

## The Endoplasmic Reticulum Glucosyltransferase Recognizes Nearly Native Glycoprotein Folding Intermediates\*

Received for publication, July 26, 2004, and in revised form, August 19, 2004  
Published, JBC Papers in Press, August 19, 2004, DOI 10.1074/jbc.M408404200

Julio J. Caramelo, Olga A. Castro, Gonzalo de Prat-Gay, and Armando J. Parodi‡

From the Fundación Instituto Leloir, Buenos Aires, Argentina C1405BWE

**The UDP-Glc:glycoprotein glucosyltransferase (GT), a key player in the endoplasmic reticulum (ER) quality control of glycoprotein folding, only glucosylates glycoproteins displaying non-native conformations. To determine whether GT recognizes folding intermediates or irreparably misfolded species with nearly native structures, we generated and tested as GT substrates neoglycoprotein fragments derived from chymotrypsin inhibitor 2 (GCI2) bearing from 53 to 64 (full-length) amino acids. Fragment conformations mimicked the last stage-folding structures adopted by a glycoprotein entering the ER lumen. GT catalytic efficiency ( $V_{\max}/K_m$ ) remained constant from GCI2-(1-53) to GCI2-(1-58) and then steadily declined to reach a minimal value with GCI2-(1-64). The same parameter showed a direct hyperbolic relationship with solvent accessibility of the single Trp residue but only in fragments exposing hydrophobic amino acid patches. Mutations introduced (GCI2-(1-63)V63S and GCI2-(1-64)V63S) produced slight structural destabilizations but increased GT catalytic efficiency. This parameter presented an inverse exponential relationship with the free energy of unfolding of canonical and mutant fragments. Moreover, the catalytic efficiency showed a linear relationship with the fraction of unfolded species in water. It was concluded that the GT-derived quality control may be operative with nearly native conformers and that no alternative ER-retaining mechanisms are required when glycoproteins approach their proper folding.**

A quality control mechanism in the endoplasmic reticulum (ER)<sup>1</sup> is in charge of sensing the folding state of glycoproteins before their transport to the Golgi apparatus (1). Proteins incorrectly folded are initially retained in the ER and eventually degraded by the proteasome, a process known as ER-associated degradation (2, 3). Recent findings on glycoprotein processing in

the ER showed that sugar moieties could be exploited to encode information on glycoprotein folding status. About 65% of Asn-X-Ser/Thr consensus sequences in proteins entering the ER lumen are N-glycosylated (4). The Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> glycan transferred is deglycosylated immediately by the sequential action of glucosidases I and II. At this stage the proteins that are not correctly folded are reglycosylated by UDP-Glc:glycoprotein glucosyltransferase (GT), generating Glc<sub>1</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>. This enzyme has a unique property as it glucosylates glycoproteins displaying native-close, molten globule-like but not random coil or compact-native conformations. Monoglucosylated glycans are specifically recognized by two ER resident lectins, calnexin (CNX) and calreticulin (CRT), in charge of retaining folding intermediates and irreparably misfolded glycoproteins in the ER. In addition, lectin binding facilitates conformational maturation of glycoproteins by preventing aggregation and allowing the labor of lectin-associated proteins such as ERp57 (endoplasmic reticulum protein 57), a protein of the protein disulfide isomerase family (5).

The single glucose added by GT is removed *in vivo* by glucosidases I and II, whose activity is independent of the folding status of glycoproteins, causing the release of the latter from the lectin anchors. The ensuing shuttle between Glc-containing, lectin-bound, and Glc- and lectin-free species, catalyzed by the opposing activities of GT and glucosidases I and II ends on acquisition of native compact conformations. The selectivity of this cycle for misfolded proteins resides in GT, since it is the only element of the mechanism that senses glycoprotein conformational states. This quality control mechanism is not the only one operating in the ER as there are alternative mechanisms for retaining proteins not displaying their fully folded structures, as those involving other conventional chaperones, such as BiP/Grp78 or reversible disulfide bonding with ER resident proteins (1).

Because GT reglycosylates a myriad of misfolded glycoproteins, the molecular determinants for glycoprotein recognition should be, in principle, independent of the primary structure and should rest upon elements common to irreparably misfolded or partially folded structures. Current evidence points to exposure of hydrophobic residues as the common determinant (6, 7). In a recent work we presented evidence showing that GT recognized local stretches of hydrophobic residues in small disordered fragments rather poorly but actively glucosylated glycoproteins showing long range interactions in molten globule-like conformations, thus, exposing anilino-naphthalene sulfonic acid (ANS) binding hydrophobic patches (8).

It remained to be established, however, if the enzyme is sensitive enough to glucosylate glycoproteins with minimal structural differences with respect to native conformations. Inability to glucosylate nearly native conformations would mean that alternative retaining systems must replace CNX and CRT at the end of the folding process or in irreparably misfolded species, showing conformations slightly differing from native ones. To

\* This work was supported by National Institutes of Health Grant RO1 GM44500/13, the Howard Hughes Medical Institute, and the Agencia Nacional de Promoción Científica y Técnica (Argentina). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Fundación Instituto Leloir, Avda. Patricias Argentinas 435, Buenos Aires, Argentina C1405BWE. Tel.: 54-11-4863-4012; Fax: 54-11-4865-2246; E-mail: aparodi@leloir.org.ar.

<sup>1</sup> The abbreviations used are: ER, endoplasmic reticulum; ANS, anilino-naphthalene sulfonic acid; BiP/Grp78, binding protein/glucose-regulated protein 78, an endoplasmic reticulum chaperone of the Hsp70 family; CI2, chymotrypsin inhibitor 2; CNX, calnexin; CRT, calreticulin; GCI2, glucosylated chymotrypsin inhibitor 2; Grp94, glucose-regulated protein 94, an endoplasmic reticulum chaperone of the Hsp90 family; GT, UDP-Glc:glycoprotein glucosyltransferase; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

address this issue we have now correlated GT catalytic efficiency with structures of neoglycoproteins of increasing sizes approaching that of the full-length species, thus mimicking the last folding conformations adopted by a glycoprotein entering the ER lumen. We, thus, found that GT is an exquisite sensor of the thermodynamic stability of nearly native structures that are to be retained for quality control iteration.

#### EXPERIMENTAL PROCEDURES

**Materials**—All chemicals were purchased from Sigma with the exception of sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate, which was purchased from Pierce, and NAP-5 desalting columns, from Amersham Biosciences. Rat liver GT was purified as described (9). UDP-[<sup>14</sup>C]Glc (300 mCi/mmol) was synthesized as previously described (9).

**Expression, Purification, and Glycosylation of Chymotrypsin Inhibitor 2 (CI2)**—This study is based on the truncated form of CI2, where the first residue corresponds to Met-20 (10). CI2-(1–25), CI2-(1–40), and CI2-(1–64) were obtained as described previously (8). The remaining fragments were produced by introducing a stop codon at the appropriate positions by the inverse PCR method using the pTZ18U-based vector (11). We also introduced the substitution E7C to chemically glycosylate the proteins. Transformed *Escherichia coli* TG2 cultures at OD 1 were induced overnight with 1 mM isopropylthiogalactoside. Cells were harvested, resuspended in 1/20 volume of 50 mM Tris-HCl buffer, pH 8.0, 2 mM EDTA, 1 mM phenylmethanesulfonyl fluoride, 0.1% Triton X-100, and 50 µg/ml lysozyme. After incubation at room temperature for 30 min, cells were placed in an ice bath and sonicated twice for 1 min with a Virsonic 550 sonicator; CI2s were found mainly in inclusion bodies after centrifugation at 16,000 rpm for 30 min. The inclusion bodies were washed three times with 20 ml of 50 mM Tris-HCl buffer, pH 8.0, 10 mM EDTA, 0.5% Triton X-100, and 100 mM NaCl and then 50 mM Tris-HCl buffer, pH 8.0, 10 mM EDTA, 0.5% Triton X-100, 0.5 M guanidinium-HCl, and 5 mM β-mercaptoethanol. Finally, inclusion bodies were solubilized with 15 ml of 50 mM Tris-HCl buffer, pH 8.0, 5 mM EDTA, and 6 M urea. After centrifugation for 30 min at 16,000 rpm the solutions were applied to a Superdex-75 HR16/50 column and eluted with 50 mM Tris-HCl buffer, pH 8.0, 4 M urea, and 150 mM NaCl. Fractions containing CI2s were pooled. Trifluoroacetic acid and acetonitrile were added to a final concentration of 0.05 and 5%, respectively. Proteins were purified by reverse phase high performance liquid chromatography using a Vydac C8 column with a linear gradient of acetonitrile/water in 0.1% trifluoroacetic acid and dried in a SpeedVac. Because GT substrates are high mannose-type glycoproteins, thyroglobulin/Pronase-derived high mannose-type glycopeptides were covalently linked to the new-engineered C7 compounds by using sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate as described previously (8). Glycosylated CI2s will be referred to as GCI2s.

**Spectroscopic Characterization of GCI2s**—CD spectra were measured with a Jasco J810 spectropolarimeter. We used 10 µM GCI2s in far UV and 40 µM in near UV CD. Intrinsic fluorescence and ANS binding were measured with an Aminco Bowman Series 2 spectrofluorimeter with 5 µM GCI2s. We used 50 µM ANS with the excitation wavelength set at 360 nm and spectra recorded from 400 to 650 nm. All spectra were recorded at 25 °C in 20 mM HEPES buffer, pH 7.4.

**Protein Stability**—Chemical denaturation experiments were performed as described (10). Fluorescence spectra were detected with the excitation wavelength set at 280 nm, and spectra were recorded from 300 to 430 nm. Excitation and emission band passes were 8 and 2 nm, respectively. The fluorescence spectra center of mass values were used to describe the conformational states.

**Activity Assays**—GT assays were carried out at 37 °C in 30 µl of 40 mM HEPES buffer, pH 7.4, 20 mM KCl, 5 mM CaCl<sub>2</sub>, and 15 µM UDP-Glc (100,000 cpm of UDP-[<sup>14</sup>C]Glc, 300 mCi/mmol). Reactions were started by the addition of 2 µg of rat liver GT and stopped after 10 min by the addition 30 µl of cracking buffer at 100 °C containing 5% SDS. The samples were immediately heated for 10 min at 90 °C. An aliquot of 15 µl was run in Tricine-SDS-PAGE electrophoresis (12). The dried gel was exposed to an LE phosphor screen (Amersham Biosciences) for 12 h and revealed in a Storm 820 PhosphorImager. A linear relationship between density and radioactivity was obtained under conditions described.

#### RESULTS

**Description of the Model Neoglycoprotein System Employed**—CI2 is an extremely useful model to study chaperone and GT recognition of nascent chains (8). However, because GT sub-

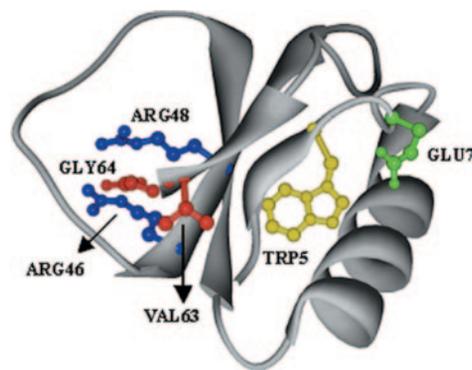


FIG. 1. Structure of CI2. Adapted from McPhalen and James (13).

strates are glycoproteins, the original E7 in CI2 was replaced by a cysteine (E7C) to which a high mannose-type glycopeptide derived from bovine thyroglobulin was attached through a bifunctional chemical cross-linker (glycosylated CI2s are to be referred to as GCI2s). The original glutamic acid residue is in a type III turn resembling many natural *N*-glycosylation sites (4). It is exposed to the solvent and has no effect on full-length CI2 conformation (13). We had previously determined that a significant GT activity was only attained with the fragment of 53 residues in length, the first ones to display long range native interactions (8). We now focused on GCI2 fragments having from 53 to 64 (full-length) amino acid residues. Their conformations mimic the last folding structures adopted by a glycoprotein entering the ER lumen. The acquisition of the native structure in CI2 is highly cooperative with the addition of the last two residues (Val-63 and Gly-64) (14). Although partially exposed to the solvent, Val-63 interacts with Val-47 and Trp-5 (the only tryptophan residue in CI2) in the hydrophobic core, whereas Gly-64 does so with Arg-46 and Arg-48 through ionic interactions (Fig. 1) (13). A point mutation (V63S) was also introduced because previous experiments had shown that certain mutations at that position have a destabilizing effect on CI2-(1–64) native structure (15).

**Structural Characterization of Last Folding Intermediates**—The center of mass of fluorescence spectra of GCI2s revealed almost no variation of tryptophan exposure to the solvent as the chain grew from GCI2-(1–53) to GCI2-(1–62), with an abrupt change upon the addition of the last two residues where the tryptophan side chain appears completely buried (Fig. 2A). Both GCI2-(1–63)V63S and GCI2-(1–64)V63S fluorescence spectra showed a shift toward longer wavelengths with respect to the canonical derivatives having Val-63. Far UV CD spectra presented only a marginal variation for derivatives ranging from GCI2-(1–53) to GCI2-(1–62), thus indicating almost no effect of the chain length on the secondary structure within that size range (Fig. 2B; only the GCI2-(1–53) is shown). On the contrary, a somewhat different spectrum was observed in GCI2-(1–64), whereas GCI2-(1–63), GCI2-(1–63)V63S, and GCI2-(1–64)V63S presented intermediate patterns between those of GCI2-(1–53) to GCI2-(1–62) and GCI2-(1–64).

Although derivatives ranging from GCI2-(1–53) to GCI2-(1–63) displayed very similar near UV CD spectra (Fig. 2C; only those of GCI2-(1–53), GCI2-(1–63), and GCI2-(1–63)V63S are shown), those of GCI2-(1–64) and GCI2-(1–64)V63S were sharply different as they displayed a positive band at about 290 nm. This implied that the full-length neoglycoproteins had the Trp-5 residue located in an asymmetric environment indicative of defined three dimensional structures, in contrast with shorter fragments. In addition, only non-full-length species showed ANS binding, although GCI2-(1–64)V63S gave a slightly positive signal when compared with GCI2-(1–64) or

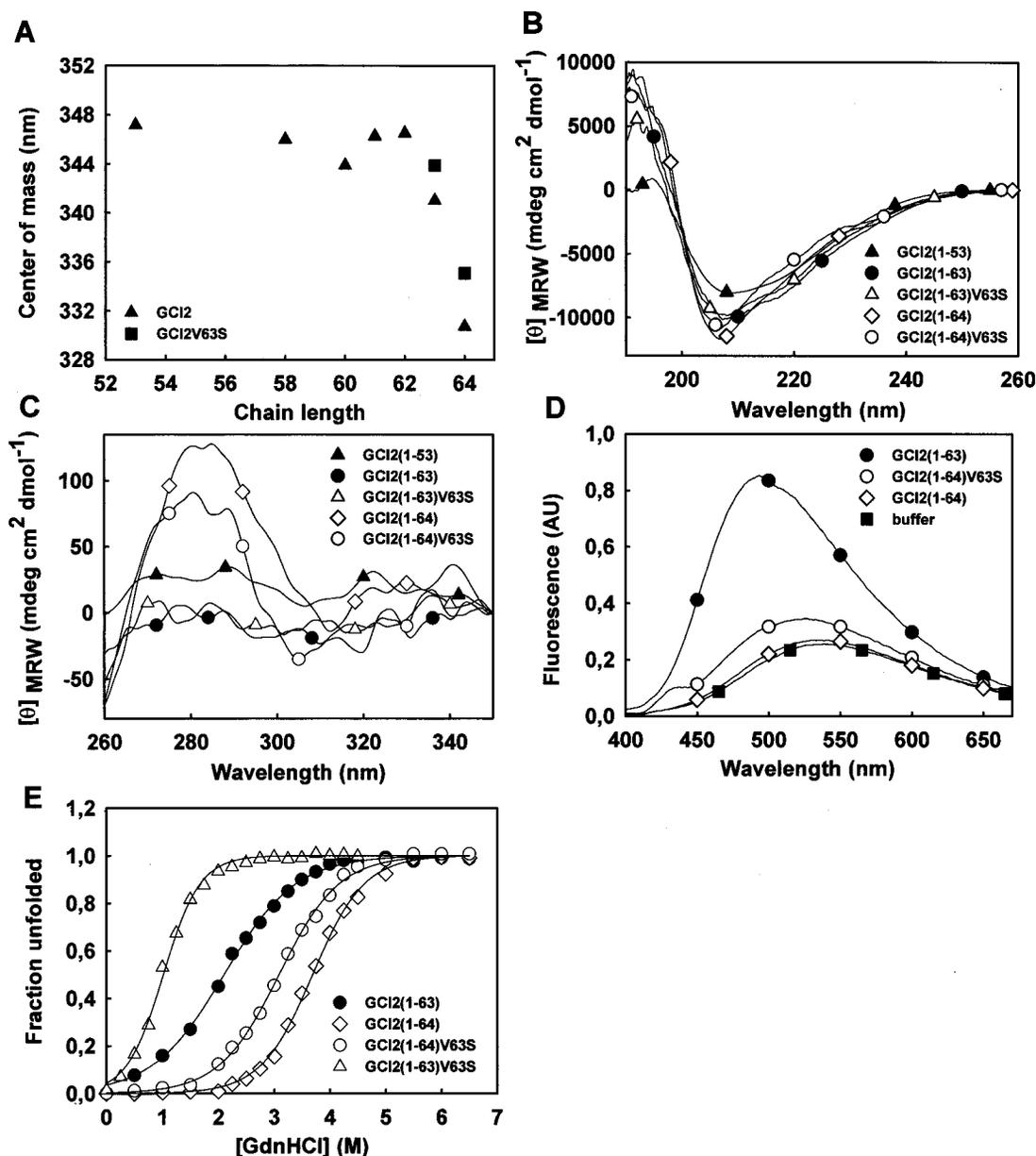


FIG. 2. Structural characterization of GCI2-(1-53), GCI2-(1-58), GCI2-(1-60), GCI2-(1-61), GCI2-(1-62), GCI2-(1-63), GCI2-(1-64), GCI2-(1-63)V63S, and GCI2-(1-64)V63S. *A*, fluorescence center of mass versus chain length. *B*, far UV CD of selected GCI2s. Non-indicated curves are similar to GCI2-(1-53). *C*, near UV CD of selected CI2s. Non-indicated curves lie between those of GCI2-(1-53) and GCI2-(1-63). *D*, ANS binding. Spectra not shown resemble that of GCI2-(1-63). *AU*, arbitrary units. *E*, unfolding of indicated neoglycoproteins with increasing guanidinium hydrochloride (*GdnHcl*) concentrations.

the buffer (Fig. 2*D*; although not shown, GCI2-(1-53) to GCI2-(1-62) also showed ANS binding). ANS is a sensitive probe for the presence of accessible hydrophobic patches. It was concluded that the onset of native secondary and tertiary structure in GCI2s takes place in parallel when 83% of sequence information is attained (GCI2-(1-53) and that the acquisition of the compact cooperatively folded native structure occurs only upon the addition of the last two residues.

**Additional Structural Features of Mutant Neoglycoproteins**—Chemical denaturation of GCI2-(1-63)V63S, GCI2-(1-64)V63S, and of their canonical homologues with guanidinium hydrochloride showed how the V63S mutation and removal of Gly-64 affected the stability of the parental proteins (Fig. 2*E*). The stability of GCI2-(1-64) and GCI2-(1-63) diminished by 1.3 and 0.2 kcal/mol, respectively, upon introduction of mutations (Table I). The destabilizing effect was more pronounced for the full-length protein, since cleavage of the C-terminal

residue has a higher destabilizing effect than the V63S mutation. As we mentioned above, CI2 structure is stabilized through ionic interactions from the C terminus to Arg-46 and Arg-48. The stability differences between wild type and mutated same-length derivatives parallels the shift to longer wavelengths in the fluorescence spectra and the increased ANS binding of the mutants.

**GT Recognizes Native-close Conformations of GCI2 Neoglycoproteins**—Quantitative kinetic experiments with GT were performed because the GCI2 family remained soluble within wide concentration ranges. As shown in Fig. 3*A* GCI2-(1-63) and GCI2-(1-64) were much poorer substrates than the shorter fragments (values yielded by fragments not shown fell between those of GCI2-(1-53) and GCI2-(1-63)V63S). The glucose acceptor capacities of GCI2-(1-63)V63S and GCI2-(1-64)V63S were higher than those of the respective canonical fragments (GCI2-(1-63) and GCI2-(1-64)). Because the C-terminal resi-

TABLE I  
Unfolding parameters of neoglycoproteins

$m$  in column 2 refers to the constant from the equation  $\Delta G[\text{GdnHCl}] = \Delta G_{\text{H}_2\text{O}} + m[\text{GdnHCl}]$ , where  $\Delta G[\text{GdnHCl}]$  (GdnHCl is guanidinium hydrochloride) and  $\Delta G_{\text{H}_2\text{O}}$  are the free energies of folding in water and at a given denaturant concentration.

	$\Delta G_{\text{H}_2\text{O}}$	$m$	$[\text{GdnHCl}]_{50}$
	$\text{kcal mol}^{-1}$	$\text{kcal mol}^{-1} \text{M}^{-1}$	$\text{M}$
GCI2-(1-64)	$4.7 \pm 0.1$	$1.28 \pm 0.05$	$3.7 \pm 0.2$
GCI2-(1-64)V63S	$3.4 \pm 0.1$	$1.10 \pm 0.03$	$3.1 \pm 0.1$
GCI2-(1-63)	$1.9 \pm 0.1$	$0.92 \pm 0.02$	$2.1 \pm 0.1$
GCI2-(1-63)V63S	$1.7 \pm 0.1$	$1.7 \pm 0.1$	$1.0 \pm 0.1$

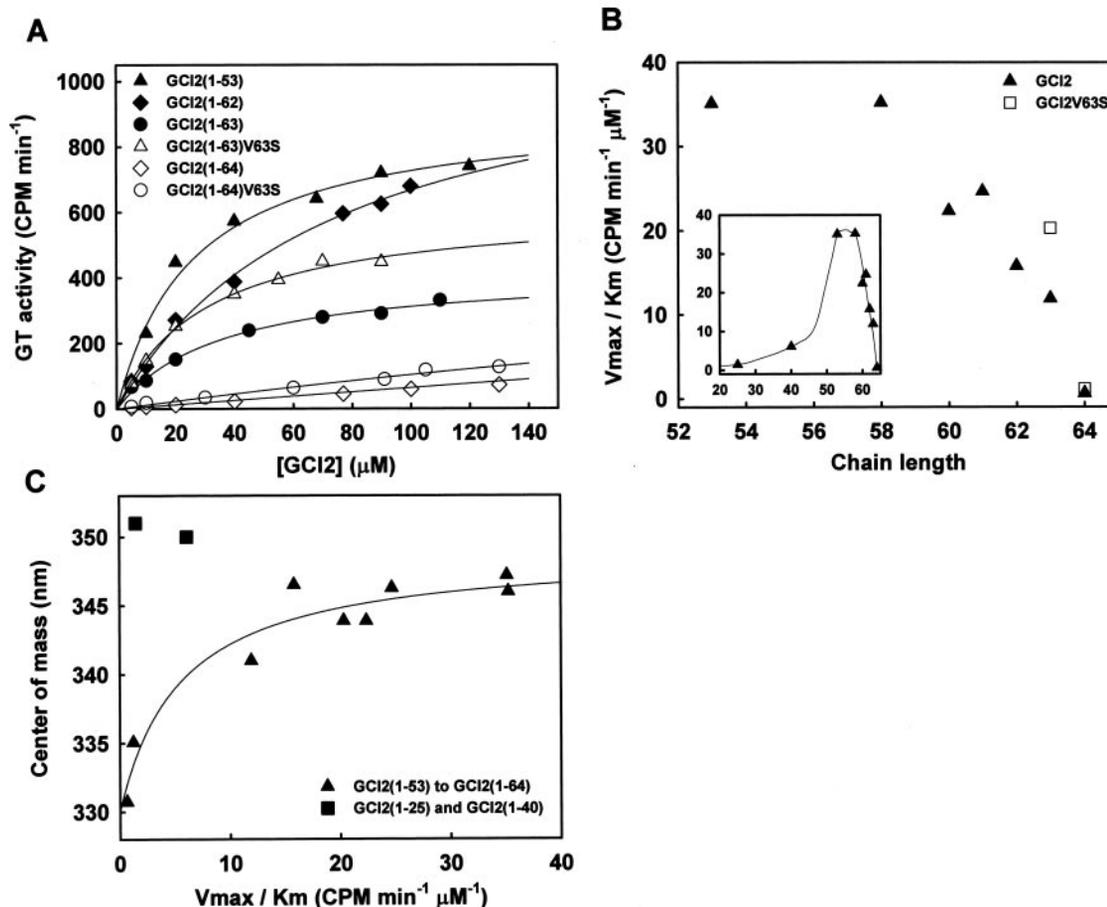


FIG. 3. GT activity. A, GT activity with different neoglycoproteins as acceptor substrates. Curves not shown lie between those of GCI2-(1-62) and GCI2-(1-53). B, GT catalytic efficiency ( $V_{\text{max}}/K_m$ ) versus chain length. Inset, catalytic efficiency versus chain length from GCI2-(1-25) to GCI2-(1-64). C, GT catalytic efficiency versus fluorescence center of mass. Both in this panel and in the inset data for GCI2-(1-25) and GCI2-(1-40) were taken from Caramelo *et al.* (8).

due in GCI2-(1-64) is a glycine, the decrease in GT activity from GCI2-(1-63) to GCI2-(1-64) is not an outcome from its inhibition by a particular amino acid side chain at the C-terminal position but, rather, is a consequence of the effect of this residue on the global structure of GCI2. All efficient substrates bound ANS, supporting the idea that GT recognizes the exposition of the hydrophobic core. The slightly higher glucose acceptor capacity of GCI2-(1-64)V63S with respect to GCI2-(1-64) also correlated with a slightly higher ANS binding capacity (Fig. 2D). It is worth mentioning that we had previously reported that neither GCI2-(1-25), GCI2-(1-40), nor GCI2-(1-64) bind ANS and were poorly or not glycosylated by GT (8).

The GT catalytic efficiency expressed as  $V_{\text{max}}/K_m$  showed a clear trend when expressed as a function of the chain length (Fig. 3B); no significant variation was observed from GCI2-(1-53) to GCI2-(1-58), but a sharp decrease appeared from this species to GCI2-(1-64). Note the higher catalytic efficiency detected with GCI2-(1-63)V63S and GCI2-(1-64)V63S as substrates when compared with the respective canonical com-

pounds. The catalytic efficiency showed a direct hyperbolic relationship with the exposure of the hydrophobic core as indicated by the intrinsic fluorescence mass center values (Fig. 3C; values for GCI2-(1-25) and GCI2-(1-40) were taken from Caramelo *et al.* (8)). Points corresponding to GCI2-(1-25) and GCI2-(1-40) fell outside the curve. As previously determined, those fragments display disordered conformations lacking ANS binding hydrophobic patches, although their Trp-5 residues are solvent-exposed. Results observed reflected, therefore, the loss of GT recognition upon progressive consolidation of the substrate compact fold.

*GT Senses the Thermodynamic Stability of Close to Native Folding Intermediates*—The free energy of unfolding of GCI2-(1-63)V63S, GCI2-(1-64)V63S, and of their wild type, same-length homologues revealed an inverse exponential dependence with GT catalytic efficiencies (Fig. 4A). The free energy of unfolding values of wild type and mutant neoglycoproteins covered a rather wide range, from 1.7 to 4.7 kcal/mol, the last of which closely resembles values observed for medium size

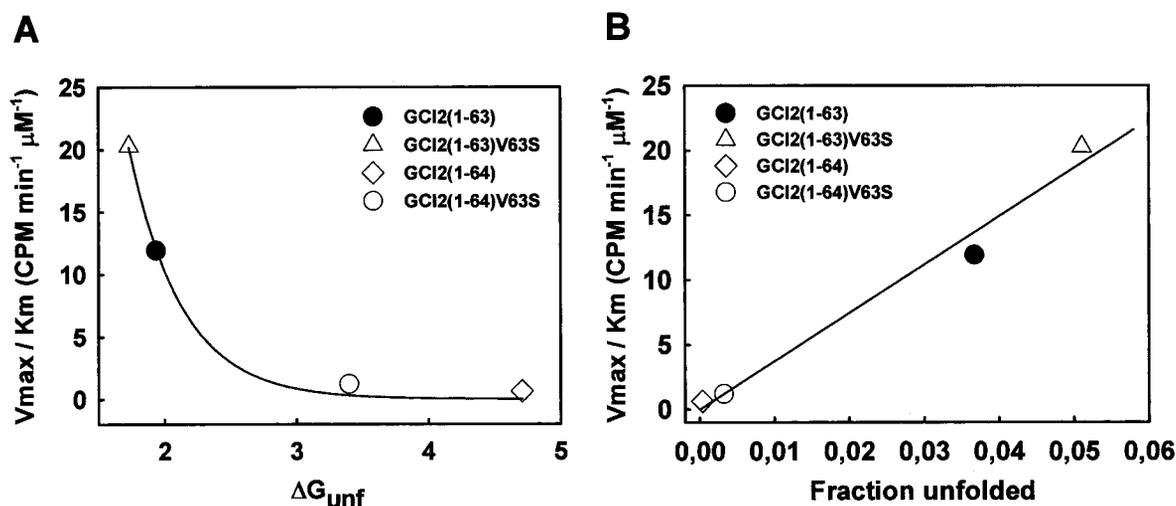


FIG. 4. Relationship between GT catalytic efficiency and neoglycoprotein structural features. A, GT catalytic efficiency versus unfolding free energy ( $\Delta G_{unf}$ ). B, GT catalytic efficiency versus fraction of unfolded neoglycoproteins in water ( $f_{u_{H_2O}}$ ).

monomeric proteins. Notably, the catalytic efficiencies showed a linear relationship with the population fraction of unfolded GCI2 in water (Fig. 4B; the fractions of unfolded GCI2s were calculated assuming similar free energy values for the unfolded states of all four neoglycoproteins). Results presented in Figs. 4, A and B, suggest that perhaps GT is one of the most exquisite biosensors of the stability of protein conformations described to date.

#### DISCUSSION

In a previous report we studied the glucose acceptor capacities of four neoglycoproteins, GCI2-(1-25), GCI2-(1-40), GCI2-(1-53), and GCI2-(1-64) (8), an unstructured hydrophilic peptide, a hydrophobic peptide devoid of long range interactions, a molten globule-like, and a properly folded species, respectively. The onset of secondary and tertiary native structure occurs in GCI2-(1-53), which was the best GT substrate, and we concluded that the enzyme preferentially recognized solvent-accessible hydrophobic regions present in native-like conformations that have not reached their final compact state. Results here reported further extend our observations to additional compounds and confirm our previous conclusion, as GT catalytic efficiency showed a hyperbolic relationship with tryptophan intrinsic fluorescence center of mass but only in ANS binding (*i.e.* exposing hydrophobic patches) compounds (Fig. 3C). For instance GCI2-(1-40), despite displaying a stretch of highly hydrophobic amino acid residues from positions 26 to 39, lacked long range interactions, did not bind ANS, and consequently was a rather poor GT substrate. Furthermore, relationships of catalytic efficiency (Fig. 3B, *inset*) or ANS fluorescence of CI2s (Fig. 5 in de Prat Gay *et al.* (14)) with chain length were practically super imposable. Of the several features of a molten globule conformation (mobility of the polypeptide chain, presence of secondary and tertiary structural elements, free access of the glycan to the solvent, exposure of hydrophobic amino acid patches, etc.) only the latter is not shared with GCI2s that are poor GT substrates.

GT glucosylated neoglycoproteins of increasing length until the substrate displayed its full size, thus acquiring its native compact structure. However, because folding of the neoglycoproteins employed is highly cooperative with the acquisition of the last two residues (14), it may be argued that in other systems in which the three-dimensional structure of fragments lacking a few residues at their C terminus does not significantly differ from that of the full-length species, glucosylation might stop before acquisition of the native structure at its most

minimal details. To test this possibility we introduced a point mutation in the amino acid before last (V63S), known to induce a minimal structural perturbation (15). Although the far UV CD spectra of GCI2-(1-64) and GCI2-(1-64)V63S showed only slight variations and changes observed for near UV CD spectra were minimal as also were their intrinsic fluorescence mass center values and the extent of ANS binding and that the energy of unfolding only showed a respective decrease from 4.7 to 3.4 kcal/mol, both compounds were differentially glucosylated by GT. Similarly, although the free energy of unfolding of GCI2-(1-63) and GCI2-(1-63)V63S only differed by 0.2 kcal/mol, both neoglycoproteins were glucosylated with sharply different catalytic efficiencies. In fact, GT catalytic efficiency showed a linear dependence with the proportion of unfolded GCI2 in water. This last parameter may be taken as an indication of the dynamic behavior of GCI2 as a higher value would indicate that a molecule occurs for longer time periods in putative unfolded conformations. The exquisite way by which GT senses the stability of glycoproteins ensures that no species displaying minimally incomplete three-dimensional structures would leave the ER. The ability of GT to discriminate between minimally differing three-dimensional structures has been recently documented also (7). Introduction of a point mutation (F280S) to exo-(1,3)- $\beta$ -glucanase elicited recognition of the glycoprotein by GT presumably without significantly affecting its three-dimensional structure, as the protein remained enzymatically active.

At this point we can visualize a complete picture of GT recognition of a single folding protein domain emerging into the ER lumen (*inset* in Fig. 3B). CI2-(1-64) has two potential BiP/Grp78 binding sites, one of them centered at Trp-5 and the other at Leu-32, the last one within a stretch of highly hydrophobic residues. Whereas the first one is abolished on linking the glycopeptide to C7, the second site remains active after glycan addition as GCI2-(1-40) (but not GCI2-(1-25) sharply enhanced BiP/Grp78 intrinsic ATPase activity (8). Moreover, a molten globule-like conformation displaying optimal glucose acceptor capacity is not attained until the molecule has about 83% of its sequence information. The protein entering the ER lumen would then be *N*-glycosylated at position seven, recognized by BiP/Grp78 when about 35–40 amino acids are already in the lumen, and collapse on reaching a 50–53-amino acid length. Formation of the molten globule-like structure will then prevent BiP/Grp78 binding because this chaperone recognizes heptapeptides with aliphatic hydrophobic amino acids in alter-

nating positions in proteins displaying extended conformations. As expected, GCI2-(1–53) triggered a much lower BiP/Grp78 ATPase activity than GCI2-(1–40), although both neoglycoproteins shared the same Grp/BiP binding site centered at Leu-32 (8). It is doubtful that immediate partial deglycosylation of the transferred glycan mediated by glucosidases I and II would elicit a long-lived CNX/CRT-GCI2 interaction; because affinity of lectins for monoglucosylated glycans is rather low (16), it would be expected that an initial lectin binding would be followed by dissociation of the complex and glucosidases I- and II-mediated complete deglycosylation. No further glucosylation and, hence, CNX/CRT binding would occur until formation of molten globule-like conformations. Formation of these last conformations would then result in glucosylation-deglycosylation cycles lasting until the native structure at its most minimal details is attained. This interpretation is somewhat at odds with the *in vivo* observation that glycoproteins bearing an *N*-glycan within the first 50 amino acids would interact with CNX/CRT and not with BiP/Grp78 (17). A short-lived interaction with the latter chaperone, perhaps not detectable by immunoprecipitation as it would last as the protein grows from about 35–40 until it reaches 50–53 amino acids, can conceivably take place. Furthermore, the sequential interaction, first with BiP/Grp78 and then with CNX/CRT, has been documented for several glycoproteins (17–20).

How does GT recognize nascent polypeptide substrates? We have previously determined that GT has at least two independent folding domains located at the N- and C-terminal ends, comprising 80 and 20% of the molecule, respectively. Both domains remained tightly but not covalently bound upon a mild proteolytic treatment, after which they could not be separated by common biochemical procedures without loss of enzymatic activity (21). The C-terminal portion is the catalytic domain as it shares sequence similarity and the ability to bind UDP-Glc analogs with glycosyltransferase family 8 enzymes (22). On the other hand, there is strong evidence suggesting that the N-terminal domain is the portion recognizing the folding status of glycoprotein substrates in addition to being required for the C-terminal domain stability (21, 23). It may be speculated that on GT recognition of molten globule-like structures, the contact surface between N- and C-terminal domains becomes somehow partially distorted, thus, allowing binding of the substrate glycoprotein hydrophobic surface. This is highly reminiscent to

what happens on Grp94 recognition of folding intermediates. Grp94 is an ER chaperone that, unlike BiP/Grp78 and the same as GT, recognizes proteins in advanced folding stages. It has been established that Grp94 is a dimer and that the contact surface between both monomers is closely adjacent if not superimposed to that recognizing folding species (24). Probably GT inserts a portion of the N-terminal domain, normally buried in the N-terminal/C-terminal domain interphase, into the substrate native hydrophobic core. This interaction would cease when the final side chain packing of the core is attained, giving place to the stable and consolidated final structure. Experiments to address these issues through spectroscopic methods are in progress.

## REFERENCES

1. Parodi, A. J. (2000) *Annu. Rev. Biochem.* **69**, 69–93
2. Fewell, S. W., Travers, K. J., Weissman, J. S., and Brodsky, J. L. (2001) *Annu. Rev. Genet.* **35**, 149–191
3. Trombetta, E. S., and Parodi, A. J. (2003) *Annu. Rev. Cell Dev. Biol.* **19**, 649–676
4. Petrescu, A. J., Milac, A.L., Petrescu, S. M., Dwek, R. A., and Wormald, M. R. (2004) *Glycobiology* **14**, 103–114
5. Molinari, M., and Helenius, A. (1999) *Nature* **402**, 90–93
6. Taylor, S. C., Thibault, P., Tessier, D. C., Bergeron, J. J., and Thomas, D. Y. (2003) *EMBO Rep.* **4**, 405–411
7. Taylor, S. C., Ferguson, A. D., Bergeron, J. J., and Thomas, D. Y. (2004) *Nat. Struct. Mol. Biol.* **11**, 128–134
8. Caramelo, J. J., Castro, O. A., Alonso, L. G., de Prat-Gay, G., and Parodi, A. J. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 86–91
9. Trombetta, E. S., and Parodi, A. J. (1992) *J. Biol. Chem.* **267**, 9236–9240
10. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M., and Fersht, A. R. (1993) *Biochemistry* **32**, 11259–11269
11. Clackson, T., Detlef, G., and Jones, P. T. (1991) *PCR, a Practical Approach*, (McPherson, M. J., Quirke, P., and Taylor, G. R., eds) p. 202, IRL Press at Oxford University Press, Oxford
12. Schagge, H., and von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379
13. McPhalen, C. A., and James, M. N. (1987) *Biochemistry* **26**, 261–269
14. de Prat Gay, G., Ruiz-Sanz, J., Neira, J. L., Corrales, F. J., Otzen, D. E., Ladurner, A. G., and Fersht, A. R. (1995) *J. Mol. Biol.* **254**, 968–979
15. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) *J. Mol. Biol.* **254**, 260–288
16. Patil, A. R., Thomas, C. J., and Surolia, A. (2000) *J. Biol. Chem.* **275**, 24348–24356
17. Molinari, M., and Helenius, A. (2000) *Science* **288**, 331–333
18. Kim, P. S., and Arvan, P. (1995) *J. Cell Biol.* **128**, 29–38
19. Tomita, Y., Yamashita, T., Sato, H., and Taira, H. (1999) *J. Biochem. (Tokyo)* **126**, 1090–1100
20. Hammond, C., and Helenius, A. (1994) *Science* **266**, 456–458
21. Guerin, M., and Parodi, A. J. (2003) *J. Biol. Chem.* **278**, 20540–20546
22. Tessier, D. C., Dignard, D., Zapun, A., Radominska-Pandya, A., Parodi, A. J., Bergeron, J. J., and Thomas, D. Y. (2000) *Glycobiology* **10**, 403–412
23. Arnold, S. M., and Kaufman, R. J. (2003) *J. Biol. Chem.* **278**, 43320–43328
24. Yamada, S., Ono, T., Mizuno, A., and Nemoto, T. K. (2003) *Eur. J. Biochem.* **270**, 146–154