



CrossMark  
click for updates

Cite this: *RSC Adv.*, 2014, 4, 56455

# Chemoenzymatic synthesis of neoglycoproteins driven by the assessment of protein surface reactivity†

T. Bavaro,<sup>\*a</sup> M. Filice,<sup>\*b</sup> C. Temporini,<sup>a</sup> S. Tengattini,<sup>a</sup> I. Serra,<sup>a</sup> C. F. Morelli,<sup>c</sup> G. Massolini<sup>a</sup> and M. Terreni<sup>a</sup>

In this paper a series of 2-iminomethoxyethyl mannose-based mono- and disaccharides have been synthesized by a chemoenzymatic approach and used in coupling reactions with  $\epsilon$ -amino groups of lysine residues in a model protein (ribonuclease A, RNase A) to give semisynthetic neoglycoconjugates. In order to study the influence of structure of the glycans on the conjugation outcomes, an accurate characterization of the prepared neoglycoproteins was performed by a combination of ESI-MS and LC-MS analytical methods. The analyses of the chymotryptic digests of the all neoglycoconjugates revealed six Lys-glycosylation sites with a the following order of lysine reactivity: Lys 1  $\gg$  Lys 91  $\equiv$  Lys 31 > Lys 61  $\equiv$  Lys 66. A computational analysis of the reactivity of each lysine residue has been also carried out considering several parameters (amino acids surface exposure and  $pK_a$ , protein flexibility). The *in silico* evaluation seems to confirm the order in lysine reactivity resulting from proteomic analysis.

Received 24th September 2014

Accepted 23rd October 2014

DOI: 10.1039/c4ra11131a

[www.rsc.org/advances](http://www.rsc.org/advances)

## Introduction

Glycoproteins play a pivotal role in many physiological and pathological processes.<sup>1</sup> Their function and versatility have long suggested the power of these glycosylated biomolecules in therapeutic strategies, with particular relevance in the development of carbohydrate-based vaccines.<sup>2,3</sup> The glycosylation or other controlled chemical modifications of proteins, such as PEGylation and acylation, can dramatically improve their physical and biological properties.<sup>4</sup> For instance, chemical conjugation with suitable glycans has been exploited for the production of therapeutic enzymes with improved activity and reduced *in vivo* metabolism.<sup>5</sup> Neoglycoproteins, which contain unnatural linkages between protein and saccharides, can also provide carbohydrate antigens and immunogens from which

immunodiagnostic and therapeutic reagents can be derived.<sup>4</sup> For example, conjugation of antigenic proteins from *Mycobacterium tuberculosis* with arabino-mannane polysaccharides has been proposed for developing highly immunogenic glycoconjugate vaccines against tuberculosis.<sup>3</sup>

Chemical routes to the synthesis of neoglycoproteins can involve random or site-selective modifications of protein surface expecting the final covalent linkage of the glycans *via* their reducing end interposed by a reactive spacer.<sup>1,6</sup> Usually, methods used for the synthesis of neoglycoproteins permit to conjugate the carbohydrate to a protein by the alkylation of nucleophilic side chains of cysteine or lysine residues.<sup>2,7,8</sup> Cysteine-based chemical ligation has been significantly used to fully synthesize several neoglycoproteins<sup>8</sup> and in some cases the incorporation, by site-directed mutagenesis, of one reactive cysteine residue as a chemoselective tag at a preselected position within the given protein was requested.<sup>9</sup> A different strategy, known as indiscriminate glycosylation, relies on the use of 2-iminomethoxyethyl thioglycosides (IME) and takes advantage from the high abundance of lysine residues on the protein surface, allowing the introduction of various saccharide units for each molecule of protein.<sup>8</sup> For example, Pearce and co-workers employed this methodology to incorporate up to 20 000 glycan moieties onto the surface of adenovirus.<sup>10</sup> Although firstly introduced about 30 years ago, this simple and well-established method can now be revalued in vaccine development in order to strengthen immune response by the use of polyantigenic glycoconjugates prepared with a carrier protein bearing a high number of lysines.<sup>11</sup> A high glycosylation degree can also be useful in order to improve the therapeutic efficiency

<sup>a</sup>Department of Drug Sciences and Italian Biocatalysis Center, University of Pavia, via Taramelli 12, I-27100 Pavia, Italy. E-mail: teodora.bavaro@unipv.it

<sup>b</sup>Departamento de Biocatalisis, Instituto de Catalisis (ICP-CSIC), Marie Curie 2 Cantoblanco Campus UAM, 28049 Madrid, Spain. E-mail: marco.filice1@gmail.com

<sup>c</sup>Department of Chemistry and Italian Biocatalysis Center, University of Milano, via Golgi 19, I-20133 Milano, Italy

† Electronic supplementary information (ESI) available: Regioselective hydrolysis of acetylated mannopyranosides 1–3 catalyzed by immobilized *Candida rugosa* lipase. Glycosylation study for the synthesis of disaccharides 5, 6 and 8 at different reaction conditions. Different cycles of hydrolysis reaction of compound 3 catalyzed by CRL-ECOD. Analysis of the solvent accessibility (ASA) of the amino acids of RNase A (PDB code 1FS3). Surface analysis and localization of Lys residues of RNase A in monomeric form. <sup>1</sup>H and <sup>13</sup>C NMR spectra of disaccharides 5, 6 and 8. ESI-MS deconvoluted spectra of neoglycoproteins 13–16. List of glycopeptides identified by LC-ESI-MS<sup>n</sup> from chymotryptic digestion of 13–16. See DOI: 10.1039/c4ra11131a

of bioactive proteins, but, in this case also the area of the protein responsible of the biological activity could be involved in the chemical modification.

The coupling reaction between oligosaccharides and a protein by a non-selective glycosylation approach, allows the formation of different randomly modified glycoforms, including the possible glycosylation of areas of the protein surface associated with its biological activity. This implies the need of a detailed characterization of the conjugation products.

In addition to the complex analysis required for the characterization of glycoconjugates, it should be also considered that the development of new drugs based on complex structures such as neoglycoproteins is still hampered by the complex multi-step strategies required for sugar preparation. Nowadays, different approaches are available to facilitate the access to pure oligosaccharides, such as the automated solid-phase synthesis.<sup>2,12</sup> Alternative enzymatic or chemoenzymatic approaches has also gained popularity because of intrinsic properties of the biocatalysts. In this vein, the lipase-assisted chemoenzymatic synthesis of biologically relevant di- and trisaccharides has been recently reported.<sup>13–15</sup>

In this work, we propose a useful multidisciplinary approach for a rapid preparation and analytical characterization of complex neoglycoproteins obtained by indiscriminate glycosylation of lysine  $\epsilon$ -amino groups of ribonuclease A (RNase A), selected as model protein. For the synthesis of the reactive glycans used for conjugation with the protein, the enzymatic regioselective deprotection of peracetylated mannopyranose derivatives 1–3 has been studied. The compound 3 carries a thiocyanomethyl group at the C-1 position as precursor of the reactive IME. The products 3a,b obtained by hydrolysis of the thiocyanomethyl mannopyranoside 3 have been then considered as intermediates for the synthesis of peracetylated dimannopyranoside and arabinomannopyranosides 5, 6 and 8, according to the general chemoenzymatic strategy previously reported.<sup>13–15</sup> The final IME-saccharides (9–12) have been used for the glycosylation of RNase A. Flow Injection Analysis (FIA) of intact glycoproteins by MS is generally carried out to establish product identity and integrity of natural glycoproteins, as well as to assess the relative abundance of individual glycoforms.<sup>16–18</sup> In this work, this MS analytical approach has been used also for the characterization of the neoglycoproteins in order to determine the number and position of glycans introduced. Accordingly, the glycosylation degree was monitored by direct infusion of intact proteins in a linear ion trap mass spectrometer (ESI-LIT-MS), while a detailed characterization of glycosylation sites and their relative abundance was performed by liquid chromatography-mass spectrometry peptide mapping after an appropriate proteolytic cleavage of the synthesised neoglycoproteins. Moreover, a computational study of the reactivity of the different lysines on the RNase A surface has been performed. The *in silico* evaluation of selected physicochemical parameters that could correlate with the reactivity of the different glycosylation sites (lysines) on the protein surface, could be useful tools for the design, preparation and characterization of neoglycoconjugates.

## Results and discussion

### Synthesis of the target IME-saccharides for neoglycoproteins preparation

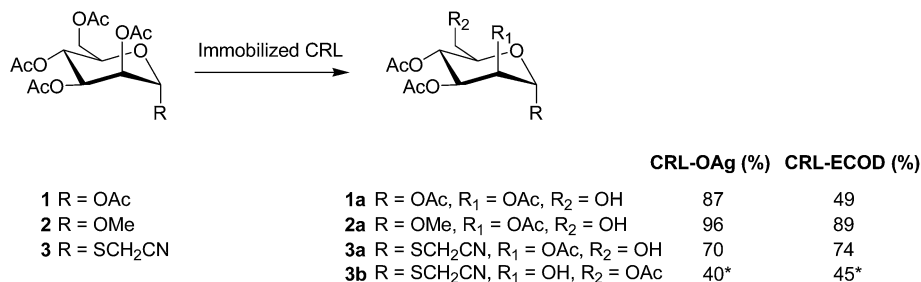
Mannopyranose acts as a pattern-recognition molecule in several human and bacterial glycoproteins.<sup>19</sup> For example, human immunoglobulins contain high mannose- and/or hybrid-type oligosaccharides linked to asparagine or serine/threonine residues on the heavy chain of the protein,<sup>20</sup> while, lipoarabinomannan (LAM), the major surface antigen of *Mycobacterium tuberculosis*, presents an arabinomannan core capped with mono-, di- and trimannosides.<sup>21</sup>

In this work, peracetylated mannose and derivatives thereof have been thus subjected to enzymatic hydrolysis, in order to obtain carbohydrate moieties bearing only one free hydroxyl group in specific positions. The obtained compounds have been then used as key building blocks for the synthesis of mannose-based disaccharides. First, the hydrolyses of the peracetylated  $\alpha$ -mannopyranoside (1) (Scheme 1), catalyzed by different lipases adsorbed on hydrophobic support, were studied (Table S1, see ESI†).

Among all the enzymes tested (lipases from *Candida rugosa*, *Pseudomonas fluorescens*, *Aspergillus niger*, *Pseudomonas cepacia*, *Pancreas porcine* and *Rhizomucor miehei*), only the lipase from *Candida rugosa* (CRL) showed both a good activity and a useful selectivity, affording 1,2,3,4-tetra-*O*-acetyl- $\alpha$ -D-mannopyranoside (1a), deriving from the hydrolysis of the acetyl group at the 6 position, as the only product, according to the results previously obtained with other pyranoses.<sup>13–15</sup> The catalytic performances of the other enzymes tested were negligible (data not shown) and, consequently, CRL was selected as the biocatalyst for the hydrolysis of acetylated  $\alpha$ -*O*-methyl mannopyranoside (2) and 1-thio-(*S*-cyanomethyl)- $\alpha$ -mannopyranoside (3).

CRL, adsorbed onto two supports with different grade of hydrophobicity such as octyl-sepharose (OAg) or decaoctyl-sepharose (ECOD),<sup>22</sup> afforded different results depending on the substrate (Table S1, see ESI†). Great differences can be observed in term of selectivity, depending on the functional group linked at the anomeric position of the sugar (acetoxy, methoxy or thiocyanomethyl group). In particular, a complete selectivity was achieved toward the primary position when 1 and 2 were used as substrate, while the presence of thiocyanomethyl group at the anomeric position strongly influenced the regioselectivity. In fact, when compound 3 was subjected to enzymatic hydrolysis, 2,3,4-tri-*O*-acetyl-1-thio-(*S*-cyanomethyl)- $\alpha$ -D-mannopyranoside 3a (deprotected at the C-6 position) together with 3,4,6-tri-*O*-acetyl-1-thio-(*S*-cyanomethyl)- $\alpha$ -D-mannopyranoside 3b (deprotected at the C-2 position) were obtained in almost equal amounts (Scheme 1 and Table S1†) In addition, the selected biocatalyst (CRL-ECOD) resulted completely stable after five cycles (Fig. S1, see ESI†).

The hydrolysis of substrate 3 was scaled up to 1.65 L affording 2.2 g and 1.8 g of products 3a and 3b, respectively. The C-6 and C-2 monodeprotected products 3a and 3b were used as building blocks to prepare the saccharides 5, 6 and 8 (Scheme 2)

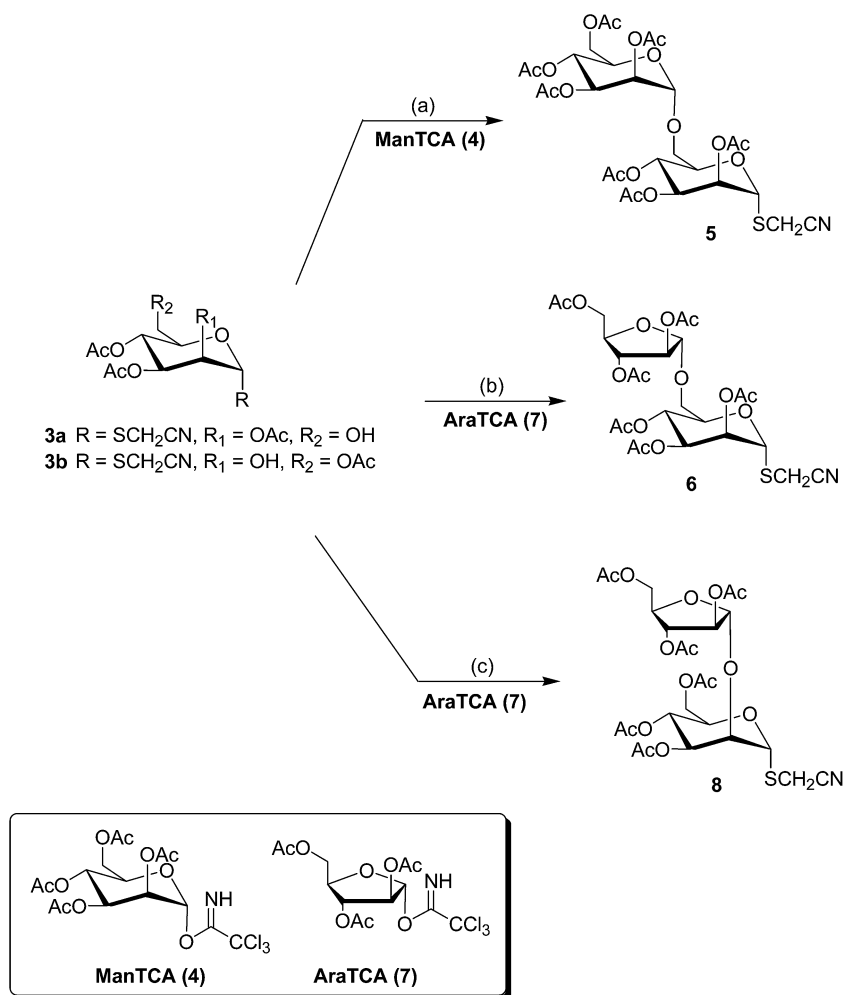


**Scheme 1** Regioselective hydrolysis of acetylated mannopyranosides **1–3** catalyzed by immobilized *Candida rugosa* lipase. The best percentage conversions of monodeprotected compounds **1a–3a** and **3b** are reported. Experimental conditions: 50 mM phosphate buffer with 20% acetonitrile; pH 4; room temperature; [substrate] = 5 mM, \*10 mM; volume = 25 mL; CRL-OAg = CRL-ECOD = 100 IU, IU =  $\mu\text{mol min}^{-1}$ .

by Schmidt glycosylation reaction (Lewis acid-catalyzed glycosylation using glycosyl trichloroacetimidate as the donor).<sup>13</sup>

The reaction was optimized with respect to nature and concentration of glycosylation promoter, time and temperature (Table S2, see ESI†) and the acetylated thiocyanomethyl disaccharides Man(1-6)Man (**5**) and Ara(1-6)Man (**6**) were obtained in

70% and 62% yields, respectively using the mild Lewis acid boron trifluoride etherate as the promoter (Scheme 2). The synthesis of the peracetylated disaccharide Ara(1-2)Man (**8**) required instead the use of the stronger promoter *tert*-butyldimethylsilyl trifluoromethanesulfonate, to afford product **8** in 50% yield (Table S2†).



**Scheme 2** Synthesis of disaccharides **5**, **6** and **8** via glycosylation reaction. (a) **3a**, BF<sub>3</sub>·OEt<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C, MS 4 Å, 5 h, 70% yield; (c) **3b**, TBSOTf, 0 °C, Et<sub>2</sub>O, MS 4 Å, 1 h; 50% yield; (b) **3a**, BF<sub>3</sub>·OEt<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 25 °C, MS 4 Å, 1 h, 62% yield. MS = molecular sieves, DMF = *N,N*-dimethylformamide, TBSOTf = *tert*-butyldimethylsilyl trifluoromethanesulfonate, DBU = 1,8-diazabicyclo[5.4.0]undec-7-ene.

Compounds **3**, **5**, **6** and **8** were subsequently submitted to Zemplén deacetylation reaction<sup>23</sup> with the double purpose to remove the acetyl esters protecting groups and to functionalize the C-1 position as 2-iminomethoxyethyl (IME) group in single a synthetic step (Scheme 3). In all the cases, the IME-saccharides **9–12** were obtained in about 50% yield, as evaluated by ESI-MS.

Conjugation of these glycans was performed with the commercially available ribonuclease A (124 amino acids, molecular mass 13.681 Da) containing 10 Lys residues (Fig. 1 and Scheme 3).

First, the reaction of Man-IME (**9**) with RNase A was studied as a model reaction in order to optimize the experimental conditions for the formation of a covalent bond involving the  $\epsilon$ -amino groups of the lysine residues (Table 1). The effects of the buffer, pH and temperature on the reaction outcomes have been investigated through ESI-MS analyses carried out by direct infusion (see Experimental section).

The nature of the buffer and the temperature seem to poorly influence the outcome of the reaction (Entries 1–3, 5 and 6). On the contrary, surprisingly increasing the pH from 9.5 to 10.5 resulted in a strong reduction of the yield from 94% to 58% of the glycosylated protein (Entries 3 and 4), while a decrease of pH value below 9.0 poorly influenced the amount of conjugated protein.

The neoglycoprotein **13** is composed of a mixture of glycoforms with a common peptide backbone but bearing a different number of sugar units. In particular, the most abundant glycoforms are those bearing two or three mannose units (Table 1).

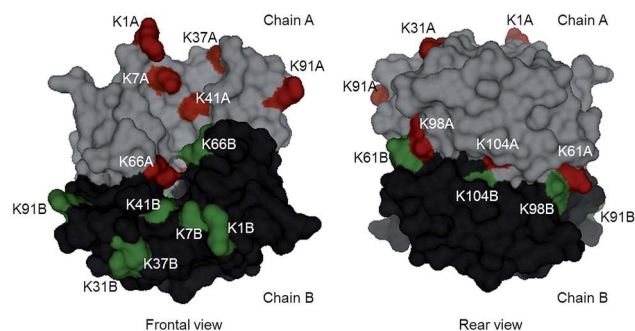
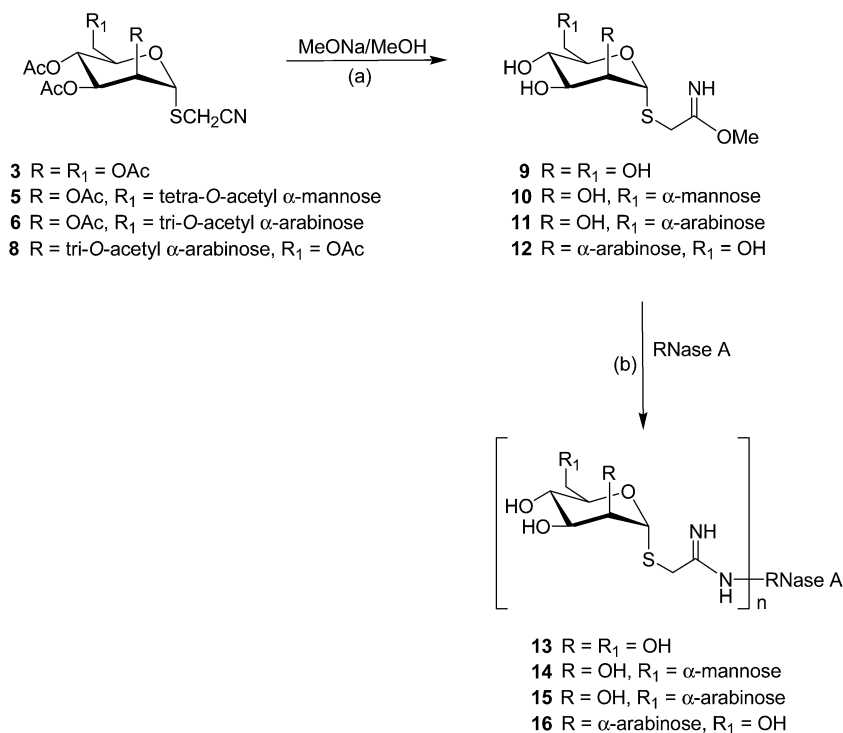


Fig. 1 Surface analysis and localization of Lys residues belonging to each monomer of RNase A dimer.

The glycosylation of RNase was also performed with activated disaccharides **10–12** at the optimal reaction conditions ( $\text{Na}_2\text{B}_4\text{O}_7$  buffer, pH 9.5, 37 °C) (Table 2).

Generally, the glycosylation yields ranged between excellent (93% for Man-IME **9**) and good reaction yields (78% for Ara(1-2) Man-IME **12**). These values were calculated as the ratio between of the sum of the relative abundance of all the glycoforms over the total ions, including the non-glycosylated protein (Table 2). Under a qualitative point of view, the glycosylation profiles varied depending on the glycoside used. In particular, when the monosaccharide **9** was used, the glycosylation ratio (the number of saccharides bound to each molecule of RNase) was 2.3. Similar results have been obtained with IME-disaccharides **9–11**, while a lower ratio (about 1.0) was observed when



Scheme 3 Synthesis of neoglycoproteins **13–16** by conjugation reaction of IME-saccharides **9–12** with ribonuclease A. (a) **3**, **5**, **6** and **8**, MeONa/MeOH, room temperature, 24 h, 50% yield; (b) **9–12**, sodium tetraborate buffer, pH 9.5, 37 °C, molar ratio glycosidic reagent/ribonuclease A 100 : 1, 24 h.

**Table 1** Optimization of experimental conditions for the conjugation reaction of RNase A with Man-IME (9)

Entry	T (°C)	pH	Buffer	N incorporated Man (% relative abundance)								Mannose bound/protein (mol/mol)	Conversion (%)
				0	1	2	3	4	5	6	7		
1	37	8	Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	15	35	25	16	7	1	—	—	1.7	85
2	37	8.5	Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	5	16	29	26	16	7	1	—	2.6	95
3	37	9.5	Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	7	19	29	24	15	5	1	—	2.4	93
4	37	10.5	Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	42	38	20	—	—	—	—	—	0.8	58
5	37	9.5	Na <sub>2</sub> CO <sub>3</sub>	6	9	28	29	20	8	—	—	2.7	94
6	25	9.5	Na <sub>2</sub> B <sub>4</sub> O <sub>7</sub>	—	3	11	18	25	20	14	9	4.0	100

**Table 2** Synthesis of neoglycoproteins 13–16. Experimental conditions: 100 mM sodium tetraborate buffer, pH 9.5, 37 °C, molar ratio glycosidic reagent/ribonuclease A 100 : 1, 24 h

Reagent	Product	N incorporated glycan (% relative abundance)								Glycan bound/protein (mol/mol) ± SD	Conversion (%) ± SD
		0	1	2	3	4	5	6			
9	13	7	19	29	24	15	5	1	2.3 ± 0.3	93 ± 3	
10	14	8	24	27	21	14	6	—	2.3 ± 0.0	92 ± 3	
11	15	9	21	31	24	11	4	—	2.2 ± 0.1	91 ± 2	
12	16	21	39	27	13	—	—	—	1.3 ± 0.1	78 ± 6	

Ara(1-2)Man-IME (12) was used as the glycosylating agent (Table 2). In the last case, the presence of a glycosidic bond in position C-2 of mannose seems to reduce the efficiency of the reaction, probably as a consequence of a steric hindrance.

### Characterization of neoglycoproteins

The analytical characterization of the obtained neoglycoproteins (13–16) was carried out by applying an LC-MS method previously developed for natural glycoproteins, which includes selective proteolysis and analysis of the glycopeptides obtained in order to locate the glycosylation sites.<sup>24,25</sup>

In order to have a direct evidence of the exact positions where the observed multi-site glycosylation has occurred and a quantitative estimate of the glycosylation relevance at each site, all the neoglycoproteins were hydrolyzed by chymotrypsin and the released glycopeptides were selectively extracted, separated and identified by ESI-MS<sup>n</sup>. In detail, glycopeptides were selectively enriched by on-line SPE on hypercarb trap column (PGC-online SPE), and subsequently separated on an Amide 80 column by HILIC-MS/MS. The obtained results (Fig. 2) indicate that the vast majority of conjugation reactions implicate those lysine residues not implicated in the interface of the dimeric form of the protein (Fig. 1), regardless the glycan used. This result confirms that RNase A in solution is mainly present in the dimeric complex. Indeed, only a minor part (<10%) of lysine residues included in the interface of the dimeric aggregate, namely K98 and K104, react with mannose and other glycans, as probable consequence of the reaction of a minor portion of monomeric protein in equilibrium with the dimeric form.

The most reactive area of the protein (almost 40% of glycosylated form) involved K1/K7 amino group. Unfortunately, the glycopeptide obtained by enzymatic digestion contains two Lys residues and the generated MS<sup>2</sup> and MS<sup>3</sup> spectra did not allow

to unambiguously identify the actual glycosylation site. However, taken together the evidences on which the K7 looks on to the dimeric interface (Fig. 1), it might be assumed that the lysine conjugated with the glycans is the K1.

The other positions involved in the reactions are almost equally distributed (about 20% each) among K91 and K31, while a minor amount of RNase was found to be glycosylated at K61-K66.

Interestingly, the same reactivity order of surface lysines was observed when RNase was coupled with either mono- or disaccharides (9–12), notwithstanding the lower glycosylation extent reached with the latter, markedly evident with Ara(1-2)Man-IME 12 (Table 2).

Finally, in attempt to modulate the reactivity of the different lysine residues, the conjugation reaction with Man-IME was performed at lower pH value (8.5 and 8). Reducing the pH of the reaction medium from 9.5 to 8, an improvement of the reactivity of K1 from 40% up to 55% of the total conjugated lysines was observed, while reducing below the 20% the glycosylation extent of K91 which was the second residue in reactivity order (Fig. 3).

### In silico analysis of RNase A lysine reactivity

In order to strengthen the experimental evidence a computational study has been carried out considering different physicochemical parameters of RNase A to gain a theoretical biophysical correlation of the reactivity of the different glycosylation sites of the protein.

To this scope, various critical parameters have been considered: intermolecular protein-protein interactions, multimeric conformations or tendency to aggregation in solution, solvent accessibility for each amino acid side chain, flexibility of different surface area and ionization of the reactive amino acids. By using the PDB structure, we assessed of the probable

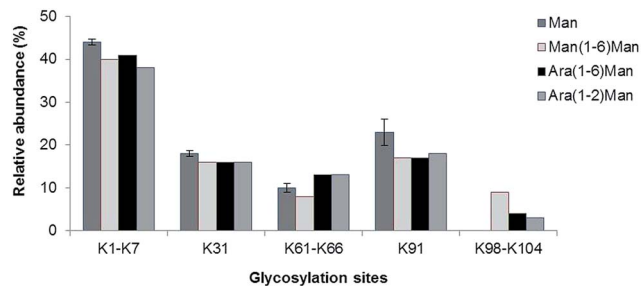


Fig. 2 Glycosylation sites involved in conjugation reaction of IME-saccharides 9–12 with RNase A.

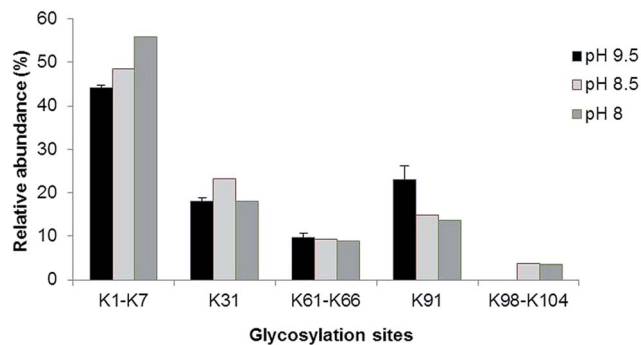


Fig. 3 Glycosylation sites involved in conjugation reaction of IME-Man (9) with RNase A at different pH values.

conformation in solution using the Protein Interfaces, Surfaces and Assemblies (PISA) service at the European Bioinformatics Institute (<https://www.ebi.ac.uk/pdbe/pisa/>). The results showed a tendency of this protein to assembly in a dimeric form in solution. The monomeric form of RNase A exhibited on its surface all lysine residues (Fig. S2, see ESI†). However, only few of them (K1, K31, K61, K66 and K91) were completely exposed on the area of the protein external at the interface (Fig. 1) while others lysines are placed at the limit of the protein surface involved in the dimeric complex, and thus, probably poorly exposed towards the reaction medium. Accordingly, the accessible surface area (ASA) of the different lysines has been analysed including also the amino acids involved in the interface between the two monomers.<sup>26</sup> With the combined PISA and ASA analyses a ranking of the different lysines related to their surface exposition has been obtained. As reported in Fig. 4 and S3 (see ESI†), the most accessible Lys resulted the K1, followed by K66 ≈ K91 > K61 ≈ K98 > K31.

Fig. 5 reports the  $pK_a$  analysis of the Lys of the protein. The lowest  $pK_a$  values are comprised between 8 and 8.5 (K41 and K66). It is noteworthy that, in the dimeric form of the protein, for K66 the  $pK_a$  value decrease from about 10.5 in the monomeric form to about 8.5 in the dimer. Another important parameter to be considered is the protein flexibility. In fact, protein structure varies among different conformations defined by the amino acidic backbone, and the more flexible zones could theoretically better react during a chemical modification.

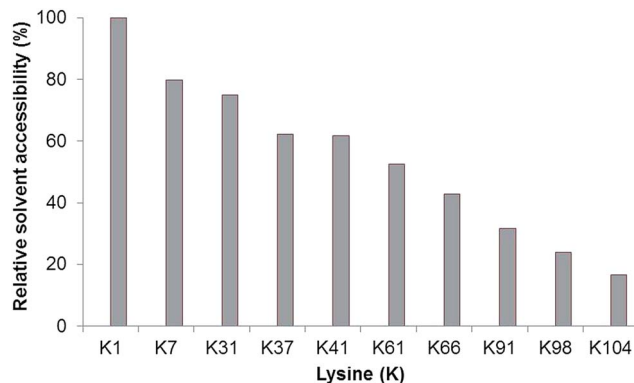


Fig. 4 Analysis of the solvent accessibility (ASA) of the lysines of RNase A (PDB code 1FS3).

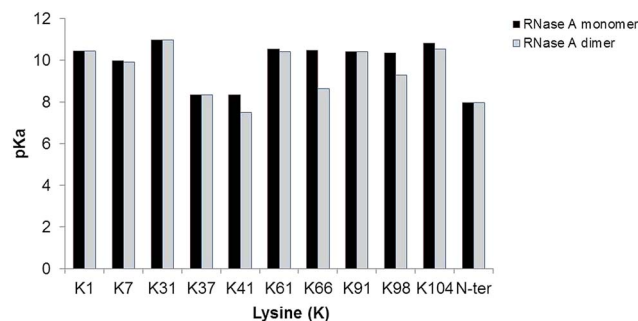


Fig. 5  $pK_a$  values of lysine groups in RNase A based on its 3D structure.

Thanks to the analysis of flexibility in biomolecules and networks service it is possible to estimate the space of vibration for each residues. The results obtained in this analysis suggest that the most flexible areas of RNase are centered on the residues 66–70, 1, 88–93 and 30–39 of the protein sequence (Fig. 6). Especially the first residue range comprised between amino acids 66–70 showed a great fluctuation variation up to  $17 \text{ \AA}^2$ . Accordingly to the evaluation this parameter the most reactive lysine residues could be outlined (K1, K66, K61, K91 or K31).

## Experimental

Lipase from *Candida rugosa* and octyl agarose (Octyl Sepharose CL-4B) was purchased from Sigma-Aldrich (Milano, Italy). Decaoctyl-Sepabeads (Sepabeads EC-OD/S) was kindly provided by Resindion s.r.l. - Mitsubishi Chemical Corporation (Binasco, Milano, Italy). Other reagents and chemicals were purchased from Fluka, Aldrich, Pharmacia Biotech (Uppsala, Sweden) and used without further purification. Solvents were dried by standard methods.

### Thin layer chromatography (TLC)

Analytical TLC was performed on silica gel F<sub>254</sub> precoated aluminum sheets (0.2 mm layer, Merck, Darmstadt, Germany). Products were detected by spraying with 5% H<sub>2</sub>SO<sub>4</sub> in ethanol, followed by heating to ca. 150 °C. Silica gel 60 (40–63 μm, Merck) was used for flash chromatography.

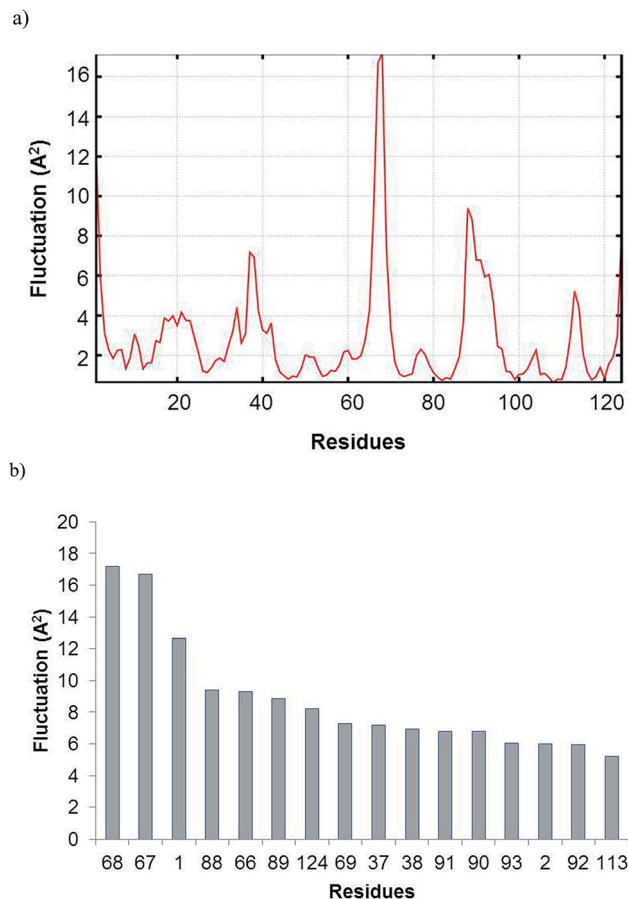


Fig. 6 Flexibility analysis of RNase A structure (PDB: 1FS3) [http://biocomp.chem.uw.edu.pl/CABSflex/]. (a) Fluctuation variation of entire protein sequence, (b) Fluctuation variation ranked by amino acids.

### Nuclear magnetic resonance (NMR)

$^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were recorded in  $\text{CDCl}_3$  at 400.1 and 100.6 MHz, respectively, on a Bruker Avance 400 Spectrometer (Bruker, Karlsruhe, Germany) equipped with a Topspin software package on a workstation running Windows operating system. Chemical shifts ( $\delta$ ) are given in ppm and were referenced to the solvent signals ( $\delta_{\text{H}}$  7.26,  $\delta_{\text{C}}$  77.00). Signal multiplicities are abbreviated as follows: s, singlet; d, doublet; t, triplet; m, multiplet; b, broad.  $^{13}\text{C}$  NMR signal multiplicities were based on attached proton test (APT) spectra. Structures assignment were performed by means of 2D-COSY (Correlation Spectroscopy), HSQC (Heteronuclear Single Quantum Correlation) experiments and HMBC (Heteronuclear Multiple Bond Correlation). Spectra analyses were carried out using iNMR reader software (http://www.inmr.net) on an Apple computer.

### High performance liquid chromatography (HPLC)

A Merck-Hitachi L-7100 liquid chromatograph equipped with a L-7400 UV detector (E. Merck, Darmstadt, Germany) was used for analytical HPLC. Chromatographic conditions: column Phenomenex Gemini C18 250  $\times$  4.6 mm (5  $\mu\text{m}$ ) (Chemtek Analytica,

Anzola Emilia (BO) Italy), eluent phosphate buffer 10 mM-acetonitrile 6 : 4, pH 4, flow rate 1.0 mL  $\text{min}^{-1}$ , UV detector  $\lambda$  220 nm, temperature 25  $^\circ\text{C}$ .

### Mass spectrometry (MS) analyses

MS experiments were carried out on a LTQ-MS (Thermo Electron, San Jose, CA) with an electrospray ionization (ESI) source controlled by X-calibur software 1.4 (Thermo Finnigan, San Jose, CA). The mono- and disaccharide samples were dissolved in methanol and introduced into the mass spectrometer with a syringe pump at a flow rate of 5  $\mu\text{L min}^{-1}$ . Full scan intact MS experiments were carried out under the following instrumental conditions: positive ion mode, mass range 150–1000  $m/z$ , source voltage 4.6 kV, capillary voltage 36 V, sheath gas 6, auxiliary gas 1, capillary temperature 250  $^\circ\text{C}$ , tube lens voltage 80 V. The intact glycoprotein and glycopeptide samples were prepared and analyzed according to the procedures below reported.

### Protein determination

Protein concentration was determined by the Bradford assay on a Shimadzu UV spectrophotometer 1601 (UV Probe 1.0 Shimadzu).

### Enzymatic activity

Enzymatic activities were evaluated by hydrolysis of ethyl butyrate. The hydrolytic activity was quantified titrimetrically at pH 7.0 and room temperature using a pH-Stat (Metrohm 718 STAT Tritino, Herisau, Switzerland).

### In silico analysis of RNase A

Different bioinformatic tools are freely offered by various research groups working in this fast growing topic (http://www.expasy.org/). It has been considered the protein data bank (PDB) structure of bovine pancreatic RNase A, 1FS3. The  $\text{pK}_a$  of each lysine residue was estimated using a bioinformatic tool developed to predict the  $\text{pK}_a$  values of ionisable groups in proteins based on the 3D structure PROPKA (http://propka.ki.ku.dk/). The space of vibration for each lysine residue was evaluate by the analysis of flexibility and networks service (http://biocomp.chem.uw.edu.pl/CABSflex/). The analysis of the solvent accessibility (ASA) was performed according to the literature.<sup>24</sup>

### Ribonuclease A glycosylation with 2-iminomethoxyethyl (IME) thioglycosides

According to B.G. Davis' protocol<sup>27,28</sup> with some minor modifications, the glycosylation reactions were carried out in sodium tetraborate buffer 100 mM, pH 9.5. Ribonuclease A was dissolved in the buffer at a concentration of 1 mg  $\text{mL}^{-1}$ , then to the solution was added IME-thioglycosides to a final glycoside/protein molar ratio of 100/1. The reaction mixtures were vortexed for 1 minute and incubated for 24 hours at 37  $^\circ\text{C}$  under continuous stirring.

## Glycoproteins purification

After glycosylation, the reaction mixtures were purified in order to make them suitable to direct infusion in ESI-MS. All the samples were submitted to ultracentrifugation using filter with loading capacity of 500  $\mu\text{L}$  and Nominal Molecular Weight Limit (NMWL) of 3 kDa. The purification protocol consists of seven centrifugation steps of 20 min at 13 000g and 4 °C using deionized water as washing solvent. After purification, the glycoproteins were collected in deionized water and stored at  $-20$  °C. For long period storage, ammonium bicarbonate 10 mM, pH 7.4 was used.

## Glycosylation yield determination

Glycosylation yields and relative abundance of glycoforms were determined by direct infusion on a ESI-LTQ-MS (Thermo Electron, San Jose, CA). A small rate of purified reaction mixture from glycosylation reaction was diluted to a final concentration of 0.3 mg mL<sup>-1</sup> in H<sub>2</sub>O : acetonitrile, 50 : 50 containing 0.05% TFA and introduced into the mass spectrometer with a syringe pump at a flow rate of 5  $\mu\text{L min}^{-1}$ . 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, sheath gas 15, auxiliary gas 2, capillary temperature 220 °C, tube lens voltage 140 V.

The spectra were deconvoluted with Bioworks Browser (Thermo Electron, revision 3.1). The identities of ribonuclease A and its glycoforms were assigned on the basis of the average molecular weight in the deconvoluted spectra and their relative abundance determined by the relative intensities of the corresponding peaks.

## Digestion with chymotrypsin

The reaction mixtures were digested with chymotrypsin as follows: to a 100  $\mu\text{L}$  of protein solution 100  $\mu\text{M}$  in water was added 90  $\mu\text{L}$  of ammonium bicarbonate 100 mM pH 8.5 and 10  $\mu\text{L}$  of dithiothreitol (DTT) solution (100 mM in ammonium bicarbonate 100 mM pH 8.5). The solution was heated at 60 °C for 30 minutes in order to reduce the disulfide bridges. Enzyme was then added to a final protein/enzyme ratio of 100 : 1. The solution was incubated overnight at 37 °C under continuous stirring. The digestion was stopped by inactivating the chymotrypsin with TFA 2.5% (v/v).

## On-line PGC-Solid-Phase Extraction and HPLC-MS analysis

Glycopeptides obtained by chymotrypsin digestion were selectively *on-line* extracted rearranging a method of Temporini *et al.*<sup>25</sup> The final conditions provide 10 minutes of desorption from the trap column with 80% A (acetonitrile with TFA 0.05%)/20% B (H<sub>2</sub>O with TFA 0.05%) at 100  $\mu\text{L min}^{-1}$ . The HPLC gradient analysis was from 30% to 57% B in 11 minutes with a flow rate of 200  $\mu\text{L min}^{-1}$ . Mass spectra were generated in positive ion mode under the following instrumental conditions: source voltage 4.5 kV, capillary voltage 31 V, sheath gas flow 40 (arbitrary units), auxiliary gas flow 10 (arbitrary units), sweep gas flow 0 (arbitrary units), capillary temperature 250 °C, tube lens voltage 95 V. MS<sup>2</sup> and MS<sup>3</sup> spectra were obtained by CID with normalized collision energy 35.0.

Glycopeptides were identified on the basis of fragmentation ions in MS<sup>2</sup> and MS<sup>3</sup> spectra by *in silico* screening with Bioworks Browser (Thermo Electron, revision 3.1) setting the glycan moieties as optional modifications of lysine residues. Only the identification with a X-corr greater than 1 were considered and to avoid false positive MS<sup>2</sup> and MS<sup>3</sup> spectra of all species recognized as glycopeptides were manually confirmed.

For quantitative measurements, the area of each glycopeptide was determined by integrating the peak in the respective single ion current chromatogram.

## 1-Thio-S-cyanomethyl-2,3,4,6-tetra-O-acetyl- $\alpha$ -D-mannopyranoside (3)

Synthesis of 1-thio-S-cyanomethyl-2,3,4,6-tetra-O-acetyl- $\alpha$ -D-mannopyranoside (3) was performed according to the procedure previously reported.<sup>27</sup>

## General procedure for enzymatic hydrolysis

The enzymatic hydrolyses of the mannose derivatives 1–3 (between 5 and 20 mM) were carried out in 25 mM phosphate buffer containing 20% acetonitrile for complete solubilisation of the substrates under mechanical stirring. The reactions started after addition of the immobilized enzyme preparation. During the reactions, pH was maintained constant by automatic titration. The course of the hydrolyses was monitored by HPLC and products were quantified using authentic standards. After complete or almost complete consumption of the substrates, the reactions were stopped by biocatalyst filtration and the obtained products were isolated by extraction with ethyl acetate followed by purification using flash chromatography, and were characterized by <sup>1</sup>H and COSY 2D NMR.

## 1,2,3,4-Tetra-O-acetyl- $\alpha$ -D-mannopyranose (1a)

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  = 6.1 (s, 1H, H1), 5.40 (dd, 1H, *J* = 3.3, 1.7 Hz, H-2), 5.30 (dd, *J* = 3.4, 1.4 Hz, 1H, H-3), 5.27 (m, 1H, H-4), 3.70 (dd, *J* = 11.9, 6.5 Hz, 1H, H-6B), 3.61 (dd, *J* = 11.6, 6.4 Hz, 1H, H-6A), 3.83 (m, 1-H, H-5), 2.15 (s, 3H), 2.10 (s, 3H), 2.05 (s, 3H), 2.00 (s, 3H), CH<sub>3</sub>COO.

## Methyl 2,3,4-tri-O-acetyl- $\alpha$ -D-mannopyranoside (2a)

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  = 5.90 (s, 1H, H1), 5.55 (dd, *J* = 3.2, 1.6 Hz, 1H, H-2), 5.25 (m, 1H, H-4), 4.98 (dd, *J* = 3.6, 1.3 Hz, 1H, H-3), 4.14 (m, 1-H, H-5), 3.80 (dd, *J* = 11.4, 6.8 Hz, 1H, H-6B), 3.66 (dd, *J* = 11.2, 6.1 Hz, 1H, H-6A), 3.30 (s, 3H, OCH<sub>3</sub>), 2.20 (s, 3H), 2.15 (s, 3H) and 2.07 (s, 3H), CH<sub>3</sub>COO.

## Cyanomethyl 2,3,4-tri-O-acetyl- $\alpha$ -D-thiomannopyranoside (3a)

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  = 5.50 (s, 1H, H-1), 5.40 (dd, *J* = 3.4, 1.7 Hz, 1H, H-2), 5.35 (t, 1H, H-4), 5.27 (dd, *J* = 3.2, 1.2 Hz, 1H, H-3), 4.10–4.18 (m, 1H, H-5), 3.82 (dd, *J* = 11.2, 6.2 Hz, 1H, H-6B), 3.69 (dd, *J* = 11.6, 6.4 Hz, 1H, H-6A), 3.49 (d, *J* = 17.22 Hz, 1H, HA SCH<sub>2</sub>CN), 3.35 (d, *J* = 17.23 Hz, 1H, HB SCH<sub>2</sub>CN), 2.19 (s, 3H), 2.12 (s, 3H) and 2.03 (s, 3H), CH<sub>3</sub>COO.



**Cyanomethyl 3,4,6-tri-*O*-acetyl- $\alpha$ -D-thiomannopyranoside (3b)**

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 5.55 (s, 1H, H-1), 5.42 (dd,  $J$  = 9.7, 9.8 Hz, 1H, H-4), 5.18 (dd,  $J$  = 3.2, 1.2 Hz, 1H, H-3), 4.30–4.40 (m, 2H, H-5, H-6A), 4.15–4.22 (m, 2H, H-2, H-6B), 3.48 (d,  $J$  = 17.2 Hz, 1H, HA  $\text{SCH}_2\text{CN}$ ), 3.35 (d,  $J$  = 17.2 Hz, 1H, HB  $\text{SCH}_2\text{CN}$ ), 2.16 (s, 3H), 2.12 (s, 3H) and 2.06 (s, 3H),  $\text{CH}_3\text{COO}$ .

**Recycling of the biocatalysts**

Recycling of CRL immobilized on decaoctyl Sepabeads was performed by evaluating the hydrolysis of substrate 2 in the same conditions above described. When the highest conversion was achieved, the reaction mixture was filtered under reduced pressure and the immobilized biocatalyst was re-used for the next reaction.

**Cyanomethyl (2,3,4,6-tetra-*O*-acetyl- $\alpha$ -D-mannopyranosyl)-(1  $\rightarrow$  6)-2,3,4-tri-*O*-acetyl- $\alpha$ -D-thiomannopyranoside (5)**

2,3,4,6-Tetra-*O*-acetyl- $\alpha$ -D-mannopyranose trichloroacetimidate 4 (0.136 g; 0.276 mmol) and cyanomethyl 2,3,4-tri-*O*-acetyl- $\alpha$ -D-thiomannopyranoside 3a (0.050 g; 0.138 mmol) were dissolved in anhydrous  $\text{CH}_2\text{Cl}_2$  (5 mL) and the solution was cooled to 0 °C. Subsequently, boron trifluoride diethyl etherate (0.01 mL) was added in the presence of activated 4 Å molecular sieves (0.300 g), under magnetic stirring under a nitrogen atmosphere. After 6 hours, the reaction mixture was neutralised with the addition of triethylamine (0.006 mL). The solvent was removed under reduced pressure and the residue was purified by flash column chromatography (hexane–ethyl acetate, 1 : 1) affording 5 (white amorphous solid, 0.067 g, 70% yield).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 5.48 (d,  $J$  = 1.4 Hz, 1H, H-1), 5.39 (dd,  $J$  = 3.2, 1.6 Hz, 1H, H-2), 5.35–5.21 (m, 5H, H-4', H-3', H-2', H-4, H-3), 4.85 (d, 1H,  $J$  = 1.4 Hz, H-1'), 4.29–4.25 (m, H-5, 2H, H-6A), 4.13 (dd,  $J$  = 11.2, 6.1 Hz, 1H, H-6B), 4.05–4.00 (m, 1H, H-5'), 3.85 (dd,  $J$  = 11.6, 6.8 Hz, 1H, H-6'B), 3.58 (dd,  $J$  = 11.6, 3.2 Hz, 1H, H-6'A), 3.54 (d,  $J$  = 17.3 Hz, 1H, HA  $\text{SCH}_2\text{CN}$ ), 3.36 (d,  $J$  = 17.3 Hz, 1H, HB  $\text{SCH}_2\text{CN}$ ), 2.19 (s, 3H), 2.16 (s, 3H), 2.11 (s, 3H), 2.08 (s, 3H), 2.06 (s, 3H), 2.00 (s, 3H) and 1.99 (s, 3H),  $\text{CH}_3\text{COO}$ .

$^{13}\text{C}$  NMR (101 MHz,  $\text{CDCl}_3$ ):  $\delta$  = 170.20, 170.03, 169.97, 169.86, 169.68 ( $\text{CH}_3\text{COO}$ ), 115.74 ( $\text{SCH}_2\text{CN}$ ); 97.36 (C-1'), 81.55 (C-1), 70.63 (C-5), 69.56 (C-2), 69.46 (C-3'), 69.08 (C-3), 68.97 (C-5'), 66.52 (C-4), 66.21 (C-6), 65.87 (C-4'), 62.51 (C-6'), 21.01, 20.94, 20.85, 20.81, 20.71 ( $\text{CH}_3\text{COO}$ ), 15.47 ( $\text{SCH}_2\text{CN}$ ).

MS:  $m/z$  = 714.29 [ $\text{M} + \text{Na}$ ] $^+$  (calcd 714.63).

**Cyanomethyl (2,3,5-tri-*O*-acetyl- $\alpha$ -D-arabinofuranosyl)-(1  $\rightarrow$  6)-2,3,4-tri-*O*-acetyl- $\alpha$ -D-thiomannopyranoside (6)**

2,3,5-Tri-*O*-acetyl- $\alpha$ -D-arabinofuranosyl trichloroacetimidate 7 (0.373 g, 0.886 mmol) and cyanomethyl 2,3,4-tri-*O*-acetyl- $\alpha$ -D-thiomannopyranoside 3a (0.160 g, 0.443 mmol) were dissolved in anhydrous  $\text{CH}_2\text{Cl}_2$  (20 mL) and the mixture was stirred to room temperature. Boron trifluoride diethyl etherate (0.028 mL) was added in the presence of activated 4 Å molecular sieves

(0.500 g), and the solution was left to stir under nitrogen. After stirring for 1 hour, the reaction mixture was neutralized by the addition of triethylamine (0.018 mL). The solvent was removed under reduced pressure and The solvent was removed under reduced pressure and the residue was purified by flash column chromatography (hexane–ethyl acetate, 1 : 1) affording 6 (white amorphous solid, 0.170 g, 62% yield).

$^1\text{H}$ -NMR (400 MHz;  $\text{CDCl}_3$ ):  $\delta$  = 5.45 (d,  $J$  = 1.4 Hz, 1H, H-1), 5.34 (dd,  $J$  = 3.5, 1.4 Hz, 1H, H-2), 5.30 (dd,  $J$  = 9.8, 9.8 Hz, 1H, H-4), 5.24 (ddd,  $J$  = 3.5, 3.5, 1.7 Hz, 1H, H-4'), 5.20 (dd,  $J$  = 9.8, 3.5 Hz, 1H, H-3), 5.17 (dd,  $J$  = 9.2, 6.7 Hz, 1H, H-2'), 5.04 (dd,  $J$  = 9.2, 3.5 Hz, 1H, H-3'), 4.51 (d,  $J$  = 6.7 Hz, 1H, H-1'), 4.24 (ddd,  $J$  = 9.8, 6.5, 2.4 Hz, 1H, H-5), 4.04 (dd,  $J$  = 13.0, 3.5 Hz, 1H, H-5'A), 3.96 (dd,  $J$  = 11.7, 6.5 Hz, 1H, H-6A), 3.66 (dd,  $J$  = 11.7, 2.4 Hz, 1H, H-6B), 3.62 (dd,  $J$  = 13.0, 1.7 Hz, 1H, H-5'B), 3.54 (d,  $J$  = 17.2 Hz, 1H, HA  $\text{SCH}_2\text{CN}$ ), 3.28 (d,  $J$  = 17.2 Hz, 1H, HB  $\text{SCH}_2\text{CN}$ ), 2.17 (s, 3H), 2.13 (s, 3H), 2.09 (s, 3H), 2.06 (s, 3H), 2.04 (s, 3H) and 1.99 (s, 3H),  $\text{CH}_3\text{COO}$ .

$^{13}\text{C}$ -NMR (101 MHz;  $\text{CDCl}_3$ ):  $\delta$  = 170.46, 170.30, 169.99, 169.87, 169.70 and 169.58,  $\text{CH}_3\text{COO}$ , 116.04 ( $\text{SCH}_2\text{CN}$ ), 100.78 (C-1'), 81.50 (C-1), 70.94 (C-5), 70.05 (C-3'), 69.74 (C-2), 69.53 (C-3), 69.15 (C-2'), 67.60 (C-4'), 66.88 (C-6), 66.47 (C-4), 63.23 (C-5'), 21.08, 21.00, 20.89 and 20.80,  $\text{CH}_3\text{COO}$ , 15.22 ( $\text{SCH}_2\text{CN}$ ).

MS:  $m/z$  = 642.33 [ $\text{M} + \text{Na}$ ] $^+$  (calcd 642.58).

**Cyanomethyl (2,3,5-tri-*O*-acetyl- $\alpha$ -D-arabinofuranosyl)-(1  $\rightarrow$  2)-3,4,6-tri-*O*-acetyl- $\alpha$ -D-thiomannopyranoside (8)**

2,3,5-Tri-*O*-acetyl- $\alpha$ -D-arabinofuranosyl trichloroacetimidate 7 (0.650 g, 1.544 mmol), and cyanomethyl 3,4,6-tri-*O*-acetyl- $\alpha$ -D-thiomannopyranoside 3b (0.280 g, 0.775 mmol) were coevaporated with anhydrous toluene and dried under reduced pressure overnight. The mixture was dissolved in  $\text{Et}_2\text{O}$  (25 mL), cooled to 0 °C and *tert*-butyldimethylsilyl trifluoromethanesulfonate (0.178 mL) was added in the presence of activated 4 Å molecular sieves (0.500 g). The reaction solution was left to stir under nitrogen to 0 °C. After stirring for 3 hours the reaction mixture was neutralized by the addition of triethylamine (0.130 mL). The solvent was removed under reduced pressure and The solvent was removed under reduced pressure and the residue was purified by flash column chromatography (hexane–ethyl acetate, 1 : 1) affording 8 (white amorphous solid, 0.240 g, 50% yield).

$^1\text{H}$ -NMR (400 MHz;  $\text{CDCl}_3$ ):  $\delta$  = 5.59 (d,  $J$  = 1.3 Hz, 1H, H-1), 5.30 (t,  $J$  = 10.0 Hz, 1H, H-4), 5.28–5.24 (m, 2H, H-4 and H-4'), 5.04 (dd,  $J$  = 3.2, 1.2 Hz, 1H, H-3), 5.02 (dd,  $J$  = 3.0, 1.3 Hz, 1H, H-3'), 4.43 (dd,  $J$  = 12.5, 4.0 Hz, 1H, H-6A), 4.38 (d,  $J$  = 7.0 Hz, 1H, H-1'), 4.25(ddd,  $J$  = 10.1, 4.0, 2.0 Hz, 1H, H-5), 4.16 (dd,  $J$  = 3.2, 1.3 Hz, 1H, H-2), 4.14–4.09 (m, 2H, H-2 and H-6B), 4.02 (dd,  $J$  = 13.1, 3.1 Hz, 1H, H-5'A), 3.59 (dd,  $J$  = 13.1, 1.6 Hz, 1H, H-5B), 3.45 (d,  $J$  = 17.2 Hz, 1H, HA  $\text{CH}_2\text{CN}$ ), 3.30 (d,  $J$  = 17.2 Hz, 1H, HB  $\text{CH}_2\text{CN}$ ), 2.17 (s, 3H), 2.16 (s, 3H), 2.15 (s, 3H), 2.07 (s, 3H), 2.04 (s, 3H) and 2.03 (s, 3H),  $\text{CH}_3\text{COO}$ .

$^{13}\text{C}$ -NMR (101 MHz;  $\text{CDCl}_3$ ):  $\delta$  = 171.03, 170.43, 170.38, 170.20, 169.69 and 169.38 ( $\text{CH}_3\text{COO}$ ), 116.05 ( $\text{SCH}_2\text{CN}$ ), 102.65 (C-1'), 84.45 (C-1), 76.30 (C-2), 71.34 (C-3), 69.87 (C-3'), 69.80

(C-5), 68.75 (C-2'), 67.51 (C-4), 65.63 (C-4'), 63.55 (C-5'), 61.56 (C-6), 29.84, 21.08, 20.87 and 20.76 (CH<sub>3</sub>COO), 16.02 (SCH<sub>2</sub>CN).

MS:  $m/z = 642.50 [M + Na]^+$  (calcd 642.58).

### General Procedure for synthesis of IME mono- and disaccharides 9–12

Compounds 3–8 (1.0 mmol) were dissolved in 5 mL of dry methanol and 1 wt% of sodium methoxide (1.6 mmol) was added. The reaction was reacted for 48 h under nitrogen, after which time the reaction mixture was concentrated *in vacuo* and the solid formed was characterized by MS analysis.

Man-IME (9) MS:  $m/z = 290.33 [M + Na]^+$  (calcd 290.29).

Man(1-6)Man-IME (10) MS:  $m/z = 452.75 [M + Na]^+$  (calcd 452.64).

Ara(1-6)Man-IME (11) MS:  $m/z = 422.33 [M + Na]^+$  (calcd 422.40).

Ara(1-2)Man-IME (12) MS:  $m/z = 422.33 [M + Na]^+$  (calcd 422.40).

## Conclusions

In this work we presented a convenient and integrated approach to the synthesis and characterization of neoglycoproteins prepared by non-selective glycosylation of lysine  $\epsilon$ -amino groups of RNase A, chosen as a model protein. More in detail, by means of a chemo-biocatalytic approach, different activated mannose-based glycans were synthesized in good or excellent yields and efficiently coupled to the model protein *via* reaction with 2-iminomethoxyethyl reactive group generated on mannose glycans. With the aim to depict a reactivity map of protein surface and to support the analytical characterization of newly generated glycoproteins, an investigation of different biophysical parameters of this protein was conducted. Thanks to suitable LC-MS analyses the reacting lysines of RNase A were detected demonstrating a good correlation between the most reactive conjugation sites and the study of the solvent accessibility and flexibility of the different areas of the protein surface containing the lysine residues. By this way, the feasibility of this proposed approach to support the design and characterization of neoglycoproteins has been clearly confirmed. In particular, among all reactive lysines of RNase A only those exposed outside the interface of the dimeric complex can react with the glycans. The solvent accessibility and the flexibility of the different reactive lysines exposed externally to the interface seems to be the most relevant parameters regulating the reactivity of the different residues. The ionization of the different residues ( $pK_a$ ) seems instead to have a very poor contribution in the differentiation of the reactivity of the different lysines. Accordingly, by modulating the pH of the reaction medium, the preferential reactivity of lysines (K1 and K31) can be exalted and an almost

selective glycosylation of these two residues can be obtained, probably because at low pH, the reactivity of all lysines is depressed (as demonstrated by the reduction of the percentage of glycosylated protein), and only the most exposed and/or flexible residues maintain a good reactivity.

The possibility to drive the conjugation with glycans toward defined areas of the protein surface can be very important in case of antigenic and other bioactive proteins, in order to avoid modification of the residues involved in the biological activity. To this scope and to widen the possible application areas, studies involving antigenic protein structures are ongoing.

## Acknowledgements

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. M. F. thanks to CSIC for a JAE-Doc contract ("Junta para la Ampliacion de estudios") cofounded by ESF (European Social Fund).

## Notes and references

- 1 R. D. Astronomo and D. R. Burton, *Nat. Rev. Drug Discovery*, 2010, **9**, 308.
- 2 G. J. L. Bernardes, B. Castagner and P. H. Seeberger, *ACS Chem. Biol.*, 2009, **4**, 703.
- 3 G. Kallenius, A. Pawlowski, B. Hamasur and S. B. Svenson, *Trends Microbiol.*, 2008, **16**, 456.
- 4 S. Frokjaer and D. E. Otzen, *Nat. Rev. Drug Discovery*, 2005, **4**, 298.
- 5 D. G. Ilyushin, I. V. Smirnov, A. A. Belogurov, I. A. Dyachenko, T. I. Zharmukhamedova, T. I. Novozhilova, E. A. Bychikhin, M. V. Serebryakova, O. N. Kharybin, A. N. Murashev, K. A. Anikienko, E. N. Nikolaev, N. A. Ponomarenko, D. D. Genkin, G. M. Blackburn, P. Masson and A. G. Gabibov, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **110**, 1243.
- 6 M. Monsigny, A. C. Roche, E. Duverger and O. Srinivas, Neoglycoproteins. *Comprehensive Glycoscience, From Chemistry to Systems Biology*, Elsevier, Amsterdam, 2007, p. 477.
- 7 B. G. Davis, *Chem. Rev.*, 2002, **102**, 579.
- 8 D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131.
- 9 E. J. Grayson, G. J. L. Bernardes, J. M. Chalker, O. Boutureira, J. R. Koeppel and B. G. Davis, *Angew. Chem., Int. Ed.*, 2011, **50**, 4127.
- 10 O. M. T. Pearce, K. D. Fisher, J. Humphries, L. W. Seymour, A. Smith and B. G. Davis, *Angew. Chem., Int. Ed.*, 2005, **44**, 1057.
- 11 J. Zhu, Q. Wan, D. Lee, G. Yang, M. K. Spassova, O. Ouerfelli, G. Ragupathi, P. Damani, P. O. Livingston and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2009, **131**, 9298.
- 12 C. H. Hsu, S. C. Hung, C. Y. Wu and C. H. Wong, *Angew. Chem., Int. Ed.*, 2011, **50**, 11872.

- 13 M. Filice, J. M. Palomo, P. Bonomi, T. Bavaro, R. Fernandez-Lafuente, J. M. Guisan and M. Terreni, *Tetrahedron*, 2008, **64**, 9286.
- 14 M. Filice, T. Bavaro, R. Fernandez-Lafuente, M. Pregnotato, J. M. Guisan, J. M. Palomo and M. Terreni, *Catal. Today*, 2009, **140**, 11.
- 15 M. Filice, J. M. Guisan, M. Terreni and J. M. Palomo, *Nat. Protoc.*, 2012, **7**, 1783.
- 16 Y. J. Kim and M. L. Doyle, *Anal. Chem.*, 2010, **82**, 7083.
- 17 K. Sandra, P. Sandra and I. Vandenheede, *LC GC Eur.*, 2013, **26**, 10.
- 18 A. Staub, D. Guillarme, J. Schappler, J. L. Veuthey and S. Rudaz, *J. Pharm. Biomed. Anal.*, 2011, **55**, 810.
- 19 L. East and C. M. Isacke, *Acta Biochim. Biophys. Sin.*, 2002, **1572**, 364.
- 20 J. N. Arnold, M. R. Wormald, R. B. Sim, P. M. Rudd and R. A. Dwek, *Annu. Rev. Immunol.*, 2007, **25**, 21.
- 21 A. Holemann, B. L. Stocker and P. H. Seeberger, *J. Org. Chem.*, 2006, **71**, 8071.
- 22 J. M. Palomo, M. Filice, R. Fernandez-Lafuente, M. Terreni and J. M. Guisan, *Adv. Synth. Catal.*, 2007, **349**, 1969.
- 23 Z. Wang, *Comprehensive Organic Name Reactions and Reagents*, John Wiley & Sons, 2010, p. 3123.
- 24 C. Temporini, E. Perani, E. Calleri, L. Dolcini, D. Lubda, G. Caccialanza and G. Massolini, *Anal. Chem.*, 2007, **79**, 355.
- 25 C. Temporini, P. Bonomi, I. Serra, A. Tagliani, T. Bavaro, D. Ubiali, G. Massolini and M. Terreni, *Biomacromolecules*, 2010, **11**, 1623.
- 26 C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 2284.
- 27 C. W. Chang, S. S. Chang, C. S. Chao and K. K. T. Mong, *Tetrahedron Lett.*, 2009, **50**, 4536.
- 28 B. G. Davis, S. I. Van Kasteren, D. Anthony and N. Sibson, Imaging agent, International Patent, WO 2007/020450 A2, 22 February 2007.