Effects of the N-Linked Glycans on the 3D Structure of the Free α-Subunit of Human Chorionic Gonadotropin†

Paul J. A. Erbel,‡ Yasmin Karimi-Nejad,§,|| J. Albert van Kuik,‡ Rolf Boelens,|| Johannis P. Kamerling,‡ and Johannes F. G. Vliegenthart*,‡

Department of Bio-Organic Chemistry and Department of NMR Spectroscopy, Bijvoet Center, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

Received December 6, 1999; Revised Manuscript Received March 2, 2000

ABSTRACT: To gain insight into intramolecular carbohydrate−protein interactions at the molecular level, the solution structure of differently deglycosylated variants of the α-subunit of human chorionic gonadotropin have been studied by NMR spectroscopy. Significant differences in chemical shifts and NOE intensities were observed for amino acid residues close to the carbohydrate chain at Asn78 upon deglycosylation beyond Asn78-bound GlcNAc. As no straightforward strategy is available for the calculation of the NMR structure of intact glycoproteins, a suitable computational protocol had to be developed. To this end, the X-PLOR carbohydrate force field designed for structure refinement was extended and modified. Furthermore, a computational strategy was devised to facilitate successful protein folding in the presence of extended glycans during the simulation. The values for ϕ and ψ dihedral angles of the glycosidic linkages of the oligosaccharide core fragments GlcNAc**2**(β1−4)GlcNAc**1** and Man**3**(β1−4)GlcNAc**2** are restricted to a limited range of the broad conformational energy minima accessible for free glycans. This demonstrates that the protein core affects the dynamic behavior of the glycan at Asn78 by steric hindrance. Reciprocally, the NMR structures indicate that the glycan at Asn78 affects the stability of the protein core. The backbone angular order parameters and displacement data of the generated conformers display especially for the β-turn 20−23 a decreased structural order upon splitting off the glycan beyond the Asn78-bound GlcNAc. In particular, the Asn-bound GlcNAc shields the protein surface from the hydrophilic environment through interaction with predominantly hydrophobic amino acid residues located in both twisted β-hairpins consisting of residues 10−28 and 59−84.

Human chorionic gonadotropin (hCG)[1] is a glycoprotein hormone involved in the maintenance of the corpus luteum during early pregnancy. It is a heterodimer consisting of noncovalently associated α- and β-subunits. Both subunits are glycosylated and contain a cystine knot motif formed by three disulfide bridges as a central structural element (*1*, *2*). Besides the role of α-subunit (αhCG or αhCG[glycan^{52,78}]) in the αβ-dimer, it has been discovered that the noncombined α-subunit of hCG (hCG-α_f), which is produced in significant amounts by the placenta during pregnancy, has an independent biological activity in being able to stimulate prolactin secretion from human decidual cells (*3*, *4*).

Glycosylation of proteins has been shown to have many functions, including stabilization of the protein structure (*5*). In the α-subunit of hCG, the two glycosylation sites at Asn52 and Asn78, respectively, have remarkably different properties with respect to the stability of the subunit as shown by site-directed mutagenesis (*6*). The absence of glycosylation at Asn52 does not affect folding and stability of the α-subunit. In contrast, mutant α-subunits lacking glycosylation at Asn78 are poorly secreted and rapidly degraded, although the precise molecular origin of this instability is unknown. For the heterodimeric hCG, X-ray structures were reported (*1*, *2*). Since these data were obtained from protein samples deglycosylated beyond the first monosaccharide unit of the glycosylation sites, very little can be concluded as to the effects of the glycans on the protein structure.

To gain insight into the glycosylation effects, our first aim was the investigation of the structure in solution of deglycosylated variants of the isolated α-subunit. This approach does not only allow a stepwise analysis of the problem, but is also relevant in view of the independent biological activity of the free α subunit. In our studies, we observed that treatment of αhCG[glycan^{52,78}] with endo-β-*N*-acetylglucosaminidase B (endo B) under native conditions results in the removal of the oligosaccharides from both Asn52 and

† This investigation was supported by The Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO).

* To whom correspondence should be addressed.

‡ Department of Bio-Organic Chemistry.

§ Present address: Solvay Pharmaceuticals, WWA A008, P.O. Box 900, 1380 DA Weesp, The Netherlands.

|| Department of NMR Spectroscopy.

[1] Abbreviations: 2D, two-dimensional; NOE, nuclear Overhauser effect; NOESY, 2D nuclear Overhauser and exchange spectroscopy; TOCSY, 2D total correlation spectroscopy; CRMSD, coordinate root-mean-square deviation; RMSD, root-mean-square deviation; Endo B, endo-β-*N*-acetylglucosaminidase B; PNGase F, peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F; αhCG[GlcNAc^{52,78}], deglycosylated αhCG containing only a single GlcNAc at Asn52 and Asn78; hCG, human chorionic gonadotropin; hCG-α_f, noncombined α-subunit of hCG; αhCG or αhCG[glycan^{52,78}], native α-subunit of hCG containing glycans at Asn52 and Asn78; αhCG[glycan^{78}], deglycosylated αhCG containing only the glycan at Asn78.

10.1021/bi992786n CCC: $19.00
Published on Web 04/29/2000

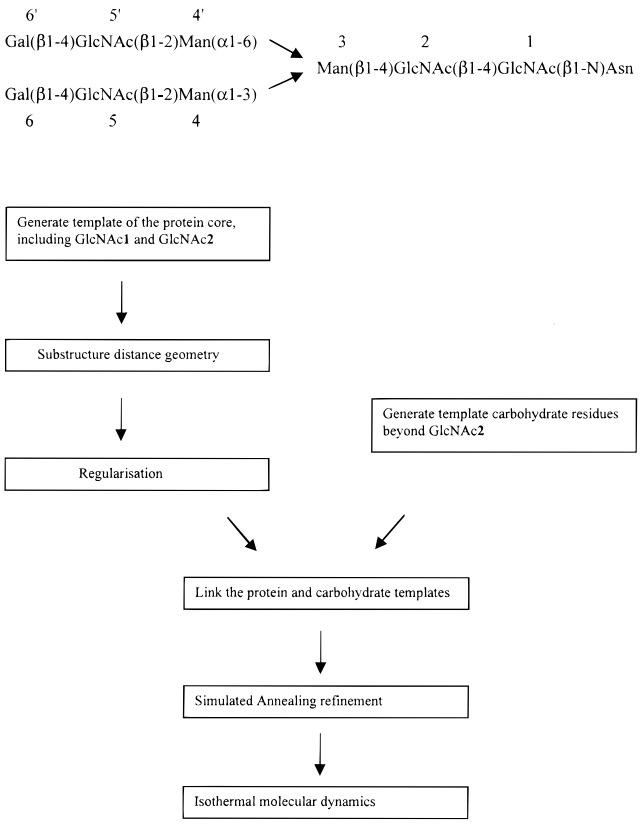


FIGURE 1: Glycan at Asn78 of αhCG; flowchart of the structure determination strategy.

Asn78 (*7*), leaving only a single GlcNAc residue at each site (αhCG[GlcNAc^{52,78}]). Interestingly, digestion of αhCG-[glycan^{52,78}] under native conditions with peptide-*N*^{4}-(*N*-acetyl-*β*-glucosaminyl)asparagine amidase F leads to the selective removal of the glycan at Asn52, leaving the glycan at Asn78 unaltered (αhCG[glycan^{78}]) (*7*). Both structural variants αhCG[GlcNAc^{52,78}] and αhCG[glycan^{78}] were investigated by NMR spectroscopy, and it was found that the overall appearance of the NMR spectra of the two species was rather similar (*8*). Although for the duration of the NMR experiments, both proteins were stable, prolonged storage in solution led to a faster deteriotation of αhCG[GlcNAc^{52,78}] than its fully glycosylated counterpart (*8*). It could also be demonstrated that the thermal stability of αhCG[glycan^{52,78}] is decreased upon trimming down the glycan at Asn78 beyond the Asn-bound GlcNAc^{78} (*9*).

On the basis of a large number of homonuclear ^{1}H-^{1}H NOEs, we determined the solution structure of αhCG-[GlcNAc^{52,78}] (*10*). In the present study, we focus on the solution structure of αhCG[glycan^{78}] to deduce effects of glycosylation by comparison of both protein structures. As no straightforward strategy is available for the calculation of the NMR structure of intact glycoproteins, to enable such a study, a suitable computational protocol had to be developed. Of particular interest are the structural details of the glycan attached to Asn78 (for structure see Figure 1) in αhCG and the impact of the glycan on the protein core of the α-subunit.

EXPERIMENTAL PROCEDURES

NMR Data. To enable the comparison of the 2D homonuclear NOESY spectra of αhCG[GlcNAc^{52,78}] and αhCG-[glycan^{78}], they were recorded under the same experimental conditions i.e., at 328 K and pH 5.1 on a Varian Unity 750 or a Bruker AMX 600 spectrometer. Details regarding the acquisition of the NMR spectra were as reported previously (*8*, *11*).

The NOESY spectra of αhCG[GlcNAc^{52,78}] and αhCG-[glycan^{78}] with mixing times of 80 ms each were analyzed utilizing routines of the in-house written program package REGINE. Distance restraints for the structure calculation of the protein core of αhCG[glycan^{78}] were determined as reported before for αhCG[GlcNAc^{52,78}] (*10*). Distance restraints within the glycan chain were increased by 1 Å, to account for the line widths, being smaller than those of the protein signals. The narrower lines are due to the higher flexibility of the glycan chains as established by relaxation measurements (*11*). In general, the lower limits for the distance restraints were set to the sum of the van der Waals radii (1.8 Å), except for very weak intra-monosaccharide NOE intensities, where the lower limits were set to 4 Å.

Structure Calculation. Structure calculations were performed using the program X-PLOR 3.851 (http://xplor.csb.yale.edu). The calculation strategy was based on the hybrid distance geometry simulated annealing (DG/SA) method (*12*−*14*). To overcome the problems associated with the NMR structure determination of proteins containing complex disulfide bonding topologies, the parameters of the protocol were extensively modified, and an additional refinement step of isothermal molecular dynamics was added (*10*).

A set of force-field parameters and topology files for constituting monosaccharide residues suitable for NMR structure determinations of glycoproteins was derived from the carbohydrate force field supplied with X-PLOR (*15*). The equilibrium values of bond lengths, bond angles, and improper dihedral angles were retained, but the following modifications were implemented.

(1) The force constants for all terms were altered to match those used in the protein force field.

(2) To maintain the chiral centers in the rings, two improper dihedrals per ring carbon were defined with a force constant of 1000 kcal mol^{-1} rad^{-2}.

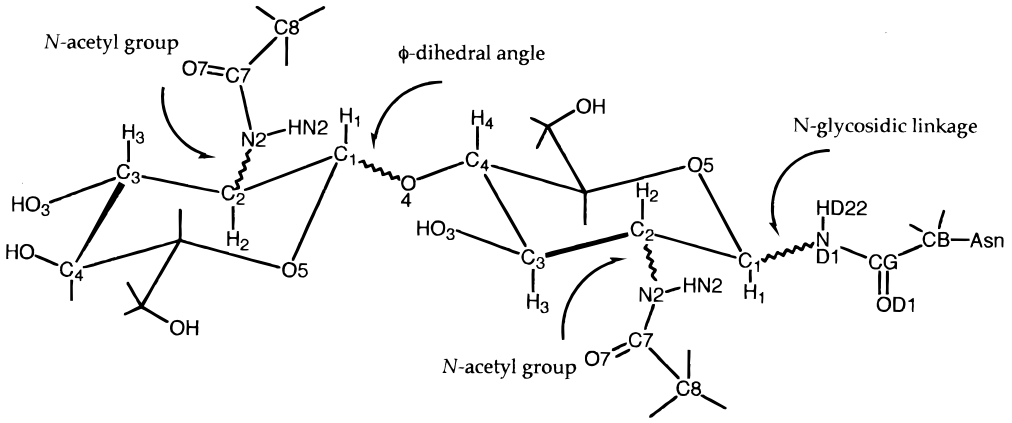
(3) To account for the exo-anomeric effect, the dihedral angle ϕ between the monosaccharide residues was modeled by a superposition of two dihedral terms with a periodicity of 3 and 2, respectively, and a phase-shift angle of 0°.

(4) Force-field parameters and topology data for the *N*-acetyl group in *N*-acetylglucosamine were added to the force field, resulting in a new residue NAG (*N*-acetylglucosamine). To define the *N*-acetyl groups, equilibrium values of bond lengths, bond angles, and improper dihedral angles were derived from the peptide bond parameters of the protein force field (file parallhdg.pro). The force field parameters were extended by a dihedral angle term, with a periodicity of 6 and a phase-shift angle of 180°, and defined 6 times over the linkage of the monosaccharide ring-atom C2 and the nitrogen atom of the *N*-acetyl group.

(5) The N-glycosidic linkage between asparagine and the *N*-acetylglucosamine residue at the reducing end of the glycan was defined via a patch protocol with parameters analogous to the force field for the *N*-acetyl group (Table 1).

Due to the extended and branched nature of the glycan at Asn78 in αhCG[glycan^{78}], the convergence rate of the initial

Table 1: Modifications of the X-PLOR Force Field for Carbohydrates[a]



	N-Acetyl group	
	Dihedral Angles for C2−N2 Linkage	
topology[b]	C1−C2−N2−HN2 C1−C2−N2−C7 C3−C2−N2−HN2 C3−C2−N2−C7 H2−C2−N2−HN2 H2−C2−N2−C7	
parameters[c]	X−CC−NH1−X	0.45 6 180.00
	Improper Dihedral Angles for Planar (H)N−C═O Linkage	
topology	HN2−N2−C7−C8	
parameters	H−NH1−C−CC	500.0 0 0.0
topology	N2−O7−C8−C7	
parameters	NH1−X−X−C	500.0 0 0.0
	N-Glycosidic Linkage (Analogues to *N*-Acetyl Group)	
	Dihedral Angles	
topology	C2−C1−ND1−HD22 etc.	
parameters	X−CCE−NH1−X	0.45 6 180.00
	Improper Dihedral Angles	
topology	CG−CB−OD1−ND1	
parameters	C−CH2E−O−NH1	500.00 0 0.0
topology	HD22−ND1−CG−CB	
parameters	NH1−X−X−CC	500.00 0 0.0
	φ-Dihedral Angle	
	Dihedral Angles to Account for Exo-Anomeric Effect	
topology	C4−O4−C1−O5 MULT 2	
parameters	CC−OA−CCE−OE MULT 2 0.4 2 0.0	1.3293 3 0.0
	Chiral Ring Carbon Atom; Example C3	
	Improper Dihedral Angles to Maintain Chiral Center	
topology	C3−C2−H3−C4 C3−O3−C2−C4 etc. for C1, C2, C4, C5	
parameters	CC−X−X−CC	1000.00 0 35.26439

[a] Files carbohydrate.top and protein-glyco.param distributed with X-PLOR 3.851. [b] New entry added to the topology file carbohydrate.top. [c] New entry added to the parameter file protein-glyco.param.

calculations was low. To overcome this problem, the standard DG/SA-based protocol we designed for αhCG[GlcNAc[52,78]] (*10*), was modified. The distance geometry calculations were initially restricted to the protein and to those monosaccharide residues that exhibit NOEs to the protein core (residue GlcNAc**1** and GlcNAc**2**), to facilitate proper folding of the protein. Subsequently, monosaccharide residues beyond GlcNAc**2** were introduced, and then the structures were refined in a simulated annealing step. A flowchart of the structure determination strategy is depicted in Figure 1.

To select an ensemble of conformers that represents the solution structure of the α-subunit, the following strategy was applied: Initially, models of the α-subunit with low conformational energies were clustered according to their cystine knot topology. The screening produced two families. The family that exhibited significantly higher average conformational energies around the disulfide linkages was discarded. For the remaining conformers, an energy-ordered root-mean-square deviation (RMSD) profile (*16*) was computed by ranking them according to their overall energy. Subsequently, the maximal pairwise backbone coordinate root-mean-square deviation (CRMSD) in consecutive clusters of conformers was determined by starting with the first two lowest-energy conformers, then including the first three, and so forth. From the resulting structural clusters, the final ensemble was selected by consideration of the overall energy. The conformers constituting this ensemble were finally screened for deviations from ideal geometries and the agreement with the experimental data. The mean RMSD from idealized covalent geometry was not allowed to exceed 0.01

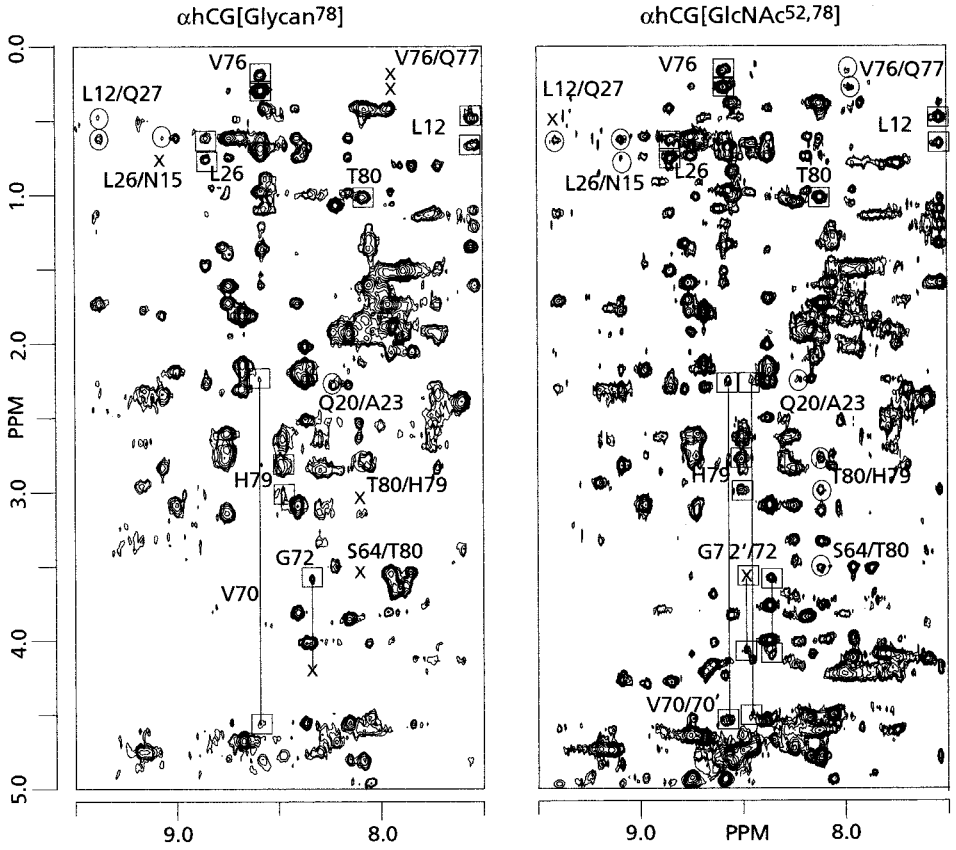


FIGURE 2: Comparison of a part of the 2D ^{1}H-NOE spectrum of αhCG[GlcNAc^{52,78}] (right) and αhCG[glycan^{78}] (left). Both spectra have a mixing time of 80 ms. The NOESY spectrum of αhCG[GlcNAc^{52,78}] was recorded at 328K at 750 MHz, the one of αhCG[glycan^{78}] was recorded at 328 K at 600 MHz ^{1}H resonance frequency. Some cross-peaks are assigned to illustrate the slight differences in terms of NOE patterns, NOE cross-peak intensity or ^{1}H chemical shift. In general, the line width of the NMR spectra of αhCG[glycan^{78}] was larger than that for αhCG[GlcNAc^{52,78}] due to the higher mass and the reduced overall tumbling time. As a consequence, the NMR spectra of αhCG-[glycan^{78}] showed lower cross-peak intensity and increased overlap problems [(*x*) no cross-peak; (□) TOCSY cross-peak; (○) NOE cross-peak].

Å for bonds, 1° for angles, and 1° for the improper dihedrals. The NOE violations were not allowed to exceed 0.5 Å.

For the validation of the glycan conformations, molecular mechanics calculations were performed using the CHEAT force field (*17*) in the INSIGHT II program (BIOSYM/Molecular Simulations, Insight II 95.0 Molecular Modeling System, 1995, San Diego). This force field is based on the treatment of hydroxyl groups as extended atoms, and contains parameters adjusted for aqueous solutions. This prevents the occurrence of unrealistic intramolecular hydrogen bonds in vacuo and reduces the number of local energy minima, because many possible orientations of the hydroxyl torsion angles are avoided. Monosaccharide residues are treated as rigid bodies, using coordinates from X-ray analyses. The ϕ and ψ dihedral angles of glycosidic bonds are the only variables. They are defined according to the IUPAC recommendations (*18*): $\phi = \theta$ (O_{5A}-C_{1A}-O_{XB}-C_XB) and $\psi = \theta$ (C_{1A}-O_{XB}-C_{XB}-$C_{(X-1)B}$). Iso-energy contour plots were calculated for glycosidic linkages by a grid search with steps of 5° in ϕ and ψ, respectively. The calculated "rigid geometry energy maps" contain iso-energy contour levels, which are plotted at intervals of 1 kcal/mol with respect to the calculated global energy minimum.

The computational work was carried out on Silicon Graphics O2 workstations. The structures were visualized by using INSIGHT II (Biosym/MSI) and MOLMOL (*19*). The subsequent numerical analyses and validations of the selected structures were performed by using the programs AQUA, PROCHECK-NMR (*19*), MOLMOL, and X-PLOR 3.851, as well as by in-house written programs in the MATLAB language (MATLAB 5.0, The Math Works, http://www.mathworks.com).

RESULTS

General. A comparison of the NMR parameters of αhCG-[glycan^{78}] (desialylated form) and αhCG[GlcNAc^{52,78}] was made on the basis of 2D NOESY spectra with various mixing times. In general, due to the presence of the extended and hydrated glycan of αhCG[glycan^{78}], the ^{1}H resonances are broader than those of αhCG[GlcNAc^{52,78}]. As a consequence of the increased line widths, NOESY spectra of αhCG-[glycan^{78}] exhibit overall a reduced cross-peak intensity and more signal overlap. The ^{1}H chemical shifts of the amino acid residues of the two α-subunit variants are in most cases similar; however, significant differences are present for the side-chain protons of Leu12 and His79 and the amide protons of Ser19, Tyr65, and Thr80.

Also differences are observed in the NOE intensities of cross-peaks of αhCG[glycan^{78}] in comparison to αhCG-[GlcNAc^{52,78}], which exceed the average intensity differences between NOEs of both α-subunit variants (Figure 2). The NOE intensities involving the residues Leu12, Leu26, Tyr65, Val76, His79, and Thr80 of the β-sheet have a reduced intensity in αhCG[glycan^{78}]. Altered NOE intensities are also

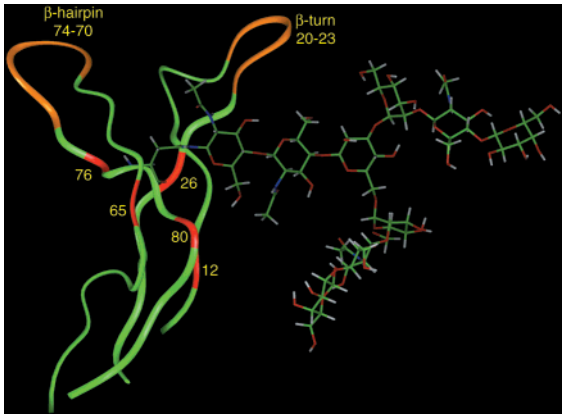


FIGURE 3: Representation of the residues influenced by the presence of the glycan at Asn78. The residues indicated in red show slightly different NOE intensities due to specific interactions with the glycan. The residues located in the heterogeneous β-turn and hairpin loop are drawn in orange. The residues that are unaffected by the presence of the glycan are represented in green.

found for residues located in the turn segments of the β-hairpins of the α-subunit (Figure 3). In detail, the β-turn comprising amino acid residues 20−23 shows increased NOE intensities in the case of αhCG[glycan^{78}] for the cross-peaks between residues Phe18/Ile25, Gln20/Ala23, and Ala23/Pro24, and decreased intensities for the NOEs between residues Ser19/Gln20 and Gly22/Ala23. In the β-hairpin loop comprising residues 70−74, increased cross-peak intensities are found between the residues Val70/Phe74 and Met71/Phe74 for αhCG[glycan^{78}]. In summary, 29 altered NOE intensities were observed resulting in 13 increased and 16 decreased upper limits in the distance restraints of αhCG-[glycan^{78}] in comparison to αhCG[GlcNAc^{52,78}].

Structure Determination and Quality of the Structures. To model αhCG[GlcNAc^{52,78}], in total 666 distance restraints were derived from the NOESY spectra as described previously (*10*). The number of intraprotein NOE distances was 620. The GlcNAc at Asn52 was modeled with 2 Asn52-GlcNAc and 4 intra-GlcNAc distance restraints, whereas the GlcNAc at Asn78 was modeled with 28 protein−GlcNAc and 12 intra−GlcNAc distance restraints. αhCG[glycan^{78}] was calculated with a similar set of NOE distances for the protein core taking into account the 29 altered NOE intensities mentioned above. The glycan at Asn78 in αhCG-[glycan^{78}] was modeled with 21 inter-monosaccharide NOEs and 39 intra-monosaccharide NOEs. The set of protein−carbohydrate NOE restraints was extended, as compared to αhCG[GlcNAc^{52,78}], by including NOEs between Leu12/GlcNAc**2** and Leu26/GlcNAc**2**. In total, 30 distance restraints were assigned between the protein and GlcNAc**1/2**.

Previously, we described (*10*) an X-PLOR based DG/SA calculation protocol specifically designed to give high convergence rates and good local geometry of the complex pattern of five disulfide bridges in the cystine knot of αhCG-[GlcNAc^{52,78}]. To model αhCG[glycan^{78}], containing an intact diantennary structure, we extended and modified the X-PLOR carbohydrate force field designed for refinement. The modifications were necessary for three reasons: (i) an *N*-acetyl group had to be defined to build the GlcNAc unit; (ii) the N-glycosyl linkage to Asn had to be defined; and (iii) the high temperature in the simulated annealing step, necessary for the correct folding of the protein part, and the addition of distance restraints require adjustments of force constants to maintain optimal ring conformations. A further difficulty was caused by the branched topology of the glycan in αhCG[glycan^{78}], which presented a serious obstacle for the convergence of the calculations. To overcome this problem, we devised a suitable computational strategy for the NMR structure determination of glycoproteins. It relies on a stepwise introduction of the glycan part, thereby facilitating the protein folding during the simulation (see Experimental Procedures). The protocol yielded a high rate of convergence (Table 2).

Out of 100 conformers of αhCG[GlcNAc^{52,78}] initially generated, an ensemble of 26 conformers was selected, based on an energy-ranked CRMSD profile (Figure 4a). Details of the selection criteria are given in the section Experimental Procedures.

To represent the solution structure of αhCG[glycan^{78}] (Figure 4b), at first an ensemble of 29 conformers was selected out of initially 150 conformers of αhCG[glycan^{78}], based on the same criteria. Since the selection criteria consider exclusively the protein geometry of the calculated conformers, an additional screening of the carbohydrate conformations of the ensemble was required. To this end, the structural quality of the glycan part of the conformers was validated by assessing two crucial conformational details: the conformation of the pyranose rings and the conformation around the glycosidic linkages. Their ϕ and ψ torsion angles were assessed by comparison to the energy minima in rigid geometry energy maps of the disaccharides GlcNAc(β1−4)GlcNAc and Man(β1−4)GlcNAc, which were determined by molecular mechanics calculations (see Experimental Procedures). Consequently, after the first protein-focused selection step of the αhCG[glycan^{78}] ensemble, three conformers were excluded that have improper geometry of the glycosidic linkages between GlcNAc**2** and GlcNAc**1** or Man**3** and GlcNAc**2** or because of a strong increase in the overall NOE energy of the glycan part.

The individual backbone conformations for the αhCG-[GlcNAc^{52,78}] and αhCG[glycan^{78}] conformers of all non-glycine amino acid residues in the well-defined regions including residues 10−28 and 59−84 are presented in the Ramachandran map shown in Figure 5 (*20*). In the case of αhCG[glycan^{78}], 35.0% of these residues is located in most favored regions and 63.0% in allowed regions of the Φ/Ψ space. In the disallowed region, only 2.0% of the amino acid residues was found. In αhCG[GlcNAc^{52,78}], the corresponding values are similar, viz., 34.2, 63.9, and 1.9%. Previously, we determined the conformation of αhCG[GlcNAc^{52,78}] with an X-PLOR protocol unable to handle monosaccharide residues (*10*). In comparison to the Φ and Ψ dihedral angles of Asn78, calculated with this elementary X-PLOR protocol, the backbone conformation of Asn78 in the conformers, generated with a X-PLOR protocol able to handle monosaccharides, improved from a conformation in the disallowed region to the allowed region of the Φ/Ψ space.

Structural Comparison of the Protein Core of αhCG-[GlcNAc^{52,78}] and αhCG[glycan^{78}]. Both αhCG species show a closely related nonglobular protein folding (Figure 6). The well-defined core formed by the residues 10−28 and 59−84 consists of two twisted β-hairpins linked by three disulfide bridges (Cys28:Cys82, Cys 10:Cys 60, and Cys32:Cys84) forming a cystine knot. The two hairpins are connected by

Table 2: Structural Statistics for the Conformer Ensembles of αhCG[GlcNAc^{52,78}] and αhCG[glycan^{78}][a]

	αhCG[GlcNAc^{52,78}]	αhCG[glycan^{78}]
no. of distance restraints		
protein-core: total non-intraresidual	620	620
protein-core → glycan at Asn78	28	30
inter-monosaccharides at Asn78		21
intra-monosaccharides at Asn78	12	39
mean RMS deviations from experimental restraints		
NOE (Å)	$43.5 \times 10^{-3} \pm 1.3 \times 10^{-3}$	$42.9 \times 10^{-3} \pm 2.1 \times 10^{-3}$
mean RMS deviations from idealized covalent geometry		
bonds (Å)	$6.10 \times 10^{-3} \pm 0.23 \times 10^{-3}$	$6.21 \times 10^{-3} \pm 0.3 \times 10^{-3}$
angles (deg)	0.763 ± 0.019	0.775 ± 0.023
impropers (deg)	0.665 ± 0.016	0.678 ± 0.025
dihedrals (deg)	27.1 ± 1.6	27.5 ± 1.5
restraint violations		
mean NOE violations per structure > 0.2 Å	7.7 ± 1.6	8.4 ± 1.6
mean energies (kcal mol^{-1})		
E_{noe}	62.8 ± 3.9	65.4 ± 6.3
E_{vdW}[b]	67.0 ± 4.5	72.1 ± 4.9
E_{bond}	24.1 ± 1.8	25.0 ± 2.5
E_{angle}	103.4 ± 5.3	106.4 ± 6.5
$E_{improper}$	24.4 ± 1.3	25.5 ± 2.0
$E_{dihedrals}$	1.40 ± 0.97	1.43 ± 0.61
coordinate RMS differences (Å)[c]	average pairwise	
backbone atoms 10−28, 59−70, 75−84	0.75 ± 0.17	0.71 ± 0.13
heavy atoms 10−28, 59−70, 75−84	1.36 ± 0.21	1.41 ± 0.22

[a] Both ensembles comprise 26 conformers. [b] The parallhdg force constant parameter set of X-PLOR 3.851 was used with the quartic repel core repulsion potential. No attractive van der Waals or electrostatic terms were used. [c] The segments 10−28, 59−70, and 75−84 constitute to the well-defined region defined by backbone angular order parameter >0.9.

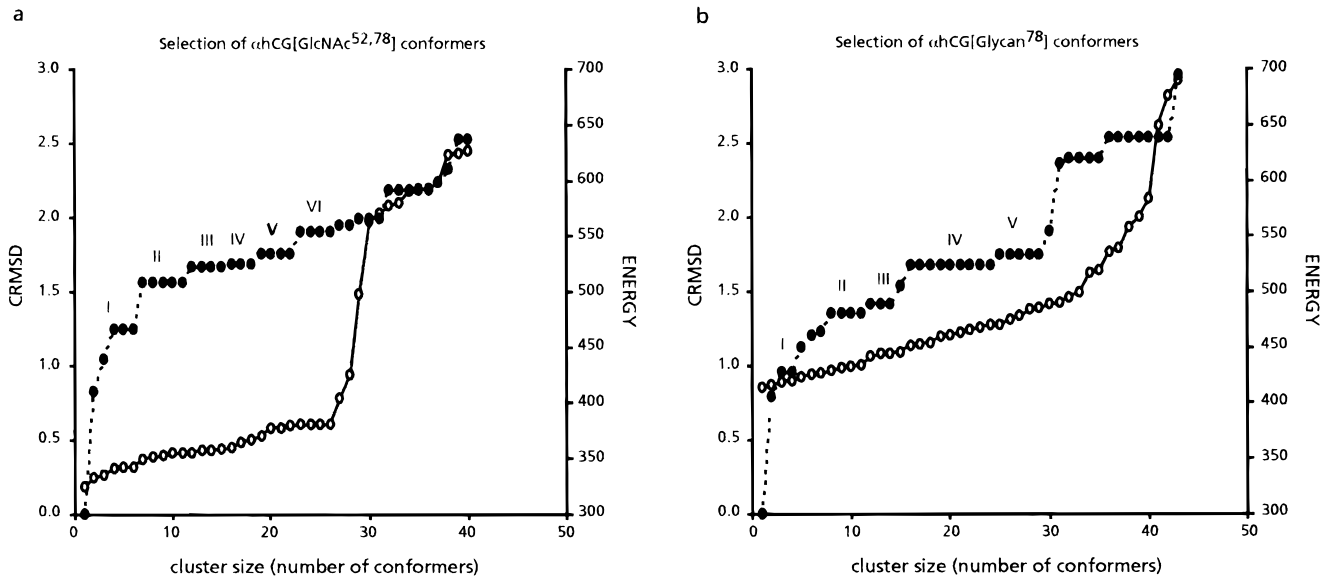


FIGURE 4: Maximal pairwise CRMSD of all backbone atoms of residues 8−32 and 58−88, respectively, as a function of the conformers, ranked to their overall energy (dotted line), is given on the left *y*-axis. (A) αhCG[GlcNAc^{52,78}], (B) αhCG[glycan^{78}]. The CRMSD values were calculated after a superposition of these backbone residues. The different clusters are denoted with Roman numbers. The second graph (continuous line) shows the overall energy of the individual conformers, quantified on the right side of the *y*-axis.

a disordered loop including the residues 33−57. The N- and C-terminal segments, comprising residues 1−9 and 85−92, respectively, are also disordered and stick out into the solvent like two arms. They are connected to the core of the protein by two disulfide bridges (Cys7:Cys31 and Cys59:Cys87) in a hinge-like manner. A detailed investigation of the two β-hairpins reveals that the first β-hairpin consists of two extended segments including residues 10−14 and 24−28. They are connected by two interlocking turns, namely a 3_{10}-type turn formed by residues 15−19 and a classic β-turn comprising residues 20−23. The other β-hairpin is formed by two strands 59−69 and 75−85 respectively, which are connected by a hairpin loop formed by residues 70−74.

Analysis of the backbone angular order parameters (*21*) of the solution structure of αhCG[GlcNAc^{52,78}] and αhCG-[glycan^{78}], giving a measure of their local precision, are shown in Figure 7, panels a−d. As compared to αhCG-[GlcNAc^{52,78}], the degrees of structural definition in the backbone near Leu12 and in the β-turn 20 to 23 are slightly higher in αhCG[glycan^{78}]. In the second hairpin, the comparison of backbone angular order parameters reveals a less uniform picture. For Met71 and Gly73 in αhCG-[glycan^{78}], the degree of local definition in the backbone is higher than in αhCG[GlcNAc^{52,78}], whereas for Gly72 and Phe74, the opposite is true. We also computed global displacements for all residues with backbone angular order

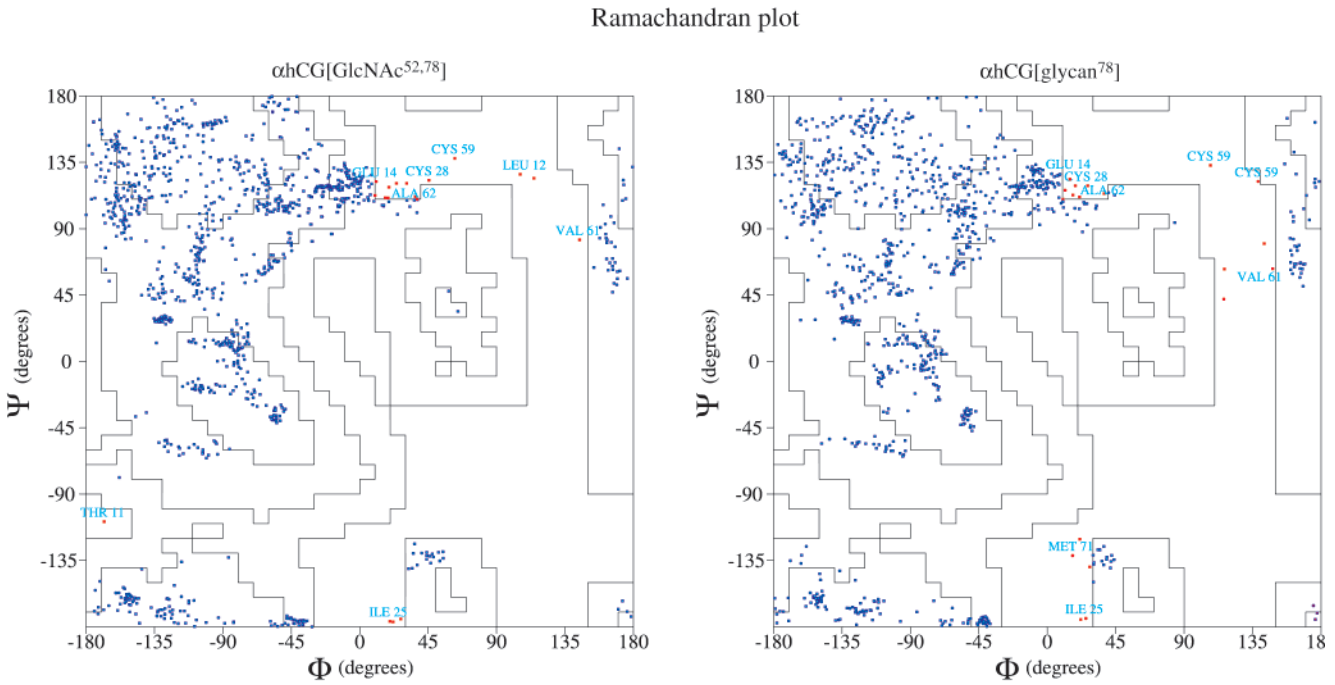


FIGURE 5: Ramachandran plot for the ensemble of solution structures of αhCG[GlcNAc^{52,78}] and αhCG[glycan^{78}]. All the nonglycine amino acid residues of the well-defined segments 10−28 and 59−84 are shown. (Φ angle definition = C'_{i-1}−N_i−C_i^{α}−C'_i, Ψ angle definition: N_i-C_i^{α}-C'_i-N_{i+1}).

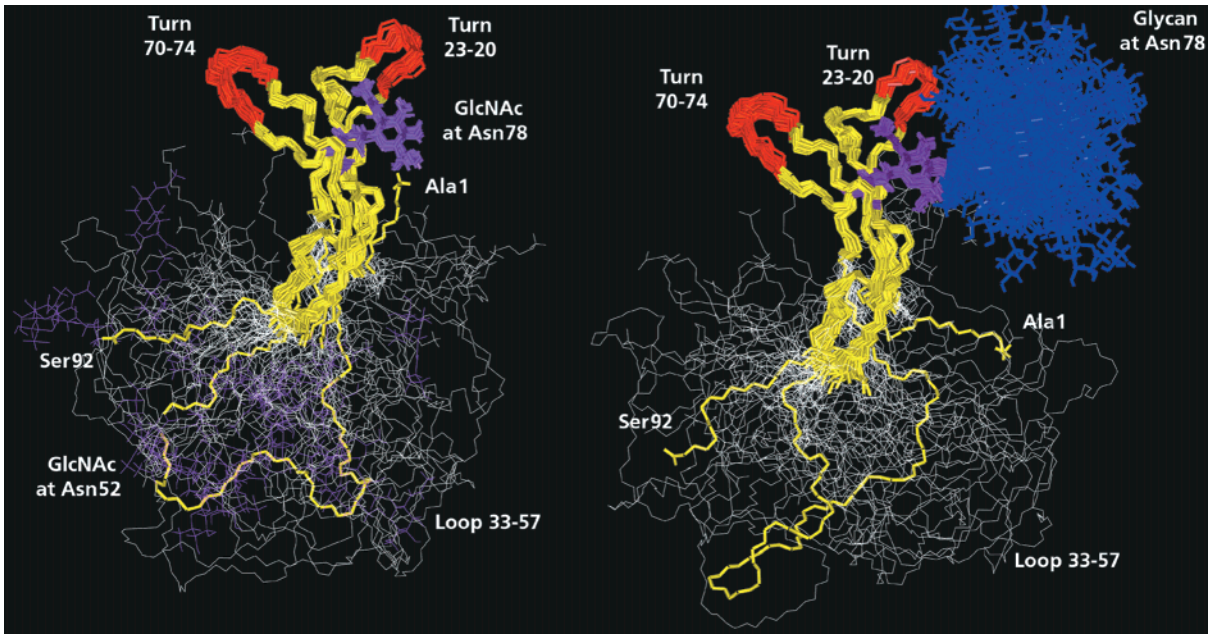


FIGURE 6: Representation of the backbone conformation of the 26 conformers of αhCG[GlcNAc^{52,78}] (left) and αhCG[glycan^{78}] (right). The superposition was carried out including the backbone atoms of the residues 10−28, 59−70, and 75−84. The residues outside this region indicated in white represent the flexible and undefined part of the α-subunit, except one conformer which is represented in a yellow ribbon. The *β*-sheet is drawn in yellow. The tight turn and the hairpin loop are colored in red. GlcNAc**1** is presented in purple and the residues beyond GlcNAc**1** in blue.

parameters > 0.9 (*22*). The result is shown in Figure 7, panels e and f. The displacement data give a picture similar to the order parameters, especially indicating decreased conformational heterogeneity of the *β*-turn 20−23 in αhCG[glycan^{78}] as compared to αhCG[GlcNAc^{52,78}].

DISCUSSION

The present investigation was aimed to determine to which extent the carbohydrate part of the free α-subunit of hCG affects the protein core. To this end, different deglycosylated samples of the α-subunit were prepared and studied by NMR spectroscopy. The glycan at Asn78 of the native α-subunit can only be split off beyond GlcNAc**1**, nevertheless significant differences occur upon deglycosylation in the tips of the twisted *β*-hairpins, which are part of the well-defined protein core of α-subunit. For the *β*-turn constituting amino acid residues 20−23, the structure of αhCG[glycan^{78}] exhibits a higher degree of backbone order than the structure

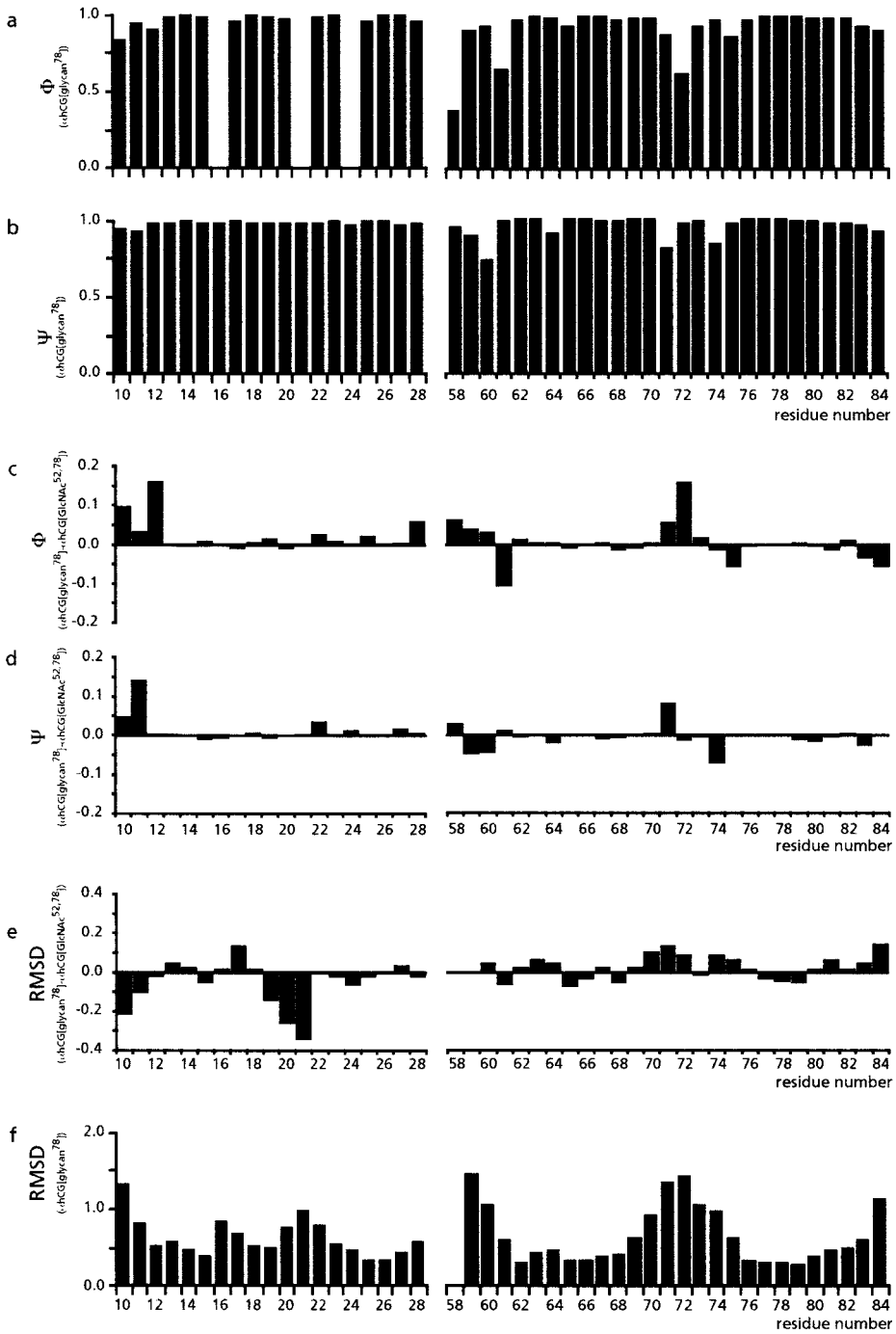


FIGURE 7: Angular order parameters for the torsional angles (a) ϕ and (b) ψ as a measure of local precision of the structure of αhCG-[glycan^{78}]. The difference of the angular order parameter between αhCG[glycan^{78}] minus αhCG[GlcNAc^{52,78}] is displayed for the torsional angles ϕ and ψ in Figure 8, panels c and d, respectively. The difference in the global displacement data between αhCG[glycan^{78}] minus αhCG[GlcNAc^{52,78}] is shown in Figure 8e. (f) The global displacement of αhCG[glycan^{78}].

of αhCG[GlcNAc^{52,78}]. Previously, we observed after prolonged storage of αhCG[GlcNAc^{52,78}] in solution two sets of ^{1}H resonances with an intensity ratio of approximately 1:4 (*8*) for the amino acid residues 69−75. This indicates that the second β-hairpin loop is involved in a conformational exchange process being slow on the NMR time scale. The ensemble of solution structures of αhCG[GlcNAc^{52,78}] only represents the predominant conformation of αhCG-[GlcNAc^{52,78}] in solution; consequently, the degree of definition present in the segment 69−75 in this ensemble might be overestimated. In contrast, no evidence for such a dynamic equilibrium was found in the spectra of αhCG[glycan^{78}].

The conformation of the glycan chain in αhCG[glycan^{78}], as obtained from structure calculations, is in good agreement with the energy minima observed in rigid geometry energy maps of free model disaccharides. Iso-energy contour plots of the ϕ and ψ dihedral angles characterizing the glycosidic linkages in GlcNAc(β1→4)GlcNAc and Man(β1→4)GlcNAc (Figure 8) show the broad conformational energy minima accessible to these disaccharides. In contrast, the glycosidic linkages in GlcNAc**2**(β1→4)GlcNAc**1** and Man**3**(β1→4)-GlcNAc**2** in αhCG[glycan^{78}] are sterically hindered by the protein, as reflected by the limited range of values for ϕ and ψ dihedral angles observed in the ensemble of solution

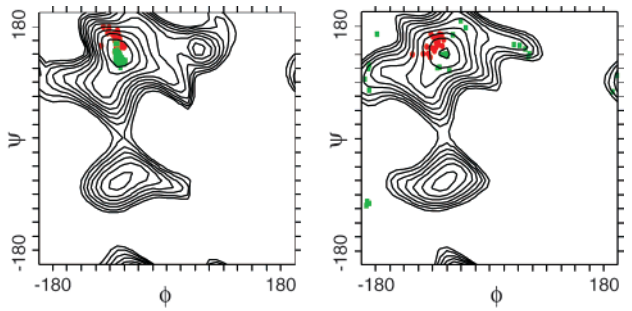


FIGURE 8: Plots in ϕ,ψ surfaces (A) GlcNAc(β1−4)GlcNAc and (B) Man(β1−4)GlcNAc. In red dots the ϕ and ψ angles of the glycosidic linkages at Asn78 in the 26 αhCG[glycan^{78}] conformers are shown. As a reference, the ϕ and ψ angles of the glycosidic linkages of a glycan attached to a protein, lacking interaction between the protein core and the glycan beyond the Asn-bound GlcNAc, are displayed in green dots. The iso-energy contour plots are obtained by CHEAT calculations at intervals of 1 kcal/mol from the global minimum. (ϕ angle definition = O5−C1−O4−C4; ψ angle definition = C1−O4−C4−C3).

structures of αhCG[glycan^{78}]. These findings fit nicely the ^{13}C relaxation parameters $T_{1\rho}$ and T_2 we determined previously for the corresponding carbohydrate units in αhCG-[glycan^{78}] (*11*). These data indicate an increased flexibility toward the outer end of its glycan chain (*23*) (Figure 9 and Table 3). Other aspects of the carbohydrate conformation in αhCG[glycan^{78}], such as the predominant $^4C_1$ chair conformation of the pyranose rings, are also excellently reproduced by the structure calculations solely based on the NOE cross-peaks between H1−H3−H5 and H2−H4. One can therefore conclude that the modifications and extensions we added to the X-PLOR carbohydrate force field result in a balance between experimental NOE restraints and conformational energy terms.

We have reported previously, based on NOE spectra of fully glycosylated αhCG including the intact glycan at Asn52, that the mobilities of the glycans at Asn52 and Asn78 are different (*8*) (Figure 9). Especially, the core structure of the glycan at Asn78 has a reduced mobility. In contrast, the

Table 3: Line Widths and the Corresponding Atomic RMS Differences (Å) of the H1 Proton of the Monosaccharide Residues in the Glycan at Asn78

residue	line width (Hz)	global displacement
GlcNAc**1**	>29	0.47
GlcNAc**2**	29	1.21
Man**3**	20	3.05
Man**4**	7.2	4.86
Man**4′**	7.2	5.01
GlcNAc**5/5′**	nd[a]	7.58/6.81
Gal**6/6′**	nd	10.96/12.97

[a] nd = not determined.

dynamic behavior of the glycan at Asn52 is comparable to the mobility of the free glycan. In agreement with these data is the finding that of the two N-glycosylation sites Asn52 and Asn78 under native conditions only the glycan at Asn52 in αhCG is susceptible to cleavage by PNGase F (*7*). The model proposed here of the glycosylation site at Asn78 clearly illustrates that PNGase F is not capable of approaching the tightly packed Asn78-GlcNAc**1** linkage. These results disprove the suggestion that, based on a single set of glycan signals measured by NMR at ^{13}C and ^{15}N labeled αhCG sample, both glycans should have a similar dynamical behavior, like a free glycan (*24*). A possible explanation for this disagreement is the extreme line broadening beyond detection of NMR signals of the GlcNAc at Asn78 in the doubly labeled samples.

Glycosylation of proteins has been found to be involved in processes such as protein folding, molecular recognition, and protein stabilization. However, these functions of glycosylation are only partially understood at a molecular level. Concerning hCG, site-directed mutagenesis experiments have shown that the glycan at Asn52 is relevant for biological activity (*25*). From comparison of the X-ray structures of intact hCG (*1*, *2*) and the NMR structure of its α-subunit (*10*), it can be inferred that dissociation of the dimer results in large structural changes in the α-subunit. The region 33−57 is dramatically affected by becoming completely unstruc-

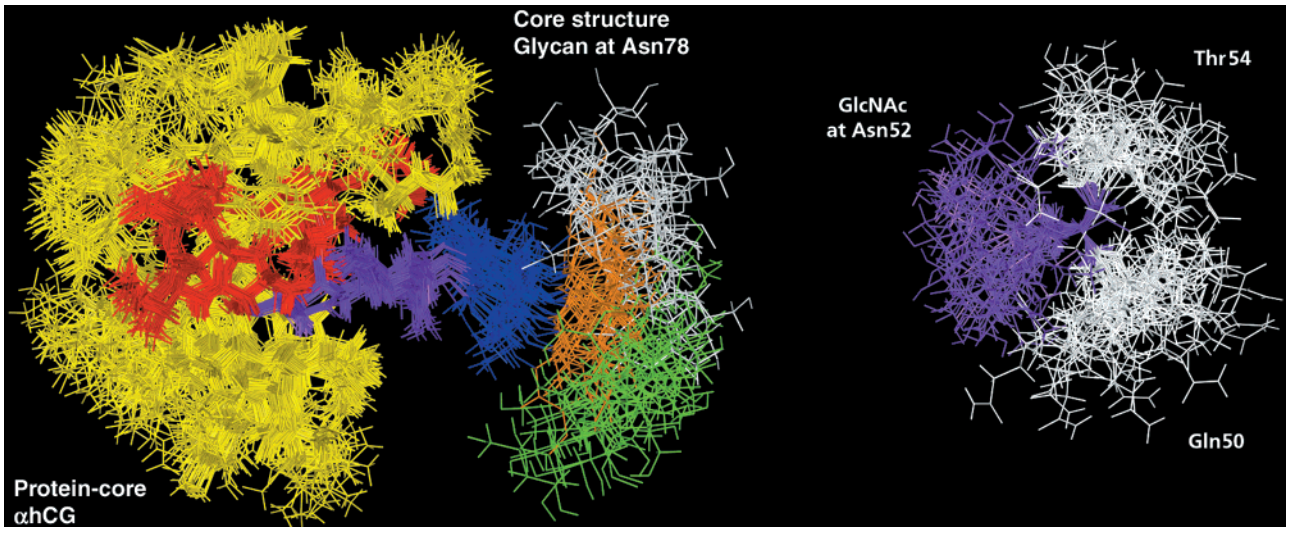


FIGURE 9: Representation of the dynamics of the carbohydrate structure. (left) Dynamics of the carbohydrate core at Asn78. The superposition was carried out including the backbone atoms of the residues 10−28, 59−70, and 75−84. The protein core, including the heavy atoms, is shown in yellow. The hydrophobic residues Leu12, Pro24, Ile25, Leu26, Val68, Thr69, Val70, and Val76 that interact with the glycan at Asn78 are colored in red. GlcNAc**1** is shown in purple. GlcNAc**2**, Man**3**, Man**4**, and Man**4′** is subsequently colored in blue, orange, green and white. (right) Dynamics of GlcNAc**1** at Asn52. The superposition was carried out including the backbone and side-chain atoms of Asn52. GlcNAc**1** and Asn52 are shown in purple and the residues 50−51 and 53−54 of the flexible loop are colored white.

tured in the dissociated α-subunit. The loss of structural definition around the glycosylation site at Asn52, together with the high mobility of the glycan linked to it, explain why the glycan at Asn52 has no influence on the stability of the free α-subunit. In the past, it has been proposed that the absence of the glycan at Asn52 decreases the stability and secretion of the heterodimer (*6*, *26*). In contrast, the removal of the carbohydrate residues beyond GlcNAc at Asn78 in αhCG[GlcNAc^{52,78}] resulted in decreased thermal stability as shown previously (*9*). On the molecular level, this can be explained by increased disorder of the protein segments, close to the carbohydrate chain at Asn78, as demonstrated in this study. Moreover, the complete removal of the glycan at Asn78 in the α-subunit by mutagenesis resulted in poor secretion (<20%) of hCG and rapid degradation of the mutant glycoprotein (*6*). The presence of the *β*-subunit partially stabilizes such mutants and allows about 45% of the subunit to exit the cell in the form of a dimer.

So far, only the stabilization of the adhesion domain of human CD2 by glycosylation has been well-studied. Upon removal of the glycans beyond GlcNAc**1** (*27*), similar features were observed as in the case of the α-subunit of hCG. Although on a macroscopic level glycosylation seems to exhibit similar effects on the stability of these glycoproteins, the underlying molecular feature is completely different. In the adhesion domains of CD2, the glycan can stabilize the protein conformation through counterbalancing the clustering of the five positively charged lysines at the protein surface (*27*). However, in the α-subnit of hCG, the glycan at Asn78 exerts its protective function by shielding the protein surface from the environment through interaction with predominantly hydrophobic amino acid residues (*8*). In both glycoproteins, the strongest interactions are observed at GlcNAc**1** of the glycan.

In summary, glycosylation of α-subunit of hCG at Asn78 plays an important role in stabilizing the protein. Besides macroscopic effects such as increased solubility and increased protease resistance, especially GlcNAc**1** is important to stabilize the 3D-structure.

ACKNOWLEDGMENT

The crude preparation of hCG was kindly provided by Prof. Dr. G. W. K. van Dedem (Diosynth B.V., Oss, The Netherlands). Endoglycosidase B was a generous gift from Dr. S. Bouquelet, Universitè des Sciences et Techniques de Lille Flandres-Artois, Villeneuve d'Ascq, France. We thank Dr. Carol Thijssen-van Zuylen for her exploring studies of the NMR spectra of αhCG[glycan^{78}].

REFERENCES

1. Lapthorn, A. J., Harris, D. C., Littlejohn, A., Lustbader, J. W., Canfield, R. E., Machin, K. J., Morgan, F. J., and Isaacs, N. W. (1994) *Nature 369*, 455−461.
2. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., and Hendrickson, W. A. (1994) *Structure 2*, 545−558.
3. Blithe, D. L., Richards, R. G., and Skarulis, M. C. (1991) *Endocrinology 129*, 2257−2259.
4. Blithe, D. L., and Iles, R. K. (1995) *Endocrinology 136*, 903−910.
5. Wyss, D. F., and Wagner, G. (1996) *Curr. Opin. Biotechnol. 7*, 409−416.
6. Matzuk, M. M., and Boime, I. (1988) *J. Cell. Biol. 106*, 1049−1059.
7. Van Zuylen, C. W. E. M., De Beer, T., Rademaker, G. J., Haverkamp, J., Thomas-Oates, J. E., Hård, K., Kamerling, J. P., and Vliegenthart, J. F. G. (1995) *Eur. J. Biochem. 231*, 754−760.
8. De Beer, T., Van Zuylen, C. W. E. M., Leeflang, B. R., Hård, K., Boelens, R., Kaptein, R., Kamerling, J. P., and Vliegenthart, J. F. G. (1996) *Eur. J. Biochem. 241*, 229−242.
9. Van Zuylen, C. W. E. M., Kamerling, J. P., and Vliegenthart, J. F. G. (1997) *Biochem. Biophys. Res. Commun. 232*, 117−120.
10. Erbel, P. J. A., Karimi-Nejad, Y., De Beer, T., Boelens, R., Kamerling, J. P., and Vliegenthart, J. F. G. (1999) *Eur. J. Biochem. 260*, 490−498.
11. Thijssen-van Zuylen, C. W. E. M., De Beer, T., Leeflang, B. R., Boelens, R., Kaptein, R., Kamerling, J. P., and Vliegenthart, J. F. G. (1998) *Biochemistry 37*, 1933−1940.
12. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
13. Nilges, M., Kuszewski, J., and Brünger, A. T. (1991) *Computational Aspects of the Study of Biological Macromolecules by NMR* (Hoch, J. C., Ed.) New York, Plenum Press.
14. Kuszewski, J., Nilges, M., and Brünger, A. T. (1992) *J. Biomol. NMR 2*, 33−56.
15. Weis, W. I., Brünger, A. T., Skehel, J. J., and Wiley, D. C. (1990) *J. Mol. Biol. 212*, 737−761.
16. Widmer, H., Widmer, A., and Braun, W. (1993) *J. Biomol. NMR 3*, 307−324.
17. Grootenhuis, P. D. J., and Haasnoot, C. A. G. (1988) *Mol. Simul. 10*, 75−95.
18. IUPAC−IUB Joint commission on biochemical nomenclature (JCBN) (1983) *Eur. J. Biochem. 131*, 5−7.
19. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
20. Laskowski, R. A., Rullman, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
21. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) *Protein Sci. 1*, 736−751.
22. Billeter, M. (1992) *Q. Rev. Biophys. 25*, 325−377.
23. Lommerse, J. P. M., Kroon-Batenburg, L. M. J., Kroon, J., Kamerling, J. P., and Vliegenthart, J. F. G. (1995) *Biochemistry 34*, 8196−8206.
24. Weller, C. T., Lustbader, J., Seshadri, K., Brown, J. M., Chadwick, C. A., Kolthoff, C. E., Ramnarain, S., Pollak, S., Canfield, R., and Homans, S. W. (1996) *Biochemistry 35*, 8815−8823.
25. Matzuk, M. M., Keene, J. L., and Boime, I. (1989) *J. Biol. Chem. 264*, 2409−2414.
26. Heikoop, J. C., Van den Boogaart, P., De Leeuw, R., Rose, U. M., Mulders, J. W. M., and Grootenhuis, P. D. J. (1998) *Eur. J. Biochem. 253*, 354−356.
27. Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanadam, A. R. N., Smolyar, A., Reinherz, E. L., and Wagner, G. (1995) *Science 269*, 1273−1278.

BI992786N