J Incl Phenom Macrocycl Chem (2007) 57:683–688
DOI 10.1007/s10847-006-9265-0

ORIGINAL ARTICLE

Molecular dynamics study of host–guest interactions in cyclodextrins: methodology and data analysis for a comparison with solution data and the solid-state structure

Giuseppina Raffaini · Fabio Ganazzoli

Received: 15 May 2006 / Accepted: 20 October 2006 / Published online: 9 March 2007

Abstract A general method is proposed to model the behavior of cyclodextrins (CDs) and of their inclusion compounds through energy minimizations and molecular dynamics (MD) simulations at a constant temperature. In this way, the formation of a host–guest compound is obtained starting from many trial geometries with the guest outside the CD cavity without any a priori assumption. The MD simulation results are analyzed through two functions: (*i*) the similarity maps of the root-mean-square distances between instantaneous conformations found in the MD runs to recognize different families of conformers; (*ii*) the pair distribution function PDF, yielding the probability density of finding appropriate atom pairs as a function of their distance at equilibrium. As an example, the inclusion compound formed by *β*-CD and (–)-menthol-*β*-D-glucoside is investigated. The lowest-energy inclusion compound is in good agreement with the results of single-crystal X-ray analysis, while at room temperature the MD runs show a closely similar arrangement with thermal fluctuations. In this case, the PDF between diagnostic hydrogen atoms of *β*-CD and of the guest molecule are fully consistent with the experimental NOE results obtained from NMR measurements in solution.

G. Raffaini (✉) · F. Ganazzoli
Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta", Sez. Chimica, Politecnico di Milano, via L. Mancinelli 7, 20131 Milano, Italy
e-mail: giuseppina.raffaini@polimi.it

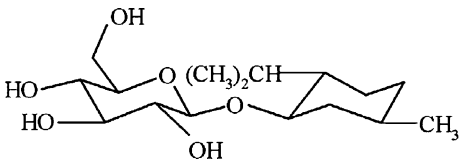
F. Ganazzoli
e-mail: fabio.ganazzoli@polimi.it

Keywords *β*-Cyclodextrin · Computer simulations · Glycoconjugate · Molecular dynamics · Pair distribution function · Similarity maps

Introduction

Atomistic computer simulations are an established tool to investigate host–guest compounds [1]. Much interest has been devoted to study the energetics of binding in inclusion compounds and the relevant contributions, be they dispersion, electrostatic or hydrophobic forces, and the release of high-energy water molecules from the host cavity [1, 2]. However, most papers focus on calculating the potential energy of the isolated molecules and of the bound system. Such calculations are quite fast and reliable, but they correspond to describe a frozen system at 0 K, whereas thermal motion is relevant for calculating average properties at a finite temperature. For such goal, only molecular dynamics (MD) simulations can provide an appropriate tool [1, 3, 4], even though the time length of the simulations [5] may sometimes become a computational problem.

In this paper, a methodology to model the formation of the inclusion compounds of cyclodextrins is described, and a new data analysis is proposed. The procedure is applied to study the inclusion compound formed by *β*-cyclodextrin (*β*-CD) with glycoconjugate (–)-menthol-*β*-D-glucoside **(1)** shown in Scheme 1. The simulations are carried out in vacuo, with implicit solvent and with explicit water assuming a 1:1 stoichiometry as found experimentally [6]. The geometries of the initial adduct obtained by placing **(1)** outside the CD cavity in different trial arrangements (as detailed in the next section) are first optimized, and then sub-



Scheme 1

jected to MD runs at a constant average temperature, until equilibrium is achieved. Final energy minimizations of many instantaneous conformations periodically sampled in the MD runs at equilibrium eventually yield the most stable state. In this way, an inclusion compound is obtained, with the main axis of the guest molecule roughly parallel to the average C_7 symmetry axis of the macrocycle. In the most stable geometry, the menthyl group is buried inside the hydrophobic cavity, with the polar glucoside moiety protruding above the secondary rim of β-CD, in agreement with the solid-state structure. Furthermore, it is shown that at equilibrium the thermal motion of the host–guest compound produces a distribution of intermolecular distances fully consistent with NMR measurements in solution, so that the proposed method can provide useful information for this technique.

The paper also shows that MD simulations are important to describe the conformational rearrangements leading to the host–guest compound and the average properties at equilibrium in solution. The rearrangements of the adduct leading to the inclusion compound can be monitored by the similarity maps, constructed through the root-mean-square distances of the same atoms in two instantaneous geometries (RMSD maps) saved during the MD runs, allowing easy recognition of distinct families of conformers. The equilibrium mobility and the distribution of the interatomic distances, for instance between diagnostic atoms for NMR experiments, are best analysed through the pair distribution function (PDF), yielding the probability density of finding the selected atoms as a function of their separation. Conversely, the most stable state found by geometry optimizations can be compared with the crystalline structure, provided packing effects do not greatly modify the host–guest geometry.

Methodological section

The simulations and the data analysis were performed with InsightII/Discover 2000 [7], using the consistent valence force field CVFF [8] with a Morse potential for bonded atoms. We note that CVFF describes the non-bonded interactions through van der Waals and Coulombic terms only, with no extra term for the hydrogen bonds (H-bonds). The structures of β-CD and of compound **(1)** were generated with the available templates of InsightII, then subjected to an MD run in vacuo and finally optimized up to an energy gradient lower than 4×10^{-3} kJ mol^{-1} Å^{-1} (see also ref. [9]). The resulting geometries turned out to be in good agreement with the X-ray crystal structures. Compound **(1)** was then placed close to β-CD choosing 12 unbiased trial geometries (Fig. 1) as starting points for the initial optimizations, so that the main different sides of **(1)** could approach the two rims and the outer surface of β-CD in different orientations. An inclusion compound was not assumed in this stage, unlike what is often done (see for instance refs. [3, 10]). All these arrangements were thoroughly optimized with an implicit water solvent in an effective dielectric medium, and the resulting adducts were subjected to MD simulations. The dynamic equations were integrated using the Verlet algorithm with a time

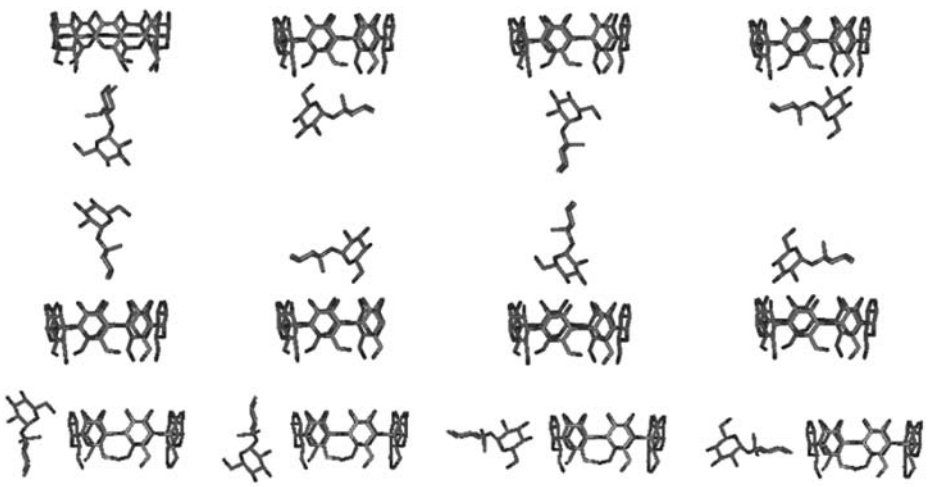


Fig. 1 The initial trial arrangements of compound **(1)** placed close to β-CD. Both molecules were separately optimized before docking

step of 1 fs at a temperature of 300 K, controlled through the Berendsen thermostat. The MD runs lasted for 2 ns, and showed significant rearrangements usually lasting for less than 300 ps, without any further change at longer times. The equilibration was monitored through the time changes of the energy contributions and of appropriate geometrical parameters, such as the distance between the centers of mass of the menthyl or glucoside moiety of the guest and the center of mass of the host. Eventually, many instantaneous conformations generated during the MD trajectory were optimized to find of the most stable one. The energy minimizations and the MD runs in explicit water were carried out as before after adding more than one thousand solvent molecules at the density of 1 g cm^{-3} in a cubic cell with axes of 33 Å using periodic boundary conditions, and lasted for 500 ps until equilibrium was achieved. Because of the great computational cost due to the large number of atoms, these simulations were only carried out for a few relevant geometries mentioned in the next section.

For the data analysis of the MD runs, the similarity maps, or RMSD maps, are constructed by calculating for a given set of n instantaneous conformations, or frames, the $n \times n$ root-mean-square distances among selected atoms, and plotting them as a function of the frame indices as a bi-dimensional map with an appropriate color coding. The RMSD between two frames is defined as the minimum value of the function [7]

$$\left\{ \frac{1}{N} \sum_{i=1}^{N} \left(\boldsymbol{r}_i^a - \boldsymbol{r}_i^b \right)^2 \right\}^{1/2} \tag{1}$$

where the index i runs over the N selected atoms, the a, b superscripts indicate two different frames, and r is the vector position of the given atom, while the minimum value is chosen to remove trivial effects related to rigid-body translations or rotations. Therefore, similar conformations belonging to the same family of conformers have a small RMSD, and unlike conformations a large RMSD. On the other hand, the PDF, often indicated with the notation $g_{ij}(r)$ (this function was less correctly denoted as Radial Distribution Function, or RDF, in ref. [9]), gives the probability density of finding atoms j at a distance r from atoms i, and for a finite system it is defined as [7]

$$g_{ij}(r) = d\langle N_{ij}(r) \rangle / dV(r) \tag{2}$$

where $d\langle N_{ij}(r) \rangle$ is the average number of times the j atoms are comprised in a spherical shell of thickness dr at a distance r from atoms i within an MD run. Thus, $g_{ij}(r)$ yields the average local density of atoms j in the shell volume $dV(r)$ comprised at a distance between r and $r + dr$ from atoms i, and it vanishes at large r.

Results and discussion

A preliminary study of the isolated molecules showed that β-CD has a truncated conical shape, with very little distortions from an average C_7 symmetry in vacuo. Minor distortions with a slight tilting of the glucoside rings are present in explicit water, but this feature does not greatly affect the symmetry of the cavity. On the other hand, compound **(1)** has an elongated shape, with a roughly orthogonal arrangement of the average planes through the sugar and the menthyl rings.

The formation of the inclusion compound in the implicit solvent and in vacuo

Initial geometry optimizations of the adducts in the dielectric medium were carried out for 12 trial arrangements shown in Fig. 1 with compound **(1)** outside the β-CD cavity. These optimizations led to a few distinct geometries only. In particular, the guest molecule placed above the β-CD rims formed in most cases an inclusion compound with relatively strong interactions, while close to the outer surface it only yielded loosely bound adducts. In the lowest-energy initial geometry, the apolar menthyl group enters the hydrophobic cavity, while the glucoside moiety stays above the secondary rim forming only a few intermolecular H-bonds. When insertion takes place from the opposite side, the energy is larger by about 10 kJ mol^{-1} partly due to some strain of the guest molecule, while insertion of the glucoside moiety leads to an even less stable state by further 2 kJ mol^{-1}. Larger relative energies were found when the guest molecule remained outside the cavity or close to the outer surface.

All the optimized geometries were then subjected to MD runs at room temperature for 2 ns in the effective dielectric medium and in vacuo. In this way, the guest molecule could explore a large part of the configurational space, and eventually find a favorable arrangement by forming an inclusion compound. However, when compound **(1)** was initially placed near the outer surface of β-CD, it could ''diffuse'' towards either rim, eventually forming again an inclusion compound, or it could fly away, thus maximizing the configurational entropy. The MD runs yielding an inclusion compound further stabilized the system with a concomitant

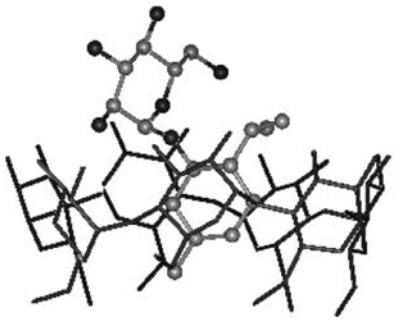


Fig. 2 The most stable inclusion compound formed by compound **(1)** with β-CD. This geometry corresponds to the absolute energy minimum found after the MD runs. For clarity, β-CD is shown by a line drawing, and the glycoconjugate in a ball-and-stick model with the O atoms indicated as darker spheres. The hydrogen atoms have been omitted for clarity

decrease of the number of unlike arrangements. The final energy minimizations in the dielectric medium eventually led to the most stable geometry shown in Fig. 2 with the menthyl group in the hydrophobic cavity, and the glucoside moiety above the secondary rim, stabilized by intermolecular H-bonds. This geometry optimized in vacuo was found to nicely agree with the crystal structure determined by X-ray analysis ([6], and L. Malpezzi, unpublished results). Therefore, energy minimizations do yield the solid-state structure, provided crystalline packing does not greatly affect the geometry of the inclusion compound.

The upside-down arrangement with the glucoside group above the primary rim is less stable by about 7 kJ mol^{-1}, due to the lack of intermolecular H-bonds and to some strain of the macrocycle. A similar destabilization (about 6–7 kJ mol^{-1}) is also found when the glucoside moiety enters the cavity, mainly due to some strain of the macrocycle and/or of the guest molecule. However, in water the arrangement with the menthyl group inside the cavity is further stabilized by the H-bonds of the glucoside group with the solvent. In fact, it was found that upon solvation with explicit water the arrangement with the included glucoside moiety is not stable, since in an MD run it slowly exits the cavity to form H-bonds with water.

The simulations in explicit water and the data analysis in solution

The main inclusion compounds obtained in the dielectric medium with the menthyl group inside the β-CD cavity were also found by MD simulations in explicit water starting from the same initial arrangements with the guest outside the cavity. In particular, the geometry shown in Fig. 2 was quickly reached starting from the first arrangement of the second row in Fig. 1, with a concomitant expulsion of a few included water molecules. The MD runs provide much important information about the dynamics of the process, including the equilibrium fluctuations and rearrangements, and the RMSD maps provide an important tool to follow these rearrangements and to detect different families of conformers. A typical RMSD map obtained from the MD simulation in water leading to the geometry shown in Fig. 2 is reported in Fig. 3. The RMSD maps of the individual host and guest molecule (center and right) are essentially featureless, showing no major intramolecular rearrangement. Conversely, in the map of the whole adduct (at left) the large RMS distances between the initial 25 frames, roughly, and all the other ones (indicated by the thin black stripes close to the two axes) correspond to the different geometry before and after guest inclusion. Afterwards, the squares along the diagonal with small RMS distances (< 1 Å, shown with blue and magenta pixels) indicate similar conformations belonging to the same family of conformers, whereas the off-diagonal squares with larger RMS distances (shown with green, cyan and black pixels) correspond to largely different conformations. In this case, the two families correspond to different orientations of the guest within the cavity obtained by a rotation of about 90° around an axis roughly parallel to the approximate C_7 symmetry axis of β-CD. The sharpness of the transition between the families is enhanced by the chosen binning (a frame was saved every 0.5 ps), but the conformational change is anyway quite fast.

The equilibrium geometries sampled in an MD run at room temperature after inclusion, and in particular the distances between diagnostic atoms, are analysed through the PDF plots. As an example, Fig. 4 reports the plots of the PDF between the hydrogens {H3} of β-CD adjacent to the secondary hydroxyls, and the three hydrogens of one methyl group of the isopropyl substituent of compound **(1)** (upper panel), or of the methyne hydrogen of the same substituent (lower panel). The distances between these hydrogens are diagnostic in NMR experiments for detecting the formation of an inclusion compound and its geometry. Fig. 4 shows that during the MD runs these distances span a wide range of values both with implicit and with explicit solvent due to thermal fluctuations and local rearrangements. In particular, the methyl hydrogens of the isopropyl group lie most likely at a distance of about 2.4 Å from the {H3} hydrogens of β-CD, as shown by the largest peak in the upper panel of Fig. 4, but they can also be found at a much larger separation. Still, they are usually found at a distance smaller than

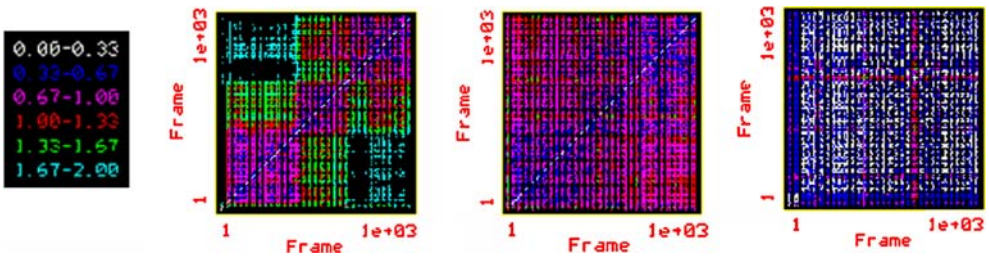


Fig. 3 The RMSD maps of the host–guest adduct, of β-CD and of compound **(1)**, from left to right, obtained from the MD runs in explicit water with concomitant expulsion of water molecules from the cavity. In each plot, the horizontal and vertical axes indicate the frames encountered in the MD run (one frame was saved every 0.5 ps), and the root-mean-square distance between the corresponding conformations (in Å) is shown in false colors (the scale is reported at left; note that the diagonal terms are equal to zero by definition). Root-mean-square distances larger than 2 Å are shown in black

4 Å, while larger values are seldom found. On the other hand, the methyne hydrogen of the same isopropyl group is always at distances larger than 4 Å from the {H3} hydrogens of β-CD (lower panel of Fig. 4), farther than the neighboring methyl hydrogens. It should be noted that in water the peaks (filled symbols) are shallower and broader than in the dielectric medium (empty symbols) because the solvent allows the system to probe for longer times the whole range of intermediate distances. However, in each panel the peak positions do not change, and the volume integrals of the curves are the same, since they yield the total number of the hydrogens involved in the calculation. Therefore, broader peaks must be significantly shallower. The results shown in Fig. 4 and those calculated for other sets of hydrogens are in excellent agreement with the NOE data measured in NMR experiments ([6] and A. Mele, unpublished results), considering that a NOE signal is observed for hydrogen atoms at a distance lower than about 4 Å. It should also be stressed that these experiments probe distances averaged over the measurement time, and therefore the PDF is best suited for a comparison with theoretical calculations.

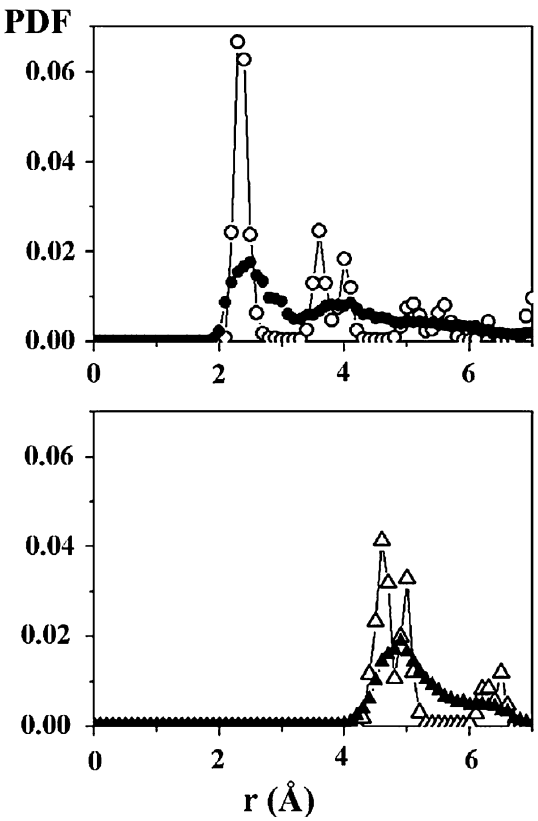


Fig. 4 The pair distribution function PDF between selected hydrogens of β-CD and of compound **(1)** obtained from the MD runs in implicit solvent (empty symbols) and in water (filled symbols). The figure shows the PDF of hydrogens {H3} of β-CD and the three hydrogens of one methyl group of the isopropyl substituent in **(1)** (upper panel), or the hydrogen of the methyne group of the same substituent (lower panel)

Conclusions

MD simulations provide many important results about the conformation and dynamics of cyclodextrins and on the stability, conformational properties and equilibrium fluctuations of their inclusion compounds [1,3,4]. The proposed methodology involves optimization of many trial adducts with the guest placed outside the CD cavity, and MD runs of all the optimized geometries at a constant temperature. In this way, any a priori assumption about the host–guest geometry is avoided. The paper shows that the RMSD similarity maps are most useful to distinguish among different families of conformers encountered in the MD trajectory, including the formation of the inclusion compound and its possible rearrangements. The most stable inclusion compound obtained by full energy minimizations of the isolated molecule in vacuo can be compared with the solid-state structure, and large

differences between the calculated and the observed geometry are symptomatic of significant intermolecular packing effects in the crystalline state. Moreover, the MD simulations are useful to establish the geometry of the host–guest compound in solution, in particular the formation of the inclusion compound from an outer adduct with concomitant expulsion of a small cluster of included water molecules. From the MD runs, the average properties of the system at equilibrium can be calculated. In particular, this paper shows that the pair distribution function PDF, yielding the probability density of finding diagnostic groups of atoms as a function of their separation, is most useful for the interpretation of the NOE data measured in NMR experiments.

Acknowledgments We gratefully acknowledge helpful discussions with prof. Andrea Mele and dott. Luciana Malpezzi, who disclosed us their experimental results before publication (NMR measurements in solution and X-ray crystal structure, respectively).

References

1. Lipkowitz, K.B.: Applications of computational chemistry to the study of cyclodextrins. Chem. Rev. **98**, 1829–1873 (1998)
2. Liu, L., Guo, Q.-X.: The driving forces in the inclusion complexation of cyclodextrins. J. Incl. Phenom. Macrocycl. Chem. **42**, 1–14 (2002)
3. Franchi, P., Lucarini, M., Mezzina, E., Pedulli, G.F.: Combining magnetic resonance spectroscopies, mass spectrometry, and molecular dynamics: investigation of chiral recognition by 2, 6-di-*O*-methyl-*β*-cyclodextrin. J. Am. Chem. Soc. **126**, 4343–4354 (2004)
4. Varady, J., Wu, X., Wang, S.: Competitive and reversible binding of a guest molecule to its host in aqueous solution through molecular dynamics simulation: benzyl alcohol/*β*-cyclodextrin system. J. Phys. Chem . **B 106**, 4863–4872 (2002)
5. Dodziuk, H., Lukin, O.: The dependence of the average energy difference for the diastereomeric complexes of α-pinene enantiomers with α-cyclodextrin on the length of dynamic simulations. Chem. Phys. Lett. **327**, 18–22 (2000)
6. Malpezzi, L., Fuganti, C., Fronza, G., Mele, A.: The structure of (–)menthol and (–)menthol-*β*-D-glucoside *β*-cyclodextrin complex by ^{1}H-NMR spectroscopy and X-ray single crystal investigation, 3rd Asian Cyclodextrin Conference ACC2005, Nankai University, Tianjin, China, 8–12 maggio 2005, pp 65–66
7. Accelrys Inc. InsightII 2000; San Diego, CA. See also the URL http://www.accelrys.com/
8. Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M., Hagler, A.T.: Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins Struct. Funct. Genet. **4**, 31–47 (1988)
9. Mele, A., Raffaini, G., Ganazzoli, F., Selva, A.: *β*-cyclodextrin and 5-methoxytryptammonium ion host-guest association *in vacuo*: simulation of non-covalent inclusion by molecular dynamics. J. Incl. Phenom. Macrocyclic Chem. **44**, 219–223 (2002)
10. Dodziuk, H., Lukin, O., Nowiński, K.S.: Molecular mechanics calculations of molecular and chiral recognition by cyclodextrins. Is it reliable? The selective complexation of decalins by *β*-cyclodextrin. J. Mol. Struct. (Theochem). **503**, 221–230 (2000)