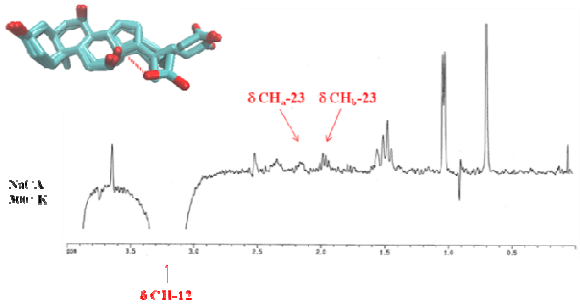
MedChemComm

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/medchemcomm

Table of Contents



DFT and NMR spectroscopy studies unveil three major minima conformations of cholic acid that may affect its biological properties.

Conformational Properties of Cholic Acid, a Lead Compound at the Crossroads of Bile Acid Inspired Drug Discovery.

Authors: Antimo Gioiello,[1] Francesco Venturoni, [1†] Sara Tamimi,[1] Chiara Custodi, [1] Roberto Pellicciari,[1,2] Antonio Macchiarulo. [1]*

Address. [1] Dipartimento di Chimica e Tecnologia del Farmaco, Università degli Studi di Perugia, 06123 Perugia, Italy. Fax: +39 075 585 5161; Tel: +39 075 585 5160. [2] TES Pharma S.r.l. via Palmiro Togliatti 22bis 06073 Loc. Terrioli, Corciano (Perugia) Italy. [†] Present address: Novartis Pharma AG, TRD, WSJ-42.2.08, 4002 Basel, Switzerland.

KEYWORDS. Bile acids, Conformation, Nuclear Receptors, GPCRs, Lipids, DFT, NMR.

* Antonio Macchiarulo

Dipartimento di Chimica e Tecnologia del Farmaco

Via del Liceo 1

06123 Perugia (Italy)

Tel +39 075 585 5160

Fax +39 075 585 5161

e.mail: antonio.macchiarulo@unipg.it

Abbreviations. Bile acids (BAs); Farnesoid X receptor (FXR); G protein-coupled bile acid receptor 1 (GPBAR1, TGR5); Cholic acid (CA); Density functional theory (DFT); Nuclear magnetic resonance (NMR); Systematic pseudo-Montecarlo routine (SPMC); Root mean square deviation (RMSD); Multidimensional scaling (MDS); Polar surface area (PSA); NMR-correlation spectroscopy (COSY); Heteronuclear multiple-quantum correlation (HMQC); Nuclear Overhauser effect spectroscopy (NOESY); Nuclear Overhauser effect (NOE).

Abstract.

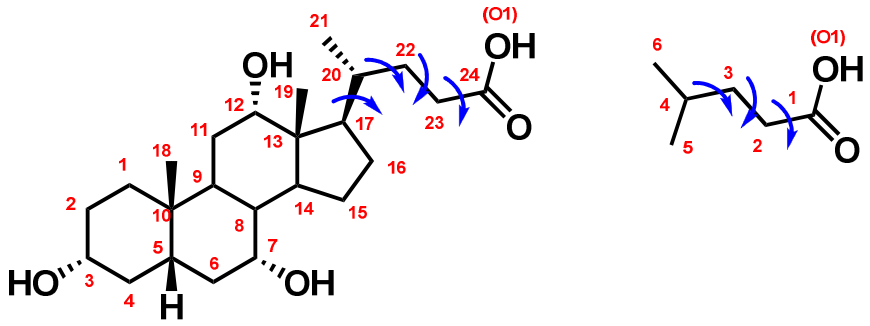
Cholic acid is an endogenous primary bile acid endowed with endocrine signaling functions and digestive properties. In this work its conformational profile and the impact of the steroid nucleus are investigated using DFT studies and NMR spectroscopy. As a result, three major minima conformations are found with an ‘extended’, ‘semi-extended’ and ‘folded’ side chain, respectively. While the ‘extended’ and ‘semi-extended’ conformations are often found as bioactive conformations in the binding site of co-crystallized biological targets, the ‘folded’ conformation is only experimentally observed in NMR studies. Overall, the results of this study provide additional basis for the interpretation of biological functions of bile acids, and disclose new pharmacophoric templates for the design of constrained bile acid analogues.

Introduction.

Primary and secondary bile acids (BAs) are essential constituents of endogenous bile, with the former originating from the catabolism of cholesterol in the liver and the latter being formed from primary BAs through the action of the intestinal bacteria.[1] Although formerly known for their digestive properties, BAs are now widely recognized as signaling molecules endowed with genomic and nongenomic endocrine functions that are mediated by their interaction with a plethora of nuclear and cell membrane receptors including the nuclear farnesoid X receptor (FXR) and the G protein-coupled bile acid receptor 1 (GPBAR1, TGR5).[2-4]

Cholic acid (CA, **1**, Figure 1) is a primary BA in human and the most conserved BA among mammal species.[5] In structural terms, it bears three hydroxyl groups oriented on the polar concave side of the molecule at positions $C_3\alpha$, $C_7\alpha$ and $C_{12}\alpha$, respectively. This feature bestows high solubility and relatively poor detergency to CA with respect to other endogenous BAs.[6] Given its facial amphipathic properties and ensuing ability to form micellar aggregates with specific supramolecular arrangement, CA has been extensively used as building block in supramolecular chemistry to transport ion and polar molecules across the membrane.[7-10] From a pharmacological point of view, CA is devoid of activity toward FXR (EC_{50}> 100 μM) and shows a moderate agonistic activity at TGR5 (EC_{50}= 13.6 μM),[11,12] that accounts for its ability to prevent metabolic syndrome in diet-induced obese mice.[13] Moreover, it has been reported to act as endogenous regulator of the intestinal microbiome, whose equilibrium is critical for human health and disease.[14,15] The European Medicines Agency has recently given a positive recommendation for the therapeutic use of CA in the treatment of inborn errors of primary bile acid synthesis.[16,17]

During its journey in human body as signaling hormone or as carrier of compounds across lipid bilayer, CA is exposed to many processes of molecular recognition by different biological targets and environments that include enzymes, receptors, transporters and cell membranes. Molecular recognition is a complex and dynamic event that involves the conformational rearrangement of both binding partners according to their energetic landscapes.[18]



1 **2**

Figure 1. Chemical structures of Cholic acid (CA) and isohexanoic acid. Arrows indicate torsional angles used for the conformational analyses.

The binding of ligand to its biological target determines a restriction of the conformational freedom to a bioactive conformation that may be different from the global energy minimum of the solvated molecule. Survey studies have reported that local minimum conformations are generally preferred by ligands when binding to proteins and acceptable conformational strain energies of bioactive conformations are very often lower than 5 kcal/mol.[19-26]

Likewise to other BAs, CA is endowed with a rigid steroid scaffold and a flexible side chain that counts four torsional angles. Assuming that each one of these angles can exist in three distinct rotational states, BAs can theoretically adopt 81 conformations, distributed between different energetic states.

Previous pioneering works have contributed to shed lights on the importance of the conformation of the side chain of BAs in affecting some of their physicochemical parameters related to detergency and biological properties.[27-30] Likewise, controlling the conformational profile of CA has been reported as pivotal to tune the property of the molecule as synthetic membrane transporter.[31] In this work, we report a thorough study on the conformational properties of CA, using density functional theory (DFT) and nuclear magnetic resonance (NMR) studies to characterize the number and properties of minima conformations. The influence of the steroid nucleus on the conformational profile of the side chain is also investigated by performing a comparative conformational analysis

with isohexanoic acid (**2**), a molecule mimicking the side chain of CA. A survey of structures co-crystallized with CA in the protein data bank (PDB)[32] is finally reported with the aim of identifying the number and properties of bioactive conformations adopted by the side chain of this versatile molecule.

Results and Discussion.

In order to locate all of the possible minima conformations of CA, a systematic pseudo-Montecarlo routine (SPMC)[33] was used to rotate the four torsional angles: C_{16}-C_{17}-C_{20}-C_{21} (α dihedral); C_{17}-C_{20}-C_{22}-C_{23} (β dihedral); C_{20}-C_{22}-C_{23}-C_{24} (γ dihedral); C_{22}-C_{23}-C_{24}-O_1 (δ dihedral). After DFT B3LYP/6-31++G** energy minimization of the resulting conformations using the standard Poisson-Boltzmann solvent model,[34] 19 unique minima conformations were found within an energy window of 15.3 kcal/mol from the global minimum (Table 1). The population (p_i) of each minimum conformation was calculated at 300 K as percentage using the Boltzmann distribution. A distance matrix of these conformations was generated using the root mean square deviation values (RMSD) as calculated on the carbon atoms and hydroxyl oxygen atoms (Table S1 of supporting information). This matrix was instrumental to perform a multidimensional scaling (MDS) analysis, and provide a map of the conformational states of CA (Figure 2) in which conformations laying in close proximity share similar geometries of side chain torsional angles. In order to get more insights into the properties of these minima points, Figure 3 depicts a plot of the distance (Å) between C_{24} and O_3 *versus* the polar surface area (PSA) for each of the conformations.

Table 1. Minima conformations of CA with relative values of torsional angles (dihedral), conformational gap energy from global minimum (ΔE), Boltzmann population (p_i), solvation energy and dipole.[a]

Name	Dihedral α	Dihedral β	Dihedral γ	Dihedral δ	ΔE (kcal/mol)	ΔE (kJ/mol)	p_i at 300K (%)	Solvation Energy (kcal/mol)	Dipole (Debye)
Conf.4	-179.4	-159.1	174.1	74.2	0.0	0.00	76.8	-92.9	42.4
Conf.1	173.4	61.9	-174.2	-66.8	1.0	4.0	15.4	-92.6	40.7
Conf.39	-60.4	-69.1	154.4	157.7	1.4	5.9	7.1	-78.5	30.1
Conf.13	85.7	-155.0	71.3	80.3	3.3	13.6	0.3	-84.2	25.1
Conf.12	-56.4	-159.3	64.5	66.5	3.7	15.3	0.2	-89.2	38.2
Conf.11	-62.3	-88.3	67.6	136.9	4.0	16.6	0.1	-79.0	26.5
Conf.54	95.3	84.8	82.2	-43.6	4.6	19.1	0.04	-91.7	39.8
Conf.38	-72.7	86.0	-145.8	-52.1	4.9	20.4	0.02	-78.7	30.2
Conf.29	112.9	-79.3	-62.8	156.7	5.2	21.7	0.01	-87.9	28.0
Conf.37	-175.1	73.7	-63.1	-55.4	6.0	25.0	0.003	-86.9	27.6
Conf.56	81.3	-147.7	172.8	87.9	6.0	25.3	0.003	-86.8	31.0
Conf.32	-61.7	-161.7	52.0	62.2	6.4	26.9	0.002	-86.4	37.5
Conf.33	66.3	62.8	65.5	-67.0	6.7	28.0	0.001	-91.1	36.5
Conf.63	171.4	-128.8	70.0	103.5	6.8	28.3	0.001	-86.1	27.2
Conf.36	56.9	-101.1	-42.6	133.0	7.9	33.1	<0.001	-80.1	30.3
Conf.15	71.3	69.1	-106.9	-59.7	8.9	37.1	<0.001	-85.3	24.5
Conf.14	-56.2	-179.8	-81.0	44.9	8.9	37.1	<0.001	-84.0	30.9
Conf.61	-60.2	86.9	70.8	-52.0	10.2	42.8	<0.001	-83.1	29.6
Conf.35	-60.2	-63.9	-54.5	161.0	15.3	63.8	<0.001	-75.8	34.7

[a] Minima conformations are sorted in ascending order of conformational gap energy (ΔE) from the global minimum (conf.4).

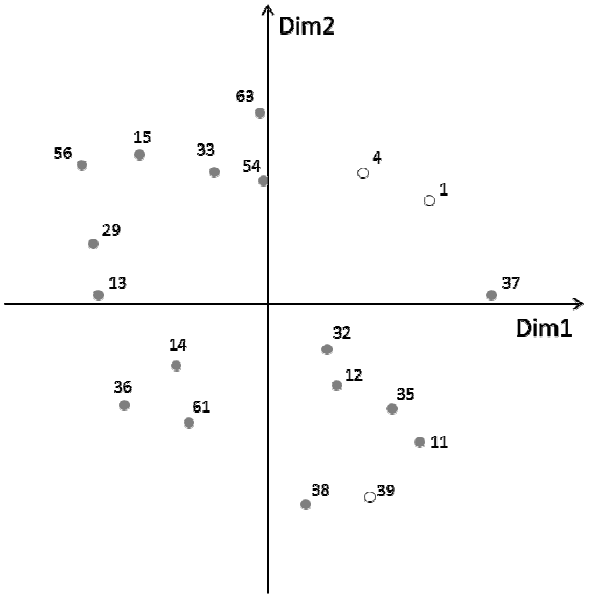


Figure 2. MDS map (Kruskal Stress = 0.128) of 19 minima conformations of CA. Top energetic favoured minima conformations (conf.4, .1, .39) are indicated in white circles. Dim1 and Dim2 are dimensional descriptors generated by the MDS analysis on the basis of the distance matrix of RMSD values (Å) between minima conformations of CA (Table S1 of supporting information).

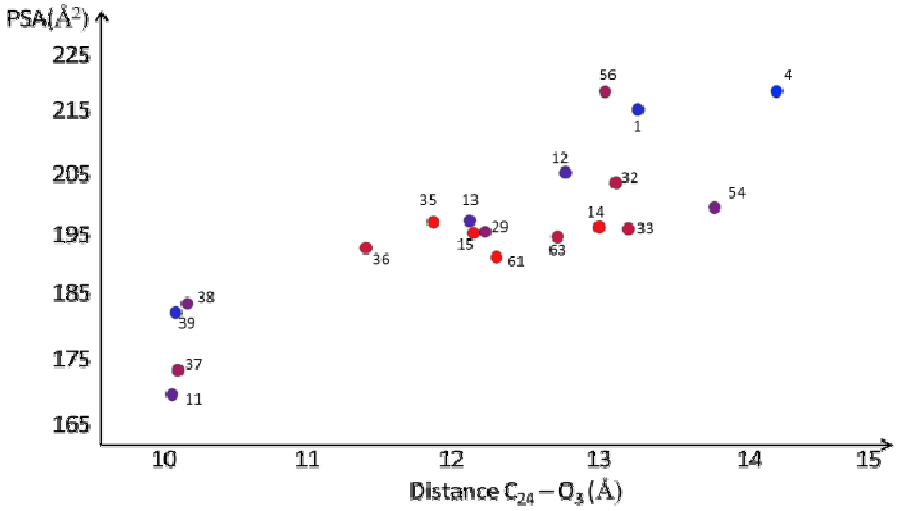


Figure 3. Plot of distance (C_{24}-O_3) versus PSA values of minima conformations. Each conformation is colored according to the Boltzmann distribution from blue (highly populated) to red (poorly populated).

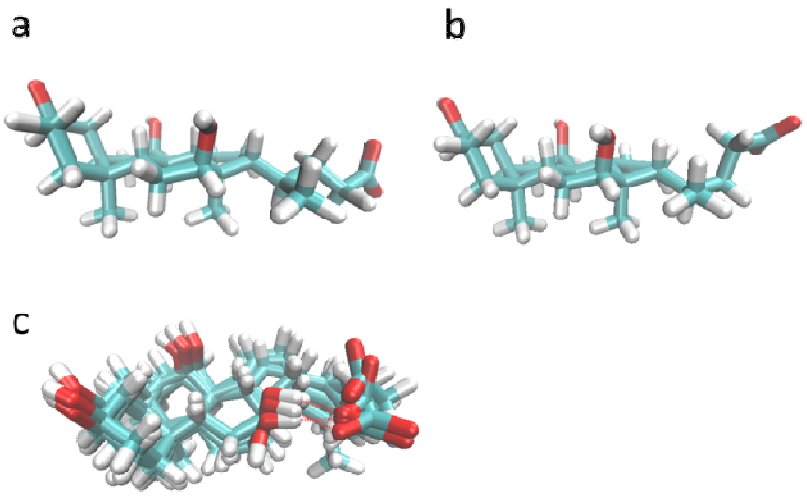


Figure 4. Most populated minima conformations of CA. a) Global minimum conf.4; b) local minimum conf.1; c) superposition of folded minima conformations (conf.39, .11, .38, .37), the intramolecular hydrogen bond is shown as red dashed line in conf.39, .11 and .38.

The inspection of this plot shows the presence of 'folded', 'semi-extended' and 'extended' minima conformations endowed with increasing PSA, that span a range of distances from 10 Å to 14 Å with PSA values from 173 $Å^2$ to 216 $Å^2$. Most of these conformations are within the interval 11-13 Å of C_{24}-O_3 distance. In particular, a fully 'extended' conformation of the side chain features the global minimum of CA (conf.4, p_i= 76.8%), with α, β, and γ dihedrals adopting *trans* conformations and δ dihedral assuming a *gauche* conformation. Of note, the plane of the carboxylic group is placed perpendicular to the steroid nucleus, with C_{23} being in *anti* with respect to C_{17}, and one carboxylic oxygen oriented on the α-face (Figure 4a). The methyl group of the side chain (C_{21}) is in *gauche* with respect to C_{19} and *anti* with respect to C_{15} as well as to the carboxylic group. As a result, the PSA (= 216.8 $Å^2$) as well as the dipole moment (Dipole = 42.4 debye) of the global minimum are the highest among all of the minima points of CA (Table 1). Given its specific disposition perpendicular to the steroid plane and contiguous to the α-face, the carboxylic group extends the PSA of the α-face in the global minimum of CA, bestowing better amphipathic property to such a conformation. The closest local minimum to the global minimum conformation of CA is conf.1 (p_i

= 15.4%, Figure 4b). In this local minimum, α and γ dihedrals of the side chain are in *trans* conformations, while β and δ dihedrals adopt *gauche* conformations. Although conf.1 is also endowed with a relative large PSA, its side chain is less extended than conf.4, being C_{23} *eclipsed* with respect to C_{17}. As a consequence, it is likely that this local minimum may have a reduced compliance to the amphipathic property of the molecule. The inspection of figure 3 reveals also four 'folded' conformations of **1** with distances between C_{24} and O_3 of 10 Å, and low values of dipole moment (Table 1) as well as PSA (conf.39, .11, .38, .37). Three of these local minima points (conf.39, .11, .38) feature an intramolecular hydrogen bond involving the carboxylic group and the hydroxyl moiety in C_{12} (Figure 4c). Although three of these conformations are not significantly populated (conf.11 p_i = 0.1%; conf.38 p_i = 0.02%; conf.37 p_i = 0.003%), conf.39 shows a p_i of 7.1%, suggesting a fair presence of this local minimum in solution at 300 K.

Of note, conf.11, .37 and .38 incorporate one *trans* torsional angle (conf.11: δ dihedral; conf.37: α dihedral; conf.38: γ dihedral) and three *gauche* angles (conf.11: α, β, γ dihedrals; conf.37: β, γ, δ dihedrals; conf.38: α, β, δ dihedrals) in the side chain, whereas the side chain of conf.39 has two *trans* torsional angles (γ, δ dihedrals) and two *gauche* angles (α, β dihedrals). This observation may account for the more energetic favored state of the latter conformation over the formers, with less strain energies associated to the *trans* torsional angles than to the *gauche* dihedrals.

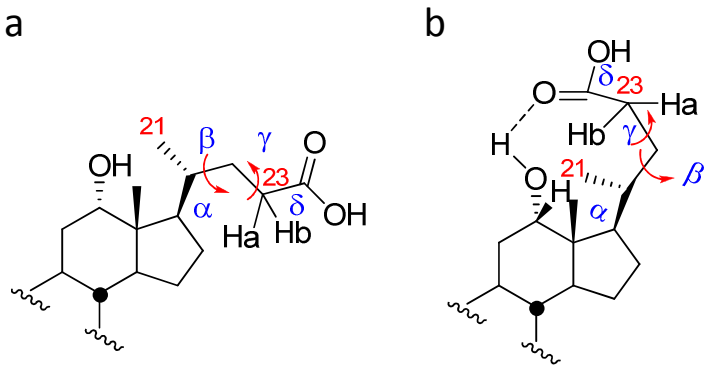


Figure 5. Extended (a) and folded conformation (b) of the side chain. NOE effects of protons in positions C_{21} and C_{23} (red labeled) are studied to assess the conformation of the side chain.

The conformational profile of CA was also investigated using NMR-correlation spectroscopy (COSY), heteronuclear multiple-quantum correlation (HMQC) and nuclear Overhauser effect spectroscopy (NOESY).[35] The aim of this part of the study was to experimentally pinpoint the presence of the folded conformation (conf.39) in CA. According to DFT studies, this conformation is not significantly populated (p_i = 7.1%) at 300 K in water. Hence, we envisaged to run NMR experiments using conditions to favor its population rate and focusing on the variation of the Overhauser effect (NOE) related to protons in C_{21} and C_{23}. As illustrated in Figure 5, these atoms can be useful to study 'extended' or 'folded' conformations of the steroid side chain. NOESY experiments were thus performed for CA and its sodium salt (concentration 0.014 M),[36-39] using an aprotic solvent at different decreasing temperatures (300 K, 280 K, 260 K and 240 K). The solvent was THF-d8/DMSO-d6 (3.5:1, v/v) because it proved to be the best solvent in the range of adopted temperatures. From a careful analysis of the results (Figure 6, Table S2 of supporting information), a variation of the NOE effect was observed as resulting from the protons located at C_{21} and C_{23}. In particular, their relative ratio varied from 1:1 (300 K) to 1:3.4 (240 K), suggesting a lower conformational freedom of the CA side chain at low temperatures as a consequence of the stabilization of the intramolecular hydrogen bond between the $C_{12}\alpha$ hydroxyl group and the carboxylic moiety of the side chain. Indeed, the 'folded' conformation (conf.39) justifies a reduced distance between the C_{23}Ha/C_{23}Hb and the C_{21} methyl group as well as a more pronounced NOE effect. On the contrary, a free rotation of the β-and γ-angles would be allowed without this hydrogen bond ensuring the spatial equivalence of the corresponding protons (Figure 5a), as observed at room temperature. Noteworthy, the NMR analysis of sodium cholate provides further experimental evidence supporting the 'folded' conformation of CA. Firstly, we found that the C_{21} methyl group has a NOE effect almost exclusively with the CH_{23}b, an effect that was nearly absent with the epimeric proton demonstrating the absence of conformational freedom of the side chain. Secondly, this NOE interaction was clearly evident even at 300 K indicating a more stable 'folded' conformation in the presence of a charged carboxylate group at this temperature (Figure S1 of

supporting information). Finally, the analysis of the row associated with the selective irradiation of the proton at $C_{12}\alpha$ position of the sodium cholate evidenced a weak NOE effect between the $C_{12}H$ and $C_{23}Hb$ (Figure S2 of supporting information), further confirming the 'folded' conformation shown in Figure 5b.

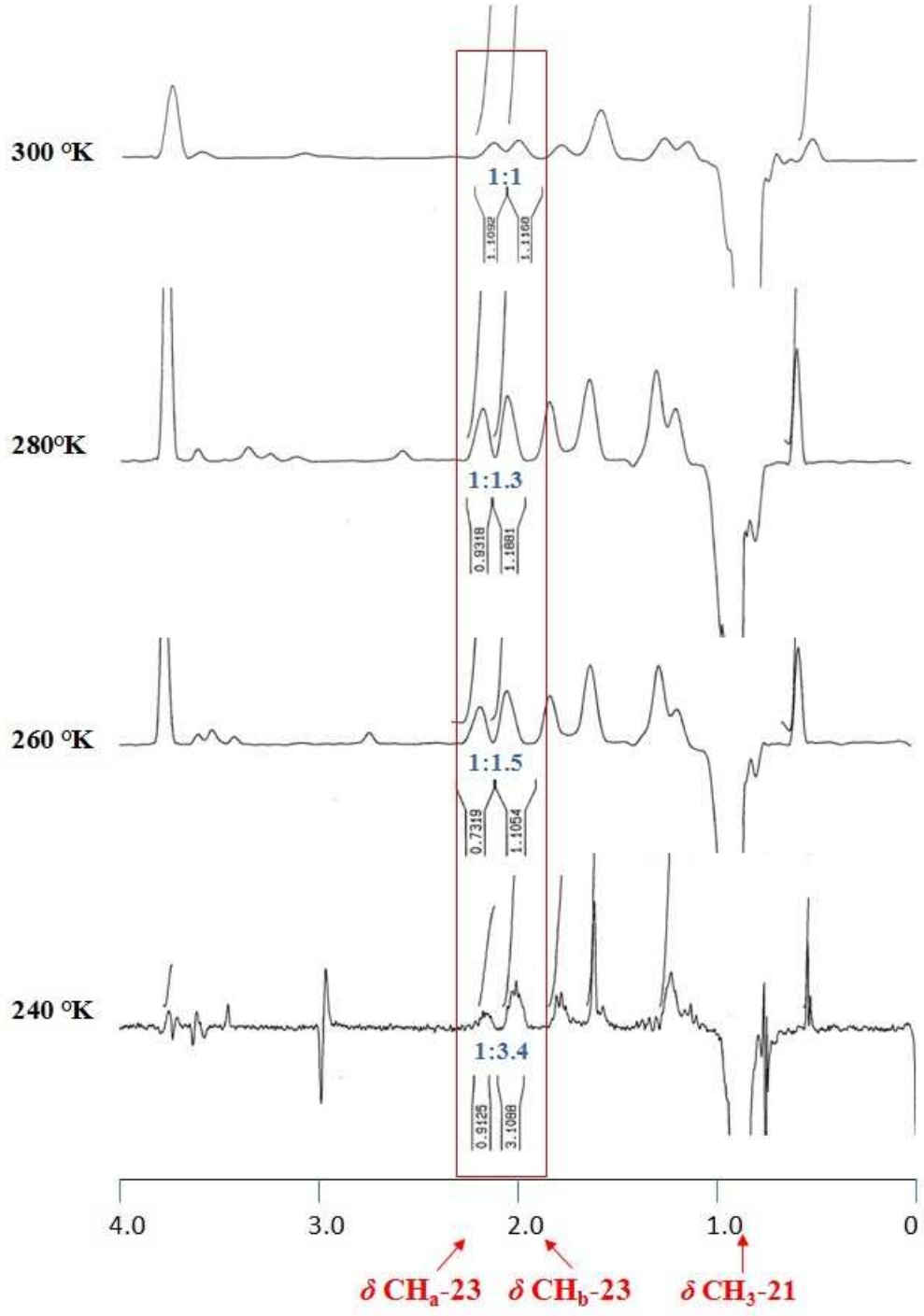


Figure 6. NOE effect between the protons at C_{21} and C_{23} position of CA at different temperatures.

Although the 'extended' global minimum (conf.4) and the 'semi-extended' local minimum (conf.1) would suggest a small influence of the steroid nucleus on the conformational profile of the side chain, the 'folded' conformation of CA (conf.39) points out that this is not the case, highlighting an important role for the steroid nucleus. To further investigate the impact of the steroid nucleus on the conformational profile of CA, we carried out a conformational analysis on isohexanoic acid using SPMC and DFT B3LYP/6-31++G**. Isohexanoic acid mimes the side chain of CA without the steroid nucleus. Using the same computational protocol described above for CA, the following dihedral angles were analyzed: C_5-C_4-C_3-C_2 (β dihedral); C_4-C_3-C_2-C_1 (γ dihedral); C_3-C_2-C_1-O_1 (δ dihedral). As a result, 9 unique minima conformations were found within an energy window of 2.4 kcal/mol from the global minimum (Table 2). Of note, the global minimum of isohexanoic acid (conf.iv, p_i = 27.3%) shows a distinct pattern of all *trans* geometries for dihedrals β-δ with respect to the same angles found in the global minimum of CA. Conversely, its closest local minimum (conf.iii, p_i = 27.0%) displays a pattern of dihedrals [*gauche*-(β)/*trans*-(γ)/*gauche*-(δ)] similar to that observed in the 'semi-extended' local minimum of CA (conf.1). Other medium populated local minima of isohexanoic acid (conf.ii, p_i = 18.9%; conf.i, p_i = 16.9%) have patterns of side chain dihedrals that are not observed in the most populated minima conformations of CA (conf.4, conf.1, conf.39). Hence, these observations suggest that the steroid nucleus does play a major role in influencing the conformational profile of the side chain of CA at different levels. Firstly, it enlarges the number of minima conformations by the rotation of the α dihedral angle (19 minima conformations of CA *versus* 9 minima conformation found in isoexanoic acid). At the same time, the steroid nucleus bestows some rigidity to the side chain, favoring the population of mainly one out of three major minima conformations (conf.4 p_i = 76.8%; conf.1 p_i = 15.4%; conf.39 p_i = 7.1%) as compared to the four populated minima conformations observed in the isohexanoic acid (conf.iv p_i = 27.3%; conf.iii p_i = 27.0%; conf.ii p_i = 18.9%; conf.i p_i = 16.9%). Lastly, it impacts on the geometry of the side chain dihedrals by providing steric interactions and stereospecific intramolecular hydrogen bond (conf.39).

Table 2. Minima conformations of isohexanoic acid with relative values of torsional angles (dihedral), conformational gap energy from global minimum (ΔE), Boltzmann population (p_i), solvation energy and dipole.[a]

Name	Dihedral β	Dihedral γ	Dihedral δ	ΔE (kcal/mol)	ΔE (kJ/mol)	p_i at 300K (%)	Solvation Energy (kcal/mol)	Dipole (Debye)
conf.iv	171.7	-176.3	169.6	0.0	0.0	27.3	-76.9	15.1
conf.iii	64.4	175.5	11.0	0.01	0.03	27.0	-76.9	15.1
conf.ii	169.9	-64.4	138.3	0.2	0.9	18.9	-75.8	12.7
conf.i	65.5	63.9	-135.4	0.3	1.2	16.9	-75.7	12.6
conf.ix	-63.3	179.7	1.1	1.2	5.2	3.5	-76.8	14.9
conf.v	175.7	81.0	45.1	1.3	5.7	2.8	-75.6	12.6
conf.vi	61.0	-80.2	-47.5	1.4	5.8	2.6	-75.6	12.6
conf.x	-61.5	77.2	48.0	2.3	9.8	0.5	-75.7	12.3
conf.xi	-65.8	-77.3	135.4	2.4	10.2	0.4	-75.6	12.3

[a] Minima conformations are sorted in ascending order of conformational gap energy (ΔE) from the global minimum (conf.IV).

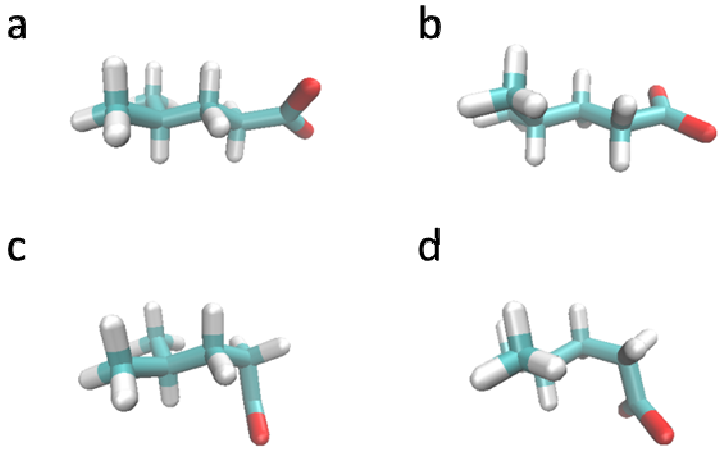


Figure 7. Most populated minima conformations of isohexanoic acid. a) Global minimum conf.iv; b) local minimum conf.iii; c) local minimum conf.ii; d) local minimum conf.i.

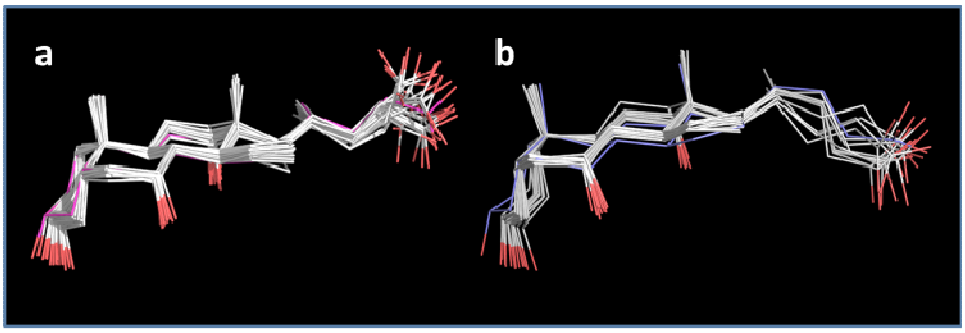


Figure 8. The global minimum of CA (conf.4, red sticks) superimposed with its closest bioactive conformations from Table S3 of supporting information. b) The high populated local minimum of CA (conf.1, blue sticks) superimposed with its closest bioactive conformations from Table S3 of supporting information.

Upon binding to a target protein, the conformational degree of freedom of CA is confined to a specific biologically active conformation. Then, in order to get insights into its bioactive conformations, we carried out a survey of the atomic coordinates of CA into crystallographic complexes. A set of 11 entries was collected including transporters and enzymes,[40-46] and comprising 39 conformations of CA. These conformations can be grouped into 7 distinct bioactive conformations according to the nearly *trans* and *gauche* geometries of dihedral angles α-δ (group A-G, Table 3). In particular, the most found bioactive conformation (group A) shows dihedral geometries in agreement with the global minimum of CA (conf.4), featuring a *trans*-(α)/*trans*-(β)/*trans*-(γ)/*gauche*-(δ) pattern of dihedrals in the side chain. Another often found bioactive conformation of CA (group B) bears the pattern of side chain dihedrals observed in conf.1 [*trans*-(α)/*gauche*-(β)/*trans*-(γ)/*gauche*-(δ)]. Of note, the remaining five groups (C-G) have patterns of side chain dihedrals that are not observed in the major populated conformations of CA (conf.4, conf.1, conf.39).

Table 3. Groups of bioactive conformations of CA according to geometries of side chain dihedrals.[a]

Group	Dihedral α	Dihedral β	Dihedral γ	Dihedral δ	Entry
A	*trans*	*trans*	*trans*	*gouche*	3ELZ-1; 3ELZ-3; 3ELZ-4; 3ELZ-7; 3ELZ-9; 3EM0-4; 3EM0-6; 2QO5-2; 1TW4-1
B	*trans*	*gouche*	*trans*	*gouche*	3ELZ-13; 3EM0-5; 3EM0-8; 2FT9-2; 2QO4-1; 2QO6-4
C	*trans*	*trans*	*gouche*	*gouche*	1EE2-1; 3ELZ-5; 3ELZ-6; 3EM0-1; 3EM0-2; 3EM0-3; 3EM0-7
D	*trans*	*trans*	*gouche*	*trans*	1EE2-2; 2DQY-2; 3ELZ-11; 2FT9-1; 1TW4-3; 1TW4-4
E	*trans*	*trans*	*trans*	*trans*	2DQY-1; 3ELZ-2; 3ELZ-12; 2QO5-3; 1TW4-2
F	*trans*	*gouche*	*trans*	*trans*	2RLC-1; 2AZY-1; 3ELZ-8; 3ELZ-10
G	*trans*	*gouche*	*gouche*	*gouche*	3EM0-9

[a] Nearly *trans* dihedrals are defined those angles with values between ±120 and ±180 degrees; nearly *gauche* dihedrals are defined those angles with values between 0 and ±120 degrees.

Although this result may suffer from the experimental uncertainties of ligand atomic coordinates associated to the use of complexes with poor or medium resolution data,[47] it points out a clear preference of CA to bind biological targets in 'extended' (conf.4) and 'semi-extended' conformations (conf.1). This observation is further supported when RMSD values are calculated to identify which one of the major populated conformations of CA (conf.4, conf.1, conf.39; pi > 99%) is closest to the bioactive conformations. While a bioactive conformation close to the 'folded' conf.39 minimum was not observed, bioactive conformations close to conf.4 (global minimum) and

conf.1 were found in the 67% and 33% of instances (Figure 8, Table S3 of supporting information), reflecting their ranking of energetic values. This trend is also observed after the removal of entries with poor resolution factors (resolution > 2 Å, Table S3 of supporting information). Collectively, these results are in agreement with some large-scale studies in literature on the bioactive conformational changes of drug-like molecules, suggesting a general trend of ligands to adopt an extended conformation when binding to target proteins with exposed hydrophobic surfaces contacting to protein residues.[25]

Once observed conf.4 and conf.1 of CA in biological environments, an open question is whether the 'folded' conf.39 may have any biological meaning since it was identified as most populated in a biologically not compatible aprotic solvent. Supporting a hypothetical biological function of conf.39, it is worth mentioning that the aprotic solvent may mimic to some extend conditions found in the environment of biological membranes. Previous studies have shown that conjugation with CA aids the passive transport pathway of small peptides across the intestinal epithelium, beside the active transport pathway mediated by the intestinal bile acid carrier.[48] Accordingly, the less polar 'folded' conformation of CA may account for better interaction properties of this compound within membrane lipid bilayers, favoring the passive transport properties of CA. Sustaining further this hypothesis, a recent study has shown the impact of stereospecific intramolecular hydrogen bond on increasing cell permeability of compounds.[49]

Conclusions.

The conformational profile of CA is found to be composed of 19 minima, with three of them accounting for more than 99% of the Boltzmann distribution at 300 K. A comparative conformational analysis with isohexanoic acid suggests a major role for the steroid nucleus in influencing the conformational profile of CA, bestowing rigidity to the BA side chain as well as determining the geometry of side chain dihedrals. The 'extended' side chain conformation of CA (conf.4) is the global minimum, and is often adopted as bioactive conformation in crystal

complexes. Bioactive conformations close to the 'semi-extended' extended conformation of CA (conf.1) are also adequately found in binding sites of proteins. Although the folded conformation of CA (conf.39) was hitherto only observed in NMR experiments using an aprotic solvent system, it may account for specific properties of this primary BA, such as better cell permeability, as well as provide a suitable template for the design of constrained analogues as chemical tools to unravel the complexity of BA signaling pathways.

Experimental Parts.

Conformational analysis. Given its pKa of 5,[50] CA is fully deprotonated at physiological pH, and as such it was modeled in this study. Likewise, isohexanoic acid was considered in the deprotonated form to mimic CA side chain properties. Conformational analyses were carried out using two stages of calculations. Firstly, the systematic pseudo-Montecarlo method (SPMC)[33] of MacroModel (version 9.9) was applied to locate all possible conformational minima of CA and isohexanoic acid. The calculation was performed using the OPLS-2005 field and implicit water solvent. The maximum number of steps was set to 100.000 and the number of steps for each of the four selected rotatable bonds to 10.000 (Figure 1). The number of unique structures to be saved for each rotatable bond was set to 150.000 within an energy window of 50 kcal/mol from the global minimum. The root mean square deviation (RMSD) cutoff to remove redundant conformers was set to 0.1 Å. Each conformation was submitted to energy minimization using the OPLS-2005 force field, the truncated newton conjugated gradient (TNCG) method, a maximum number of 100.000 iteration steps, and a convergence gradient criteria of 0.05 kcal/mol/Å. RMSD values calculated on C.sp3, C.sp2, O.sp3 atoms of CA, and C.sp3, C.sp2 of isohexanoic acid were used for the selection of the relative unique conformations (RMSD < 0.1 Å). As a result, 23 conformations for CA and 16 conformations for isohexanoic acid were found and submitted to the second stage of calculations, consisting in geometry optimizations with DFT B3LYP/6-31G** and diffuse++ function. All DFT calculations were carried out using Jaguar (version 7.9) as implemented in Schrodinger suite, with a maximum

grid density and an ultrafine accuracy level. Solvation free energies in water were calculated using the standard Poisson-Boltzmann solvent model.[34] Resulting DFT optimized geometries of CA and isohexanoic acid minima conformations were respectively compared on the basis of RMSD values and visual inspection to discard multiple copies of the same minima conformations (RMSD < 0.1 Å, calculated using C.sp3, C.sp2, O.sp3 atoms for CA, and C.sp3, C.sp2 atoms for isohexanoic acid). A total of 19 and 9 unique geometric structures with relative energies were thus obtained for CA and isohexanoic acid, respectively. The RMSD values of CA were also used to construct a distance matrix of 19 conformations necessary to perform a Multidimensional Scaling (MDS) analysis.[51] The aim of this method is to build a mapping of a series of individual conformations from the distance matrix between these individuals. To build an optimal representation, the MDS algorithm minimizes a criterion called "Stress". The closer is the stress to zero, the better is the representation. The polar surface area for each of the minima conformations was calculated using the fraction of hydrophilic surface accessible to the solvent as implemented in Qikprop program. Boltzmann distribution at 300 K was calculated with a script implemented in Maestro.

NMR Studies. NMR spectra were obtained with a Bruker AC 400 MHZ spectrometer and the chemical shifts are reported in parts per million (ppm). Spectroscopic analyses were performed using the following conditions:

- Solvent system: THF-d8 and THF-d8/DMSO-d6 (3.5:1, v / v) were found to be the best solution at the different ranges of temperature employed for the analysis of cholic acid and sodium cholate, respectively.

- Concentration: the bile acid concentration was fixed below 0.5 M to avoid the formation of micellar aggregates and head-to-tail interactions. Each experiment was thus conducted dissolving 5 mg of compound in 0.9 mL of THF-d8 or of a mixture THF-d8/DMSO-d6 (3.5:1, v/v), yielding a concentration of 0.014M.

- Temperature: the experiments were performed at different temperatures (300 K → 240 K). Noteworthy, NMR experiments at 300 K were instrumental to assign the correct frequency for

protons and carbons of CA and its sodium salt, and to compare them with those previously reported in the literature.[35]

Survey of bioactive conformations of CA. Experimentally determined atomic coordinates of CA were taken from Ligand Expo,[32] a database linked to the PDB which provides atomic coordinates of ligands found in complexes deposited in the main PDB. We then assessed the proximity of the bioactive conformations to the 19 minima conformations found in the side chain of CA (Table S1), calculating the RMSD on C.sp3, C.sp2 and O.sp3 atoms of the molecule.

Supporting Information. Table S1 reports the distance matrix of 19 minima conformations of CA. Table S2 contains ^{13}C and ^{1}H resonance assignments for CA and its sodium salt at 300 K. Table S3 contains RMSD values (Å) of bioactive conformations of CA from minima conformations. Figure S1 shows NOE effect between the methyl at C_{21} position and the protons located at C_{23}Ha and C_{23}Hb at 300 K. Figure S2 shows row associated with the selective irradiation of the proton at $C_{12}\beta$ position.

References

1 M. J. Monte, J. J. Marin, A. Antelo and J. Vazquez-Tato, *World J. Gastroenterol.*, 2009, **15**, 804-816.

2 P. Lefebvre, B. Cariou, F. Lien, F. Kuipers and B. Staels, *Physiol. Rev.*, 2009, **89**, 147-191.

3 P. B. Hylemon, H. Zhou, W. M. Pandak, S. Ren, G. Gil and P. Dent, *J. Lipid Res.*, 2009, **50**, 1509-1520.

4 S. M. Houten, M. Watanabe and J. Auwerx, *EMBO J.*, 2006, **25**, 1419-1425.

5 A. F. Hofmann, L. R. Hagey and M. D. Krasowski, *J. Lipid. Res.*, 2010, **51**, 226-246.

6 A. F. Hofmann and A. Roda, *J. Lipid Res.*, 1984, **25**, 1477-1489.

7 P. R. Brotherhood and A. P. Davis, *Chem. Soc. Rev.*, 2010, **39**, 3633-3647.

8 A. P. Davis, *Molecules*, 2007, **12**, 2106-2122.

9 A. P. Davis and J. J. Walsh, *Chem. Commun.*, 1996, 449-451.

10 P. A. Brady, R. P. Bonar-Law, S. J. Rowan, C. J. Suckling and J. K. M. Sanders, *Chem. Commun.*, 1996, 319–320.

11 H. Sato, A. Macchiarulo, C. Thomas, A. Gioiello, M. Une, A. F. Hofmann, R. Saladin, K. Schoonjans, R. Pellicciari and J. Auwerx, *J. Med. Chem.*, 2008, **51**, 1831-1841.

12 R. Pellicciari, A. Gioiello, A. Macchiarulo, C. Thomas, E. Rosatelli, B. Natalini, R. Sardella, M., Pruzanski, A. Roda, E. Pastorini, K. Schoonjans and J. Auwerx, *J. Med. Chem.*, 2009, **52**, 7958-7961.

13 M. Watanabe, S. M. Houten, C. Mataki, M. A. Christoffolete, B. W. Kim, H. Sato, N. Messaddeq, J. W. Harney, O. Ezaki, T. Kodama, K. Schoonjans, A. C. Bianco and J. Auwerx, *Nature*, 2006, **439**, 484-489.

14 K. B. Islam, S. Fukiya, M. Hagio, N. Fujii, S. Ishizuka, T. Ooka, Y. Ogura, T. Hayashi and A. Yokota, *Gastroenterology*, 2011, **141**, 1773-1781.

15 A. Yokota, S. Fukiya, K. B. Islam, T. Ooka, Y. Ogura, T. Hayashi, M. Hagio and S. Ishizuka, *Gut Microbes*, 2012, **3**, 455-459.

16 K. D. Setchell, J. E. Heubi, *J. Pediatr. Gastroenterol. Nutr.*, 2006, **43**, S17-22.

17 *www.ema.europa.eu*

18 R. G. Smock and L. M. Gierasch, *Science*, 2009, **324**, 198-203.

19 M. C. Nicklaus, S. Wang, J. S. Driscoll and G. W. Milne, *Bioorg. Med. Chem.*, 1995, **3**, 411-428.

20 M. Vieth, J. D. Hirst and C. L. Brooks, 3rd, *J. Comput. Aided Mol. Des.*, 1998, **12**, 563-572.

21 E. Perola and P. S. Charifson, *J. Med. Chem.*, 2004, **47**, 2499-2510.

22 K. T. Butler, F. J. Luque and X. Barril, *J. Comput. Chem.*, 2009, **30**, 601-610.

23 M. Sitzmann, I. E. Weidlich, I. V. Filippov, C. Liao, M. L. Peach, W. D. Ihlenfeldt, R. G. Karki, Y. V. Borodina, R. E. Cachau and M. C. Nicklaus, *J. Chem. Inf. Model.*, 2012, **52**, 739-756.

24 Q. Wang, Y. P. Pang, PLoS One., 2007, **2**, e820.

25 E. Perola, P. S. Charifson, *J. Med. Chem.*, 2004, **47**, 2499-510.

26 W. L. Jorgensen, *Science*, 1991, **254**, 954-955.

27 A. Roda, B. Grigolo, R. Aldini, P. Simoni, R. Pellicciari, B. Natalini and R. Balducci, *J. Lipid. Res.*, 1987, **28**, 1384-1397.

28 R. Pellicciari, B. Natalini, S. Cecchetti, B. Porter, A. Roda, B. Grigolo and R. Balducci, *J. Med. Chem.*, 1988, **31**, 730-736.

29 R. Pellicciari, S. Cecchetti, B. Natalini, A. Roda, B. Grigolo and A. Fini, *J. Med. Chem.*, 1985, **28**, 239-242.

30 R. Pellicciari, S. Cecchetti, B. Natalini, A. Roda, B. Grigolo and A. Fini, *J. Med. Chem.*, 1984, **27**, 746-749.

31 Y. Zhao, H. Cho, L. Widanapathirana and S. Zhang, *Acc. Chem. Res.*, 2013.

32 Z. Feng, L. Chen, H. Maddula, O. Akcan, R. Oughtred, H. M. Berman and J. Westbrook, *Bioinformatics*, 2004, **20**, 2153-2155.

33 J. M. Goodman and W. C. Still, *J. Comp. Chem.*, 1991, **12**, 1110-1117.

34 G. Fogarasi, X. Zhou, P. W. Taylor and P. Pulay, *J. Am. Chem. Soc.*, 1996, **114**, 8191-8201.

35 D. V. Waterhous, S. Barnes and D. D. Muccio, *J. Lipid Res.*, 1985, **26**, 1068-1078.

36 A. M. Seldes, M. E. Deluca and E. G. Gros, *Magn. Res. Chem.*, 1986, **24**, 185-190.

37 Y. Murata, G. Suglhara, K. Fukushlma and M. Tanaka, *J. Phys. Chem.*, 1982, **86**, 4690-4695.

38 N. Funasaki, M. Fukuba, T. Hattori, S. Ishikawa, T. Okuno and S. Hirota, *Chem. Phys. Lipids*, 2006, **142**, 43-57.

39 N. Funasaki, M. Fukuba, T. Kitagawa, M. Nomura, S. Ishikawa, S. Hirota and S. Neya, *J. Phys. Chem. B*, 2004, **108**, 438-443.

40 H.W. Adolph, P. Zwart, R. Meijers, I. Hubatsch, M. Kiefer, V. Lamzin, E. Cedergren-Zeppezauer. *Biochemistry*, 2000, **39**, 12885-12897.

41 S. Bencharit, C.C. Edwards, C.L. Morton, E.L. Howard-Williams, P. Kuhn, P.M. Potter, M.R. Redinbo. J. Mol. Biol., 2006, **363**, 201-214.

42 Y.H. Pan, B.J. Bahnson. *J. Mol. Biol.*, 2007, **369**, 439-450.

43 S. Capaldi, G. Saccomani, D. Fessas, M. Signorelli, M. Perduca, H.L. Monaco. *J. Mol. Biol.*, 2009, **385**, 99-116.

44 S. Capaldi, M. Guariento, M. Perduca, S.M. Di Pietro, J.A. Santomé, H.L. Monaco. *Proteins*. 2006, **64**, 79-88.

45 S. Capaldi, M. Guariento, G. Saccomani, D. Fessas, M. Perduca, H.L. Monaco. *J. Biol. Chem.* 2007, **282**, 31008-31018.

46 D. Nichesola, M. Perduca, S. Capaldi, M.E. Carrizo, P.G. Righetti, H.L. Monaco. *Biochemistry*. 2004, **43**,14072-14079.

47 A.K. Malde, A.E. Mark, *J. Comput. Aided Mol. Des.* 2011, **25**, 1−12.

48 P.W. Swaan, K.M. Hillgren, F.C. Jr Szoka, S. Oie. *Bioconjug. Chem.* 1997, **8**, 520-525

49 B. Over, P. McCarren, P. Artursson, M. Foley, F. Giordanetto, G. Grönberg, C. Hilgendorf, M.D. Lee, P. Matsson, G. Muncipinto, M. Pellisson, M.W. Perry, R. Svensson, J.R. Duvall, J. Kihlberg. *J. Med. Chem.* 2014, DOI: 10.1021/jm500059t.

50 A. Fini, A. Roda, *J. Lipid. Res.*, 1987, **28**, 755-759

51 J. De Leeuw, Applications of Convex Analysis to Multidimensional Scaling. In *Recent Developments in Statistics.*; Barra, J. R. a. o., Ed. North Holland Publishing Company: Amsterdam, The Netherlands, **1977**; pp 133-146.