Biochem. J. (1974) **143**, 197–205
Printed in Great Britain

The Molecular and Crystal Structures of 4-*N*-(2-Acetamido-2-Deoxy-β-D-Glucopyranosyl)-L-Asparagine Trihydrate and 4-*N*-(β-D-Glucopyranosyl)-L-Asparagine Monohydrate

THE X-RAY ANALYSIS OF A CARBOHYDRATE–PEPTIDE LINKAGE

By L. T. J. DELBAERE*
Chemical Crystallography Laboratory, South Parks Road, Oxford OX1 3QS, U.K.

(*Received* 1 *April* 1974)

X-ray analyses have shown that the glucopyranose rings of GlcNAc-Asn [4-*N*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine] and Glc-Asn [4-*N*-(β-D-glucopyranosyl)-L-asparagine] both have the C-1 chair conformation and also that the glucose–asparagine linkage of each molecule is present in the β-anomeric configuration. The dimensions (the estimated standard deviations of the last digit are in parentheses) of the glycosidic bond in GlcNAc-Asn and Glc-Asn are, respectively, $C_{(1)}$-$N_{(1)}$ 0.1441(6) nm, 0.146(2) nm; angle $O_{(5)}$-$C_{(1)}$-$N_{(1)}$ 106.8(3)°, 105.7(8)°; angle $C_{(2)}$-$C_{(1)}$-$N_{(1)}$ 111.1(4)°, 110.4(9)°; angle $C_{(1)}$-$N_{(1)}$-$C_{(9)}$ 121.4(4)°, 120.5(9)°. The glycosidic torsion angle $C_{(9)}$-$N_{(1)}$-$C_{(1)}$-$C_{(2)}$ is 141.0° and 157.6° in GlcNAc-Asn and Glc-Asn respectively. Hydrogen-bonding is extensive in these two crystal structures and does affect one torsion angle in particular. Two very different values of χ_1(N-C^{α}-C^{β}-C^{γ}) occur for the asparagine residue of the two different molecules; the values of χ_1, −69.0° in GlcNAc-Asn and 61.9° in Glc-Asn, correspond to two different staggered conformations about the C^{α}-C^{β} bond as the NH_3^+ group is adjusted to different hydrogen-bonding patterns. The two *trans*-peptide groups in GlcNAc-Asn show small distortions in planarity whereas that in Glc-Asn is more non-planar. The mean plane through the atoms of the amide group at $C_{(2)}$ in GlcNAc-Asn is approximately perpendicular (69°) to the mean plane through the $C_{(2)}$, $C_{(3)}$, $C_{(5)}$ and $O_{(5)}$ atoms of the glucose ring and that at $C_{(1)}$ is less perpendicular (65°). The mean plane through the atoms of the amide group in Glc-Asn makes an angle of only 55° with the mean plane through these same four atoms of the glucose ring. The $N_{(1)}$-H bond of the amide at $C_{(1)}$ is *trans* to the $C_{(1)}$-H bond in these two compounds; the $N_{(2)}$-H bond of the amide at $C_{(2)}$ is *trans* to the $C_{(2)}$-H bond in GlcNAc-Asn. The values of the observed and final calculated structure amplitudes have been deposited as Supplementary Publication SUP 50035 (26 pages) at the British Library (Lending Division), (formerly the National Lending Library for Science and Technology), Boston Spa, Yorks. LS23 7BQ, U.K., from whom copies may be obtained on the terms given in *Biochem. J.* (1973) **131**, 5.

The compound GlcNAc-Asn [4-*N*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine], shown in Fig. 1, is a model for the carbohydrate–peptide linkage which occurs in several glycoproteins (Marshall & Neuberger, 1972). Further, the presence of the 2-acetamido group of the glucose moiety of GlcNAc-Asn enhances the possibility of the molecule being an inhibitor of lysozyme; crystallographic investigations (Beddell, 1970) indicate that the 2-acetamido-glucosyl residue of GlcNAc-Asn is bound in the crystalline state to site C of lysozyme with the 1-aspartamido group of the molecule being located in site D (Blake *et al.*, 1967). The crystal structure of GlcNAc-Asn was therefore investigated to obtain the exact dimensions of this important linkage compound and also the overall conformation of the molecule. Subsequently, the X-ray structure of Glc-Asn [4-*N*-(β-D-glucopyranosyl)-L-asparagine] was determined in order to have two independent sets of parameters for both the carbohydrate–protein linkage and also for the conformation of the aspartyl-glucose moiety. Glc-Asn has been given the same numbering scheme throughout this paper as that indicated for GlcNAc-Asn in Fig. 1 with the exception that $O_{(2)}$ in the former compound replaces the *N*-acetyl group at $C_{(2)}$ in the latter molecule.

Different hydrogen bonding and packing patterns would provide an indication of the conformational

*Present address: Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2E1, Canada.

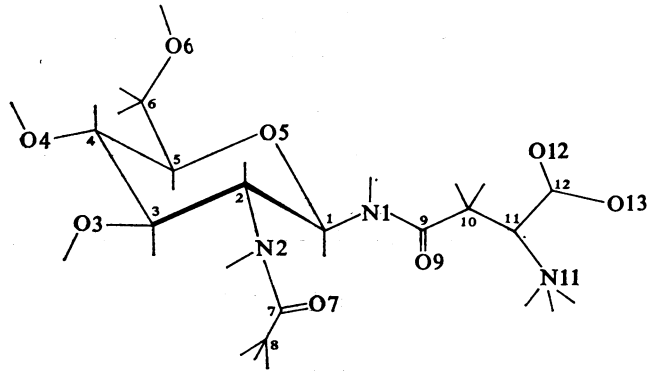

Fig. 1. 4-*N*-(2-*Acetamido*-2-*deoxy*-β-D-*glucopyranosyl*)-L-*asparagine*

flexibility of the residues common to both GlcNAc-Asn and Glc-Asn, which should aid in establishing the conformation in solution and in the complex of GlcNAc-Asn with lysozyme. Ramachandran *et al.* (1973) have shown that the theoretical state of minimum energy of the *trans*-peptide unit is a non-planar conformation and in these two molecules there are three such units which may be examined to see how the observed solid-state conformations compare with this theory. Austen & Marshall (1970, 1974) have related the chiroptical properties of these molecules to their conformation in solution.

Experimental

Crystal data

Crystals of the two compounds were kindly provided by Dr. B. M. Austen and Dr. R. D. Marshall of the Department of Chemical Pathology, St. Mary's Hospital Medical School, London, U.K.; the GlcNAc-Asn trihydrate had been crystallized from an aqueous ethanol solution at 4°C and the Glc-Asn monohydrate was crystallized from an aqueous ethanol solution at 20°C. The crystal data that were obtained are given in Table 1. A single crystal of each compound was mounted about the *a* axis. The intensity data were collected for the 0*kl*–4*kl* levels on a Hilger–Watts linear diffractometer by a moving-crystal stationary-counter method with MoK_α radiation ($\lambda = 0.07107$ nm) and balanced filters. The peak scan range was twice that of each background scan. Lorentz and polarization corrections factors were applied, but no absorption corrections were made.

Structure determination and refinement

(*i*) *GlcNAc-Asn.* A Wilson (1942) plot was used to convert the observed structure amplitudes into normalized structure factors, E values (Karle & Hauptman, 1956); the overall isotropic temperature factor is 0.0285 nm^2. Initial symbolic addition by hand on a Σ_2-listing (Karle & Karle, 1966) gave the following starting set of phases:

h	*k*	*l*	E	Phase	
1	6	0	2.9	0°	Origin-defining reflexions
0	3	0	2.1	0°	
0	6	1	2.3	0°	
2	6	0	2.2	0°	From Σ_1-type indications (Karle & Hauptman, 1956)
1	3	18	3.9	*a*	(0°, 90°, 180°, 270°)
3	$\bar{5}$	7	3.4	*b*	(45°, 135°) defines enantiomorph
1	2	17	2.1	*c*	(0°, 90°, 180°, 270°)

Tangent formula (Karle & Hauptman, 1956) refinement on the phases of the 223 reflexions with E ≥ 1.4 for the 32 solutions gave R_{Karle} values (Karle & Karle, 1966) of 0.196, 0.211, etc. up to a maximum value of 0.268. An E map based on the phases determined for the solution ($a = 90°$, $b = 45°$ and $c = 270°$) with the lowest R_{Karle} value (0.196) contained all of the non-hydrogen atoms of the molecule of GlcNAc-Asn in addition to one water molecule. An initial structure-factor calculation based on these atoms gave an *R* index of 0.28 where

$$R = \frac{\Sigma||F_o| - |F_c||}{\Sigma|F_o|}$$

A difference map revealed a second water molecule with the third water molecule distributed equally over two different crystallographic sites. Three cycles of full-matrix least-squares isotropic refinement gave an *R* index value of 0.129. Three further cycles of least-squares refinement with anisotropic temperature factors decreased the *R* index value to 0.068;

Table 1. *Crystal data*

Compound	GlcNAc-Asn	Glc-Asn
Chemical formula	$C_{12}O_8N_3H_{21}$, $3H_2O$	$C_{10}O_8N_2H_{18}$, H_2O
Formula weight	389.37 daltons	312.29 daltons
Crystal system	Monoclinic	Monoclinic
Systematic absences	$00l(l=2n+1)$	$00l(l=2n+1)$
Space group	$P2_1$(*c*-unique)	$P2_1$(*c*-unique)
Unit-cell dimensions:		
a	0.494(1) nm	0.494(1) nm
b	0.777(1) nm	0.808(1) nm
c	2.426(4) nm	1.668(2) nm
γ	97.7(3)° nm	96.1(3)°
V	0.9227 nm^3	0.6618 nm^3
Unit-cell content *Z*	2	2
Measured density (by flotation)	1.42 $g \cdot cm^{-3}$	1.57 $g \cdot cm^{-3}$
Predominant crystal forms	{100}, {010} {001}, {101}	{100}, {010} {001}, {00$\bar{1}$}
Linear absorption coefficient μ_{Mokx}	1.3 cm^{-1}	1.5 cm^{-1}
Size of crystal	0.45 mm × 0.20 mm × 0.050 mm	0.90 mm × 0.35 mm × 0.017 mm
Number of reflexions with intensities $>3\sigma(I)$	1438	892
Total scan range	2.5°	3.0°
F(000)	404 electrons	328 electrons

two matrix blocks were used, one for the refinement of the scale and temperature factors and the other for the positional parameters. A difference electron-density map then revealed the positions of the hydrogen atoms of the GlcNAc-Asn molecule. The hydrogen atoms were assigned the isotropic temperature factors which occurred, after the isotropic refinement, for the atoms to which they were bonded. Further least-squares refinement of only the non-hydrogen atom parameters converged to a final *R* index of 0.054. The weighting scheme used was $W=1$ if

$$|F_o| < 9.88 \text{ and } W = \frac{9.88}{|F_o|} \text{ if } |F_o| > 9.88.$$

(*ii*) *Glc-Asn.* The observed structure amplitudes were converted into E values by means of a Wilson (1942) plot. The overall isotropic temperature factor is 0.0245 nm^2. The following starting set of phases was obtained by carrying out symbolic addition by hand (Karle & Karle, 1966):

h	*k*	*l*	E	Phase	
1	6	0	2.8	0°	Origin-defining reflexions
2	7	0	2.4	0°	
0	5	1	2.3	0°	
0	10	0	1.7	180°	From Σ_1-type indications
1	$\bar{6}$	7	3.2	*a*	(45°, 135°) defines enantiomorph
1	5	15	2.9	*b*	(90°, 270°)
1	$\bar{4}$	0	2.5	180°	
4	$\bar{10}$	1	3.0	0°	

Tangent formula (Karle & Hauptman, 1956) refinement on the phases of the 176 reflexions with $E \geqslant 1.4$ gave R_{Karle} values of 0.202, 0.204, 0.220 and 0.236 for the four solutions. The E map of the solution with the lowest R_{Karle} value was not chemically meaningful whereas the E map of the solution ($a=135°$, $b=90°$) with $R_{Karle}=0.204$ contained all of the non-hydrogen atoms of the structure. Three cycles of full-matrix least-squares isotropic refinement gave an *R* index of 0.158. An additional three cycles of least-squares refinement with anisotropic temperature factors lowered the *R* index to 0.091; the positional parameters were refined in one matrix block, with the scale and temperature parameters refined in a second matrix block. The locations of the hydrogen atoms were determined from a subsequent difference map; the hydrogen atoms were given the final isotropic temperature factors of the atoms to which they were bonded. Three more cycles of least-squares refinement of the non-hydrogen atom parameters converged at an *R* index of 0.072. Unit weights were used for the first three cycles of refinement and then the weighting scheme used was:

$$W = \frac{1-\exp\left(\frac{\sin\theta}{\lambda}\right)^2}{1-0.044|F_o| \times 0.0011|F_o|^2}$$

The atomic scattering factors used were obtained from the International Tables for X-ray Crystallography (1962). Computations were carried out on the Oxford University KDF9 computer with the Oxford University Crystallographic Programs written by O. J. Hodder, G. Ford and J. S. Rollett.

Table 2. *Atomic parameters with the estimated standard deviation of the last digit in parentheses*

(*a*) GlcNAc-Asn, $3H_2O$

(i) Non-hydrogen atom fractional co-ordinates and anisotropic temperature factors. Temperature factors are of the form

$$T = \exp[-2\pi^2(U_{11}h^2a^{*2} + U_{22}k^2b^{*2} + U_{33}l^2c^{*2} + 2U_{23}klb^*c^* + 2U_{13}hla^*c^* + 2U_{12}hka^*b^*)]$$

where a^*, b^* and c^* are the reciprocal lattice axial lengths.

Atom	$10^4x/a$	$10^4y/b$	10^4z/c	10^3U_{11}	10^3U_{22}	10^3U_{33}	10^32U_{23}	10^32U_{13}	10^32U_{12}
$O_{(3)}$	−638(9)	12(4)	1700(2)	79(3)	24(1)	31(2)	−1(2)	−34(3)	16(2)
$O_{(4)}$	2668(9)	2929(4)	1233(2)	79(3)	36(1)	21(1)	5(2)	23(2)	40(3)
$O_{(5)}$	1711(6)	4325(3)	2665(–)	44(2)	25(1)	17(1)	2(2)	6(2)	16(2)
$O_{(6)}$	126(8)	6642(3)	1880(2)	66(3)	26(1)	25(1)	−5(2)	−19(2)	29(2)
$O_{(7)}$	−4392(9)	−658(5)	3137(2)	35(3)	45(2)	74(2)	28(3)	20(3)	19(3)
$O_{(9)}$	4890(7)	3293(4)	3900(2)	29(3)	60(2)	25(1)	−4(2)	1(2)	0(3)
$O_{(12)}$	−890(9)	4990(5)	5404(2)	58(3)	57(2)	35(2)	−32(3)	2(3)	20(3)
$O_{(13)}$	2286(13)	4001(10)	5912(2)	68(4)	153(5)	24(2)	−25(4)	2(3)	56(6)
$O_{(14)}$	−978(21)	7972(11)	4730(3)	118(7)	102(5)	82(4)	15(7)	−11(7)	60(8)
$O_{(15)}$	4517(20)	9465(10)	4336(3)	143(7)	102(4)	62(3)	24(6)	44(6)	115(8)
$O_{(16)}$†	3561(41)	9365(22)	5758(5)	156(12)	127(9)	51(5)	79(9)	110(8)	148(15)
$O_{(17)}$†	9378(34)	648(12)	5514(4)	160(13)	53(4)	34(4)	8(7)	−29(10)	6(10)
$N_{(1)}$	746(8)	2863(4)	3481(2)	34(2)	41(2)	15(1)	4(2)	−4(2)	14(3)
$N_{(2)}$	−1(8)	−404(4)	2873(2)	37(2)	24(1)	25(1)	15(2)	−1(2)	19(2)
$N_{(11)}$	4489(10)	2519(6)	5089(2)	45(3)	69(3)	28(2)	−3(3)	7(3)	28(4)
$C_{(1)}$	1798(9)	2701(4)	2932(2)	38(3)	24(2)	16(1)	4(2)	−4(2)	7(2)
$C_{(2)}$	44(8)	1281(4)	2609(2)	31(2)	24(2)	20(2)	9(2)	−3(2)	12(2)
$C_{(3)}$	1144(10)	1218(4)	2015(2)	48(3)	20(1)	19(2)	0(2)	−1(3)	15(3)
$C_{(4)}$	1270(9)	2998(4)	1750(2)	47(3)	24(1)	17(2)	6(2)	10(3)	23(3)
$C_{(5)}$	2876(9)	4379(5)	2117(2)	33(3)	24(1)	22(2)	9(2)	9(2)	4(2)
$C_{(6)}$	2820(11)	6209(5)	1905(2)	50(3)	25(2)	27(2)	2(3)	1(3)	5(3)
$C_{(7)}$	−2233(10)	−1266(5)	3104(2)	38(3)	28(2)	32(2)	5(3)	−2(3)	13(3)
$C_{(8)}$	−1949(14)	−3037(6)	3321(3)	62(4)	27(2)	61(3)	23(4)	37(4)	15(3)
$C_{(9)}$	2443(9)	3166(5)	3921(2)	31(3)	32(3)	18(2)	3(3)	−2(2)	18(3)
$C_{(10)}$	879(10)	3271(6)	4465(2)	27(3)	57(2)	22(2)	−10(3)	−1(3)	8(3)
$C_{(11)}$	2771(9)	3852(6)	4942(2)	29(3)	50(2)	21(2)	−12(3)	2(2)	−11(3)
$C_{(12)}$	1234(11)	4310(7)	5469(2)	38(3)	56(2)	26(2)	−21(4)	6(3)	−5(4)

† Has an occupation factor of 0.5.

(ii) Hydrogen atom fractional co-ordinates and isotropic temperature factors.

	$10^3x/a$	$10^3y/b$	$10^3z/c$	10^3U_{iso}
$H_{(1)}$	355	241	298	22
$H_{(2)}$	−160	166	258	25
$H_{(3)}$	275	78	206	27
$H_{(4)}$	−75	322	170	27
$H_{(5)}$	475	409	214	22
$H_{(6)}$	350	634	156	34
$H_{(7)}$	390	697	222	34
$H_{(8)}$	−340	−381	329	48
$H_{(9)}$	−50	−294	359	48
$H_{(10)}$	−100	−328	302	48
$H_{(11)}$	−45	391	443	34
$H_{(12)}$	−15	219	453	34
$H_{(13)}$	405	475	490	33
$H_{(14)}$	−100	281	354	27
$H_{(15)}$	140	−88	289	26
$H_{(16)}$	610	316	520	42
$H_{(17)}$	540	209	490	41
$H_{(18)}$	300	172	526	41
$H_{(19)}$	50	−94	176	40
$H_{(20)}$	325	416	104	40
$H_{(21)}$	−60	647	164	32

Table 2.—*continued*

(*b*) Glc-Asn, H_2O

(i) Non-hydrogen atom fractional co-ordinates and anisotropic temperature factors. Temperature factors are of the form

$$T = \exp[-2\pi^2(U_{11}h^2a^{*2} + U_{22}k^2b^{*2} + U_{33}l^2c^{*2} + 2U_{23}klb^*c^* + 2U_{13}hla^*c^* + 2U_{12}hka^*b^*)$$

Atom	$10^4x/a$	$10^4y/b$	$10^4z/c$	10^3U_{11}	10^3U_{22}	10^32U_{33}	10^32U_{23}	10^32U_{13}	10^32U_{12}
$O_{(2)}$	11009(17)	−1661(9)	4153(6)	50(6)	34(4)	18(4)	9(7)	7(7)	18(7)
$O_{(3)}$	10882(15)	−3721(6)	2712(7)	64(5)	13(3)	29(3)	−5(6)	−2(9)	28(6)
$O_{(4)}$	10988(19)	−1687(9)	1262(6)	73(7)	23(3)	26(4)	8(6)	47(8)	28(7)
$O_{(5)}$	7333(14)	646(7)	2695(–)	49(5)	23(3)	15(3)	1(6)	14(8)	13(6)
$O_{(6)}$	7803(22)	2916(8)	1386(7)	95(8)	17(3)	38(4)	13(6)	15(9)	34(8)
$O_{(9)}$	8224(16)	3413(7)	4089(7)	49(5)	16(3)	38(4)	−14(6)	16(8)	−7(6)
$O_{(12)}$	7834(20)	4921(8)	5928(6)	82(7)	21(3)	29(4)	−3(6)	−25(8)	15(7)
$O_{(13)}$	7087(19)	6901(8)	5046(7)	76(7)	22(3)	27(4)	1(6)	−11(8)	−17(7)
$O_{(14)}$	5588(18)	4643(9)	2595(7)	65(7)	51(5)	33(5)	−13(8)	−22(8)	47(8)
$N_{(1)}$	6824(20)	683(8)	4073(6)	59(7)	12(3)	23(4)	−7(7)	4(9)	1(7)
$N_{(11)}$	3342(19)	5052(9)	4209(7)	40(6)	24(4)	21(4)	3(7)	−15(8)	6(7)
$C_{(1)}$	8705(23)	339(11)	3437(8)	32(7)	19(4)	15(5)	7(8)	1(9)	9(8)
$C_{(2)}$	9305(23)	−1468(11)	3473(8)	35(7)	20(5)	17(5)	−9(8)	−18(9)	17(8)
$C_{(3)}$	10824(19)	−1969(9)	2723(8)	27(6)	15(4)	25(4)	6(9)	1(10)	10(7)
$C_{(4)}$	9392(24)	−1435(11)	1951(8)	43(8)	17(5)	19(5)	−7(7)	11(10)	11(9)
$C_{(5)}$	9131(23)	457(11)	2025(8)	26(7)	22(5)	19(5)	5(7)	6(9)	4(8)
$C_{(6)}$	7820(24)	1161(11)	1286(7)	49(8)	21(4)	24(5)	17(8)	−17(10)	−4(9)
$C_{(9)}$	6699(24)	2249(12)	4332(7)	45(8)	24(5)	19(5)	−3(7)	−42(8)	21(9)
$C_{(10)}$	4526(23)	2526(10)	4928(7)	45(8)	13(4)	20(5)	−5(7)	9(9)	−8(8)
$C_{(11)}$	4211(27)	4365(11)	5010(8)	70(9)	12(4)	21(5)	9(7)	5(11)	29(9)
$C_{(12)}$	6498(23)	5461(11)	5359(8)	36(7)	15(4)	23(5)	−11(7)	−6(9)	6(8)

(ii) Hydrogen atom fractional co-ordinates and isotropic temperature factors.

Atom	$10^3x/a$	$10^3y/b$	$10^3z/c$	10^3U_{iso}	Atom	$10^3x/a$	$10^3y/b$	$10^3z/c$	10^3U_{iso}
$H_{(1)}$	1090	84	343	21	$H_{(11)}$	535	−22	417	27
$H_{(2)}$	735	−194	352	18	$H_{(12)}$	280	603	422	20
$H_{(3)}$	1185	−119	260	23	$H_{(13)}$	175	453	404	20
$H_{(4)}$	750	−203	188	16	$H_{(14)}$	435	491	383	20
$H_{(5)}$	1080	83	212	20	$H_{(15)}$	950	−291	442	28
$H_{(6)}$	885	72	82	39	$H_{(16)}$	970	−447	265	32
$H_{(7)}$	630	50	125	39	$H_{(17)}$	1130	−275	128	29
$H_{(8)}$	540	219	540	25	$H_{(18)}$	910	369	132	37
$H_{(9)}$	290	209	477	25	$H_{(19)}$	365	453	243	43
$H_{(10)}$	295	467	533	31	$H_{(20)}$	665	417	221	43

Results

The values of the observed and final calculated structure amplitudes are given in the Supplementary Publication (SUP 50035). The final atomic parameters are given in Table 2; the bond lengths and angles are given in Table 3. The relevant dihedral angles are in Table 4. Table 5 contains the hydrogen-bonding parameters of the two structures. Figs. 2(*a*) and 2(*b*) are diagrams of the two molecules, GlcNAc-Asn and Glc-Asn, respectively, viewed along the $-a$ axes of the respective unit cells. The crystal structure (hydrogen atoms are omitted) of GlcNAc-Asn trihydrate is shown projected down the *b* axis of the unit cell in Fig. 3(*a*); Fig. 3(*b*) is a projection of the crystal structure (omitting hydrogen atoms) of Glc-Asn monohydrate down the *a* axis of the unit cell. Hydrogen bonds are denoted by dashed lines in Figs. 2 and 3.

Discussion

The X-ray analyses have shown that the glucopyranose rings of GlcNAc-Asn and Glc-Asn both have the C-1 chair conformation and also that the glucose–asparagine linkage of each molecule is present in the β-anomeric configuration (Fig. 2). These facts agree with the earlier chemical assignments (Marshall & Neuberger, 1972) which were based on the structures of the intermediates used in the

Table 3. *Bond dimensions*

Bond distances (nm) with the estimated standard deviation of the last digit in parentheses.

	GlcNAc-Asn	Glc-Asn
$C_{(1)}$-$C_{(2)}$	0.1524(5)	0.152(1)
$C_{(2)}$-$C_{(3)}$	0.1543(6)	0.154(2)
$C_{(3)}$-$C_{(4)}$	0.1619(5)	0.155(2)
$C_{(4)}$-$C_{(5)}$	0.1532(5)	0.155(1)
$C_{(5)}$-$C_{(6)}$	0.1516(5)	0.153(2)
$C_{(7)}$-$C_{(8)}$	0.1497(6)	—
$C_{(1)}$-$N_{(1)}$	0.1441(6)	0.146(2)
$C_{(2)}$-$N_{(2)}$	0.1455(5)	—
$C_{(7)}$-$N_{(2)}$	0.1334(6)	—
$C_{(2)}$-$O_{(2)}$	—	0.143(2)
$C_{(3)}$-$O_{(3)}$	0.1420(5)	0.142(1)
$C_{(4)}$-$O_{(4)}$	0.1435(6)	0.142(1)
$C_{(6)}$-$O_{(6)}$	0.1417(6)	0.143(1)
$C_{(1)}$-$O_{(5)}$	0.1424(4)	0.144(1)
$C_{(5)}$-$O_{(5)}$	0.1446(4)	0.145(1)
$C_{(7)}$-$O_{(7)}$	0.1226(6)	—
$N_{(1)}$-$C_{(9)}$	0.1360(6)	0.134(1)
$C_{(9)}$-$O_{(9)}$	0.1201(6)	0.121(1)
$C_{(9)}$-$C_{(10)}$	0.1536(6)	0.150(2)
$C_{(10)}$-$C_{(11)}$	0.1517(6)	0.152(1)
$C_{(11)}$-$N_{(11)}$	0.1468(7)	0.153(2)
$C_{(11)}$-$C_{(12)}$	0.1554(7)	0.148(2)
$C_{(12)}$-$O_{(12)}$	0.1246(7)	0.126(2)
$C_{(12)}$-$O_{(13)}$	0.1231(7)	0.128(1)

preparation of 4-*N*-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine. The dimensions (the estimated standard deviations of the last digit are in parentheses) of the glycosidic bond in GlcNAc-Asn and Glc-Asn are, respectively, $C_{(1)}$-$N_{(1)}$ 0.1441(6) nm, 0.146(2) nm; angle $O_{(5)}$-$C_{(1)}$-$N_{(1)}$ 106.8(3)°, 105.7(8)°; angle $C_{(2)}$-$C_{(1)}$-$N_{(1)}$ 111.1(4)°, 110.4(9)°; angle $C_{(1)}$-$N_{(1)}$-$C_{(9)}$ 121.4(4)°, 120.5(9)°. Thus the bond distances and angles of the carbohydrate–peptide linkage in both GlcNAc-Asn and Glc-Asn are equivalent to within the estimated standard deviations. The torsion angles about the $N_{(1)}$-$C_{(1)}$ bond differ by 16.6°, however; the torsion angle $C_{(9)}$-$N_{(1)}$-$C_{(2)}$ is 141.0° and 157.6° and the torsion angle $C_{(9)}$-$N_{(1)}$-$C_{(1)}$-$O_{(5)}$ is −100.2° and −83.6° in GlcNAc-Asn and Glc-Asn respectively. This torsion-angle difference is a result of $O_{(9)}$ and $N_{(1)}$ being involved in different hydrogen-bonding patterns in the two compounds (Table 5).

The bond dimensions of the asparagine moiety in GlcNAc-Asn and Glc-Asn are similar to those found in the neutron-diffraction study of L-asparagine monohydrate (Verbist *et al.*, 1972) and also to those obtained in the X-ray structure determination of glycyl-L-asparagine (Pasternak *et al.*, 1954). However, one torsion angle (the other torsion angles agree within 22°) of the asparagine moiety is very different in GlcNAc-Asn from that occurring in Glc-Asn and

Table 4. *Dihedral angles*

The estimated standard deviations of the dihedral angles are 0.5° and 0.9° for those of GlcNAc-Asn and Glc-Asn respectively except for the estimated standard deviations of θ_N which are 3° and 5°, for GlcNAc-Asn and Glc-Asn respectively. The convention and notation of the peptide torsional angles is that of Edsall *et al.* (1966) whereas the notation of Ramachandran *et al.* (1973) is used for the dihedral angles θ_C' and θ_N. The value of the dihedral angle θ_C' ($C_{(1)}^{\alpha}$, 0; C'N) is 180° less than the clockwise (positive) rotation of $C_{(1)}^{\alpha}$ to O, viewed along the N-C' axis of a peptide unit; the value of the dihedral angle θ_N ($C_{(2)}^{\alpha}$, H;NC') is 180° less than the clockwise rotation of $C_{(2)}^{\alpha}$ to H viewed along the C'-N axis.

	GlcNAc-Asn	Glc-Asn
(*a*) Glucose residue		
$O_{(5)}$-$C_{(1)}$-$C_{(2)}$-$C_{(3)}$	59.5°	53.8°
$C_{(1)}$-$C_{(2)}$-$C_{(3)}$-$C_{(4)}$	−55.3	−48.9
$C_{(2)}$-$C_{(3)}$-$C_{(4)}$-$C_{(5)}$	53.1	54.2
$C_{(3)}$-$C_{(4)}$-$C_{(5)}$-$O_{(5)}$	−55.1	−65.6
$C_{(4)}$-$C_{(5)}$-$O_{(5)}$-$C_{(1)}$	61.3	71.5
$C_{(5)}$-$O_{(5)}$-$C_{(1)}$-$C_{(2)}$	−63.4°	−65.1
(*b*) Asparagine moiety		
$\psi_1(N_{(11)}$-$C_{(11)}$-$C_{(12)}$-$O_{(12)})$	201.3°	194.5°
$\psi_2(N_{(11)}$-$C_{(11)}$-$C_{(12)}$-$O_{(13)})$	23.1	12.1
$\chi_1(N_{(11)}$-$C_{(11)}$-$C_{(10)}$-$C_{(9)})$	−69.0	61.9°
$\chi_{21}(C_{(11)}$-$C_{(10)}$-$C_{(9)}$-$N_{(1)})$	187.9	166.5
$\chi_{22}(C_{(11)}$-$C_{(10)}$-$C_{(9)}$-$O_{(9)})$	9.8	−12.0
(*c*) Amide groups		
$\theta_C'(C_{(10)}, O_{(9)}; C_{(9)}N_{(1)})$	2.0°	−1.6°
$\theta_N(C_{(1)}, H_{(N1)}; N_{(1)}C_{(9)})$	2.3	18.9
$\omega(C_{(10)}$-$C_{(9)}$-$N_{(1)}$-$C_{(1)})$	180.6	174.0
$\theta_C'(C_{(8)}, O_{(7)}; C_{(7)}N_{(2)})$	−0.1°	—
$\theta_N(C_{(2)}, H_{(N2)}; N_{(2)}C_{(7)})$	−0.7°	—
$\omega(C_{(8)}$-$C_{(7)}$-$N_{(2)}$-$C_{(2)})$	175.8	—
$C_{(9)}$-$N_{(1)}$-$C_{(1)}$-$O_{(5)}$	−100.2°	−83.6°
$C_{(9)}$-$N_{(1)}$-$C_{(1)}$-$C_{(2)}$	141.0	157.6
$C_{(7)}$-$N_{(2)}$-$C_{(2)}$-$C_{(1)}$	112.7	—
$C_{(7)}$-$N_{(2)}$-$C_{(2)}$-$C_{(3)}$	−125.6	—

also from that in L-asparagine monohydrate. This torsion angle, $\chi_1(N_{(11)}$-$C_{(11)}$-$C_{(10)}$-$C_{(9)})$, is −69.0° in GlcNAc-Asn, 61.9° in Glc-Asn and 72.2° in L-asparagine monohydrate. The occurrence of these two staggered conformations about the $C_{(11)}$-$C_{(10)}$ bond indicate that the asparagine moiety can rotate approximately 130° about this bond in order to adjust the NH_3^+ group to different hydrogen-bonding environments. The $N_{(11)}\cdots O_{(13)}$ intramolecular distance is 0.261 nm in GlcNAc-Asn and 0.265 nm in Glc-Asn, but although there is an electrostatic attraction between the NH_3^+ and the CO_2^- groups, such contacts are common between the α-amino group and the carboxyl group in amino acids and should not really be considered as hydrogen bonds (Koetzle *et al.*, 1972).

Table 5. *Intermolecular distances and relevant angles*

(*a*) GlcNAc-Asn, $3H_2O$

(i) Between molecules of GlcNAc-Asn

$O_{(6)}\cdots O_{(13^{I})}$	0.265 nm
$H[O_{(6)}]\cdots O_{(13^{I})}$	0.197 nm
Angle $O_{(6)}$-H$\cdots O_{(13^{I})}$	174°
$O_{(3)}\cdots O_{(6^{II})}$	0.273 nm
$H[O_{(3)}]\cdots O_{(6^{II})}$	0.189 nm
Angle $O_{(3)}$-H$\cdots O_{(6^{II})}$	140°
$O_{(4)}\cdots O_{(12^{I})}$	0.280 nm
$H[O_{(4)}]\cdots O_{(12^{I})}$	0.209 nm
Angle $O_{(4)}$-H$\cdots O_{(12^{I})}$	121°
$N_{(2)}\cdots O_{(7^{II})}$	0.288 nm
$H[N_{(2)}]\cdots O_{(7^{II})}$	0.215 nm
Angle $N_{(2)}$-H$\cdots O_{(7^{II})}$	147°
$N_{(11)}\cdots O_{(12^{III})}$	0.288 nm
$H_{(16)}\cdots O_{(12^{III})}$	0.194 nm
Angle $N_{(11)}$-$H_{(16)}\cdots O_{(12^{III})}$	157°
$N_{(1)}\cdots O_{(9^{IV})}$	0.313 nm
$H[N_{(1)}]\cdots O_{(9^{IV})}$	0.229 nm
Angle $N_{(1)}$-H$\cdots O_{(9^{IV})}$	162°

(ii) Involving water molecules

$O_{(16)}\cdots O_{(17^{V})}$	0.248 nm
$O_{(3)}\cdots O_{(16^{I})}$	0.278 nm
$O_{(14)}\cdots O_{(15^{IV})}$	0.281 nm
$O_{(14)}\cdots O_{(17^{V})}$	0.281 nm
$O_{(12)}\cdots O_{(14)}$	0.284 nm
$N_{(11)}\cdots O_{(16^{II})}$	0.292 nm
$H_{(18)}\cdots O_{(16^{II})}$	0.223 nm
Angle $N_{(11)}$-$H_{(18)}\cdots O_{(16^{II})}$	125°
$N_{(11)}\cdots O_{(17^{IV})}$	0.293 nm
$H_{(18)}\cdots O_{(17^{IV})}$	0.197 nm
Angle $N_{(11)}$-$H_{(18)}\cdots O_{(17^{IV})}$	163°
$O_{(7)}\cdots O_{(15^{VI})}$	0.296 nm
$O_{(13)}\cdots O_{(17^{IV})}$	0.296 nm
$O_{(14)}\cdots O_{(15)}$	0.296 nm
$O_{(16)}\cdots O_{(17^{VII})}$	0.297 nm
$O_{(4)}\cdots O_{(16^{VIII})}$	0.298 nm
$N_{(11)}\cdots O_{(15^{II})}$	0.300 nm
$H_{(17)}\cdots O_{(15^{II})}$	0.245 nm
Angle $N_{(11)}$-$H_{(17)}\cdots O_{(15^{II})}$	131°

Symmetry codes

I	$-x$	$1-y$	$-\frac{1}{2}+z$
II	x	$-1+y$	z
III	$1+x$	y	z
IV	$-1+x$	y	z
V	$-1+x$	$1+y$	z
VI	$-1+x$	$-1+y$	z
VII	x	$1+y$	z
VIII	$1-x$	$1-y$	$-\frac{1}{2}+z$
IX	$1-x$	$-y$	$-\frac{1}{2}+z$

(*b*) Glc-Asn, H_2O

$O_{(2)}\cdots O_{(13^{I})}$	0.262 nm
$H[O_{(2)}]\cdots O_{(13^{I})}$	0.169 nm
Angle $O_{(2)}$-H$\cdots O_{(13^{I})}$	170°
$N_{(11)}\cdots O_{(9^{II})}$	0.273 nm
$H_{(13)}\cdots O_{(9^{II})}$	0.188 nm
Angle $N_{(11)}$-$H_{(13)}\cdots O_{(9^{II})}$	159°
$O_{(6)}\cdots O_{(12^{III})}$	0.274 nm
$H[O_{(6)}]\cdots O_{(12^{III})}$	0.190 nm

Table 5.—*continued*

(*b*) Glc-Asn, H_2O

Angle $O_{(6)}$-H$\cdots O_{(12^{III})}$	165°
$O_{(14)}\cdots O_{(6)}$	0.274 nm
$H_{(20)}\cdots O_{(6)}$	0.183 nm
Angle $O_{(14)}$$H_{(20)}\cdots O_{(6)}$	164°
$O_{(4)}\cdots O_{(12^{IV})}$	0.279 nm
$H[O_{(4)}]\cdots O_{(12^{IV})}$	0.194 nm
Angle $O_{(4)}$-H$\cdots O_{(12^{IV})}$	160°
$O_{(14)}\cdots O_{(3^{V})}$	0.280 nm
$H_{(19)}\cdots O_{(3^{V})}$	0.212 nm
Angle $O_{(14)}$-H$\cdots O_{(3^{V})}$	124°
$O_{(3)}\cdots O_{(14^{I})}$	0.281 nm
$H[O_{(3)}]\cdots O_{(14^{I})}$	0.208 nm
Angle $O_{(3)}$-H$\cdots O_{(14^{I})}$	151°
$N_{(11)}\cdots O_{(14)}$	0.294 nm
$H_{(14)}\cdots O_{(14)}$	0.217 nm
Angle $N_{(11)}$-$H_{(14)}\cdots O_{(14)}$	158°
$N_{(11)}\cdots O_{(2^{V})}$	0.301 nm
$H_{(12)}\cdots O_{(2^{V})}$	0.215 nm
Angle $N_{(11)}$-$H_{(12)}\cdots O_{(2^{V})}$	173°
$N_{(1)}\cdots O_{(2^{II})}$	0.327 nm
$H[N_{(1)}]\cdots O_{(2^{II})}$	0.233 nm
Angle N-$_{(1)}$H$\cdots O_{(2^{II})}$	159°

Symmetry codes

I	x	$-1+y$	z
II	$-1+x$	y	z
III	$2-x$	$1-y$	$-\frac{1}{2}+z$
IV	$2-x$	$-y$	$-\frac{1}{2}+z$
V	$-1+x$	$1+y$	z

The dimensions of the glucopyranose ring in both GlcNAc-Asn and Glc-Asn do not differ significantly from the values found in crystal structures containing glucose residues (Arnott & Scott, 1972). The 2-acetamido-2-deoxyglucopyranosyl moiety in GlcNAc-Asn is similar to that in GlcNAc [2-acetamido-2-deoxy-α-D-glucopyranose (Johnson, 1966)], and also to that in GlcNAc-ONp [(2-acetamido-1-β-*p*-nitrophenyl)-2-deoxy-β-D-glucopyranose monohydrate; L. Brehm, personal communication], especially with regard to the 2-acetamido group. In all of these three structures, the $N_{(2)}$-H bond is *trans* to the $C_{(2)}$-H bond and is approximately parallel to the axial group at $C_{(1)}$, i.e. *cis* to $C_{(1)}$-H in both GlcNAc-Asn and GlcNAc-ONp and *cis* to $C_{(1)}$-$O_{(1)}$ in GlcNAc. The conformation about the $C_{(2)}$-$N_{(2)}$ bond in each of these three compounds may be a consequence of intramolecular steric factors and/or intermolecular hydrogen-bonding. Steric factors would involve the minimum energy for the intramolecular contacts such as those of $H_{(N2)}$ and $O_{(7)}$ with the axial and equatorial atoms bonded to atoms $C_{(1)}$, $C_{(2)}$ and $C_{(3)}$. In these three crystal structures, this conformation allows similar hydrogen bonds to occur between the *N*-acetyl groups at atom $C_{(2)}$ in neighbouring unit cells, i.e. the N-H$\cdots$O, amide$\cdots$amide interactions form infinite chains of hydrogen bonding along one axis of the unit cell and

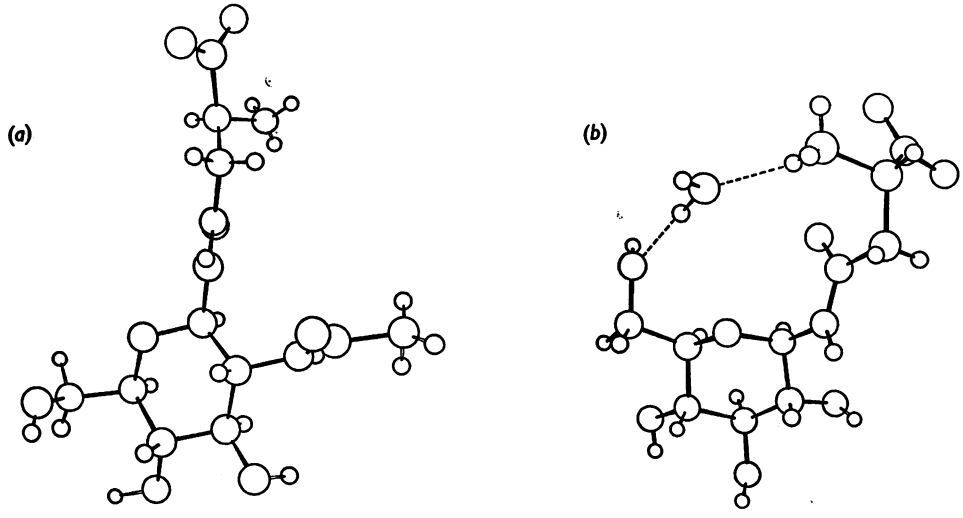

Fig. 2. *Molecular structure of (a) GlcNAc-Asn and (b) Glc-Asn*

(*a*) View of the molecule of GlcNAc-Asn down the $-a$ axis of the unit cell. (*b*) View of the molecule of Glc-Asn and the water molecule down the $-a$ axis of the unit cell; hydrogen bonds are denoted by dashed lines.

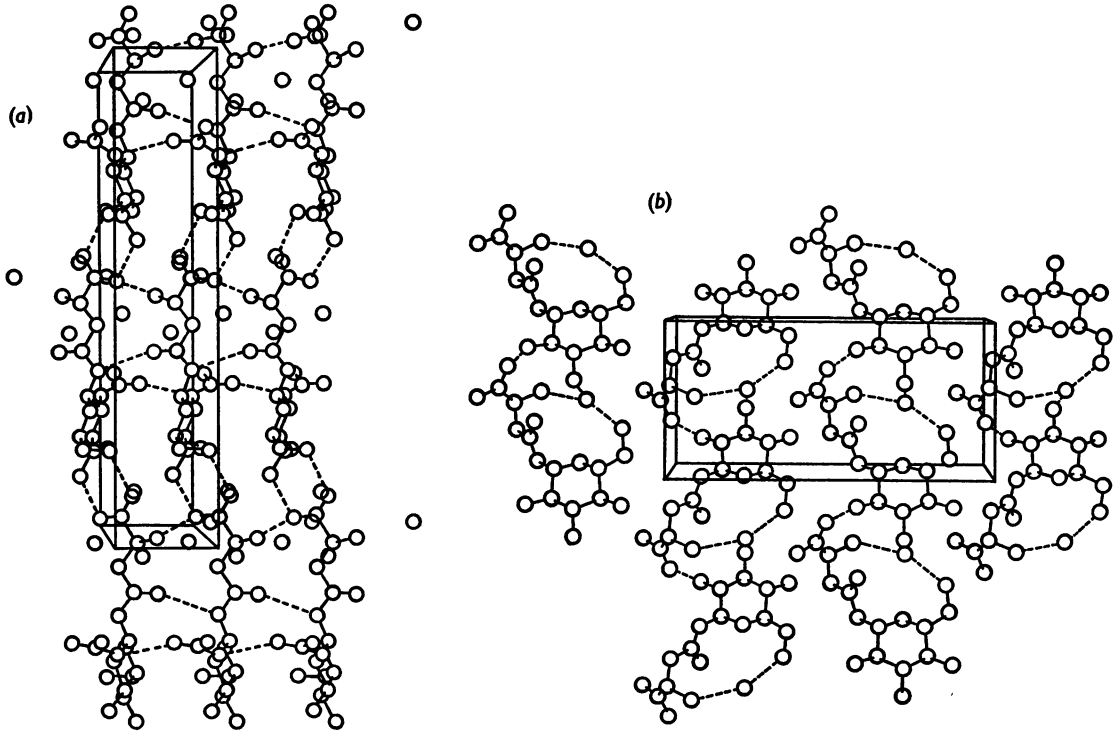

Fig. 3. *Crystal structure of (a) GlcNAc-Asn trihydrate and (b) Glc-Asn monohydrate*

(*a*) Projection of the crystal structure of GlcNAc-Asn trihydrate (hydrogen atoms are omitted) down the *b* axis of the unit cell. (*b*) Projection of the crystal structure of Glc-Asn monohydrate (hydrogen atoms are omitted) down the *a* axis of the unit cell. Hydrogen bonds are denoted by dashed lines.

the repeat period is approx. 0.5nm along this axis. The mean plane through the atoms of the amide group at $C_{(2)}$ in GlcNAc-Asn forms an angle of 69° to the mean plane through the $C_{(2)}$, $C_{(3)}$, $C_{(5)}$ and $O_{(5)}$ atoms of the glucose ring and that at $C_{(1)}$ is less perpendicular (65°). The mean plane through the atoms of the amide group in Glc-Asn makes an angle of only 55° with the mean plane through these same four atoms of the glucose ring. The $N_{(1)}$-H bond at $C_{(1)}$ is *trans* to the $C_{(1)}$-H bond in both GlcNAc-Asn and Glc-Asn.

The *a* unit-cell dimension of 0.494nm in GlcNAc-Asn trihydrate and in Glc-Asn monohydrate is similar to that occurring in the crystal structures of the *trans*-amides of 9-, 10-, 11- and 12-membered lactam rings (Winkler & Dunitz, 1971). This repeat period in these latter structures is a consequence of the amide···amide hydrogen-bonding between adjacent unit cells to form infinite chains through the crystal along this axis. In GlcNAc-Asn trihydrate, this pattern of hydrogen-bonding from the amide group at $C_{(2)}$ along the *a* axis of the unit cell has been described in the preceding paragraph; in addition, the amide group at $C_{(1)}$ demonstrates this similar hydrogen-bonding. There also is an infinite chain of hydrogen-bonding along the *a* axis owing to $N_{(11)}$-H···$O_{(12III)}$ where III refers to the symmetry code in Table 5(*a*). These three chains of hydrogen-bonding are shown in Fig. 3(*a*). In Glc-Asn monohydrate there is no corresponding amide···amide interaction at $C_{(1)}$ but the $N_{(11)}$-H···$O_{(9II)}$ and $N_{(1)}$-H···$O_{(2II)}$ interactions, where II refers to the symmetry code in Table 5(*b*), form infinite chains of hydrogen-bonding along the *a* axis.

Theoretical calculations by Ramachandran *et al.* (1973) have shown that the state of minimum energy of the *trans*-peptide unit is a non-planar conformation; in general crystal-structure data (Ramachandran & Kolaskar, 1973) agree with the predictions of this theory. The dihedral angles of the three amide groups present in GlcNAc-Asn and Glc-Asn are given in Table 4. The torsion angle ω is the conventional ω-rotation angle (Edsall *et al.*, 1966). The value of the dihedral angle $\theta_{C'}$ ($C_{(1)}^{\alpha}$, O; C′N), viewed along the N-C′ axis of a peptide unit, is 180° less than the clockwise (positive) rotation of $C_{(1)}^{\alpha}$ to O; the value of the dihedral angle θ_N ($C_{(2)}^{\alpha}$, H; NC′), viewed along the C′-N axis, is 180° less than the clockwise rotation of $C_{(2)}^{\alpha}$ to H. These values agree with the predictions that $|\theta c'|$ is small and that the greatest distortion is at the nitrogen atom of the amide moiety. The most non-planar of the amides in the structures of these two compounds is that of the aspartamido group in Glc-Asn.

The crystal density and unit-cell dimensions (Table 1) have confirmed that the compound GlcNAc-Asn crystallizes as the trihydrate under the conditions used (Marshall & Neuberger, 1972). Each of the two crystallographically independent sites [$O_{(16)}$ and $O_{(17)}$ in Table 2(*a*)] of the one disordered water molecule are involved in several hydrogen bonds (Table 5) but since the two sites are less than 0.25nm apart (sum of the van der Waals radii is 0.28nm), only one can be occupied in a given unit cell; the result is a random distribution, throughout the crystal structure, of one water molecule over the two sites.

I thank Dr. B. M. Austen, Dr. L. N. Johnson and Dr. R. D. Marshall for valuable discussions, Dr. C. K. Prout for facilities and the National Research Council of Canada for a fellowship.

References

Arnott, S. & Scott, W. E. (1972) *J. Chem. Soc. Perkin Trans. II*, 324–335

Austen, B. M. & Marshall, R. D. (1970) *Biochim. Biophys. Acta* **215**, 559–561

Austen, B. M. & Marshall, R. D. (1974) *Proc. Int. Symp. Glycoconjugates 2nd* in the press

Beddell, C. R. (1970) D.Phil. Thesis, University of Oxford

Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967) *Proc. Roy. Soc. Ser. B* **167**, 378–388

Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G. & Ramachandran, G. N. (1966) *J. Mol. Biol.* **15**, 399–407

International Tables for X-ray Crystallography (1962) (Kasper, J. S. & Lonsdale, K., eds.), vol. 3, pp. 202–203, The Kynoch Press, Birmingham

Johnson, L. N. (1966) *Acta Crystallogr.* **21**, 885–891

Karle, J. & Hauptman, H. (1956) *Acta Crystallogr.* **9**, 635–651

Karle, J. & Karle, I. L. (1966) *Acta Crystallogr.* **21**, 849–859

Koetzle, T. F., Hamilton, W. C. & Parthasarathy, R. (1972) *Acta Crystallogr. Sect. B* **28**, 2083–2089

Marshall, R. D. & Neuberger, A. (1972) in *Glycoproteins* (Gottschalk, A., ed.), pp. 453–470, Elsevier, Amsterdam, London and New York

Pasternak, R. A., Katz, L. & Corey, R. B. (1954) *Acta Crystallogr.* **7**, 225–236

Ramachandran, G. N. & Kolaskar, A. S. (1973) *Biochim. Biophys. Acta* **303**, 385–388

Ramachandran, G. N., Lakshminarayanan, A. V. & Kolaskar, A. S. (1973) *Biochim. Biophys. Acta* **303**, 8–13

Verbist, J. J., Lehmann, M. S., Koetzle, T. F. & Hamilton, W. C. (1972) *Acta Crystallogr. Sect. B* **28**, 3006–3013

Wilson, A. J. C. (1942) *Nature (London)* **150**, 151–152

Winkler, K. F. & Dunitz, J. D. (1971) *J. Mol. Biol.* **59**, 169–182