ELSEVIER

www.elsevier.com/locate/carres

Carbohydrate Research 336 (2001) 141–153

CARBOHYDRATE RESEARCH

An orthorhombic crystal form of cyclohexaicosaose, CA26·32.59 H_2O: comparison with the triclinic form

Olaf Nimz,[a] Katrin Geßler,[b] Isabel Usón,[c] Wolfram Saenger[a,*]

[a]*Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, D-14195 Berlin, Germany*
[b]*Xerox GmbH, Vor dem Lauch 15, D-70567 Stuttgart, Germany*
[c]*Institut für Anorganische Chemie, Universität Göttingen, Tammannstrasse 4, D-37077 Göttingen, Germany*

Received 4 January 2001; received in revised form 30 July 2001; accepted 14 September 2001

Abstract

Cycloamylose containing 26 glucose residues (cyclohexaicosaose, CA26) crystallized from water and 30% (v/v) polyethyleneglycol 400 in the orthorhombic space group $P2_12_12_1$ in the highly hydrated form CA26·32.59 H_2O. X-ray analysis of the crystals at 0.85 Å resolution shows that the macrocycle of CA26 is folded into two short left-handed V-amylose helices in antiparallel arrangement and related by a twofold rotational pseudosymmetry as reported recently for the $(CA26)_2$·76.75 H_2O triclinic crystal form [Geßler, K. et al. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4246–4251]. In the orthorhombic crystal form, CA26 molecules are packed in motifs reminiscent of V-amylose in hydrated and anhydrous forms. The intramolecular interface between the V-helices in CA26 is dictated by formation of an extended network of interhelical C–H···O hydrogen bonds; a comparable molecular arrangement is also evident for the intermolecular packing, suggesting that it is a characteristic feature of V-amylose interaction. The hydrophobic channels of CA26 are filled with disordered water molecules arranged in chains and held in position by multiple C–H···O hydrogen bonds. In the orthorhombic and triclinic crystal forms, the structures of CA26 molecules are equivalent but the positions of the individual water molecules are different, suggesting that the patterns of water chains are perturbed even by small structural changes associated with differences in packing arrangements in the two crystal lattices rather than with differences in the CA26 geometry.

Keywords: Cyclohexaicosaose; Cycloamylose; V-amylose; Hydrophobic channel; Disordered water chain

1. Introduction

The linear polysaccharide amylose consists of α-(1→4)-linked D-glucose units in the common $^4C_1$-chair form. It is found in starch granules as double helices with intertwined, parallel oriented strands with 6 × 2 D-glucoses per pitch of 21.6 Å. Depending on hydration, two packing modes of these double helices are found: the densely packed A-form in cereals and the more hydrated, loosely packed B-form in tubers. Since no crystal structure at atomic resolution is available for amylose fragments in the A- or B-form, the presently known structures rely largely on X-ray fiber diffraction and model building.[1,2]

Another polymorph of amylose is the V-form (V = Verkleisterung = gelation) in which the polymer is folded into a left-handed single helix with six D-glucoses per turn of 8.1 Å pitch[3] and an ~5 Å wide central channel. It is obtained from aqueous solution of amylose by drying or if small and/or slim molecules are

* Corresponding author. Tel.: +49-30-83853412; fax: +49-30-83856702.
E-mail address: saenger@chemie.fu-berlin.de (W. Saenger).

0008-6215/01/$ - see front matter
PII: S0008-6215(01)00249-X

added to the solution like iodine, fatty acids, alcohols, Me_2SO and many others.[4] These molecules form inclusion complexes with V-amylose that have been studied extensively by X-ray fiber diffraction and by spectroscopic methods, with 'iodine's blue' of interest since its discovery in 1814.[5] The only obvious requirement for inclusion formation is that these molecules as guest compounds fit into the ~ 5 Å wide channel provided by V-amylose as host.

In the absence of single crystals of amylose fragments showing V-conformation, cyclodextrins were used as model systems, the smallest α-cyclodextrin with six D-glucoses in the macrocycle simulating one turn of the V-amylose helix. Recently, very large cyclodextrins (cycloamyloses, CA) became available,[6] and the first X-ray crystal structure of one member of this family of molecules, the hydrated form of CA26 (a cycloamylose with 26 D-glucoses in the macrocycle)[7] showed that it is folded into a shape already foreseen by the Dutch painter M.C. Escher in his picture 'Bond of Union' (Fig. 1).

In this crystal structure of triclinic space group $P1$, two CA26 and 76.75 H_2O molecules are contained in the unit cell. The folding pattern of cycloamylose in CA26 is such that two short V-amylose helices are linked by two D-maltotetraose segments containing 'band-flip' motifs that are typical of cyclodextrins with ten or more D-glucoses in the macrocycle.[8,9] Of particular interest is the hydration of CA26 featuring disordered water molecules located in channels formed by the V-amylose helices.

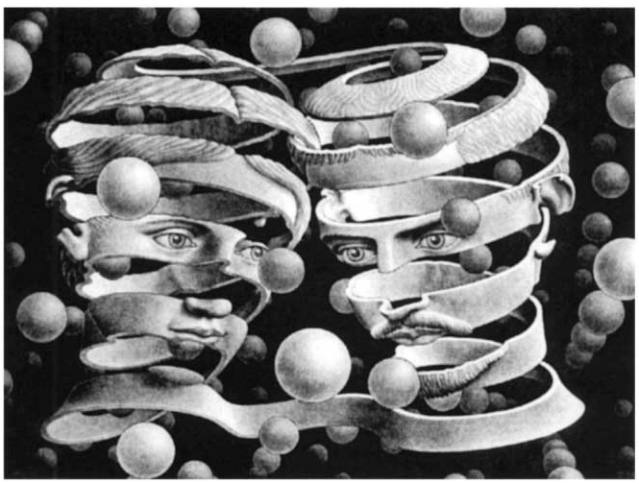


Fig. 1. The folding of the ribbon in 'Bond of Union' by M.C. Escher resembles closely the folding of the oligosaccharide chain of CA26, with some differences: one of the helices is left- and the other right-handed, there is intertwining of the helices in the top of the figure, the band-flip is missing and the two helices are slightly inclined to indicate affection. © 2000 Cordon Art B.V.–Baarn–Holland. All rights reserved.

Using identical conditions as reported previously,[7] a new form of hydrated CA26 was obtained with orthorhombic space group $P2_12_12_1$ and CA26·32.59 H_2O in the asymmetric unit. Here we compare this form with $(CA26)_2$·76.75 H_2O in the triclinic space group $P1$ with focus on the conformations of the CA26 molecules, their intermolecular interactions and the positions and disorder of the water molecules located in the central channels of the short V-helices and between the CA26 molecules.

2. Results and discussion

The overall structure of CA26 in the present orthorhombic space group is comparable to that of CA26 in the triclinic space group with two molecules in the asymmetric unit, $(CA26)_2$·76.75 H_2O. In the following, these will be referred to as CA26*o* (orthorhombic) and CA26*ta*, CA26*tb* (triclinic, molecules *a* and *b*). In all three molecular structures, CA26 is clearly divided into two halves, glucose residues G1–G13 and G14–G26, which are related to each other by a pseudo-twofold rotation axis (pseudo-$C2$) (Fig. 2).

In each of the two half segments of CA26, a 10 glucoses long stretch is folded into nearly two ($1\ ^2/_3$) turns of single-stranded left-handed helices with six glucoses per turn, characteristic of V-amylose.[3] The helices are oriented antiparallel and related by the pseudo-$C2$ axis. Their central channel-like cavities with van der Waals diameter of 5–5.5 Å are filled by disordered water molecules. Because all glucoses are oriented the same way (syn), the folding of the CA26 molecules is stabilized by intramolecular hydrogen bonds $O(3)_n\cdots O(2)_{n+1}$ between adjacent glucose residues along the chain and $O(6)_n\cdots O(2)_{n+6}/O(3)_{n+6}$ between the helical turns. By contrast, the glucose pairs G13–G14 and G26–G1 connecting the half

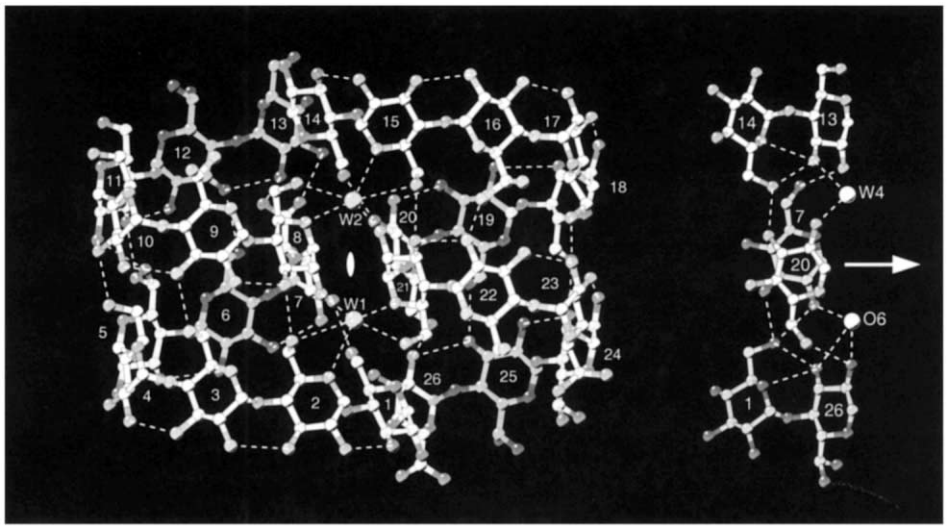


Fig. 2. Two views of CA26*o* in the orthorhombic space group, CA26·32.59 H_2O. Those parts of the molecule located at the front are light gray, those at the back dark gray, glucose residues G are numbered 1–26. Hydrogen bonds between $O(3)_n\cdots O(2)_{n+1}$ and $O(6)_n\cdots O(2)_{n+6}/O(3)_{n+6}$ are depicted as dashed lines. (Left) CA26 and water W1, W2. The location of the pseudo-twofold rotation axis (symbolized as ellipse) is between water molecules W1 and W2 and vertical to the plane of the paper. (Right) view as in left panel, but rotated 90° so that the pseudo-twofold axis is now horizontal (arrow). Shown is the interface between the two V-helices where the band-flip segments, glucoses G26–G1 and G13–G14, are connected by direct and $O(6)'_{22}$- and W4-mediated hydrogen bonds to glucoses G7 and G20.

segments of CA26 are oriented anti in a conformation termed band-flip,[8] (Fig. 3). The band-flip is stabilized by three-center (bifurcated) hydrogen bonds between $O(3)_nH$ as a donor and $O(5)_{n+1}$, $O(6)_{n+1}$ as acceptors.[9]

The two helical segments are cross-linked by four conserved water molecules W1–W4 which may be replaced by O(6) of symmetry-related CA26 molecules (see below), depending on crystal packing. W1 and W2 are located in strategic positions and confer stability to the folding of CA26 by hydrogen bonding in bidentate mode to O(5)/O(6) atoms; W1 to G21 and G2, W2 to G8 and G15, with additional single hydrogen bonds $W1\cdots O(2)_8$ and $W2\cdots O(2)_{21}$. Water W4 and $O(6)'_{22}$ (of a symmetry related CA26 molecule and in position of water W3 in the triclinic form of CA26) are coordinated at the 'back' of the molecule (right panel of Fig. 2), $O(6)_{22}$ hydrogen bonded to O(2), O(3) of G26 and O(2) of G7, and W4 hydrogen bonded to O(3) of G13 and O(2) of G20.

If the two V-helices in each of the three CA26 molecules (CA26*o*, CA26*ta*, CA26*tb*) that are formed by glucoses G3–G12 and G16–G25 are superimposed by least-squares fits, the pseudo-*C*2 symmetry is indicated by only slight deviations from the ideal 180° required for strict antiparallel arrangement, i.e. 172.3°, for CA26*o*, 179.8 and 170.5° for CA26*ta* and CA26*tb*, respectively. The structural similarity among the three CA26 is also evident from root mean square (r.m.s.) deviations, if half and whole molecules are superimposed (all O(6)- and H-atoms omitted). The r.m.s. deviations are in the range 0.31–0.54 Å for half-molecules (G1–G13 on G14–G26) and 0.55–0.82 Å for superposition of whole molecules. These data suggest that the conformations of the three crystallographically independent molecules CA26*o*, CA26*ta* and CA26*tb* are virtually identical and only little, if at all, influenced by crystal packing forces.

In the three CA26 molecules, all glucose residues are in the commonly observed $^4C_1$ chair conformation. For CA26*o*, the Cremer–Pople puckering parameters Q and θ are in the ranges 0.520–0.581 Å and 0.6–10.8°, respectively, (see Table 1 for geometrical data) indicating little or no steric strain. This is corroborated by the virtual $O(4)_n\cdots O(4)_{n+1}$ distances which fall into two groups. Nine of the 10 glucose residues in the V-helices (G4–G12 and G17–G25) show an average distance of 4.27 Å (range 4.064–4.491 Å) while most of the 2×4 glucoses in the connecting band-flip segments are longer, with virtual distances in

the range 4.424–4.790 Å, with one located 'before' (G13 and G26) and three 'after' the band-flip (the two extreme $O(4)_n \cdots O(4)_{n+1}$ distances correspond to MM3 energies of 1.53 kcal/mol for 4.064 Å and 1.39 kcal/mol for 4.790 Å; A.D. French, personal communication). The relatively undistorted geometry of CA26*o* is further supported by glycosidic linkage conformations ϕ and ψ (for definition see Table 1). In an energy map obtained for D-maltose on the basis of hybrid MM3-HF/6-31G* calculations,[11] all of the values for the three CA26 molecules are located in the global minimum at ϕ_H $-20°$, ψ_H $+20°$ (defined as ϕ_H, H1–C1–O1–C4′ and ψ_H, H4′–C4′–O1–C1[11]) within the 1.0 kcal/mol energy contour except for the glucoses around the two band-flips that correspond to a secondary minimum within a 3.0 kcal/mol contour at $-30°$, $-160°$ for ϕ_H,ψ_H. The energy barrier between this secondary minimum and the global minimum is less than 5 kcal/mol, in agreement with NMR data showing that in solution, the band-flips migrate around the macrocycle[7] (we thank A.D. French for his comments on this matter).

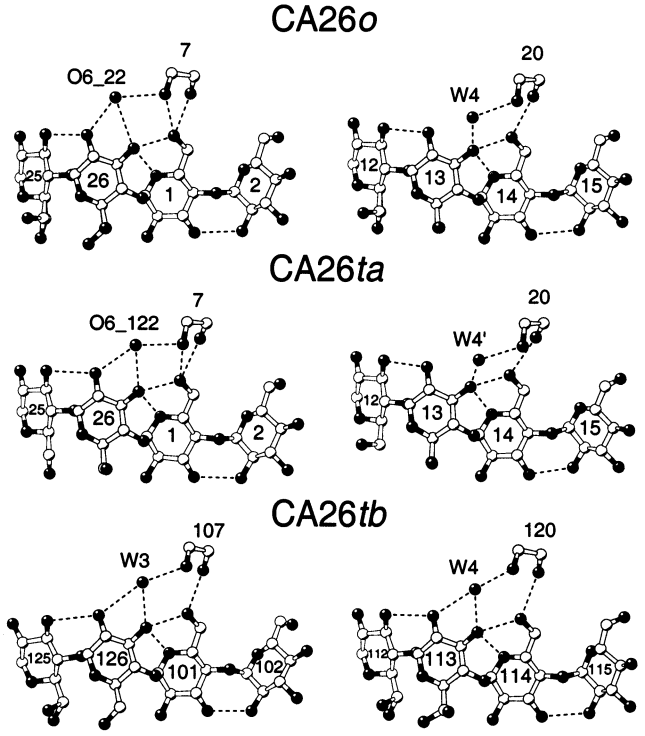


Fig. 3. Band flips in CA26*o*, CA26*ta* and CA26*tb*. The segments of four glucoses G25–G2 and G12–G15 show the band-flip sites defined by G26–G1 and G13–G14; these glucose pairs are stabilized in anti orientation by three-center hydrogen bonds $O(3)_{26}$–H$\cdots O(5)_1$,$O(6)_1$ and $O(3)_{13}$–H$\cdots O(5)_{14}$,$O(6)_{14}$. Additional stabilization is provided by formation of a hydrogen bond network involving water W4 and oxygen atom $O(6)'_{22}$ of a second CA26 molecule (or water molecule W3 in the triclinic crystal form), and O(2), O(3) of G7 or G20 in the same CA26 molecule (see also Fig. 2(B)). Adjacent glucoses on both sides of the band flips are oriented syn as usually found in amylose chains and hydrogen bonded $O(3)_n \cdots O(2)_{n+1}$.

The torsion angles χ (O(5)–C(5)–C(6)–O(6)) defining the orientation of the C(6)–O(6) bond is predominantly $(-)$gauche, sometimes $(+)$gauche and never trans, as found previously in cyclodextrin crystal structures.[13] This pattern is also observed in the three CA26, with four χ angles strictly conserved. They are systematically $(+)$gauche for G1, G7, G14, G20 (Table 2) because glucoses G1, G14 are associated with the two band-flips $O(6)_1 \cdots O(3)_{26}$/$O(3)_7$; $O(6)_{14} \cdots O(3)_{13}$/$O(3)_{20}$ (Fig. 3) and G7, G20 with the interface between the two helical segments $O(6)_7 \cdots O(2)_{13}$; $O(6)_{20} \cdots O(2)_{26}$ (Fig. 2(B)), thereby contributing to the three-dimensional folding of CA26.

The torsion angles χ of the other glucoses are predominantly $(-)$gauche or disordered $(-)/(+)$gauche, with some differences in CA26*o*, CA26*ta*, CA26*tb* because O(6) hydroxyl groups may be engaged in intermolecular hydrogen bonding which depends on crystal packing and is different for each of the three CA26 molecules.

The band-flips of CA26*o* (Fig. 3) are stabilized not only by the three-center interglucose hydrogen bonds $O(3)_{26} \cdots O(5)_1$/$O(6)_1$ and $O(3)_{13} \cdots O(5)_{14}$/$O(6)_{14}$. In addition, intramolecular hydrogen bonds $O(6)_1 \cdots O(2)_7$/$O(3)_7$ and $O(6)_{14} \cdots O(3)_{20}$ and hydrogen bonds mediated by $O(6)'_{22}$ of a second, symmetry related CA26 molecule $O(2)_7 \cdots O(6)'_{22} \cdots O(2)_{26}$/$O(3)_{26}$ and by a water molecule $O(3)_{13} \cdots W4 \cdots O(2)_{20}$ contribute to the conformational stability. This geometry is conserved in all three CA26 molecules (Fig. 3), suggesting that it is an intrinsic feature of the folding pattern characteristic of CA26.

In contrast to the smaller CA10 and CA14, the band-flips in CA26 are not associated with 'kinks', in agreement with the notion[8] that kinks will become smaller with increasing chain length of the CA because residual strain is dissipated along the chain. In CA26*ta* and

Table 1
Geometrical parameters defining glucose shape and conformation in CA26*o*: Cremer–Pople[10] parameters Q and θ, intramolecular $O(3)_n\cdots O(2)_{n+1}$ and intraglucose $O(4)_n\cdots O(4)_{n-1}$ distances, glycosidic angles $C(4)_n$–$O(4)_n$–$C(1)_{n+1}$ and torsion angles ϕ, ψ

Glucose number	Q (Å)	θ (°)	$O(4)_n\cdots O(4)_{n-1}$ (Å) [a]	$C(4)_n$–$O(4)_n$–$C(1)_{n+1}$ (°) [b]	ϕ (°)[c]	ψ (°) [c]	$O(3)_n\cdots O(2)_{n+1}$ [d] (Å)
1	0.577	10.2	4.499	117.0	102.4	112.9	2.831
2	0.560	8.0	4.705	114.7	99.8	115.7	2.856
3	0.543	3.9	4.522	118.0	106.8	120.4	2.747
4	0.548	5.2	4.206	119.2	107.3	114.8	2.794
5	0.560	7.8	4.137	119.7	97.7	108.6	3.023
6	0.544	1.0	4.340	119.4	105.6	126.2	2.958
7	0.565	7.0	4.242	117.5	106.4	108.9	2.724
8	0.541	2.3	4.345	118.1	104.4	114.0	2.861
9	0.559	6.8	4.237	121.5	101.9	117.9	2.940
10	0.524	7.9	4.252	119.3	104.1	117.1	2.843
11	0.543	1.5	4.375	116.2	104.8	114.9	2.746
12	0.544	0.6	4.394	115.4	102.3	109.2	2.898
13	0.543	3.6	4.505	117.2	87.8	−51.0	5.617
14	0.581	8.1	4.597	117.0	105.7	123.3	2.778
15	0.575	5.3	4.581	114.6	101.2	99.7	2.761
16	0.577	6.3	4.569	117.0	107.2	114.6	2.748
17	0.520	4.5	4.302	121.6	110.5	127.1	2.890
18	0.542	10.8	4.043	116.0	95.5	105.0	3.131
19	0.569	2.4	4.398	117.7	103.8	116.8	2.836
20	0.547	7.9	4.202	118.6	107.8	114.3	2.737
21	0.552	4.5	4.322	117.5	105.1	102.3	2.786
22	0.547	9.5	4.139	120.8	112.1	111.8	2.780
23	0.566	8.6	4.235	118.8	110.8	126.6	2.877
24	0.560	3.3	4.488	117.6	103.2	109.0	2.773
25	0.552	2.8	4.376	118.1	101.1	119.2	2.930
26	0.543	3.2	4.422	117.2	89.6	−47.6	5.663

[a] Average e.s.d. 0.005 Å.
[b] Average e.s.d. 0.4°.
[c] Defined as ϕ, $O(5)_{n+1}$–$C(5)_{n+1}$–$O(4)_n$–$C(4)_n$, ψ, $C(1)_{n+1}$–$O(4)_n$–$C(4)_n$–$C(3)_n$, average e.s.d. 0.5°.
[d] Average e.s.d. 0.006 Å.

CA26*tb*, some small kinks were observed just *prior* to the band-flips with distances $O(3)_{12}\cdots O(2)_{13}$ at 2.98 Å and $O(3)_{25}\cdots O(2)_{26}$ at 3.28 Å. These distances are, however, 2.88 and 2.92 Å, respectively, in CA26*o* (Table 1) and in the usual range (2.72–2.95 Å) observed for the helical segments in CA26*o*. There are two slightly larger $O(3)_n\cdots O(2)_{n+1}$ distances one turn down the chain, $O(3)_5\cdots O(2)_6$, 3.02 Å and $O(3)_{18}\cdots O(2)_{19}$, 3.13 Å, which we associate with crystal packing forces and not with the 'kinks' accompanying 'band-flips'.

Each of the two helical segments of CA26 is internally stabilized by seven hydrogen bonds formed between glucoses in adjacent turns, $O(6)_n\cdots O(2)_{n+6}$ and/or $O(6)_n\cdots O(3)_{n+6}$ (Table 3). The hydrogen bond distances depend on the torsion angle χ.[7] If $O(6)_n$ is in (+)gauche orientation, it is closer to $O(2)_{n+6}$; if it is (−)gauche, it is closer to $O(3)_{n+6}$. The only exceptions to this rule are the band-flip glucoses G1 and G14, which belong to the segments connecting the helices and are not part of the helices.

The intramolecular interface between helices is governed by C–H···O hydrogen bonds.—The two short helices in CA26 face each other across the pseudo-twofold rotation axis such that the pair of glucoses G7–G8 of one helix is in contact with the pair G20–G21 of the other helix (Fig. 4). This arrangement is geometrically well defined by short interhelix C–H···O hydrogen bonds in the range H···O, 2.53–3.03 Å, between C(1)–H and C(4)–H as donors and O(2), O(3) as acceptors and formed likewise in all three CA26 molecules. The observed H···O distances are in agreement with molecular mechanics and quantum chem-

ical calculations on the system methane–dimethyl ether which indicate that the C–H···O potential energy has a minimum in this distance range.[12]

Water molecules enclosed in the helical channels.—The helical channels formed by segments G1–G13 and G14–G26 are filled by disordered water molecules with site occupancies in the range 0.29–0.68 (Fig. 5(A–C) and Scheme 1). Within the channels, water sites are mainly stabilized by C3–H, C5–H and C6–H, each forming two to six C–H···O_W hydrogen bonds in the H···O distance range 3.13–3.59 Å (Fig. 5(B)). Hydrogen bonds between water and glucose O–H groups are rare within the channel of CA26, ($O(6)_{20}$···O_{W27}, 3.22 Å; $O(6)_{16}B$···O_{W33}, 3.34 Å), but they are found frequently close to the openings of the channels.

Table 2
O(5)–C(5)–C(6)–O(6) torsion angles χ of CA26*o*

Glc [a]	χ [b] (°)	Occ [c]	χ [b] (°)	Occ [c]
1	68.6			
2	−59.1			
3	−63.0			
4	−63.3			
5	−59.4	0.79	72.1	0.21
6	−63.3			
7	68.4	0.87	−40.8	0.13
8	−58.7			
9	−61.8	0.66	66.5	0.34
10	−57.1			
11	−60.6			
12	−60.1	0.78	61.7	0.22
13	−58.5			
14	66.5			
15	−59.4			
16	−63.2	0.78	56.6	0.22
17	60.3	0.55	−58.5	0.45
18	−67.1			
19	−63.0			
20	65.2			
21	−62.4			
22	−62.8			
23	−56.9	0.87	52.9	0.13
24	61.3	0.55	−61.5	0.45
25	−61.2	0.63	62.8	0.37
26	−54.4	0.63	64.2	0.37

[a] Number of glucose residue.
[b] O(5)–C(5)–C(6)–O(6) torsion angles χ, e.s.d. are around 0.7°.
[c] Positional disorder for O(6) is represented as major and minor occupancy, respectively.

The distances between the water sites in channels A and B are depicted in Scheme 1. They are in the range 1.01–1.79 Å in channel A and 1.16–1.91 Å in channel B, too short for normal O_W···O_W hydrogen bonds in the range 2.5–3.2 Å. These data suggest three possible interpretations:

1. Each water site is occupied at about 0.5 and mutually exclusive with the adjacent neighbor, i.e. if site m is occupied, sites $m+1$ and $m-1$ are unoccupied, and vice versa.
2. The sites attributed to water are, in fact, partially occupied PEG400 that was used for crystallization (see Experimental). This view is supported by distances between the sites which are in the range typical of C–C (1.44 Å) and C–O (1.43 Å) bonds. In addition, PEG400 forms channel-type inclusion complexes with α-cyclodextrin.[14] As shown, however, with fast atom bombardment (FAB) mass spectrometry, no PEG400 could be traced in crystals of CA26. This was confirmed by Raman spectra taken from CA26 crystals (data not shown) in which two bands of PEG400 (812 and 886 cm^{-1}) attributed to 'lattice vibrations' (700–1500 cm^{-1}) are not found.
3. The interpretation that we favor is illustrated in Scheme 1. We assume that the water sites in channels A and B appear disordered due to statistical superposition of chains of water molecules with normal O–H···O hydrogen bonding distances of 2.5–3.2 Å. These chains are indicated by vertical lines to the left and to the right of the water chains (Scheme 1), and short horizontal lines crossing the vertical lines connect every second water site, e.g. for channel A the sites W40, W42, W44, W46 etc. for one chain and the sites W41, W43, W45 etc. for the other, with O_W···O_W distances in the range 1.73–2.82 Å and 1.64–2.62 Å, respectively. Likewise, the two chains in channel B are formed by sites W23, W25, W27, W29 etc. and by sites W24, W26, W28, W30 etc. with O_W···O_W distances within 2.34–2.49 Å and 1.96–2.68 Å, respectively.

Table 3
Intramolecular $O(6)_n\cdots O(2)_{n+6}/O(3)_{n+6}$ distances <3.4 Å in CA26*o*

Residue/atom I occ [a]		χ [b]	Residue/atom [a] J		Distance [c] IJ
1	O6	+	7	O2	3.036
1	O6	+	7	O3	2.770
2	O6	−	8	O3	2.765
3	O6	−	9	O2	3.316
3	O6	−	9	O3	2.926
4	O6	−	10	O2	3.348
4	O6	−	10	O3	2.722
5	O6 0.79	−	11	O3	2.809
5	O6 0.21	+	11	O2	2.800
5	O6 0.21	+	11	O3	3.001
6	O6	−	12	O3	3.208
7	O6 0.87	+	13	O2	2.809
7	O6 0.13	−	13	O2	2.779
7	O6 0.13	−	13	O3	2.751
14	O6	+	20	O3	2.693
15	O6	−	21	O3	2.753
16	O6 0.22	+	22	O2	2.764
17	O6 0.55	+	23	O2	2.659
17	O6 0.55	+	23	O3	3.303
17	O6 0.45	−	23	O3	2.995
18	O6	−	24	O2	3.183
18	O6	−	24	O3	2.718
19	O6	−	25	O2	3.266
19	O6	−	25	O3	2.817
20	O6	+	26	O2	3.014

[a] Glucose residue and atom number of hydroxyls with occupancy of disordered components.
[b] (−) or (+) gauche orientation of O6.
[c] Distance between atoms I and J in Å., average e.s.d. around 0.005 Å.

The occupancies of the disordered water sites vary considerably, from 0.29 to 0.68 for the chain starting with site W23 in channel B and from 0.32 to 0.56 for the chain starting with W24, whereas the average occupancies are comparable, 0.50 and 0.45, respectively. This suggests that both alternative chains are equally preferred, probably because water sites are nearly equivalent energetically with up to six partners for C–H···O interactions in the cavities of the CA26 host molecules (Fig. 5(B)).

Outside the channels, water molecules are chelated to glucoses in two different motifs.— Since CA26 crystallizes as high hydrate, triclinic $(CA26)_2\cdot 76.75\ H_2O$ and orthorhombic $CA26\cdot 32.59\ H_2O$, a large number of $CA26\cdots O_W$ interactions are expected. The most frequently occurring hydration motifs are the chelation of a water molecule in bidentate mode to O(2)–H and O(3)–H or to O(6)–H and O(5) of the *same* glucose to form five-membered chelate rings, see also Refs. 8, 15, and 16. Both of these motifs occur 18 times in CA26*o*. The angles at the water oxygen, $O(3)_n\cdots O_W\cdots O(2)_n$ or $O(6)_n\cdots O_W\cdots O(5)_n$ are always in the range 49–59° and the $O_W\cdots O$ distances are in the range 2.62–3.36 Å. There is no clear preference of the $O_W\cdots O$ distances to one of the two hydrogen bonding partners of the glucose as this is dependent on additional hydrogen bonding contacts in which the water molecules O_W are engaged.

Other common motifs concern water molecules bridging adjacent glucoses along the chain, $O(6)_n\cdots O_W\cdots O(5)_{n+1}$ with a short component indicated by $O(6)_n\cdots O_W$ distances in the range of 2.57–2.83 Å (mean 2.70 Å) and a longer component for $O_W\cdots O(5)_{n+1}$ in the range 2.75–3.29 Å (mean 3.04 Å), in agreement with the notion that the ether-like O(5) is a weak hydrogen bond acceptor; the $O(6)_n\cdots O_W\cdots O(5)_{n+1}$ angle is in the range 77–87°. A number of water molecules connect

adjacent turns of the V-helix by hydrogen bonding $O(6)_n\cdots O_W\cdots O(2)_{n+6}/O(3)_{n+6}$ and $O(5)_n/O(6)_n\cdots O_W\cdots O(2)_{n+6}/O(3)_{n+6}$. Distances $O(6)_n\cdots O_W$ and $O_W\cdots O(2)_{n+6}$ are in the range 2.45–3.35 Å (mean 2.91 Å), distances $O(5)_n\cdots O_W$ in the range 2.62–3.39 Å (mean 3.01 Å) and distances $O_W\cdots O(3)_{n+6}$ in the range 3.16–3.36 Å (mean 3.28 Å).

Packing of CA26 molecules in the crystal lattice.—In the orthorhombic unit cell, CA26*o* molecules are arranged in layers parallel to the crystallographic *a*, *c* plane (Fig. 6). They are oriented such that their helical axes are perpendicular to this plane and form a hexagonal pattern that is stabilized by intermolecular direct (O···O) and water-mediated (O···O_W···O) hydrogen bonds and by C–H···O interactions. The latter are of the type and geometry described in Fig. 4, and involve G3, G4, G9, G10 from two turns of one CA26*o* which interact with two turns G4, G5, G10, G11 of a symmetry related one, according to Scheme 2(a).

Likewise, in the triclinic crystal form, similar intermolecular interactions are found for molecule CA26*tb* (Scheme 2(b)) (but not for

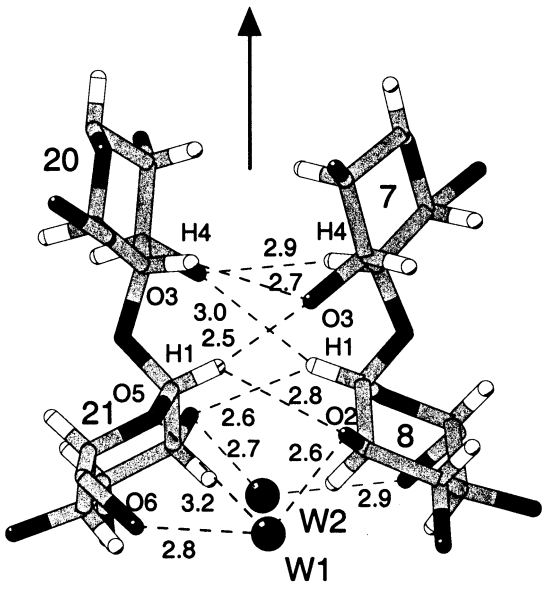
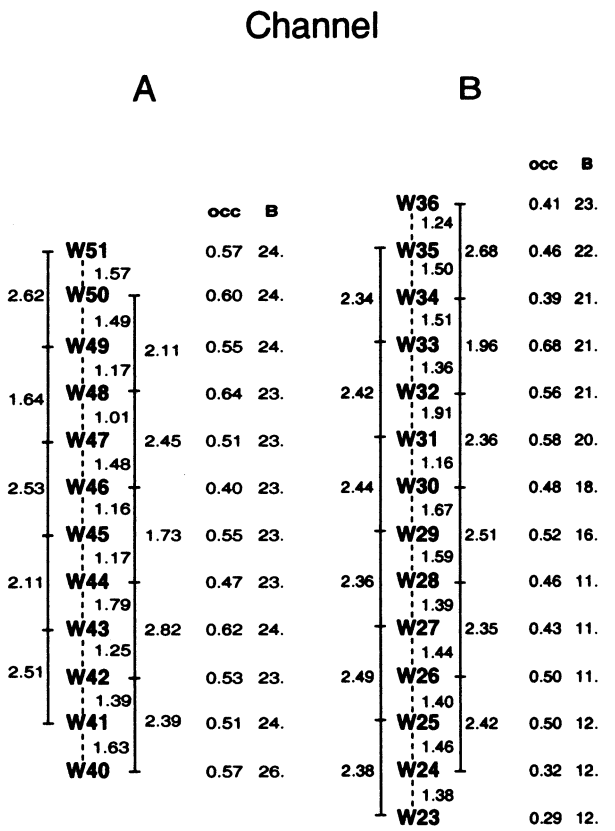


Fig. 4. Interface between the two V-amylose helices of CA26*o* which are located so close together that glucoses G7, G8 and G20, G21 interact with an extended network of C–H···O hydrogen bonds. These bonds are shown as dashed lines with H···O distances in Å. Additional hydrogen bonds with water molecules W1 and W2 (spheres) are shown. The pseudo-twofold axis of the CA26 molecule (arrow) lies in the plane of the paper.



Scheme 1. $O_W\cdots O_W$ distances of disordered water molecules in both channels G1–G13 (right, B) and G14–G26 (left, *A*). After anisotropic refinement of these water positions, the occupation factors and B-factors (Å^2) converged at values tabulated here. 1–2 distances are indicated directly between water molecules, 1–3 distances at vertical lines to the left and to the right of the water chains; small horizontal bars indicate waters for which the 1–3 distances (Å) are relevant.

CA26*ta*), suggesting that these short contacts between V-helices are of more general importance and may be of significance for packing of V-amylose helices as well.

Along the crystallographic *b*-axis, the layers of CA26*o* molecules are stacked. The axes of the short V-helices are not collinear but displaced by ~ 6 Å in the direction of the *a*-axis and form angles of $\sim 160°$ with each other, giving rise to 'infinite zigzag helices' arranged parallel to the *b*-axis. Along this axis, adjacent CA26*o* molecules are oriented head-to-head and hydrogen bonded mainly between glucose O2, O3 hydroxyls according to Scheme 3(a).

The main features of this packing scheme are also found in the triclinic space group for molecules CA26*ta* and CA26*tb*, including the hexagonal pattern formed by the short V-helices stacked in the form of 'infinite zigzag helices'. However, adjacent molecules are ori-

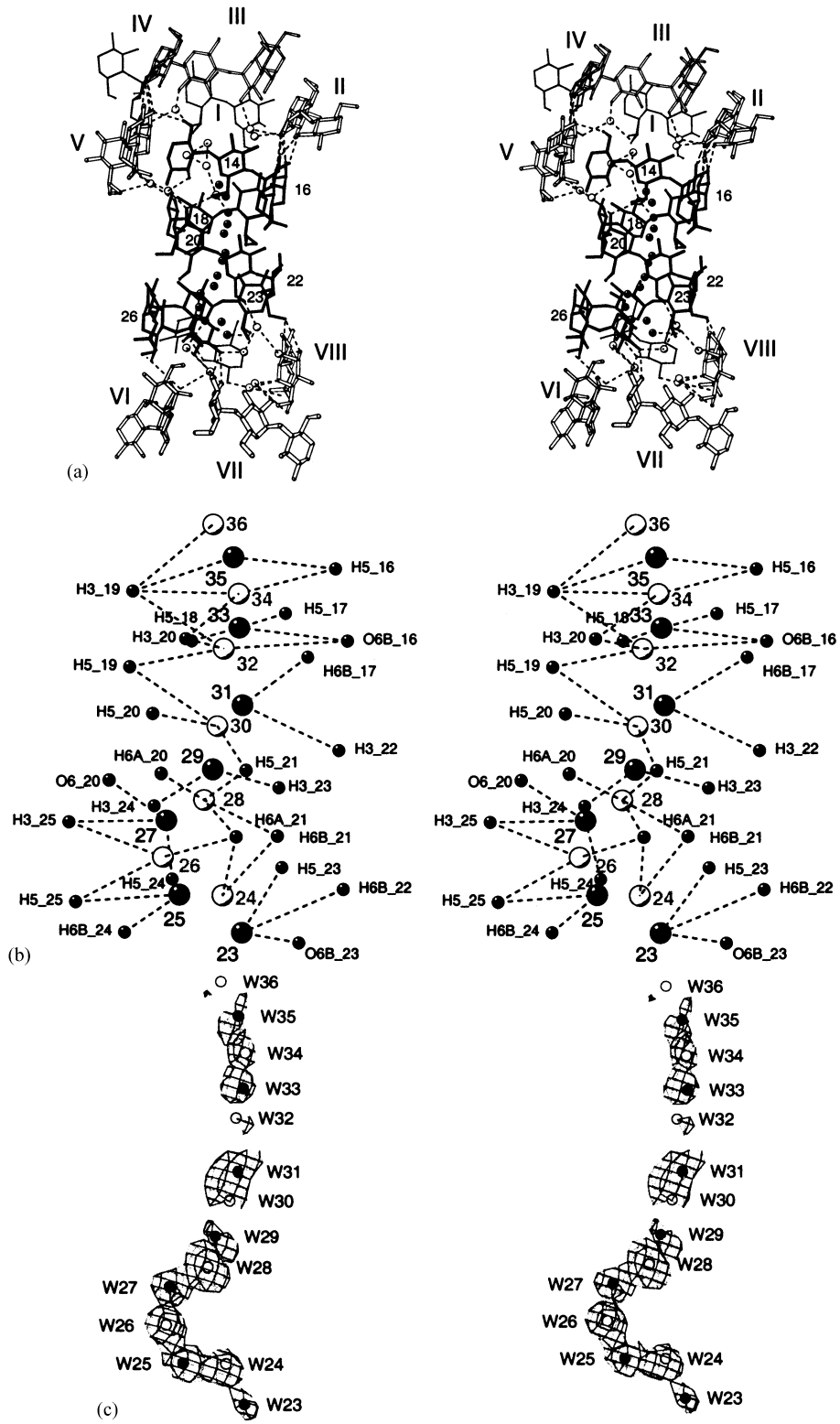


Fig. 5. Stereographic representation of disordered water molecules in the channel B formed by V-helix segment G13-G26 in CA26*o*. (a) Water molecules in the channel are represented as black and water outside the channel as white spheres, respectively. The glucose residues shown with open bonds belong to symmetry related CA26 (I–VIII). (b) The contacts (C–H···O_W and O···O_W) between water molecules and CA26 in channel G13–G1 are shown as dashed lines. Mutually exclusive chains are represented as black and white spheres, respectively. (c) Difference ($F_{obs} - F_{calc}$) electron density located in the channel (G13–G1) of CA26*o* and representing the disordered water chains. The map is contoured at 3σ level, positions of modeled water molecules are indicated for the two chains by open and filled circles (see text).

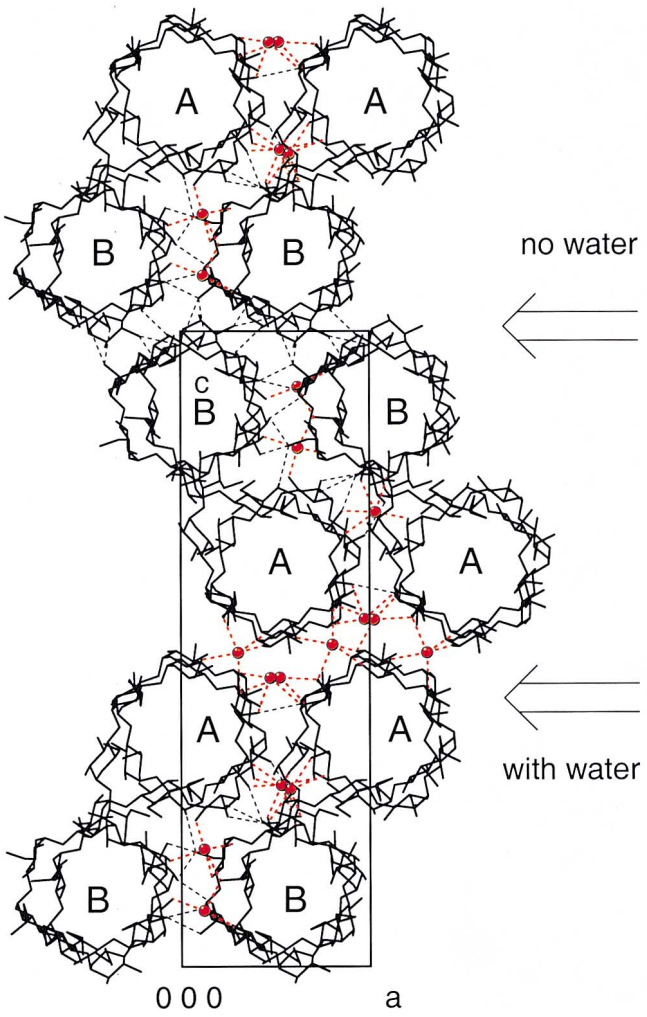
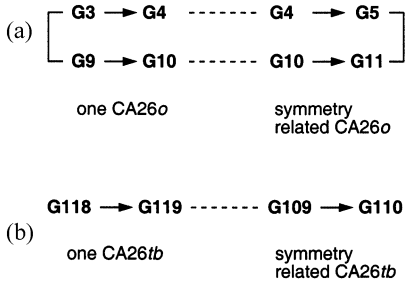


Fig. 6. Packing of six symmetry related CA26*o* molecules in the crystallographic *a*, *c*-plane. The helices form a hexagonal pattern comparable to the packing of V-amylose in fibers.[3] An extended network of direct (black dashed lines) and water-mediated hydrogen bonds (red dashed lines) interconnects the molecules. The interconnection in direction parallel to the *a*, *b*-plane follows two different schemes, with direct hydrogen bonds (at $c = 0$, indicated by arrow 'no water') or exclusively water mediated hydrogen bonding (at $c = 1/2$, indicated by arrow 'with water'). The molecular packing in direction of the *b*-axis is stabilized only by few direct intermolecular contacts and by many water-mediated-bridges. The two channels in CA26*o* are indicated by A and B.

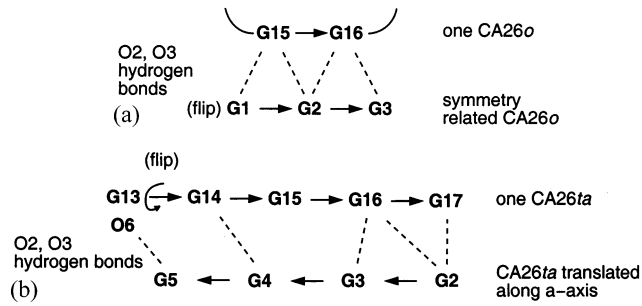


Scheme 2.

ented head-to-tail. They are arranged along the *a*-axis and hydrogen bonded between O2, O3 hydroxyl groups similar to Scheme 3(a) except for the directions of the maltooligosaccharide chains, as indicated by the short arrows (→) in Scheme 3(b). Since the band-flip occurs at G13–G14, G14 is hydrogen bonded with O2, O3 but G13 with O6.

In Fig. 7 the packing of CA26*o* molecules in the *a*, *c* plane is compared with the packing of V-amylose helices in crystalline fibers. The hexagonal arrangement of the short (CA26) and long (V-amylose) helices is similar. There is one noteworthy feature that concerns the intermolecular interactions: on one side of the CA26*o* molecule (indicated by the arrow 'with water' in Fig. 6), these interactions are mediated by hydrogen-bonded water molecules, and supported by C–H···O_W contacts. On the other side of these CA26*o* molecules (arrow 'no water' in Fig. 6), there are direct contacts of C–H···O and O–H···O–H type, without bridging water. The C–H···O contacts are of a comparable geometry as shown in Fig. 4.

The two packing modes, with and without water, are compared in Fig. 7 with the packing of V-amylose in hydrated form (V_h) and anhydrous form (V_a). There are striking similarities in the arrangement of CA26*o* and V-amylose molecules and even in the clustering of water molecules and their interaction with the surrounding carbohydrates. This is also reflected in the distances between helical axes, 12.9 Å in V_a-amylose and 13.6 Å in V_h-amylose, in good agreement with the two distances observed in CA26*o*, 13.3 and 14.2 Å, respectively.



Scheme 3.

3. Conclusions

The folding of the CA26 macrocycle into two short V-helices connected by band-flips is unexpected. It even escaped a modeling study that was supported by X-ray small angle scattering to provide gross structural information on the radius of gyration of CA molecules of different size.[17] The helical parts of CA26 show all features of V-amylose, in particular the formation of inclusion complexes with water, aliphatic (fatty) acids and alcohols and polyiodide that are accommodated in the ~5 Å wide channels of the V-amylose helix (to be

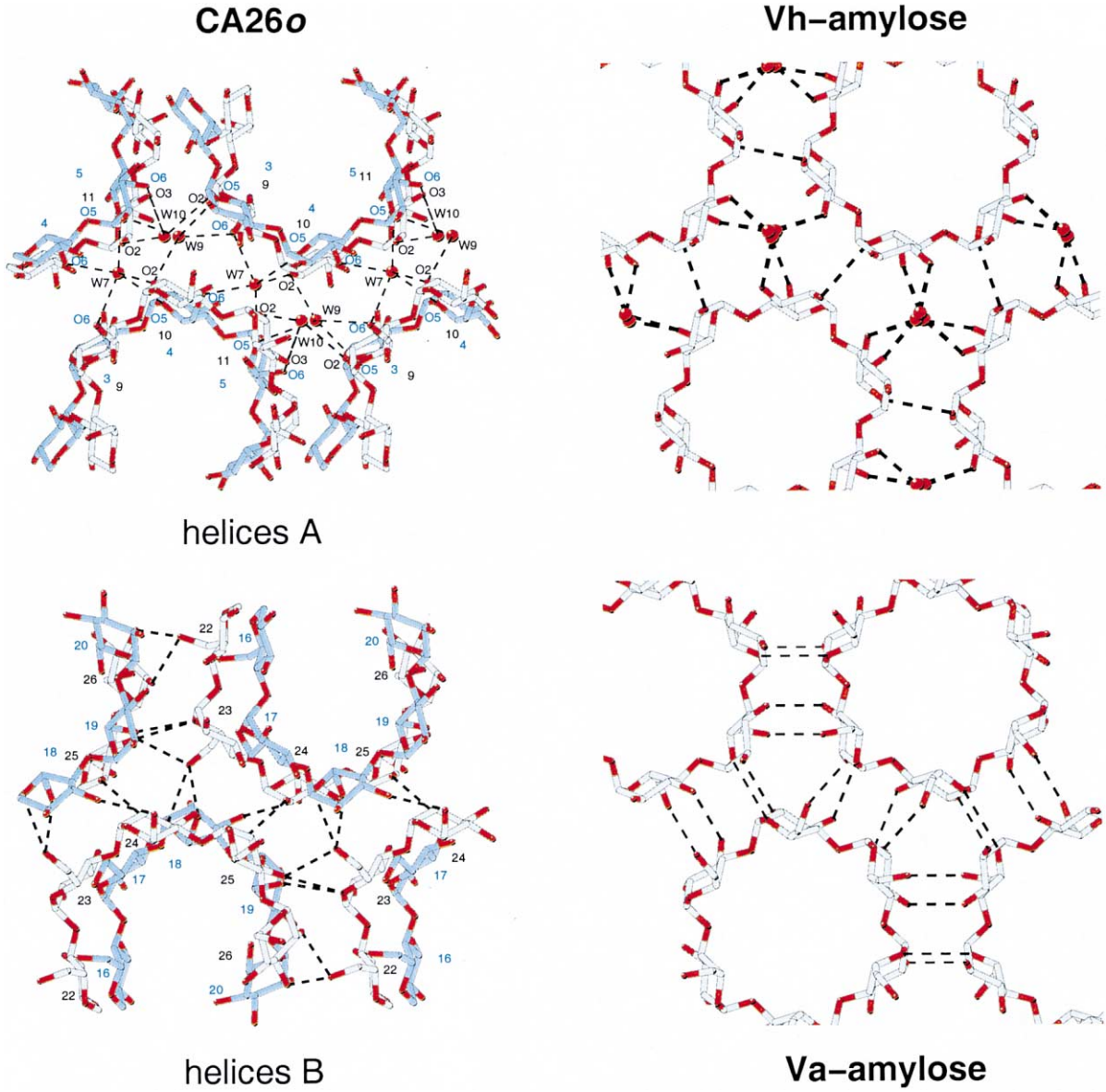


Fig. 7. Interstitial spaces in CA26*o* compared with fiber structures of V_a- and V_h-amylose. The interstitial spaces between antiparallel oriented helices A and B of CA26*o* molecules in the crystal lattice differ significantly. The interstitial space formed between symmetry related G1–G13 of CA26*o* (helices A) is filled with water and all intermolecular contacts are mediated by water (upper left), turn G2–G6 in blue and turn G8–G12 in gray. Water molecules (red spheres) located between the helices are hydrogen bonded (dashed lines). Helices B (G14–G26) are directly connected to symmetry related mates, with no water (lower left); hydrogen bonds formed between the helices are shown by dashed lines (turn G22–G26 in gray; turn G16–G20 in blue). These different modes of connectivity observed in the crystal lattice of CA26*o* resemble the situation in V-amylose fibers. V-amylose in hydrated form (V_h) shown in the upper right; red spheres are water molecules. V_h converts to the anhydrous form (V_a) after intensive drying, lower right. In the V_h form the centers of helices are 13.6 Å apart and water molecules mediate contacts between the V-helices; in the V_a form the distance between helical centers reduces to 12.9 Å and direct hydrogen bonds are formed (lower right).

published). Many attempts to crystallize larger and smaller CA failed except for thin needle crystals of CA27 (Kayo Imamura, unpublished), suggesting that the size of CA26 is unique as it folds into a stable structure not accessible for other CA.

4. Experimental

CA26 was prepared and purified as described.[7] Using the hanging drop vapor diffusion method, a drop containing 3 μL of 20 mg/mL CA26 in water (w/v) and 3 μL of 30% (v/v) PEG 400 (E. Merck) in water was equilibrated at 18 °C against a 10 mL reservoir containing 37.5% (v/v) PEG 400 in water. After several days, colorless needles of CA26 grew with dimensions $1.0 \times 0.1 \times 0.1$ mm. One of the crystals was sealed with a drop of mother liquor in a quartz capillary and used for X-ray diffraction measurements at beamline BW7B (EMBL Outstation/Deutsches Elektronen Synchrotron DESY, Hamburg), at a wavelength $\lambda = 0.97$ Å and 4 °C using a Mar Research image plate area detector. The individual images were processed with the HKL package[18] to yield 151,663 reflection intensities to a resolution of 0.85 Å. Symmetry equivalent data were merged to 23,556 unique data ($R_{sym} = 0.051$); for crystallographic data see Table 4.

The crystal structure was determined with the program SHELXD.[19] The data were extended to the triclinic hemisphere, and a rotation search was performed with the CA26*ta* molecule from the triclinic crystal structure. The maxima of the rotation function were found using E-amplitudes between 1.8 and 1.5 Å resolution (6819 reflections) and for each of these, the correlation coefficient (CC)[20] was calculated using data between 1.5 and 1.3 Å resolution. After 99 cycles of rotation function evaluation, the solution with the highest cross validation CC, based on the high resolution data, was expanded by nine cycles of progressive incorporation of more atoms and peaklist optimization[21] from bottom to top, eliminating atoms whenever this led to an improvement of the CC between observed and calculated normalized structure factors based on all data. The best solution was characterized by a CC of 83% and contained 1220 correctly positioned atoms of the four molecules present in the unit cell. Location of the symmetry elements in the cell led to the reduction of the four molecules initially located to the one present in the asymmetric unit in space group $P2_12_12_1$. After several cycles of full-matrix least-squares refinement with SHELX[22] in which the glucose interatomic distances 1–2 and 1–3 were restrained, water oxygen atoms were located from electron and difference-electron density maps using XTALVIEW.[23] A total of 32.59 water molecules is found distributed over 54 positions of which only 11 are fully occupied.

Table 4
Unit cell constants and refinement

Chemical formula	$(C_{156}H_{260}O_{130})\cdot 32.59\ H_2O$
Formula weight	4727.1
Space group	orthorhombic $P2_12_12_1$
Unit cell constants	
a (Å)	13.823(3)
b (Å)	41.023(8)
c (Å)	46.511(9)
α (°)	90
β (°)	90
γ (°)	90
V (Å^3)	26,374
D_{calcd} (g/cm^3)	1.29
Wavelength (Å)	0.97
Resolution range (Å)	23–0.85
Measured reflections	151,663
Unique reflections:	
All data	23,556
$\geq 4\sigma(F_{obs})$	19,905
Completeness (%)	97.2
R_{sym} (%)	5.1
Reflections used for R_{free}	1186
$\geq 4\sigma(F_{obs})$	1040
Number of variables	3279
Final R-factor:	
All data	0.089
With $\geq 4\sigma(F_{obs})$	0.084
R_{free}:	
All data	0.103
With $\geq 4\sigma(F_{obs})$	0.095
Goodness-of-fit	1.202
$\Delta(\rho)$max, $\Delta(\rho)$min (e/Å^3)	0.55, −0.33
r.m.s. deviation from ideality:	
Bond lengths (Å)	0.010
Bond angles (°)	0.58
Torsion angles (°)	0.70

Some of the glucose O(6) atoms are twofold disordered; during refinement, the sums of their occupation factors were fixed at 1.0. About 20% of the O–H hydrogen atoms could be identified from difference-electron density maps. They were included in the refinement and their parameters were varied according to the 'riding' model. C–H hydrogen atoms were placed into calculated positions at a C–H bond length of 0.97 Å.[22] In the last refinement cycles, no restraints were applied for the glucose residues and 3279 parameters were allowed to vary. The crystallographic R-factor converged at 0.084 for 19,905 data with $F_{obs} > 4\sigma(F_{obs})$ and for all 23,556 data at $R = 0.089$, with a goodness-of-fit of 1.202 Before initiating refinement, 5% of the reflections were flagged for calculation of R_{free}[24] to avoid overfitting of the model to the data. R_{free} converged at 0.095 for 1040 data with $F_{obs} > 4\sigma(F_{obs})$ and at 0.103 for all 1186 flagged data (see Table 4).

5. Supplementary material

Tables of atomic coordinates, bond length and bond angles have been deposited with the Cambridge Crystallographic Data Center under CCDC 167755. They may be obtained on request from CCDC, 12 Union Road, Cambridge CB2 1EZ, UK (fax: +44-1223-336033; e-mail: deposit@ccdc.cam.ac.uk or www: http://www.ccdc.cam.ac.uk).

Acknowledgements

We thank Dr Takeshi Takaha for providing purified samples of CA26, Stefan Laettig, Professor H. Welfle and Dr W.-D. Hunnius for Raman measurements and discussions. Dr Alfred D. French is acknowledged for indicating to us recent publications on oligosaccharide conformation and on C–H···O contacts (Refs. 11 and 12). We are grateful to the Deutsche Forschungsgemeinschaft and Fonds der Chemischen Industrie for financial support of this work and we thank EMBL/DESY Hamburg for synchrotron beamtime.

References

1. Imberty, A.; Chanzy, H.; Perez, S.; Buleon, A.; Tran, V. *J. Mol. Biol.* **1988**, *201*, 365–378.
2. Sarko, A.; Zugenmaier, P. In *Fiber Diffraction Methods*; French, A. D.; Gardner, K. C. H., Eds.; ACS Symposium Series No. 141; American Chemical Society: Washington, DC, 1980; pp 459–482.
3. Rappenecker, G.; Zugenmaier, R. P. *Carbohydr. Res.* **1981**, *89*, 11–19.
4. Tomasik, P.; Schilling, C. H. *Advanced Carbohydrate Chemistry and Biochemistry*; Academic Press: New York, 1998; Vol. 53, pp. 236–426.
5. Colin, J. J.; de Claubry, H. G. *Ann. Chim.* **1814**, *90*, 87–92.
6. Takaha, T.; Yanase, M.; Takata, M.; Okada, S.; Smith, S. M. *J. Biol. Chem.* **1996**, *271*, 2902–2908.
7. Geßler, K.; Usón, I.; Takaha, T.; Krauss, N.; Smith, S. M.; Okada, S.; Sheldrick, G. M.; Saenger, W. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4246–4251.
8. Jacob, J.; Gessler, K.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T.; Saenger, W. *Angew. Chem. Int. Ed.* **1998**, *37*, 606–609.
9. Jacob, J.; Geßler, K.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T.; Saenger, W. *Carbohydr. Res.* **1999**, *322*, 228–246.
10. Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
11. French, A. D.; Kelterer, A.-M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J. Comput. Chem.* **2000**, *22*, 65–78.
12. Allinger, N. L.; Durkin, K. A. *J. Comput. Chem.* **2000**, *21*, 1229–1242.
13. Saenger, W.; Jacob, J.; Geßler, K.; Steiner, Th.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T. *Chem. Rev.* **1998**, *98*, 1787–1802.
14. (a) Harada, A.; Kamachi, M. *Macromolecules* **1990**, *23*, 2823–2824;
(b) Harada, A.; Li, J.; Kamachi, M. *Nature* **1993**, *364*, 516–518.
15. (a) Hinrichs, W.; Büttner, G.; Steifa, M.; Betzel, C.; Zabel, V.; Pfannemüller, B.; Saenger, W. *Science* **1987**, *238*, 205–208;
(b) Hinrichs, W.; Saenger, W. *J. Am. Chem. Soc.* **1990**, *112*, 2789–2796.
16. Saenger, W. In *Inclusion Compounds*; Atwood, J. L.; Davies, J. E. D.; MacNicol, D. D., Eds.; Academic: London, 1984; Vol. 2, pp. 231–259.
17. Shimada, J.; Handa, S.; Kaneko, H.; Terada, T. *Macromolecules* **1996**, *29*, 6408–6422.
18. Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1996**, *276*, 307–326.
19. (a) Sheldrick, G. M. In *Direct methods for solving macromolecular structures*; Fortier, S., Ed.; Kluwer Academic: Dordrecht, The Netherlands, 1998; pp. 401–441;
(b) Usón, I.; Sheldrick, G. M. *Curr. Opin. Struct. Biol.* **1999**, *9*, 643–648.
20. Fujinaga, M.; Read, R. J. *J. Appl. Cryst.* **1987**, *20*, 517–521.
21. Sheldrick, G. M.; Gould, R. O. *Acta Crystallogr., Sect. B* **1995**, *51*, 423–431.
22. G.M. Sheldrick, CCP4 Meeting York, January 1997, 147–157.
23. McRee, D. E. *Practical Protein Crystallography*; Academic: San Diego, 1993.
24. Brünger, A. T. *Nature (London)* **1992**, *355*, 472–475.