SUPPORTING INFORMATION

Insights into the Structure of Large-Ring Cyclodextrins through Molecular Dynamics Simulations in Solution

Petko M. Ivanov and Carlos Jaime

Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Bellaterra

(Barcelona), Spain

## TABLE OF CONTENTS

Some additional computational notes

We feel necessary to make two notes about *(i)* the generation of starting coordinates with molecular graphics, and *(ii)* the preparation of the system for the simulation. If not properly manipulated until positioning very close the two ends of the long chain before executing the 'ring-closure' of the macro-ring with the graphical utility, then such starting geometry can produce after the first minimization an optimized structure that contains artificial local deformations in the vicinity of the closure, e.g. some of the glucoses can transform from *chair* conformation into deformed *boat* conformation. It is not easy to detect such an error even for a structure of the size of CD30, and this could initiate long time MD simulations with final meaningless results. We detected such cases. Other authors also mentioned this problem.[1] Besides, simpler equilibration protocols that could work for gas phase or continuum solvent model simulations are of doubtful applicability for the simulation of the dynamics of a large molecule embedded in the bulk of explicit solvent molecules. The multi-step equilibration suggested by the developers of the AMBER program and tested here is recommended.

## Geometrical parameters used for the structural analysis

See Figure 1 in the main text.
O4(n)···O4(n-1): the distance between the glycosidic oxygen atoms;
O2(n)···O3(n-1): the distance between secondary hydroxyl groups of adjacent glucoses;
C1(n)-O4(n-1)-C4(n-1): the angle at the glycosidic oxygen connecting two glucoses;
O4(n)···O4(n-1)···O4(n-2): the angle formed by three neighbor glycosidic oxygen atoms;
O5(n)-C1(n)-O4(n-1)-C4(n-1): the dihedral angle ϕ;
C1(n)-O4(n-1)-C4(n-1)-C3(n-1): the dihedral angle ψ;
O4(n)···O4(n-1)···O4(n-2)···O4(n-3): the dihedral angle of four consecutive glycosidic oxygen atoms;
O3(n)···C4(n)···C1(n+1)···O2(n+1): the dihedral *flip* between secondary hydroxyls of adjacent glucoses.
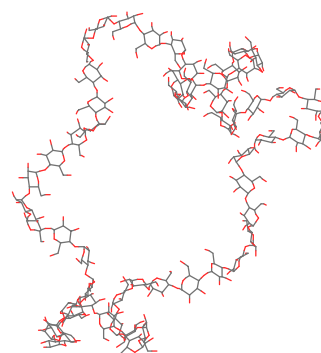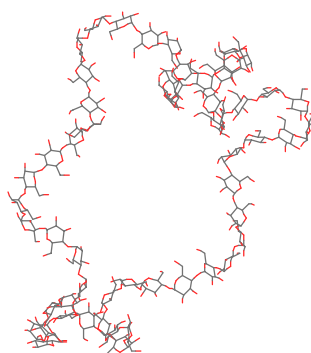
CD26(X-ray)
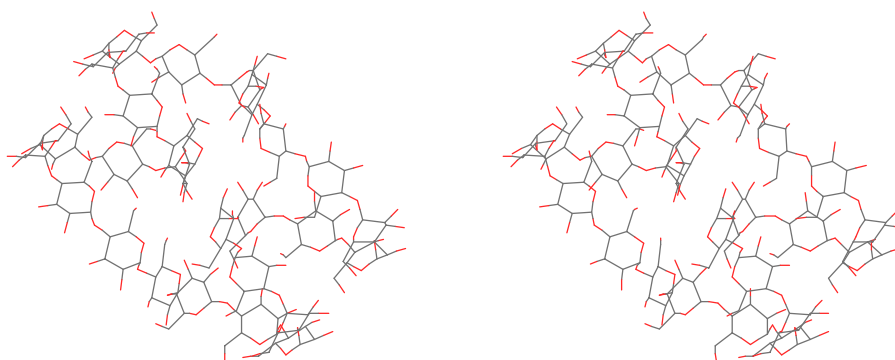
CD30
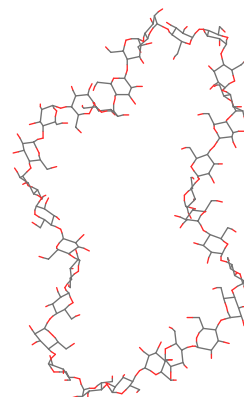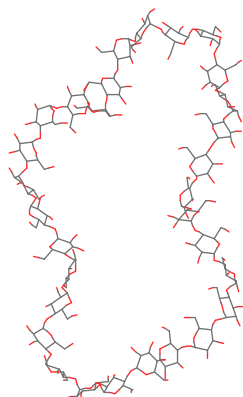
CD55

CD70

CD85
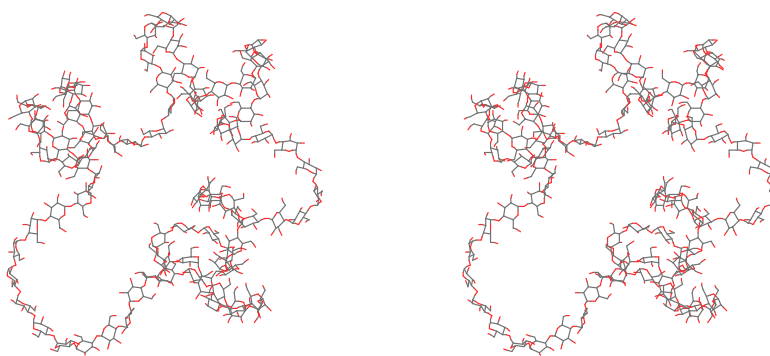
CD100
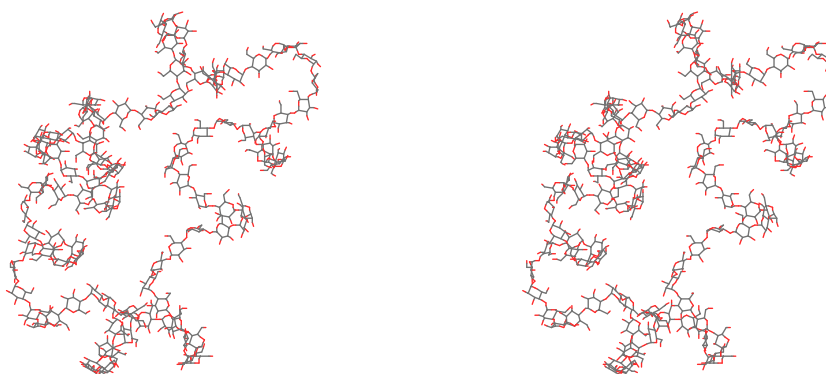
Figure 1S. Stereo displays of the initial geometries used in the simulations (geometries after the 20.0 ps MD simulations of the equilibration step).

CD26

CD30
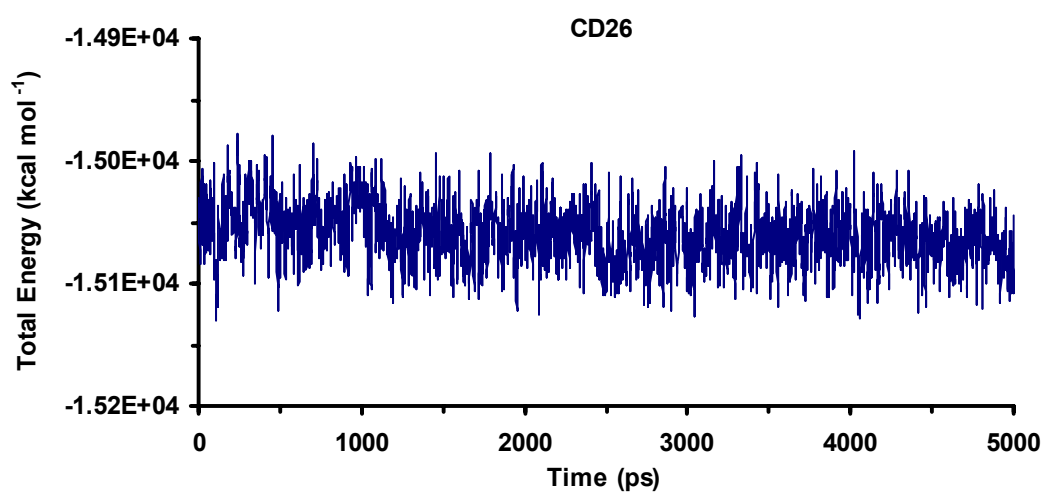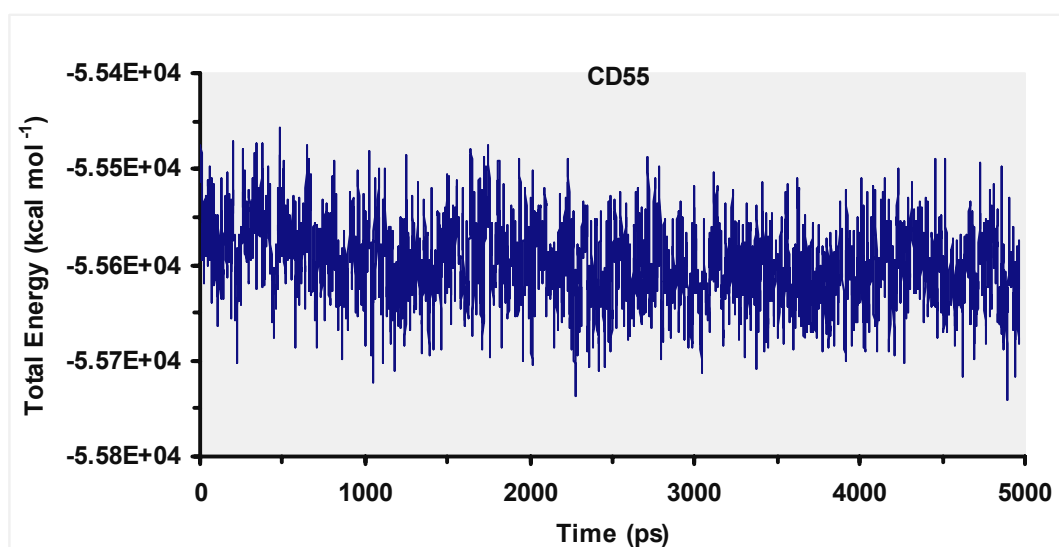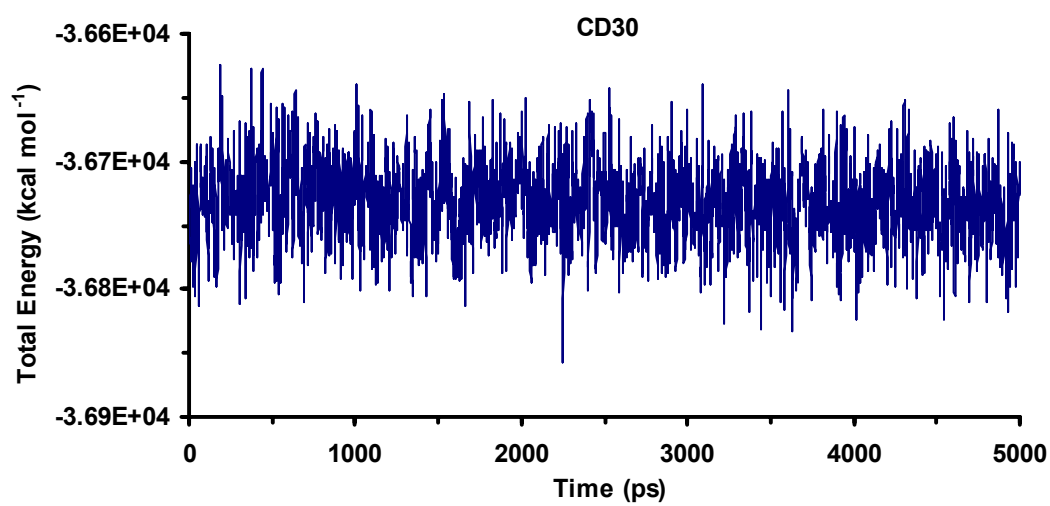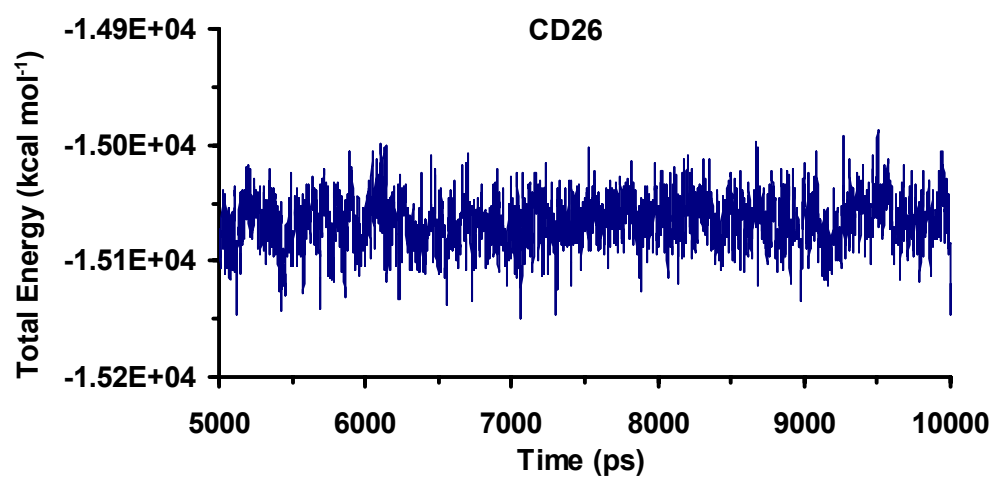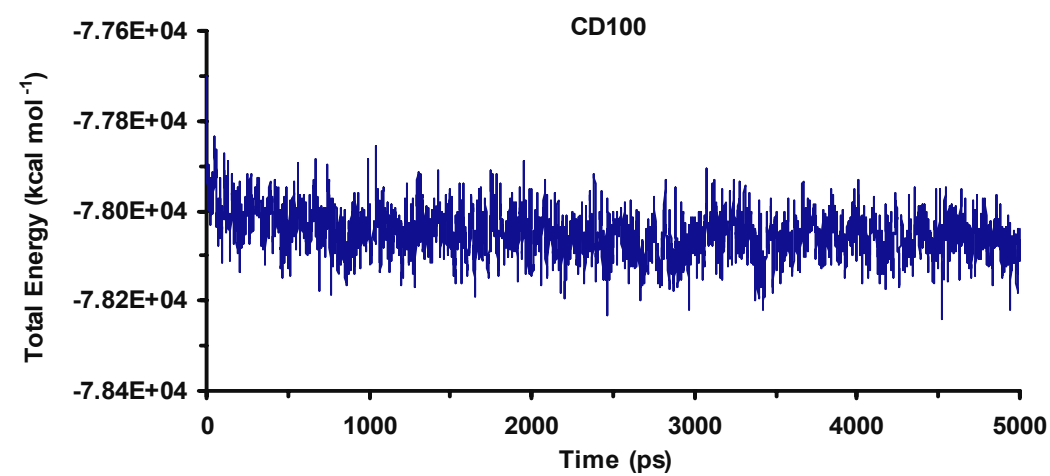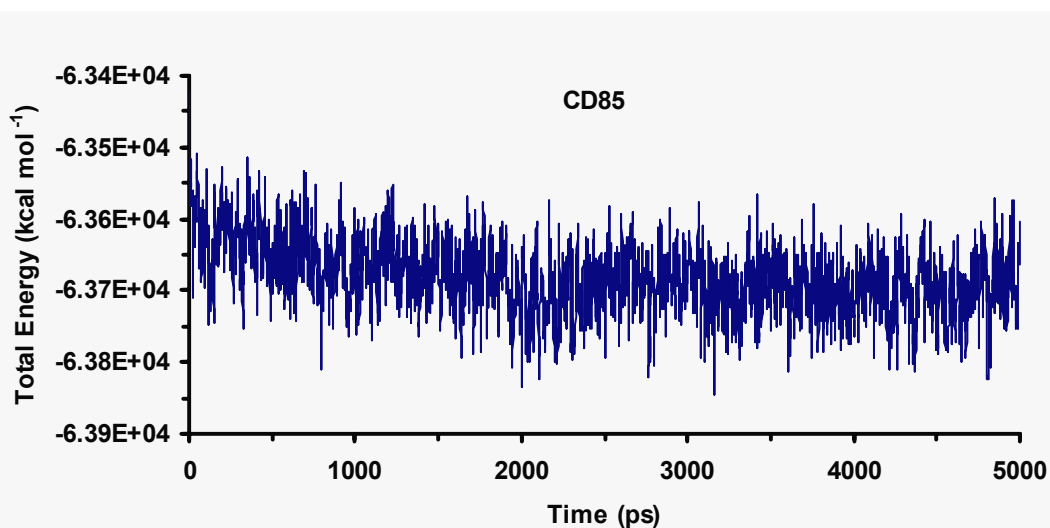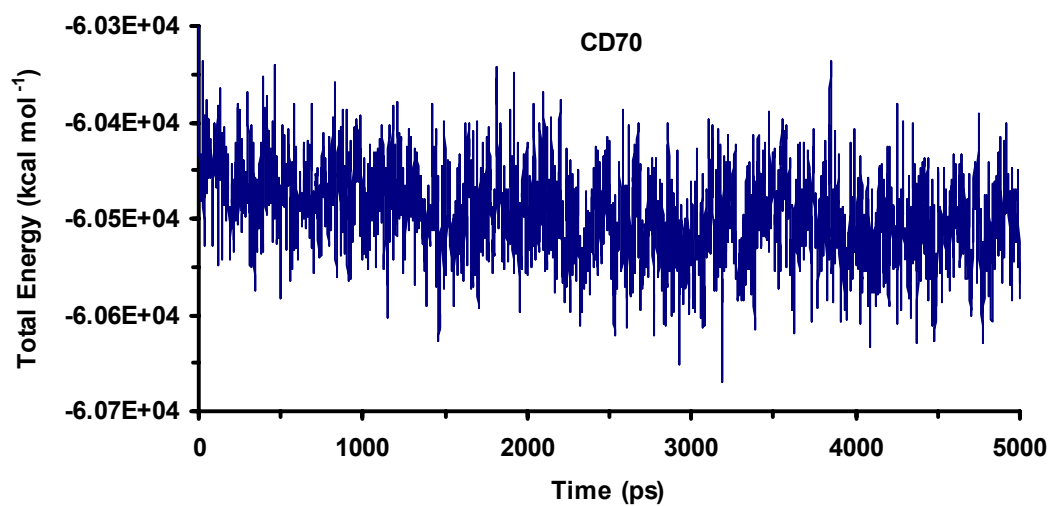
CD55

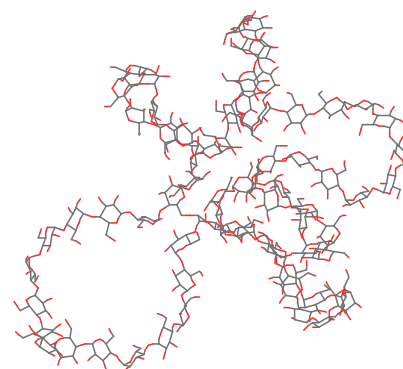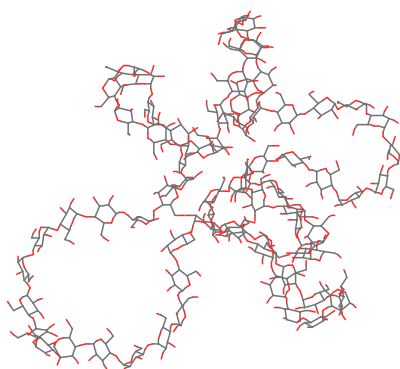**Figure 2S.** Variations of the total energies with simulation time.

CD26

CD26

CD30

CD55

CD70

CD85

CD100
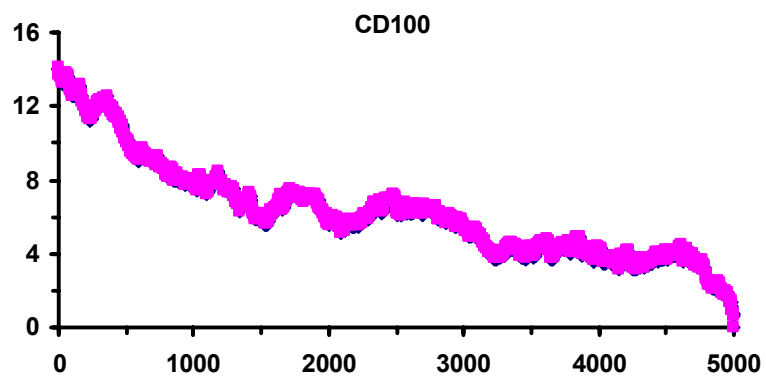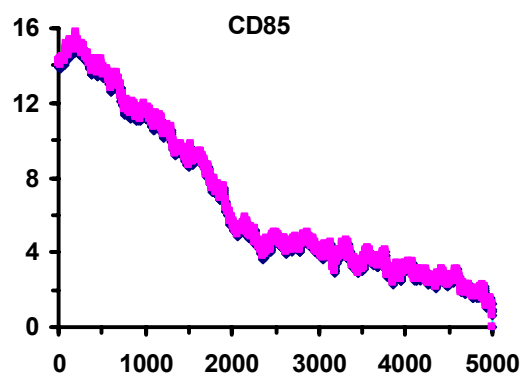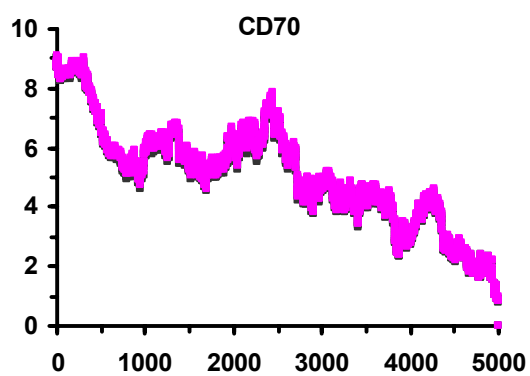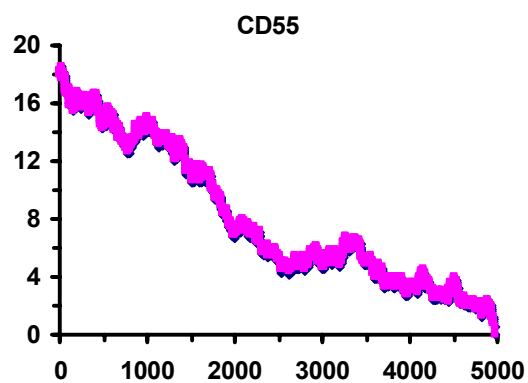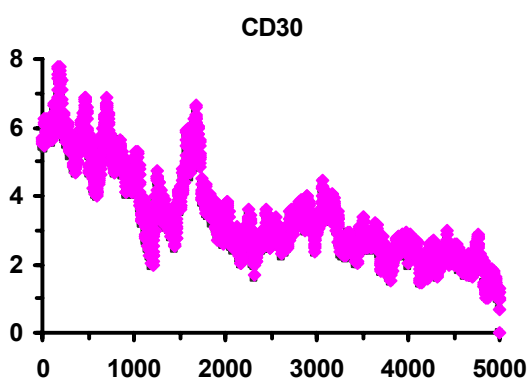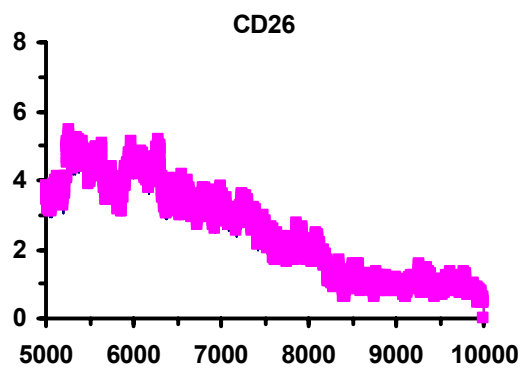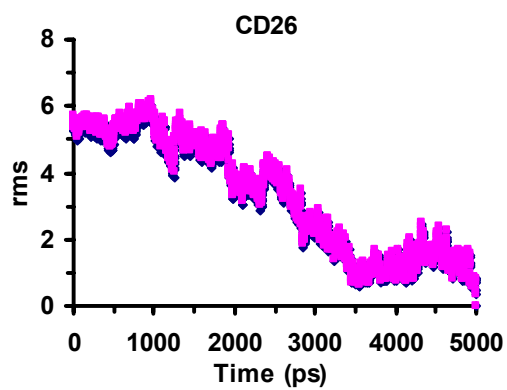
Figure 3S. Rms deviations of atomic coordinates (in Å) relative to the final set of coordinates saved from the simulation. The C1, O4 and C4 atoms were used in the fit.

CD26(X-ray)   **CD26-0.75**   **CD26-2.0**

CD26-3.0   **CD26-4.0**   **CD26-5.0**

CD30-0.2   **CD30-3.0**   **CD30-5.0**
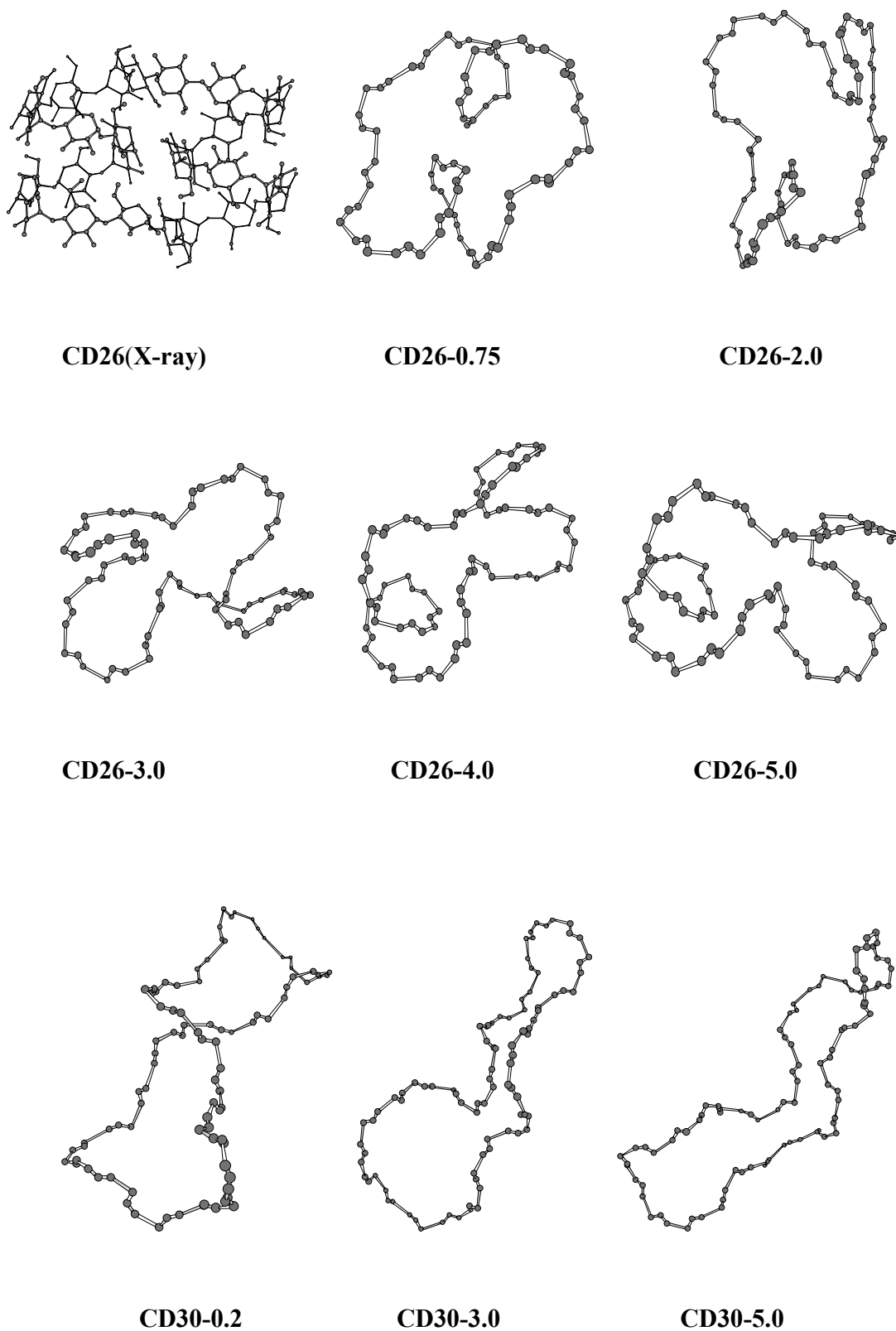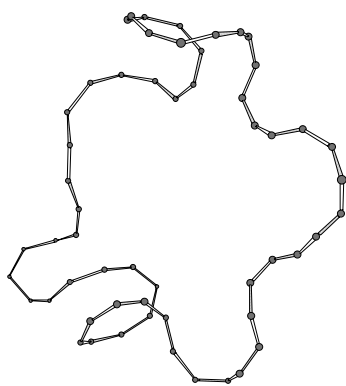
CD55-0.25
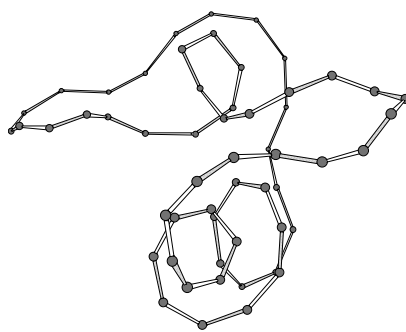
CD55-3.0
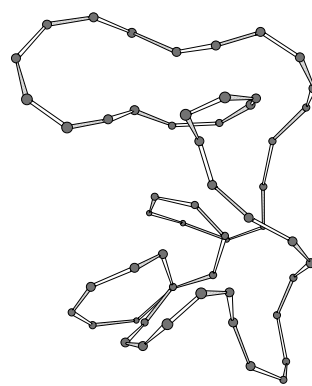
CD55-4.7
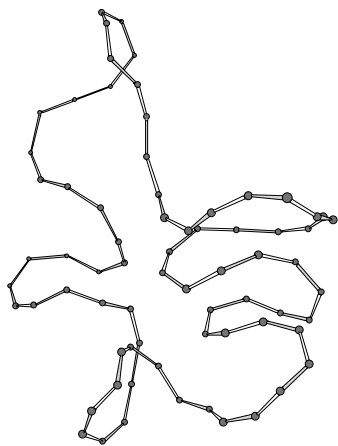
CD70-0.25
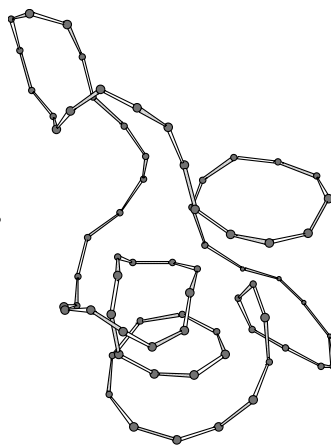
CD70-3.0
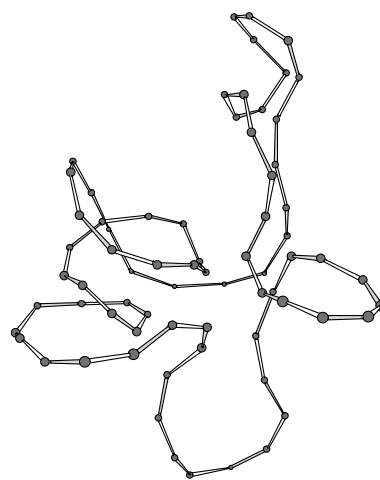
CD70-5.0

CD85-0.2

CD85-3.0

CD85-4.5

CD100-0.25   **CD100-3.0**   **CD100-5.0**
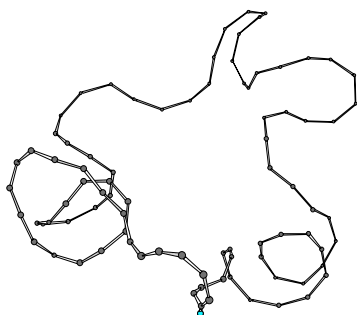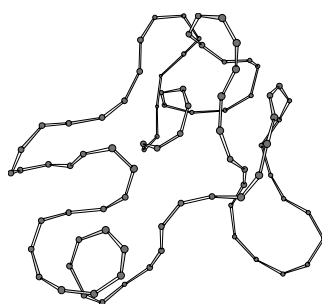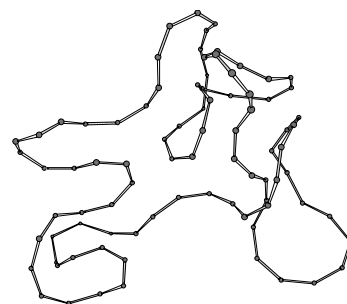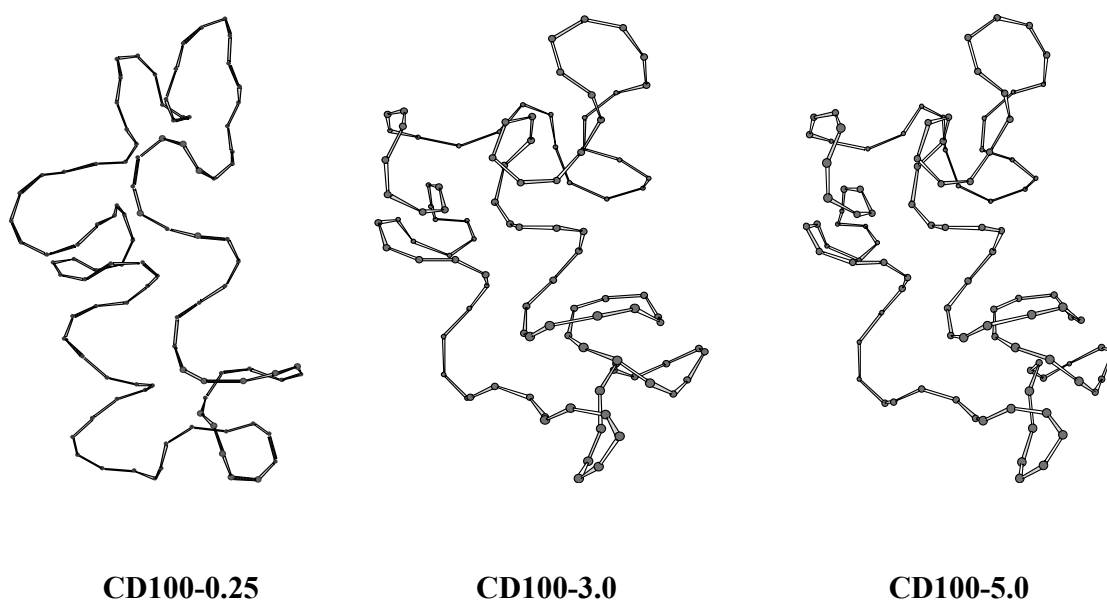
Figure 4S. Simplified views of the geometries at an earlier, intermediate, and the final stages of the simulations of all CDs. The (C1(n)—O4(n-1)—C4(n-1)) portions were presented for the smaller CDs, while only the O4 atoms were utilized for CDn (n = 55, 70, 85, 100).
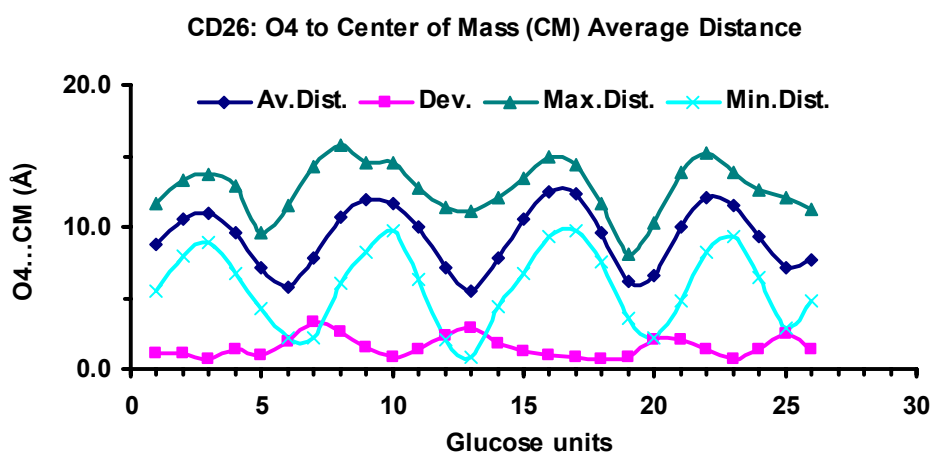
Figure 5S. Variation of the time averaged values of the O4 to the center of mass distance evaluated for each residue of CD26 (5.0 ns simulation). Rms deviations, maximum and minimum distances are also given.

CD30: O4 to Center of Mass (CM) Average Distance
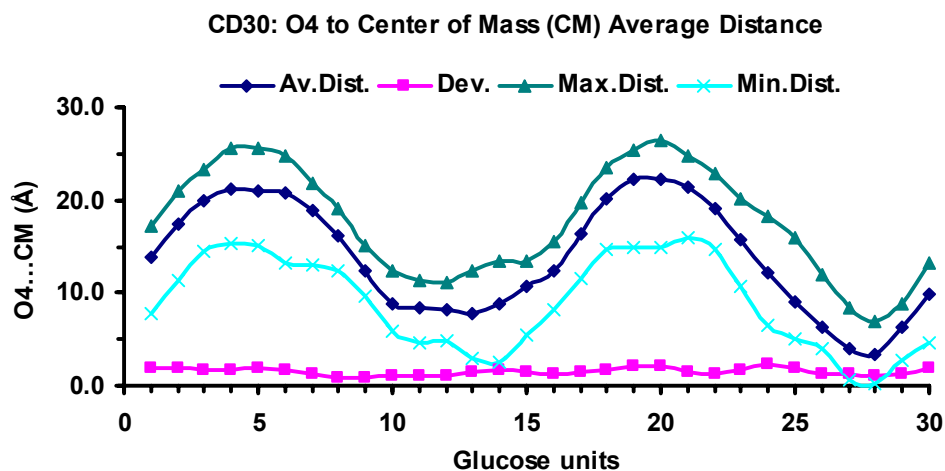
Figure 6S. Variation of the time averaged values of the O4 to the center of mass distance evaluated for each residue of CD30 (5.0 ns simulation). Rms deviations, maximum and minimum distances are also given.

CD55: O4 to Center of Mass (CM) Average Distance
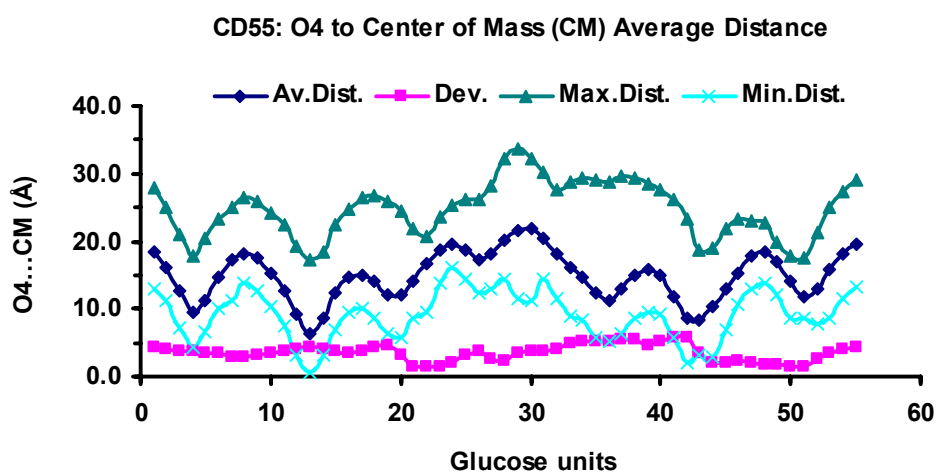
Figure 7S. Variation of the time averaged values of the O4 to the center of mass distance evaluated for each residue of CD55 (5.0 ns simulation). Rms deviations, maximum and minimum distances are also given.

CD70: O4 to Center of Mass (CM) Average Distance
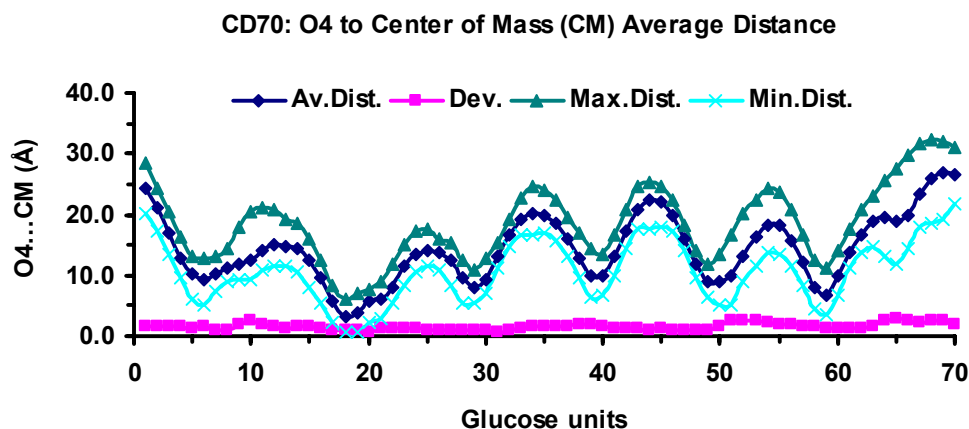
Figure 8S. Variation of the time averaged values of the O4 to the center of mass distance evaluated for each residue of CD70 (5.0 ns simulation). Rms deviations, maximum and minimum distances are also given.

CD85: O4 to Center of Mass Distance (CM) Average Distance
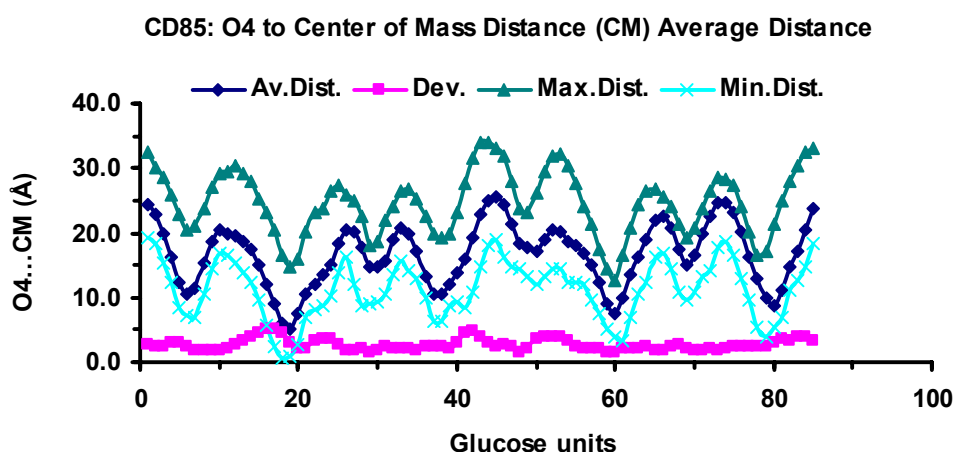
Figure 9S. Variation of the time averaged values of the O4 to the center of mass distance evaluated for each residue of CD85 (5.0 ns simulation). Rms deviations, maximum and minimum distances are also given.

CD100: O4 to Center of Mass (CM) Average Distance

Av.Dist ◆ — Dev. ■ — Max.Dist. ▲ — Min.Dist. ✕

O4...CM (Å)

50.0
40.0
30.0
20.0
10.0
0.0

0   20   40   60   80   100

Glucose units
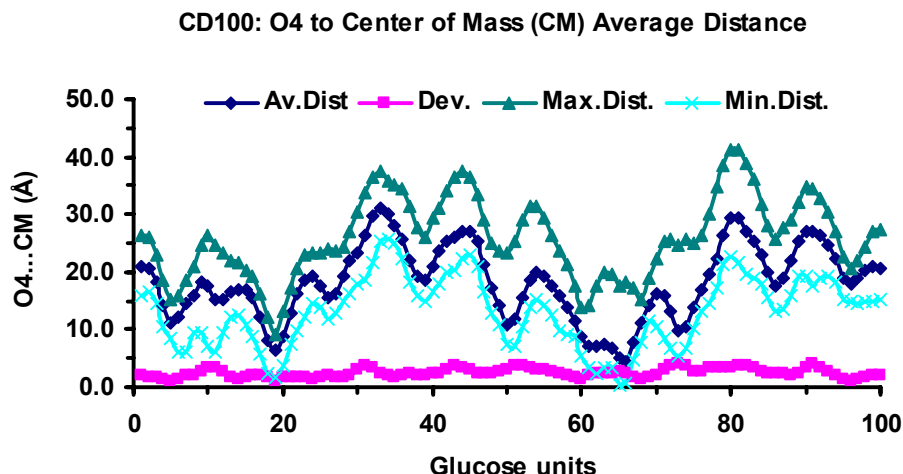
Figure 10S. Variation of the time averaged values of the O4 to the center of mass distance evaluated for each residue of CD100 (5.0 ns simulation). Rms deviations, maximum and minimum distances are also given.

Experimental Structural Characteristics of Cyclodextrins

In order to comment the structural characteristics of the large-ring cycodextrins, we first have to recall also the peculiarities of the native cyclodextrins and look for the differences we may find when comparing the two groups of macrocycles (Table 1S).

The most interesting structural features could be summarized in the following way:[2]

1. The O4 atoms defining the macrorings of CD6-CD8 are nearly coplanar, while larger deviations are observed for the nanogon in CD9.[3]

2. The O4(n)...O4(n-1) distances shaping the edges of the macrorings do not vary noticeably along the ring of a particular member of the CD family. They increase, however, from CD6 to CD8 because the glucose unit has to adjust to the respective radii of the CD, and are roughly constant for CD8 and the larger CDs, ~4.5 Å.

3. The O4(n)···O4(n-1)···O4(n-2) angles have values around 120º in CD6, 128º in CD7, 135º in CD8, and close to 138º in the larger CDs.

4. The average O2(n)···O3(n-1) distances are not the same for all CDs. They decrease from CD6 (2.98 Å) through CD7 (2.88 Å) to CD8 (2.82 Å), implying, in accord with experimental data,[4,5] that the hydrogen bonds are stronger in CD8 than in CD6 and CD7.

5. Experimental evidence exists[6] for interchange of hydrogen bonds in CD7, i.e. the O3-H···O2 and O3···H-O2 hydrogen bonds are in a dynamical equilibrium.[2]

6. The O6-H hydroxyl groups are frequently engaged in intra-glucose hydrogen bonding of the type O6(n)-H···O5(n), and usually represent the minor component of three-center hydrogen-bonding interactions.

7. The torsion angles ϕ (O5(n)—C1(n)—O4(n-1)—C4(n-1)) and ψ (C1(n)—O4(n-1)—C4(n-1)—C3(n-1)) of CD6 to CD8 are very similar.

8. Cooperative networks are formed by the O-H···O hydrogen bonds with the participation of the CDs hydroxyl groups and the water molecules.[2] The cooperative effect,[7] as shown by quantum chemical calculations, contributes about 25% of additional energy to the individual hydrogen bond.[8,9]

9. There is a large number of C-H···O hydrogen bonds that stabilize the host-guest interaction and the crystal lattice.[10,11] In carbohydrate crystal structures, 25% of the hydrogen bonds are of the C-H···O type.[12]

The CD9 macromolecule is distorted such that the O4 atoms describe an ellipse shaped like a boat. The cavity is considerably collapsed compared with CD6 to CD8. CD9 has the maximum number of glucose residues that can form an annular structure.[2] The molecular shapes of CD10 and CD14 are very different compared to the shapes of the smaller CD6 to CD9. This is due to the ~180º flipping of two diametrically opposed glucoses so that the ring of intramolecular O2(n)···O3(n-1) hydrogen bonds, which is still present in CD9, is disrupted.[13] At the flip site, two adjacent glucoses are oriented *trans*, the other glucoses still remaining *cis* (Figure 1 in main text). The two symmetrically arranged 'band flips' divide the molecules into two halves connected at the flip sites. The central cavities are no longer open and round and the elliptically distorted molecules adopt shapes resembling butterflies with the band flips located at the 'body' and the wings formed by CD-like segments.[2] The *trans* orientations of adjacent glucoses are stabilized by O6(n)···O3(n-1) and O5(n)···O3(n-1) hydrogen bonds.

In large CDs, two strings in opposite orientation may approach each other and associate across the center of the molecule through hydrogen bonds between O2 and O3 hydroxyl groups.[2] CD10 crystallized from 1:1 mixture of water and acetonitrile, but a complex with the organic solvent is not formed.[14] Thus the inclusion properties of CD10 are different to the properties of the native cyclodextrins, which may accommodate many kinds of guest molecules, among them acetonitrile.[15] The cavities in CD10 and larger CDs are distorted into narrow grooves that only guest molecules with geometries complementary to these grooves might be able to form inclusion complexes.[2]

Table 1S. Some Structural Parameters of the CD6, CD7, CD8, CD9, CD10, CD14 and CD26

Hydrates (data from Table 1 of Ref. [3] and Ref. [15a] in main text)[a]

		CD6	CD7	CD8	CD9	CD10	CD14	CD26[b]
Number of glucose units		6	7	8	9	10	14	26
cavity diameter		4.7-5.3	6.0-6.5	7.5-8.3				
height of the cone		7.9±1	7.9±1	7.9±1				
volume of the cavity		174	262	472				
	av.	109.2	109.8	108.9	112.1	99.4	103.4	103.6
Φ	min.	102.0	102.3	103.6	88.4	94.1	96.6	91.1
	max.	114.9	118.6	123.2	141.2	102.1	110.2	115.3
	av.	128.8	127.6	127.1	124.7	106.1	112.6	115.3
Ψ	min.	115.1	114.2	111.9	97.6	96.3	103.6	97.4
	max.	148.7	140.0	138.5	144.5	122.0	135.2	131.4

	av.	119.9	128.3	134.9	136.6	138.2	138.2	126.4
O4(n)..O4(n-1)..O4(n-2)	min.	116.9	125.2	133.5	125.7	126.7	131.6	117.9
	max.	122.3	132.5	136.9	149.9	145.9	142.5	136.7
	av.	4.24	4.38	4.50	4.49	4.49	4.54	4.39
O4(n)…O4(n-1)	min.	4.16	4.27	4.43	4.26	4.36	4.45	4.11
	max.	4.30	4.50	4.59	4.73	4.63	4.61	4.56
	av.	2.98	2.88	2.82	2.91	2.93	2.83	2.86
O2(n)…O3(n-1)	min.	2.90	2.80	2.76	2.74	2.85	2.76	2.65
	max.	3.15	2.98	2.91	3.23	3.01	2.90	3.10

[a] Units are: distances (Å), angles (°), volumes (Å^3). [b] The data refer to the helical part.

The crystal structure of CD26 hydrate[16] shows that the CD chain is not folded as proposed on the basis of the structure of CD14 with central antiparallel, left-handed double helix and two band flips in the loops.[2] CD26 adopts the shape of a figure eight in which each half consists of two left-handed, single helical turns with six glucoses per repeat (Figure 1S, CD26(X-ray)). At the 'upper' and at the 'lower' sides, the short helices are connected by two stretches of three glucoses containing one band flip each. Thus, the structure of CD26 hydrate is modular and contains elements taken from CD6 and CD10 and from V-amylose.[2] The two short single helices in CD26 present channel-like cavities with a width similar to the one found in CD6. They accommodate disordered water molecules but could also enclose other molecules of suitable size.

Table 2S. Computed Distances (Å), Bond Angles and Dihedral Angles (in degrees) of Neighboring Glucose Units Obtained from the MD Simulations in Water Solution

	O4(n) …O4(n-1)				O2(n)…O3(n-1)			
	Av.	rms	max.	min.	av.	rms	max.	min.
CD26	4.3	0.2	5.1	3.0	3.7	0.6	6.1	2.4
CD30	4.4	0.2	5.1	3.2	4.1	0.5	6.3	2.4
CD55	4.4	0.2	5.1	2.9	3.6	0.6	5.6	2.4
CD70	4.4	0.2	5.4	3.0	3.6	0.6	5.8	2.4
CD85	4.4	0.2	5.7	3.1	3.6	0.6	5.8	2.4
CD100	4.4	0.2	5.7	2.7	3.6	0.6	6.3	2.4

	C1(n)—O4(n-1)—C4(n-1)				O4(n)…O4(n-1)…O4(n-2)			
	av.	rms	max.	min.	av.	rms	max.	min.
CD26	116.2	3.5	130.6	101.1	129.4	9.2	179.6	83.3
CD30	116.7	3.5	133.9	100.6	139.8	8.5	178.6	83.0
CD55	116.1	3.5	133.0	100.9	134.5	8.6	175.1	75.7
CD70	116.0	3.5	131.6	101.1	136.3	8.0	176.6	79.8
CD85	116.1	3.5	133.0	101.6	136.6	8.1	179.2	83.1
CD100	116.1	3.5	133.0	101.8	134.9	8.7	179.9	64.3

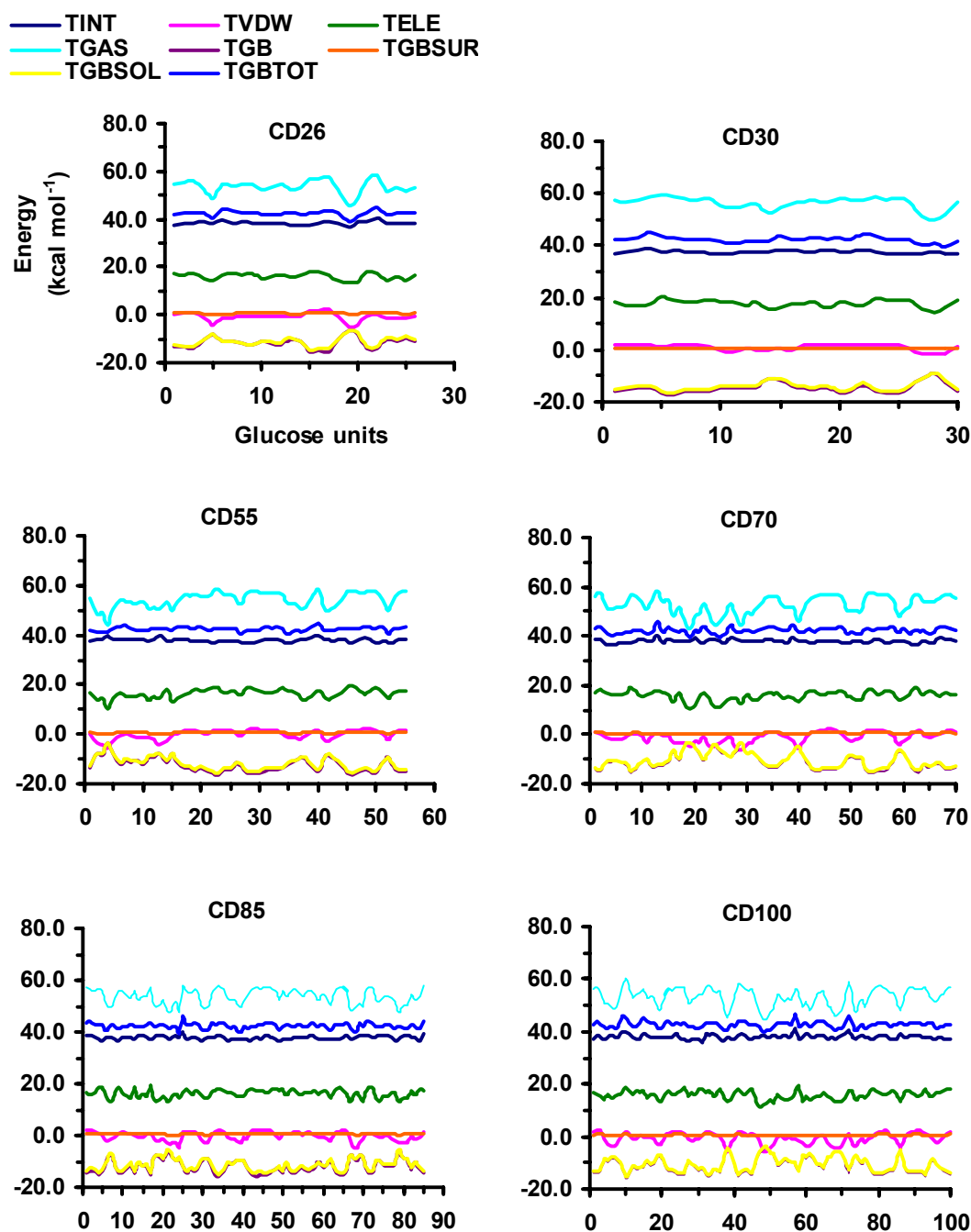|	O4(n)···O4(n-1)···O4(n-2)···O4(n-3)		O3(n)···C4(n)···C1(n+1)···O2(n+1)	
	av.	rms	av.	rms
CD26	-19.0	31.9	-42.0	32.4
CD30	-39.7	30.9	-42.3	28.2
CD55	-25.7	33.7	-37.1	32.3
CD70	-18.8	34.7	-29.5	33.9
CD85	-26.5	31.7	-38.2	29.3
CD100	-21.8	32.6	-33.9	31.4
	Φ O5(n)-C1(n)-O4(n-1)-C4(n-1)		Ψ C1(n)-O4(n-1)-C4(n-1)-C3(n-1)	
CD26	98.6	20.3	98.0	23.4
CD30	87.3	18.2	75.0	16.5
CD55	103.2	22.7	99.7	17.8
CD70	107.6	22.8	103.2	18.5
CD85	104.1	20.2	99.0	16.1
CD100	104.0	21.6	99.2	17.7

Figure 11S. Variation of time averaged values of individual energy terms determined for each residue of the CDs in the 5.0 ps simulations (TINT – total internal energy, TVDW – total van der Waals energy, TELE – total electrostatic energy, TGAS – total gas phase energy, TGB – total GB polarization energy, TGBSUR – total surface area energy, TGBSOL – total GB solvation energy, TGBTOT – total energy (TGAS + TGBSOL)).
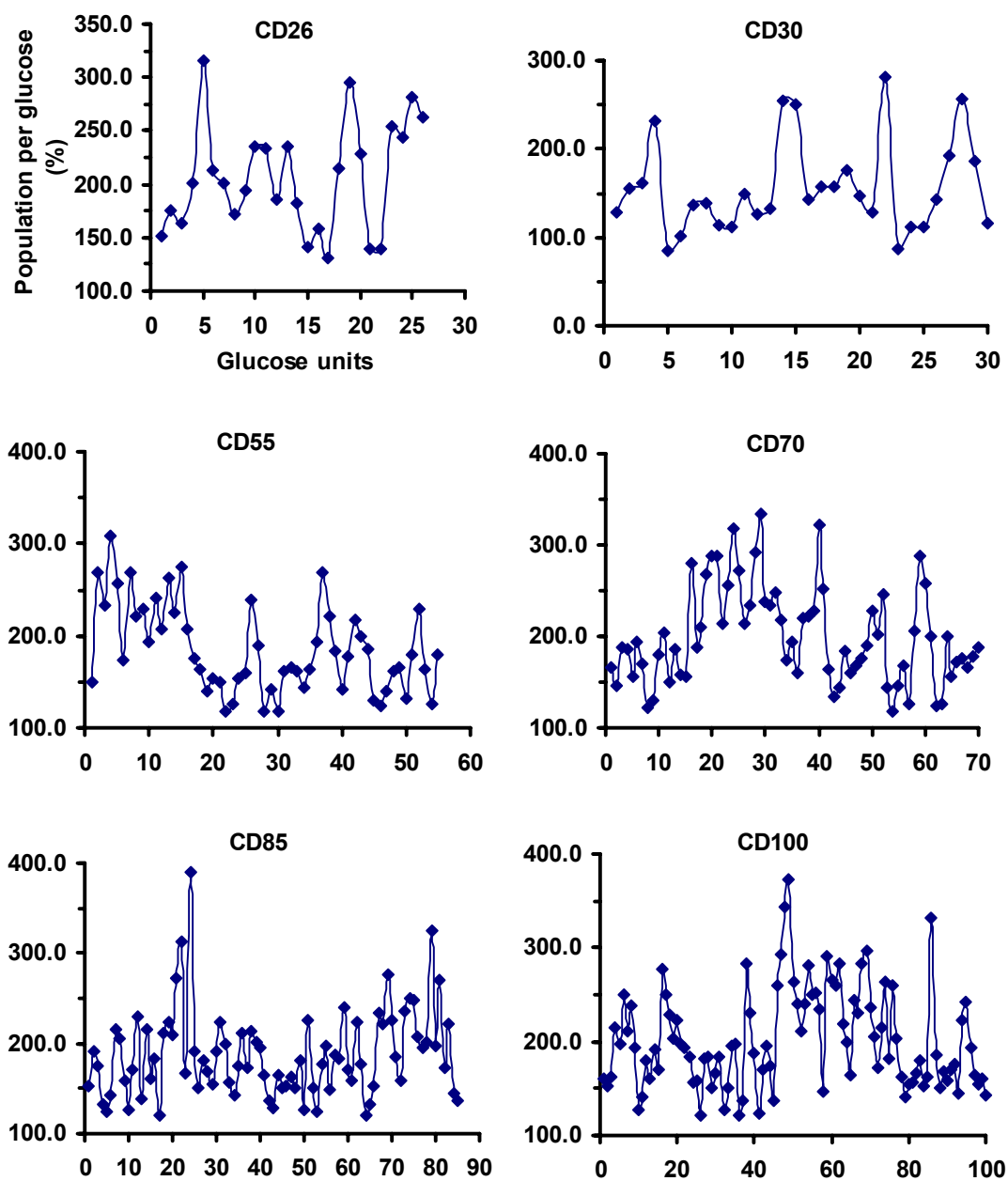
Figure 12S. Time averaged populations of intramolecular hydrogen bonds estimated for each glucose residue.
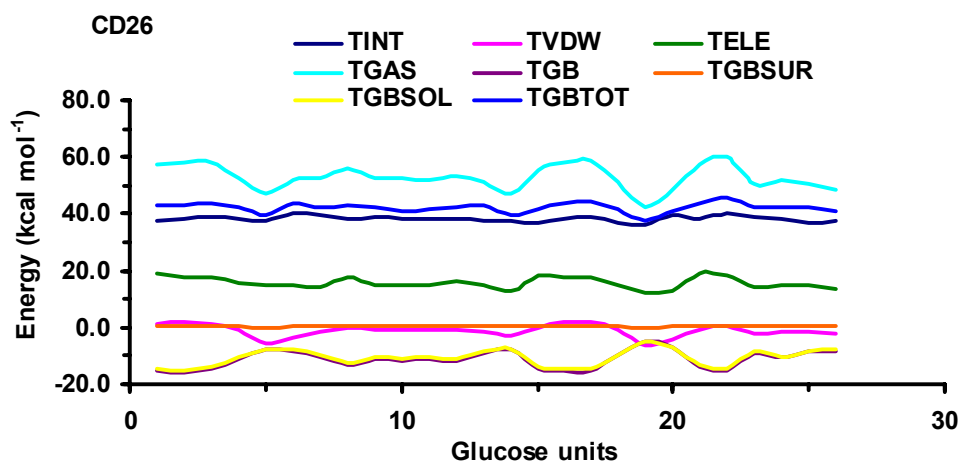
Figure 13S. Variations of time averaged values of individual energy terms evaluated for each residue of CD26 for the simulation interval from the 5th to the 10th ns.
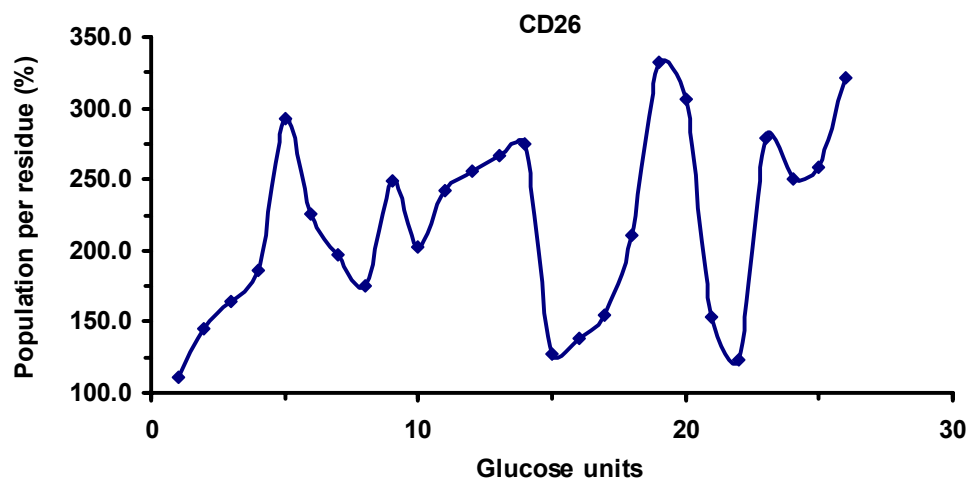

Figure 14S. Time averaged populations of intramolecular hydrogen bonds estimated for each glucose residue of CD26 for the simulation interval from the 5th to the 10th ns.

References

[1] Momany, F.A.; Willett, J.L. *Carbohydrate Research* **2000**, *326*, 210-226.

[2] Saenger, W.; Jacob, J.; Gessler, K.; Steiner, T.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T. *Chem. Rev*. **1998**, *98*, 1787-1802.

[3] Fujiwara, T.; Tanaka, N.; Kobayashi, S. *Chem. Lett*. **1990**, 739-742.

[4] Casu, B.; Reggiani, M.; Gallo, G.G.; Vigevani, A. *Chem. Soc. Spec. Publ.* **1968**, *23*, 217-226.

[5] Bergeron, R.; Channing, M.A. *Bioorg. Chem.* **1976**, *5*, 437-449.

[6] Hanabata, K.; Matsuo, T.; Suga, H. *J. Inclusion Phenom*. **1987**, *5*, 325-333.

[7] Saenger, W. *Nature* **1979**, *279*, 343-344.

[8] Lesyng, B.; Saenger, W. *Biochim. Biophys. Acta* **1981**, *678*, 408-413.

[9] Koehler, J. E. H.; Saenger, W.; Lesyng, G. *J. Comput. Chem.* **1987**, *8*, 1090-1098.

[10] a) Steiner, T. *Crystallogr. Rev.* **1996**, *6*, 1-51; b) Steiner, T. *Chem. Commun.***1997**, 727-734.

[11] Takahashi, H.; Tsuboyama, S.; Umezawa, Y.; Honda, K.; Nishio, M. *Tetrahedron,* **2000**, *56*, 6185-6191.

[12] Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*, Springer-Verlag: Heidelberg, 1991.

[13] Jacob, J.; Gessler, K.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T.; Saenger, W. *Angew. Chem. Int. Ed. Eng.,* **1998**, *37*, 606-609.

[14] Ueda, H.; Endo, T.; Nagase, H.; Kobayashi, S.; Nagai, T. *J. Inclusión Phenom. Mol. Recognit. Chem.* **1996**, *25*, 17-20.

[15] Aree, T.; Jacob, J.; Saenger, W.; Hoier, H. *Carbohydr. Res.* **1998**, *307*, 191-197.

[16] 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.