Research Article

Molecular Modeling and Docking Study on Glycoproteins of Tomato Spotted Wilt Virus and its Inhibition by Antiviral Agents

Soundararajan P[1], Manivannan A[1], Muneer S[2], Park YG[1], Ko CH[1] and Jeong BR[1,2,3*]

[1]Department of Horticulture, Division of Applied Life Science (BK21Plus Program), Graduate School, Gyeongsang National University, Jinju 660-701, Korea
[2]Institute of Agricultural & Life Science, Gyeongsang National University, Jinju 660-701, Korea
[3]Research Institute of Life Science, Gyeongsang National University, Jinju 660-701, Korea

*Corresponding author: Jeong BR, Department of Horticulture, Division of Applied Life Science (BK21Plus Program), Graduate School, Gyeongsang National University, Jinju 660-701, Korea

Abstract

Tomato Spotted Wilt Virus (TSWV) belongs to the genus *Tospovirus* and has a tripartite, single-stranded negative-sense RNA genome. It is the only genus infects plants in the *Bunyaviridae* family. During viral infection, a complementary strand code for the envelope glycoprotein (GP) precursor is post-translationally cleaved into two spikes or glycoproteins such as GN (amino-terminal) and GC (carboxyl- terminal). Importantly, GN and GC plays key role in particle assembly, maturation, and release in infected cells. Among the two proteins, GN is most important because during maturation process GC retained in the endoplasmic reticulum (ER) unless co-expressed with GN. It enables GN-GC co-migration from ER to the golgi complex. In order to elevate the broad spectrum of viral inhibitors, it is necessary to understand the viral proteins structural details and its interaction with antiviral drugs. The compounds used for protein-ligand docking are tunicamycin, distamycin-A, tiazofurin, actigard, admire, and ribavirin. In order to elucidate their structural interaction, the three-dimensional structure of both GN and GC was predicted. Protein-protein (GC-GN) docking indicates that the C-terminal of GN is necessary for heterodimerization with GC and localization from ER to golgi complex. Consequently, protein-ligand (GN-antiviral compounds) showed that the compounds such as tunicamycin and distamycin-A were considered as the most efficient drugs. On conclusion, in this study the structural details and the docking interaction between GC-GN and GN-antiviral compounds have been explored in TSWV.

Keywords: Glycoprotein; Modeling; Docking; Antiviral compounds; Inhibition

Abbreviations

GP: Glycoprotein; GN: Amino-terminal of Glycoprotein; GC: Carboxy-terminal of Glycoprotein; ER: Endoplasmic Reticulum; TSWV: Tomato Spotted Wilt Virus

Introduction

Up to our knowledge this is the first modeling and docking report on glycoprotein (GP) of tomato spotted wilt virus (TSWV). TSWV belongs to the genus *Tospovirus* which has broad host range and worldwide distribution [1]. Unlike other genus in the *Bunyaviridae* family, *Tospovirus* is unique in its ability to contain the plant infecting counter parts [2]. This enveloped plant virus cause great economic losses in most of the (sub) tropical and temperate crops. Approximately 1080 plant species are affected by TSWV and most of the *Tospovirus* is transmitted by an insect belongs to the order *Thysanoptera* vector [3-4]. The genome of *Tospovirus* consists of tripartite RNA, composed of large (L) (<9 kb), medium (M) (~4.8 kb) and small (S) segment (~3 kb). The L segment encodes RNA-dependent RNA polymerase (RdRp), the M segment codes for GP and the non-structural protein (NSm) followed by the S segment that encodes another non-structural protein (NSs) and the nucleoprotein (N) [5]. Among the above mentioned proteins, the envelope GP plays crucial role in particle assembly, maturation, and release in the host

organism [6]. During viral infection, complementary strand codes for the GP precursor is post-translationally cleaved into two spikes or glycoproteins such as GN and GC (Where N- and C- refers to the amino- and carboxyl-terminal) [7]. Initially, envelope protein such as GN and GC are separately accumulated in endoplasmic reticulum (ER) [8]. Viruses of this family obtain their lipid membrane and maturation at the golgi complex of the host cell. Similarly, animal infecting bunyavirus also have the same mechanism during infection [9]. Firstly, GC protein is arrested in the ER, and able to migrate to the golgi complex only by forming dimer with the GN protein. Transient expression studies of TSWV in plant cells have shown that GN is able to migrate itself and also co-translate with GC from the ER to the golgi complex. Unlike GN, GC is unable to leave the ER on its own. Therefore, GN-GC interaction is rather unique feature among plant viruses and required for the replication of the virus in the vector and also for virus particle assembly, development, maturation, and organization in the host [6,10,11].

In order to elevate the broad spectrum of viral inhibitors, it is necessary to understand the structural details of viral proteins and its interaction with antiviral drugs. So, the practically used antiviral agents against TSWV such as tunicamycin, distamycin-A, tiazofurin, actigard, admire, and ribavirin have been virtually screened in the current study to identify the elite compounds to inhibit TSWV.

Tunicamycin, an N-glycosylation inhibitor was reported to affect the GC/GN exit from ER [12]. Distamycin-A was reported to delay the spreading of TSWV in tobacco leaves [13]. Carner et al. [14] mentioned that tiazofurin showed efficient activity against TSWV in tomato. Ribavirin significantly reduced the overall growth rate of the virus by increasing the duration of lag phase in the TSWV affected cells in both tomato and tobacco [15]. Actigard and admire application in early infection showed some effective reduction of TSWV in flue-cure tobacco [16].

However, till date none of the reports have documented the three-dimensional (3D) model, and the protein-protein interaction and protein-antiviral compound(s) interaction of TSWV GN and GC. In the biological database amino acid sequence of many important proteins are available but their 3D structure(s) are limited. Structural details of protein are necessary to study the exact function and its nature. Application of computational technology on the prediction of protein structure and their interaction/recognition process has been successfully applied in biological research. This technique could provide the proper guide to carryout vast array of experiments in less time and reduced cost [17-19]. Hence, to understand the GN-GC interaction and inhibitory mechanism of antiviral compounds, molecular modeling, and docking strategy have been carried out. Further, the current work will provide theoretical guidance on the research and development of antiviral agents against TSWV.

Materials and Methods

Molecular modeling

The sequence of GN and GC was retrieved separately from the whole length of precursor envelope glycoprotein (ID: O55647) from UniProt database (http://www.uniprot.org/). I-TASSER, ranked as No. 1 server for protein structure prediction, was used to predict the 3D model [20]. For ab-initio modeling, I-TASSER generates the full length model of proteins by excising continuous fragments from threading alignments and then reassembles those using replica-exchanged Monte Carlo simulations. Low temperature replicas (decoys) generated during the simulation are clustered by SPICKER and top five cluster centroids are selected for generating full atomic models. Cluster density is defined as the number of structure decoys at a unit of space in the SPICKER cluster. A higher cluster density means the structure occurs more often in the simulation trajectory and therefore indicates a better quality model. The confidence (C)-score is used to estimate the quality of the models based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. To find the accuracy of the predicted model, transmembrane (TM) score and root mean square deviation (RMSD) are also measured based on the C-score to know the structural similarity between two structures.

Validation, transmembrane and active site prediction

The best 3D model selected was validated by Structural Analysis and Verification Server (SAVeS) (http://services.mbi.ucla.edu/SAVES/). In SAVes, the stereochemical quality of a protein structure was checked residue-by-residue geometry and overall structure geometry by PROCHECK [21] and WHAT_CHECK. Non-bonded interactions between different atom types were evaluated by ERRAT. VERIFY_3D and PROVE was used to determine the compatibility of the atomic model (3D) with its own amino acid sequence (1D) by

assigning a structural class based on its location and environment (alpha, beta, loop, polar etc) and comparing the results to good structures. Volumes of atoms in macromolecules using an algorithm treats the atoms like hard spheres and calculates a statistical Z-score deviation for the model from highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better) PDB-deposited structures was calculated by PROVE. Transmembrane region of both GN and GC have been predicted by TMHMM server (v. 2.0) [22]. Active site was predicted by the Q-SiteFinder, an online web server [23].

Ligand preparation and validation

Chemical formula of all the anti-viral compounds such as distamycin-A (3003), tiazofurin (403014), actigard (77928), admire (77934), and ribavirin (34439) were retrieved from ChemSpider database (http://www.chemspider.com) except tunicamycin. The chemical formula of tunicamycin (6433557) has been retrieved from PubChem (http://www.ncbi.nlm.nih.gov/pccompound). Compound identification number from their respective databases has been mentioned in the brackets. All the structures were sketched and optimized using universal force field till the optimized geometry is achieved in ArgusLab (v. 4.0.1) [24]. After optimization, all compounds were converted to PDB format for further docking studies.

Protein–protein docking simulation

To find the essential region of GN for the dimerization between GN and GC during rescue of GC from ER, protein-protein docking stimulation was carried out by using GRAMM-X (v.1.2.0), a protein-protein docking web server from Vakser Lab [25]. Whole protein structure of both GN and GC was chose for protein-protein interaction. Maximum number of 100 output models was given using default parameters. Generally, in GRAMMX intermolecular energy function has been used to smoothen by changing the range of the atom-atom potentials. This technique locates the area of the global minimum of intermolecular energy for structures of different accuracy. GRAMM-X was able to detect the near native matches in complexes with large conformational changes

Protein-ligand docking simulation

Autodock (v. 4.2) was used to analyze the interaction between GN and anti-viral compounds [26]. Firstly, Kollman united atom charges, solvation parameters, and polar hydrogens were added to the protein. Protonated species were merged with the non-polar hydrogens after Gasteiger charge was assigned. The rigid roots of selected flexible molecules were defined and ligands were allowed to rotate freely by applying Auto-Tors. To dock the compounds against macromolecules grid maps need to be set on the receptor to define the region for interaction. The grid box x, y, and z coordinates were varied according to the size of the ligand. Lamarckian genetic algorithm (LGA) is the hybrid of the Genetic Algorithm (GA) method with the adaptive Local Search (LS) method to enhance the relative performance of GA. The spacing between grid points was set to 0.375 Å. Maximum of 10 conformers was used for each compound. The population size was set to 100 and individuals were initialized with the following parameters: maximum number of energy evaluation 55×10^5, maximum number of generations 1000, maximum number of top individual that automatically survived 1, mutation rate 0.02, crossover rate 0.8, step sizes were 0.2 Å for translations, 5.0^0 for
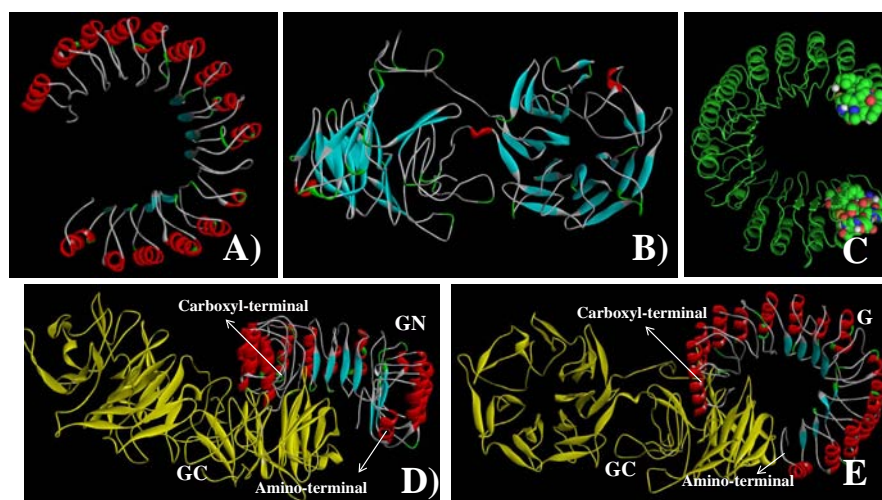
Figure 1: Modeling and interaction of glycoprotein of tomato spotted wilt virus. Three-dimensional structure of tomato spotted wilt virus envelope glycoprotein such as GN (A) and GC (B), where N and C refer to the amino- and carboxyl-terminal. C) Active site regions of GN involved in anchoring of GC are represented in spheres. D) Involvement of carboxyl-terminal of GN for dimerization process with GC. E) Both carboxyl- and amino-terminal forms dimer for stabilization and trafficking to golgi complex. GC was colored in cyan and secondary based coloring pattern was used for GN.

quaternions and 5.0^0 for torsions, cluster tolerance 0.5 Å, external grid energy 1000, maximum initial energy 0.0 and maximum number of retries 10000. Totally, 50 LGA runs were performed to each anti-viral compound and the best configuration was selected based on the least global energy values, H-bonds and No. of electro statistically interacted amino acids.

Visualization

The Discovery Studio Visualizer (v. 4.0) has been used to visualize the modeled proteins and docked molecules. The active sites of GN predicted from the Q-site finder was viewed in PyMol (v. 1.7) software.

Result and Discussion

Molecular modeling of glycoprotein

To date, neither crystallographic/NMR structure nor theoretically predicted 3D model of envelope TSWV glycoprotein is available. However, several experimental studies on TSWV have suggested that the envelope GP is the major contributor for the pathogenicity of TSWV and it is associated with the attachment/receptor binding on the host cell surface. Therefore, the present report has attempted to utilize the existing experimental results with the bioinformatics approaches to understand the interaction between both GN-GC and GN-antiviral agents, respectively. In addition, the structural information of the TSWV proteins would present a significant step in forward to identifying efficient anti-viral agent(s) for TSWV. Due to the less homology of TSWV proteins even within the genus, the theoretical structural prediction of GP protein is very difficult. In computational biology, prediction of 3D structure of protein from the amino acid sequence is relatively easy when a candidate protein exhibits significant sequence similarity (30%) to the already known protein structure [17,18]. The total length of the glycoprotein is 1135

amino acids which is divided into three parts i.e., 1-35 amino acids act as a signal peptide, 36-484 amino acids codes the GN and 485-1135 amino acids codes the GC. The modeling was done only on the region encodes GN and GC proteins.

In the current work, five models were predicted by the I-TASSER server and based on the C-score, TM-score, and RMSD, one best model for GN and GC was selected (Figure 1A and B) for further analysis. In a benchmark set of 500 non-homologous proteins C-score of TSWV GN protein is highly correlated with TM-score and RMSD. Selected GN and GC protein's I-TASSER parameters value such as C-Score, No. of decoys, cluster density, TM-score, and RMSD are shown in table 1. The CHARM force field has been applied to minimize the energy of GN and GC i.e., -1096.92 and -5506.352 KJ·mol^{-1}, respectively. Modeled protein φ-ψ torsion angles were evaluated by PROCHECK. Ramachandran plot of the energy minimized 3D model of GN protein displayed 93.0% in favorable region (75.2% in core and 17.8% in additionally allowed region), 4.7% in generously allowed and only 2.2% in disallowed region. Percentage of residues in favoured regions (74.7% in core and 16.9% in additional allowed region) was 91.6%, a generously allowed region was 5.9% and disallowed region was 2.6%. Theoretically predicted models almost satisfied all the parameters of WHAT_CHECK, ERRAT, VERIFY_3D, AND PROVE (Data not shown).

In detail, the structure of GN protein consisted of 15 helices along with parallel sheets. Notably, 3 and 4 parallel sheets were present in between the 1 to 4 and 6 to 10 helix region, respectively (Figure 1A). In contrast to the GN, GC possessed few misfolding in the modeled structure. Interestingly, two monomers arranged in a barrel like fashion with four anti-parallel beta sheets were observed in GC model (Figure 1B). First and second monomer was found between the region

Table 1: I-TASSER parameter values of selected model of GN and GC protein of TSWV.

Protein	C-Score	No. of decoys	Cluster density	Tm-Score	RMSD
GN	-2.73	943	0.0979	0.40±0.13	13.7±4.0
GC	-0.93	600	0.3478	0.60±0.14	10.1±4.6

of 1 to 298 amino acids and 299 to 612 amino acids, respectively. Considerably, region between 613-651 amino acids were not folded properly. According to previous report, GC is unable to exist on its alone from the ER due to the improper folding [11]. Concordantly, previous experiments reported that unlike GN, absence of signaling peptides in N-terminal region of GC resulted in ER arrest [2]. In agreement with our result, Ribeiro et al. [5] reported that the primary reason for the failure of ER arrest and golgi retention of GC could be due its improper folding The improper alignment in between the 613-651 amino acids could lead GC to retain in ER (Figure 1B). Therefore, transport competence of membrane proteins like GC need to leave the ER was referred to be regulated by the proper folding in its structure.

Protein-protein docking

Since experimental evidence indicates that the GC protein was unable to leave the ER and it requires interacting with GN for transportation to the golgi complex from ER. Therefore, to elucidate the structural interaction between the GN-GC, protein-protein docking simulation has been investigated. The dimerization process illustrated in Figure 1C and D showed that for the interaction with GC, C-terminal of the GN plays a major role than the other regions [2,11,15]. The docking results of GRAMMX demonstrated that C-terminal of GN significantly interact maximum number of times with GC followed by N-terminal region than the other regions of the GN. The protein-protein docking affirmed the importance of C-terminal region of GN in the initialization and formation of dimerization (Figure 1D). On the other hand, the monomer present in the N-terminal region of GC (Figure 1B) was involved maximum number of times in the complex formation with the GN. The additional involvement of N-terminal along with C-terminal could aids for further stabilization and trafficking to the golgi complex (Figure 1E). Studies performed in mammalian cells have shown that the last 20 amino acids of the C-terminal of GN were essential for their co-translocation to the Golgi complex [27]. Further, the removal of first 25-30 amino acids of the cytoplasmic, along with the transmembrane domain (TMD) locating in C-terminal of GN does not lead the protein to translocate into plasma membrane [11,28]. Moreover, active site predicted in GN by Q-SiteFinder is located in this region (Figure 1C). Correspondingly, the amino acids in C-terminal of GN are essential for the secretory pathway, and also lead to malfunction/extremely slow in ER retention [29]. The GN region involved in the interaction consisted of a short helix between 425 to 438 amino acids and loop between the 439-449 amino acids. Notably, small consensus region observed in between 425-429 helix can act as a potential signal peptide cleavage site [4]. Recent discoveries states that signal peptides are very useful for the translocation of secretory and membrane-bound proteins across the membrane of the ER [12]. Moreover, among ten active sites predicted by Q-SiteFinder, two sites are exactly present in the N- and C-terminal of the GN protein, where the protein-protein docking between GC-GN interactions occurred (Figure 1D and E) frequently. For understanding the virus budding and protein compartmentalization, the signaling region is also a vital one. Hence, protein-ligand docking studies were carried on the N-terminal region of GN (Figure 2A) due to its interaction with GC and the presence of TMD, cytoplasmic tail, signaling peptides, and N-glycosylation site.
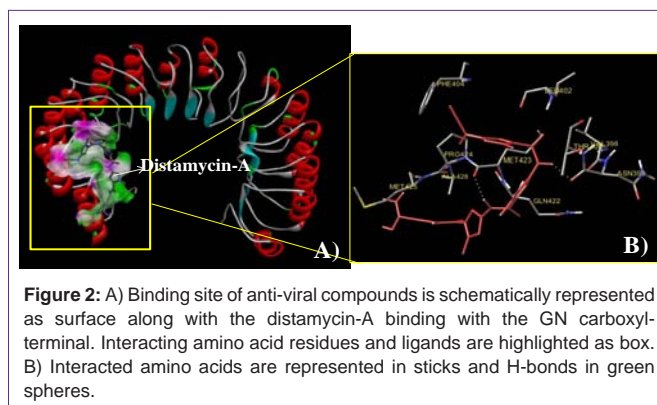
Figure 2: A) Binding site of anti-viral compounds is schematically represented as surface along with the distamycin-A binding with the GN carboxyl-terminal. Interacting amino acid residues and ligands are highlighted as box. B) Interacted amino acids are represented in sticks and H-bonds in green spheres.

Protein-ligand docking

Progressive deletions in the C-terminal tail of GN proved that the golgi retention signal maps to a region encompassing approximately 10 to 50 residues in TMD and the cytoplasmic tail. Expression of the construct surprisingly showed the TMD of GN was sufficient to enable ER-exit and golgi targeting of GC [11]. The antiviral compounds docked in C-terminal of GN have been represented as surface and the docked pose was demonstrated by highlights the distamycin-A in the rectangle [Figure 2A]. Binding affinity between the protein and the ligand is ranked mainly based on the binding energy, and number of hydrogen bond formed between the receptor and the donor (Table 1). In addition, amino acids involved in electrostatic interaction were also analyzed (Table 2). Upon the protein-ligand docking study, the compounds tunicamycin and distamycin-A were considered to be the most efficient. Though, distamycin-A possess lesser binding energy than tunicamycin, latter compound formed three H-bonds similar to the former. Remarkably, distamycin-A binds with the amino acids PRO424 and MET425 just before the consensus signaling peptides such as Ser, Met, Ala, and Gln (SMAQ) (426-429) (Table 1, Figure 2B) [4]. Involvement of H-bond with the Pro424 and Met425 amino acids possibly blocks the signaling for its own golgi retention and also rescuing mechanism of GC. Most importantly distamycin-A could stop the signaling process of GN [29] because of its binding affinity on amino acids just before the SMAQ (signaling region) [4]. Accordingly, N-Glycosylation of GN can be altered by distamycin-A due to its efficient and stable binding with the GN. Consequently, N-glycosylation site of the envelope membrane proteins of Bunyaviridae family viruses are modified by the covalent addition of asparagines-linked oligosaccharides could be affected [30]. This mechanism is vital for protein folding and function of envelope proteins such as GN and GC [31]. Additionally, N-glycosylation plays important role for the receptor binding, membrane fusion, and penetration into cells, virulence, directing virus morphogenesis at the budding site [32]. The experimental evidence from the Kikkert et al. (2001) revealed that when the tunicamycin was added during expression of the precursor, it blocked the exit of GC and GN from the ER by inhibiting the N-glycosylation [12]. During docking, distamycin-A shared the similarity with tunicamycin by forming the H-bond with the Thr400 and electrostatically interacted with ASN399, an N-glycosylation site in C-terminal TMD of GN (Table 1). Unlike tunicamycin, role of distamycin-A on controlling the TSWV are not explored or documented well. In this study, probable inhibition pattern or mechanism by distamycin-A has been virtually

18. Soundararajan P, Sakkiah S, Sivanesan I, Lee KW, Jeong BR. Macromolecular docking simulation to identify binding site of FGB1 for antifungal compounds. Bull Korean Chem Soc. 2011; 32: 3675-3681.

19. Sekhar PN, Amrutha RN, Sangam S, Verma DPS, Kishor PBK. Biochemical characterization, homology modeling and docking studies of ornithine δ-aminotransferase-an important enzyme in proline biosynthesis of plants. J Mol Graph Model. 2007; 26: 709-719.

20. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008; 9: 40, 1-8.

21. Laskowski RA, MacArthur MW, Moss DS, Thronton JM. PROCHECK: a program to check the stereo-chemical quality of protein structures. J Appl Cryst. 1993; 26: 283-291.

22. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001; 305: 567-80.

23. Laurie AT, Jackson RM. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. Bioinformatics Bioinformatics. 2005; 21: 1908-1916.

24. Oda A, Takahashi O. Validation of Arguslab efficiencies for binding free energy calculations. Chem-Bio Informatics J. 2009; 9: 52-61.

25. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. Nucleic Acids Res. 2006; 34; 310-314.

26. Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit. 1996; 9: 1-5.

27. Shi X, Brauburger K, Elliott RM. Role of N-linked glycans on bunyamwera virus glycoproteins in intracellular trafficking, protein folding, and virus infectivity. J Virol. 2005; 79: 13725-13734.

28. Snippe M, Smeenk L, Goldbach R, Kormelink R. The cytoplasmic domain of tomato spotted wilt virus Gn glycoprotein is required for golgi localisation and interaction with Gc. Virol. 2007; 363: 272-279.

29. Persson R, Pettersson RE. Formation and intraceuular transport of a heterodimeric viral spike protein complex. J Cell Biol. 1991; 112: 257-266.

30. Doms RW, Lamb RA, Rose JK, Helenius A. Folding and assembly of viral membrane proteins. Virology. 1993; 193: 545-562.

31. Braakman I, van Anken E. Folding of viral envelope glycoproteins in the endoplasmic reticulum. Traffic. 2000; 1: 533-539.

32. Parodi AJ. Protein glucosylation and its role in protein folding. Annu. Rev Biochem. 2000; 69: 69-93.