination of UV and MS detection allows not only to characterize protein samples and their structural variants or fragments and aggregates, but also to carry out release testing and stability monitoring as it facilitates the quantification of aggregates [17]. However, viscometers are becoming more and more common to identify structure changes such as denaturation and aggregation this because concentration-sensitive detector are highly precise but the accuracy cannot be so good. This is caused by calibration curves that are normally constructed using narrow-polydispersity linear standards that usually do not possess the same chemistry or architecture as the analyte.

New trends show a great interest for the combination of SEC with multiple detectors, including mainly UV, RI, LS, multiangle light scattering (MALS), multiangle laser light scattering (MALLS) and viscometer. The multi-detection approach is now possible due to the small dimensions of the flow cells which allow the analysis without peak broadening and sensitivity loss. The combination of SEC separation with three or four detectors can enhance protein characterization by generating accurate molecular weights and by giving insight into the degradation pathways and products associated with aggregation.

In the next section we report some new trends in SEC applications published in the last three years using different detectors for the characterization of proteins, large protein aggregates and small protein aggregates and oligomers. A critical summary of the applications, including column type, detection method and type of analysed sample is reported in Table 3.

Applications of SEC to the Characterization of Proteins

Conventional SEC columns packed with 3 or 5 µm with small pores (150 Å) or wide pores (300 Å) are typically used for small proteins (erythropoietin) and large proteins (monoclonal antibody, mAb), respectively [17]. These columns can find promising applications in HP-SEC during the manufacturing process to rapidly monitor sialic acid content and glycosylation in glycoproteins, which are considered as an index for product quality of therapeutic glycoproteins, in particular for fc-fusion proteins [52]. In a recent work, Liu et al. proposed a new rapid method in SEC-HPLC/UV to simultaneously monitor protein oligomerization from monomers to large aggregates and to determine their nitration degrees (NDs). Tyrosine (Tyr) nitration is associated with an enhancement of the allergenic potential of proteins and dityrosine formation promotes protein aggregation. This method allowed to use ND data to directly quantify bovine serum albumin (BSA) monomers and oligomers (especially dimers and trimers) [53]. Another interesting new application concerns the separation and quantification of dextran sulphate (average MW 8000 Da), an additive widely used during protein production to enhance cellular productivity. Due to its toxicity, dextran sulphate is necessarily removed during the purification process. Tazi et al. set up a rapid and simple HP-SEC/UV method, which includes protein precipitation and dextran sulphate separation, followed by post-column derivatization. This method allows detecting the residual dextran sulphate in highly concentrated purified protein products, as an alternative to ELISA and fluorometric assays [54]. A common disadvantage of conventional SEC columns concerns the non-specific hydrophobic and electrostatic interactions which can occur between protein samples and stationary phase, causing peak tailing, low recovery and baseline drift. To overcome this limitation, the use of chaotropic salts, such as sodium perchlorate ($NaClO_4$) in the mobile phase, was recently explored. A good peak shape, higher peak resolution and recovery and more stable baseline were observed [55].

New studies explored the potentiality of columns, packed with classical pore size particles (3–5 µm) or sub-2 µm

Table 3 Summary of principal new trends in SEC applications published in the last three years

Techniques	Estimation/objective	Sample/substrate	SEC columns length × i.d.; particle size; porosity	Advantages (A)/disadvantages (D)
Protein A affinity-SEC-HPLC/UV [79]	Aggregation (characterization and quantification of mAb monomers, mAb aggregates and HCP in a complex mixture)	mAb and HCP in growing cell cultures	Tosoh Bioscience TSKGel GW3000SWxl 300 × 7.8 mm; 5 µm; 250 Å	**A**-Simplicity: simultaneous quantification of mAb, HCP and aggregates in a single HPLC run **A**-rapidity: by avoiding the additional capture chromatography steps necessary for SEC alone, that requires purified products **A**-improved accuracy in quantifying relative HCP content in comparison with traditional methods, such as Bradford **A**-reduced cost: by decreasing the quantity of ELISA assays required to quantify HCP **A**-applicability: for different samples, in high-throughput process development
SEC-HPLC/UV-ESI MS [85]	Aggregation (characterization and quantification of monomers, small aggregates and oligomers) Aggregation (interconversion: dissociation/reassociation)	BSA, human transferrin (Tf), human arylsulfatase A (rhASA)	Waters Biosuite high resolution 300 × 4.6 mm; 5 µm; 250 Å	**A-**Novelty: set-up of an on line SEC ESI MS system for real-time aggregate monitoring **A**-simplicity: characterization of complex mixture of proteins without sample purification or desalting before the analysis **A**-differentiation of sample aggregates from ESI process artifacts **A**-versatility: for different protein assemblies **D-**use of MS-compatible denaturants
AIEX-SEC-HPLC/UV [76]	Aggregation (characterization and quantification of aggregates in a protein mixture)	mAbs in presence of a proteinaceous excipient (rHSA)	Tosoh Bioscience TSKGel G3000SWxl 300 × 7.8 mm; 5 µm; 250 Å Supelco Discovery BIO GFC 300 300 × 7.8 mm; 5 µm; 300 Å	**A**-Novelty: proof-of-concept for the serial coupling of AIEX-SEC for aggregation studies in protein mixtures **A**-versatility: potential application for different protein mixtures **D-**the pI of the protein of interest should not be >8.5 **D**-the pI of excipient protein and protein of interest should differ of 0.5 units **D**-the aggregation levels refer to aggregates that can be detected by SEC **D**-optimization required **D**-robustness not evaluated
SEC-UHPLC/UV-MALS-QELS-RI [66]	M, Mw, -D_M, hydrodynamic radius (Rh)	mAbs (rituximab, trastuzumab) fusion protein (etanercept) synthetic copolymer	Waters Acquity Ethylene Bridged Hybrid 200 300 × 4.6 mm; 1.7 µm; 200 Å	**A**-Rapidity **A-**high sensitivity, efficiency and accuracy **A-**versatility: for studying different types of samples **A-**applicability: for comparing products of different manufacturers

Table 3 (continued)

Techniques	Estimation/objective	Sample/substrate	SEC columns length × i.d.; particle size; porosity	Advantages (A)/disadvantages (D)
SEC-2D HPLC/UV [86]	Direct analysis of free drugs and related small molecule impurities size variants analysis separation, identification and quantification of small molecular impurities from the intact ADC	ADCs	Tosoh Bioscience TSKgel SW3000 300 × 4.6 mm; 10 µm; 250 Å	**A-**Simplicity and rapidity: sample preparation not required; simultaneous separation of size variants and small molecular species and identification and quantification of small molecular species **A**-information on ADC aggregation and on ADC small molecular degradation products **A**-validation **A-**applicability: for ADC characterization and stability studies, during conjugation and formulation development
SEC–HPLC/UV [81]	Reaction equilibrium and kinetic aggregation (reversibility monomer–dimer)	IgG1	GE Healthcare Superdex 200 prep grade 600 × 26 mm; 34 µm; 10 µm	**A-**Novelty: prediction of aggregation under different process conditions and study of changes in aggregation during SEC elution **A-**validation
SEC-UHPLC/UV [44, 57]	Size variants analysis [44]	Multiple mAbs (IgG1 and IgG4 types) [44]	Waters Acquity BEH200 150 and 300 × 4.6 mm; 1.7 µm; 200 Å [44]	**A-**Rapidity and high-throughput **A**-high resolution in separating also minor size variants **A-**the generated pressures do not induce aggregation; the samples are analyzed at intact level **A**-versatility: to analyze a wide range of mAbs **A**-robustness **D**-partially denatured monomers may be unresolved from the main peak
	Aggregation (characterization of monomers, low and high MW species) [57]	mAbs (pertuzumab, adalimumab, belimumab, bevacizumab, denosumab, infliximab, ofatumumab, palivizumab, rituximab, trastuzumab) ADCs (~150 kDa) (brentuximab-Vedotin) [57]	Tosoh Bioscience TSKgel SW3000 150 × 4.6 mm; 2 µm; 250 Å Agilent Technologies AdvanceBioSEC 150 × 4.6 mm; 2.7 µm; 300 Å [57]	
SEC–HPLC/FL [52]	Total sialic acid content Glycosylation (amount of glycans)	fc-fusion protein from cell cultures	Tosoh Bioscience TSKGel G3000SWxl 300 × 7.8 mm; 5 µm; 250 Å	**A**-Simplicity: the sialic acid content is inversely proportional to SEC-HPLC retention time **A**-rapidity and high-throughput: the traditional sialic acid content measurements requires purification and titer detection prior to sialic acid content analysis **A**-applicability: to monitor sialic acid content, during the cell culture and biomanufacturing processes **D-**further applications are necessary to demonstrate the universality of the method

Table 3 (continued)

Techniques	Estimation/objective	Sample/substrate	SEC columns length × i.d.; particle size; porosity	Advantages (A)/disadvantages (D)
SEC–HPLC/UV [72, 73]	Size distribution analysis Aggregation (characterization and quantification)	Virus and VLPs vaccines [72] Virus vaccine (influenza A virus strains) [73]	Tosoh Bioscience TSKGel G5000PWxl 300 × 7.8 mm; 17 µm; 1000 Å; TSKGel G4000SWxl 300 × 7.8 mm; 8 µm; 450 Å [72] TSKGel G6000PWxl 300 × 7.8 mm; 13 µm; >1000 Å [73]	**A-**Simplicity **A-**rapidity **A-**versatility and applicability: to analyze final product, during different steps of downstream processing and storage
SEC–HPLC/UV [82]	Aggregation (characterization and quantification of aggregates induced by light)	mAbs (bevacizumab, cetuximab,infliximab, rituximab,trastuzumab) in medicines and in diluted administration forms	Waters Acquity Ethylene Bridged Hybrid 200 300 × 4.6 mm; 1.7 µm; 200 Å	**A-**Applicability: important information on photodegradation and stability of products, during in-use hospital conditions
SEC-HPLC/UV [54]	Residual dextran sulfate (average Mw 8000 Da) content	Highly concentrated purified protein products	Sepax Technologies Zenix, SEC-150, 300 × 7.8 mm; 3 µm; 150 Å	**A-**Novelty: the possibility to quantify dextran sulfate in highly concentrated samples **A-**rapidity: the sample treatment procedure is simple in comparison to ELISA based assays **A-**validation
SEC-UPLC/UV [74]	Aggregation (characterization and quantification)	VLPs vaccines (HPV, EV71, HBcAg)	Tosoh Bioscience TSKGel G5000 PWxl 300 × 7.8 mm; 17 µm; 1000 Å; TSKGel G6000 PWxl 300 × 7.8 mm; 13 µm; >1000 Å Waters Acquity UPLC Protein BEH450 300 × 4.6 mm; 2.5 µm; 450 Å Sepax Technologies SRT SEC-1000 300 × 4.6 mm; 5 µm; 1000 Å	**A-**Rapidity and high-throughput: the application of iSEC mode allows to reduce the lag phase (i.e. the time range from injection to elution of the first species) and consequently analysis time **A-**high precision **A-**applicability: to obtain a high-throughput pharmaceutical process development, as an alternative to other techniques (DLS, AF4 and TEM)
SEC-UHPLC/UV [56]	Separation of a protein mixture	Intact proteins (17–670 kDa) under native conditions	Biosep Biosep-SEC-s2000 and s3000 300 × 7.8 mm; 5 µm; 150 and 290 Å Phenomenex Yarra SEC-3000 300 × 4.6 mm; 3 µm; 290 Å	**A-**Rapidity **A-**high selectivity, efficiency, resolution **D-**the elevated pressures can lead to protein denaturation/aggregation

Table 3 (continued)

Techniques	Estimation/objective	Sample/substrate	SEC columns length × i.d.; particle size; porosity	Advantages (A)/disadvantages (D)
Fast-SEC/UV [71]	Structural dynamics Aggregation (monitoring of oligomeric profile)	sHSP (18 kDa)	GE Healthcare Superdex 200 Increase 100/300 GL 300 × 10 mm; 8.6 µm; 7 µm	**A-**Novelty: the first example of purification and characterization of a phage sHSP **A-**simplicity: addition of non-ionic detergents increase solubility and lead to the formation of small aggregates (dimers and tetramers) **A-**applicability: to prevent protein aggregation during protein production and purification
SEC-UHPLC/UV/ESI MS [49]	Size variants analysis Aggregation (characterization and quantification of monomers-tetramers)	CrossMAb	Waters AcquityBEH200 300 × 4.6 mm; 1.7 µm; 200 Å	**A-**Rapidity: the online system allows a direct characterization without SEC fractionation **A-**differentiation of sample aggregates from ESI–MS artifacts **A-**identification of low abundant and non-covalent variants **A-**versatility and applicability: to study antibody-antigen interactions, fc-fusion proteins and protein scaffolds, during bio-process and formulation developments **D-**use of MS-compatible denaturants
SEC–HPLC/FL/FACS [87]	Aggregation (monitoring of protein crystal suspension aggregates)	IgG1 in PEG crystallization system	Tosoh Bioscience TSKGel G3000SWxl 300 × 7.8 mm; 5 µm; 250 Å	**A-**Sensitivity: detection of labeled protein aggregates down to 500 nm, thanks to FL detector **A-**important information on crystalline suspensions: aggregation is not an intrinsic condition of crystalline state and crystallization condition stabilizes protein samples **A-**applicability: as an alternative approach to standard tools, to analyze protein crystal suspension aggregates
SEC-HPLC/DLS [69]	Aggregation kinetics Aggregation (mechanism of formation of higher-order aggregates)	IgG1	GE Healthcare Superdex 200 100/300 GL 300 × 10 mm; 13 µm; 7 µm	**A-**The evaluation of different external factors (buffer type, salt concentration, temperature and pH) allows to find that pH has the most significant effect on aggregation **A-**applicability: during pre-production to control and prevent aggregation
PF-SEC/UV [62]	Refolding and simultaneous purification	rhPTH 1–34	GE Healthcare Superdex 75 100/300 GL 300 × 10 mm; 13 µm; 7 µm	**A-**Simplicity: a solution additive combination (proline and maltose) increases protein refolding, inhibiting aggregation, with simultaneous purification **A-**high efficiency and rapidity of refolding and purification process **A-**high mass recovery, purity and specific bioactivity of sample product **A-**applicability: during large-scale production

Table 3 (continued)

Techniques	Estimation/objective	Sample/substrate	SEC columns length × i.d.; particle size; porosity	Advantages (A)/disadvantages (D)
SEC-HPLC/MALS-UV-RI [68]	M Root-mean-square (RMS) radius (R) Conjugation ratio (n. of proteins bound per biopolymer molecule)	MVCs in dilute solutions	Shodex Shodex OH pak SB-804 HQ 300 × 8 mm; 10 µm; 200 Å	**A**-Simplicity: complete characterization of MVCs with reduced cost **A**-rapidity: this method, named as branching analysis, requires the total M of MVC and not the M of each component, as traditional conjugate analysis **A-less** accuracy, but good approximation in comparison with traditional conjugate analysis **D**-applicability: not as an a alternative, but as a complementary approach to traditional conjugate analysis
SEC-HPLC/MALS-UV-RI [75]	Size distribution analysis Aggregation (characterization and quantification of monomer, aggregates and fragments)	Clostridium difficile vaccine components and intermediates	Tosoh Bioscience TSKGel Super SW3000 300 × 4.6 mm; 4 µm; 250 Å	**A-**Applicability: during process and development studies to control aggregation
SEC-UHPLC/MALS-UV-RI [67] SEC-UHPLC/UV-ESI MS-ΔHDX [60] SEC-UHPLC/UV-IMS-MS-ΔHDX [61]	Extinction coefficients Conformational changes of protein higher-order structures in a protein mixture	mAbs (rituximab, rastuzumab, infliximab) fusion protein (etanercept) synthetic copolymer protein mixture (insulin, cytochrome C, ubiquitin, myoglobin) [60] (cytochrome C, ubiquitin, myoglobin) [61]	Waters Acquity Ethylene Bridged Hybrid 200 300 × 4.6 mm; 1.7 µm; 200 Å Waters Protein BEH SEC-125 150 × 4.6 mm; 1.7 µm; 125 Å	**A**-Accuracy and precision in the determination of extinction coefficients in comparison with theoretical values **A**-versatility: for studying different types of samples **A-**applicability: during development studies **A**-simplicity: the addition of TFA in the mobile phase overcomes restrictions for HDX in the solvent accessible parts of the protein **A**-rapidity: comparison of global conformational changes in a mixed sample solution, simultaneously **A**-screening for conformational changes vs. solution conditions **A**-applicability: to compare bach-to-batch higher-order structures and optimize reaction conditions, during purification **D**-use of MS-compatible denaturants

Table 3 (continued)

Techniques	Estimation/objective	Sample/substrate	SEC columns length × i.d.; particle size; porosity	Advantages (A)/disadvantages (D)
SEC-HPLC/UV [55]	Mw Molecular size distribution (MSD) Aggregation (characterization and quantification of monomer, dimers, tetramers, aggregates)	IgG	Tosoh Bioscience TSKGel Super SW3000 300 × 4.6 mm; 4 µm; 250 Å	**A-**Complete elution of aggregates and tetramer **A**-simplicity: the addition of $NaClO_4$ in the mobile phase overcomes non-specific interactions protein/stationary phase **A**-high resolution and improved peak shape **A**-rapidity **A**-baseline stabilization **A**-high recovery of aggregates **A**-validation **A**-applicability: as an alternative approach to the method described in EP monographs
SEC-HPLC/UV [53]	NDs (tyr nitration detection and quantification) Aggregation (characterization and quantification)	BSA mono, di-, tri-, and higher oligomers	Agilent Technologies AdvanceBioSEC 300 × 4.6 mm; 2.7 µm; 300 Å	**A-**Novelty and simplicity: proof-of-concept for the direct quantification of oligomers by their determined NDs **A-**rapidity and low sample consumption: the detection of mono and oligomers, as well as their NDs is simultaneous **A**-validation
SEC-HPLC/UV [63, 65]	Aggregation (characterization and quantification of tetramers, octamers and higher-order aggregates made up of three or more ErA tetramers bound together) [65] (Characterization and quantification of mAb fragments) [63]	Basic protein higher-order aggregates from Erwinia chrysanthemi L-asparaginase (ErA) IgG1 (fragments and impurities)	Agilent Technologies ProSEC 300S 300 mm × 7.8 mm, 5 µm; 300 Å [65] Tosoh Bioscience TSKGel G3000SWxl 300 mm × 7.8 mm; 5 µm; 250 Å [63, 65]	**A-**Good efficiency **A**-improved resolution and reproducibility in comparison with exposed-silanol or diol-silica-based SEC columns, in which higher-order aggregate recovery and detection are not reproducible **A**-robust quantification of higher-order aggregates **A**-applicability: to compare bach-to-batch higher-order structures [65]; to study forced degradation samples and clarify the regions and the peptide bond involved in degradation [63]
SEC-UHPLC/FL [16, 45]	Separation of mAb monomer from ADCs Aggregation (characterization of monomers and dimers)	mAbs (ipilimumab, adalimumab) ADCs (~150 kDa) (trastuzumab-emtansime and brentuximab-Vedotin) [45] IgG1 (adalimumab) [16]	Agilent Technologies AdvanceBioSEC 150 × 4.6 mm; 2.7 µm; 300 Å Tosoh Bioscience TSKGel UP-SW3000 150 × 4.6 mm; 2.0 µm; 250 Å Phenomenex Yarra SEC X-150 150 × 4.6 mm; 1.8 µm; 150 Å Waters AcquityBEH200 150 × 4.6 mm; 1.7 µm; 200 Å	**A-**Optimized column particle size and pore size and mobile phase conditions **A-**rapidity and high throughput **A**-high efficiency and resolution **A-**versatility: for a wide range of mAb products **D**-the samples must be analyzed at moderate pressures and ambient temperature to avoid the formation of aggregates or degradation products

Table 3 (continued)

Techniques	Estimation/objective	Sample/substrate	SEC columns length × i.d.; particle size; porosity	Advantages (A)/disadvantages (D)
SEC-HPLC/UV–MS/MS [58]	Separation of a protein mixture purification and characterization of MHC isoform quantification	Large intact proteins (MHC ~ 223 kDa and β-MHC isoform)	PolyLC Inc. PolyHydroxyethyl A 200 × 9.4 mm; 5 µm; 500 Å	**A-**High-efficient purification and characterization, as an alternative to the use of gel electrophoresis and specific antibodies **A-**high purity **A-**high mass accuracy **A-**applicability: to characterize large proteins (200 kDa) **D-**use of MS-compatible denaturants
sSEC-2D UHPLC/UV–MS/MS [59]	Separation of high Mw proteins detection of low-abundance PTMs	Intact proteins (10–223 kDa) in complex mixtures	PolyLC Inc. PolyHydroxyethyl A 200 × 9.4 mm; 3 µm; 300–1000 Å	**A-**High resolution **A-**high reproducibility **A-**high-efficient purification **A-**high coverage (particularly for low abundance and high MW proteoforms) **A-**applicability: to characterize native protein complexes **D-**use of MS-compatible denaturants

particles, in combination with ultra-high-pressure SEC (UHPLC-SEC) and fast protein liquid chromatography (FPLC), also named as Fast-SEC. The only critical aspect of the use of high pressures might be the enhancing effect on protein denaturation, associated to a viscous heating of the mobile phase under such conditions. UHPLC-SEC with UV or fluorescence detector (FL) has been successfully used to obtain rapid (within 10 min) and efficient analysis of different samples, such as mixtures of intact proteins (17 kDa < MW < 670 kDa) or mAbs and antibody–drug conjugates (ADCs), a complex biotherapeutic composed of mAb, a linker and a small molecule drug covalently bound [16, 45, 56, 57]. UHPLC-SEC has proven to be also powerful for the analysis of multiple mAbs types. Conventional SEC separations of mAbs result in low resolution between the different sized species, long analysis time (up to 1 h) and low sample throughput. Yang et al. have recently shown that UHPLC-SEC with MS detection can overcome conventional SEC limitations and allows to resolve mAb size variants (IgG1 and IgG4 types) with a three time higher throughput compared to conventional SEC, and without inducing aggregation under a careful optimization of mobile phase conditions (i.e. ionic strength and presence of salt additives) [44]. Haberger et al. set-up an UHPLC-SEC approach, coupled to native ESI–MS, for fast and robust analysis of bispecific mAb size variants. Bispecific products are more complex biomolecules compared to standard antibodies, as they are developed with complicated production processes and technologies (i.e. CrossMAb design). Given their intrinsic structural complexity, the search for their size variants during bioprocess and formulation development is very important. The developed Fast-SEC approach demonstrated improvements in terms of run time and chromatographic separation resolving power compared to the in-house standard HPLC-SEC approach as can be seen in Fig. 3. The method was suitably robust for the analysis of CrossMAb size variants derived from stress stability and bio-process development studies. Native SEC/UV/MS method also allows to obtain the characterization of low-abundant and non-covalent size variants and represents a powerful platform for studying antibody-antigen interactions and other major classes of biopharmaceuticals, such as fc-fusion proteins and protein scaffolds [49].

Very recent MS-compatible SEC methods have shown their potential for the characterization of large intact proteins and for the fractionation of intact proteins from complex mixtures [58, 59]. Jin et al. set-up a SEC middle-down MS method for purifying myosin heavy chain (MHC) from heart tissue and for characterizing MHC (~223 kDa) and its main isoform (β-MHC) [58]. Cai et al. have applied a 2D UHPLC based on a serial size exclusion chromatography (sSEC) with columns with different pore size connected in series, in combination with a RP-UHPLC/MS system, to analyse cardiac

protein extract over a broad molecular weight range (from 10 up to 223 kDa) and low abundance protein post-translational modifications (PTMs). The 2D strategy, coupled with sSEC and top–down MS, allowed to obtain reproducible and high resolution separations of complex protein mixtures [59]. The combination of differential hydrogen–deuterium exchange (ΔHDX) with ESI–MS or ion mobility spectrometry–mass spectrometry (IMS–MS) in the SEC platform can be successfully applied to study protein conformational changes in protein mixtures, such as cytochrome C, ubiquitin, myoglobin and insulin. This allows to have a rapid comparison of protein batch-to-batch higher-order structures and consequently to optimize conditions for enzymatic reactions or for protein chemical modification reactions eventually required to reach the mature form of a biotechnological product [60, 61].

Recombinant proteins can also be prone to aggregation during the refolding and purification process. Vemula et al. set-up a simple and rapid Protein Folding-SEC (PF-SEC) method, that was able to obtain simultaneously in vitro refolding and purification of recombinant human parathyroid hormone (rhPTH 1–34), preventing sample aggregation. PF-SEC, using solution additive combinations (i.e. proline and maltose), prevents aggregation and increases the efficiency of refolding and purification process in terms of high mass recovery, purity and specific bioactivity [62].

To evaluate the integrity and purity of mAbs, an important issue concerns the monitoring and the quantification of mAb fragments, which also can form aggregates or non-covalent/covalent complexes with intact antibodies. The identification of fragments requires orthogonal analytical methods, such as SEC-HPLC, CE and LC–MS [63].

Applications of SEC to the Characterization of Protein Aggregates

High-molecular weight aggregates, derived from incorrect light or heavy chain association, typically represent critical product-related impurities for mAb production and bispecific mAb formats [49]. An important concern about high MW species is their potential to increase product immunogenicity. SEC/MALS has been the method of choice to characterize and monitor the formation of higher-order structures, complexes and aggregates for many years [64]. However, it has been demonstrated that the determination of higher-order aggregates for basic proteins using exposed-silanol or diol-silica-based SEC conventional columns is not reproducible. For this reason, recently coated-silica SEC columns, which contain hydrophobic silica particles, are proposed as an interesting alternative to improve reproducibility and resolution of basic protein larger aggregates [65]. Figure 4 shows the resolution of biopharmaceutical enzyme *Erwinia chrysanthemi* l-asparaginase (ErA) octamer and monomer, with a high degree of consistency between replicate injections using the ProSEC 300S column. A stationary phase with an appropriate surface modification which is able to remove acidic silanols and reduce interaction effects.

Recently, new interesting studies showed the potentiality of the combination of UHPLC-SEC with multi-detection

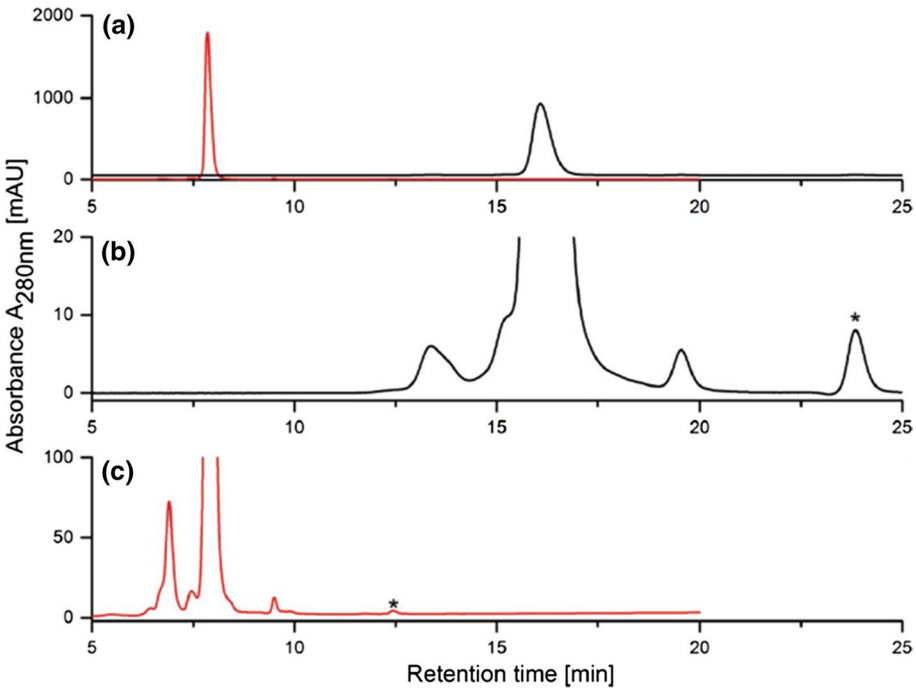


Fig. 3 Comparison of standard HPLC-SEC (*black*) and Fast-SEC (*red*) chromatograms of a CrossMab monomer (**a**) and magnification of standard HPLC-SEC and Fast-SEC chromatograms, respectively (**b**) and (**c**), during stability studies of the same sample (storage for 24 months at 5 °C). Fast-SEC shows rapid analysis times and a major ability to resolve non-covalent low and high MW aggregates formed during stability studies. Peaks with *asterisks* are derived from sample matrix [49]

systems including MALS. UHPLC-SEC/UV-MALS-quasi elastic light scattering (QELS)-RI is a rapid and sensitive method, which provides accurate results of molar mass (M), mass-average molar mass (Mw) and molar mass dispersity ($-D_M$) of different types of biotherapeutic proteins, such as mAbs and fusion proteins [66]. Likewise, a UHPLC-SEC/UV-MALS-RI configuration was developed by Miranda et al., which obtained an accurate estimation of extinction coefficients of some therapeutic mAbs (rituximab, rastuzumab, infliximab, etanercept) with important results for the determination of protein quantity, useful to characterize samples during the manufacturing process [67].

Some interesting applications in the field of large protein aggregates have been reported using columns with high particle size (≥8 µm), showing high resolution. These columns have been applied to study multivalent conjugates (MVCs) which are molecules formed by multiple proteins and high molecular weight polymer chain, useful to improve pharmacokinetics and to decrease degradation of molecules, such as growth factors. The bioactivity of MVCs is dependent on the conjugation ratio (number of proteins bound per biopolymer molecule), and important information can be obtained by a combination of SEC/MALS and UV-RI in-line detectors [68]. In addition, these SEC columns and DLS have been usefully applied in the monitoring of aggregation kinetics from monomer to aggregates of mAbs, under different conditions (pH, temperature, salt additives and buffer type) [69].

It is important also to focus the attention on the role of mobile phase composition (solvents, buffers and additives). In the literature, the potential effect of mobile phase salt additives on sample aggregation was less treated. Singla et al. show that the use of high ionic strength buffers can induce protein aggregation [69]. Only recent publications comment the role of salt additives in SEC, for example on mAb and ADC aggregates. A recent SEC-UHPLC/FL method, in fact, shows that NaCl concentration up to 200 mM can create oligomeric artefacts, measured by mAb dimer/monomer area ratios [45]. Therefore, the addition of stabilizing excipients in downstream processing to inhibit aggregation can represent an important issue. For example, trehalose can reduce aggregation and stabilize monomer content in presence of salt additives [70].

In a recent study, small heat shock proteins (18 kDa) produced from cloned genes were characterized for the first time by using Fast-SEC, SDS-PAGE and DLS detection. These tools allowed to study protein intrinsic ability to form large aggregates, by monitoring polydisperse oligomers also under different conditions (temperature, protein concentration, detergents) [71].

In the last years, the interest of biopharmaceutical industry has also been directed toward the development of virus-like particle-based (VLP) vaccines. These products are large size and complex protein assemblies, deriving from recombinant expressions of viral structural proteins. Also in this case the control of product-related impurities such as aggregates is higly important. HPLC-SEC/UV and UHPLC-SEC/UV methods were developed for characterization of virus and VLPs samples [72–74]. A recent UHPLC-SEC/UV method was able to characterize five recombinant protein-based VLPs including human papillomavirus (HPV) VLPs, human enterovirus 71 (EV71) VLPs, and chimeric hepatitis B core antigen (HBcAg) VLPs pointing out the generic applicability of the assay. The method uses an interlaced injection mode (interlaced SEC, iSEC), that consists of injecting subsequent samples prior to the complete elution of previous sample components. The proposed UHPLC-SEC/UV is a rapid (within 3.1 min per sample), high-throughput and versatile method for quantitatively characterize VLP aggregates during manufacturing process or storage, and it is an alternative to the traditionally used techniques such as DLS, asymmetrical flow field-flow fractionation (AF4), and transmission electron microscopy (TEM) [74].

A recent paper describes the use of SEC-HPLC/MALS-UV-RI approach to characterize aggregation levels for four different protein antigens for a Clostridium difficile vaccine, which is an example of protein-based vaccine [75].

Aggregation of biopharmaceutical products can be scarcely monitored in samples containing other proteinaceous ingredients, such as recombinant human serum albumin (rHSA) used to improve product stability. In this specific case, SEC does not allow to separate mAb monomer from rHSA aggregates due to similar MW and/or hydrodynamic volume of the two proteins or their aggregates. This problem can be overcome with a serial coupling of anion exchange chromatography (AIEX) and SEC. Gargani-Weisbjerg et al. proposed the first in-series combination of AIEX-SEC, to study mAbs mixed with rHSA. AIEX separates proteins and captures rHSA and SEC allows to resolve mAbs aggregates, investigating the role of excipient rHSA on therapeutic protein aggregation level [76].

Recently SEC combined in-line with small-angle X-ray scattering (SAXS) and small-angle neutron scattering instrument (SANS) has been presented as a novel approach to characterize protein samples prone to aggregation and low abundance mixture of macromolecular species, due to important structural information given by protein scattering [77, 78].

To quantify mAb in a cell culture process containing a complex mixture of proteins, the combination of Protein A affinity-HPLC with HP-SEC can be very promising. This approach allowed to simultaneously and rapidly quantify in a single HPLC run mAb monomers, mAb aggregates and host cell proteins (HCPs) in growing cell cultures, representing an accurate, simple and economic method to quantify relative HCP content as an alternative to conventional methods such as ELISA and Bradford [79].

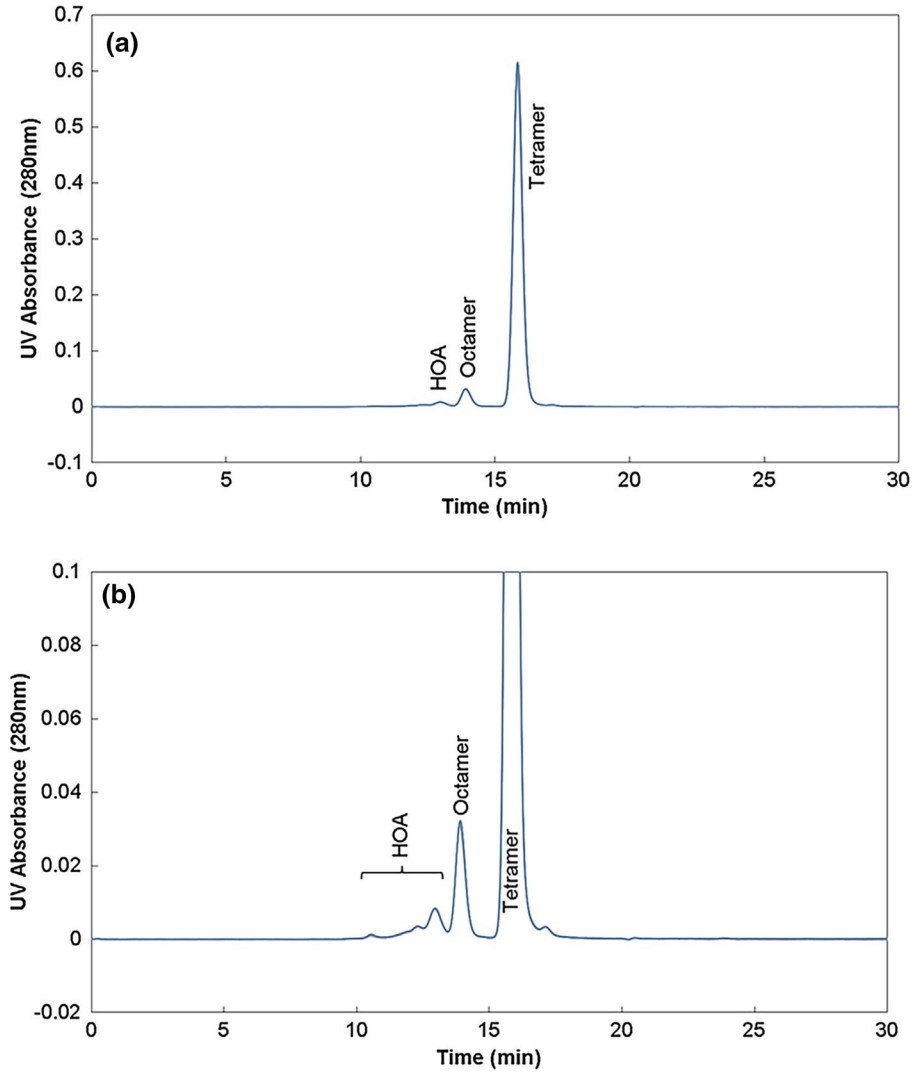


Fig. 4 Example of SEC Chromatograms for ErA Using Coated-Silica Column. Each chromatogram represents an overlay of 10 consecutive injections. The data are presented both for full-scale (**a**) and close-up of the region of interest (**b**) [65]

The interest in formation of small aggregates in biotechnology and, consequently, the search for methods able to analyse and characterize small protein aggregates, is recent. In fact, also the formation of very small protein clusters can affect the viscosity and the stability of protein solutions and represents a problem for production, purification and delivery of biopharmaceutical products, such as mAb solutions [80]. SEC/UV methods are applied to rapidly and simply evaluate the photo-degradation and/or the formation of small aggregates (dimers) in mAbs, both in medicines and diluted administration forms [81, 82]. In addition, Ojala et al. have shown the potentiality of SEC/UV in the estimation of the reversibility of the aggregation, the reaction equilibrium and kinetic, and in the prediction of IgG1 aggregation (mainly dimers) using different processing conditions [81].

For small protein aggregates and oligomers, MS can represent an important tool complementary to HP-SEC. The first attempts of SEC-HPLC combined with native ESI-TOF MS to characterize intact small oligomers (dimers-tetramers) of proteins, such as IgG1 or antithrombin III dates 2010–2012 [83, 84]. In these works, SEC-HPLC was used to analyse and fractionate oligomers, and then native ESI–MS was applied to characterize fractions. With this approach transient small oligomers can be dissociated by SEC.

To avoid oligomer dissociation and to monitor the evolution of small oligomers in real time, online SEC/native ESI MS systems represents a powerful tool. This approach provides important information on protein conformational stability and on transient small aggregates that undergo rapid dissociation/re-association on the chromatographic time scale, depending on different conditions. Online detection with native ESI MS allows also to measure the mass distribution of non-covalent protein assemblies and to distinguish oligomers present in samples from the artefacts produced during the ESI process. Studies at different pH (7.0 and 5.0)

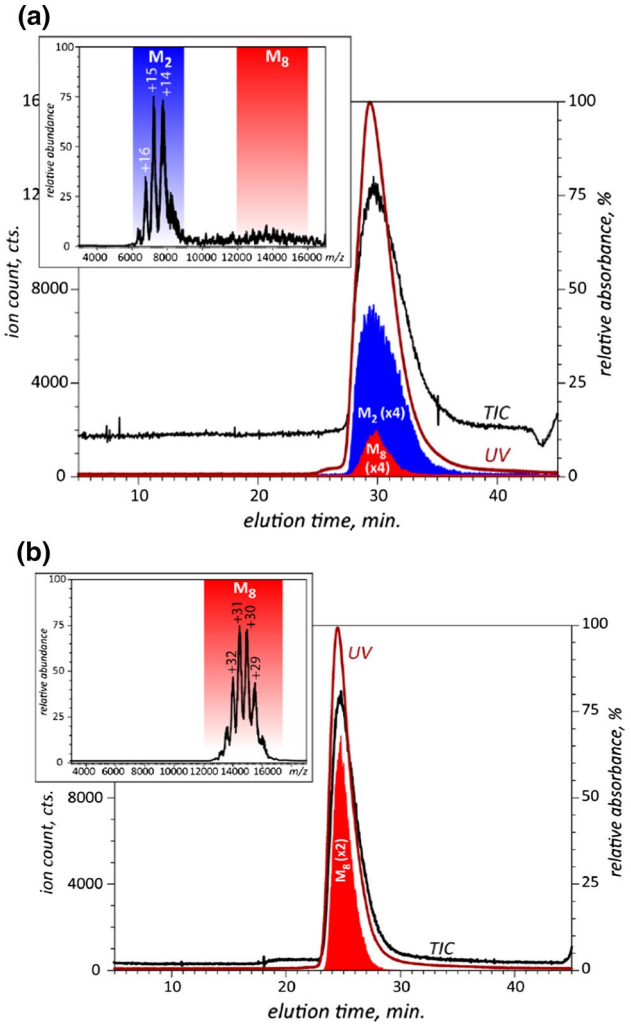


Fig. 5 Online SEC/MS analysis of human arylsulfatase A (rhASA) at pH 7.0 (**a**) and at pH 5.0 (**b**): UV (*brown trace*), total ion (*black trace*) and the cumulative extracted ion (*coloured areas*) chromatograms. *Inset* mass spectrum averaged across the SEC peak. The total ion chromatograms reveal the contributions from both octameric (M8 *red area*) and dimeric (M2 *blue area*) ions at pH 7.0 (**a**) and octameric species alone at pH 5.0 (**b**, M8 *red area*) [85]

on human arylsulfatase A (rhASA) show a different protein oligomerization, revealing the evolution of dimers into octamers under acidic conditions [85] (Fig. 5).

An interesting novel approach in which SEC is coupled with 2D-LC/MS allowed the direct analysis of free drugs and related small molecule impurities (free drug, linker drug and potential adducts) in ADCs. SEC 1D-LC separates ADC from small molecular species, giving information about ADC monomer ad aggregates. RP-HPLC in the 2nd dimension separates the different small molecule impurities present in the ADC sample, as illustrated in Fig. 6. The proposed SEC-RP 2D-LC can be used to study ADC stability, also providing the identification and quantification of small molecule degradation products [86].

New approaches concerning the use of HP-SEC equipped with a fluorescence (FL) detector and flow cytometry (Fluorescence-Activated Cell Sorting, FACS) allow to study crystalline state of biopharmaceuticals. Crystallization represents an important formulation platform for biopharmaceuticals, but the addition of polyethylene glycol (PEG) can cause an increase in aggregation. FACS represents a new tool to study protein crystal suspensions and allows also to detect protein crystal aggregates down to 500 nm. This study shows that IgG1 aggregate formation, which occurs in PEG crystallization system, is not an intrinsic problem of protein crystalline state, but it takes place in the supernatant. This demonstrates that it is wrong to consider that a sample protein can be less stable in crystal than in solution [87].

Conclusions

Size exclusion chromatography is an essential method for determining molecular weights of proteins and for investigating the extent of aggregation in protein therapeutics.

Over the last few years we witnessed an improvement in stationary phases devoted to SEC, characterized by innovative hybrid silica and organosiloxane particle materials with advanced chemical stability, reduced silanol activity, and larger pore sizes compared to classical SEC. In particular, superficially porous (core shell or porous shell) particles with very large (1000 Å) pores are useful for separating very large biomolecules such as monoclonal antibodies and large proteins.

Progress in SEC columns, shorter and narrower columns packed with smaller-sized particle used in low dispersion UHPLC equipment, have led to fast and efficient separations with advantages in several fields of application, including high-resolution analysis of oligomers and comprehensive two-dimensional (2D-LC) of high-molecular-weight macromolecules separations. In addition, comprehensive 2D-LC can represent a good alternative to make SEC separations compatible with MS detection.

Multi-detector SEC platform has come of age and many applications have been reported in the last years. The latest SEC applications include the growing role of MS detection. In particular, the use of native ESI–MS enables both the identification of abundant size variants (like dimers) by accurate mass determination and the characterization of various low-abundant and non-covalent aggregates. Few examples of small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) instrument combined with SEC have been reported to obtain information on the presence of aggregates or oligomers in protein

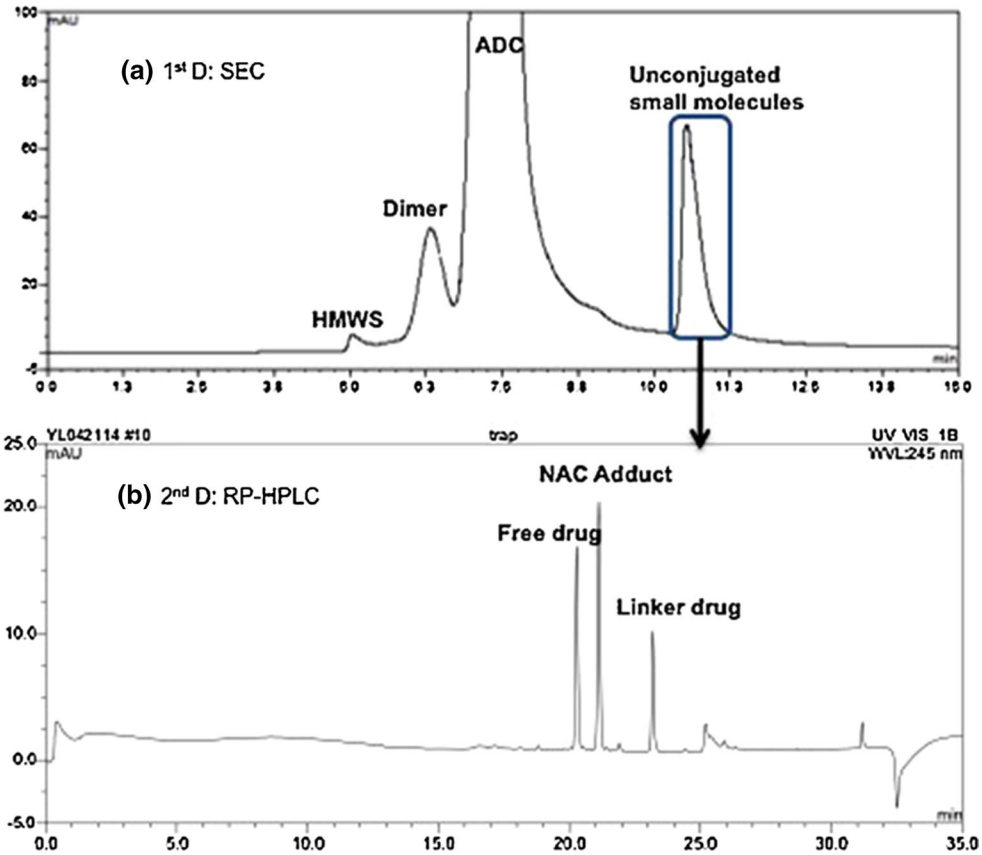


Fig. 6 Analysis of unconjugated small molecules in ADC sample by SEC-RP 2D-LC. SEC 1D separates the unconjugated small molecules from the ADC and aggregates (dimer and high MW species) (**a**), and RP 2D-LC separates the three unconjugated small molecule impurities (free drug, NAC adduct and linker drug) (**b**) [86]

samples. The proper combination of detectors provides implemented information on intact proteins, their variants or aggregates and protein complexes.

The application of SEC methods such as those described in this review will have an important role in advancing the characterization of biopharmaceuticals and reducing some of the risks associated with biopharmaceuticals.

References

1. Dimitrov DS (2012) In: Voynov V, Caravella JA (eds) Methods mol biol. Springer, New York
2. Strohl WR, Knight MD (2009) Curr Opin Biotechnol 20:668–672
3. Mahler HC, Friess W, Grauschopf U, Kiese S (2009) J Pharm Sci 98:2909–2934
4. Rathore N, Rajan RS (2008) Biotechnol Prog 24:504–514
5. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS (2010) Pharm Res 27:544–575
6. Patel J, Kothari R, Tunga R, Ritter NM, Tunga BS (2011) Bio-Process Int 1:20–31
7. Staub A, Guillarme D, Schappler J, Veuthey JL, Rudaz S (2011) J Pharm Biomed Anal 55:810–822
8. Berkowitz SA, Engen JR, Mazzeo JR, Jones GB (2012) Nat Rev Drug Discov 11:527–540
9. Kaltashov IA, Bobst CE, Abzalimod RR, Wang G, Baykal B (2012) Wang S Biotechnol Adv 30:210–222
10. Oliva A, Fariña JB, Llabrés M (2007) Curr Pharm Anal 3:230–248
11. Parr MK, Montacir O, Montacir H (2016) J Pharm Biomed Anal 130:366–389
12. Zhang H, Cui W, Gross ML (2014) FEBS Lett 588:308–317
13. Zhang Z, Pan H, Chen X (2009) Mass Spectrom Rev 28:147–176
14. Fekete S, Veuthey JL, Guillarme D (2015) LCGC Special issues Oct 02: 8–15
15. Bobály B, Sipkó E, Fekete J (2016) J Chromatogr B 1032:3-2
16. Fekete S, Veuthey JL, Guillarme D (2017) J Pharm Biomed Anal 141:59–69
17. Sandra K, Vandenheede I, Sandra P (2014) J Chromatogr A 1335:81–103
18. Oliva A, Llabres M, Farina JB (2001) J Pharm Biomed Anal 25:833–841
19. Hong P, Koza S, Bouvier ESP (2012) J Liq Chromatogr RT 35:2923–2950
20. Patten PA, Schellekens H (2003) In: Brown F, Mire-Sluis AR (eds) The immunogenicity of biopharmaceuticals: lessons learned and consequences for protein drug development. Karger, Basel
21. Rosenberg AS (2006) AAPS J 8:E501–E507
22. Philo JS (2009) Curr Pharm Biotechnol 10:359–372
23. Narhi LO, Schmit J, Bechtold-Peters K, Sharma D (2012) J Pharm Sci 101:493–498
24. Hamrang Z, Rattray NJW, Pluen A (2013) Trends Biotechnol 31:448–458
25. Smulders R, Koll H, Smith B, Bassarab S, Seidl A, Hainzl O, Jiskoot W (2011) Pharm Res 28:920–933
26. Khodabandehloo A, Chen DD (2017) Bioanalysis 9:313–326
27. Gandhi AV, Pothecary MR, Bain DL, Carpenter JF (2017) J Pharm Sci 106:2178–2186
28. Fekete S, Beck A, Veuthey JL, Guillarme D (2014) J Pharm Biomed Anal 101:161–173
29. Uliyanchenko E (2014) Anal Bioanal Chem 406:6087–6094

30. Striegel A, Yau WW, Kirkland JJ, Bly DD (2009) Modern size-exclusion liquid chromatography: practice of gel permeation and gel filtration chromatography, 2nd edn. Wiley, New York
31. Wu CS (2003) Handbook of size-exclusion chromatography and related techniques. Marcel Dekker, New York
32. Saunders GD, Barth HG (2012) LCGC North Am 30:544–563
33. Hudalla CJ, Alden B, Walter T, Walsh D, Bouvier E, Iraneta PC, Lawrence N, Wyndham K (2012) LCGC North Am 30:20–29
34. Wyndham KD, Walter TH, Iraneta PC, Alden BA, Bouvier ESP, Hudalla CJ, Lawrence NL, Walsh D (2012) LCGC North Am 30(S4):20–29
35. Unger KK, Liapis AI (2012) J Sep Sci 35:1201–1212
36. Wagner BM, Schuster SA, Boyes BE, Kirkland JJ (2012) J Chromatogr A 1264:22–30
37. Fekete S, Ganzler K, Guillarme D (2013) J Pharm Biomed Anal 78–79:141–149
38. Schure MR, Moran RE (2017) J Chromatogr A 1480:11–19
39. Pirok BWJ, Breuer P, Serafine JM, Hoppe SJM, Chitty M, Welch E, Farkas T, van der Wal S, Peters R, Schoenmakers PJ (2017) J Chromatogr A 1486:96–102
40. Wagner BM, Schuster SA, Boyes BE, Shieldsa TJ, Miles WL, Haynes MJ, Moran RE, Kirkland JJ, Schure MR (2017) J Chromatogr A 1489:75–85
41. Fekete S, Oláh E, Fekete J (2012) J Chromatogr A 1228:57–71
42. Bouvier ESP, Koza SM (2014) Trends Anal Chem 63:85–94
43. Fekete S, Guillarme D (2014) Trends Anal Chem 63:76–84
44. Yang R, Tang Y, Zhang B, Lu X, Liu A, Zhang YZ (2015) J Pharm Biomed Anal 109:52–61
45. Goyon A, Beck A, Colas O, Sandra K, Guillarme D, Fekete S (2017) J Chromatogr A 1498:80–89
46. Goyon A, Guillarme D, Fekete S (2017) J Pharm Biomed Anal 135:50–60
47. Krull IS, Rathore AS (2015) LCGC Europe 28:54–58
48. Haberger Taraban MB, DePaz RA, Lobo B, Yu YB (2017) Anal Chem 89:5494–5502
49. Leiss M, Heidenreich AK, Pester O, Hafenmair G, Hook M, Bonnington L, Wegele H, Haindl M, Reusch R, Bulau P (2016) Mabs 8:331–339
50. Striegel AM (2005) Anal Chem 77:104A–113A
51. Fekete S, Guillarme D, Sandra S, Sandra K (2016) Anal Chem 88:480–507
52. Liu J, Chen X, Fan L, Deng X, Fai Poon H, Tan WS, Liu X (2015) Biotechnol Lett 37:1371–1377
53. Liu F, Reinmuth-Selzle K, Lai S, Weller MG, Pöschl U, Kampf CJ (2017) J Chromatogr A 1495:76–82
54. Tazi LM, Shiranthi J (2016) J Chromatogr B 1011:89–93
55. Wang H, Levi MS, Del Grosso AV, McCormick WM, Bhattacharyya LJPBA (2017) J Pharm Biomed Anal 138:330–343
56. De Vos J, Kaal ER, Swart R, Baca M, Vander Heyden Y, Eeltink S (2016) J Sep Sci 39:689–695
57. Goyon A, D'Atria V, Bobaly B, Wagner-Rousset E, Beck A, Fekete S, Guillarme D (2017) J Chromatogr B 1058:73–84
58. Jin Y, Wei L, Cai W, Lin Z, Wu Z, Peng Y, Kohmoto T, Moss RL, Ge Y (2017) Anal Chem 89:4922–4930
59. Cai W, Tucholski T, Chen B, Alpert AJ, McIlwain S, Kohmoto T, Jin S, Ge Y (2017) Anal Chem 89:5467–5475
60. Makarov AA, Helmyet R (2016) J Chromatogr A 1431:224–230
61. Pierson NA, Makarov AA, Strulson CA, Mao Y, Mao B (2017) J Chromatogr A 1496:51–57
62. Vemula S, Vemula S, Dedaniya A, Ronda SR (2016) Anal Bioanal Chem 408:217–229
63. Li W, Yang B, Zhou D, Xu J, Li W, Suen W-C (2017) J Chromatogr B 1048:121–129
64. Beck A, Wagner-Rousset E, Ayoub D, Van Dorsselaer A, Sanglier-Cianferani S (2013) Anal Chem 85:715–736
65. Gervais D, Downer A, King D, Kanda P, Foote N, Smith S (2017) J Pharm Biomed Anal 139:215–220
66. Espinosa-de la Garza CE, Miranda-Hernandez MP, Acosta-Flores L, Perez NO, Flores-Ortiz LF, Medina-Rivero E (2015) J Sep Sci 38:1537–1543
67. Miranda-Hernandez MP, Elba R, Valle-González ER, Ferreira-Gómez D, Pérez NO, Flores-Ortiz LF, Medina-Rivero E (2016) Anal Bioanal Chem 408:1523–1530
68. Svedlund FL, Eda I, Altiok EI, Healy KE (2016) Biomacromolecules 17:3162–3171
69. Singla A, Bansal R, Joshi V, Rathore AS (2016) AAPS J 18:689–702
70. Bickel F, Herold EM, Signes A, Romeijn S, Jiskoot W, Kiefer H (2016) Eur J Pharm Biopharm 107:310–320
71. Bourrelle-Langlois M, Morrow G, Finet S, Tanguay RM (2016) PLoS One 11:1–21
72. Yang Y, Li H, Zhengjun LZ, Zhanga Y, Zhanga S, Chena Y, Yua M, Ma G, Su Z (2015) Vaccine 33:1143–1150
73. Vajda J, Weber D, Brekel D, Hundt B, Müller E (2016) J Chromatogr A 1465:117–125
74. Ladd-Effio C, Stefan A, Oelmeier SA, Hubbuch J (2016) Vaccine 34:1259–1267
75. Lancaster C, Rustandi RR, Pannizzo P, Ha S (2016) Methods Mol Biol 1476:279–287
76. Gargani-Weisbjerg PL, Bjerg Caspersen M, Cook K, Van De Weert M (2015) J Pharm Sci 104:548–556
77. Malaby AW, Srinivas Chakravarthy S, Thomas C, Irving TC, Sagar V, Kathuria SV, Bilsel O, Lambright DG (2015) J Appl Crystallogr 48:1102–1113
78. Jordan A, Jacques M, Merrick C, Devos J, Forsyth VT, Porcar L, Martel A (2016) J Appl Crystallogr 49:2015–2020
79. Gjoka X, Schofield M, Cvetkovic A, Gantier R (2014) J Chromatogr B 972:48–52
80. Yearley EJ, Godfrin PD, Perevozchikova T, Zhang H, Falus P, Porcar L, Nagao M, Curtis JE, Gawande P, Taing R, Isidro E, Zarraga IE, Wagner NJ, Yun Liu Y (2014) Biophys J 106:1763–1770
81. Ojala F, Sellberg A, Hansen TH, Broberg Hansen E, Nilsson B (2015) Biotechnol J 10:1814–1821
82. Hernandez-Jimenez J, Salmeron-García A, Cabeza J, Velez C, Capitan-Vallvey LF, Navas N (2016) J Pharm Sci 105:1405–1418
83. Kukrer B, Vasco Filipe V, van Duijn E, Kasper PT, Vreeken RJ, Heck AJR, Jiskoot W (2010) Pharm Res 27:2197–2204
84. Wang G, Johnson AJ, Kaltashov IA (2012) Anal Chem 84:1718–1724
85. Muneeruddin K, Thomas JJ, Salinas PA, Kaltashov IA (2014) Anal Chem 86:10692–10699
86. Li Y, Gu C, Gruenhagen J, Zhang K, Yehl P, Chetwyn NP, Medley CD (2015) J Chromatogr A 1393:81–88
87. Hildebrandt C, Mathaes R, Saedler R, Winter G (2016) J Pharm Sci 105:1059–1065

CONCLUSIONS

The present PhD project originates from the growing global need of new effective treatments against TB, classified as the deadliest infectious disease in 2015.

To this aim, two strategies were followed: the rational design of a *neo*-glycoconjugate vaccine and the investigation of a new therapeutic target for the development of innovative drugs.

In the context of this PhD work, different and complementary analytical techniques were used and innovative analytical methods were developed for the identification, characterization and design of new candidates for innovative anti-TB therapies.

The glycoprotein vaccine consists of an immunogenic protein from *Mtb* conjugated with carbohydrate moieties mimicking the saccharidic surface antigen LAM. TB10.4 and Ag85B are two mycobacterial proteins known for their strong immunoreactivity, inducing both T cell mediated and humoral responses. For this reason, they were examined as carrier proteins and conjugated with mono-, di- and tri-saccharides of mannose. More complex saccharides will be considered to better simulate LAM structure.

The glycoconjugate vaccine was examined in several respects. A systematic study was performed by HILIC-UV to ensure the maintenance of stability during the production process. HILIC technique was also coupled to high resolution MS and the developed method allowed to identify several product impurities. The degradation phenomena observed by the two HILIC methods were attributed to the residual presence of rEK in protein solutions and prompted us to synthesize a rEK-IMER in order to prevent product contamination. Then, the immunogenicity of the conjugation compounds was investigated considering both protein and glycan moieties. Ag85B mutants were designed to avoid the masking of putative antigenic amino acids by glycosylation and different carbohydrates were evaluated for the conjugation. Proteins, sugars and the resulting glycoconjugates were structurally characterized by intact MS measurements of the compounds and LC-MS analyses of the deriving peptides and glycopeptides. Furthermore, a correlation between structure and function was derived by antibody epitope identification and SPR determination of affinity binding constants.

*Mtb*PNP represents an interesting target for the development of selective inhibitors as new drugs potentially active against resistant and latent forms of TB. Several purine analogues were tested for the identification of agents inhibiting the mycobacterial enzyme and not the human counterpart.

In order to find new antimycobacterial drugs, a small library of 6- and 8-substituted nucleosides was screened towards PNPs from *Mtb* and human. The inhibition induced by the compounds was

evaluated through an enzymatic activity assay followed by HILIC-UV analyses. The screening procedure allowed to select the most interesting compounds, for which K_i and IC_{50} parameters (indicative for the inhibitor potency) were defined. The selected nucleosides were also submitted to *in vitro* tests on *Mtb* cultures. Molecular modelling studies will be performed to rationalize the experimental data and assist the design of more active inhibitors.

This PhD project showed that pharmaceutical analysis can play a fundamental role in the drug discovery phase, allowing to address various aspects and answering questions about the possible biological activity of new potential drug candidates.

Since a unique, ideal analytical platform is not available, the development of different methodologies able to give complementary information is undoubtedly useful, as demonstrated in this work.

Overall, pharmaceutical analysis proved to be a powerful tool to support the various phases of the glycoconjugate vaccine preparation and to lay the basis for the rational design of antitubercular lead-candidates.