Eur. J. Biochem. **260**, 490–498 (1999) © FEBS 1999

Solution structure of the α-subunit of human chorionic gonadotropin

Paul J. A. Erbel[1], Yasmin Karimi-Nejad[2], Tonny De Beer[1], Rolf Boelens[2], Johannis P. Kamerling[1] and Johannes F. G. Vliegenthart[1]

[1]*Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, the Netherlands;* [2]*Bijvoet Center, Department of NMR Spectroscopy, Utrecht University, the Netherlands*

The three-dimensional solution structure of the α-subunit in the α,β heterodimeric human chorionic gonadotropin (hCG), deglycosylated with endo-β-*N*-acetylglucosaminidase-B (dg-αhCG), was determined using 2D homonuclear and 2D heteronuclear ^{1}H, ^{13}C NMR spectroscopy at natural abundance in conjunction with the program package XPLOR. The distance geometry/simulated annealing protocol was modified to allow for the efficient modelling of the cystine knot motif present in αhCG. The protein structure was modelled with 620 interproton distance restraints and the GlcNAc residue linked to Asn78 was modelled with 30 protein-carbohydrate and 3 intraresidual NOEs. The solution structure of dg-αhCG is represented by an ensemble of 27 structures. In comparison to the crystal structure of the dimer, the solution structure of free dg-αhCG exhibits: (a) an increased structural disorder (residues 33–57); (b) a different backbone conformation near Val76 and Glu77; and (c) a larger flexibility. These differences are caused by the absence of the interactions with the β-subunit. Consequently, in free dg-αhCG, compared to the intact dimer, the two hairpin loops 20–23 and 70–74 are arranged differently with respect to each other. The β-GlcNAc(78) is tightly associated with the hydrophobic protein-core in between the β-hairpins. This conclusion is based on the NOEs from the axial H1, H3, H5 atoms and the *N*-acetyl protons of β-GlcNAc(78) to the protein-core. The hydrophobic protein-core between the β-hairpins is thereby shielded from the solvent.

Keywords: chorionic gonadotropin; chorionic gonadotropin free α subunit; glycoprotein structure; cystine knot; NMR; XPLOR.

Human chorionic gonadotropin (hCG) is a placental glycoprotein hormone which acts through binding to a G-protein-coupled receptor resulting in increased adenylate cyclase activity [1–3]. The subsequent increase in cAMP levels stimulates the *corpus luteum* to produce progesterone until the placenta itself acquires the ability to produce this pregnancy sustaining steroid [4]. The hormone is a heterodimer consisting of two noncovalently associated, glycosylated subunits, α (αhCG) and β (βhCG). The α-subunit consists of 92 amino acids and is N-glycosylated at Asn52 and Asn78. Previously, site-directed mutagenesis experiments revealed that the glycan at Asn52 is relevant for biological activity [5], and recently it has been proposed that this glycan is involved in the stability of the heterodimer [6]. The glycan at Asn78 stabilizes the α-subunit [7–9]. Two crystal structures have been reported for HF-treated recombinant hCG [10,11]. A striking structural element of both subunits is the cystine knot motif comprised of three disulfide bonds.

Correspondence to J. F. G. Vliegenthart, Bijvoet Center, Department of Bio-Organic Chemistry, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, the Netherlands. Fax: +31 30 2540980. E-mail: vlieg@accu.uu.nl

Abbreviations: crmsd, coordinate root-mean-square deviation; dg-αhCG, deglycosylated α-subunit of hCG; GlcNAc, *N*-acetyl-β-D-glucosamine; $GlcNAc^{78}$, GlcNAc linked to Asn78; $GlcNAc^{52}$, GlcNAc linked to Asn52; hCG, human chorionic gonadotropin; hCG-α_f, noncombined α-subunit of hCG; αhCG, α-subunit of hCG; HSQC, heteronuclear single-quantum coherence; PNGase F, peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F.

Enzymes: Peptide-N^4-(*N*-acetyl-β-glucosaminyl)asparagine amidase F (EC 3.5.1.52)

(Received 15 October 1998, accepted 10 December 1998)

Structural analysis of intact glycoproteins by NMR spectroscopy is complicated [12], because many protons of the carbohydrate chains overlap with the Cα resonances of the amino acids. Furthermore, the presence of the extended and hydrated glycans leads to line broadening of the protein signals. In addition, the assignment of the carbohydrate resonances in intact glycoproteins is difficult due to (micro)heterogeneity of the glycans and the occurrence of more than one glycosylation site like in αhCG. Isotopic labelling of proteins is frequently used to circumvent some of these problems. However, in the case of glycoproteins that should have a human type of glycosylation pattern, isotopic labelling has to be performed in mammalian cells. This labelling technique is still in the beginning of its development [13] and its usefulness for the structure elucidation of glycoproteins has still to be shown. So far, a few NMR studies on the solution structure of intact glycoproteins have been performed [14], but the direct contacts between the glycan and the protein backbone have only rarely been identified [15,16]. The crystallization of glycoproteins is hampered by the flexibility and the heterogeneity of the glycans. Therefore, X-ray studies are usually carried out on glycoproteins with largely truncated glycans. In several cases partial deglycosylation resulted in the loss of the bioactivity [10,12].

Previously, we have described that treatment of urinary hCG-derived αhCG with endo-β-*N*-acetylglucosaminidase-B under native conditions results in the removal of the oligosaccharide chains from both Asn52 and Asn78, leaving only a single GlcNAc residue at each site [17]. An almost complete assignment of the ^{1}H NMR spectra of this deglycosylated αhCG (dg-αhCG) was achieved [18]. In the present report the solution structure of dg-αhCG is described, based on a large

number of homonuclear ^1H-^1H NOEs and using the program package XPLOR.

EXPERIMENTAL PROCEDURES

Distance restraints

2D homonuclear NOESY, 2D heteronuclear ^1H-^{13}C HSQC-NOESY and 3D homonuclear ^1H TOCSY-NOESY spectra were recorded at 328 K and pH 5.1 with a Varian UnityPlus 750 spectrometer (SON NMR Large Scale Facility, Utrecht University) as previously reported [18], and they were used to determine inter-proton distances. The majority of these inter-proton distances were estimated from NOE intensities observed in the two-dimensional NOESY spectrum acquired with a mixing time of 80 ms. The NOE intensities were calibrated to distances on the basis of interresidual distances of amino acids participating in a regular β-structure and using intraresidual distances. The NOE intensities of these intraresidual $d_{\alpha N}$ (i, i) cross-peaks were calibrated to 2.8 Å. The NOE cross-peaks of $d_{\alpha N}$ (i, i + 1) and d_{NN} (i, i + 1) from β-strands were calibrated to 2.2 and 4.3 Å, respectively. The interstrand $d_{\alpha\alpha}$ (i, j), $d_{\alpha N}$ (i, j) and d_{NN} (i, j) cross-peaks from β-strands were then calibrated to 2.3, 3.2 and 3.3 Å, respectively. NOE cross-peak intensities were estimated and classified accordingly, resulting in upper distance limits of 2.6, 3.1 and 4.0 Å. A fourth category corresponding to upper distance limits of 5.0 Å was established from the subset of additional cross-peaks observed in the NOESY spectra with a mixing time of 150 ms. All lower bounds were set to the sum of the van der Waals radii, i.e. 1.8 Å. A small number of additional NOE restraints was obtained from the 2D heteronuclear and 3D homonuclear experiments.

Appropriate multiplicity correction factors for equivalent spins were defined as given by Constantine *et al*. [19]:

$$Z = (n_I n_S)^{1/6}$$

where Z is the multiplicity correction factor, n_I is the number of equivalent spins I and n_S is the number of equivalent spins S.

The values for the pseudoatom corrections of methyl groups and aromatic H_2/H_6 or H_3/H_5 atoms were 0.4 and 2.0 Å, respectively [20,21]. In cases where both cross-peaks originating from NOEs to a non-stereoassigned diastereotopic group could be observed, the longer upperbound was used without multiplicity correction and treated as a single constraint with centre averaging. The pseudoatom corrections used for non-stereoassigned CH_2-groups and $CH(CH_3)_2$-groups were 0.88 and 1.5 Å, respectively [20]. Except for the special case of the glycosylated $CONH_2$-group of Asn78, none of the diastereotopic groups were stereospecifically assigned.

Apart from these experimental NMR data, covalent linkages for the disulfide bridges were used for the disulfide pairs known from the X-ray structure [10,11]. Accordingly, disulfide bridges are located between the cystine residues 10–60, 28–82 and 32–84, forming a cystine knot, and between the residues 7–31 and 59–87.

Structure calculation

Structure calculations were performed using the program XPLOR version 3.851 (http://xplor.csb.yale.edu). The calculation strategy was based on the hybrid distance geometry simulated annealing (DG/SA) method [22–24]. The parameters of the protocol, however, were extensively modified, and an additional refinement step was added. The force field parameters (developed by Nilges *et al*.) were downloaded from the EMBL-server (http://www.nmr.embl-heidelberg.de/nilges/). Details of the computations were as follows: An initial ensemble of 100 conformers was generated in a distance geometry step based on substructure metric matrix embedding, during which the disulfide bridges are treated as distance constraints [25]. These DG conformers were then regularised by a simulated annealing molecular dynamics run which deviates in several aspects from the standard regularisation protocol dgsa.inp: The initial temperature was raised to 3000 K, and the number of steps performed at high temperature was set to 2000. Similarly, the cooling phase was extended to 2000 steps. The time-step for the regularisation was 0.003 ps during the high temperature dynamics, and 0.005 ps during the cooling phase, which was finished at a temperature of 100 K. Covalent bonds for the disulfide bridges were gently introduced by increasing their force constant slowly during the high temperature dynamics. The increase was performed in a stepwise manner, starting with 12.5% of the equilibrium value of the force constant for the S-S bond lengths, then increasing it to 25% and 37.5% of its equilibrium value, respectively. Each of these values was applied during 250 steps of molecular dynamics. Subsequently, the S-S bond force constant was raised to 87.5% of its equilibrium value, and applied for 1000 steps of molecular dynamics. The final 200 steps of the high temperature phase were run with the full force constant for the S-S bond length. The conformers resulting from this regularisation were then submitted to further refinement. The refinement procedure comprised two steps, the first of which was another simulated annealing run of 40 ps duration, during which the system was cooled from 3000 K to 100 K in 8000 steps, followed by 500 steps of conjugate gradient energy minimization. The force constant for the S-S bond lengths was ramped up linearly from 12.5% to 100% of its equilibrium value during the cooling stage, where at the same time the scaling factor for the van der Waals radii was decreased linearly from 0.9 to 0.75 and the weight on the van der Waals energy term, C_{rep}, was increased from 0.0125 to 16.7 $kJ \cdot mol^{-1} \cdot Å^{-4}$. The final stage of the refinement comprised 10 ps of restrained molecular dynamics at 300 K, followed by 250 steps of conjugate gradient energy minimization. During all dynamics runs, the inter-proton distances were restrained to their experimentally determined values by an NOE energy term consisting of a square-well quadratic potential with a force constant k_{noe} = 209 $kJ \cdot mol^{-1} \cdot Å^{-2}$. Only the final energy minimization was carried out with k_{noe} = 313.5 $kJ \cdot mol^{-1} \cdot Å^{-2}$.

In order to select an ensemble of conformers which represents the solution structure of αhCG, the following strategy was applied: initially, models of αhCG with low conformational energies were grouped according to their cystine knot topology. The screening produced two groups and the group that exhibited significantly higher average conformational energies around the disulfide linkages was discarded. For the remaining conformers, an energy-ordered rmsd profile [26] was computed by ranking them according to their overall energy. Subsequently, the maximal pairwise 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 agreement with experimental data The mean r.m.s. deviations from idealised covalent geometry were not allowed to exceed 0.01 Å in case of bonds, 1 ° in case of angles, and 1 ° for the improper dihedrals, while the NOE violations were not allowed to exceed 0.5 Å.

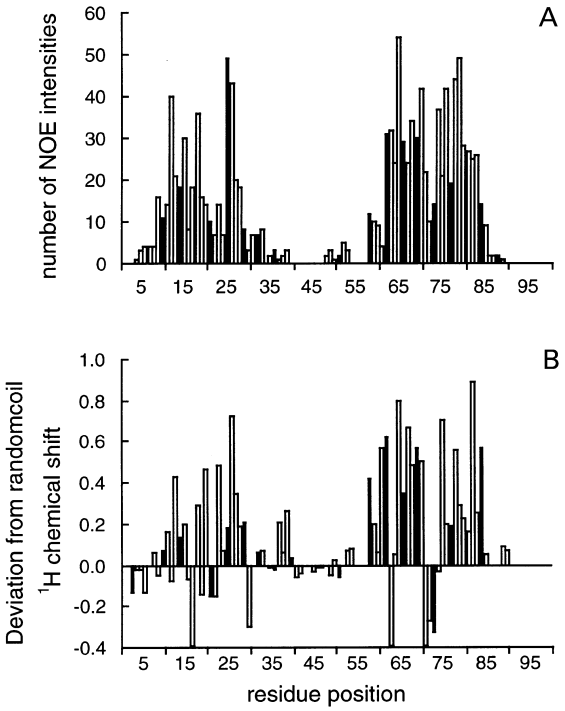

Fig. 1. The number of assigned interresidual NOE intensities (A) and the deviation from random coil chemical shift values [45] of the Cα protons (standard value – measured value) as a function of the sequence position in αhCG (B).

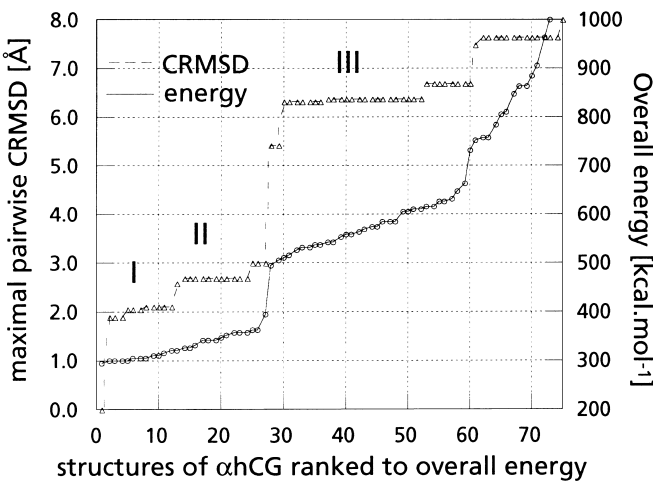

Fig. 2. Crmsd values. The maximal pairwise crmsd of all backbone atoms of residues 8–32 and 58–88, respectively, as a function of the structures, ranked to their overall energy (dotted line), is given using the first *y*-axis. The plateaux of the graph for all backbone atoms are denoted by roman numbers, indicating the corresponding structure clusters. The crmsd values were calculated after a superposition of these backbone residues. The second graph (continuous line) shows the overall energy of the individual structures, quantified on the right side of the *y*-axis.

Table 1. Structural statistics for the family of 27 αhCG conformers.

Statistic	Value
Distance restraints:	
sequential (|i-j| = 1)	259
medium-range (2 < |i-j| = 4)	81
long-range (| i-j | ⩾ 5)	280
total	620
Mean r.m.s. deviations from experimental restraints:	
NOE (Å)	0.045 ± 0.0027
Mean r.m.s. deviations from idealised covalent geometry:	
bonds (Å)	0.0019 ± 0.0004
angles (°)	0.333 ± 0.042
impropers (°)	0.218 ± 0.003
Restraint violations:	
Mean NOE violations per structure > 0.2 Å	8.63 ± 1.57
Maximum NOE violation (Å)	0.47
Mean energies (kJ·mol^{-1}):	
E_{noe}	259.33 ± 32.19
E_{vdW}[a]	317.89 ± 34.23
E_{bond}	102.54 ± 13.84
E_{angle}	498.80 ± 27.71
$E_{improper}$	75.82 ± 5.98
E_{total}	1337.98
Conformational energies of the cystine knot (kJ·mol^{-1}):	
standard DG/SA protocol	122.52
modified DG/SA protocol	43.81
Atomic r.m.s. differences (Å) [b]	average pairwise
backbone atoms 11–28 and 59–84	0.88 ± 0.27
heavy atoms 11–28 and 59–84	1.60 ± 0.35
backbone atoms 11–19, 24–28, 59–70, 80–84	0.71 ± 0.18
heavy atoms 11–19, 24–28, 59–70, 80–84	1.49 ± 0.31

[a] The parallhdg force constant parameter set of XPLOR 3.851 was used with the quartic repel core repulsion potential. No attractive van der Waals or electrostatic terms were used. [b] The segments 11–28, 59–84 and the segments 11–19, 24–28, 59–70, 80–84 correspond to the well-defined region including and excluding the conformationally heterogeneous hairpin loops, respectively.

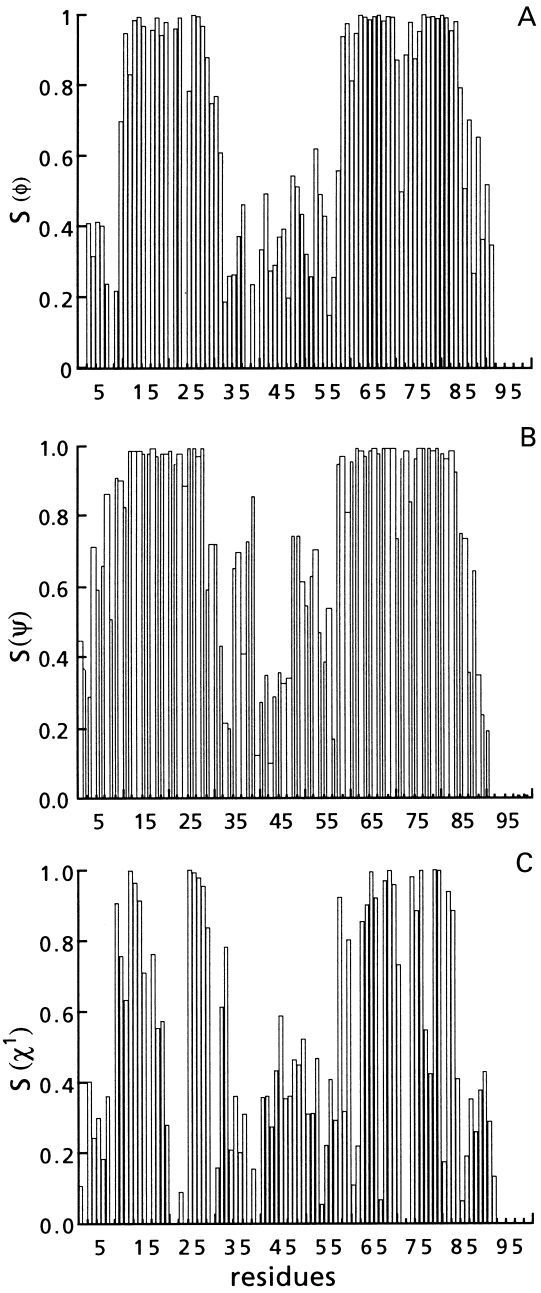

Fig. 3. Angular order parameters for the torsional angles φ (A), ψ (B) and χ^1(C) as a measure of local precision of the structure of αhCG.

The computational work was carried out on a Silicon Graphics O2. The structures were visualised using INSIGHT II (Biosym/MSI) and MOLMOL [27]. The subsequent numerical analyses and validations of the selected structures were performed using the programs AQUA, PROCHECKNMR [28], MOLMOL and XPLOR 3.851, as well as home-written programs in the MATLAB language (MATLAB 5.0, The Math Works, http://www.mathworks.com).

Modelling of GlcNAc(78)-Asn78

The NOE intensities between the GlcNAc residue linked to Asn78 and the protein core were estimated from the aforementioned NOESY spectra. The NMR restraint set for the carbohydrate moiety comprised 30 NOE constraints between the protein core and GlcNAc(78) and three short-range constraints within the GlcNAc residue. Three different αhCG conformers, reflecting the conformational space accessible to the hairpin loop comprising residues 20–23, were selected as starting conformations. The modelling of GlcNAc(78) was carried out using the INSIGHT II and DISCOVER molecular modelling packages (MSI). The AMBER force field including specific parameters for carbohydrates [29,30] was used with a dielectric constant of 4, taking into account distance dependence. During the minimization, the atoms of the amino acid residues 1–13, 27–64 and 81–92 were fixed. The residues 14–26 and 65–80 of the protein core were subsequently allowed to move, first only the side-chains (1000 steps) and then also the backbone in 1000 steps of restrained energy minimization using the steepest descent algorithm. The GlcNAc moiety was then minimized applying only intracarbohydrate distance restraints. This energy minimization was performed using the conjugate gradients algorithm and convergence was accepted when the r.m.s. gradient was less than 0.418 kJ·mol^{-1}, within a maximum of 1000 steps. Finally, 20 steps of unrestrained energy minimization were carried out with the steepest descent algorithm.

RESULTS

General

For the structure calculation of dg-αhCG in solution the program package XPLOR 3.851 was used. The experimental restraints are derived from a variety of NMR data and supplementary data from the crystal structure concerning the cystine pairing.

A total number of 620 nonredundant NOE restraints could be inferred from the NOESY spectra (Table 1). In Fig. 1A the number of interresidual NOE restraints are shown as a function of the residue position. The distribution of distance restraints over the amino acid sequence is inhomogeneous since apart from the lack of medium- and long-range restraints around the N- and C-termini of the protein, nonsequential NOE restraints are missing for the loop formed by the residues 33–57.

Previous attempts to estimate the amide proton exchange behaviour of αhCG revealed that at a temperature range between 300 K to 328 K and at pH 5.1, all amide protons of αhCG are exchanging rapidly with the solvent [8,18]. Therefore, no hydrogen-bonding restraints were used in the calculations.

The disulfide bonding pattern of αhCG, known from the two crystal structures of the intact hormone [10,11] was translated into distance restraints and applied during the distance geometry stage of the calculations. Subsequently, throughout all regularization and refinement steps, the disulfide bridges were introduced as covalent bonds (see Experimental procedures).

Structure determination and analysis

The complex pattern of the five cystine bridges of αhCG, spatially very close to each other, complicates the structure determination considerably. Our initial attempts to model the three-dimensional structure of αhCG from the NOE restraints, based on the standard DG/SA calculation protocol [22–24] implemented in the program package XPLOR, were not satisfactory. The calculations had a rather low success rate since only about 66% of the conformers had the correct cystine knot topology. Moreover, even those conformers, which possessed a proper folding around the disulfide linkages, still exhibited undesirably high conformational energies, indicating that the calculations had not converged. Given the low success rates reported for other solution structure determinations of

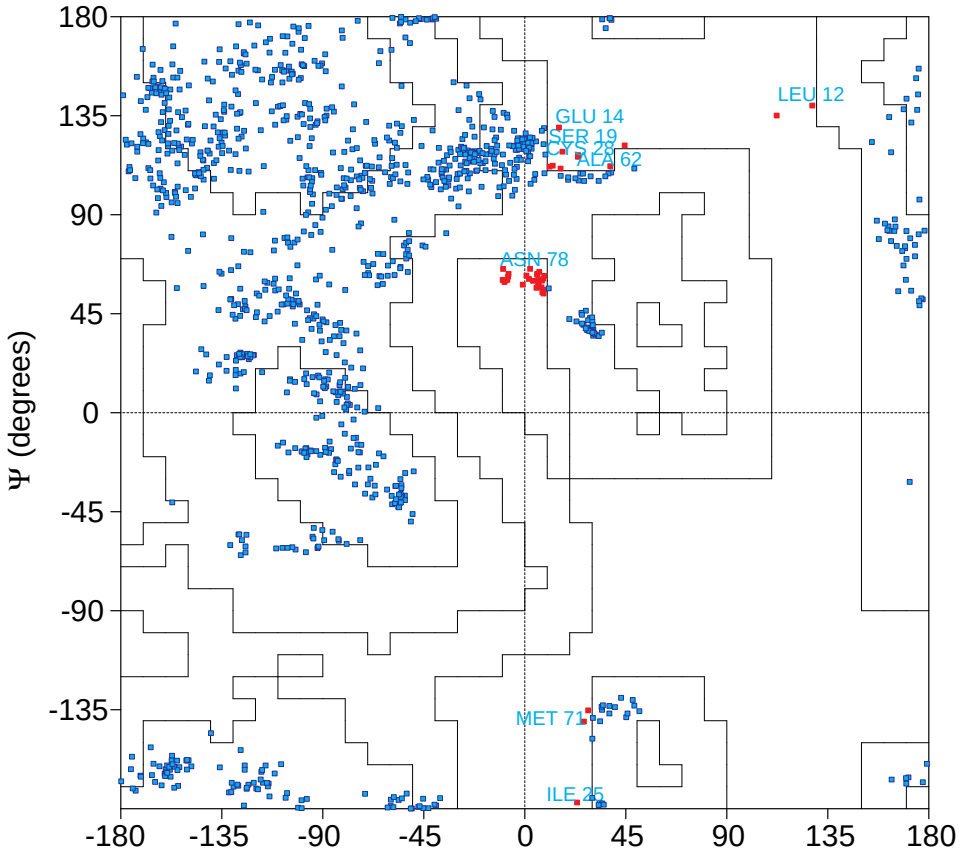

Fig. 4. Ramachandran plot for the ensemble of solution structures of αhCG. All the nonglycine residues of the segments 11–28 and 59–84 are shown.

cystine knot proteins [31] these difficulties seem to be a common problem when modelling cystine knot motifs from NMR data. To overcome the problems associated with NMR structure determinations of proteins with complex disulfide bonding topologies, we modified the DG/SA-based calculation protocol (see Experimental procedures). This modification afforded a success rate of more than 90% and also gives better geometries than the standard DG/SA protocol (Table 1), only at the cost of a slight increase in computational time.

Out of initially 100 generated conformers, an ensemble was selected to represent the solution structure of αhCG. A priori nine conformers, containing an incorrect mirror image knot topology were discarded on the basis of their significantly higher conformational energy around the knot motif. A further reduction was achieved to an ensemble of 27 conformers comprising the two distinct structural families (denoted in Fig. 2 with I and II), based on an energy-ranked crmsd profile (Fig. 2). Details of the selection criteria are given in the Experimental procedures section.

Backbone and side-chain angular order parameters [32,33] of the solution structure of dg-αhCG, giving a measure of its local precision, are shown in Fig. 3A,B,C. It is obvious from Fig. 3 that the degree of structural definition within the protein is inhomogeneous. The disorder present in the segment comprising residues 33–57 reflects the absence of medium- and long-range NOEs. This is in good agreement with the random coil ^{1}H chemical shifts of the corresponding residues (Fig. 1B). In contrast, the segments comprising residues 11–28 and 59–84 represent the well-defined parts of the solution structure of αhCG. The angular order parameters of these segments S(ϕ) and S(ψ) are ⩾ 0.8, and the average pairwise rmsd values are 0.88 Å for the backbone atoms and 1.60 Å for the heavy atoms. If residues 20–23 and 71–79, located in conformationally heterogeneous hairpin loops, are excluded, the rmsd values for backbone and heavy atoms drop to 0.71 Å and 1.49 Å, respectively.

Quality of the structure

The individual backbone conformations of all nonglycine residues in the defined regions (11–28 and 59–84) are presented in the Ramachandran map shown in Fig. 4. According to analysis with the program PROCHECK-NMR [28], 34% of these residues are located in most-favoured regions and 62.1% in allowed regions of the ϕ/ψ space. In the disallowed region 3.9% of the amino acid residues were found, comprising almost all backbone conformations of the glycosylated Asn78 (GlcNAc(78)). The last feature is probably caused by the interaction of the GlcNAc residue attached to Asn78 with the surrounding protein core.

Three-dimensional structure of dg-αhCG in solution

The nonglobular folding of dg-αhCG is highly unusual. The well-defined core formed by residues 11–28 and 59–84 consists of two twisted β hairpins linked by three disulfide bridges (Cys28 : Cys82, Cys10 : Cys60 and Cys32 : Cys84) forming a cystine knot (Fig. 5) [34]. The two hairpins are connected by a disordered loop consisting of the residues 33–57. The N- and C-terminal segments, comprising residues 1–10 and 85–92, respectively, are also disordered and stick out into the solution like two arms. The segments are connected to

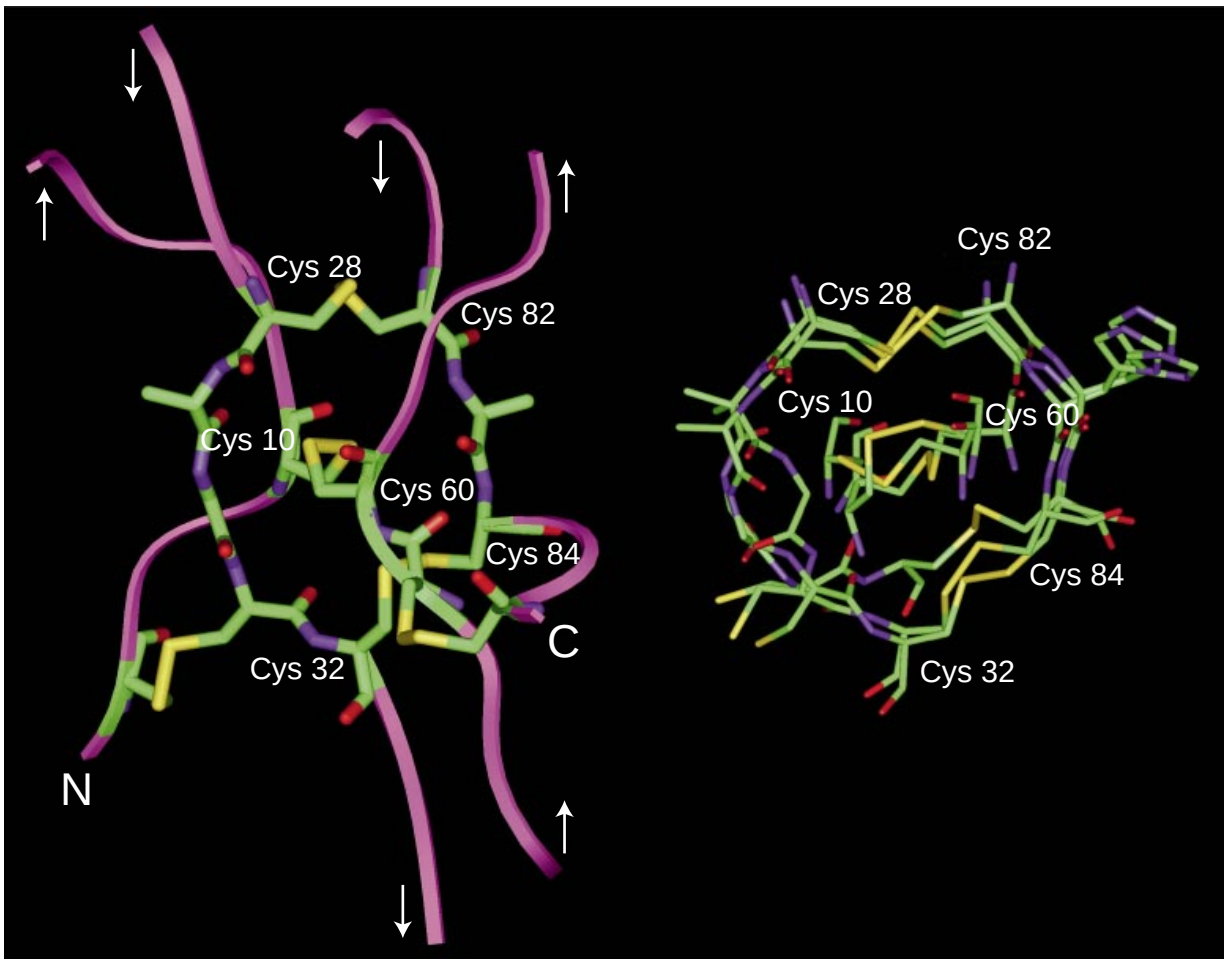

Fig. 5. Representation of the cystine knot motif of αhCG. Three representative NMR structures (right), crystal structure of hCG (left). The backbone strands are indicated in purple. Only the heavy atoms of the cystine knot are displayed.

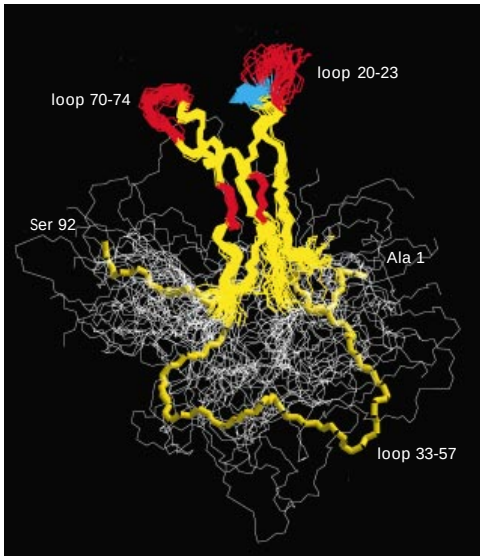

Fig. 6. Representation of the backbone conformation of the 27 conformers of αhCG. The superposition was carried out including the backbone atoms of the residues 8–32 and 58–85. The residues outside this region indicated in white represent the flexible and undefined part of αhCG, except one conformer which is represented in a yellow ribbon. The β-sheet is drawn in yellow and the 3_{10}-helix in cyan. The β-bulges, the tight turn and the hairpin loop are coloured in red.

the protein-core by two disulfide bridges (Cys7 : Cys31 and Cys59 : Cys87) in a hinge-like manner. In Fig. 6 the protein backbone representation of the solution structure of dg-αhCG is shown. A detailed investigation of the two β-hairpins constituting the core of αhCG reveals that the first one consists of two extended segments including residues 11–14 and 24–28. The segments 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, which are connected by a hairpin loop formed by residues 70–74. The two opposite strands of the second hairpin exhibit irregularities in their hydrogen bonding pattern due to a β-bulge located at residue Lys63 and two interlocked β-turns involving residues 76–82. Moreover, neither the first nor the second β hairpin possesses a complete hydrogen bonding network as judged by analysis of the consensus hydrogen bonds present in the solution structure of dg-αhCG (Table 2).

Analysis of the cystine side chain conformations reveals that in most cases the disulfide bridges exhibit conformational heterogeneity (Fig. 5). Possibly, the overlap often occurring in homonuclear NMR spectra of cystine knot proteins [35–37] and consequently the scarcity of unambiguously assigned NOEs for the spatially close cystine side chains might have caused this conformational variability. On the other hand these results are in agreement with the NMR spectra, namely the resonance of the Cys59 shows extensive line broadening, while the signals that

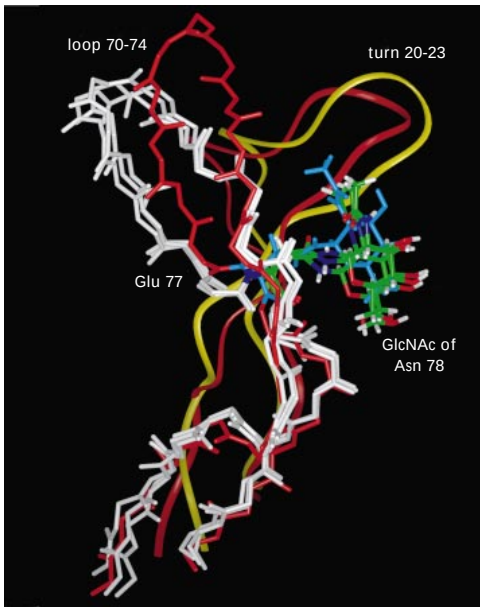

Fig. 7. Overlay of the backbone atoms of the β-sheets of the NMR structure onto the crystal structure. In the X-ray structure the protein-core and the GlcNAc(78) residue are coloured in red and blue, respectively. The β-sheet of three representative NMR structures comprising GlcNAc(78) is displayed in white, while the second β-sheet is indicated with a yellow ribbon.

should stem from Cys7 and Cys87 are not observed. These features suggest that these residues, linking the flexible ends and disordered loop, might be involved in a conformational exchange process. This is possibly due to slow rotational isomerism of the disulfide bridges, as has been observed previously in disulfide bridges of small protease inhibitors such as BPTI [38].

The conformational properties of the GlcNAc residues of dg-αhCG linked to Asn52 and Asn78 are different. Previously, we reported that the glycan chain at Asn52 is flexible and solvent exposed and does not influence the structure of free αhCG [8,17]. However, the spatial position of GlcNAc(78) is well-defined due to 30 NOEs between the protein-core and this monosaccharide (Fig. 7). The majority of these NOEs are observed between the hydrophobic residues, such as Leu12, Ile25, Val68, Val76 and the axial H1, H3 and H5 and the *N*-acetyl protons of β-GlcNAc. Through this orientation GlcNAc(78) shields the hydrophobic core in between the β hairpins from the solvent.

Table 2. Hydrogen bonds present within the protein moiety of dg-αhCG in solution. Hydrogen bonds were identified with MOLMOL if the distance between donor and acceptor was < 2.4 Å, and the corresponding angle $< 35°$ in $> 50\%$ of the conformers constituting the solution structure ensemble of dg-αhCG. Only the backbone atoms are taken into account.

Donor	Acceptor	Donor	Acceptor
Asn15	Ile25	Asn66	Glu77
Gln20	Ala23	Gly73	Val70
Gly22	Gln20	Phe74	Val70
Ala23	Gln20	His79	Val76
Met29	Thr11	Thr80	Ser64
Ser64	Ala81	Ser85	Cys59

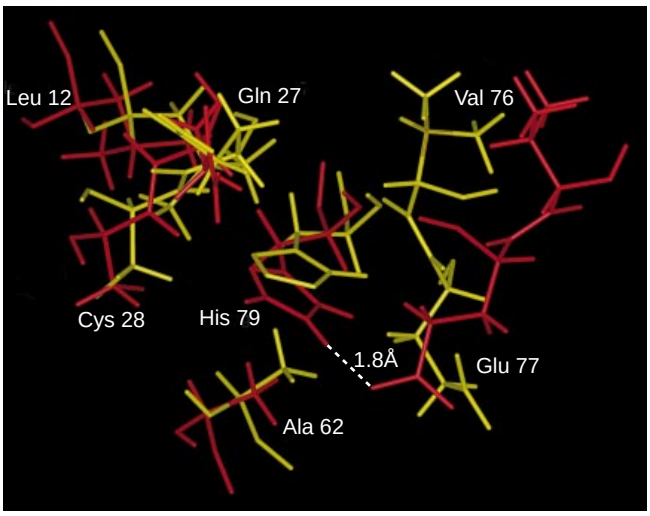

Fig. 8. Superposition of the microenvironment of His79 in αhCG in an average NMR structure (yellow) and the X-ray structure of hCG (red). In earlier studies Parsons & Pierce [41] have reported that Hε2 of His79 in αhCG is not exchangeable in ^2H_2O at pH 5, whereas this proton becomes exchangeable upon binding to the β-subunit. The small deviation concerning the negatively charged side chain of Glu77 and the orientation of the hydrophobic side chain of Val76 can support an increase of exchange rate in intact hCG.

Accession numbers

Coordinates of the NMR structures of endo-β-*N*-acetylglucosaminidase-B deglycosylated αhCG isolated from urinary hCG will be deposited in the Protein Data Bank, Brookhaven National Laboratory, USA.

DISCUSSION

Comparison of the solution structure of free αhCG with the crystal structure of recombinant HF-dg-αhCG in the intact hormone

Crystallographic studies of heterodimeric HF-treated hCG (N-glycans mainly trimmed to GlcNAc-GlcNAc) revealed a tight association of the α-and β-subunits. At the heart of the subunit interface a short seven-stranded β-barrel is formed, consisting of five β-strands of the β-subunit and two β-strands of the α-subunit, comprising residues α30–37 and α53–57. Apart from this central region hydrogen bonds are found in a short stretch of an antiparallel β-sheet between the residues β44–46 and α77–75 and hydrophobic contacts are observed between valine and leucine at β44–45 and the triplet of phenylalanine residues (α17,18,74) [10,11].

From Fig. 6 it is obvious that free dg-αhCG is a rather flexible molecule. This is reflected by the disorder we observed for the loop comprising residues α33–57, and also by the conformational heterogeneity of the hairpin loop α70–74 and the tight turn α20–23. Indications for the flexible nature of the core moiety of free αhCG, as is suggested by Wu *et al.* [11], can be inferred from its amide exchange behaviour. It was observed that all amide protons exchange at a temperature of 300 K with the solvent within 30 min [8]. This observation is in good agreement with the relatively incomplete hydrogen bonding network identified in the solution structure of dg-αhCG (Table 2). We can conclude that the increased flexibility of free dg-αhCG in comparison to the dimer, is due to the

absence of interactions with the β-subunit. Although interpretation of this observation in terms of biological functioning is beyond the scope of this study, it should be mentioned that other investigations have shown [5,39,40] that this loop of the α-subunit is essential for receptor recognition and binding. Since the primary structure of the α-subunit has been conserved throughout the glycohormone family of CG, FSH and LH, we suggest that in each of the hormones the structure of this loop of the α-subunit is induced by the specific β-subunits. Taking into account that the biological function of LH and CG are similar and that both glycohormones are recognised by the same receptor, it is tempting to hypothesise that the 3D structure of this loop in the α-subunit of intact LH is similar to that in the X-ray structure of αhCG in the complex with βhCG, but different from the α-subunit in intact FSH.

It has been reported that the glycan beyond the first GlcNAc(78) may have a small effect on the local tertiary structure of αhCG, also causing an increase in the thermal stability [8,18]. On the basis of the solution structure of dg-αhCG the latter feature can be explained by a stabilising effect of this glycan on the heterogeneous tight turn α20–23 and the heterogeneous hairpin loop α70–74.

A further difference between αhCG in the X-ray structure of HF-treated hCG and dg-αhCG is the orientation of the segments comprising the residues of the tight turn α20–23 and the residues of the hairpin loop α70–74 (Fig. 7). This is mainly the result of a different conformation around His79 (Fig. 8). The pK_a value of this His79 in the intact hormone and the free α-subunit has been determined previously to be 6.6 and 2.5, respectively [41]. In the solution structure of the free α-subunit the environment of His79 is less polar than in the X-ray structure of the intact hormone, which is due to different orientations of the side chains of Glu77 and Val76. Moreover, in intact hCG the orientation of the side chain of Glu77 allows for the formation of a hydrogen bond with the Hε2 atoms in the imidazole ring of His79. This is in agreement with the suggestion that the microenvironment of His79 in the α-subunit changes upon the formation of the hCG dimer [41].

A model for the noncombined α-subunit of hCG (hCG-α_f)

Besides the role of αhCG 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, is able to stimulate prolactin secretion from human decidual cells [42,43]. Although the carbohydrate composition of hCG-α_f partially differs from αhCG [43,44], the α-subunit derived through dissociation of hCG is also able to stimulate prolactin secretion cells [42]. This observation strongly indicates that the solution structure of dg-αhCG is almost identical to that of fully glycosylated hCG-α_f. It should be noted that neither the glycan at Asn52 nor the glycan chain beyond GlcNAc(78) influences the structure of the protein-core of αhCG.

ACKNOWLEDGEMENTS

This investigation was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO). The crude preparation of hCG was kindly provided by 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 Florence Casset for her help with the modelling of the carbohydrate part of dg-αhCG.

REFERENCES

1. Pierce, J.G. & Parsons, T.F. (1981) Glycoprotein hormones: structure and function. *Annu. Rev. Biochem.* **50**, 465–495.
2. McFarland, K.C., Sprengel, R., Philips, H.S., Kohler, M., Rosemblit, N., Nikolics, K., Segaloff, D.L. & Seeburg, P.H. (1989) Lutropin-choriogonadotropin receptor: an unusual member of the G protein-coupled receptor family. *Science* **245**, 494–499.
3. Taylor, C.W. (1990) The role of glycoproteins in transmembrane signalling. *Biochem. J.* **272**, 1–13.
4. Talwar, G.P. (1979) Human chorionic gonadotropin and ovarian and placental steroidogenesis. *J. Steroid Biochem.* **11**, 27–34.
5. Matzuk, M.M., Keene, J.L. & Boime, I. (1989) Site specificity of the chorionic gonadotropin N-linked oligosaccharides in signal transduction. *J. Biol. Chem.* **264**, 2409–2414.
6. Heikoop, J.C., Van den Boogaart, P., De Leeuw, R., Rose, U.M., Mulders, J.W.M. & Grootenhuis, P.D.J. (1998) Partially deglycosylated human choriogonadotropin, stabilized by intersubunit disulfide bonds, shows full bioactivity. *Eur. J. Biochem.* **253**, 354–356.
7. Matzuk, M.M. & Boime, I. (1988) The role of asparagine-linked oligosaccharides of the α subunit in the secretion and assembly of human chorionic gonadotropin. *J. Cell. Biol.* **106**, 1049–1059.
8. Van Zuylen, C.W.E.M., Kamerling, J.P. & Vliegenthart, J.F.G. (1997) Glycosylation beyond the Asn78-linked GlcNAc residue has a significant enhancing effect on the stability of the α subunit of human chorionic gonadotropin. *Biochem. Biophys. Res. Commun.* **232**, 117–120.
9. Thijssen-van Zuylen, C.W.E.M., De Beer, T., Leeflang, B.R., Boelens, R., Kaptein, R., Kamerling, J.P. & Vliegenthart, J.F.G. (1998) Mobilities of the inner three core residues and the Man (α1->6) branch of the glycan at Asn78 of the α-subunit of human chorionic gonadotropin are restricted by the protein. *Biochemistry* **37**, 1933–1940.
10. Lapthorn, A.J., Harris, D.C., Littlejohn, A., Lustbader, J.W., Canfield, R.E., Machin, K.J., Morgan, F.J. & Isaacs, N.W. (1994) Crystal structure of human chorionic gonadotropin. *Nature* **369**, 455–461.
11. Wu, H., Lustbader, J.W., Liu, Y., Canfield, R.E. & Hendrickson, W.A. (1994) Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein. *Structure* **2**, 545–558.
12. Lustbader, J.W., Pollak, S., Lobel, L., Trakht, I., Homans, S., Brown, J.M. & Canfield, R.E. (1996) Three-dimensional structures of gonadotropins. *Mol. Endocrinol.* **125**, 21–31.
13. Lustbader, J.W., Birken, S., Pollak, S., Pound, A., Chait, B.T., Mirza, U.A., Ramnarian, S., Canfield, R.E. & Brown, J.M. (1996) Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: an advance toward rapid determination of glycoprotein structures. *J. Biomol. NMR* **7**, 295–304.
14. Fletcher, C.M., Harrison, R.A., Lachmann, P.J. & Neuhaus, D. (1994) Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure* **2**, 185–199.
15. Wyss, D.F., Choi, J.S., Li, J., Knoppers, M.H., Willis, K.J., Arulanadam, A.R.N., Smolyar, A., Reinherz, E.L. & Wagner, G. (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* **269**, 1273–1278.
16. Metzler, W.J., Bajorath, J., Fenderson, W., Shaw, S., Constantin, K., Naemura, J., Leytze, G., Peach, R.J., Lavoie, T.B., Mueller, L. & Linsley, P.S. (1997) Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved CD28. *Nature Struct. Biol.* **4**, 527–531.
17. Van Zuylen, C.W.E.M., De Beer, T., Rademaker, G.J., Haverkamp, J., Thomas-Oates, J.E.H., Hård, K., Kamerling, J.P. & Vliegenthart, J.F.G. (1995) Site-specific and complete enzymic deglycosylation of the native human chorionic gonadotropin α-subunit. *Eur. J. Biochem.* **231**, 754–760.
18. De Beer, T., Van Zuylen, C.W.E.M., Leeflang, B.R., Hård, K., Boelens, R., Kaptein, R., Kamerling, J.P. & Vliegenthart, J.F.G. (1996) NMR studies of the free α subunit of human chorionic gonadotropin; structural influences of N-glycosylation and the α subunit on the conformation of the α subunit. *Eur. J. Biochem.* **241**, 229–242.

19. Constantine, K.L., Madrid, M., Bányai, L., Trexler, M., Patthy, L. & Llinás, M. (1992) Refined solution structure and ligand-binding properties of Pdc-109 domain-B–A collagen-binding type-II domain. *J. Mol. Biol.* **223**, 281–298.
20. Wüthrich, K., Billeter, M. & Braun, W. (1983) Pseudo-structures for the 20 common amino- acids for use in studies of protein conformation by measurement of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.
21. Koning, T.M.G., Boelens, R. & Kaptein, R. (1990) Calculation of the nuclear Overhauser effect and the determination of proton-proton distances in the presence of internal motion. *J. Magn. Reson.* **90**, 111–123.
22. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1988) Determination of three- dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical stimulated annealing calculations. *FEBS Lett.* **229**, 317–324.
23. Nilges, M., Kuszewski, J. & Brünger, A.T. (1991) Sampling properties of simulated annealing and distance geometry. In *Computational Aspects of the Study of Biological Macromolecules by NMR* (Hoch, J.C., ed.), pp. 451–455. Plenum Press, New York.
24. Kuszewski, J., Nilges, M. & Brünger, A.T. (1992) Sampling and efficiency of metric matrix distance geometry: a novel partial metrization algorithm. *J. Biomol. NMR* **2**, 33–56.
25. Brünger, A.T. (1992) *XPLOR, Version 3.1. A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT.
26. Widmer, H., Widmer, A. & Braun, W. (1993) Extensive distance geometry calculations with different NOE calibrations: New criteria for structure selection applied to sandostatin and BPTI. *J. Biomol. NMR* **3**, 307–324.
27. Koradi, R., Billeter, M. & Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–55.
28. Laskowski, R.A., Rullman, J.A.C., MacArthur, M.W., Kaptein, R. & Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486.
29. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986) An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**, 230–252.
30. Homans, S.W. (1990) A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Manα1–3Manβ1–4GlcNAc. *Biochemistry* **29**, 9110.
31. Fletcher, J.I., Smith, R., O'Donoghue, S.I., Nilges, M., Connor, M., Howden, M.E.H., Christie, M.J. & King, G.F. (1997) The structure of a novel insecticidal neurotoxin, ω- atracotoxin-HV1, from the venom of an Australian funnel web spider. *Nature Struct. Biol.* **4**, 559–566.
32. Hyberts, S.G., Goldberg, M.S., Havel, T.F. & Wagner, G. (1992) The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.
33. Markley, J.L., Bax, A., Arata, Y., Hilberts, C., Kaptein, R., Sykes, B.D., Wright, P.E. & Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Biomol. NMR* **12**, 1–23.
34. Sun, P.D. & Davies, D.R. (1995) The cystine-knot growth-factor superfamily. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 269–291.
35. De Beer, T., Van Zuylen, C.W.E.M., Leeflang, B.R., Hård, K., Boelens, R., Kaptein, R., Kamerling, J.P. & Vliegenthart, J.F.G. (1994) Rapid and simple approach for the NMR resonance assignment of the carbohydrate chains of an intact glycoprotein; application of gradient-enhanced natural abundance $^{1}H–^{13}C$ and HSQC-TOCSY to the α-subunit of human chorionic gonadotropin. *FEBS Lett.* **348**, 1–6.
36. Campos-Olivas, R., Bruix, M., Santoro, J., Lacadena, J., Martinez del Pozo, A., Gavilanes, J.G. & Rico, M. (1995) NMR solution structure of the antifungal protein from *Aspergillus giganteus*: Evidence for cysteine pairing isomerism. *Biochemistry* **34**, 3009–3021.
37. Vervoort, J., van den Hooven, H.W., Berg, A., Vossen, P., Vogelsang, R., Joosten, M.H.A.J. & de Wit, P.J.G.M. (1997) The race-specific elicitor AVR9 of the tomato pathogen *Cladosporium fulvum*: a cystine knot protein. *FEBS Lett.* **404**, 153–158.
38. Otting, G., Liepiush, E. & Wüthrich, K. (1993) Disulfide bound isomerization in BPTI and BPTI (G365), a NMR study of correlated mobility in proteins. *Biochemistry* **32**, 3571–3582.
39. Moyle, W.R., Campbell, R.K., Rao, S.N.V., Ayad, N.G., Bernard, M.P., Han, Y. & Wang, Y. (1995) Model of human chorionic gonadotropin and lutropin receptor interaction that explains signal transduction of the glycoprotein hormones. *J. Biol. Chem.* **270**, 20020–20031.
40. Moyle, W.R., Myers, R.V., Wang, Y., Han, Y., Lin, W., Kelley, G.L., Ehrlich, P.H., Rao, S.N.V. & Bernard, M.P. (1998) Functional homodimeric hormones: implications for hormone action and evolution. *Chem. Biol.* **5**, 241–254.
41. Parsons, T.F. & Pierce, J.G. (1983) Studies of histidine residues of human and bovine glycoprotein hormones by nuclear magnetic resonance. *Int. J. Peptide Protein Res.* **21**, 522–535.
42. Blithe, D.L., Richards, R.G. & Skarulis, M.C. (1991) Free alpha molecules from pregnancy stimulate secretion of prolactin from human decidual cells: a novel function for free alpha in pregnancy. *Endocrinology* **129**, 2257–2259.
43. Blithe, D.L. & Iles, R.K. (1995) The role of glycosylation in regulating the glycoprotein hormone free α-subunit endocrinology and free β-subunit combination in the extraembryonic coelomic fluid of early pregnancy. *Endocrinology* **136**, 903–910.
44. Kawano, T., Endo, T., Nishimura, R., Mizuochi, T., Mochizuki, M., Kochibe, N. & Kobata, A. (1988) Structural differences found in the sugar chains of eutopic and ectopic free α-subunits of human glycoprotein hormone. *Arch. Biochem. Biophys.* **267**, 787–796.
45. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons Inc., New York.