Structural glycobiology of the major allergen of *Artemisia vulgaris* pollen, Art v 1: O-glycosylation influence on the protein dynamics and allergenicity

Laercio Pol-Fachin2 and Hugo Verli1,2,3

1Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970 RS, Brazil and 2Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000 RS, Brazil

Received on January 18, 2012; revised on February 13, 2012; accepted on February 14, 2012

Art v 1 is the major allergen of mugwort (*Artemisia vulgaris*) pollen. It is formed by an N-terminal globular defensin-like part and a C-terminal proline-rich domain. As the structure and the dynamics of Art v 1 have been mostly described for its recombinant, non-glycosylated form, which does not occur in normal plant physiology, the present work intends to obtain a three-dimensional model for Art v 1 native O-glycosylation structure and to evaluate the influence of such glycans over the protein dynamics and allergenicity through molecular dynamics simulations in triplicates. Structural insights into the mutual recognition of Art v 1 protein and carbohydrate moieties recognition by antibodies were obtained, in which glycan chains remained close to the previously identified epitopes in the defensin-like domain, thus pointing to potential interferences with antibodies recognition. To our knowledge, this is the first structural report of an entire furanose-containing glycoprotein. As well, together with the previously determined NMR structures, the obtained results contribute in the comprehension of the effect of glycosylation over both proline-rich and defensin-like domains, providing an atomic representation of such alterations.

**Keywords:** allergy / antibody recognition / arabinofuranose / glycosidic linkage / molecular dynamics

**Introduction**

Art v 1 is the major allergen of mugwort (*Artemisia vulgaris*) pollen (Gadermaier et al. 2004), which is a widely spread plant in temperate and humid zones of the Northern hemisphere and along the Mediterranean Basin (Spieksma et al. 1980). Structurally, Art v 1 is a 108-amino acid glycoprotein formed by two domains (Himly et al. 2003). At the N-terminal, there is a globular defensin-like, disulfide-bond rich structural unit, which is the main target of IgE antibodies in atopic individuals (Himly et al. 2003; Dedic et al. 2009) and contains the main epitope for T-cell recognition (Jahn-Schmid et al. 2002). Conversely, a proline-rich domain containing several (Ser/Ala)-(Pro)2-4 extensin-like repeats is located at Art v 1 C-terminal, which is post-translationally modified through proline 4-hydroxylation and O-glycosylation (Himly et al. 2003). The 4-hydroxyproline (HyP) residues at the C-terminal domain may be added by two types of glycans: the most abundant consist of single β-arabinofuranoses, also partially or mainly recognized by antibodies in the subgroups of *Artemisia*-allergic patients (Himly et al. 2003; Leonard et al. 2005); and the less abundant comprises a large arabinogalactan, composed of three galactopyranose (Galp) and at least 10 arabinofuranose (Araf) residues (Himly et al. 2003; Leonard et al. 2005). Accordingly, Art v 1 may be found in a minor prevalence 13-kDa glycoform, only presenting β-Araf-linked glycans, or in the mainly occurring 15 kDa glycosylation state, also containing, beyond β-Araf, one or two arabinogalactan oligosaccharides (Himly et al. 2003). Although such glycans may represent from 30 to 50% of the mature protein molecular mass, mostly no three-dimensional (3D) information is available for them, and Art v 1 protein moiety could only be completely solved in its recombinant form through NMR techniques in the absence of O-linked glycans (Razzera et al. 2009). From such assignments, it could be observed that, although the defensin-like domain shows a relatively unvariable conformation, its C-terminal part presents a high flexibility (Razzera et al. 2010), thus assuming several conformational states. As well, distinct chemical shifts have been identified for some regions of the recombinant (non-glycosylated) and natural (glycosylated) Art v 1 (Razzera et al. 2010), although the structural source of these differences could not be distinguished.

In this context, the current work intends to obtain a 3D model for the completely glycosylated Art v 1 O-linked and to evaluate the effects of glycosylation over the structure and dynamics of its protein moiety through molecular dynamics (MD) simulations in aqueous solutions. The oligosaccharides...
were assessed and constructed based on their composing disaccharidic units, through energy contour maps and solution MD simulations conformational sampling. The obtained glycan chains were attached to the protein core and further refined through MD. The obtained results were discussed in terms of the differences between Art v 1 non-glycosylated and glycosylated forms, as well as its (hydroxy)proline-rich domain flexibility and dynamics. Such results were compared with experimental data, mainly derived from NMR results, supporting further insights into glycosylation role on Art v 1 dynamics and allergenicity.

Results

Systems preparation

In order to explore Art v 1 conformation and flexibility in aqueous solutions, including its glycan chains, we have studied such protein in increasing complexity levels, from non-glycosylated states, containing either Pro or HyP in the C-terminal segment, to both its minor and major prevalence glycoforms. Additionally, in order to provide a higher conformational sampling, each condition was assessed in triplicates, each one starting from a distinct protein conformation, as derived from clustering analysis based on NMR data (Razzera et al. 2010). Six non-glycosylated systems were simulated, being three for Pro-containing structures and three for Art v 1 containing HyP residues. Six systems were assembled for an Art v 1 minor prevalence glycoform, thus lacking the arabinogalactan oligosaccharide, but presenting mono-β-Araf at every HyP located within (Ser)-(HyP)_{2-4} motifs. Three of these structures contained Araf in its \(1T_2\) form, and the other three presented such residue in its \(4T_3\) conformation. Finally, the Art v 1 major prevalence glycoform was considered in six additional systems, also accounting for both Araf-puckering states, presenting the arabinogalactan glycan linked to Hyp79 (see the Materials and methods section) and mono-β-Araf at all other HyP residues within (Ser)-(HyP)_{2-4} peptides.

Concerning Art v 1-glycosylated states, mostly no 3D data were available for constructing its O-linked glycans (Himly et al. 2003; Leonard et al. 2005). In this context, for the structures of the large arabinogalactan oligosaccharide (Figure 1A) and the mono-β-Araf (Figure 1B), considering both φ-ψ and ω-ψ combinations and varying the Araf puckering in its two possible conformations (i.e., \(1T_2\) and \(4T_3\)), 3 O-glycosidic linkages (between monosaccharides and 4-HyP) and 12 disaccharides were studied through energy contour maps and MD.
Glycosidic-linkage conformational profile

Based on the data presented in Supplementary data, Figure S1 and Table I, it may be observed that all the α-anomers preceded by Araβ residues were mainly populated at negative φ values, independently on the furanose-puckering form, as previously observed for the α-Araβ(1→5)-Araβ in previous works (Cros et al. 1994; Pérez et al. 2000). Regarding (1→5) and (1→6) linkages, the φ angle only adopted three main conformational peaks (Supplementary data, Figure S1H, N, O, R and S), around −180°, −60° and 60°, in accordance with Shefter and Trueblood’s convention (Shefter and Trueblood 1965). Although only minor differences in glycosidic-linkage geometries were observed between disaccharides presenting distinct Araβ-puckering states, the global shape of the complete arabinogalactan oligosaccharides, when composed by such conformations, is quite different (Figure 1C and D).

Furthermore, from both disaccharides and glycoproteins MD simulations, as presented in Table I, no major differences within standard deviations were observed when comparing glycosidic linkages on different complexity levels, i.e. either as isolated disaccharides or composing glycoproteins, as previously observed for N-glycans composing disaccharides (Fernandes et al. 2010).

Effects of glycosylation over Art v 1 protein moiety structure and dynamics

For all the glycosylation states and systems evaluated, the defensin-like domain maintained its original secondary structure content, as monitored by PROCHECK analyses (Supplementary data, Figure S2), as well as its tertiary structure and globularity during MD simulations (Supplementary data, Figure S3). In addition, interconversions could not be observed within the simulated time scale between the three clusters of conformations derived from the NMR data in either of the simulated conditions, as evaluated by performing clustering analyses with the same cut-off value employed under the available NMR models (see Materials and methods section).

Regarding the (Ser)-(Pro)4 extensin-like motifs conformation, only point differences between the evaluated conditions could be observed as a consequence of mono-β-arabinosylation (Supplementary data, Table S1), which agrees with previous findings suggesting that mono-Araβ are not capable of modifying the conformational behavior of (Ser)-(Pro)4-isolated glycopeptides (Velasquez et al. 2011). Still, a clear influence of such motifs vicinity may be observed in the context of the entire proteins, as observed from both NMR (Razzera et al. 2010) and the present MD simulation data, in which Pro86-Pro89 sequences mostly populate geometries from −90° to −120°, whereas the Pro102-Pro105 peptide ranges from −130° to −160° (Supplementary data, Table S1).

Furthermore, the initial part of the proline-rich domain, comprising residues 57–70, identified as “transition region” (Razzera et al. 2010), has been suggested to present a reduced flexibility when compared with the following 71–108 C-terminal region (Razzera et al. 2010). In fact, from a principal component analysis (PCA) on the polyproline-rich region (Figure 2A), the derived eigenvector 1 root-mean-square deviation (RMSD) plots (see Materials and methods section) indicate that a great increase in flexibility may be observed in the final part of the C-terminal region (Figure 2B), but starting approximately at residue 77. This profile, while mainly observed in cluster 1, which represents 75% of NMR-derived structures (Supplementary data, Table SII), point out the averaged nature of NMR-derived motions (Woods 1998). As a consequence, minor sources of flexibility around the polyproline-rich domain may have remained unseen during NMR experiments, as those in cluster 2 (Figure 2C), focused on the 90–108 C-terminal portion, and cluster 3, mostly around residues 90–100 (Figure 2D). In addition, such data suggest that glycosylation is not capable of modifying the main mobility tendencies of such a region of the protein.

The O-glycosylation at the proline-rich domain had been described to influence the defensin-like part, thus causing differences on the NMR chemical shifts of specific amino acid residues when comparing the natural and recombinant proteins (Razzera et al. 2010). Accordingly, the conformational behaviors of some amino acid residues (His34 and Ala42) identified from NMR experiments, contained within the N-terminal segment, or their interactions with the proline-rich domain were observed to change as due to glycosylation (Supplementary data, Tables SIII–SIV). Namely, although Ala42 was in close contact with unmodified prolines and their vicinal amino acid residues in the non-glycosylated protein (Supplementary data, Tables SIII and SIV), Val97 and the O-linked carbohydrate chains were the main interaction partners of Ala42 during glycosylated Art v 1 simulations. Such modifications involved a reduction in the interaction energy involving Ala42 and the proline-rich domain as due to glycosylation, when comparing Art v 1 non-glycosylated and glycosylated states (Supplementary data, Table SV). On the other hand, and at the same time, the presence of O-linked glycans appeared to increase the interaction of other residue located within the defensin-like domain, His34, with the C-terminal segment as due to glycosylation (Supplementary data, Table SVI). In these systems, His34 side chain also populated a larger number of conformational states, in comparison with its non-glycosylated counterpart (Supplementary data, Figure S4).

Art v 1 glycosylation and allergenicity

The proliferative T cell response to “mugwort” pollen has been recognized to mainly occur due to a single Art v 1 immunodominant epitope, comprising residues 25–36 (Jahn-Schmid et al. 2002). From NMR experiments, two B cell, IgE-recognized epitopes could be also identified, mostly located within the Art v 1 defensin-like domain, each one
allow a proper antibody constant size for an epitope has been observed as necessary to of the others, could not be determined. Considering that a flexible proline-rich domain (Gly74-Gly75-Ser85-Gly93-His108), also presumably recognized by IgE antibodies (Razzera et al. 2010), exhibited perturbations in NMR chemical shifts, although their role in IgE binding, either forming (Razzera et al. 2010), exhibited perturbations in NMR chem-
ical shifts, although their role in IgE binding, either forming (Razzera et al. 2010), exhibited perturbations in NMR chem-
ical shifts, although their role in IgE binding, either forming (Razzera et al. 2010), exhibited perturbations in NMR chem-
ical shifts, although their role in IgE binding, either forming (Razzera et al. 2010), exhibited perturbations in NMR chem-
ical shifts, although their role in IgE binding, either forming (Razzera et al. 2010), exhibited perturbations in NMR chem-
ical shifts, although their role in IgE binding, either forming (Razzera et al. 2010), exhibited perturbations in NMR chem-
ical shifts, although their role in IgE binding, either forming (Razzera et al. 2010), exhibited perturbations in NMR chem-

Table I. Comparison of the dihedral angles on the studied glycosidic linkages, in the context of isolated disaccharides and within glycoproteins

<table>
<thead>
<tr>
<th>Glycosidic linkage</th>
<th>Dihedral angle (°)</th>
<th>Arabinose-containing simulations</th>
<th>Arabinose-absent simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$^{1}T_{2}$</td>
<td>$^{4}T_{1}$</td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\beta$-Araf-(1→O)-HyP</td>
<td></td>
<td></td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>103 ± 26</td>
<td>110 ± 29</td>
<td>99 ± 19</td>
</tr>
<tr>
<td>$\psi$</td>
<td>−167 ± 11</td>
<td>−135 ± 42</td>
<td>−168 ± 12</td>
</tr>
<tr>
<td>$\beta$-Galp-(1→O)-HyP</td>
<td></td>
<td></td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>−101 ± 38</td>
<td>−81 ± 27</td>
<td>−95 ± 48</td>
</tr>
<tr>
<td>$\psi$</td>
<td>106 ± 37</td>
<td>121 ± 44</td>
<td>100 ± 61</td>
</tr>
<tr>
<td>$\alpha$-Araf-(1→3)-Araf$^{f}$</td>
<td></td>
<td></td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>−96 ± 37</td>
<td>−90 ± 34</td>
<td>−75 ± 36</td>
</tr>
<tr>
<td>$\psi$</td>
<td>−130 ± 40</td>
<td>−126 ± 40</td>
<td>−95 ± 28</td>
</tr>
<tr>
<td>$\alpha$-Araf-(1→3)-Galp</td>
<td></td>
<td></td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>−119 ± 44</td>
<td>−144 ± 39</td>
<td>−98 ± 46</td>
</tr>
<tr>
<td>$\psi$</td>
<td>−141 ± 27</td>
<td>−142 ± 27</td>
<td>−139 ± 26</td>
</tr>
<tr>
<td>$\alpha$-Araf-(1→5)-Araf$^{f}$</td>
<td></td>
<td></td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>−114 ± 40</td>
<td>−110 ± 41</td>
<td>−91 ± 41</td>
</tr>
<tr>
<td>$\psi$</td>
<td>169 ± 45</td>
<td>167 ± 49</td>
<td>175 ± 48</td>
</tr>
<tr>
<td>$\alpha$-Araf-(1→6)-Galp</td>
<td></td>
<td></td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>−105 ± 38</td>
<td>−103 ± 39</td>
<td>−92 ± 37</td>
</tr>
<tr>
<td>$\psi$</td>
<td>−181 ± 46</td>
<td>−149 ± 29</td>
<td>−175 ± 48</td>
</tr>
<tr>
<td>$\beta$-Galp-(1→6)-Galp</td>
<td></td>
<td></td>
<td>Disaccharide Glycoprotein$^{a}$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\psi$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Average value obtained from the conformational behavior of each glycosidic linkage in all oligosaccharide structures of all glycoproteins MD simulations.

containing five residues (Ser3-Lys4-Lys55-Ser56-Ala63 and Ser14-Arg40-Glu41-Glu45-Ser46; Razzera et al. 2010). Additionally, a set of five amino acid residues located within the flexible proline-rich domain (Gly74-Gly75-Ser85-Gly93-His108), also presumably recognized by IgE antibodies (Razzera et al. 2010), exhibited perturbations in NMR chemical shifts, although their role in IgE binding, either forming an independent epitope or being recognized together with one of the others, could not be determined. Considering that a constant size for an epitope has been observed as necessary to allow a proper antibody–epitope interaction (Sun et al. 2011), the Ser3-Lys4-Lys55-Ser56-Ala63 and Ser14-Arg40-Glu41-Glu45-Ser46 moiety (Himly et al. 2003). On the other hand, a second group of patients have been also identified, which, although still binding to the glycosylated form, present lower or no reactivity to recombinant Art v 1, which suggests an important role of its O-glycans for IgE binding (Himly et al. 2003). As derived from the performed MD simulations, several intramolecular contact differences between non-glycosylated and glycosylated Art v 1 systems could be observed, possibly interfering with antibody recognition (Figure 3). Accordingly, during the trajectories starting from clusters 1 and 3, the 102–105-linked oligosaccharides stood beside the Ser14-Arg40-Glu41-Glu45-Ser46 epitope (Figure 3A1 and C1). Additionally, elements of 79–80 and 86–89 glycan chains came close to the epitope composed by the residues Ser3-Lys4-Lys55-Ser56-Ala63 (Figure 3B1).

Discussion

Although it has been proposed that half of known proteins are potentially glycosylated (Apweiler et al. 1999; Ben-Dor et al. 2004), the structure and conformation of glycoproteins have been studied in only a very small number of works (Pol-Fachin and Verli 2011). Such glycans may affect several properties of their linked polypeptides, as folding and biological function (Pol-Fachin and Verli 2011). Although no
data are available about the role of Art v 1 glycan chains for plant physiology, the effects of glycosylation over its protein moiety structure, as well as over Art v 1 allergenicity, were previously described by NMR experiments (Razzera et al. 2010) and have been further explored by the present study. Accordingly, the NMR-derived results regarding the protein global structure, as secondary structure content, and Art v 1 mobility, as the transition state flexibility in comparison with the remaining C-terminal part, were employed for validating the obtained data, and the conformational source of some NMR chemical shift alterations, as due to glycosylation, could be proposed. Moreover, to our knowledge, this is the first structural report of an entire furanose-containing glycoprotein.

Regarding Art v 1 carbohydrate moieties, only minor differences could be observed in its glycosidic-linkage geometries and flexibilities as due to modifications in the AraF-puckering form. In this context, the $4T_3$ conformation appeared to allow the generation of minor prevalence conformational states in solution for some disaccharides (Supplementary data, Figure S1I in comparison with Supplementary data, Figure S1E; Supplementary data, Figure S1J in comparison with Supplementary data, Figure S1F) and to present a larger number of minimum-energy conformations in vacuum, especially in $(1 \rightarrow 5)$ and $(1 \rightarrow 6)$ linkages (Supplementary data, Figure S1M in comparison with Supplementary data, Figure S1L; Supplementary data, Figure S1Q in comparison with Supplementary data, Figure S1P), as a previously observed employing MM3 force field (Cros et al. 1994; Pérez et al. 2000). Such behavior indicates a higher conformational flexibility related to AraF-containing oligosaccharides when assuming the $4T_3$-puckering state. With reference to disaccharide-glycoprotein comparisons, no major differences in glycosidic-linkage geometries could be observed between the isolated structures and the entire $O$-linked glycans (Table I), suggesting that the protein moiety is not capable of inducing major conformational changes in its linked oligosaccharides, as proposed previously (Fernandes et al. 2010). Such data reinforce the suggestion that disaccharides solution conformations may
be employed for constructing 3D models of protein-linked oligosaccharides.

In order to better comprehend Art v 1 protein moiety recognition by antibodies, chemical modifications (Perovic et al. 2009; Stanic et al. 2009) and site-directed mutagenesis studies (Gadermaier et al. 2010) have been conducted for Art v 1. Several studies have also been able to identify epitopes within other proteic plant allergen surfaces for IgE binding (Suphioglu et al. 1998; Flicker et al. 2006). Moreover, although glycosylation has also been identified as an important allergenic factor for Art v 1 (Himly et al. 2003), it is recognized that antibodies recognizing glycoproteic antigens may be specific for their carbohydrate units, irrespective of the protein carrier, or may mutually bind both peptidic and glycan moieties (Lisowska 2002). Accordingly, although the “sera” from patients presenting no reactivity to recombinant Art v 1 presumably only recognize its O-linked glycans, those showing reduced reactivity may also bind amino acid residues from either polyproline-rich or defensin-like domains. In this context, assuming that such shared recognition occur comprising amino acid residues composing the previously identified epitopes at its defensin-like domain, the data presented in the present work provide the first structural, atomic-level proposal for such binding. During MD simulations, several forthcomings between defensin-like domain epitopes and glycan chains could be observed, which may suggest such regions as putative sites for O-linked glycans mutual recognition by IgEs: (i) Ser14-Arg40-Glu41-Glu45-Ser46 epitope and 102–105 glycans, which occurred almost systematically in eight independent MD simulations, and (ii) Ser3-Lys4-Lys55-Ser56-Ala63 residues and 79–80- and/or 86–89-linked oligosaccharides, occurring in four trajectories. Such data, together with previous studies on recombinant Art v 1 recognition by IgEs (Razzera et al. 2010) and HLA Class II/Peptide-TCR interactions (Jahn-Schmid et al. 2005, 2008), may be employed in the future for exploring natural, glycosylated, Art v 1 interactions with antibodies and the consequences of this binding to allergy triggering.

Concluding remarks

The comprehension of the mutual influence between bounded and/or unbounded protein and carbohydrate moieties is still a challenging task mainly due to the difficulties associated with the available experimental techniques in dynamically evaluate such interactions (Pol-Fachin and Verli 2011). In recent years, MD simulations have emerged as a promising tool, when combined with appropriate biochemical and structural experimental data, to study glycoproteins, glycoconjugates and protein-carbohydrate complexes in their biological solutions.

Fig. 3. Representative average distance matrices, consisting on the smallest distance between residue pairs, for glycosylated Art v 1 (A) cluster 1, (B) cluster 2 and (C) cluster 3 from MD trajectories. Gray points indicate no difference between non-glycosylated and glycosylated systems, and white and black indicate approximation and withdrawal in the glycosylated systems, respectively. In the structures, O-glycans are shown as sticks; the protein core is presented as cartoon; and the previously identified epitopes are illustrated as dots.
In the present study, we have built 3D models for the native Art v 1 glycoprotein, the major allergen of *A. vulgaris* pollen, and applied MD simulations in order to assess such protein–carbohydrate recognition. Accordingly, in conjunction with previous NMR experiments, an atomic-level representation of O-glycosylation effects over Art v 1 protein moiety could be obtained. Although to date most of the MD studies on glycoproteins reported N-linked oligosaccharides (Pol-Fachin and Verli 2011), the present study comprises one of the earliest studies containing an O-glycosylated protein and, to our knowledge, the first furanose-containing glycoprotein. Moreover, it may be expected that a better comprehension of carbohydrate interference on the immunologic response is potentially capable of contributing in future efforts to develop new therapeutic agents and, ultimately, for allergen-specific immunotherapy.

**Materials and methods**

Nomenclature and software

The nomenclature recommendations and symbols were used as proposed by IUPAC (1996). The relative orientation of a pair of contiguous carbohydrate residues is described, for different types of linkages, by two or three torsional angles at the glycosidic linkage. For a (1 → X) linkage, where “X” is “2”, “3”, “4”, “5” or “6” for the (1 → 2), (1 → 3), (1 → 4), (1 → 5) or (1 → 6) linkages, respectively, the φ and ψ are defined as shown in equations below:

\[ \phi = O5 - C1 - O1 - CX \]  \hspace{1cm} (1)

\[ \psi = C1 - O1 - CX - C(X - 1) \]  \hspace{1cm} (2)

For a (1 → 6) linkage, the ω is defined as shown below:

\[ \omega = O1 - C6' - C5' - C4' \]  \hspace{1cm} (3)

Finally, for a (1 → 5) linkage, the ω is defined as shown below:

\[ \omega = O1 - C5' - C4' - C3' \]  \hspace{1cm} (4)

The manipulation of structures was performed with MOLDEN (Schaftenaar and Noordik 2000), VMD (Humphrey et al. 1996) and PyMol (The PyMOL Molecular Graphics System), the secondary structure content analyses were performed with PROCHECK (Laskowski et al. 1993) and all the MD calculations and the remaining analyses were performed using GROMACS simulation suite, version 4.5.1 (Hess et al. 2008), and GROMOS96 43a1 force field (Scott et al. 1999). Additionally, the (Ser)-(Pro)₄ extensin-like motif chirality was assessed via torsional handedness of the four consecutive Pro/HyP Ca atoms, also known as the ζ angle (Witte and Shakhnovich 1994).

Glycosidic-linkage geometries assessment

The conformational characterization of the glycosidic linkages composing Art v 1 glycan chains was initially performed by varying the φ–ψ and ω–ψ angles, formed by two consecutive monosaccharide residues, from −180° to 150° with a 30° step, in a total of 144 conformers for each linkage, as described previously (Becker et al. 2007). This was performed by using a constant force to restrict only the φ and ψ proper dihedrals during energy minimization with conjugate gradients in each of the above-mentioned values, allowing the search of the conformational space associated with the given linkage. Then, using the minimized output geometries, a series of MD simulations was performed for 20 ps at 10 K, with an integration step of 0.5 fs, to further support the search for minimum-energy conformations (Pol-Fachin and Verli 2008). The relative stabilities of each conformation, obtained from the 10 K MD last frame, were therefore used to construct relaxed energy contour plots. Subsequently, each of the obtained energy maps was smoothed using SigmaPlot 11 software, employing the Loess smoother with a sampling proportion of 0.1 and a polynomial degree of 2. The minimum-energy conformations, obtained from such contour plots, were further refined through aqueous solutions MD simulations (see below).

Principal component analysis

The general idea of the PCA is to reduce the dimensionality of a given data set composed by interrelated variables, retaining as much as possible of the variation presented within it (Jolliffe 2002). For this purpose, the input data is transformed into a new set of variables, known as the principal components, which are uncorrelated, but are ordered so that the first few of them retain most of the variation contained within the original variables (Jolliffe 2002). In the case of the present work, the PCA was employed in order to better assess the high flexibility of the polyproline-rich domain, thus separating its main sources of motion into a new set of flexibility variables, that is, RMSF graphs. Therefore, instead of a single RMSF graph containing the global flexibility related to such domain, the eight eigenvectors obtained from such investigation through GROMACS suite were able to describe the main...
sources of Art v 1 C-terminal part flexibility, being eigenvector 1 the principal flexibility basis around such domain.

Building of the glycosylated Art v 1 atomistic models

The Art v 1 glycosylated structures were constructed, as previously described (Pol-Fachin and Verli 2011), based on the combination of structural data for (1) its protein part, derived from the NMR solution data presented in PDB ID 2KPY (Razzera et al. 2010), comprising recombinant, nonglycosylated Art v 1; and for (2) its carbohydrate moiety, as described in the Glycosidic-linkage geometries assessment section. Regarding the protein moiety, due to the high flexibility related to the proline-rich domain, as observed from NMR experiments (Razzera et al. 2010), and in order to provide a higher conformational sampling for the evaluated systems, the 20 Art v 1 NMR models were clustered (employing g cluster from GROMACS suite, with a 0.95-nm cutoff; Supplementary data, Table SIII) to obtain three groups (Supplementary data, Figure S5A) and allow each condition to be assessed as triplicates: (i) cluster 1 (including 15 models, represented by model 20); (ii) cluster 2 (including four models, represented by model 2); and (iii) cluster 3, represented by model 5 (Supplementary data, Figure S5B–D). Each of these clusters was represented in six conditions, to be assessed as triplicates: (i) cluster 1 (including 15 models, represented by model 20); (ii) cluster 2 (including four models, represented by model 2); and (iii) cluster 3, represented by model 5 (Supplementary data, Figure S5B–D). Each of these clusters was represented in six conditions, to be explored through MD: (i) non-glycosylated, containing Pro residues at the C-terminal part of the protein; (ii) nonglycosylated, containing HyP residues at the proline-rich domain; (iii) O-glycosylated with mono-β-Araf (Figure 1B) in all HyP residues located within (Ser)-(HyP)2-4 motifs, in 1T2 or (iv) in 2T3 conformations; (v) O-glycosylated with a large arabinogalactan oligosaccharide (Figure 1A) at HyP79 and with mono-β-Araf at the remaining (Ser)-(HyP)2-4, also accounting for both 1T2 and (vi) 4T3 conformations. The HyP79 residue was chosen as the arabinogalactan glycosylation site due to the adequate arrangement and surface fitting of the oligosaccharide in relation to Art v 1 protein moiety when linked to such position (Supplementary data, Figure S5E–R). For the glycosylated systems, Art v 1 oligosaccharides were built based on the most prevalent geometries for its composing disaccharides, as obtained from solution MD simulations, being subsequently added to the protein structures.

MD simulations

The 18 Art v 1 starting structures, representing their different glycosylation states, as well as the minimum-energy conformations obtained from the energy maps for the disaccharides, were solvated in triclinic boxes using periodic boundary conditions and an SPC water model (Berendsen et al. 1987). Counter ions (Cl−) were added to neutralize the systems, whenever needed. The employed MD protocol was based on previous studies (de Groot and Grubmüller 2001), as described (Verli and Guimarães 2004; Becker et al. 2005; Pol-Fachin et al. 2009). The Lincs method (Hess et al. 1997) was applied to constrain covalent bond lengths, allowing an integration step of 2 fs after an initial energy minimization using the Steepest Descents algorithm. Electrostatic interactions were calculated with the Particle Mesh Ewald method (Darden et al. 1993), and the dielectric constant was treated as e = 1. Temperature and pressure were kept constant by coupling (glyco)proteins or disaccharides, ions and solvent to external temperature and pressure baths with coupling constants of τ = 0.1 and 0.5 ps (Berendsen et al. 1984), respectively. The reference temperature was adjusted to 310 K after an initial and slow heating of the systems from 50 to 310 K, in steps of 5 ps, each one increasing the reference temperature by 50 K. Although each Art v 1 simulation was extended to 50 ns, the isolated disaccharides trajectories were extended to 0.1 μs.

Supplementary data

Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

Funding

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) from the Ministério de Ciência e Tecnologia, by Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) from the Ministério da Educação e Cultura, Brasília, DF, Brazil.

Conflict of interest

None declared.

Abbreviations

3D, three-dimensional; Araf, arabinofuranose; Galp, galactopyranose; HyP, 4-hydroxyproline; MD, molecular dynamics; PCA, principal component analysis; RMSF, root-mean-square deviation.

References


