New Type of Linkage between a Carbohydrate and a Protein: *C*-Glycosylation of a Specific Tryptophan Residue in Human RNase U_s[†]

Jan Hofsteenge,[*,‡] Dieter R. Müller,[§] Tonny de Beer,[||] Andreas Löffler,[‡] Wilhelm J. Richter,[§] and Johannes F. G. Vliegenthart[||]

Friedrich Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland, Research Services, Ciba-Geigy Ltd., CH-4002, Basel, Switzerland, and Bijvoet Center, Department of Bio-organic Chemistry, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, The Netherlands

Received August 15, 1994; Revised Manuscript Received September 28, 1994[⊗]

ABSTRACT: We report a new type of linkage between a carbohydrate and a protein, involving the rarely modified side chain of a tryptophan residue. An aldohexopyranosyl residue was found to be linked via a C−C bond to the indole ring of the tryptophan residue at position 7 of human RNase U_s. Mass spectrometric analysis of peptides containing this residue showed a molecular mass 162 Da higher than that expected for tryptophan. The fragmentation pattern of the modified amino acid side chain was reminiscent of that of aromatic *C*-glycosides, suggesting a direct attachment of a hexose residue to a C-position of the tryptophan indole moiety. ^1H and ^{13}C NMR spectroscopic data confirmed this inference and unequivocally demonstrated the substituent to be an aldohexopyranosyl residue, *C*-glycosidically linked to the C2 atom of the indole. This mode of attachment differs from the ones known so far, in which carbohydrates are linked to an amino acid side chain by *N*- or *O*-glycosidic bonds.

A renewed interest in RNases of the pancreatic type was stimulated by the finding that some of them, have potent physiological effects, other than digestion of RNA in food (D'Alessio, 1993). For instance, eosinophil derived neurotoxin (EDN)[1] was originally identified by its neurotoxic effect after injection into the cerebellum (Durack et al., 1981). This protein was subsequently shown to contain RNase activity and to be structurally homologous to bovine pancreatic RNase A (Gleich et al., 1986; Slifman et al., 1986). Its amino acid sequence, deduced from the DNA sequence (Hamann et al., 1990), is identical to the chemically determined sequence of RNase U_s found in human urine (Beintema et al., 1988), except for the amino acid residue at position 7. Whereas a tryptophan residue was predicted from the DNA sequence of EDN, this amino acid could not be identified by chemical means in RNase U_s (Beintema et al., 1988) or in EDN (Gleich et al., 1986). This suggests a posttranslational modification of this residue.

Two types of posttranslationally modified tryptophan residues in proteins have been reported so far. The bacterial enzyme methylamine dehydrogenase contains the unusual redox cofactor tryptophan tryptophylquinone, which is thought to be the product of cross-linking of two tryptophan residues (McIntire et al., 1991). Another modification has recently been reported in a neuropeptide isolated from the stick insect *Carausius morosus*. Based on mass spectrometric data, it was suggested that a hexose is attached to the N1 atom of the indole ring of the tryptophan (Gäde et al., 1992).

Here we report the modification of Trp-7 of human RNase U_s by the *C*-glycosidic attachment of an aldohexopyranosyl residue to the C2 atom of the indole ring.

MATERIALS AND METHODS

Protein Purification and Protein Chemistry. RNase U_s was purified by chromatography on SP-Sephadex at pH 3.0 (Iwama et al., 1981), heparin Sepharose (Reimert et al., 1991), and SP-Sephadex at pH 7.5 (Iwama et al., 1981). Final purification was achieved by reversed-phase HPLC, using a C_4 column (Vydac, Hispania, CA) equilibrated in 0.1% trifluoroacetic acid. Protein was eluted with a linear gradient of 10.5−56% CH_3CN in 80 min, at a flow rate of 1 mL/min. Its N-terminal amino acid sequence was in complete agreement with the previously published primary structure (Beintema et al., 1988).

Peptides from the N-terminal region of RNase U_s were obtained by digestion of reduced and carboxymethylated (Hofsteenge et al., 1991) or sulfitolyzed (Chan, 1968) protein with the protease from *Staphylococcus aureus* (2% w/w; 18 h), thermolysin (3% w/w; 18 h), or elastase (5% w/w; 4 h). Digestion of peptide 5−8 with aminopeptidase M (10 units/nmol of peptide; 2 h) yielded the tripeptide 6−8. All digestions were performed in 50 mM NH_4HCO_3, pH 7.8, at 37 °C. Peptides were purified on a C_{18} reversed-phase column as described (Hofsteenge et al., 1991). The phenylthiohydantoin-derivative (PTH-derivative)[1] of modified

[†] Part of this investigation was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO). A.L. is supported by a grant from the Deutsche Forschungsgemeinschaft.

[*] To whom correspondence should be addressed. Telephone: (061) 697 47 22. Fax: (061) 697 39 76.

[‡] Friedrich Miescher-Institut.

[§] Ciba-Geigy Ltd.

[||] Utrecht University.

[⊗] Abstract published in *Advance ACS Abstracts*, November 1, 1994.

[1] Abbreviations: CID, collision induced dissociation; 1D, one dimensional; 2D, two dimensional; DCI, desorption chemical ionization; EDN, eosinophil derived neurotoxin; EI, electron impact ionization; ESI, electrospray ionization; HSQC, ^1H-detected heteronuclear single-quantum coherence spectroscopy; MLEV, composite pulse devised by M. Levitt; PTH, phenylthiohydantoin; ROESY, rotating-frame nuclear Overhauser spectroscopy; TOCSY, total correlation spectroscopy; TPPI, time-proportional phase increment.

0006-2960/94/0433-13524$04.50/0 © 1994 American Chemical Society

Trp-7 was prepared by manual Edman degradation (Tarr, 1977) and purified by reversed-phase HPLC on a 5-μm Partisil ODS 3 column (4.6 × 250 mm) equilibrated in 20 mM NH_4-acetate, pH 4.5, with a gradient of 7−63% CH_3CN over 60 min in the same buffer, at a flow rate of 1 mL/min. Two PTH-derivatives were obtained that corresponded to X_1 and X_2 in Figure 1.

Methods for automated Edman degradation, identification of PTH amino acids, and amino acid composition analysis have been described previously (Hofsteenge et al., 1991).

Peptides were synthesized using the solid-phase method and N^{α}-fluorenyl-9-methoxycarbonyl-protected amino acids.

Mass Spectrometry. Electrospray ionization MS (ESI-MS) and ESI-MS/MS measurements were carried out in the positive ion mode using a PE Sciex API III triple quadrupole mass spectrometer fitted with an articulated, pneumatically assisted nebulization probe. For the measurements carried out in the electron impact (EI) or desorption chemical ionization (DCI) mode, a Finnigan MAT TSQ 70 triple stage quadrupole mass spectrometer was used. In the DCI mode deuterated ammonia gas (N^2H_3, Merck, Sharp & Dohme, Montreal, ON) served as a reagent gas. For collision induced dissociation (CID) of mass selected ions in ESI-MS/MS experiments, argon was used as a target gas at a density of 5.5×10^{14} molecules/cm^2, and also for the EI-MS/MS and DCI-MS/MS experiments conducted at a gas pressure of 0.5 mTorr. Collision energies were as given in the figure legends.

Ultraviolet and NMR Spectroscopy. Ultraviolet absorption spectra were measured on a Hewlett Packard 8452A spectrophotometer. The peptide concentrations used to calculate molar extinction coefficients were determined by amino acid analysis. Emission fluorescence spectra were recorded on a Perkin Elmer LS-3 fluorescence spectrophotometer, at an excitation wavelength of 280 nm.

NMR spectra were recorded at 295 or 300 K using a Bruker AMX-500 or AMX-600 spectrometer (Bijvoet Center, Department of NMR Spectroscopy, Utrecht University or NSR Center, University of Nijmegen, The Netherlands). Samples were dissolved in 95% (v/v) $H_2O/^2H_2O$, containing 5 mM potassium phosphate, pH 5.4, in ^2H_2O (99.99% ^2H_2O, MSD Isotopes) or in DMSO-d_6 (Merck). The ^1H carrier frequency was placed on the water resonance, unless otherwise indicated. Suppression of the water resonance was achieved either by a water eliminated Fourier transform-pulse (WEFT-pulse) sequence (1D measurements on the ^2H_2O sample) (Hård et al., 1992) or by low-power water irradiation during the relaxation delay. ^1H NMR spectra were recorded with a spectral width of 5400−8500 Hz, depending on solvent and magnetic field strength. For 1D ^1H NMR spectra 16−512 free induction decays were collected of 8K or 16K complex data points. 2D experiments were recorded with 256−512 experiments in the t_1 dimension, and 8−240 free induction decays of 2048 data points were collected per experiment. Quadrature detection in the t_1 dimension was achieved by either the time-proportional phase increment (TPPI) method (Marion & Wüthrich, 1983) or the States−TPPI method (Marion et al., 1989). 2D ^1H-TOCSY spectra were recorded using a "clean" MLEV-17 TOCSY transfer step of 20, 50, or 60 ms (Braunschweiler & Ernst, 1983; Bax & Davis, 1985a; Griesinger et al., 1988; Hård et al., 1992), preceded by a trim pulse of 2.5 ms. 2D ^1H-ROESY spectra (Bax & Davis, 1985b) were recorded with a ROESY spin-lock pulse of 300 ms, and the ^1H carrier frequency was placed at about 6 ppm. The 2D gradient-enhanced natural abundance $^1H-^{13}C$ HSQC spectrum was recorded essentially as described (de Beer et al., 1994). NMR data sets were processed using the Bruker UXNMR software package as described previously (Hård et al., 1992; de Beer et al., 1994).

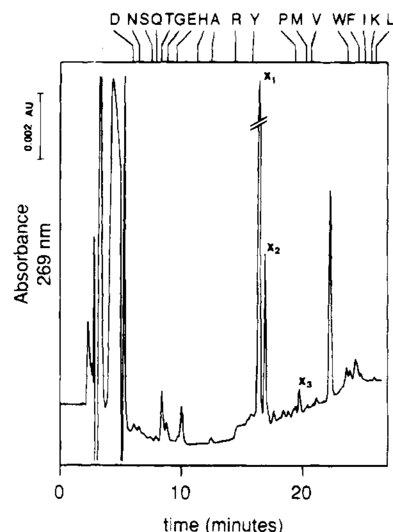


FIGURE 1: Reversed-phase HPLC of the PTH-derivative of the amino acid residue at position 7 of human RNase U_s. RNase U_s was degraded by automated Edman degradation, and the PTH-derivatives were separated by reversed-phase HPLC as described under Materials and Methods. The unknown derivatives are indicated with X_{1-3}. The elution position of the standard PTH-derivatives has been indicated at the top, using the one-letter code.

RESULTS

Protein Chemistry. Edman degradation of native RNase U_s failed to reveal the identity of the amino acid residue at position 7, in agreement with previously reported results (Beintema et al., 1988). At the corresponding cycle one major and two minor PTH-derivatives were observed (Figure 1), none of which coincided with that of any of the known amino acids. The area of the major peak (X_1) comprised 80% of the sum of the three peaks, but was considerably lower than that of, e.g., the stable amino acid Phe at position 5. X_2 and X_3 comprised 18% and 2%, respectively. X_2 was concluded to be a diastereomer of X_1, since it had the same ultraviolet and EI-MS spectrum.[2] Four N-terminal peptides were analyzed by Edman degradation and found to yield the same three PTH-derivatives as the parent protein, at the residue corresponding to position 7 (Table 1). These results indicated that the procedures used for cleavage and purification of the peptides had not altered the residue at this position. The ultraviolet spectrum of the tetrapeptide F-T-X-A (**4**, Table 1) has a maximum absorbance at 282 nm and a distinct shoulder at 290 nm. The molar extinction coefficient of this peptide ($\epsilon_{282} = 9017\ M^{-1}\ cm^{-1}$) is about 1.5-fold higher than that of the synthetic tetrapeptide F-T-W-A ($\epsilon_{280} = 5869\ M^{-1}\ cm^{-1}$). The fluorescence emission spectrum of **4** is very similar to that of F-T-W-A, but its maximum has shifted 2

[2] Most likely X_2 results from racemization of the C^{α} atom during Edman degradation. This reaction will only lead to diastereomers with amino acid residues containing at least one asymmetric C atom in the side chain [Ile and Thr, cf. Inman and Apella (1977) and Trp^{hex} in this communication].

Table 1: Characterization of N-Terminal Fragments from RNase U_s by Edman Degradation and ESI-MS[a]

residues	sequence	PTH-derivative of X	mass $[M+H]^+$ observed	mass $[M+H]^+$ expected
1. 1−12	K-P-P-Q-F-T-X-A-Q-W-F-E	X_1, X_2, X_3	1726.8	1564.8
2. 1−12	K-P-P-Q-F-T-X-A-Q-W-F-E	X_1, X_2[b]	1727.0	1564.8
3. 5−10	F-T-X-A-Q-W	X_1, X_2, X_3	1000.4	838.4
4. 5−8	F-T-X-A	X_1, X_2, X_3	686.4	524.3
5. 6−8	T-X-A	nd	539.2	377.2
6. 7	X	X_1, X_2	nd[c]	322.1[d]

[a] Peptides 1, 3, and 4 were prepared from reduced and carboxymethylated RNase U_s, whereas peptide 2 was obtained from sulfitolyzed RNase U_s. X_{1-3} refer to the PTH-derivatives obtained upon Edman degradation as shown in Figure 1. [b] The minor component X_3 was not detected due to the low amount of material sequenced. [c] Only EI data available [*m/z* 292, base peak, due to R(X) vs *m/z* 130 in PTH-Trp]. [d] $[M+H]^+$ of PTH-Trp.

nm toward lower wavelength and is about 2.8-fold higher. These results are consistent with the presence of a modified tryptophan residue at position 7 of RNase U_s. Reversed-phase HPLC of the peptide F-T-X-A showed that the modifying group was of a polar nature, since the peptide eluted considerably earlier than the unmodified, synthetic peptide (data not shown).

Mass Spectrometry. ESI-MS analysis of the five peptides listed in Table 1 showed that in each case the observed mass is 162 Da higher than that expected for tryptophan. It is important to note that peptides **1** and **2** have identical molecular masses but were obtained from RNase U_s prepared for enzymatic cleavage by entirely different methods (see the legend to Table 1). This reinforces the conclusion that the methods used to generate peptides did not modify the residue at position 7.

In order to obtain structural information on the substituent responsible for this increment, the tripeptide T-X-A (**5**, Table 1) was peracetylated with deuterioacetic anhydride (Ac^*_2O, where $Ac^* = C^2H_3CO$) in pyridine. The resulting derivative (**5**-Ac^*_6 $[M+H]^+$ at *m/z* 809.6, spectrum not shown) had incorporated six deuterioacetyl groups compared to only two in synthetic T-W-A (one into the threonine β-OH and one into the α-NH_2-group). This demonstrated that X in T-X-A contains four groups that are reactive toward Ac^*_2O. Together with the mass of 162 Da, this strongly suggested that the modifying substituent on Trp-7 is a hexosyl residue and excluded hydroxylation followed by the conjugation with deoxyhexose. Monoglycosylation is usually corroborated by MS/MS and CID. Low energy CID in an ESI-MS/MS experiment with **5**-Ac_6^* using mass selected $[M+H]^+$ ions (*m/z* 809) gave the daughter ion spectrum shown in Figure 2 (upper panel). Since the N-terminal B_2, the C-terminal Y_2, and an "inner" immonium ion $A_1(X)$ (see Chart 1, R = Ac*) all were 342 Da heavier than those obtained from synthetic T-W-A, they had to contain the hexose unit in question. While these ions merely showed the tryptophan residue to be modified, an R(X) ion at *m/z* 472 was more informative in that it corresponded to a modified indolyl-methyl cation, thus limiting the modification to sites within the side chain of this residue (see Chart 1). Fragmentation processes indicating glycosidic attachment of a six-carbon building block more directly, i.e., elimination of an "intact" $C_6H_6O_5Ac^*_4$ anhydrosugar moiety (loss of 342 Da) and/or formation of a highly characteristic $C_6H_7O_5Ac^*_4$ "sugar ion" (*m/z* 343) were, however, conspicuously absent. This strongly suggested that the hexose was not *N*-, but *C*-glycosidically linked to the indole moiety. The presence of a free indole N1 was demonstrated by H/^{2}H exchange in a DCI-MS and DCI-MS/MS experiment using N^2H_3 as the

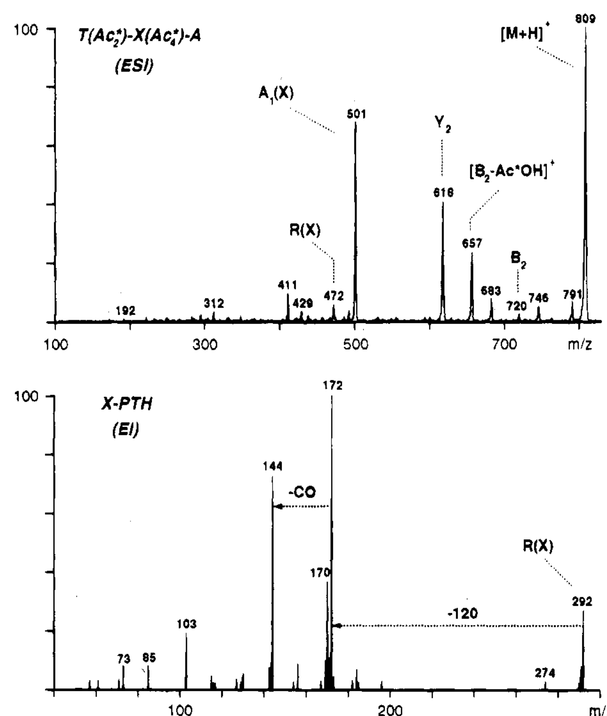
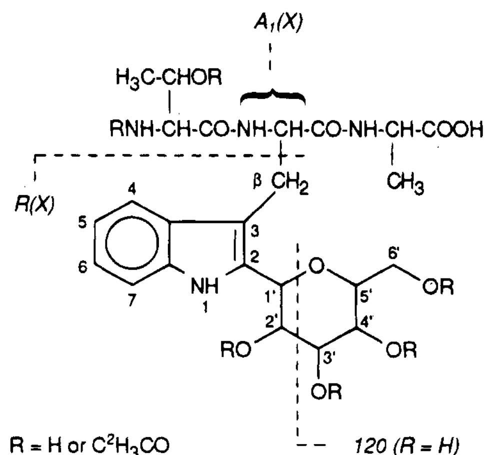


FIGURE 2: Mass spectrometric analysis of perdeuterioacetylated T-X-A and X-PTH. ESI-MS/MS daughter ion spectrum of mass selected $[M+H]^+$ ions of **5**-Ac_6^* (*m/z* 809, upper panel) and EI-MS/MS daughter ion spectrum of mass selected R(X) fragment ions of X-PTH (**6**, *m/z* 292, lower panel). The collision energies used were 55 and 35 eV, respectively.

Chart 1



reagent gas ($[M+H]^+$ at *m/z* 809 → $[M(d_5)+^2H]^+$ at *m/z* 815; data not shown).

Evidence in favor of *C*-glycosidic attachment of an aldohexosyl residue to the indole moiety was derived from

ESI-MS/MS of the PTH-derivative of X (**6**-PTH, Table 1) using EI (spectrum not shown). When subjected to low energy CID (Figure 2, lower panel), the R(X) ions decomposed excessively by loss of 120 Da (m/z 292 → 172). This reflects the net loss of four $CH_2{=}O$, indicating that only part of the $C_6H_{10}O_5$ unit is readily released (see Chart 1, R = H). Corresponding losses of 120 Da were also observed in the analysis of peptides **1**−**5**. The facile loss of four oxygenated carbon atoms can be taken as evidence for a glycosidic attachment of an aldohexose via an aryl C−C bond, i.e., a bond not readily cleaved, since exactly such CID behavior has been reported for flavone *C*-glycopyranosides having a C1′ attachment (Becchi & Fraisse, 1989; Li et al., 1992; Prox, 1968). In another important fragmentation channel the ions at m/z 172 decompose even further, i.e., by loss of 28 Da, due to secondary loss of CO, retaining only the C atom by which the hexose moiety is attached (C1′). This transition is also reported as being pronounced in analogous flavone *C*-glycopyranosides (Becchi & Fraisse, 1989; Li et al., 1992; Prox, 1968).

NMR Spectroscopy. To elucidate the structure of X-7 NMR analysis was performed on the hexapeptide F-T-X-A-Q-W. Resonances originating from the five amino acids F, T, A, Q, and W were readily assigned on the basis of characteristic spin systems observed in the 2D TOCSY spectra and by sequential ROEs between α and backbone amide protons. Intraresidue ROEs correlated the aromatic side chain protons to their respective α and β/β' protons (not shown; Wüthrich, 1986). The backbone amide proton (δ 8.31) of residue X-7 was identified by a sequential ROE to the α proton (δ 4.24) of the preceding Thr-6 and correlated to its respective α (δ 4.54) and β/β' protons (δ 3.36/3.34) by the through-bond connectivity pattern observed in the TOCSY spectrum (Figure 3). The sequential ROE from the α proton of X-7 to the backbone amide proton of Ala-8 (δ 7.91) corroborated these assignments. H4 of X-7 (δ 7.67), resonating at a chemical shift typical for tryptophan H4s (Wüthrich, 1986), was identified by the ROE to the β/β' protons of X-7 and shows a TOCSY connectivity pattern to H5 (δ 7.14), H6 (δ 7.20), and H7 (δ 7.41) characteristic of a tryptophan residue (Figure 3; Wüthrich, 1986). The signal at 10.47 ppm in the ^{1}H NMR spectrum, a chemical shift typical of indole amide protons (Wüthrich, 1976, 1986), was only observable when spectra were recorded of the hexapeptide dissolved in H_2O or DMSO and was, therefore, assigned to an indole NH. It was identified as the N1H of X-7 on the basis of an intraresidue ROE from the H7 (Figure 3). Furthermore, ^{13}C chemical shifts of the indole moiety of X-7 (Table 2) closely resemble those of Trp-10 (Figure 3) and tryptophan residues in other peptides (Wüthrich, 1976). Thus, the data clearly demonstrate that X-7 is a modified tryptophan residue.

The presence of the indole protons, N1H and H4−H7, only leaves the possibility of substitution at the 2-position of Trp-7. The absence of H2 was concluded from the following observations. In the 1D ^{1}H NMR spectrum of the unmodified synthetic peptide F-T-W-A a well-resolved H2 signal (δ 7.30) was observed, but this signal was clearly absent in the spectrum of F-T-X-A (not shown). The N1H of Trp-7 appeared as a singlet in the 1D ^{1}H NMR spectrum, whereas the N1H resonance of Trp-10 displayed the expected coupling constant (1 Hz) due to the presence of the vicinal H2. Moreover, through-bond connectivity from α and β/β' protons to H2 were not observed for Trp-7 but were clearly visible for Trp-10. Furthermore, intraresidue ROEs were observed from the Trp-10 β/β' protons to its respective H2 proton and from the Trp-10 N1H to its H2, but these ROEs were conspicuously absent in the case of Trp-7 (Figure 3). Finally, only a single tryptophan H2/C2 correlation was observed in the HSQC spectrum, belonging to Trp-10 (Figure 3). In summary, all of the protons of Trp-7, except H2, were observed, demonstrating that this residue is substituted at C2.

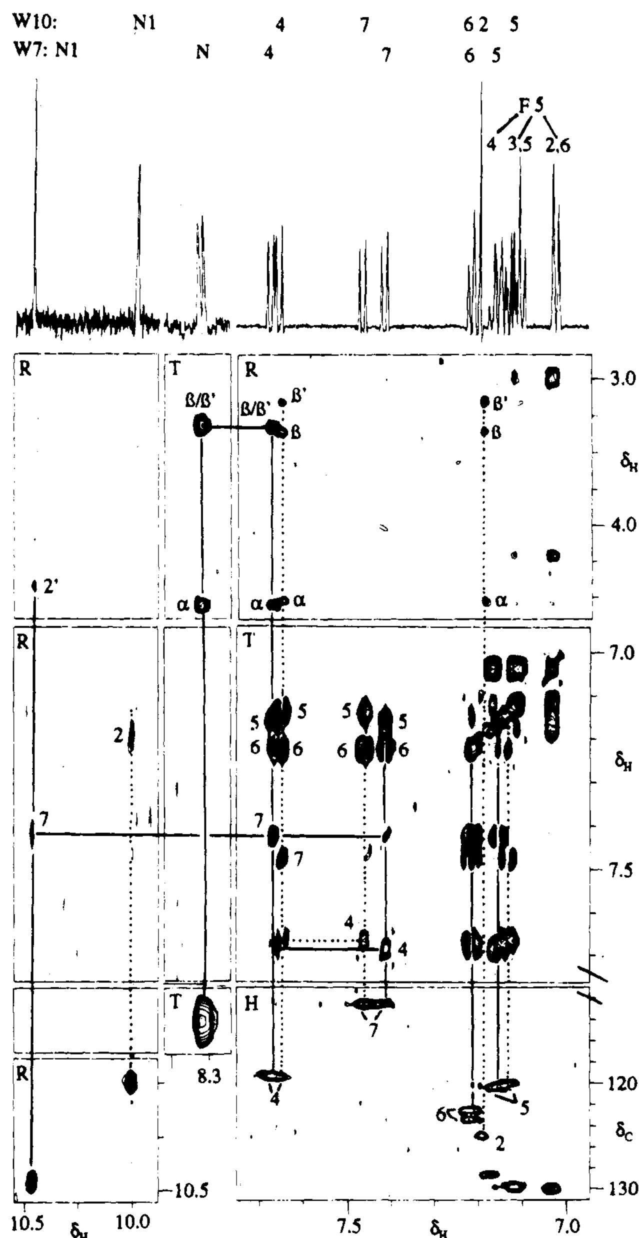


FIGURE 3: Assignment pathway of the tryptophan moiety of X-7. Fragments of 1D, 2D ROESY (R), 2D TOCSY (T), and 2D ^{1}H−^{13}C HSQC (H) spectra of FTXAQW. The assignment pathway is depicted for the tryptophan moiety of X-7 (solid lines) and for W-10 (dotted lines). The ROE between N1H of W-7 and H2′ is also shown. Other cross-peaks originate from F-5; their aromatic proton resonances are shown in the 1D spectrum. The notation for tryptophan residues is as depicted in Chart 1 and as described by Wüthrich (1986). Symbols in the 1D spectrum represent resonances of respective protons. In the HSQC spectrum they represent corresponding ^{1}H−^{13}C correlations. In ROESY and TOCSY spectra, α, β, or a number indicate a correlation of the corresponding proton to the proton indicated by the vertical line.

Table 2: ^{1}H and ^{13}C Chemical Shifts (ppm) of Residue 7 in RNase U$_s$ and the Homonuclear Vicinal Coupling Constants (Hz) of the Hexosyl Moiety[a]

proton	δ_H ($J_{n,n+1}$)	carbon	δ_c
NH	8.31		
Hα	4.54	Cα	55.5
Hβ	3.36	Cβ	26.7
Hβ′	3.34		
H4	7.67	C4	119.3
H5	7.14	C5	120.1
H6	7.20	C6	123.4[b]
H7	7.41	C7	112.4
N1H	10.47		
H1′	5.22 (7.8)	C1′	68.1
H2′	4.42 (3.2)	C2′	68.9
H3′	4.09 (5.5)	C3′	71.3
H4′	3.96 (3.8)	C4′	69.5
H5′	3.87 (8.3; 3.3)	C5′	79.6
H6′	4.21	C6′	60.3
H6″	3.77		

[a] ^{1}H chemical shifts are given at 300 K by reference to internal acetone (δ 2.225; Vliegenthart et al., 1983), while ^{13}C chemical shifts are given by reference to the α-anomeric resonance of ^{13}C-labeled glucose (δ 92.9; Bock & Pedersen, 1983). The ^{1}H chemical shifts and the vicinal ^{1}H−^{1}H coupling constants were determined from 600-MHz 1D spectra, except for H5 and H6 whose ^{1}H chemical shifts were determined from the 2D TOCSY spectrum. The ^{13}C chemical shifts were determined from the ^{1}H−^{13}C correlations in the 2D ^{1}H−^{13}C HSQC spectrum. The notation for tryptophan is as described by Wüthrich (1986), and for the hexose moiety as described by Vliegenthart et al. (1983). [b] Might be interchanged with the ^{13}C6 chemical shift of W-10 (δ 122.6).

Inspection of the 1D ^{1}H NMR spectrum of F-T-X-A-Q-W revealed that the substituent linked to C2 of Trp-7 contains seven protons, which are scalarly coupled to each other as demonstrated in the TOCSY spectrum (Figure 4). It should be noted that the ^{1}H chemical shifts and the spin-coupling topology are similar to those observed for aldohexoses (Bock et al., 1983). The doublet at 5.22 ppm was assigned to H1′ of the hexose substituent; the other protons were assigned by the TOCSY pattern and vicinal ^{1}H−^{1}H coupling constants (Table 2). The ^{1}H−^{13}C HSQC spectrum revealed the presence of one hydroxymethyl group (C6) and five methine carbons (Figure 4), consistent with an aldohexosyl moiety. This was further substantiated by the ^{13}C chemical shifts of C2′−C6′ (Table 2), which are comparable to those observed for aldohexoses (Bock et al., 1983; Bock & Pedersen, 1983; King-Morris & Serianni, 1987). The identity of the hexose could not be determined by comparison with chemical shift and coupling constant data of monosaccharides (Bock et al., 1983; Bock & Pedersen, 1983; Adinolfi et al., 1993). The 1D ^{1}H NMR spectrum of acetylated F-T-X-A showed large downfield shifts (about 1 ppm) for H2′, H3′, H4′, H6′, and H6″ with respect to the chemical shifts of the nonacetylated form, indicating the presence of hydroxyl groups at the respective carbons, and corroborating that the substituent is indeed an aldohexosyl group. In contrast, H5′ experienced only a slight downfield shift upon acetylation, indicating the lack of a free hydroxyl group at C5′. Together with the ^{13}C-chemical shift of C5′ (79.6 ppm), which is characteristic for C5s of aldohexoses in the pyranose form, this shows that the substituent is an aldohexopyranose (Bock & Pedersen, 1983; King-Morris & Serianni, 1987). Acetylation did not affect the chemical shift of H1′ significantly, indicating that the hexopyranosyl moiety is linked via C1′. H1′ and H2′ resonate further downfield than in other hexoses, which is probably caused by their proximity to the indole ring. Intense ROEs from H1′ and H2′ to the Trp-7 β/β′ protons and the Trp-7 N1H, respectively, are consistent with linkage via C1′. The ^{13}C-chemical shift of C1′ (δ 68) is far upfield of those for C1 in free aldohexoses and *O*-glycosides (90−105 ppm; Bock & Pedersen, 1983; King-Morris & Serianni, 1987), indicating that the substituent is not linked through an *O*-glycosidic bond. The presence of the N1H of Trp-7 in the NMR spectra already excluded an *N*-glycosidic bond to the indole nitrogen, as does a comparison with the chemical shift of the anomeric carbon atom of the ribose residue in guanosine (δ 85; Leupin et al., 1987). Comparison with the ^{13}C-chemical shifts of C1′ (δ 70−71) in compounds

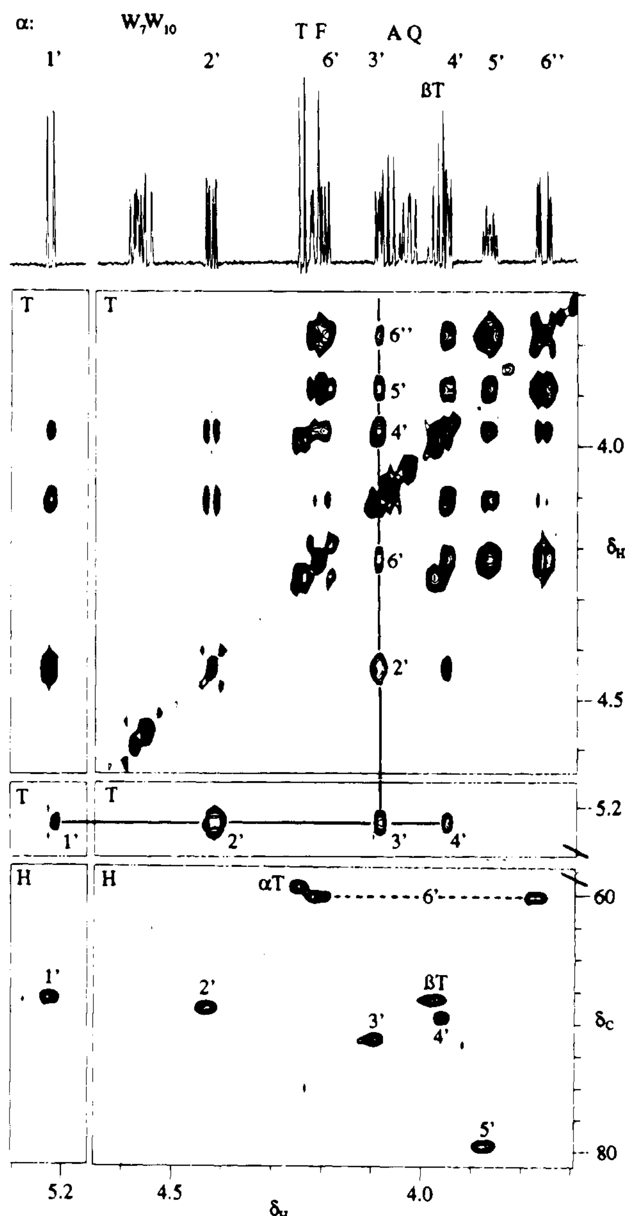


FIGURE 4: Assignment pathway of the aldohexopyranosyl moiety of X-7. Fragments of a 1D, a 2D TOCSY (T), and a 2D HSQC (H) spectrum of FTXAQW. The assignment pathway of the aldohexopyranosyl moiety is depicted starting at the diagonal peak, indicated by 1′. The notation for the aldohexopyranosyl moiety is as shown in Chart 1 and for amino acids as described by Wüthrich (1986). Symbols in the 1D spectrum represent resonances of their respective protons. In the HSQC spectrum they represent the corresponding ^{1}H−^{13}C correlations. Numbers at off-diagonal cross-peaks in the TOCSY spectrum indicate a correlation between the corresponding proton and the proton indicated by the solid line.

having a monosaccharide residue in C1′−C linkage to an aryl ring (Farr et al., 1992; Kato & Morita, 1993) shows that the aldohexopyranosyl moiety is *C*-glycosidically linked to Trp-7.

DISCUSSION

Different modes of attachment of carbohydrates to amino acid side chains of proteins are known to occur [see Stults and Cummings (1993) and references therein]. Oligosaccharides may be linked *N*-glycosidically to the asparagine side chain, or *O*-glycosidically to serine and threonine side chains. In addition, *N*-acetylglucosamine bound *O*-glycosidically to the side chain of serine has been found in a large number of nuclear and cytoplasmic proteins (Hart et al., 1989). Hydroxylated amino acids like 5-hydroxylysine in collagen and 4-hydroxyproline in plant cell wall proteins can also be *O*-glycosylated (Stults & Cummings, 1993). It remains to be determined whether or not the formation of the *C*-glycosidic bond to tryptophan reported here is enzyme-catalyzed. If so, it would require a different biosynthetic pathway than the above-mentioned sugar−protein linkages, since it involves the formation of a carbon−carbon bond. Although aromatic *C*-glycosides in proteins have not been reported before, low molecular weight *C*-glycosides are very abundant in plants. At least five major classes can be distinguished on the basis of the aglycon involved (Franz & Grün, 1983); attachment of the carbohydrate in all cases occurs to a carbon atom of nucleophilic character, and it has been shown in one case that the sugar donor is UDP-glucose (Franz & Grün, 1983). Activation of sugars by the formation of diphosphate derivatives is also used in the biosynthesis of *N*- and *O*-glycosylated proteins. It will be of interest to see whether in the synthesis of Trp^{hex} a diphosphate-activated sugar derivative is the precursor for the glycosylation at the nucleophilic C2 atom.

It is unlikely that Trp^{hex} in RNase U_s is an artifact of the purification of the enzyme from urine for a number of reasons. First, the unknown residue at position 7 of RNase U_s from erythrocytes (Yasuda et al., 1990a) was also found to be Trp^{hex} (A. Löffler and J. Hofsteenge, unpublished results). Second, RNase U_s contains a second tryptophan at position 10, which is not modified. Third, human RNases that are very similar, if not identical, to RNase U_s have been purified from liver (Sorrentino et al., 1988), spleen (Yasuda et al., 1990b), placenta (Shapiro & Vallee, 1991), and eosinophils (Gleich et al., 1986). Since the human genome contains only one locus for EDN/RNase U_s (Hamann et al., 1990), it seems likely that these RNases are the product of the same gene. In all of these enzymes the residue at position 7 seems to be modified, since it could not be identified by Edman degradation. Whether the substituent is an aldohexopyranosyl residue in these cases as well remains to be determined. In this connection, it should be noted that the enzyme from placenta has spectroscopic properties similar to those reported here for the peptide F-T-W^{hex}-A. It displayed an increase in the absorbance around 280 nm, as well as in the fluorescence emission spectrum (Shapiro & Vallee, 1991). An unusual RNase U_s isolated from the urine of pregnant women is of particular interest. The protein is highly glycosylated and contains four additional N-terminal amino acids (Sakakibara et al., 1992). Furthermore, at position 7 an unmodified tryptophan was reported to be present. This raises the possibility that the modification of Trp-7 in RNase U_s by an aldohexopyranosyl residue is linked to other posttranslational modifications and is physiologically regulated.

Recently, Gäde et al. (1992) have proposed that a hexosyl residue is attached to the tryptophan in a neuropeptide from the stick insect *C. morosus*. The site of attachment was not established, but it was thought likely to be the indolic N atom (Gäde et al., 1992). From our results it is clear that in RNase U_s the hexose is linked to the C2 position, and it will be of interest to establish whether the same is true for the insect neuropeptide. It is intriguing to note that the primary structure of RNase U_s is identical to that of EDN, and that also in this protein Trp-7 appears to be modified. It remains to be seen whether the modification of the tryptophan in this protein and in the neuropeptide is functionally related to their biological activity. The finding of a hexose-modified tryptophan residue in both a peptide from insects and a human enzyme raises the possibility that this kind of posttranslational modification is widespread both at the protein and the phylogenetic level.

ACKNOWLEDGMENTS

We thank Renate Matthies for technical help and Drs. A. Matus, B. Ernst, F. Meins, and J. Thomas for carefully reading the manuscript.

REFERENCES

Adinolfi, M., Barone, G., Corsaro, M. M., Lanzetta, R., & Mangoni, L. (1993) *J. Carbohydr. Chem. 12*, 903−911.

Bax, A., & Davis, D. G. (1985a) *J. Magn. Reson. 65*, 355−360.

Bax, A., & Davis, D. G. (1985b) *J. Magn. Reson. 63*, 207−213.

Becchi, M., & Fraisse, D. (1989) *Biomed. Environ. Mass Spectrom. 18*, 122−130.

Beintema, J. J., Hofsteenge, J., Iwama, M., Morita, T., Ohgi, K., Irie, M., Sugiyama, R. H., Schieven, G. L., Dekker, C. A., & Glitz, D. G. (1988) *Biochemistry 27*, 4530−4538.

Bock, K., & Pedersen, C. (1983) *Adv. Carbohydr. Chem. Biochem. 41*, 27−65.

Bock, K., Pedersen, C., & Pedersen, H. (1983) *Adv. Carbohydr. Chem. Biochem. 42*, 193−225.

Braunschweiler, L., & Ernst, R. R (1983) *J. Magn. Reson. 53*, 521−528.

Chan, W. W.-C. (1968) *Biochemistry 7*, 4247−4253.

D'Alessio, G. (1993) *Trends Cell Biol. 3*, 106−109.

de Beer, T., van Zuylen, C. W. E. M., Hård, K., Boelens, R., Kaptein, R., Kamerling, J. P., & Vliegenthart, J. F. G. (1994) *FEBS Lett. 348*, 1−6.

Durack, D. T., Ackerman, S. J., Loegering, D. A., & Gleich, G. J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5165−5169.

Farr, R. N., Kwok, D.-I., & Daves, G. D. (1992) *J. Org. Chem. 57*, 2093−2100.

Franz, G., & Grün, M. (1983) *Planta Medica 47*, 131−140.

Gäde, G., Kellner, R., Rinehart, K. L., & Proefke, M. L. (1992) *Biochem. Biophys. Res. Commun. 189*, 1303−1309.

Gleich, G. J., Loegering, D. A., Bell, M. P., Checkel, J. L., Ackerman, S. J., & McKean, D. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3146−3150.

Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870−7872.

Hamann, K. J., Ten, R. M., Loegering, D. A., Jenkins, R. B., Heise, M. T., Schad, C. R., Pease, L. R., Gleich, G. J., & Barker, R. L. (1990) *Genomics 7*, 535−546.

Hård, K., van Zadelhoff, G., Moonen, P., Kamerling, J. P., & Vliegenthart, J. F. G. (1992) *Eur. J. Biochem. 209*, 895–915.

Hart, G. W., Haltiwanger, R. S., Holt, G. D., & Kelly, W. G. (1989) *Annu. Rev. Biochem. 58*, 841–874.

Hofsteenge, J., Servis, C., & Stone, S. R. (1991) *J. Biol. Chem. 266*, 24198–24204.

Inman, J. K., & Apella, E. (1977) *Methods Enzymol. 47*, 374–385.

Iwama, M., Kunihiro, M., Ohgi, K., & Irie, M. (1981) *J. Biochem. (Tokyo) 89*, 1005–1016.

Kato, T., & Morita, Y. (1993) *Heterocycles 35*, 965–973.

King-Morris, M. J., & Serianni, A. S. (1987) *J. Am. Chem. Soc. 109*, 3501–3508.

Leupin, W., Wagner, G., Denny, W. A., & Wüthrich, K. (1987) *Nucleic Acids Res. 15*, 267–275.

Li, Q. M., van den Heuvel, H., Dillen, L., & Claeys, M. (1992) *Biol. Mass Spectrom. 21*, 213–221.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 393–399.

McIntire, W. S., Wemmer, D. E., Christoserdov, A., & Lidstrom, M. E. (1991) *Science 252*, 817–823.

Prox, A. (1968) *Tetrahedron 24*, 3697–3700.

Reimert, C. M., Minuva, U., Kharazmi, A., & Bendtzen, K. (1991) *J. Immunol. Methods 141*, 97–104.

Sakakibara, R., Hashida, K., Kitahara, T., & Ishiguro, M. (1992) *J. Biochem. (Tokyo) 111*, 325–330.

Shapiro, R., & Vallee, B. L. (1991) *Biochemistry 30*, 2246–2255.

Slifman, N. R., Loegering, D. A., McKean, D. J., & Gleich, G. J. (1986) *J. Immunol. 137*, 2913–2917.

Sorrentino, S., Tucker, G. K., & Glitz, D. G. (1988) *J. Biol. Chem. 263*, 16125–16131.

Stults, N. L., & Cummings, R. D. (1993) *Glycobiology 3*, 589–596.

Tarr, G. E. (1977) *Methods Enzymol. 47*, 335–357.

Vliegenthart, J. F. G., Dorland, L., & van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem. 41*, 209–374.

Wüthrich, K. (1976) in *NMR in Biological Research: Peptides and Proteins*, North Holland Publishing Company, Amsterdam.

Wüthrich, K. (1986) in *NMR of Proteins and Nucleic acids*, John Wiley, New York.

Yasuda, T., Mizuta, K., & Kishi, K. (1990a) *Arch. Biochem. Biophys. 279*, 130–137.

Yasuda, T., Mizuta, K., Sato, W., & Kishi, K. (1990b) *Eur. J. Biochem. 191*, 523–529.