NJC

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/njc

CREATED USING THE RSC ARTICLE TEMPLATE (VER. 3.0) - SEE WWW.RSC.ORG/ELECTRONICFILES FOR DETAILS

ARTICLE TYPE www.rsc.org/xxxxxx | XXXXXXXX

Conformational Features of 4-(*N*)-Squalenoyl-Gemcitabine in solution: a combined NMR and Molecular Dynamics Investigation

Ceruti Maurizio*[a], Rocco Flavio[a], Di Pietro Maria Enrica[b], Tocci Elena[c], De Luca Giuseppina*[b]

A combination of NMR spectroscopy and molecular modelling techniques has been used to investigate the conformational behaviour in solution of 4-(*N*)-Squalenoyl-Gemcitabine (Sq-Gem), a promising anticancer new derivative whose pharmacological and biological properties and its use in the medical applications are heavily influenced by its three-dimensional structure. NMR scalar coupling constants (J-couplings) have been experimentally measured, by mono- and bi-dimensional methods, and, subsequently, compared with the theoretically values estimated, *via* DFT calculations, on the most representative conformations obtained and optimized by molecular dynamics (MD) simulations in methanol solution. Following this approach, we found a good agreement between calculated and experimental NMR data for the major conformations in which the flexible squalenoyl chain is strongly folded.

1. Introduction

Gemcitabine is an antitumoral drug that has a slow half-life because it is rapidly metabolized intracellularly and extracellularly by cytidine deaminase into the chemotherapeutically inactive uracil derivative [1,2] and most of an administered dose is recovered as the inactive derivative in the urine. For this reason, recently, the synthesis and encapsulation in liposomes of a series of gemcitabine 4-(*N*)-acyl derivative prodrugs were reported[3] and their interaction with phospholipid bilayers and monolayers was studied[4,5]. A more recent approach to improve the biopharmaceutical properties is the conjugation with the acyclic isoprenoid chain of squalene, a precursor in the biosynthesis of sterols[6,7]. It is very interesting to note that squalene inhibits proliferation of cancer cells by decreasing farnesyl pyrophosphate levels and does not influence normal biochemical pathways[8,9]. It has been shown that the squalenoylation of some bioactive compounds allows stable nanostructures to self-assemble in an aqueous environment[6,7]; these nanostructures are characterized by a hexagonal conformation in which the drug compound is surrounded by a shell[10]. In addition, it has been observed that linking to a squalenoyl moiety improves the biopharmaceutical properties of various compounds, thus increasing their antitumoral or antiviral effects, prolonging the half-lives, and enhancing their interaction with biomembranes[6,7,11,12].

For example the interaction of the prodrug Sq-Gem with biomembranes has been studied through differential scanning calorimetry in processes of capture and release within lipid vesicles[13]. As synthetic biomembrane models, liposomal MLV and LUV were used, composed of dimyristoylphosphatidylcholine (DMPC) and of distearoylphosphatidylcholine (DSPC) that give a clear phase transition from an ordered gel-type structure to a disordered fluid-type structure. These studies concerned also the transfer of these compounds through an aqueous medium and the migration from liposomes containing a known quantity of gemcitabine or of Sq-Gem to empty liposomes. The comparison of the results obtained showed the importance of the prodrug Sq-Gem in providing better interaction with the biological membranes of Sq-Gem *versus* free gemcitabine and more controlled and gradual release of Sq-Gem over time *versus* free gemcitabine. To further confirm that squalenoylation is a general new technology that improves activity of selected drugs, recently, 1,1',2-tris-nor-squalenoyl-citarabine (Sq-Ara-C) was synthesised and formulated in nanoparticle form[14,15]. It is seen that this new nanomedicine improves the in *vitro* antitumoral activity and it is more active respect to the free drug against drug-resistant leukemic cells.

All these studies seem to indicate that improvements in the pharmaceutical properties of this new class of drug-bioconjugates, basically, depend on an increase of their flexibility and lipophilicity due to the presence of the squalenoyl moiety that allows them to spontaneously self-assemble in an aqueous medium as stable nanoparticles, having diameters in the 100-200 nm range. It should be noted that this supramolecular organization is the result of complex cooperative effect of multiple weak interactions (hydrophobic forces and polar head interactions) between a large number of components. On the other part, it is well know that the spatial arrangement a pharmaceutically active compound adopts can deeply affects its biochemical activity, guiding the interactions it creates with endogenous ligands as well as specific systems providing the controlled delivery of the active agent [16,17].

In this optic, the detailed knowledge of three-dimensional structure and conformational equilibrium of flexible drug molecules in solution becomes of utmost importance to clarify their role in chemical and biological activities. As a consequence, the last decades have witnessed intense research devoted to

structure and dynamics studies of bioactive molecules involving the use of multiple techniques, both experimental and theoretical. Among the experimental techniques, the high-resolution NMR spectroscopy is a particularly adapted and widely exploited method allowing to obtain detailed information on both molecular structure and conformational behaviour in solution. Several parameters derived from NMR experiments, such as chemical shifts and coupling constants, have been traditionally used for this purpose since they are extremely sensitive indicators of local molecular conformation and are a source of structural evidence. For example, information on dihedral angles has been obtained for a variety of biological systems by the vicinal J scalar coupling using Karplus-type relationships[18]. Nevertheless, the most important NMR parameter to estimate molecular structure is the nuclear Overhauser effect (NOE)[19]. This method relies on a dense network of distance restraints derived from NOEs data that connect pairs of atoms close in space (internuclear distance smaller than ~ 5Å). Furthermore, a combined approach of dipolar correlations (NOE) and J coupling constants as geometric constraints has been used for many complex biological systems[20]. However, these methods have limitations for obtaining connectivity information between atoms which are far apart and are often insufficient for an unequivocal structural and conformational determination. Other advanced NMR applications including the use of residual dipolar couplings (RDCs) have recently emerged as a new tool with which to study molecular structure and conformational distribution [21]. However, structural and conformational interpretation of RDCs is complicated by overall molecular motion and although specialized software can be used it is difficult to achieve accurate results for complex bioactive molecules.

In addition to a wide array of experimental possibilities, structural studies benefit also from theoretical techniques. Molecular simulations provide truly outstanding opportunities to exploit molecular recognition principles that are very difficult to observe with experimental techniques. There are several examples in the literature where experimental data are supplemented with quantum–mechanical (QM) and molecular dynamics (MD) simulations to obtain valuable information regarding biologically active molecules and flexible drugs[22]. Biomolecular simulations in structure-based drug design and molecular modelling approaches for the study of the interaction between bioactive molecules and membrane have progressed greatly and will continue expanding its reach to larger time and length scales[23]. The high level of complexity required by a dynamic description of the processes underlying molecular recognition requires a multidisciplinary investigation approach. In this perspective, the combination of nuclear magnetic resonance spectroscopy with molecular dynamics simulations has given new insights to the identification and design of new and effective drugs [24].

Both NMR measurements and computational methods have been already used to investigate the conformational behaviour of squalene and several squalene derivatives (vinyloxidosqualenes, azasqualenes, etc.) in solution. These previous studies performed on very flexible molecules without rigid fragments, indicate the presence of dynamically folded structures in solutions in which the mobile tails move around the stiffest central portion[25]. The proposed model also explains the selective reactivity in a polar medium of the mobile chain endings of squalene with respect to the central part, which is protected by these moving ends[26]. The new class of flexible drugs, 4-(*N*)-squalenoyl-gemcitabine (Sq-Gem) and 1,1',2-tris-nor-squalenoyl-citarabine (Sq-Ara-C) and analogues are less flexible than previous squalene derivatives, in fact they consist of a rigid fragment (the gemcitabine or the citarabine) linked to the flexible squalenoyl chain, and thus they may adopt different conformations compared to the other derivatives already studied.

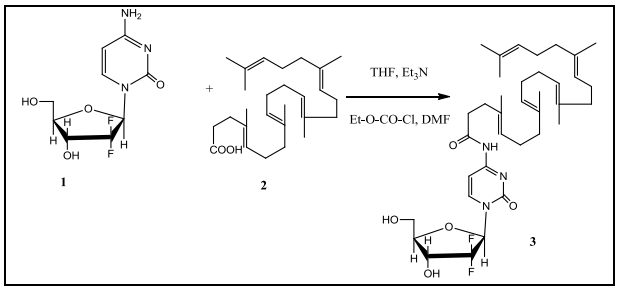
In order to get insight into the structure and the bioactive behaviour of this new class of drugs, we undertook in this work the conformational analysis of the Sq-Gem in solution using an elegant combination of experimental NMR data with molecular modelling conformational searches based on Molecular Dynamics simulations and on density functional theory (DFT). In particular we have chosen to test the conformational distribution of Sq-Gem using the J-couplings constants, J_{H-H}, J_{H-F}, J_{F-F}, $J_{C-H,}$ and J_{C-F}, which are sensitive indicators of local molecular conformation, by comparing the experimental values in solution with those calculated using DFT B3LYP/3-21 G calculations performed on the most representative conformers selected and optimized by MD simulations and on a weighted average based of their relative populations. It is necessary to emphasize that, as the molecule of Sq-Gem is practically insoluble in water, the conformational analysis was performed in methanol using the experimental data extracted from a diluted sample. This means that we are quite far from the conditions of self assembly that occur in aqueous media and that our study is mainly focused in shedding light on the chemical behavior of this substance in solution rather than in understanding the properties of self-assembly as nanoparticles in aqueous media.

2. Experimental

2.1. Synthesis of 4-[(*N*)-1,1',2-tris-nor-squalenoyl] gemcitabine

All the intermediate compounds synthesized were characterized by ^{1}H NMR spectra, registered on a Bruker 300 MHz spectrometer, for samples in $CDCl_3$ solution at room temperature, with Me_4Si (TMS) as internal standard and Mass spectra were recorded on a Finnigan MAT TSQ 700 spectrometer (San Jose, CA). The reactions were monitored by TLC on F_{254} silica gel precoated sheets (Merck, Damstadt, Germany); after development, the sheets were exposed to iodine vapour. Flash-column chromatography was performed on 230-400 mesh silica gel. Tetrahydrofuran was dried over sodium benzophenone ketyl. All solvents were distilled prior to flash chromatography.

Gemcitabine hydrochloride was purchased from Sequoia Research Products Ltd (United Kingdom). Squalene was purchased from VWR (Italy). 1,1',2-*tris-nor*-squalene acid (2 in Scheme 1) was obtained from 1,1',2-*tris-nor*-squalene aldehyde using the method previously described[14].



Scheme 1. 4-(*N*)-Squalenoyl-gemcitabine or 4-[(*N*)-1,1',2-tris-nor-squalenoyl]-2',2'-difluoro-2'-deoxycytidine (3) synthesis.

1,1',2-*tris*-*nor*-squalene acid **2** (338 mg, 0.845 mmol) was dissolved in anhydrous THF (5 ml) in a three necked flask under nitrogen, with stirring, followed by triethylamine (85.5 mg, 0.845 mmol) in anhydrous THF (2 ml) and it was cooled at 0 °C. Ethyl chloroformate (Et-O-CO-Cl) (91.7 mg, 0.845 mmol) dissolved in anhydrous THF (82 ml) was then added and left at 0 °C for 20 min, with stirring, followed by the addition of gemcitabine hydrochloride (205.52 mg, 0.845 mmol) dissolved in anhydrous DMF (5 ml). The reaction mixture was allowed to reach room temperature and allowed to react for 3 days, with stirring, under nitrogen. The reaction mixture was controlled by silica gel TLC with light dichloromethane / acetone, 95:5. The crude product was purified by flash chromatography with dichloromethane, then dichloromethane / acetone, 85:15 and finally dichloromethane / acetone, 75:25 as eluant, to give 300 mg of squalenoylgemcitabine **3** (scheme 1) (55% yield), as a colorless viscous oil. CIMS (isobutane) m/z 646 (100) EIMS. *m/z* 645 (10), 577 (8), 523 (7), 509 (18), 494 (10), 454 (15), 429 (24), 372 (100).

2.2. NMR spectra

A dilute solution (1,25 wt%) was prepared by dissolving Sq-Gem in 99,9 % pure deuterated methanol (CD_3OD). On this sample ^1H, ^{13}C and ^{19}F NMR spectra, high-resolution 1D and correlation experiments, were recorded on a Bruker Avance 500MHz (11.74 T) spectrometer. ^1H, ^{13}C and ^{19}F spectra were obtained at 500.13, 125.76 and 470.59 MHz respectively. The probe temperature was 300 K and the ^1H and ^{13}C signal of TMS were used as internal standard reference peaks for proton and carbon spectra, respectively.

2.3. MD calculations

Molecular mechanics (MM) and molecular dynamics (MD) simulations were performed by using Materials Studio 6.0 software using CVFF (Consistent Valence Force Force Field)[27] as implemented by the Material Studio codes[28]. The calculations were carried out using the *canonical ensemble* (NVT, constant temperature, number of molecules and volume) on the isolated system and in solution with CH_3OH. The first cycle of MD simulations *in vacuo* was carried out at 300K for 10 ns with 1 fs time step. The positions and the velocities of all atoms were saved on the last 4 ns of simulations in a history file. The simulation time was chosen to be long enough to overcome the approach to the equilibrium (when trajectories move from arbitrary assigned initial conditions) to span the equilibrium state regions. The snapshots were taken every 100 ps, each structure was subjected to a geometry optimization at MM level, with steepest descent, and conjugate gradient algorithm until the maximum derivative was of 0.01 kcal/mol.

The solvating model used is a 3-dimensional box with cubic periodic boundary conditions. Here the same ratio between the number of molecules of Sq-Gem and methanol as used in the NMR experiments has been reproduced. This corresponds to one single molecule of Sq-Gem solvated within 100 molecules of CH_3OH. The MD simulation has been run for 6 ns at T= 300 K with 1 fs time step. Snapshots were captured after 1ns of simulation. A total of 51 conformers were selected and optimized with the same procedure before described (steepest descent and conjugate gradient algorithm with a maximum derivative of 0.01 kcal/mol).

The van der Waals were calculated with the atom based method (cutoff 12,0 A). In all simulations the Andersen method that control the thermodynamic temperature (kept constant) and generate the correct statistical ensemble has been used[29].

2.4. DFT calculations

The coupling constants were calculated using the 'spin-spin' option implemented in GAUSSIAN 03[30] suite program. This option calculates J's including contributions from the spin-dipole (SD) and both diamagnetic (DSO) and paramagnetic (PSO) spin-orbit terms in addition to the Fermi contact (FC) term. Generally the FC contribution is the most important, but to obtain a good agreement with experiment all four contributions must be computed. The calculations were carried out on the selected conformers, obtained from the MD simulations and MM optimized, at the DFT theoretical level by using the B3LYP functional with the 3-21 G basis set. It should be emphasized that the DFT method represents a powerful tool for obtaining electronic structure and properties in medium-large size compounds and it is currently the most cost-effective method for the spin-spin calculations[31]. Among all available density functional, we chose to utilize the hybrid B3LYP since it has been shown that it provides good results for spin-spin couplings[32]. As basis set we used the 3-21 G which allowed us to get in a reasonable time values of J-couplings in good qualitative agreement with the experimental ones.

3. Results and discussion

3.1. NMR spectral analysis

In spite of the complexity of the 1D high-resolution ^1H, ^{13}C and ^{13}C-{^1H} spectra and signal overlappings, an almost complete assignment of the peaks was performed on the basis of ^1H-^1H COSY, ^1H-^{13}C HMBC, ^1H-^{13}C HSQC correlation experiments. In Table 1 the experimental values of ^1H and ^{13}C chemical shifts are listed. The structure and the numeration of Sq-Gem used in the text are shown in figure 1.

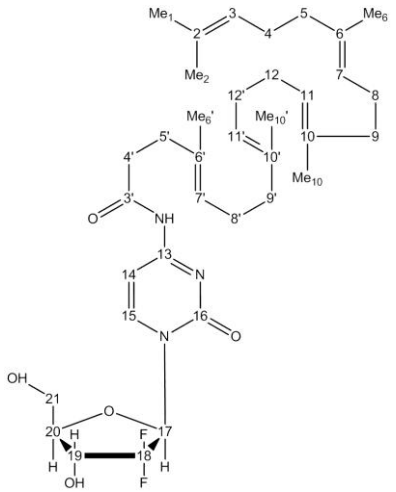


Figure 1. Structure and labelling scheme of carbons for the molecule of Sq-Gem.

Protons and fluorine nuclei are numbered after the carbon to which they are bound. The correct assignment and edition of J_{ij} was achieved by combining information from ^{1}H, ^{19}F and ^{13}C 1D NMR spectra and from different 2D experiments as ^{1}H-^{13}C *J*-resolved, ^{1}H-^{1}H J-resolved (figure 2 and figure 3) and ^{1}H-^{1}H SERF [33] (figure 4). The 2D experiments were performed in the case of nuclei which give signals not resolved in the 1D spectrum in order to extract not only the maximum number of data but mainly some J-couplings that are crucial to access the conformation of the molecule. A brief description of the application of these not trivial experiments is given below for the protons. A standard homonuclear J-resolved pulse sequence allows to separate the chemical shift and coupling constant information in separate dimensions (F_2 and F_1, respectively) of the 2D spectrum. This means that the signal from each group of isochronous protons is seen in a separate column of the 2D map, allowing its individual extraction and the quite precise measurement of the coupling constants. The benefit is evident if one considers for instance protons $H_{12}/H_{12'}$, H_5, $H_9/H_{9'}$ in the range 1.95 to 2.10 ppm: albeit their signals are superimposed in the 1D spectrum, the expansion of the ^{1}H-^{1}H J-resolved experiment shown in Fig. 3 proves their signals can be efficiently separated in different columns, as testified by the internal projection reported in the F_2 dimension, where each peak corresponds to a different chemical shift.

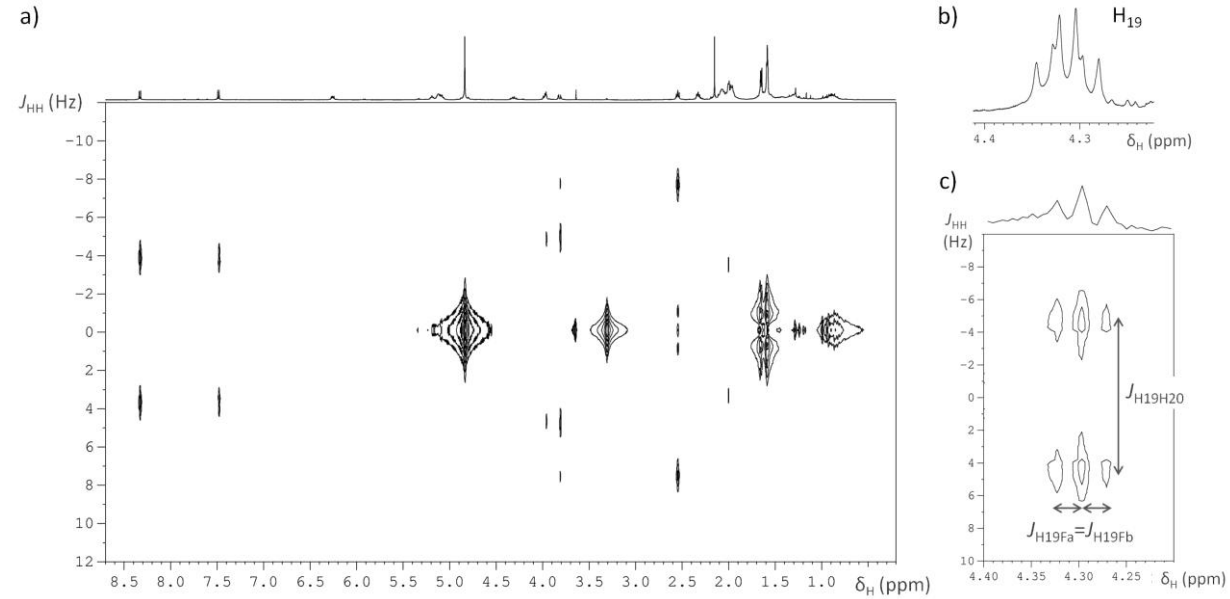


Figure 2 (a) Tilted 2D ^{1}H-^{1}H *J*-resolved spectrum recorded on Sq-Gem in CD_3OD at 300 K. The 1D ^{1}H spectrum is reported as projection in the F_2 dimension. ^{1}H-^{1}H couplings are easily read in the indirect (vertical) dimension, while ^{1}H-^{19}F couplings are refocused and appear in the direct (horizontal) dimension. Expansions of H_{19} signal in the (b) ^{1}H spectrum and (c) *J*-resolved spectrum are given as an example. The ^{1}H broadband spectrum was recorded by using 65536 points and 32 scans. The 2D *J*-resolved spectrum was recorded in 3h 24min using a data matrix of 2048 (t_2) x 256 (t_1) with 16 scans per t_1 increment. The relaxation delays were 1.2 s. Data were processed using zero-filling up to 512 points and a sine filter in F_1 dimension.

When the proper assignment of a coupling constant to a pair of nuclei is ambiguous or not possible because of poor resolution, SERF experiments reveal particularly useful. By replacing the 90° and 180° pulses of the J-resolved sequence by selective E-BURP and RE-BURP pulses, this experiment allows a desired coupling to evolve during the t_1 delay, whereas all the other couplings and the chemical shifts are refocused. This means that, contrarily to the *J*-resolved method, where one obtains for each protons the full multiplicity information in the F_1 dimension, in the SERF experiment one gets on the 2D map the signal corresponding to the proton(s) selected with the full coupling multiplicity in the F_2 dimension and only the coupling between the selected nuclei in the F_1 domain (figure 4). After tilt of the 2D spectrum the selected couplings are removed from the direct dimension allowing a simplification of the spectral pattern.

CREATED USING THE RSC ARTICLE TEMPLATE (VER. 3.0) - SEE WWW.RSC.ORG/ELECTRONICFILES FOR DETAILS

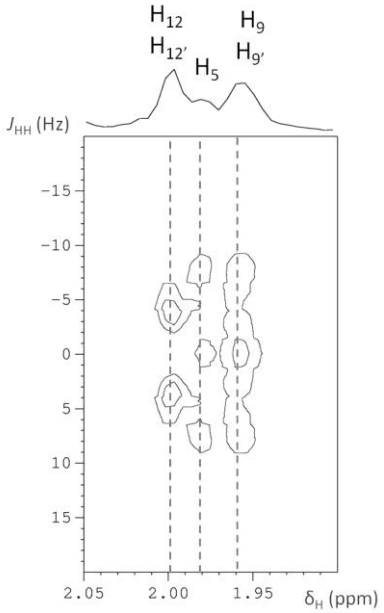


Figure 3 Expansion of the tilted 2D ^{1}H-1H *J*-resolved spectrum recorded on Sq-Gem in CD_3OD at 300 K including signals from $H_{12}/H_{12'}$, H_5, $H_9/H_{9'}$. The internal projection after tilt is reported in the F_2 dimension. ^{1}H-^{1}H couplings are easily read in the indirect (vertical) dimension.

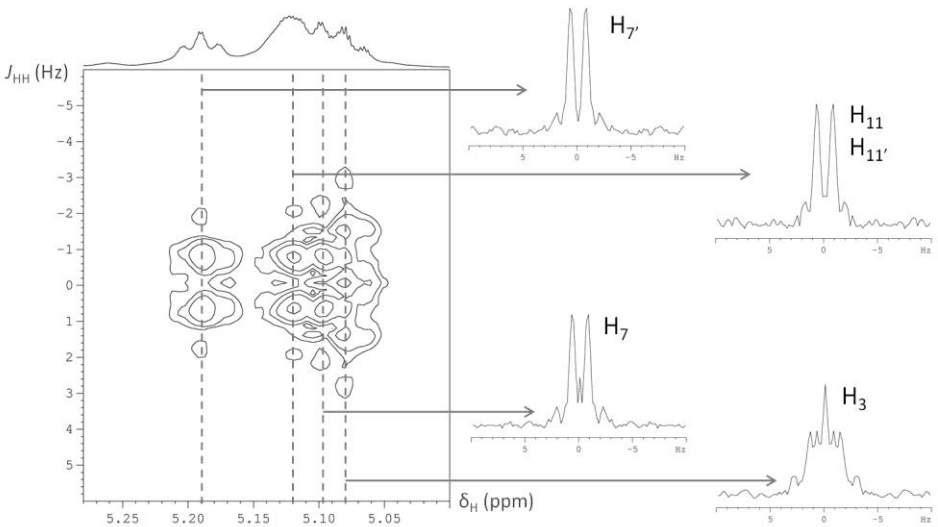


Figure 4 Tilted ^{1}H-^{1}H SERF 2D spectrum on the vinyl protons $H_{7'}$, $H_{11}/H_{11'}$, H_7 and H_3 coupled with the methyl protons H_{Me1}, H_{Me2}, H_{Me6}, $H_{Me6'}$, H_{Me10} and $H_{Me10'}$ of the squalenoyl chain of Sq-Gem in CD_3OD at 300 K. The corresponding portion of the 1D ^{1}H spectrum is reported as projection in the F_2 dimension. ^{1}H-^{1}H couplings between the selected protons are easily read in the F_1 dimension. Columns extractions are also shown. Semi-selective E-BURP excitation and RE-BURP refocusing pulses are applied on the proton channel at $\delta_{7',11,11',7,3}$ and at $\delta_{Me1,Me2,Me6,Me6',Me10,Me10'}$ with duration of 29.1 ms, corresponding to a frequency width of 170 Hz. Spectrum was recorded in 5 h using a data matrix of 1024 (t_2) x 176 (t_1) with 32 scans per t_1 increment. The relaxation delays were 1s. Data were processed using zero-filling up to 256 points in t_1 and a qsine filter in both dimensions.

In Table 2 all the experimental J_{ij} obtained from the spectral analysis of proton and carbon 1D and 2D spectra are collected, for a total of 76 values.

The vicinal coupling constants, $^3J(H_i - H_j)$, of the squalenoyl moiety are corrected of ± 0. 4, and are in the range 7-7.6 Hz in agreement with the data reported in the literature on similar compounds[25]. Although the Sq-Gem compared to the previously studied compounds shows a much more rigid part (the gemcitabine fragment) linked to the flexible squalenoyl chain, the conformational analysis, albeit inaccurate, using these experimental vicinal ^3J couplings in a Karplus relationships[18] seems to exclude a linear conformation and indicates, rather, a folded structure with an increase of mobility moving from the rigid part to the end of squalenoyl chain. As we have extracted from the analysis of the spectra a very high number of experimental J_{ij} (see Table 2), both vicinal and long-range, which are weighted averages on the conformations, we decided to compare their experimental values with those calculated *via* DFT on the conformers selected and optimized by MD simulations as explained in the following.

Table 1. Experimental NMR proton and carbon chemical shifts of Sq-Gem in CD_3OD.

^1H_i	δ_i / ppm	$^{13}C_j$	δ_j / ppm
N*H*CO	10.50	$C_{3'}$	175.33
H_{15}	8.33	C_{16}	164.67
H_{14}	7.48	C_{13}	157.54
H_{17}	6.25	C_{15}	145.85
$H_{7'}$	5.19	C_6 / C_{10} / $C_{10'}$	135.90 / 135.76 / 135.73
H_{11}, $H_{11'}$	5.13	$C_{6'}$	134.30
H_7	5.10	C_2	131.94
H_3	5.08	$C_{7'}$	126.43
H_{19}	4.30	C_{11} / $C_{11'}$	125.56 / 125.50
H_{21a}	3.97	C_3 / C_7	125.43 / 125.40
H_{20}	3.95	C_{18}	123.82
H_{21b}	3.81	C_{14}	98.26
$H_{4'}$	2.54	C_{17}	83.31
$H_{5'}$	2.33	C_{20}	82.79
H_4 / H_8 / $H_{8'}$	2.08 / 2.05	C_{19}	70.09
H_{12}, $H_{12'}$	2.00	C_{21}	60.18
H_5 / H_9 / $H_{9'}$	1.98 / 1.96	C_5 / C_9 / $C_{9'}$	40.83 / 40.81 / 40.63
H_{Me1}	1.66	$C_{4'}$	37.09
$H_{Me6'}$	1.65	$C_{5'}$	35.60
H_{Me2}	1.58	C_{12} / $C_{12'}$	29.20 / 29.18
H_{Me6} / H_{Me10} / $H_{Me10'}$	1.59 / 1.57	C_4 / C_8 / $C_{8'}$	27.79 / 27.64 / 27.56
		C_{Me1}	25.97
		C_{Me2}	17.85
		C_{Me6} / C_{Me10} / $C_{Me10'}$	16.28 / 16.26 / 16.21
		$C_{Me6'}$	16.14

Table 2. Experimental J_{H-H}, $J_{C-H,}$ J_{H-F}, J_{C-F} and J_{F-F} (in Hz) of Sq-Gem in CD_3OD. Errors are given from spectral resolution.

$\mathbf{H}_i$	$\mathbf{H}_j$	$\mathbf{J}_{H-H}$(exp)
H_3	H_4	7.5 ± 0.2
H_3	H_{Me1}	1.5 ± 0.4
H_3	H_{Me2}	1.5 ± 0.4
H_4	H_5	7.0 ± 0.3
H_4	H_{Me1}	1.2 ± 0.4
H_4	H_{Me2}	0.8 ± 0.4
$H_{4'}$	$H_{5'}$	7.6 ± 0.3
$H_{5'}$	$H_{Me6'}$	1.2 ± 0.4
H_7	H_8	7.5 ± 0.4
H_7	H_{Me6}	1.5 ± 0.4
$H_{7'}$	$H_{8'}$	7.0 ± 0.4
$H_{7'}$	$H_{Me6'}$	1.5 ± 0.4
H_8	H_9	7.0 ± 0.3
$H_{8'}$	$H_{9'}$	7.0 ± 0.3
H_{11}	H_{12}	7.5 ± 0.3
H_{11}	H_{Me10}	1.5 ± 0.4
$H_{11'}$	$H_{12'}$	7.5 ± 0.3
$H_{11'}$	$H_{Me10'}$	1.5 ± 0.4
H_{14}	N*H*CO	0.7 ± 0.4
H_{14}	H_{15}	7.6 ± 0.3
H_{17}	H_{19}	0.6 ± 0.3
H_{19}	H_{20}	8.5 ± 0.3
H_{19}	H_{21a}	0.9 ± 0.4
H_{19}	H_{21b}	0.9 ± 0.4
H_{20}	H_{21a}	2.8 ± 0.4
H_{20}	H_{21b}	2.8 ± 0.4
H_{21a}	H_{21b}	12.8 ± 0.4
H_{Me1}	H_{Me2}	0.5 ± 0.3

$\mathbf{H}_i$	$\mathbf{F}_j$	$\mathbf{J}_{H-F}$(exp)
H_{17}	F_a e F_b	7.4 ± 0.3
H_{19}	F_a e F_b	12.2 ± 0.3

$\mathbf{C}_i$	$\mathbf{F}_j$	$\mathbf{J}_{C-F}$(exp)
C_{17}	F_a / F_b	38.2 / 26.9 ± 0.6
C_{18}	F_a e F_b	258.7 ± 0.6
C_{19}	F_a e F_b	23.5 ± 0.6
C_{20}	F_a o F_b	5.0 ± 0.6

$\mathbf{F}_i$	$\mathbf{F}_j$	$\mathbf{J}_{F-F}$(exp)
F_a	F_b	240.0 ± 1.2

$\mathbf{C}_i$	$\mathbf{H}_j$	$\mathbf{J}_{C-H}$(exp)
C_2	H_{Me1}	5.9 ± 0.6
C_2	H_{Me2}	5.9 ± 0.6
C_3	H_3	150.0 ± 0.6
$C_{3'}$	$H_{4'}$	3.8 ± 0.6
$C_{3'}$	$H_{5'}$	6.2 ± 0.6
C_4	H_4	125.0 ± 0.6
$C_{4'}$	$H_{4'}$	128.0 ± 0.6
$C_{4'}$	$H_{5'}$	4.0 ± 0.6
C_5	H_5	128.0 ± 0.6
$C_{5'}$	$H_{5'}$	128.0 ± 0.6
C_7	H_7	150.0 ± 0.6
$C_{7'}$	$H_{7'}$	150.0 ± 0.6
C_8	H_8	125.0 ± 0.6
$C_{8'}$	$H_{8'}$	125.0 ± 0.6
C_9	H_9	128.0 ± 0.6
$C_{9'}$	$H_{9'}$	128.0 ± 0.6
C_{11}	H_{11}	150.0 ± 0.6
$C_{11'}$	$H_{11'}$	150.0 ± 0.6
C_{12}	H_{12}	126.0 ± 0.6
$C_{12'}$	$H_{12'}$	126.0 ± 0.6
C_{13}	H_{14}	1.3 ± 0.6
C_{13}	H_{15}	6.2 ± 0.6
C_{14}	H_{14}	181.7 ± 0.6
C_{14}	H_{15}	3.8 ± 0.6
C_{15}	H_{15}	185.9 ± 0.6
C_{16}	H_{15}	9.8 ± 0.6
C_{17}	H_{17}	173.3 ± 0.6
C_{19}	H_{19}	147.9 ± 0.6
C_{20}	H_{20}	148.4 ± 0.6
C_{21}	H_{19}	3.9 ± 0.6
C_{21}	H_{21}	142.9 ± 0.6
C_{Me1}	H_3	8.0 ± 0.6
C_{Me1}	H_{Me1}	125.0 ± 0.6
C_{Me1}	H_{Me2}	4.0 ± 0.6
C_{Me2}	H_3	8.0 ± 0.6
C_{Me2}	H_{Me1}	4.0 ± 0.6
C_{Me2}	H_{Me2}	125.5 ± 0.6
C_{Me6}	H_{Me6}	125.0 ± 0.6
$C_{Me6'}$	$H_{Me6'}$	125.0 ± 0.6
C_{Me10}	H_{Me10}	125.0 ± 0.6
$C_{Me10'}$	$H_{Me10'}$	125.0 ± 0.6

3.2. Structural theoretical calculations

The MD simulations, as well as the NMR data, show a strong evidence of folded structures both *in vacuo* and in solution, albeit the introduction of the solvent increases further the tendency toward molecular folding and tends to stabilize some of the folded (bent) conformations. Figure 5 shows the 3- dimensional box with the methanol molecules obtained from the MD simulation and the comparison between the structure of the most stable conformer in solution and *in vacuo*. In solution, the folding maximizes the intramolecular lipophilic interactions and minimizes the contact with the molecule of CH_3OH solvent. The solvent molecules create a sort of network *via* hydrogen bonds interactions that take place only in the hydrophilic site of the molecule. For our further investigation, we will consider only the 51 conformers extracted and optimized from the MD simulations in the solvent, neglecting the data *in vacuo*, because this situation is clearly closer to the NMR experimental conditions.

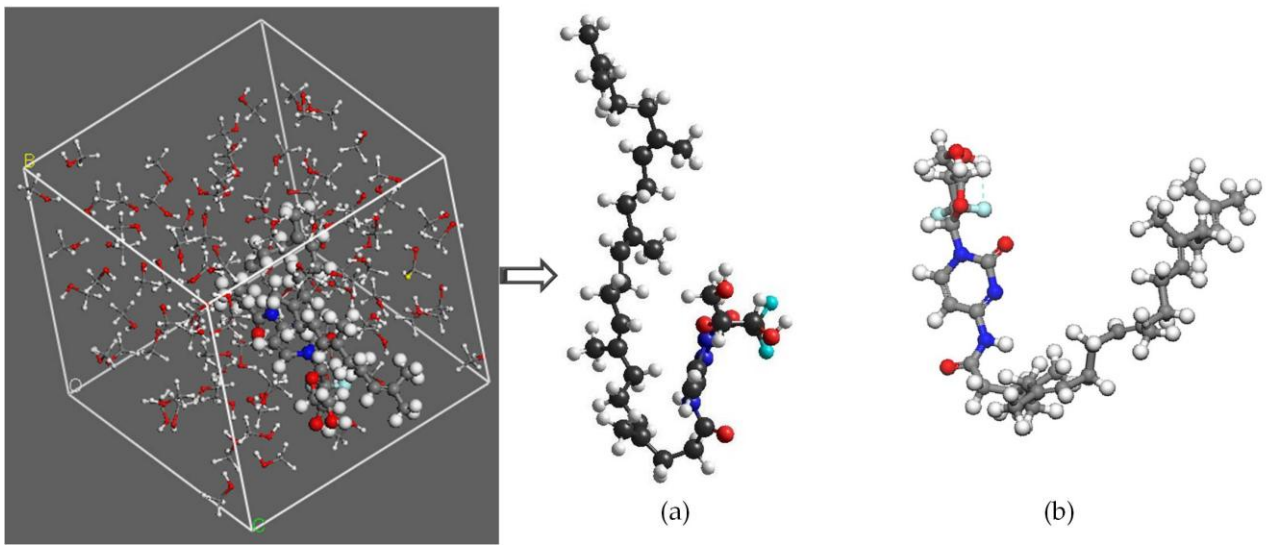


Figure 5. 3-dimensional box with methanol molecules obtained from the MD simulations, extracted structure of the most stable conformer in solution (a) and comparison with a minimum structure *in vacuo* (b).

On the 51 conformers obtained from the MD simulations, the relative populations have been computed according to the classical Boltzmann distribution:

$$P_i = e^{\left(-\frac{E_i}{RT}\right)} \Big/ \sum e^{\left(-\frac{E_i}{RT}\right)} \qquad 1)$$

Where R is the universal gas constant, T denotes the absolute temperature, (in our case T = 300 K) and E_i is the relative energy of each conformer. The results of this calculation are reported in table I of Support Information and in Figure 6 where conformers have been differentiated on the basis of similar energy content. The analysis of these relative populations shows the existence of a dominant conformer at 3.8 ns (relative energy fixed at E = 0.0 kJ/mol), which represents ~ 20% of the global population. Moreover, it can be observed a decrease in the probability of existence of the other conformers with the increase of their relative energy. This means that the conformers with high energy, above the threshold value of 8 kJ/mol, have a low existence probability, and therefore can be neglected. In this way, there are 27 conformers to be considered compared to the previous 51 (see figure 6 and figure I of Support Information). Among them, the conformers with close values of energy are similar in structure and, then, we decided to select only four major conformations to be sampled: the dominant conformer at 3.8 ns (**A**), the conformer at 3.6 ns (E ~ 4.5 kJ/mol) (**B**), the conformer at 1.0 ns (E ~ 6.5 kJ/mol) (**C**) and the conformer at 2.9 ns (E ~ 7.5 kJ/mol) (**D**).

Note that in the figure 6, where the relative existence probability of the 51 conformers is reported, each color corresponds to a group of conformers with similar energy and structure. Four different energetic/structural situations (4 colors) with energy higher than the threshold were identified and in order to sample each of them the conformers (A), (B), (C) and (D) selected were chosen one for each group.

Figure 7 reports the structures of the four conformers extracted from the MD simulations and used in the following assuming that only these four conformational states are thermally populated.

It is worth noting that the structure, shown in Figure 7, for all the conformers selected is a sign that seems to exclude a linear structure. In order to predict the scalar coupling constants the following strategy has been adopted.

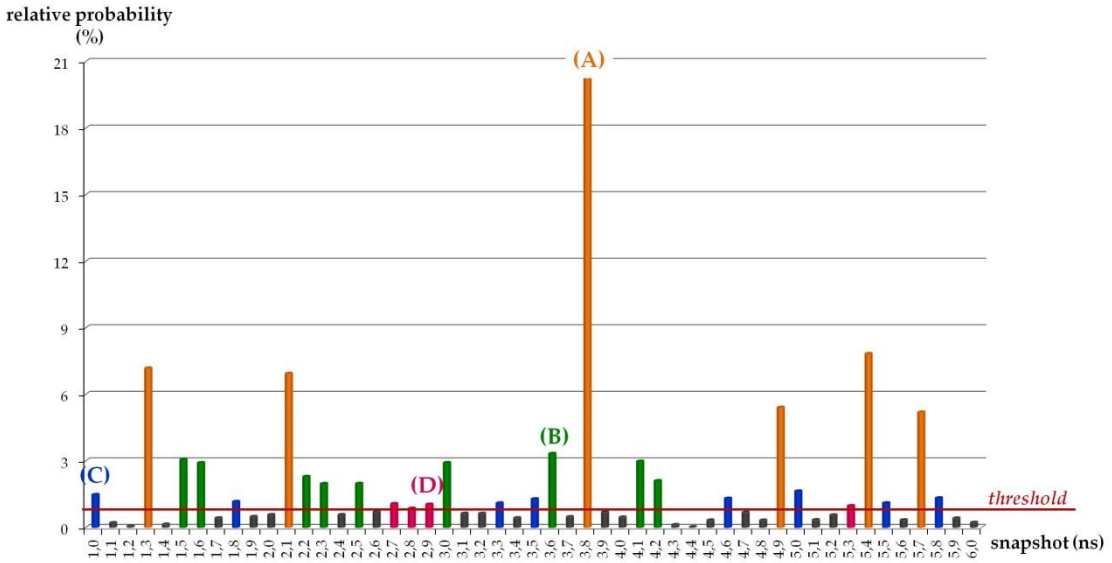


Figure 6. The relative existence probability of the 51 conformers obtained from the MD simulations, differentiated on the basis of similar energy content. Each color corresponds to a group of conformers with similar energy (and structure) and higher than the threshold. The four conformers (**A**), (**B**), (**C**), and (**D**) to be sampled were chosen among them.

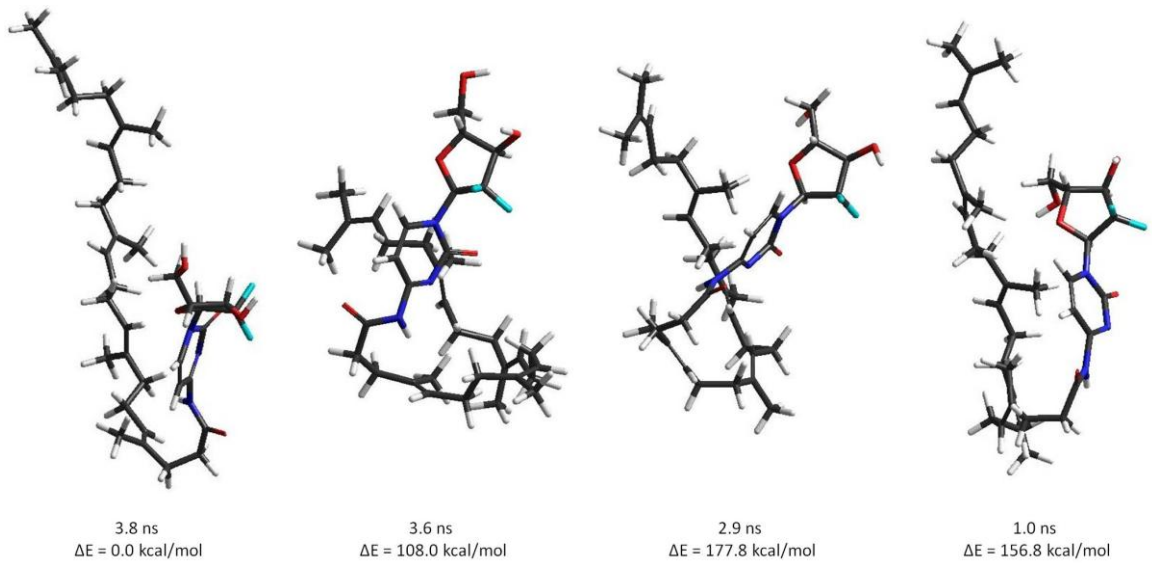


Figure 7. Structures of the conformers, (**A**), (**B**), (**C**) and (**D**), extracted from the MD simulations.

3.3. J-Couplings prediction from DFT/B3LYP calculations

As first step we extracted the geometries of the four conformers (in term of bond-lengths, angles and dihedral angles) from the MD simulations; then assuming these geometries and using the B3LYP functional at the 3-21 G level, as implemented in Gaussian 03[30], we predicted for each conformer the magnitudes and the signs of all the J couplings. Afterwards, according to their Boltzmann populations, we calculated the weighted average of these J couplings as reported in Tables 3 and 4, where for reasons of clarity the data referred to the gemcitabine fragment (rigid part) of Sq-Gem have been separated from the data referred to the squalenoyl chain. Since the analysis of the MD simulations suggests that the conformer (**A**) is the most populated, with the intent to test whether this single dominant conformer could give us a good reproducibility of the experimental J couplings, taking into account only the conformer (**A**) the scalar couplings have been calculated and reported in the same tables. Furthermore, in order to exclude the hypothesis of a linear structure, with the same method, we predicted the values of J couplings for the hypothetical linear conformer achieved by fixing the geometry of conformer (**A**) (Table 5) and changing only the dihedral angles of the squalenoyl chain. These last predicted values of J couplings are reported again in tables 3 and 4, where we labeled this linear hypothetical conformation simply as conformer (**E**). In the same tables, also, the experimental scalar couplings achieved from the spectral analysis have been listed as a comparison.

Table 3. The scalar coupling constants in Hertz calculated for the 4-(*N*)-squalenoyl-gemcitabine molecule in methanol referred to the squalenoyl chain. In the last column for comparison the corresponding experimental values are listed.

nuclei pairs	Weighted average on conformers (A), (B), (C) and (D) J^{av}_{calc}	Conformer (A) J_{calc}	Conformer (E) J_{calc}	J_{exp} (absolute value)
H_{Me1}-H_{Me2}	-0.9	-0.7	-0.7	0.5
H_4-H_{Me2}	1.1	0.8	1.7	0.8
$H_{5'}$-$H_{Me6'}$	-0.7	-0.6	-5.8	1.2
H_4-H_{Me1}	0.9	0.5	1.5	1.2
$H_{7'}$-$H_{Me6'}$	-2.4	-2.5	14.6	1.5
H_7-H_{Me6}	-2.4	-2.6	-2.7	1.5
H_{11}-H_{Me10}	-1.9	-1.9	0.0	1.5
H_{11}-$H_{Me10'}$	-1.9	-1.9	0.0	1.5
H_3-H_{Me1}	-1.4	-1.5	-1.4	1.5
H_3-H_{Me2}	-2.5	-2.5	-2.5	1.5
$C_{3'}$-$H_{4'}$	-4.5	-4.3	-14.9	3.8
$C_{4'}$-$H_{5'}$	-3.3	-3.1	-22.8	4.0
C_{Me1}-H_{Me2}	3.5	3.6	3.6	4.0
C_{Me2}-H_{Me1}	3.0	3.2	3.2	4.0
C_2-H_{Me1}	-5.7	-5.5	-5.6	5.9
C_2-H_{Me2}	-6.1	-6.0	-6.1	5.9
$C_{3'}$-$H_{5'}$	4.2	4.4	-0.3	6.2
$H_{7'}$-$H_{8'}$	4.6	4.2	21.8	7.0
H_8-H_9 = $H_{8'}$-$H_{9'}$	6.6	6.7	-1.3	7.0
H_4-H_5	6.0	6.2	-11.1	7.0
H_7-H_8	4.4	3.4	3.3	7.5
H_{11}-H_{12}	5.1	5.1	3.1	7.5
$H_{11'}$-$H_{12'}$	5.1	5.1	3.1	7.5
H_3-H_4	4.9	6.1	4.3	7.5
$H_{4'}$-$H_{5'}$	5.2	5.3	-4.3	7.6
C_{Me1}-H_3	5.7	5.5	5.7	8.0
C_{Me2}-H_3	5.9	5.8	6.2	8.0
C_4-H_4	95.3	104.8	100.2	125.0
C_8-H_8	104.8	104.9	102.1	125.0
$C_{8'}$-$H_{8'}$	104.5	104.1	101.1	125.0
C_{Me1}-H_{Me1}	101.1	101.3	100.7	125.0
C_{Me6}-H_{Me6}	101.8	101.1	101.7	125.0
C_{Me10}-H_{Me10}	101.5	101.3	101.2	125.0
$C_{Me10'}$-$H_{Me10'}$	101.5	101.3	101.2	125.0
$C_{Me6'}$-$H_{Me6'}$	101.6	101.0	104.6	125.0
C_{Me2}-H_{Me2}	101.7	101.8	101.1	125.5
C_{12}-H_{12}	104.7	104.2	103.2	126.0
$C_{12'}$-$H_{12'}$	104.7	104.2	103.2	126.0
C_9-H_9	104.8	104.7	101.0	128.0
$C_{9'}$-$H_{9'}$	105.1	104.9	111.7	128.0
C_5-H_5	105.1	104.9	111.7	128.0
$C_{4'}$-$H_{4'}$	105.4	106.4	116.9	128.0
$C_{5'}$-$H_{5'}$	106.3	106.2	35.1	128.0
$C_{7'}$-$H_{7'}$	110.6	111.0	98.9	150.0
C_7-H_7	113.0	115.1	114.9	150.0
C_3-H_3	114.1	113.7	112.3	150.0
C_{11}-H_{11}	114.5	115.5	74.4	150.0
$C_{11'}$-$H_{11'}$	114.5	115.5	74.4	150.0

Table 4. The scalar coupling constants in Hertz calculated for the 4-(*N*)-squalenoyl-gemcitabine molecule in methanol referred to the gemcitabine fragment. In the last column for comparison the corresponding experimental values are listed.

nuclei pairs	Weighted average on conformers (A), (B), (C) and (D) J^{av}_{calc}	Conformer (A) J_{calc}	Conformer (E) J_{calc}	$J_{\exp}$ (absolute value)
H_{17}-H_{19}	2.0	1.7	1.6	0.6
N*H*CO-H_{14}	-1.0	-0.9	-1.2	0.7
C_{13}-H_{14}	-0.5	-0.5	-0.5	1.3
H_{20}-H_{21}a	3.8	5.4	5.2	2.8
H_{20}-H_{21}b	3.8	5.4	5.2	2.8
C_{14}-H_{15}	-2.3	-2.8	-2.6	3.8
C_{21}-H_{19}	4.9	5.8	5.8	3.9
C_{20}-F	6.0	5.8	5.8	5.0
C_{13}-H_{15}	3.0	2.8	2.8	6.2
H_{17}-F_a	2.9	2.1	2.0	7.4
H_{17}-F_b	2.9	2.1	2.0	7.4
H_{14}-H_{15}	6.7	6.5	6.5	7.6
H_{19}-H_{20}	5.3	5.5	5.4	8.5
C_{16}-H_{15}	7.2	7.3	7.8	9.8
H_{19}-F_a	5.4	6.4	6.5	12.2
H_{19}-F_b	5.4	6.4	6.5	12.2
H_{21a}-H_{21b}	-10.9	-9.1	-9.7	12.8
C_{19}-F_a	23.0	24.8	24.7	23.5
C_{19}-F_b	23.0	24.8	24.7	23.5
C_{17}-F_a	15.2	14.7	14.5	26.9
C_{17}-F_b	22.6	25.7	25.6	38.2
C_{21}-H_{21}	110.3	110.3	109.8	142.9
C_{19}-H_{19}	118.8	113.0	113.1	147.9
C_{20}-H_{20}	120.6	120.9	121.1	148.4
C_{17}-H_{17}	133.1	133.4	133.2	173.3
C_{14}-H_{14}	122.8	123.4	123.2	181.7
C_{15}-H_{15}	147.5	148.3	149.6	185.9
F_a-F_b	181.6	160.8	160.8	240.0
C_{18}-F_a	-330.0	-346.8	-347.1	258.7
C_{18}-F_b	-330.0	-346.8	-347.1	258.7

Table 5. Dihedral angles of squalenoyl chain of conformer (**A**) that have been changed to give the structure of conformer (E)

numbering	value (°)	numbering	value (°)	numbering	value (°)
$C_4C_3C_2C_{Me1}$	-172.176	C_{12},$C_{12}C_{11}C_{10}$	167.284	C_3,C_4,C_5,C_6,	63.084
$C_5C_4C_3C_2$	-120.355	C_{11},C_{12},$C_{12}C_{11}$	-160.031	$C_{Me2}C_2C_{Me1}C_3$	-178.496
$C_6C_5C_4C_3$	177.121	C_{10},C_{11},C_{12},C_{12}	-59.915	$C_{Me6}C_6C_5C_4$	-117.737
$C_7C_6C_5C_4$	59.897	C_9,C_{10},C_{11},C_{12},	-12.483	$C_{Me10}C_{10}C_9C_8$	-94.163
$C_8C_7C_6C_5$	-163.573	C_8,C_9,C_{10},C_{11},	95.644	C_{Me10},C_{10},C_{11},C_{12},	165.505
$C_9C_8C_7C_6$	172.732	C_7,C_8,C_9,C_{10},	-174.985	C_{Me6},C_6,C_7,C_8,	4.165
$C_{10}C_9C_8C_7$	-177.260	C_6,C_7,C_8,C_9,	139.132	OC_3,C_4,C_5,	119.956
$C_{11}C_{10}C_9C_8$	86.746	C_5,C_6,C_7,C_8,	-173.119	NC_3,C_4,C_5,	-58.032
$C_{12}C_{11}C_{10}C_9$	171.580	C_4,C_5,C_6,C_7,	64.169	$C_{13}NC_3$,C_4,	168.272

3.4. Comparison between experimental and theoretical J couplings

A visual comparison between the experimental and theoretical absolute values of J couplings, obtained by DFT as explained above, is shown in the figures 8 and 9. In these graphs the J scalar couplings have been separated according to the two fragments, squalenoyl chain (figure 8) and gemcitabine (figure 9), and, within the fragments, to their numeric value in order to have an immediate comparison and to better follow the trend. From the comparison between experimental and calculated scalar couplings some general comments can be deduced and rationalized. For the gemcitabine fragment, the J couplings values calculated are in good agreement with the experimental data and they are almost completely independent from the considered conformers. This is not surprising since this fragment is the most rigid part of the molecule and, thus, the J couplings are not affected by the dihedral angles trend. The J_{H-F}, J_{F-F} and J_{C-F} couplings are those worse reproduced but again this is not surprising since it is known from the literature[34] that it is difficult to predict J couplings by DFT method when fluorine nuclei are involved. The biggest difference between experimental and calculated values is observed for the J couplings of the squalenoyl chain whose values heavily depend on the dihedral angles. It can be seen from the data reported in Table 3 and in the graph 8 that the calculated scalar couplings for the folded conformers (conformer (**A**) and the weighted average of conformers (**A**), (**B**), (**C**) and (**D**)) are in good qualitative agreement with the experimental values, whereas, the values calculated for the conformer (**E**) are those which deviate more from the experimental data.

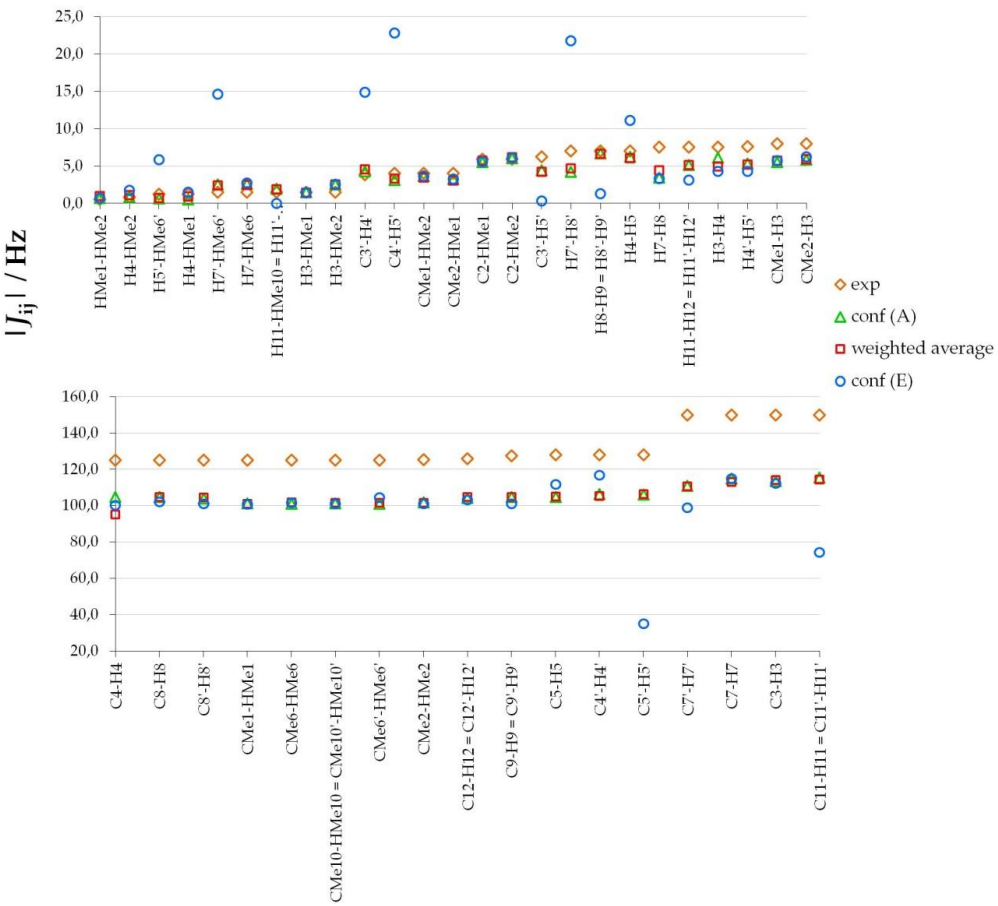


Figure 8. Observed *versus* theoretical *J*-couplings, absolute values, obtained for the squalenoyl chain of 4-(*N*)-squalenoyl-gemcitabine, respectively from the NMR experiments (orange rhombus) and by DFT approach on: conformer (**A**) (green triangle), weighted average of conformers (**A**), (**B**), (**C**) and (**D**) (red square), and conformer (**E**) (blue circle).

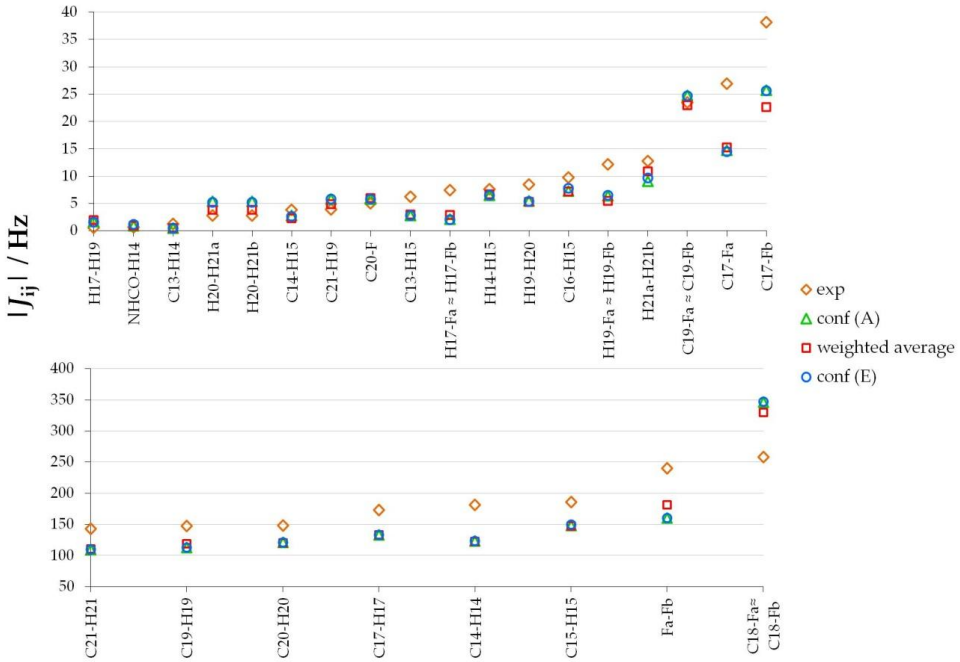


Figure 9. Observed *versus* theoretical *J*-couplings, absolute values, obtained for the gemcitabine fragment of 4-(*N*)-squalenoyl-gemcitabine, respectively from the NMR experiments (orange rhombus) and by DFT approach on: conformer (**A**) (green triangle), weighted average of conformers (**A**), (**B**), (**C**) and (**D**) (red square), and conformer (**E**) (blue circle).

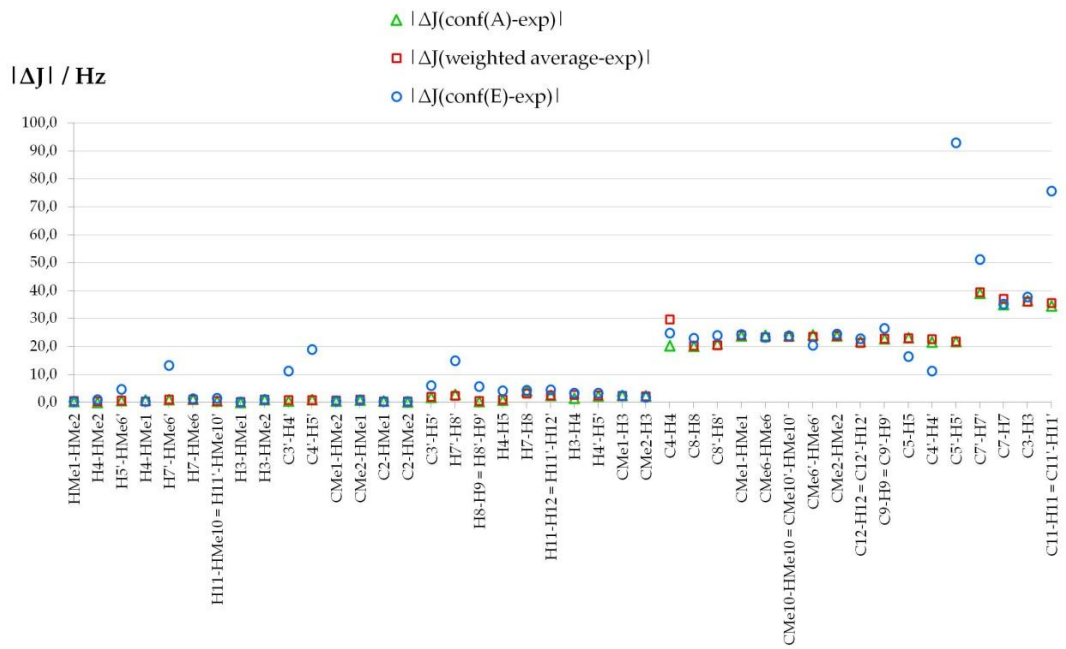


Figure 10. Absolute value of the difference between experimental and theoretical *J*-couplings, calculated for the squalenoyl chain of 4-(*N*)-squalenoyl-gemcitabine on: conformer (**A**) (green triangle), weighted average of conformers (**A**), (**B**), (**C**) and (**D**) (red square), and conformer (**E**) (blue circle).

The similarities and the discrepancies between the experimental and the calculated values can be assessed immediately in figure 10, where the errors, $\Delta J = | J_{ij}(calc) - J_{ij}(exp) |$ (J_{ij} of squalenoyl chain), are plotted for all the conformational situations tested. The figure shows that the errors on the J_{H-H} present a random distribution with a mean of zero and a small deviation (~0.5 Hz) in all the cases, except, for the conformer (**E**) whose errors on these couplings are much larger (mean >10). A similar situation, even more pronounced, it is also visible for the errors on the J_{C-H} couplings, although in this case, also the J_{C-H} of folded conformers are worse reproduced respect to the J_{H-H}. Therefore, comparison between calculated and experimental J-couplings indicates that a folded structure is the most plausible, while the hypothesis of a linear conformer can be excluded. It should be noted that our calculated J-couplings for the folded conformers reveals a remarkable overall agreement with the experimental values although it does not allow us to discriminate between the single most populated conformer (**A**) (~20% of population) and the weighted average of the four conformers (**A**), (**B**), (**C**) and (**D**). It can be said that the J-couplings measured for the Sq-Gem in methanol are consistent with structures in which the flexible squalenoyl chain is strongly folded.

Then, our results explicitly indicate that, in solution, the presence of squalenic fragment involves an increase of conformational mobility of the entire molecule of Sq-Gem compared to the molecule of gemcitabine. Hence, as suggested[13], the improvements in the pharmaceutical properties of this drug and its ability to self-assemble in nanoparticles in aqueous media, are really attributable to its high flexibility. We are aware that our modeling in methanol is far from the physiological condition in which the Sq-Gem acts as a drug, but it is reasonable to assume that these outcomes are also valid in other solvents with similar polarity, since other studies on several squalene derivatives[25] agree with this hypothesis. Furthermore, it is necessary to point out that although our molecular dynamics was relatively short (6 ns) it has been able to give meaningful indications on the conformational distribution.

4. Conclusions

The conformational distribution of squalenoyl-gemicitabine in methanol solution has been investigated applying a combined approach of NMR measurements, MD simulations and quantum mechanical calculations. We chose to test the conformational distribution by measuring the scalar couplings constants, which have been directly related to predictions derived from MD simulations. MD simulations were used to explore the conformational preferences of the molecule in methanol, the NMR parameters were calculated on the dominant conformer (**A**) and on the other most populated conformations (**B**), (**C**) and (**D**) extracted from the MD simulations, and on the linear hypothetical conformer (**E**). Quantum mechanical calculations were performed with the DFT/B3LYP method at the 3-21 G level. Following this theoretical strategy, we found that the J-couplings calculated for the most populated conformer (**A**) and for the weighted average of the four conformers (**A**), (**B**), (**C**) and (**D**) compare favourably with the experimental values, while the J-couplings calculated for the conformer (**E**) are those which deviate more from the experimental data. Therefore, based on these results, the hypothesis of a linear conformer (**E**) can be discarded with a certain confidence. On the contrary, the good agreement between the experimental and calculated data found for the conformers (**A**), (**B**), (**C**) and (**D**) proves the marked flexibility of this compound. It should be noted that although we are not able to discriminate between the most stable conformer (**A**) and the weighted average of the four, our results are in agreement with previous studies on the conformational behaviour of several squalene derivatives[25] that show dynamically folded structures in solution. It is right to point out that these results disagree with the nearly linear structure of squalenoyl chain that occurs in supramolecular assemblies upon nanoprecipitation in water. This is not surprising since our system is quite far from the conditions of self assembling that is mainly driven by hydrophobic forces between several squalene chains. In our case the only interactions to be considered are those between the molecule of Sq-Gem and the solvent molecules, the results of which are consistent with structures in which the flexible squalenoyl chain is strongly folded. It is plausible to assume that these results in methanol are also valid in other solvent with similar polarity, including water, and this would suggest that the flexibility of the squalene to form different stable conformations can help the different hydrophobic chains to organize themselves into supramolecular structures upon nanoprecipitation.

Moreover, it should be emphasized that the experimental J-couplings refer to the average conformation of the molecule, while the calculated values were obtained considering the most populated conformations which represent a population of about 60% excluding the remaining conformations. Therefore, we are obviously far from claiming that our four conformers ensure a real description of the conformational situation in solution, but the combination of experimental NMR data with the MD/DFT approach seems to be a promising and rapid procedure for investigating the three-dimensional structure of bioactive molecules in solution.

Acknowledgements

C.M. and R.F. thank MIUR and, the University of Turin (Ricerca locale 2013-14) for the financial support.

D.L. G. and D.P. M.E. thank the University of Calabria for the financial support and the L'Oréal-UNESCO grant financed by the "L'Oréal Italia per le Donne e la Scienza" program.

Notes and references

[a] *Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via P. Giuria 9, 10125 Torino, Italy*

[b] *Dipartimento di Chimica e Tecnologie Chimiche, Università degli Studi della Calabria, Via P. Bucci, cubo 14C, 87036 Rende (CS), Italy*

[c] *Institute on Membrane Technology (ITM-CNR), Via P. Bucci 17/C, 87036 Rende (CS), Italy*

* *Corresponding authors E-mail: maurizio.ceruti@unito.it, giuseppina.deluca@unical.it*

Electronic Supplementary Information (ESI) available: Table I and Figure I showing the relative energy and the corresponding relative existence probability of the 51 conformers obtained from MD calculations.

1 V. Heinemann, Y.Z. Xu, S. Chubb, A. Sen, L.W. Hertel, G.B. Grindey, W. Plunkett, *Cancer Res.*, 1992, **52**, 533.
2 D.Y. Bouffard, J. Laliberte, R.LMomparler., *Biochem. Pharmacol.*, 1993, **45**, 1857.
3 M.L. Immordino, P. Brusa, F. Rocco, S. Arpicco, M. Ceruti, L. Cattel, *J. Control. Release*, 2004, **100**, 331.
4 F. Castelli, M.G. Sarpietro, M. Ceruti, F.Rocco, L. Cattel, *Mol. Pharmaceut.*, 2006, **3**, 737.
5 F. Castelli, M.G. Sarpietro, F. Rocco, M. Ceruti, L. Cattel, *J. Colloid. Interf. Sci.*, 2007, **313**, 363.
6 B. Stella, F. Rocco, V. Rosilio, J.M. Renoir, L. Cattel, P. Couvreur, Nanoparticles of Gemcitabine derivatives. French Patent deposited on 6/6/2004, n° 04 51365.
7 P. Couvreur, B. Stella, H. Reddy, H. Hillaireau, C. Dubernet, D. Desmaele, S. Lepetre-Mouelhi, F. Rocco, L. Dereuddre-Bosquet, P. Clayette, V. Rosilio, V. Marsaud, J.M. Renoir, L. Cattel, *Nano Lett.*, 2006, **6**, 2544.
8 L.H. Reddy, P. Couvreur, *Adv. Drug Deliv. Rev.*, 2009, **61**, 1412.
9 H.L. Newmark, *Cancer Epidemiol Biomarkers Prev.*, 1997, **6**, 1101.
10 T.J. Smith, *Expert Opin. Investig. Drugs.*, 2000, **9**, 1841.
11 P. Couvreur, L.H. Reddy, S. Mangenot, J.H. Poupaert, D. Desmaele, S. Lepetre-Mouelhi, B. Pili, C. Bourgaux, H. Amenitsch, M. Ollivon, *Small*, 2008, **4**, 247.
12 M.G. Sarpietro, D. Micieli, F. Rocco, M. Ceruti, F. Castelli, *Int. J. Pharm.*, 2009, **382**, 73.
13 S. Voyron, F. Rocco, M. Ceruti, P. Forni, A. Fiorio Pla, M.G. Sarpietro, G.C. Varese, V. Filipello Marchisio, *Mycoses*, 2010, **53**, 481.
14 F. Castelli, M.G. Sarpietro, D. Micieli, B. Stella, F. Rocco, L. Cattel, *J. Colloid Interf. Sci.*, 2007, **316**, 43.
15 D. Cosco, F. Rocco, M. Ceruti, M. Vono, M. Fresta, D. Paolino, *Int. J. Nanomedicine*, 2012, **7**, 2535.
16 D. P. Fairlie, J. D. A. Tyndall, R. C. Reid, A. K. Wong, G. Abbenante, M. J. Scanlon, D. R. March, D. A. Bergman, C. L. Chai, B. A. Burkett, *J. Med. Chem.*, 2000, **43,** 1271.
17 (a) J. Z. Hilt, M. E. Byrne., *Adv. Drug Deliv. Rev.* 2004, **56**, 1599; (b) D. Boehr, R. Nussinov, P. Wright, *Nat. Chem. Biol*., 2009, **5**, 789.
18 (a) M. Karplus, *J. Phys. Chem*., 1959, **30**, 11; (b) M. Karplus, *J. Am. Chem. Soc.*, 1963, **85**, 2870.
19 D. Neuhaus, M. P. Williamson**,** *The Nuclear Overhauser Effect in Structural and Conformational Anaysis*, VCH, New York, 1989.
20 K.Wuthrich, *Angew. Chem. Int. Ed*., 2003, **42**, 3340.
21 (a) E. de Alba, N. Tjandra, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2002, **40**, 175; (b) B. Luy, A. Frank, H. Kessler in *Molecular Drug Properties. Measurement and Prediction*, R. Mannhold ed., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2008, chap. 9, 207; (c) A. Schuetz, J. Junker, A. Leonov, O. F. Lange, T. F. Molinski, C. Griesinger, *J. Am. Chem. Soc.*, 2007, **129**, 15114; (d) H. Sun, U. M. Reinscheid, E. L. Whitson, E. d'Auvergne, C. M. Ireland, A. Navarro-Vázquez, C. Griesinger, *J. Am. Chem. Soc.*, 2011, **133**, 14629; (e) M. E. Di Pietro, C. Aroulanda, D. Merlet, G. Celebre, G. De Luca, *J. Phys. Chem. B*, 2014, **118**, 9007.
22 O. Fisette, P. Lague, S. Gagné, S. Morin, *J. Biomed. Biotechnol.*, 2012, Article ID 254208,12 pages
23 (a) J. A. McCammon, B. R. Gelin, M. Karplus, *Nature*, 1977, **267**, 585; (b) T. J. Lane, D. Shukla, K. A. Beauchamp, V. S. Pande, *Curr. Opin. Struct. Biol*., 2013, **23**, 58; (c) J. Michel, *Phys. Chem. Chem. Phys.*, 2014, **16**, 4465; (d) M. Deleu, J.M. Crowet, M. N. Nasir, L. Lins, *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 2014, **1838**, 3171.
24 M. Ghitti, G. Musco, A. Spitaleri in K. Han et al. (eds.), *Protein Conformational Dynamics, Advances in Experimental Medicine and Biology 805*, Springer International Publishing Switzerland 2014.
25 (a) L. Pogliani, M. Ceruti, G. Ricchiardi, D. Viterbo, *Chem. Phys. Lipids*, 1994, **70**, 21; (b) L. Pogliani, P. Rossi, P. Paoli, M. Ceruti, *Chem. Phys. Lipids* , 1999, **103**, 67; (c) L. Pogliani, M. Milanesio, M. Ceruti, D. Viterbo, *Chem. Phys. Lipids*, 1999, **103**, 81.
26 F. Viola, P. Brusa, G. Balliano, M. Ceruti, O. Boutaud, F. Schuber, L. Cattel, *Biochem. Pharmacol.*, 1995, **50**, 787.
27 T. Marino, N. Russo, E. Tocci, M. Toscano, *Theor. Chem. Acc.*, 2001, **107**, 8.
28 Material Studio (6.0) package of Accelrys. Accelrys Software Inc.: San Diego, 2009.
29 H. C. Andersen, *J. Chem. Phys*., 1980, **72**, 2384 .
30 Gaussian 03, Revision C.02, Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., RobbM. A., Cheeseman J. R., Montgomery J. A. Jr., Vreven T., Kudin K. N., Burant J. C., MillamJ. M., Iyengar S. S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G. A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J. E., Hratchian H. P., Cross J. B., Adamo C., Jaramillo J., Gomperts R., Stratmann R. E., Yazyev O., Austin A. J., Cammi, R., Pomelli C., Ochterski J. W., Ayala P. Y., Morokuma K., Voth G. A., Salvador P., Dannenberg J. J., Zakrzewski V. G., Dapprich S., Daniels A. D., Strain M. C., Farkas O., Malick D. K., Rabuck A. D., Raghavachari K., Foresman J. B., Ortiz J. V., Cui Q., Baboul A. G., Clifford S., Cioslowski J., Stefanov B. B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R. L., Fox D. J., Keith T., Al-Laham M. A., Peng C. Y., Nanayakkara A., Challacombe M., Gill P. M. W., Johnson B., Chen W., Wong M. W., Gonzalez C., Pople J. A., Gaussian, Inc.: Pittsburgh PA, 2003.
31 T. Hegaker, M., Jaszuński, M. Pecul, *Prog. NMR Spectrosc.*, 2008, **53**, 249.
32 R. H. Contreras, V. Barone, J. C. Facelli, J. E. Peralta, *Ann. Rep. NMR Spectrosc.*, 2003, **51**, 167.

33 (a) T. Fäcke, S. Berger, *J. Magn. Reson. Ser. A*, 1995, **113**, 114; (b) J. M. Nuzillard, *J. Magn. Reson.*, 2007, **187**, 193.

34 P. Lanto, J. Vaara, T. Helgaker, *J. Chem. Phys.*, 2002, **117**, 9630.

5 .