Theor Chem Acc (2007) 117: 85–97
DOI 10.1007/s00214-006-0137-x

REGULAR ARTICLE

Itziar Maestre · Ivan Beà · Petko M. Ivanov
Carlos Jaime

Structural dynamics of some large-ring cyclodextrins. A molecular dynamics study: an analysis of force field performance

Received: 1 February 2006 / Accepted: 11 April 2006 / Published online: 22 June 2006

Abstract Molecular dynamics simulations were performed on four large ring cyclodextrins (LR-CDs), CD14, CD21, CD26 and CD28, both in gas phase and in water solution. Four different force fields (parm94, parm99, glycam2000a and MM3*) were used to examine their dependence with the results. The differences were not significant when parm99 and glycam2000a were used. Parm94 and MM3* results differ considerably due to inadequate bending parameters (parm94) or due to a strong stabilization by the intramolecular hydrogen bonds between hydroxyl groups. Simulations in the gas phase show that the CDs are elongated and twisted, and that cavities typical for the native CDs (CDn, $n = 6, 7, 8$) are not present. Simulations in water solution produced average structures that do not correspond to the conformations in the crystalline state. The LR-CDs are highly flexible and this could favor the formation of inclusion complexes through the intermediation of some other molecules that could fit to the smaller and more specific cavities they have.

Keywords Cyclodextrins · Molecular dynamics · Conformational analysis · Force field dependence

1 Introduction

There has been an increasing interest during the last decade in the large-ring cyclodextrins (hereinafter named as LR-CDs) [1,2] and their physicochemical properties [3–5], in spite of existing difficulties in their synthesis [6–8], isolation and purification [9–11]. Although LR-CDs were firstly prepared by French et al. [12], the publication of the crystal structure for the CDs composed by 9 [13], 10 [14–16], 14 [14, 15], and 26 [17,18] glucose units has been clearly determinant for a better understanding of these molecules. The X-ray structures showed new structural motifs, and it became really evident that they are structurally very different from the native CDs (those composed of 6, 7 and 8 glucose units). More recently the thermal and structural characterization of δ-CD, ε-CD and ι-CD have been carried out [19], and references to CDs with more than 60 [20] and several hundred [21] glucose units have been made. Much larger CDs, even with more than 100 glucoses in the ring, have been prepared by the action of disproportionating enzyme on amylase [21] but crystallographic structures of such large CDs have not been reported so far. The crystal structure determinations describe the geometrical properties and characteristics of two large CDs. However, 'their solution and inclusion properties have yet to be elucidated' [1,2].

The complexation properties of LR-CDs have started to be studied. The inclusion complexes of CDs of up to 17 glucose units with several benzoates, salicylate, ibuprofen anion and 1-adamantane carboxylate have been studied and their formation constants have been determined [22]. Association constants for those CDs having more than ten glucoses are usually small ($<40\,M^{-1}$). However, the LR-CDs with 21–33 glucose units form stable inclusion complexes with iodine in aqueous solution as demonstrated by isothermal titration calorimetry (formation constants of about $1–7\times10^3\,M^{-1}$) [23], and δ-CD has demonstrated to form a stable 1:1 complex with C_{70} that allows its solubilization in water [24]. More recently, η-CD (12 glucoses) has been proved to be effective in the partial separation of carbon nanotubes [25].

The structural features of the native and the LR-CDs, mainly X-ray structural analyses, have been surveyed and computational studies have also been quoted [2,14,26–28]. A review [29] was especially devoted to the applications of molecular modeling techniques to the study of the static and the dynamical features of CDs, as well as their participation in host–guest complexation. However, it is mainly centered in native CDs. Macrocycles containing 14, 18, 24 and 48 glucose units were studied by molecular dynamics simulations either in vacuum or in water solution, but with rather

On sabbatical leave from the Institute of Organic Chemistry with Center of Phytochemistry, Bulgarian Academy of Science, ul. Acad. G. Bonchev, bloc 9, 1113 Sofia, Bulgaria

I. Maestre · I. Beà · P. M. Ivanov · C. Jaime (✉)
Departament de Química, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
E-mail: carlos.jaime@uab.es

short simulation times (<100 ps) [30,31]. The circularized three-turn single helical structure proposed for CD21 from small-angle X-ray scattering was shown, with rather short MD simulations (100 ps), to persist in water at 300 K [28]. Molecular dynamics simulations (with lengths of 100–400 ps) using a DFT/ab initio-derived empirical force field, AMB99C [32,33], were used in studies of the molecular properties of the CDs containing 10 and 21 glucoses. More recently, a study on CDs with degree of polymerization of 26, 30, 55, 70, 85 and 100, using much longer molecular dynamics (MD) simulations (5 ns for all except the one of 26 for which the MD was 10 ns), has been published [34].

Since a growing experimental activity on the LR-CDs can be deduced from the literature, it is worth devoting efforts to study these promising macrocycles using computational methodologies once these have proved its goodness for studying the native CDs. We present in this paper results from long MD simulations (5 ns) of the LR-CDs with 14, 21, 26 and 28 glucose units as isolated molecules (gas phase) and in water solution, using four different force fields. The degree of polymerization of most of the CDs studied is a multiple of 7 because the change in ^{13}C - NMR chemical shifts of the large CDs above CD10, especially for the ^{13}C1 and ^{13}C4 signals, presents an oscillation with a periodicity of approximately six or seven. (See Fig. 4 of Ref. [2]. A careful look at this figure suggests 7 as the most probable periodicity.) In addition, and as reference structures, the native CDs were also studied at the same conditions of the simulation experiments.

Circular dynamics are usually named using greek letters as prefixes. For a better understanding, hereinafter the nomenclature used will be based on that used for the cycloamyloses: a number added after CD designates the number of units in the macrocycle, e.g., CD6 designates α-CyD, while CD26 is the cyclodextrin with 26 glucose residues. Figure 1 displays a schematic representation of a cyclodextrin fragment with the numbering of the atoms, as well as examples for *cis* and *trans* orientations of neighboring glucose units.

2 Computational details

The computations with the MM3* force field [35][1] were performed within Macromodel [36], while those with the parm94 [37], parm99 [38] and glycam2000a [39] force fields were carried out with the AMBER program (version 5 with parm94 and version 7 with the rest; modules LEaP, SANDER and CARNAL were used, respectively, for the preparation of the input data, the minimization and the MD simulation steps, and for the analysis of the MD trajectories) [40]. Atomic charges used for the glucose unit in the AMBER computations were those previously published [41]. In the MD simulations with parm94 force field, one additional bending parameter was used, not present in the original force field parameter file: OS–CT–OS, $k_0 = 80.0$ kcal/mol, $\phi_0 = 126.0°$. The AMBER MD simulations were always run with a cut-off of 8.0 Å for the non-bonded interactions. For simulations in water, the "solvateBox" command of LEaP has been used to create a rectangular parallelepiped solvent box around the CD with buffer distances of 10.0 Å between the walls of the box and the closest atoms of the solute (TIP3P [42] water molecules and periodical boundary conditions at the walls were used). The dimensions of the periodic TIP3P water boxes and the number of water molecules were, respectively: CD6 (30.1, 29.1, 23.7 Å; 1,959), CD7 (31.3, 30.5, 24.3 Å; 2,208), CD8 (33.1, 33.7, 24.2 Å; 2,577), CD14 (43.4, 39.5, 31.9 Å; 1,470), CD21 (55.4, 46.7, 33.8 Å; 2,321), CD26 (46.9, 41.2, 34.1 Å; 1,694), CD28 (50.0, 42.6, 34.9 Å; 1,964). The SHAKE option was used for constraining bonds involving hydrogen atoms. Starting geometries were taken from crystal structures when available (CD14: CCDC100656, [14] CD26: CCDC115146 [17]), otherwise molecular graphics was used for building the initial geometries (CD21 was constructed by adding half CD14 to another unit of CD14, while CD28 was built by adding two new glucose units in the linear part of CD26). In all cases, the structures were fully minimized, with the steepest-descent minimization scheme, prior to their use for the MD runs, eliminating as much as possible the influence of the starting geometry quality on the computed results. The systems were heated from 0 to 300 K in 20 ps, and equilibrated for 30 ps. The productive runs were executed with a time-step of 1.0 fs at constant temperature and pressure (1 bar). The simulation time was 5.0 ns for both gas phase and water solution. Samplings were taken each 1 ps.

Information about structural variations during the simulations was obtained from analysis of rms deviation from the last structure.

[1] MM3* is a modified and adapted version of Allinger's MM3 force field

3 Results and discussion

3.1 Relative conformation of glucose units and nomenclature

All glucose units in small CDs (from CD5 to CD9) present a *syn* relative arrangement between them. However, the X-ray structures for the CD10 and CD14 already show some *anti* arrangements and also a new structural motif, the *kink* arrangement. In this motif, neighbor glucoses are in a *clinal* (or gauche) disposition (Fig. 2).

A nomenclature is proposed for designating the conformation of the macroring in LR-CDs based on two nomenclatures used for cycloalkanes (using the torsional angles signs [43] and the nomenclature of Dale [44]). A character (*s*, *a*, + or −) is assigned to each relative position of a pair of glucoses depending on the value of a virtual dihedral angle (called *flip* and defined by $O3(n)\cdots C4(n)\cdots C1(n+1)\cdots O2(n+1)$, Fig. 3). The ranges for the different portions were decided after the analyses of the results from the MDs were completed.

All glucoses are in principle equivalent by symmetry. The IUPAC rule for chiral molecules (i.e. as many *R* as possible at the beginning) will be followed and the name should

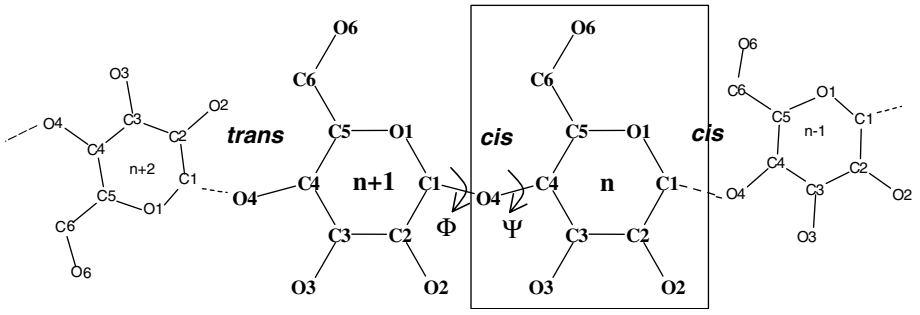

Fig. 1 Schematic representation for a CD fragment showing the atomic numbering. Each individual glucose unit is designated by a number "*n*". Atoms are identified by a *letter* and a *number* that indicates its position in the glucose

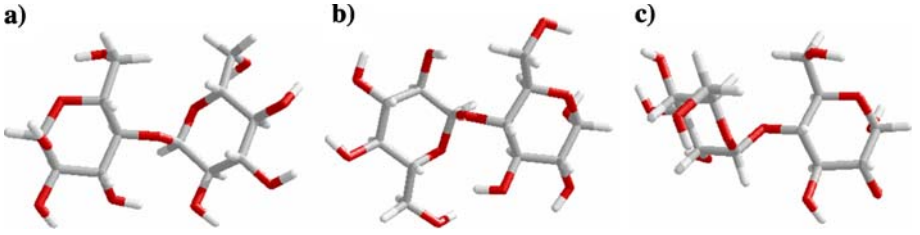

Fig. 2 Relative conformations of neighbor glucoses: **a** *syn*; **b** *anti*; **c** *kink*

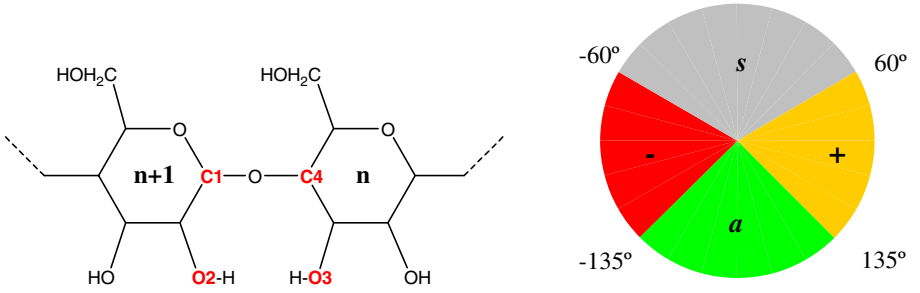

Fig. 3 Definition of the relative orientation for a pair of glucose units as derived from the *flip* dihedral angle (O3(n)$\cdots$C4(n)$\cdots$C1($n+1$)$\cdots$O2($n+1$))

start by that glucose residue having the largest number of *s* arrangement at the beginning of the name. When two or more names fulfill this requirement, the one with the *a*, + or − (in this order) as near as possible to the starting position, should be used. Since, this nomenclature produces long names (as many letters separated by comma as glucose units form the CD), whenever several identical characters are found, they are replaced by a number. As an example, the X-ray conformation of the CD10, ($s, s, s, s, a, s, s, s, s, a$), will be transformed into the ($4s, a, 4s, a$).

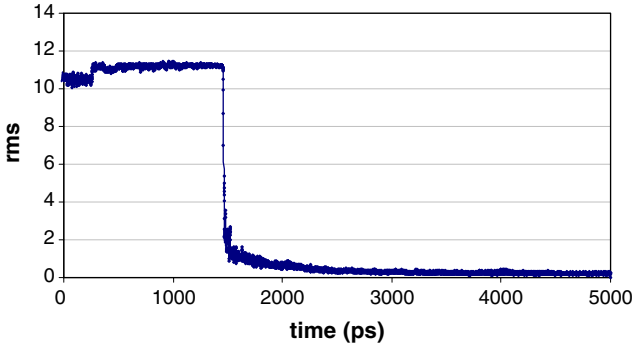

Fig. 4 Variation of the rms (Å) along the MD simulation (parm94) of the CD21 in gas phase. Two plateaus are observed: from 0 to 1.5 ns and from 2 to 5 ns

3.2 Number of conformers registered in the course of the molecular dynamics simulation

The analysis of the rms deviations from the last structure registered for each CD served to check and confirm the dynamism of the structural variations of the CDs (see Fig. 4 as an example). The outcome from the analysis indicates that conformational changes indeed take place. The most significant changes of the molecular shape were monitored because the skeletal C1–C2–C3–C4–O4 atoms were used for this analysis. Besides, the number of atoms used for estimating the rms is different in each CD (5×N, *N*=number of glucoses). Consequently, different criteria have to be used in each CD

for deciding when two structures are different (different plateaus)[2]. One average structure for each plateau was obtained with the CARNAL module of the AMBER program. For the sake of clarity in their representation, all averaged structures were energy minimized. Then they automatically become "not averaged", but rather we arrive at one of the representative structures of the set.

3.3 Starting conformations

Figure 5 depicts the starting structures used for the large CDs. A ribbon shows the shape created by the C4–O4–C1 skeleton and the glucoses are shown in colors accordingly to their relative conformation with the (n+1) glucose (see Fig. 3).

3.4 Conformational analysis with the parm94 force field

The native CDs were also studied with this force field, and their structures always remained at the *all-s* conformations.

The rms analysis of the gas phase simulations for CD14 indicates that two main structures (at 0.3 and 3.0 ns) are present, both having the shape of the number eight, producing two pseudocavities encircled by seven glucoses each. Both structures have the same nomenclature (i.e. they are the same) and present large number of *kink* arrangements, and a pair of *anti* arrangement (Fig. 6a). Two conformations were found in water solution (Fig. 6b, c). They are more unfolded, compared to the starting geometry (the two chains of the macroring do not cross as in the gas phase simulation), and differ in the conformation of the residues 9 and 11 (*s* in one and + in the other). The more unfolded geometry of the 3.0 ns conformation is due to the increased number of *kink* arrangements.

The simulations for CD21 in the gas phase rendered two structures (at 1.0 and 4.0 ns). The structure at 1.0 ns presents a folded macroring but it stretches during the simulation and appears almost completely extended at 4.0 ns. Crossings of chains and loops are not present. The elongation of the 4.0 ns structure can be associated with the increased number of consecutive glucoses adopting a − conformation (red in color). In contrast, only one conformation is present in water solution that closely resembles the starting structure.

Only two conformations were found in the gas phase simulation of CD26. The structure at 0.5 ns already presents important deviations from the initial geometry, due to the emergence of more *kink* arrangements. The macroring is completely extended at 4.0 ns, although the presence of a small loop of approximately six glucoses is observed at the extreme right side. Only one conformation is found in water solution which resembles the starting geometry. It seems that the water molecules play a decisive role for preserving the LR-CD conformation with this force field.

[2] The rms was computed considering only five atoms (C1-C2-C3-C4-O4) from each glucose. Structures were considered equal if their rms differ less than 0.015*5*(number of glucoses)

The simulations for CD28 in gas phase present three conformations (at 1.0, 2.0 and 4.0 ns, Fig. 9a–9c, respectively). The structure at 1.0 ns still maintains the two turns of the helix of the initial structure, although slight deviations are indeed present. The presence of *kink* glucoses is enhanced. Clearly, the structure becomes more extended with advancing the simulation. However (and this is valid also for the 2.0 and 4.0 ns snapshots), two loops are still observed. The difference between the latter two structures and the one at 1.0 ns remains in the ribbons that interconnect the two loops. These fragments are folded in the 1.0 ns structure (two glucoses have a + conformation), while they are extended in the other two structures (glucoses with a − conformation). As in the case of CD26, no significant deviations from the starting structure were found in the simulation of CD28 in water solution. The two structures (at 0.05 and 3.0 ns, Fig. 9d, e, respectively) have the same nomenclature and appear more disordered than the initial geometry and display more folded-over-itself fragments of the macroring, with larger number of kink glucoses.

3.5 Conformational analysis with the parm99 force field

As in the case of parm94, only one prevailing conformation is present in the three studied native CDs (CD6, CD7 and CD8). However, a deeper analysis of snapshots from the MD trajectory displays a slight bending of two diametrically opposed glucoses towards the interior of the cavity, loosing the pure truncated cone shape, and producing geometry in the form of an ellipse. The glucose units do not adopt a *kink* arrangement and the structures retain the crystalline conformation [8*s*].

The structure of CD14 during the gas phase simulation remains folded without significant deviations from the starting crystal geometry. However, the analysis of the rms indicates large movements of the macroring in water solution. The fluctuation of the rms is such that no clear definition of different structures could be done. This behavior is general for all the studied CDs with this force field. Figure 10 shows the rms fluctuation of CD14 in gas phase and in water solution, as an example.

To understand what happens in this system, the description of some snapshots we think are representative will be given. The geometry of the molecule remains similar to the initial structure at the beginning of the simulation, with a tendency for increasing the participation of *kink* arrangements. Lately, the structure begins to unfold (0.38 ns, Fig. 11a) and the presence of the +,−,+ sequence for three neighbor glucoses makes the central glucose to hinder the cavity entrance. Shortly later (0.68 ns, Fig. 11b), the +,−,+ arrangement disappears and a more extended cavity results, due to pairs of glucoses that adopt − arrangements. The cavity keeps this shape until the sequence −,+,− is again generated (2.2 ns), and one glucose unit blocks the entrance to the cavity. Finally, a structure similar to the starting one reappears (3.4 ns, Fig. 11c). In summary, tendencies for closure and opening of the cavity and its entrance were noticed. These result from changes

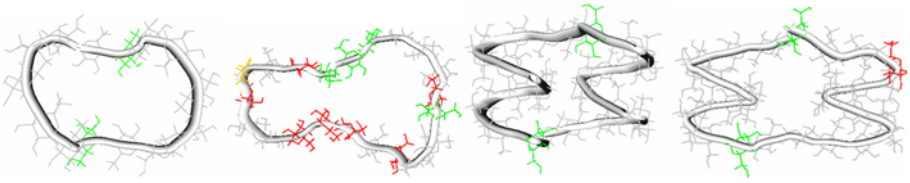

Fig. 5 Starting conformations: CD14 [6*s*, *a*, 6*s*, *a*]; CD21 [3*s*, 3*a*, −, *s*, +, −, 2*s*, 3−, *s*, −, 2*s*, *a*, −]; CD26 [12*s*, *a*, 12*s*, *a*]; CD28 [12*s*, *a*, 11*s*, −, 2*s*, *a*]

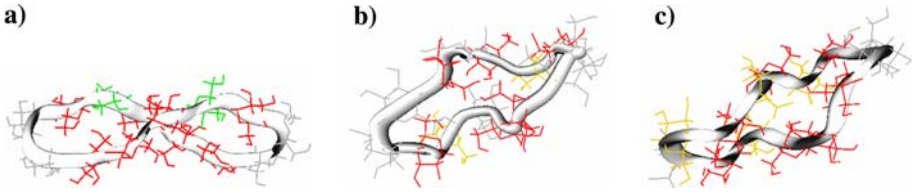

Fig. 6 Conformations obtained in the simulations of CD14 with parm94; **a** gas phase 3.0 ns [2*s*, 3−, *a*, −, 2*s*, 3−, *a*, −]; **b** water 0.5 ns [2*s*, 3−, +, −, 2*s*, −, *s*, −, +, −]; **c** water 3.5 ns [2*s*, 3−, +, −, *s*, +, −, +, −, +, −]

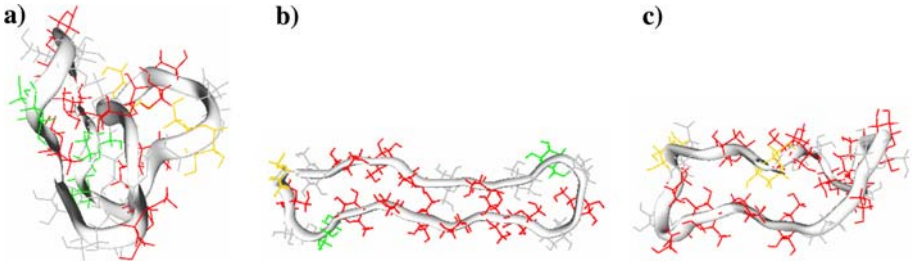

Fig. 7 Conformations obtained in the simulations of CD21 with parm94: **a** gas phase 1.0 ns [2*s*, −, *s*, −, *s*, +, −, *s*, +, 2−, *s*, 2*a*, −, *s*, −, *s*, *a*, −]; **b** gas phase 4.0 ns [*s*, 2−, *s*, +, −, *s*, *a*, 6−, *s*, −, *s*, *a*, *s*, −]; **c** water [*s*, −, *s*, 2−, +, −, *s*, +, −, *s*, 4−, *s*, −, *s*, 3−]

of the conformations of neighbor glucoses in a manner that the geometry of the macroring at the end of the simulation is similar to the starting crystal state geometry.

Only one conformation is obtained in the gas phase simulation of CD21 (Fig. 12a). Two loops of about six glucoses each and one larger loop positioned above the other two are present. It looks like the bigger loop covers the two smaller loops as a lid cover. As in the case of CD14, the conformational variations during the simulation of CD21 in water solution were examined by analyzing several snapshots. The structure undergoes variation along the simulation, adopting a shape from a rather square form (0.43 ns) to a more elongated geometry that results from an increase of the number of glucoses in conformation − (2.5 ns, Fig. 12b). It seems, however, that this geometry is not stable enough, and the structure at 2.8 ns is very similar to the 0.43 ns geometry (Fig. 12c). Finally, four loops are formed at 3.9 ns (Fig. 12d). Three of these loops are clearly defined by the neighborhood of three glucoses in − relative arrangement that point inward the cavity.

The CD26 does not present significant variations from the starting X-ray coordinates in the gas phase simulations. The initial loops are still kept although, as a whole, the structure is slightly more disordered. As in the two previous cases, the macroring experiences a tendency to unfold in water. The two helix turns present in the starting geometry can still be recognized in the structure at 2.15 ns (Fig. 13a). The structure at 3.45 ns becomes more elongated. (Fig. 13b). The stretch is due to the sequence of glucoses with prevailing − arrangement. The overall shape of the macroring remains almost the same until the end of the simulation (4.95 ns, Fig. 13c). The conformation of CD26 in aqueous solution presents a cavity with an elongated shape. Therefore, only in the case of an additional agent (external), like the crystal field packing forces, or a complexing molecule that can provoke a conformational change of CD26 in water solution, can we attain a structure that presents a double antiparallel helix as observed in the crystal state.

Two conformations can be distinguished from the simulation of CD28 in gas phase. Both are very similar to the starting structure. The CD28 also presents a tendency to unfold in water solution, although to a lesser extent than CD26 does. The two turns of the antiparallel helix of the starting structure still can be seen at 3.0 ns, although an increase in the number of glucoses in *s* arrangements is observed (Fig. 14a). The structure unfolds at 3.9 ns (Fig. 14b). A loop is present in one part of the structure, while a lobe composed of a set of − arrangements is also formed. The macroring retains the loop

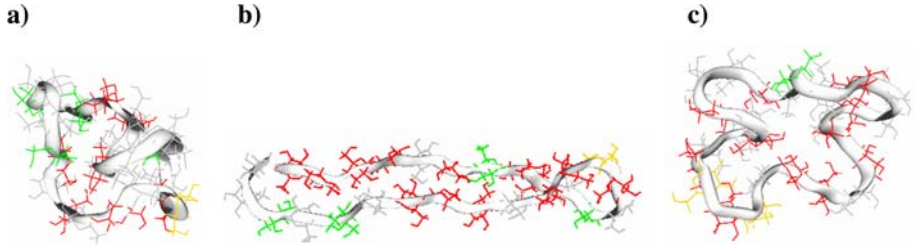

Fig. 8 Conformations obtained in the simulations of CD26 with parm94: **a** gas phase 0.5 ns [3*s*, −, 2*s*, −, *s*, −, *a*, 2*s*, *a*, 2*s*, *a*, −, *s*, −, *s*, −, *s*, +, −, *s*, *a*]; **b** gas phase 4.0 ns [2*s*, *a*, *s*, *a*, *s*, 3−, *s*, 2−, +, −, *s*, *a*, *s*, 2−, *a*, −, *s*, 4−]; **c** water [2*s*, −, *s*, −, 2*s*, −, +, −, +, *s*, 2−, *s*, 2−, *s*, −, 2*s*, −, *s*, *as*, −]

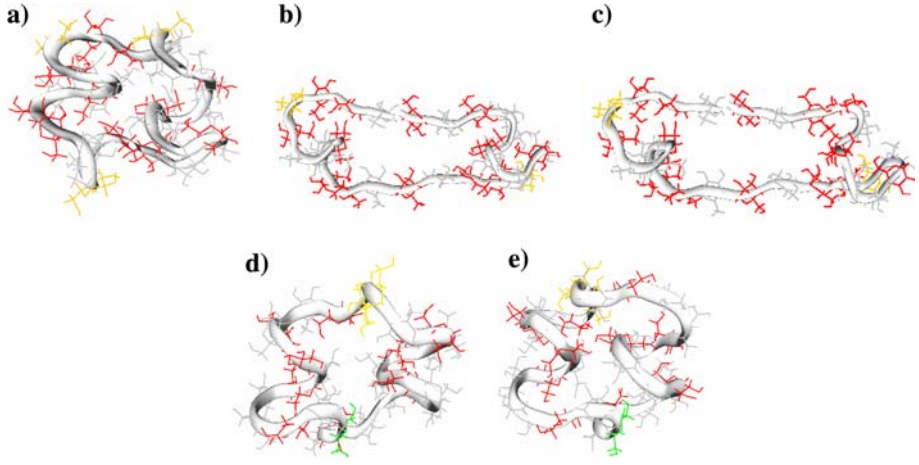

Fig. 9 Conformations obtained in the simulations of CD28 with parm94: **a** gas phase 1.0 ns [2*s*, −, 2*s*, −, 2*s*, −, 2*s*, +, *s*, 2−, +, −, 2*s*, −, *s*, 2−, *s*, +, *s*, 2−]; **b** gas phase 2.0 ns [3*s*, −, *s*, −, *s*, 2−, +, −, 2*s*, −, *s*, 2−, *s*, −, *s*, 2−, +, *s*, −, 2*s*, −]; **c** gas phase 4.0 ns [2*s*, 2−, *s*, 2−, *s*, −, *s*, 2−, +, −, 2*s*, −, *s*, 2−, *s*, −, *s*, 2−, +, *s*, −]; **d** water 0.05 ns [2*s*, −, 2*s*, −, *s*, −, 2*s*, −, *s*, 2+, −, *s*, −, *s*, −, *s*, −, 2*s*, −, *s*, *a*, *s*, −]; **e** water 3.0 ns [2*s*, −, 2*s*, −, *s*, −, 2*s*, −, *s*, 2+, −, *s*, −, *s*, −, *s*, −, 2*s*, −, *s*, *a*, *s*, −]

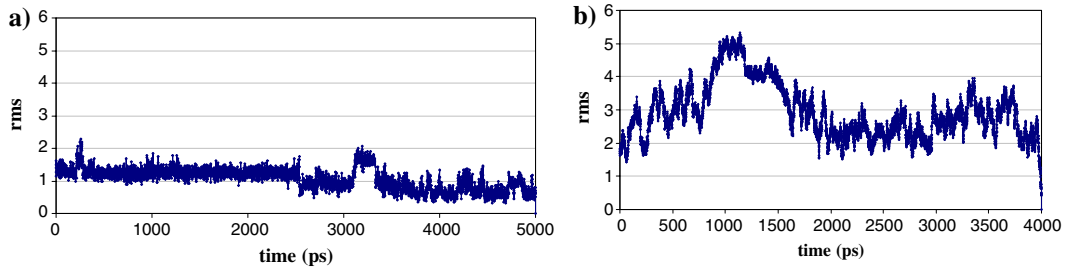

Fig. 10 Fluctuation of rms (Å) for the simulations of CD14: **a** gas phase; **b** in water solution

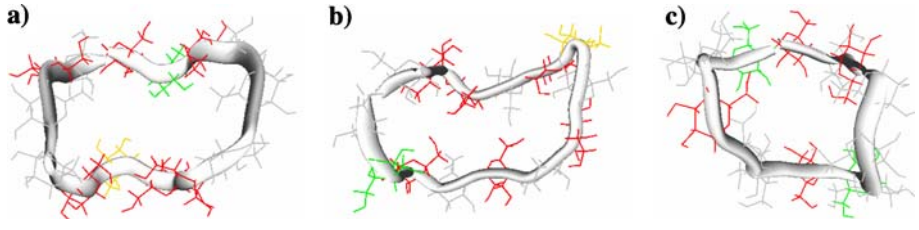

Fig. 11 Conformations obtained in the simulations of CD14 with parm99 in water solution: **a** 0.38 ns [3*s*, 2−, +, −, 2*s*, −, *s*, −, *a*, −]; **b** 0.68 ns [2*s*, *a*, −, *s*, −, *s*, −, *s*, +, −, *s*, 2−]; c) 3.4 ns [3*s*, *a*, −, 2*s*, −, 2*s*, *a*, −, *s*, −]

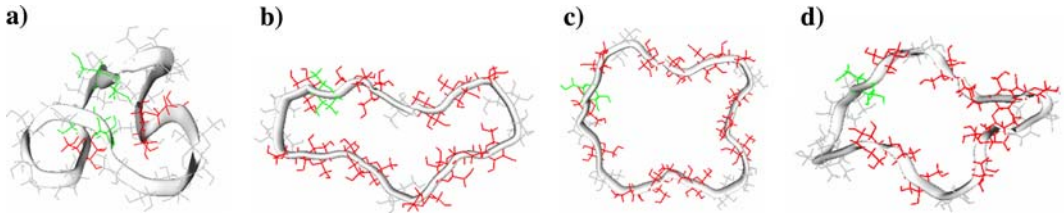

Fig. 12 Conformations obtained in the simulations of CD21 with parm99: **a** gas phase [5*s*, *a*, 2*s*, *a*, −, 4*s*, *a*, 4*s*, 2−]; **b** water solution 2.5 ns [2*s*, 3−, *s*, 2−, *a*, −, 2*s*, 4−, *s*, 4−]; **c** water solution 2.8 ns [2*s*, 3−, *s*, 4−, *s*, −, *a*, −, *s*, −, *s*, 4−]; **d** water solution 3.9 ns [5*s*, 3−, *s*, 3−, 2*s*, 4−, *s*, −, *a*]

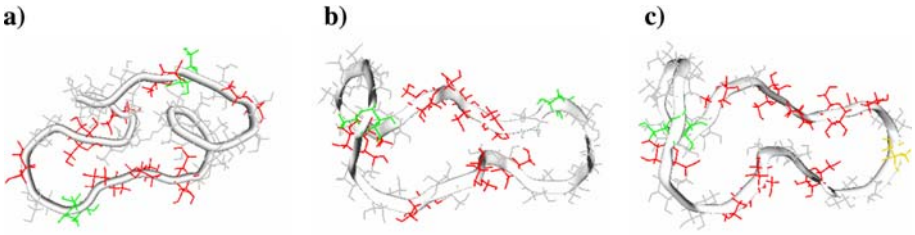

Fig. 13 Conformations obtained in the simulations of CD26 with parm99 in water solution: **a** 2.15 ns [7*s*, −, *s*, *a*, −, 4*s*, −, 2*s*, −, *s*, −, *s*, *a*, 3−]; **b** 3.45 ns [5*s*, *a*, −, 2*s*, −, *s*, −, *s*, −, 4*s*, *a*, *s*, 4−, *s*, −]; **c** 4.95 ns [7*s*, *a*, −, 2*s*, 2−, 2*s*, −, 2*s*, +, *s*, 4−, *s*, −]

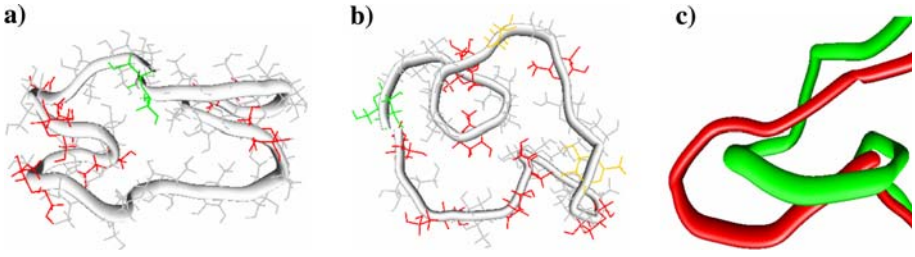

Fig. 14 Conformations obtained in the simulations of CD28 with parm99 in water solution: **a** 3.06 ns [5*s*, *a*, 3*s*, −, *s*, −, *s*, −, 2*s*, −, *s*, *a*, 5*s*, −, 2*s*, −]; **b** 3.91 ns [4*s*, −, *s*, +, 2*s*, −, *s*, +, −, 2*s*, −, 2*s*, −, 2*s*, *a*, −, *s*, −, *s*, 2−]; **c** fragment representing two different spatial arrangements of the lobe (*green* at 3.9 ns and *red* at 4.65 ns)

until the end of the simulation. Meanwhile the lobe slightly changes its position in space (Fig. 14c). The geometry at the end of the simulation closely resembles that at 3.9 ns.

3.6 Conformational analysis with glycam2000a force field

Parm94 and parm99 force fields are mostly oriented to applications on proteins, while glycam2000a is a force field specifically developed for applications on oligosaccharides and glycoproteins. This parametrization has developed specific parameters for the anomeric carbon atom C1 and for the glycosidic oxygen O4. The need to define new parameters for these atoms emerged from the observation that the lengths of bonds C1–O1 and C1–O4 vary with the configuration of the pyranose ring [39]. Thus, in order to arrive at more adequate modeling of these bonds and to better reproduce vibrational frequencies, it appeared necessary, in analogy with MM3 and MM4, to consider also cross-terms of the type stretch-torsion. However, such cross-terms had never been included in the set of potential energy functions used in AMBER. An alternative approach is to define new atom types specific for each anomer, that has the advantage of being mathematically simpler and, accordingly, computationally faster.

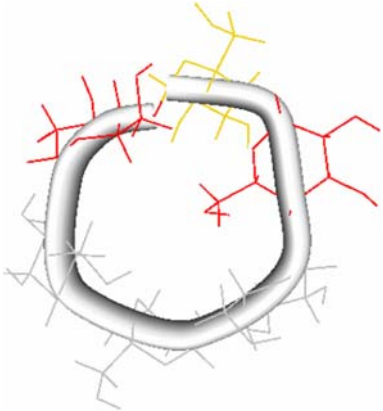

Fig. 15 Structure of the CD6 at 0.43 ns in aqueous solution with the glycam2000a force field [3*s*, −, +, −]

MD runs for the native CDs were performed only in water solution. In contrast with the results previously obtained, they are now more flexible and glucoses with − and + relative arrangements were detected. Figure 15 displays, as

an example, the geometry of CD6 at 0.43 ns. The macroring lost the well-known shape of a truncated cone and contains the sequence −,+,− that places one glucose residue blocking the cavity entrance.

Two different conformations of CD14 were distinguished in the gas phase simulations (Fig. 16). One of them is present during the first 1.25 ns simulation, while the other characterizes the remaining time until the end of the simulation. Unfolding from the starting geometry characterizes the two forms, with more pronounced variation in the second conformation that contains more glucoses in *kink* arrangement.

The deviation from the starting structure is not appreciable during about one half of the simulation in water, although greater number of *kink* glucoses is present. More important changes appear at 1.7 ns. The structure acquires the shape of number eight (Fig. 16c) with two subunits of six glucoses each. Undergoing small variations, this structure remains until the end of the simulation.

The simulation of CD21 in the gas phase required to be expanded from the original 4.0 ns to 7.0 ns due to a significant change in the rms curve at about 5.0 ns. Two conformations were detected (the first exists in the ranges 0–4.3 ns and 5.4–7.0 ns, and the second in the range 4.3–5.4 ns, Fig. 17a, b, respectively). Both structures are similar and contain three consecutive curls of approximately six glucoses each. Larger number of *kink* glucoses are present in the second structure.

The simulation of CD21 in water solution follows the previous behavior: continuous conformational changes without defined conformations. The deviations from the starting geometry start at 0.33 ns where quite open geometry is observed containing three curls (Fig. 17c). The structure at 2.0 ns presents the curls more elongated, as a result of the increase of glucoses in + and − conformations (Fig. 17d). With very small changes the molecule remains in this conformation until the end of the simulation (Fig. 17e).

The rms analysis yielded three different conformations from the gas phase simulations of CD26. Again similarities between the three structures are present. All have three consecutive loops that form a left-handed single helix (Fig. 18a). This finding is in accordance with the study of Kitamura and collaborators [45] which describes CD26 in solution to have a circular single helix.

In water solution, CD26 lost the crystal conformation (as with parm99). One of the loops disappeared at 1.1 ns, and a curl was formed in its place (Fig. 18b). CD26 keeps changing slowly and the structure at 2.7 ns is more stretched as a consequence of the presence of more glucoses in a − relative arrangement (Fig. 18c). Finally, the structure at 4.3 ns is practically the same as the previous one, and thus it seems that this structure is preferentially retained in the simulation (Fig. 18d).

Three conformations were detected in the gas phase of CD28. Two loops are present in the three cases. The size and the location of the loops vary from one to the other structure. Figure 19a shows the overlap of two structures and demonstrates that the molecule remains basically with the same shape.

As in all pre1ceding cases, the rms analysis of the simulations in water for CD28 does not allow to determine clear conformations. What is the most representative is that the structure lost one of the loops at 0.7 ns, and step by step increased the number of *kink* glucoses (Fig. 19b). The structure has stretched at 3.3 ns and, as in the other cases the elongation is accompanied by an increase of consecutive glucoses with − relative arrangement (Fig. 19c). Finally, this structure is preserved during the rest of the simulation.

3.7 Conformational analysis with MM3* force field

The MM3 force field is one of the most extensively used for computing structures of organic molecules. Although it is not included into the AMBER package, many molecular modeling systems contain this force field. The MD simulations of CD14, CD21, CD26 and CD28 were also performed using the Macromodel software package.

The analysis of the rms indicated that only one single conformation was obtained for all cases. CD14 remained unchanged along the simulation. CD26 and CD28 kept the shape, although some more *kink* glucoses were present in the conformation observed. Finally, CD21 was the one suffering from larger changes originating from the presence of several glucoses in − relative arrangement.

3.8 Analysis of the shape

The conformations discussed earlier were characterized also with their radius of gyration (Rgyr) and asphericity (ASP) [46]. The program DRAGON [47] was used to estimate the descriptors of the shape. The following approach was used: (1) the average structure was computed for each conformation, (2) the resulting structure was minimized with 50 iterations and (3) due to limitations for the number of atoms in the program, only the fragment C1–O4–C4 was used in each case.

Table 1 contains the results obtained in all the gas phase simulations. Most of the Rgyr values increase from CD6 to CD28, except for CD28 with parm94, in which case the value is smaller than for CD26. This fact indicates that CD28 is more compact than CD26 (due to the loops, CD28 does not have a totally extended shape). The comparison with the experimentally available values for native CDs suggests that the computed data are always smaller than the experimental ones, probably due to the fact of considering only the C1–O4–C4 fragment. This tendency is followed also by the LR-CDs. It is worth to mention here that the Rgyr for CD21 and CD26 with parm94 largely exceeds the corresponding experimental values. It can be concluded that these conformations do not correspond to the experimental conformations. The computed Rgyr with the other force fields are larger for CD21 and smaller for CD26. We have to make a note here that when parm99 and MM3* are used, the number of conformers covered is much smaller than with the other force fields. Parm99 and glycam2000a produce values for Rgyr reasonably near to

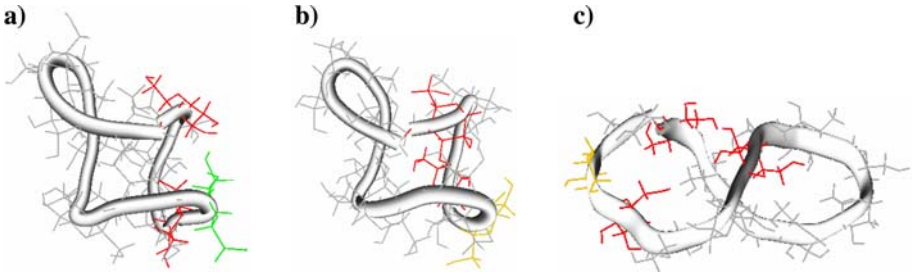

Fig. 16 Conformations obtained in the simulations of CD14 with glycam2000a: **a** gas phase 0.7 ns [6*s*, *a*, −, 3*s*, −, 2*s*]; **b** gas phase 4.0 ns [6*s*, +, −, 2*s*, 2−, 2*s*]; **c** water solution 1.7–4.0 ns [7*s*, −, *s*, +, −, 2*s*, −]

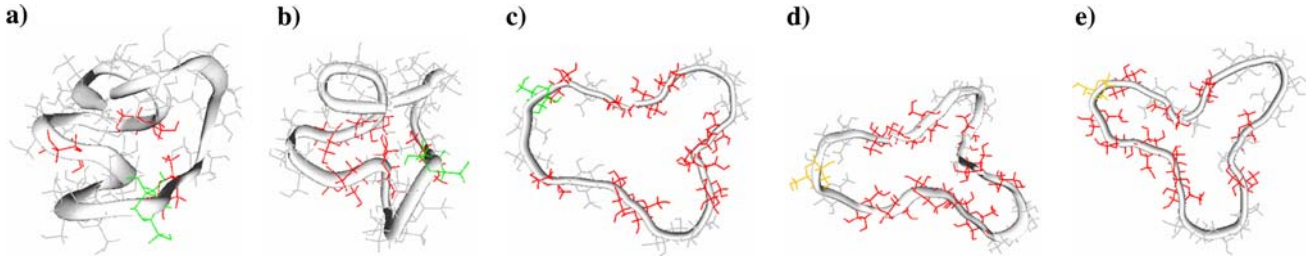

Fig. 17 Conformations obtained in the simulations of CD21 with glycam2000a: **a** gas phase 2.0 ns [9*s*, −, 3*s*, −, 5*s*, *a*, −]; **b** gas phase 4.5 ns [9*s*, 2−, 2*s*, 2−, 4*s*, *a*, −]; **c** water solution 0.33 ns [3*s*, 3−, *s*, −, *a*, 2*s*, −, *s*, 2−, 2*s*, 4−]; **d** water solution 2.0 ns [3*s*, 3−, 2*s*, +, *s*, 5−, 2*s*, 2−, *s*, −]; **e** water solution 3.9 ns [3*s*, −, *s*, −, *s*, −, +, 5−, 3*s*, −, *s*, −, *s*]

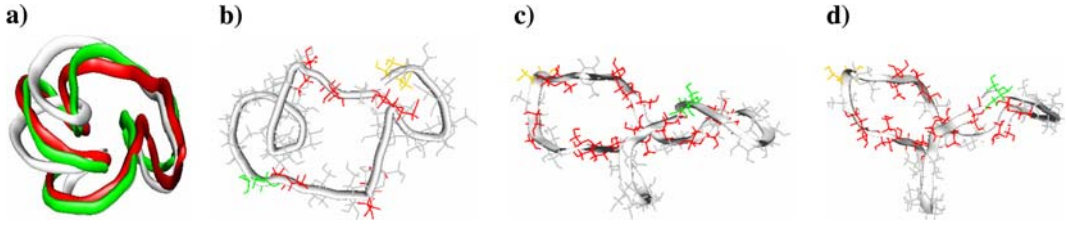

Fig. 18 Conformations obtained in the simulations of CD26 with glycam2000a: **a** gas phase [25*s*, *a*]; **b** water solution 1.1 ns [9*s*, *a*, −, 2*s*, −, 2*s*, −, 5*s*, +, *s*, 2−]; **c** water solution 2.7 ns [5*s*, 2−, *s*, −, *s*, +, −, *s*, 2−, 2*s*, 2−, 4*s*, *a*, 2−]; **d** water solution 4.3 ns [5*s*, 3−, *s*, +, *s*, 2−, 3*s*, 2−, 4*s*, *a*, 3−]

the experimental ones in all cases except for CD21 (probably due to the selected starting geometry). The ASP values of the native CDs are 0.2–0.3. Thus, the skeleton C1–O4–C4 can be described as a cylinder. No clear tendencies are estimated for the larger CDs, but in general the value of this descriptor is smaller than the corresponding value in the small CDs. Thus, their C1–O4–C4 skeleton is more spherical. The values of ASP tend to 1 for most of the LR-CDs when parm94 is used.

The results from the same analysis for the simulations in water solution are presented in Table 2. Different numbers of values are given for each CD, depending on the graph obtained in rms analysis. The radius of gyration increases from CD6 to CD21 with parm94, but from this value further remains a constant value about 17.6. These trends indicate that the size of the macromolecule increases with the number of residues, but at the same time the compactness of the structures (the CDs tend to fold over themselves) also increases. Other force fields do not follow this trend. For example, CD26 and CD28, which tend to fold with glycam2000a, have values for the Rgyr and ASP which increase from one to another structure. Finally, the computed Rgyr of CD21 and CD26 is usually larger than the corresponding experimental value, and thus, from the structural point of view, none of the computed structures correctly represents the experimental conformation.

3.9 Evaluation of the force fields

The stretching parameters do not vary noticeably between the force fields and will not be discussed here. However, differences exist between AMBER and MM3* (under Macromodel) programs in the equations used for evaluating the bending contribution. Although significant differences are not observed for the values of most of the reference angles, the most significant variations are in the values of the reference angle for the O1–C1–O4 fragment. This reference angle is similar in MM3* (108.6°) and in glycam2000a (110.7°), but extremely low and high values characterize parm99 (101.0°) and parm94 (126.0°), respectively. Moreover, the corresponding force constants (k_σ) also vary significantly from one force field to another (80.0, 160.0 and 110.7 kcal/mol rad^2 for

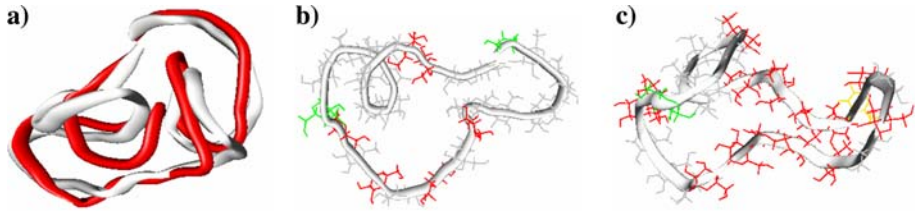

Fig. 19 Conformations obtained in the simulations of CD28 with glycam2000a: **a** gas phase [7*s*, −, 2*s*, −, 3*s*, *a*, 2−, 2*s*, −, +, 6*s*, *a*,]; **b** water solution 0.67 ns [8*s*, *a*, −, *s*, −, *s*, −, *s*, −, 7*s*, *a*, 2*s*, 2−]; **c** water solution 3.29 ns [4*s*, −, 2*s*, *a*, 2−, 3*s*, 5−, *s*, −, *s*, −, +, 5−]

Table 1 Shape descriptors for the average structures of CD6, CD7, CD8, CD14, CD21, CD26 and CD28 obtained in gas phase

	CD6	CD7	CD8	CD14	CD21	CD26	CD28
RGyr (Å)							
Exp.[a]	6.0	6.7	7.3	–	11.5	19.6	–
Parm94	4.2	6.7	6.2	12.8	23.3	31.4	28.3
Parm99	4.2	5.4	6.3	9.1	13.2	16.6	18.3
Glycam 2000a	4.3	5.5	6.5	9.2	13.0	15.9	17.0
MM3*	–	–	–	10.4	15.5	17.5	18.9
ASP							
Parm94	0.2	0.2	0.3	0.7	0.9	0.9	0.7
Parm99	0.2	0.2	0.2	0.1	0.1	0.2	0.2
Glycam 2000a	0.2	0.2	0.2	0.1	0.1	0.1	0.1
MM3*	–	–	–	0.2	0.2	0.1	0.1

[a]Determined by small angle X-ray scattering at 25°C [45]

parm94, parm99 and glycam2000a, respectively, and 0.54 mdyn/rad^2 for MM3*). The energy cost to vary slightly the value of this angle in MM3* is lesser than in the other three force fields, and this results from the utilization of different functions in this force field. In summary, the deformation of angle O1–C1–O4 with the parameters used in parm94 requires less energy and the equilibrium value of the angle exceeds the reference values for this angle in the other force fields. The effect of these different bending parameters on the final average structure obtained for LR-CDs was dramatic. A new additional simulation with parm94 was performed on CD21 (as an example). The stretched average structure obtained in the gas phase simulation was used as a starting point and the parm99 bending parameters for the O1–C1–O4 angle were used. The macrocycle folds during the simulation. Figure 20 displays the structure obtained at the end of this simulation. Thus, despite obtaining positive results with parm94 in the modeling of small CDs, this force field does not model the LR-CDs correctly.

More important differences between force fields are found in the torsional parameters. We will focus mostly on the torsional parameters for the rotations about the glycosidic bonds (C1–O4–C4) which actually determine the conformational behavior of the CDs. We have to note firstly that the dihedral O1–C1–O4–C4 was not parameterized within parm94, and rather generic values were used for this parameter (X–CT–OS–X; $N = 3$, $V_{n/2} = 1.15$, $\gamma = 0$, $n = 3$). The differences between force fields in equations and parameters for estimating the torsional contribution to the total energy are important, and we evaluated the profiles for the rotation energies around bonds C1–O4 and O4–C4 (Fig. 21) taking into account all dihedral angles that contribute.

Figure 21a shows the profiles for the rotation around bond C1–O4. The torsional profiles of parm94 and MM3* are similar: both are symmetrical and display three energy maxima and three minima. In contrast, those of parm99 and glycam2000a are not symmetrical: (1) the maximum is at about −90° with glycam2000a, while it is at about −120° for parm99; (2) parm99 presents energy minima at about −50°, 60° and 160°, while glycam2000a has the energy minima around −10°, 110° and −150°.

Figure 21b shows the profile for the rotation around the O4–C4 bond. Very similar curves characterize this rotation with parm94, parm99 and MM3*, while glycam2000a deviates significantly (one energy maximum at 0° and one single energy minimum at 180° are present). It becomes evident from these differences why the number of tilted glucoses in the MD simulations in gas phase is smaller with glycam2000a than with parm99. Interestingly, in water solution the behavior of this force field is reversed. Thus it is reasonable to conclude that the non-bonded interactions between the solvent and the solute are strong enough as to over-balance the energy cost for variations of this angle.

We have to remind that the molecules studied have large number of hydroxyls that form many intramolecular hydrogen bonds. Besides, hydrogen bonds exist also between the solute and the solvent molecules. The formation of these hydrogen bonds will be examined, although in a qualitative sense. This analysis could provide the basis to explain the different behavior observed between the simulations in gas phase and in solution using the glycam2000a force field.

Let us take CD14 as an example. The total hydrogen bond population per residue for CD14 in gas phase is 3.2 hydrogen bonds in average (all intramolecular), while the same quantity

Table 2 Shape descriptors for the average structures of CD6, CD7, CD8, CD14, CD21, CD26 and CD28 obtained in water solution

		CD6	CD7	CD8	CD14		CD21		CD26		CD28	
	Experimental	6.0	6.7	7.3	-		11.5		19.6		-	
	Parm94	4.0	4.8	5.9	10.9		17.8		17.7		17.5	
RGyr (Å)	Parm99	4.2	5.2	6.3	(158)	12.1	(150)	19.6	(650)	17.6	(365)	20.6
					(381)	11.3	(434)	20.7	(2157)	18.6	(2507)	20.7
					(676)	11.4	(1150)	17.3	(3451)	24.2	(3067)	19.8
					(1125)	11.7	(2285)	16.7	(4046)	22.3	(3917)	20.7
							(2492)	20.6	(4952)	23.4	(4659)	23.0
							(3881)	19.7			(4950)	21.2
	Glycam 2000a	4.7	5.4	6.5	(186)	12.0	(60)	21.1	(421)	18.4	(145)	19.7
					(3820)	10.8	(334)	21.9	(1469)	19.6	(671)	23.6
							(2025)	19.4	(4250)	22.3	(4930)	26.0
							(3930)	19.0				
	Parm94	0.2	0.2	0.2	0.4		0.3		0.2		0.1	
ASP	Parm99	0.3	0.2	0.2	(158)	0.2	(150)	0.2	(650)	0.2	(365)	0.2
					(381)	0.1	(434)	0.3	(2157)	0.3	(2507)	0.2
					(676)	0.2	(1150)	0.3	(3451)	0.5	(3067)	0.3
					(1125)	0.3	(2285)	0.1	(4046)	0.4	(3917)	0.2
							(2492)	0.5	(4952)	0.4	(4659)	0.3
							(3881)	0.4			(4950)	0.2
	Glycam 2000a	0.3	0.2	0.2	(186)	0.1	(60)	0.4	(421)	0.1	(145)	0.2
					(3820)	0.4	(334)	0.3	(1469)	0.2	(671)	0.3
							(2025)	0.3	(4250)	0.3	(4930)	0.4
							(3930)	0.2				

Time of existence of each structure is given in parenthesis (in ps)

in water solution is 14.1 (1.9 intramolecular hydrogen bonds and 12.2 hydrogen bonds with the solvent). It is reasonable to assume that this extra stabilization from hydrogen bonding in solution is good enough to compensate the energy expenditure for passing the barrier of 6.0 kcal/mol glucose observed for the rotations about bonds O4–C4.

4 Conclusions

The results of the conformational study of the CDs with degree of polymerization of 14, 21, 26 and 28 vary significantly depending on the force field used.

The observed tendency of parm94 to stretch the molecule (extended conformation) in gas phase is due to the bending parameters used for angle O1–C1–O4. This force field was not producing structures capable of explaining the existing experimental data.

On the other hand, the gas phase simulation with MM3* and parm99 force fields yields structures that maintain a conformation which is very similar to the initial conformation. The reason is enhanced energy stabilization from intramolecular hydrogen bonding.

Although a much more extended region of the potential energy hypersurface was explored with parm99 in water solution, the radius of gyration exceeds the experimental values. It is thus hardly believable that these structures could explain the experimental data. The observed fluctuations in the computed radius of gyration and asphericity may indicate not converged simulations (not long enough).

Finally, it has been concluded that the higher flexibility of the computed CDs, when glycam2000a is used, is due to the hydrogen bond interactions between the solvent and the solute. In contrast, in the gas phase simulations with glycam2000a structures were found which better correlate with the experimental data (see as an example the structure of CD26 which presents a sequence of three consecutive loops that form left-oriented helix with close-to-the-experimental radius of gyration). At the same time, the simulations of CD21 with glycam2000a did not afford structures that correlate with the experimental data, suggesting the starting geometry as being responsible for the failure. Although MD can be seen as a tool for conformational search, it can overcome

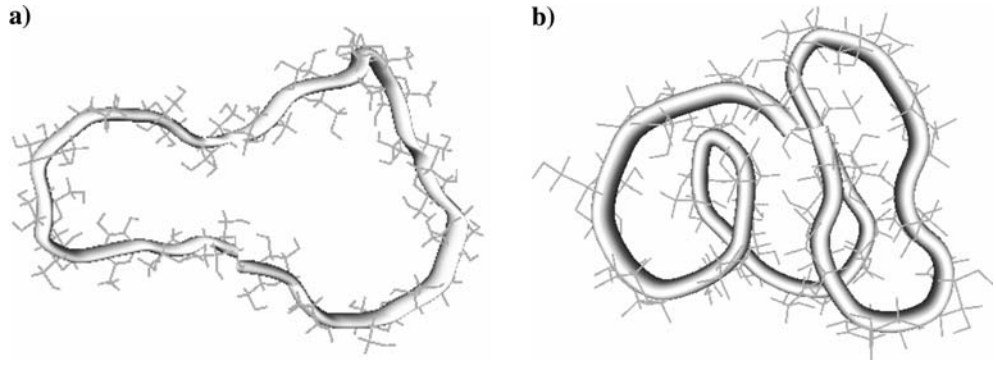

Fig. 20 Effect of using two different O1–C1–O4 bending parameters in the parm94 force field over the structure of CD21: **a** starting structure; **b** average structure when the O1–C1–O4 bending parameters of parm99 are used

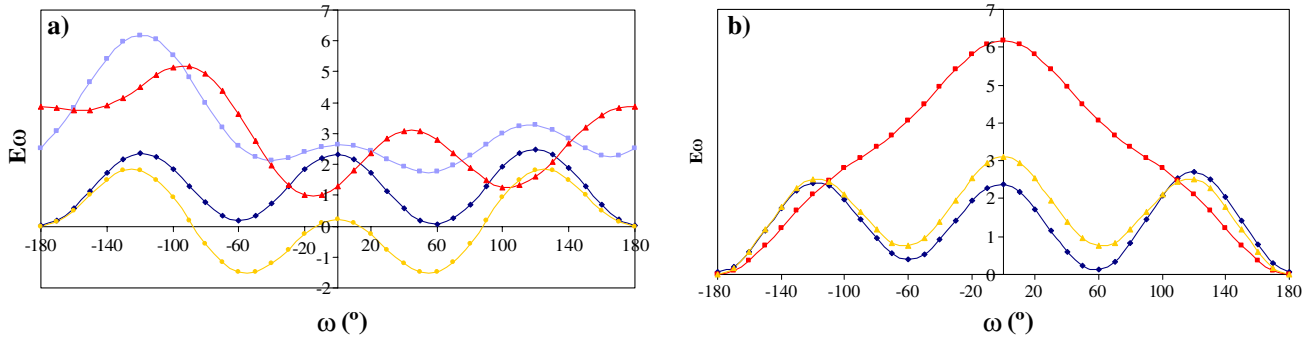

Fig. 21 Energy profiles (kcal/mol) corresponding to the rotation around bonds (*dark blue*=parm94; *light blue*= parm99; *red*=glycam2000a; *yellow*=mm3*): **a** C1–O4; **b** O4–C4

only low energy barriers, and thus the initial geometry can significantly influence the results.

The simulations performed in this work are too short and insufficient to simulate what it is observed at the NMR time scale (one signal for each type of carbon). However, we have enough information to conclude that LR-CDs are continuously changing the macroring shape. Since all glucoses are identical, all should average in the NMR time scale. Moreover, each one of the glucoses can occupy any position in the macroring, thus they will finally appear equivalent. There is a fast exchange among all glucoses that averages the chemical shifts for each carbon atom.

References

1. Saenger W, Jacob J, Gessler K, Steiner T, Hoffmann D, Sanbe H, Koizumi K, Smith SM, Takaha T (1998) Chem Rev 98:1787
2. Larsen KL (2002) J Incl Phenom 43:1
3. Ueda H (2002) J Incl Phenom 44:53
4. Motohama S, Ishii E, Endo T, Nagase H, Ueda H, Takaha T, Okada S (2001) Biologicheskii Zhurnal Armenii 53:27
5. Ueda H, Wakisaka M, Nagase H, Takaha T, Okada S (2002) J Incl Phenom 44:403
6. Zheng M, Endo T, Zimmermann W (2002) J Incl Phenom 44:387
7. Endo T, Zheng M, Zimmermann W (2002) Austr J Chem 55:39
8. Qi Q, She X, Endo T, Zimmermann W (2004) Tetrahedron 60:799
9. Endo T, Ueda H, Kobayashi S, Nagai T (1995) Carbohydr Res 269:369
10. Ueda H, Endo T, Nagase H, Kobayashi S, Nagai T (1996) J Inc Phenom 25:17
11. Larsen KL, Mathiesen F, Zimmermann W (1997) Carbohydr Res 298:59
12. French D, Pulley AO, Effenberger JA, Rougvie MA, Abdullah M (1965) Arch Biochem Biophys 111:153
13. Fujiwara T, Tanaka N, Kobayashi S (1990) Chem Lett 739
14. Jacob J, Gessler K, Hoffmann D, Sanbe H, Koizumi K, Smith SM, Takaha T, Saenger W (1998) Angew Chem Int Ed Eng 37:606
15. Ueda H, Endo T, Nagase H, Kobayashi S, Nagai T (1996) J Incl Phenom Mol Recognit Chem 25:17
16. Jacob J, Gessler K, Hoffmann D, Sanbe H, Koizumi K, Smith SM, Takaha T, Saenger W (1999) Carbohydr Res 322:228
17. Gessler K, Usón I, Takaha T, Krauss N, Smith SM, Okada S, Sheldrick GM, Saenger W (1999) Proc Natl Acad Sci USA 96:4246
18. Nimz O, Gessler K, Usón I, Saenger W (2001) Carbohydr Res 336:141
19. Bettinetti G, Sorrenti M (2002) Thermochim Acta 385:63
20. Terada Y, Yanase M, Takata H, Takaha T, Okada S (1997) J Biol Chem 272:15729
21. Takaha T, Yanase M, Takata H, Okada S, Smith SM (1996) J Biol Chem 271:2902
22. Larsen KL, Endo T, Ueda H, Zimmermann W (1998) Carbohydr Res 309:153
23. Kitamura S, Nakatani K, Takaha T, Okada S (1999) Macromol Rapid Commun 20:612
24. Furuishi T, Endo T, Nagase H, Ueda H, Nagai T (1998) Chem Pharm Bull 46:1658
25. Dodziuk H, Ejchart A, Anczewski W, Ueda H, Krinichnaya E, Dolgonos G, Kutner W (2003) Chem Commun 986
26. Koehler JEH, Saenger W, van Gunsteren WF (1987) Eur Biophys J 15:211
27. Koehler JEH, Saenger W, van Gunsteren WF (1998) Eur Biophys J 16:153

28. Kitamura S, Isuda H, Shimada J, Takada T, Takaha T, Okada S, Mimura M, Kajiwara K (1997) Carbohydr Res 304:303
29. Lipkowitz KB (1998) Chem Rev 98:1829
30. Shimada J, Handa S, Kaneko H, Takada T (1996) Macromolecules 29:6408
31. Shimada J, Kaneko H, Takada T, Kitamura S, Kajiwara K (2000) J Phys Chem B 104:2136
32. Momany FA, Willett JL (2000) Carbohydr Res 326:194
33. Momany FA, Willett JL (2000) Carbohydr Res 326:210
34. Ivanov PM, Jaime C (2004) J Phys Chem B 108:6261
35. Allinger NL, Yuh YH, Lii J-H (1989) J Am Chem Soc 111:8551
36. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC (1990) J Comput Chem 11:440
37. Cornell WD, Cieplack P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179
38. Wang J, Cieplack P, Kollman PA (2000) J Comput Chem 21:1049
39. Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B (1995) J Phys Chem 99:3832
40. Case DA, Pearlman DA, Caldwell JW, Cheathan TE III, Wang J, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Tsui V, Gohlke H, Radmer RJ, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner PK, Kollman PA (2002) AMBER 7, University of California, San Francisco
41. Beà I, Cervelló E, Kollman PA, Jaime C (2001) Comb Chem High Throughput Screening 4:605
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926
43. Burkert U, Allinger NL (1982) Molecular mechanics. ACS Monograph 177, Washington, ACS, pp 107–108
44. Dale J (1973) Acta Chem Scand 27:1115
45. Kitamura S (2000) Cyclic polymers, J.A. Semlyen Edition, p 125
46. Arteca GA (1996) Rev Comput Chem 9:191
47. Todeschini R, Consonni V, Mauri A, Pavan M. http://www.disat.unimib.it/chm/