Understanding the Conformational Properties of Fluorinated Polypeptides: Molecular Modelling of Unguisin A

Natalia Díaz and Dimas Suárez*

Departamento de Química Física y Analítica. Universidad de Oviedo. Avda. Julián Clavería 8, 33006 Oviedo (Asturias). Spain.

diazfnatalia@uniovi.es

ABSTRACT

In this work, we investigate the conformational properties of unguisin A, a natural macrocyclic heptapeptide that incorporates a γ-aminobutyric acid (Gaba), and of four of its difluorinated stereoisomers at the Gaba residue. According to nuclear magnetic resonance (NMR) experiments, their secondary structure depends dramatically on the stereochemistry of the fluorinated carbon atoms. However, many molecular details of the structure and flexibility of these systems remain unknown so that a rationale of the conformational changes induced by the fluorine atoms in the macrocycle is still missing. To fill this gap, we apply enhanced molecular dynamics (MD) techniques to explore the peptide conformational space in dimethyl sulfoxide solution followed by 4-8 μs of conventional MD simulations that provide extensive equilibrium sampling. The simulations, which compare reasonably well with the NMR-based observations, show that the secondary structure of the macrocycle is altered substantially upon fluorination, except for the (*S*,*S*) diastereomer. It also turns out that the conformations of the fluorinated peptides are visited during the enhanced MD simulation of natural unguisin A, suggesting thus that conformations accessible to the unsubstituted macrocyclic peptide may be selected by fluorination. Therefore, the computational characterization of the macrocyclic peptides could be helpful in the rational design of stereoselective fluorinated peptides with fine-tuned conformation and activity.

INTRODUCTION

Peptides in drug design offer new opportunities for therapeutic intervention.[1] Besides presenting a low toxicity and high affinity, they are well suited for binding to relatively large and shallow protein surfaces like those involved in protein-protein interactions, thus widening the accessible targets for clinical applications.[2, 3] But peptides are susceptible to proteolytic degradation, present a low membrane permeability and unfavorable oral adsorption profiles. These weaknesses can be effectively overcome by using different strategies.[4] Particularly, macrocyclization is a well-established approach to increase peptide stability against proteases and, simultaneously, improve their transfer across membranes.[5] Numerous bioactive macrocyclic peptides with antibacterial, anticancer, antifungal, antiviral, and anti-inflammatory properties have been discovered.[3]

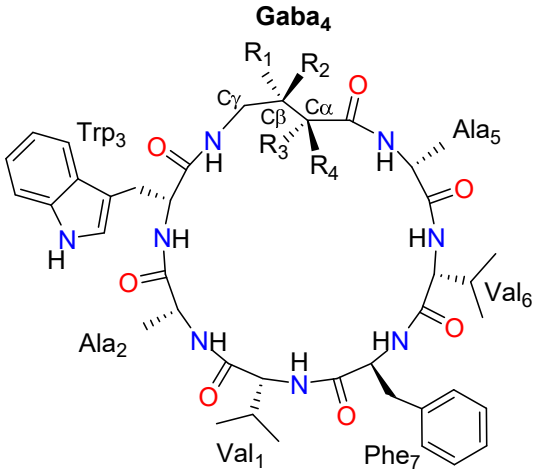
Drug molecules are expected to have well-defined bio-active conformations to maximize favorable binding enthalpy and minimize binding entropy penalty. In this respect, macrocyclization restraints the conformational space accessible to linear peptides and pre-organizes functional groups, improving target engagement due to stronger and more selective interactions.[6] However, macrocycles should also retain some conformational flexibility (*e.g.* by transiently forming intramolecular hydrogen bonds), which improves aqueous solubility, cell permeability, and oral absorption profiles.[5, 7-9] Thus, to optimize the biological activity of peptide macrocycles, it is important to understand the structural factors that stabilize their conformations and control the rate of interconversion among them. Unfortunately, conformational analysis in macrocycles is still a challenging problem due, in part, to a strong interplay between bond rotations and intramolecular interactions. In addition, numerous examples show that relatively

small structural modifications in a macrocycle result in local conformational changes that propagate along the ring to affect distal structural features.[10]

Another strategy adopted in the design of bioactive molecules with well-defined conformations consists of the selective fluorination of C-H bonds. For example, some of the best performing, top-selling drugs on the market, contain fluorine atoms in their structures.[11]. The inclusion of fluorine atoms improves the metabolic stability and increases the lipophilicity of lead molecules. In addition, fluorine modifies the pK_a values of adjacent functional groups and plays an important and unique role in influencing molecular conformation.[12] The conformational effects exerted by fluorine atoms stem from both electrostatic and stereoelectronic effects, like the so-called *gauche* effect, which is defined as the preferential stabilization of a *gauche* arrangement in a two carbon unit with F bonded vicinally to electronegative elements (*i.e.* F-C-C-X angles around ± 60º).[13, 14] However, predicting the impact of fluorine in the conformation and other properties of complex molecules remains also as a challenging problem.[15, 16]

In this work, we analyze the conformational effects associated to sterically conservative H→F substitutions in unguisin A, a heptapeptide macrocycle isolated from the marine fungus *Emericilla unguis* that incorporates a quite flexible γ-aminobutyric acid (Gaba) residue within the ring.[17, 18] Nuclear magnetic resonance (NMR) experiments performed for unguisin A and its four diastereomers fluorinated at the Cα and Cβ backbone atoms of the Gaba residue (see **3**-**7** in Scheme 1) have confirmed that the presence of the fluorine atoms modifies the secondary structure and dynamical properties of the various difluorinated macrocycles. The NMR coupling constants ($^3J_{H,H}$ and $^3J_{H,F}$) reported for the Gaba residue are consistent with a *gauche* F-Cβ-Cα-F arrangement, with compounds **4**, **6**, and **7** in Scheme 1 presenting some degree of interconversion between the two *gauche* forms (*i.e.* g+/g- disorder).[19, 20] In contrast, the coupling

between the α-fluorine atom and the adjacent NH proton ($^4J_{H,F}$) has not been observed for **4**-**7**, suggesting that expected *anti*-parallel C-F/C=O alignment is lacking. Taking into account that, in simple models, the energetic preference for the *gauche* F-C-C-F arrangement is substantially lower than that of the *anti*-parallel F-C-C=O one (*e.g.*, 0.8 kcal/mol in 1,2-difluoroethane *vs* 5.6 kcal/mol in 2-fluoro-*N*-methylacetamide),[21] it turns out that the conformational influence of fluorine atoms in unguisin A is poorly understood.



Gaba$_4$ substitutions

R_1	R_2	R_3	R_4	
H	H	H	H	unguisin A (**3**)
H	F	H	F	syn-(R,R)-difluoro (**4**)
F	H	F	H	syn-(S,S)-difluoro (**5**)
H	F	F	H	anti-(R,S)-difluoro (**6**)
F	H	H	F	anti-(S,R)-difluoro (**7**)

Scheme 1

Herein, we characterize in detail the molecular structure and dynamics of unguisin A and its difluorinated derivatives by means of extensive molecular dynamics (MD) simulations. More specifically, we take advantage of the large conformational sampling and various MD tools analysis to help disentangle the various factors that determine the conformations adopted by the cyclic polypeptides (solute entropy, *gauche* effect, intramolecular interactions, solvation, …). To this end, we derive first new molecular mechanics (MM) parameters for the Gaba residue from high level ab initio calculations. Then we resort to enhanced sampling MD methodologies, which have been shown as a reliable tool to gather a diverse and thermodynamically meaningful

conformational ensembles for cyclic peptides.[22-24] In particular, we explore the phase spaces of unguisin A and its four difluorinated **4**-**7** variants by carrying out Gaussian-accelerated molecular dynamics (GaMD) simulations (2.0 μs) followed by extended conventional molecular dynamics (cMD) simulations (4.0/8.0 μs). To validate the significance of the simulations, we carefully address our computational results about intramolecular H-bonding and NMR coupling constants with the available experimental data. To assess in detail the effects associated to fluorine substitution, we perform comparative analysis of the cMD trajectories complemented with entropy and molecular mechanics calculations. Overall, our results not only provide further molecular details about the structure and dynamics of the unguisin A macrocycle, but also suggest that computational modeling may yield useful guidelines for the *de novo* computational design of fluorinated polypeptides with targeted conformational properties.

METHODS

Set up of the systems

Initial coordinates for unguisin A (**3** in scheme 1) were obtained from the ChemSpider chemical structure database (identifier 8828057).[25] After reordering and grouping the atoms according to the Val_1-Ala_2-Trp_3-$Gaba_4$-Ala_5-Val_6-Phe_7 sequence, the structure was relaxed with the *Avogadro* program[26] to alleviate some structural problems. Subsequently, four difluorinated variants of unguisin A were obtained by replacing the appropriate H atoms in the Gaba residue of **3** by F atoms. This resulted in molecule **4** with an (*R, R*) stereochemistry at the Cβ and Cα atoms of Gaba, the alternative *syn* structure **5** with an (*S, S*) stereochemistry, and the two possible *anti* configurations (*R, S*) in molecule **6** and (*S, R*) in molecule **7**.

Standard residues in the **3-7** compounds (*i.e.*, Val, Ala, Trp and Phe) were described with the ff14SB force field.[27] For the non-standard Gaba residue and its fluorinated variants in molecules **4-7**, appropriate atom types were selected from the ff14SB force field (see Table S1). Atomic charges were computed using the Restrained Electrostatic Potential (RESP) methodology.[28, 29] To this end, the various -NH-CH_2-$(CHX)_2$-CO- Gaba residues were capped with terminal acetyl (Ace) and *N*-methylamine (Nme) groups. Then 50 conformers were generated for each molecule with the *Ms-dock* program.[30] Structural and energetic filters were applied to remove high energy and redundant conformers. For each Ace-Gaba-Nme molecule, the selected conformers were optimized at the HF/6-31G* level of theory using the *Gaussian03* package.[31] The HF/6-31G* electrostatic potential of all the conformers were processed with the *Antechamber* software[32] in order to derive the RESP atomic charges for the five different Gaba residues in **3-7**.

Taking into account that the placement of a fluorine atom adjacent to a carbonyl group has a remarkable conformational influence,[21] we refined the dihedral parameters for rotation around the Cα-C bond in Gaba (see Scheme 1) using the *Paramfit* program included in the *Amber16* package.[33] The Gaba conformers utilized for the RESP calculations were reoptimized at the RI-MP2/aug-cc-pVTZ level of theory with the *Orca* program.[34] The 4-5 most stable conformers were then perturbed by rigid rotation around the Cα-C bond in 30º steps. After removing the geometries with steric clashes, the quantum mechanical (QM) energies of the remaining structures were evaluated by means of single-point RI-MP2/aug-cc-pVTZ calculations, and the energies and geometries were introduced into *Paramfit* to optimize the Cβ-Cα-C-N and F-Cα-C-N terms by applying a least-squares fitting procedure. To include the refined dihedral parameters without affecting the standard CT-CT-C-N torsion, a new atom type termed CG was defined and

assigned to the Cβ of Gaba (see Scheme 1). All the refined parameters are collected in Table S1. The Amber *prep* and *frcmod* files are also included in the Supplementary Material.

Previous NMR experiments were performed for unguisin A and its four difluorinated derivatives in dimethyl sulfoxide (DMSO).[20] For this reason, the initial structures of **3**-**7** were surrounded by an octahedral box of DMSO molecules that extended 16 Å from the solute atoms using the *tLeap* program in *Amber16*.[35] For the DMSO solvent, the parameters by Fox and Kollman[36] and the box available in the *Amber* parameter database were loaded into *tLeap*.

Molecular Dynamics simulations

Energy minimization and molecular dynamics (MD) simulations were carried out using the *Sander* and *Pmemd* programs included in the *Amber16* suite of programs.[33, 35] The DMSO molecules were initially relaxed by means of energy minimizations and 200 *ps* of MD. Then the full systems were minimized and heated gradually to 300 *K* carrying out 60 *ps* of constant volume (NVT) MD with a 1 fs time step. Subsequently, the density was adjusted by means of 2.0 ns of constant pressure (NPT) MD with a 2 fs time step and using the Monte Carlo barostat as implemented in *Pmemd*. Langevin dynamics was employed to control the temperature (300 *K*) with a collision frequency of 2 ps^{-1}. The SHAKE algorithm[37] was employed to constraint all the R-H bonds, and periodic boundary conditions were applied to simulate a continuous system. A non-bonded cutoff of 10.0 Å was used whereas the Particle-Mesh-Ewald method[38] was employed to include the contributions of long-range interactions.

The conformational space of the unguisin **3**-**7** models was initially explored by Gaussian Accelerated Molecular Dynamics (GaMD),[39] which performs an enhanced and unconstrained sampling of the systems thanks to harmonic boost potentials that smooth out the potential energy

surface. We applied a dual potential boost on the torsion and potential energetic terms. The upper limits of the standard deviation of the total and dihedral boost potentials were set to their default values (6.0 *kcal/mol*). The parameters that define the boost potentials, a threshold energy *E* and a harmonic force constant *k*, were estimated in preliminary MD and GaMD runs from the average and standard deviation of potential energies. The number of simulation steps (NTAVE) used to update the potential energy statistics was 50000. The preliminary MD run comprised a total of 18xNTAVE simulation steps, yielding an initial guess of the *E* and *k* parameters. After an equilibration GaMD phase (2xNTAVE steps), *E* and *k* were refined during a GaMD run with a length of 10x18xNTAVE steps. Finally, the production phase of the GaMD NTP simulations performed for **3**-**7** was extended up to 2 μs. Cartesian coordinates and the values of the boost potentials were saved for analysis every 1250 simulation steps (2.5 *ps*). The graphics-processing-unit (GPU) accelerated version of the *Pmemd* code[40] was employed both in the conventional MD and in the GaMD runs.

To characterize the shape and conformation of **3**-**7** along the production GaMD simulations, we monitored the evolution of two dihedral angles, Val_1@Cα-Ala_2@Cα-Trp_3@Cα-$Gaba_4$@Cγ (RC1) and $Gaba_4$@Cα-Ala_5@Cα-Val_6@Cα-Phe_7@Cα (RC2). These two angles were used to reweight the GaMD simulations and estimate 2D free energy maps for **3**-**7**. We use the python scripts developed by Miao *et al.*, selecting the exponential average reweighting and 60 bins along each RC coordinate.[41]

Each GaMD simulation revealed 2 or 3 highly-populated basins in the 2D free energy maps. To further sample the equilibrium properties of the conformational GaMD basins, we performed conventional MD (cMD) simulations that were started by randomly selecting a structure within the corresponding basins. After minimization, thermalization and pressurization of the selected

structures, the cMD simulations extended up to 4 or 8 μs depending on the flexibility displayed by the macrocycle. More specifically, two 8 μs simulations named **3**$_1$ and **3**$_2$ were run for compound **3**, two 4 μs simulations (**4**$_1$ and **4**$_2$) for **4**, two 4 μs simulations (**5**$_1$ and **5**$_2$) for **5**, one 4 μs (**6**$_1$) and one 8 μs (**6**$_2$) simulation for **6** and, finally, three 8 μs simulations (**7**$_1$, **7**$_2$, and **7**$_3$) were run for **7**. This combination of multiple MD trajectories performed for long simulation times should provide a good sampling of the conformational space accessible to these relatively small and rigid macrocycle peptides.[42]

Analysis of the simulations

Structural analyses of the simulations were performed using the *Cpptraj* module of *Amber16*,[43] and other specific software developed locally. The coordinates of **3**-**7** along the MD trajectories were clustered using *Cpptraj* with the average-linkage clustering algorithm and a sieve of 50 frames. The distance metric between frames was calculated *via* best-fit coordinate root-mean-square deviation (RMSD) using all the backbone heavy atoms, and the clustering was finished when the minimum distance between clusters was greater than 1.0 Å.

Intramolecular H-bonds were characterized along the cMD simulation of **3**-**7**. Cutoffs of 3.5 Å in the N···O distance and 110º in the NH···O anglewere defined as the geometric criteria for the occurrence of a H-bond.[44] H-bonds between the backbone amido groups of **3**-**7** and solvent molecules were also counted considering the same geometric criteria. Spin-spin coupling constants were estimated for the Gaba residue using Karplus-type relationships. For the $^3J_{H,H}$ values, we use a Karplus-type relationship that expressed the coupling constant as a function of the H-C-C-H torsion angle. It includes three adjusted parameters with values C_0=6.01 Hz, C_1=-1.37 Hz, and C_2=3.61 Hz.[45] For the $^3J_{H,F}$ constants, we implemented the seven-parameter function proposed by Thibaudeau et al.[46] Besides the H-C-C-F torsion angle, this Karplus-type

function makes use of correction terms for the substituent electronegativity as well as for H-C-C and F-C-C bond angles. We selected λ_1=1.4, λ_2=0.9, λ_3=0.9, and λ_4=0.0 for the substituent electronegativities of the fluorine, carbon, and hydrogen atoms present in the CHF-CHF moiety of the Gaba residue in **4**-**7**. Torsion angles were computed along the cMD simulations with *Cpptraj*. The Probability Density Functions (PDFs) of the different torsions were obtained from a histogram representation and a von Mises kernel estimator. For a particular torsion angle, the location of the maxima and minima of the corresponding PDF allows to split the sampled configurational space in different conformational states that interchange along the dynamic simulation.[47] The Chimera visualization system was employed to compute molecular areas and draw molecular models.[48]

MM-PBSA and Entropy Calculations

Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) calculations[49] were carried out on 1500/3000 equally-spaced snapshots along the 4/8 μs cMD trajectories. The MM-PBSA energy of the solute atoms was computed as:

$$G_{MMPBSA} = E_{MM} + \Delta G_{solv}^{PB} + \Delta G_{solv}^{non-polar} \quad (1)$$

where E_{MM} is the molecular mechanics energy of the solute molecule including the $3RT$ contribution due to six translational and rotational degrees of freedom, ΔG_{solv}^{PB} is the electrostatic solvation energy obtained from Poisson-Boltzmann calculations,[50] and $\Delta G_{solv}^{non-polar}$ is the non-polar part of solvation energy due to cavity formation and dispersion interactions between the solute and the solvent molecules.

The E_{MM} term in eq. (1) was calculated with no cutoff. The ΔG_{solv}^{PB} term was obtained by solving the non-linear Poisson-Boltzmann (PB) equation[51] on a cubic lattice with a grid spacing of 0.33 Å and using an iterative finite-difference method. The solute is represented by the atomic charges and radii taken from the *Amber*/ff14SB representation (ε_{in}=1). The DMSO solvent was represented with ε_{out}=46.1 and the solvent accessible surface was constructed considering a solvent probe radius of 2.1 Å. The non-polar solvation $\Delta G_{solv}^{non-polar}$ was estimated using the approximation introduced by Gohlke and Case,[52]

$$\Delta G_{solv}^{non-polar} = \Delta H_{solute-solvent}^{vdW} + \gamma MSA$$

where $\Delta H_{solute-solvent}^{vdW}$ is the van der Waals interaction energy between solute and the DMSO molecules within a shell of 15 Å thickness, while the cavitation free energy contribution to the non-polar solvation energy is determined by a molecular surface area (MSA) dependent term. For the DMSO solvent, γ was determined by applying $\gamma_o\left(\gamma_{DMSO}/\gamma_{wat}\right)$, where γ_o is the original surface tension parameter (0.069 kcal / mol Å^2) proposed for aqueous solutions.[52] All these energy components were evaluated with the *Sander* and *Pbsa* programs included in the *Amber16* suite.

Assuming that the potential energy surface of the solute can be described as a collection of harmonic wells, the absolute entropy can be estimated as:

$$S \approx \left\langle S_{RRHO} \right\rangle + S_{conform}$$

where $\left\langle S_{RRHO} \right\rangle$ is the average entropy derived from statistical mechanics within the rigid-rotor harmonic-oscillator approximations (RRHO), and $S_{conform}$ is the conformational contribution to the total entropy that arises from the population of the different energy wells.[53]

The RRHO entropy calculations were performed on the 1500/3000 MD snapshots extracted from the cMD trajectories which were post-processed through the removal of all solvent molecules. Prior to the normal mode calculations yielding the harmonic frequencies, the structures were subject to geometry optimization setting a tight tolerance in the energy gradient and using the Hawkins, Cramer, and Truhlar generalized Born (HCT-GB) implicit solvent model for removing the explicit consideration of solvent degrees of freedom. The geometries of the systems were minimized in the implicit solvent until the RMSD in the gradient elements was below 10^{-6} kcal mol^{-1} $Å^{-1}$. These GBSA minimizations were carried out via the *Sander* program while the second derivatives of the MM and GBSA energies included in the Hessian matrix were calculated analytically with the *Nab* program.[54]

The instantaneous values of the MM-PBSA components and S_{RRHO} entropies oscillate on the sub-nanosecond time scales and exhibit lower-amplitude oscillations on longer time scales. The statistical uncertainty of their average values was determined by computing the block-averaged standard error of the mean.[55] Thus, the MD frames were divided into segments ("blocks") with a block size *M* that ranges from 1 up to a quarter of the total number of frames and the limiting value of the block-averaged error is reported as an upper limit to the statistical uncertainty.

We estimated the conformational entropies (S_{conform}) of the macrocycles using the *Cencalc* program.[56] For each rotatable dihedral angle of the heptapeptide molecules, which are selected using the *Amber* parameter file, *Cencalc* discretizes its time evolution by evaluating the continuous PDFs represented by a von Mises kernel density estimator, which depends on a concentration parameter κ (a κ=0.50 value was chosen here). By finding the maxima and minima of the PDF, the time series containing the values of the corresponding dihedral angle during the MD simulation is transformed into an array of integer numbers labelling the accessible

conformational states along the MD trajectory. Then, *Cencalc* calculates the unidimensional probability mass functions p_i of each torsion angle and computes the corresponding first-order $S_{conform}$ contribution as $S_{conform} = k_B \sum_i -p_i \ln p_i$. Correlation entropy among the torsional degrees of freedom was taken into account using a reformulated expansion technique termed correlation-consistent multibody local approximation (cc-MLA),[57] which is effectively independent of any expansion order and can be formally expressed as a sum of conditional entropies that closely resembles the so-called Maximum Information Spanning Tree (MIST) approach.[58] The bias of the cc-MLA entropy due to finite sampling was minimized by shuffling the elements of the arrays of integer numbers labelling the conformational states prior to the entropy estimations.

RESULTS

Preliminary GaMD sampling

There is no experimental structure available for unguisin A or for any of its **4**-**7** difluoroderivatives. In addition, it has been noted that, although macrocycles are conformationally-restricted molecules, they can still be flexible and exist as ensembles of low energy conformations connected by relatively high-energy barriers.[10] For these reasons, the GaMD enhanced sampling technique was selected to initially explore the conformational space of **3**-**7** in DMSO solution. From the GaMD trajectories, we obtained 2D free energy maps by estimating the canonical probability distribution using the Val_1@Cα–Ala_2@Cα–Trp_3@Cα–$Gaba_4$@Cγ (RC1) and $Gaba_4$@Cα–Ala_5@Cα–Val_6@Cα–Phe_7@Cα (RC2) torsions as reaction coordinates. The maps,

shown in Figure 1, exhibit well-defined minima, which in turn, delineate the most-likely macrocyclic conformations defined in terms of the selected backbone dihedral angles.

<Figure 1 here>

The comparison of the free energy maps suggests that the different fluorination stereochemistry at the Gaba residue of **4**-**7** modifies the conformational dynamics of the unguisin A ring. In terms of the global bending of the macrocycle described by the RC1 and RC2 coordinates, the difluorinated derivatives **6** and **4** present a less flexible macrocycle than that of the non-fluorinated compound **3** (see Figure 1). In contrast, **7** would be quite flexible as suggested by the presence of several low-energy basins throughout the RC1/RC2 phase space. Further insight into the conformations sampled by the GaMD **3**-**7** simulations can be obtained from clustering calculations based on the RMSD of the backbone heavy atoms (see Table S2). Considering an RMSD threshold of 1.0 Å, a large number of clusters is obtained for each trajectory, confirming thus that the GaMD simulations effectively explored multiple regions in the conformational landscape of the **3**-**7** macrocycles. Moreover, the total number of clusters and the percentage of abundance of the most populated ones differ across the GaMD simulations, which indicate a different dynamical behavior in **3**-**7**. The clustering results also point to **7** as the most flexible difluorinated variant of unguisin A.

Basing on the free energy maps shown in Figure 1, we carried out conventional MD simulations of **3**-**7** aimed to sample the equilibrium states of the unguisin macrocycle. Thus, one GaMD snapshot was randomly selected from each of the two lowest free energy basins of **3**-**6** (see Figure 1) to start the $\mathbf{3_1}$/$\mathbf{3_2}$, $\mathbf{4_1}$/$\mathbf{4_2}$, $\mathbf{5_1}$/$\mathbf{5_2}$, and $\mathbf{6_1}$/$\mathbf{6_2}$ cMD simulations. For the highly flexible compound **7**, we selected three different basins to randomly extract the initial point for the $\mathbf{7_1}$, $\mathbf{7_2}$, and $\mathbf{7_3}$ cMD simulations. Visual inspection of the initial structures, which are shown in Figure

S1, confirm that they correspond to different arrangements of the unguisin macrocycle. Even for the structures with similar RC1/RC2 values (see for instance structures **3**$_1$, **6**$_2$, and **7**$_3$ in Figure S1), the specific orientation of some peptide bonds results in different intra-molecular H-bonds.

cMD simulations: Evolution of the trajectories

To analyze the conformational space preferentially sampled by **3**-**7** during the cMD simulations, we built again 2D histogram representations of the RC1/RC2 torsion angles of the macrocycle (see Figure 1). Overall, the comparison between the GaMD and cMD histograms points out that the main GaMD basins were exhaustively sampled during the cMD simulations, allowing thus to characterize their equilibrium properties. A closer inspection of these plots shows that some of the cMD simulations (**4**$_2$, **5**$_1$, and **5**$_2$) maintain their RC1/RC2 coordinates around those of the starting GaMD structures, revealing that their secondary structures are quite stable and/or that the energy barriers for the conformational interconversion are relatively high, perhaps due to the restraints imposed by the macrocycle. In contrast, the initial structures in the **3**$_1$, **3**$_2$, **6**$_2$, **7**$_1$, **7**$_2$, and **7**$_3$ cMD simulations evolved to sample other regions in the RC1/RC2 phase space, and they populate different conformational basins in the same cMD trajectory. It is also interesting to note that several compounds preferentially sample the same conformations in terms of the RC1/RC2 coordinates. For instance, RC1 and RC2 values around 100º are highly populated along the **3**$_1$, **5**$_1$, **7**$_1$, and **7**$_3$ simulations, that is, the fluorinated **5** and **7** derivatives populate macrocycle conformations characteristic of the native unguisin A.

<Figure 2 here>

The time evolution of the macrocycle conformations along the cMD simulations was monitored in terms of their backbone RMSD values. The length of simulations (4.0 μs or 8.0 μs)

was decided depending on the persistence (or not) of specific conformers along the first 4.0 μs. Figure 2 presents the final RMSD plots, colored for the most abundant clusters as obtained by accumulating all the sampling obtained for each compound. For the non-fluorinated molecule **3**, the RMSD plots reveal that the initial configurations extracted from the GaMD calculations can interconvert on the μs scale with configurations that are structurally quite different (*i.e.* with RMSD > 2 Å) and that the $\mathbf{3_2}$ trajectory evolves after 3.3 μ towards the configurations appearing along the $\mathbf{3_1}$ simulation. Considering the two *syn*-difluorinated variants of unguisin A, **4** (*R,R*) and **5** (*S,S*), the structure of the macrocycles remained very close to the initial GaMD structures and the time evolution of the RMSD values was quite steady along the cMD simulations. There is a rapid interconversion on the *ns*-time scale between different clusters along $\mathbf{4_1}$ that, according to the RMSD values, correspond to structurally similar conformations of the backbone atoms. In contrast, only one major conformer is observed for the $\mathbf{4_2}$ (91%), $\mathbf{5_1}$ (99%), and $\mathbf{5_2}$ (89%) trajectories. The coloring pattern in Figure 2 further emphasizes that the main $\mathbf{4_2}$ conformer differs from the $\mathbf{4_1}$ ones. Similarly, the most abundant conformers of $\mathbf{5_1}$ and $\mathbf{5_2}$ are also structurally different (see below). Concerning the *anti*-difluorinated variants of unguisin A, compound **6** (*R,S*) seems to be more rigid. On one hand, a highly abundant conformation (93%) was observed during the 4.0 μs of the $\mathbf{6_1}$ simulation. On the other hand, the alternative $\mathbf{6_2}$ trajectory, which explored other conformations during the first 4.7 μs, displays again the $\mathbf{6_1}$ conformation during the last 3.3 μs. For the other *anti*-difluorinated derivative, compound **7**, three independent 8-μs long simulations were required to better characterize its phase space. According to the $\mathbf{7_1}$ RMSD plot, the cyclic peptide drastically changed its initial configuration during the first 1.4 μs of the cMD and, thereafter, it adopted a quite stable conformation that undergoes only small structural fluctuations (see Figure 2). A similar behavior is observed along

the **7₃** trajectory, with a first part of intense conformational change, while the last 4.0 μs mostly sampled the same type of configurations than in **7₁**. In contrast, trajectory **7₂** displayed frequent transits between conformations with low abundance (< 15%), which also appear in the first part of the **7₁** and **7₃** simulations. Curiously, the most abundant conformation in **7₁** and **7₃** (in red in Figure 2) is only marginally observed in **7₂** after 8.0 μs of simulation time.

cMD simulations: MM-PBSA and entropy calculations

To complement the structural characterization of the cMD simulations with an energetic assessment of their relative stabilities, we computed the MM-PBSA energies and the solute entropies as described in Methods. Table 1 collects the resulting *G* values obtained from the average MM-PBSA energies and the S^{RRHO} and $S_{conform}$ entropies, while Figures S3 and S4 display the time evolution of the MM-PBSA values and the $S_{conform}$ convergence plots, respectively.

<Table 1 here>

For the unfluorinated unguisin A, the *G* scorings for the whole trajectories indicate that the **3₁** ensemble is 1.9 kcal/mol more stable than **3₂**, which results from more favorable MM-PBSA and entropy contributions. In addition, the MM-PBSA plot along the first half of the **3₂** trajectory exhibits a pronounced lowering drift whereas the **3₁** MM-PBSA energy seems better equilibrated, fluctuating around the mean value. These observations suggest that the first half of the **3₂** simulation corresponds to a search phase throughout high energy conformations and, therefore, only the trajectory **3₁** was considered to characterize the equilibrium properties of the native unguisin A.

For the fluorinated derivatives **4** and **5**, the time evolution of their MM-PBSA values displays a steady evolution along the simulations and the conformational entropy plots are well-converged. In terms of the *G*-scoring difference (with statistical uncertainty in parentheses), $\mathbf{4_1}$ is only 0.3 (0.2) kcal/mol more stable than $\mathbf{4_2}$ while $\mathbf{5_2}$ is more stable than $\mathbf{5_1}$ by 0.7 (0.4) kcal/mol. These are quite small differences so that all the conformers sampled by these trajectories would be significantly populated at room temperature. Therefore, we accumulated the $\mathbf{4_1}$ and $\mathbf{4_2}$ trajectories, on one hand, and the $\mathbf{5_1}$ and $\mathbf{5_2}$ ones, on the other, in order to characterize the equilibrium properties of the **4** and **5** difluorinated compounds.

For the first *anti*-difluoro derivative, the ensemble of structures sampled by the $\mathbf{6_2}$ simulation turns out to be 2.3 kcal/mol more stable than that of $\mathbf{6_1}$ owing to the larger S_{conform} of $\mathbf{6_2}$, which presents several conformational changes in contrast with the rigidity of $\mathbf{6_1}$ (see Figure 2). As noticed above, the last 3.3 μs of $\mathbf{6_2}$ matches the conformation observed in $\mathbf{6_1}$. The MM-PBSA plot in Figure S3 and Table 1 indicate that this $\mathbf{6_1}$ conformation would be slightly stabilized by 1.5 kcal/mol in terms of the MM-PBSA energies, although this difference would be compensated by entropic contributions favoring the other $\mathbf{6_2}$ conformations. Therefore, we considered that the whole $\mathbf{6_2}$ trajectory would provide a more complete energetic and entropic description than $\mathbf{6_1}$.

The cMD simulations of compound **7** may deserve a more particular attention. The three trajectories $\mathbf{7_1}$, $\mathbf{7_2}$, and $\mathbf{7_3}$ present a similar stability in terms of the *G* scorings collected in Table 1, the corresponding differences being below 0.5 kcal/mol, close to their statistical uncertainties. The average MM-PBSA values and the time evolution plots show a slightly stabilization of the peptide macrocycle along the $\mathbf{7_1}$ and $\mathbf{7_3}$ simulations, but such stabilization is accompanied by a reduction in the conformational entropy occurring upon removal of the other conformations (1.5 μs for $\mathbf{7_1}$ and 4.0 μs for $\mathbf{7_3}$). Overall, the conformational entropy plots are reasonably converged

in the three trajectories. Consequently, the three simulations **7_1**, **7_2**, and **7_3** were considered to analyze the structure and dynamics of compound **7**, which turns out to exhibit a similar flexibility as that of the parent unguisin A peptide.

cMD simulations: Structural features of the most populated conformers

Figure 3 displays the cluster representatives of the backbone conformations adopted by the **3-7** systems. For the simulation **3_1**, which describes the most likely conformations of unguisin A, three major conformations (**3_{1A}**, **3_{1B}**, and **3_{1C}**) account for 41%, 21%, and 14% of the MD frames. It turns out that **3_{1C}** exhibits an open macrocycle ring structure whereas **3_{1A}** and **3_{1B}** adopt a more compact shape stabilized by intramolecular H-bonds involving backbone peptide groups. The most persistent H-bond interaction corresponds to Val_1@NH$\cdots$O-Ala_3 H that is observed in 60% of the whole **3_1** simulation. Other less abundant H-bonds link the Ala_2/Ala_5 and $Gaba_4$/Val_1 residues in the **3_{1A}** and **3_{1B}** clusters, respectively (see Figure 3 and Table S3).

<Figure 3 here>

As noticed above, the **4_1** simulation of the *syn-R,R* difluoro-peptide fluctuates between two clusters that are similarly populated 43% (**4_{1A}**) and 36% (**4_{1B}**). The second simulation, **4_2**, which is slightly less stable in terms of the MM-PBSA/entropy scoring, is dominated by a single conformation (91%). The three conformers are relatively compact and some intramolecular H-bonds are present in two of the three conformers (*e.g.*, Val_1@NH$\cdots$O-Ala_3 in **4_{1A}** and **4_{1B}**; $Gaba_4$@NH$\cdots$O-Val_1 in **4_{1A}** and **4_{2A}**; see Figure 3). However, given that the approximated free energy difference between the **4_1** and **4_2** ensembles is very small (0.4 kcal/mol), it is not feasible to highlight a particular H-bond interaction as being highly persistent in all the conformations adopted by **4**.

According to the GaMD/cMD simulations, the *syn-R,R* difluoro derivative can access to two different backbone conformations, **5$_{1A}$** that accounts for 99% of the **5$_1$** MD frames and **5$_{2A}$** that is also highly abundant (89%) in the other trajectory. Although the **5$_{1A}$** and **5$_{2A}$** representative structures differ in their backbone arrangement (RMSD =1.94 Å), the two conformers exhibit an extended Gaba residue with a *g-* conformation at the F-C-C-F moiety and the same Val$_1$@NH···O-Ala$_5$ intramolecular contact, which turns out to be highly abundant (~90%) in the two trajectories.

Basing on the MM-PBSA and entropy calculations, the equilibrium properties of the *anti-(R,S)* stereoisomer (**6**) are better described by the **6$_2$** simulation alone, which populates two major clusters, **6$_{2A}$** with 40% of the frames and **6$_{2B}$** with 33%. These clusters present a quite different pattern of H-bond contacts (see Figure 3). Thus, Val$_1$@NH···O-Ala$_5$ (32%) and Gaba$_4$@NH···O-Val$_1$ (30%) are the most relevant interactions in **6$_{2A}$**, while the Trp$_3$@NH···O-Phe$_7$ (32%) and Gaba$_4$@NH···O-Ala$_5$ (27%) contacts predominate in **6$_{2B}$**.

From the extended GaMD/cMD simulations, the *(S,R)* difluorination at the Gaba residue would result in a flexible macrocycle. For example, in terms of the total number of different clusters, **7** has more conformational variability than unguisin A (see Table S2). Considering just the most abundant cluster representatives shown in Figure 3, it turns out that Val$_1$@NH···O-Ala$_5$ (46%), Ala$_2$@NH···O-Ala$_5$ (67%), and Gaba$_4$@NH···F2-Gaba$_4$ (68%) are the most abundant contacts in **7$_{1A}$** (similar percentages are computed for **7$_{3A}$**). In **7$_{3B}$**, the Ala$_2$@NH···O-Val$_1$ (24%) and Val$_6$@NH···O-Ala$_2$ (25%) H-bond contacts are also relevant.

Some structural features of the computationally-predicted structures may be addressed against experimental data. More specifically, the temperature gradients of the ^{1}H-NMR chemical shifts ($\Delta\delta/\Delta T$) have been determined for the backbone NH groups of **3**-**7**.[20] Since the presence of stable

intramolecular H-bonds is usually ascribed to $\Delta\delta/\Delta T$ values below $3\cdot10^{-3}$ ppm/K, NH···O=C intramolecular contacts have been assigned for Val_1@NH in **3**, Ala_2@NH and Val_6@NH in **5**, Val_1@NH, $Gaba_4$@NH, and Val_6@NH in **6**, and Val_1@NH in **7**. However, a certain over-interpretation of chemical shift/temperature coefficients as arising exclusively from intramolecular hydrogen bonding effects may be possible, as noticed in the literature.[10, 59, 60] Nevertheless, there is a partial agreement between the intramolecular H-bonds characterized by the cMD simulations and the assigned NH···O=C intramolecular contacts in ref. 20. For example, the $\mathbf{3_1}$ simulation and the NMR $\Delta\delta/\Delta T$ values measured for **3** agree in pointing to Val_1@NH as the amido group preferentially involved in an intramolecular H-bond contact. On the other hand, the $\Delta\delta/\Delta T$ values obtained for **4** have been considered compatible with the lack of persistent intramolecular H-bonds, what might be explained in terms of a mixture of conformers of similar stability as suggested by our simulations and the absence of a common intramolecular H-bond contact among them. For the other difluoro derivatives **5**, **6**, and **7**, there are some coincidences between the identity of the interacting NH groups signaled-experimentally and the most populated H-bonds characterized *in silico* (*e.g.*, Ala_2@NH in **5**, Val_1@NH and $Gaba_4$@NH in **6**, and Val_1@NH in **7**). However, the overall abundances of the corresponding H-bonds during the MD simulations are below 50% (see Figure 3) and/or the accurate weight of the corresponding MD ensemble is not known. Therefore, a closer comparison between computational and experimental observations seems not feasible at this point.

Conformational properties of the Gaba residue

Further structural and dynamical details on the effects associated to the fluorine atoms can be obtained from the analysis of the backbone torsion angles involving the Gaba residue. Hence, we first analyzed whether or not the usual conformational preferences associated to the presence of fluorine atoms operate in **4**-**7**. The PDF plots and the time evolution of the conformational states sampled by the N-Cγ-Cβ-F, F-Cβ-Cα-F and F-Cα-C=O torsion angles of the Gaba residue are collected in the Supporting Information (Figure S5).

Considering the N-Cγ-Cβ-F torsion, the expected *gauche* arrangement is the most abundant conformation in the $\mathbf{4_2}$, $\mathbf{5_2}$, $\mathbf{6_2}$, $\mathbf{7_1}$, $\mathbf{7_2}$, and $\mathbf{7_3}$ simulations, while $\mathbf{4_1}$ presents an important interconversion between *gauche* and *anti* forms and $\mathbf{5_1}$ is mainly fixed in an *anti*-alignment. The two fluorine atoms in the F-Cβ-Cα-F moiety are preferentially placed in the *gauche* relative orientation in all the simulations. There is a frequent interconversion between both *gauche* forms (*g*+ and *g*–) in **4**, while the g+ conformer is preferred in $\mathbf{4_2}$ and $\mathbf{6_2}$ and *g*– is the main form in $\mathbf{5_1}$, $\mathbf{5_2}$, $\mathbf{7_1}$, $\mathbf{7_2}$, and $\mathbf{7_3}$.

In principle, adjacent C-F and C=O bonds are expected to adopt an *anti*-parallel arrangement (*i.e.*, F-C-C=O angle of 180º), what is rationalized by invoking the high polarity of the C-F bond. However, we found that the PDF for the F-Cα-C=O torsion angle in the $\mathbf{4_1}$, $\mathbf{5_1}$, $\mathbf{5_2}$, and $\mathbf{6_2}$ simulations clearly deviates from the *anti*-arrangement. Along the $\mathbf{4_2}$, $\mathbf{7_1}$, $\mathbf{7_2}$, and $\mathbf{7_3}$ trajectories, the Cα-F and C=O bonds present dihedral angles around 200-250º, which are somehow closer to the *anti*-arrangement (180º). In this respect, we note that none $^4J_{H,F}$ couplings between the α-fluorine atom and the adjacent Ala_5@NH proton has been observed in the NMR experiments performed for **4**-**7**.[20] This result suggests that the expected *anti*-parallel C-F/C=O alignments cannot be realized in **4**-**7** in consonance with our cMD results.

<Table 2 here>

From the simulation of the ^{1}H NMR spectra of **3**-**7**, the $^{3}J_{H,H}$ and $^{3}J_{H,F}$ coupling constants have been obtained for the Gaba residue.[20] In principle, these $^{3}J_{H,X}$ coupling constants reflect the underlying H-C-C-X conformation, with torsion angles close to 0º/180º resulting in larger spin-spin couplings than values close to 90º. To compare with these results, we provide in Table 2 an estimation of the selected Gaba coupling constants computed in the cMD simulations by means of Karplus-type relationships. The $^{3}J_{H,H}$ and $^{3}J_{H,F}$ values in Table 2 are given for each cMD simulation as well as their Boltzmann-averaged values.

For unguisin A (**3** in Table 2), all the $^{3}J_{H,H}$ values computed for the $\mathbf{3_1}$ trajectory are close to the results obtained from the ^{1}H NMR spectrum. In addition, the Boltzmann-averaged $^{3}J_{H,H}$ and $^{3}J_{H,F}$ values computed for compounds **4** and **5** also show a good agreement with those derived from the experimental ^{1}H NMR data. Thus, the average $^{3}J_{H,F}$ values resulting from the $\mathbf{4_1}$ and $\mathbf{4_2}$ simulations (21.0/20.1) are very close to the coupling constants reported by Hu *et. al.* (20.8/20.0). Their intermediate magnitude evidences *g+/g−* disorder associated to H-C-C-F angles around 180º having a high coupling constant (*g+* in the *syn-R,R* stereoisomer **4**), and to H-C-C-F angles around 60º with a low coupling constant (*g-* in **4**). In fact a high degree of *g+/g−* interconversion is observed all along the $\mathbf{4_1}$ simulation (see Figure S5) and further *g+/g−* interconversion may result from the conformational exchange between $\mathbf{4_{1B}}$ (*g−*) and $\mathbf{4_{2A}}$ (mostly *g+*). On the other hand, the large $^{3}J_{H,F}$ average values obtained for **5** (23.6/23.2) are also close to those of Hu *et. al.* (27.6/25.0). In our simulations, these coupling constants result from the high abundance of *g−* conformers both in $\mathbf{5_1}$ and $\mathbf{5_2}$ (see Figure S5) that, combined with the (*S,S*) stereochemistry of Gaba, determine an almost unimodal distribution of the H-C-C-F torsions centered around 190-200º values.

When comparing our estimations of the $^3J_{H,X}$ coupling constants for compounds **6** and **7** with those of Hu *et. al.* data, we find significant deviations, specially for the F-Cβ-Cα-H coupling constant (see Table 2). For instance, the $\mathbf{7_1}$, $\mathbf{7_2}$ and $\mathbf{7_3}$ simulations give unimodal H-C-C-F PDFs that peak at low (~70º for F-Cβ-Cα-H) and high torsion (~190º for H-Cβ-Cα-F) angles resulting in Boltzmann-averaged $^3J_{H,F}$ values of 7.2 and 25.2, respectively. The corresponding Hu e*t. al.* values are 18.3 and 21.5 indicating some degree of *g+/g-* disorder at the Gaba residue. Although a few *g+/g−* interchanges occur during the cMDs of **6** and **7** (see Figure S5), their frequency is low and the dominant F-C-C-F conformations (*g+* for **6** and *g−* for **7**) determine the appearance of a pair of low and high $^3J_{H,F}$ values. However, our computational results coincide with the NMR data in the sense that both sets of data show nearly identical $^3J_{H,F}$ values for **6** and **7**, suggesting that the dynamics at the Gaba residue would be similar in the two *anti*-diastereomers. Our results also agree with the NMR ones in showing that the F-Cβ-Cα-H and H-Cβ-Cα-F coupling constants turn out to be dissimilar, although the difference in the $^3J_{H,F}$ values of the same compound is clearly more accentuated in our calculations.

Overall similarities and differences between unguisin A and its difluoro-derivatives

The 2D histogram plots of the torsional coordinates RC1 and RC2 delimit the phase space regions for the various macrocyclic conformations produced by the GaMD/cMD simulations (see Figure 1). As aforementioned, the peptides **3**-**7** tend to populate distinct RC1/RC2 areas although there are noteworthy exceptions (*i.e.*, $\mathbf{3_{1A}}$, $\mathbf{5_{1A}}$, $\mathbf{7_{1A}}$). The cluster representatives in Figure 3 further illustrate the unique structural features of the **3**-**7** conformers, but they also exhibit some similarities concerning the identity of intramolecular H-bonds and the compactness of the structures. To further assess to what extent the difluorination at the Gaba residue modifies the

structure of the unguisin A macrocycle, we computed the backbone RMSD values between the most populated cluster representatives of **4**-**7** with respect to those of **3** (see Table S4). Considering a 0.5 Å RMSD threshold as indicative of close structural proximity, it turns out that the macrocycle conformation in $\mathbf{5_{1A}}$ matches that in $\mathbf{3_{1A}}$ (0.3 Å). In addition, $\mathbf{7_{1A}}$ is similar to $\mathbf{3_{1A}}$ (0.7 Å) while clusters $\mathbf{5_{2A}}$ and $\mathbf{6_{2A}}$ resemble the $\mathbf{3_{1B}}$ structure (0.9 and 0.8 Å, respectively). The rest of the RMSD values are well above 1.0 Å showing that, in general, the stereoselective fluorination of unguisin A alters substantially the geometry of this cyclic peptide, yielding new structures as previously claimed.[20] However, it is equally important to note that exceptions arise and that specific difluorination patterns (e.g., *syn*-(S,S) in this case) may retain a structural similarity with the parent system.

We also evaluated whether or not the GaMD ensemble of the non-fluorinated unguisin A (**3**) includes the conformations favored by the fluorination patterns in **4**-**7**. Inspection of the lowest RMSD values obtained by superposing all the clusters produced by the GaMD simulation of **3** onto the cluster representatives obtained from the cMD simulations of **4-7** (Tabe S5), shows that the majority of the backbone conformations of **4-7** for were also sampled during the enhanced-sampling simulation of **3**, specially when considering the *syn* derivatives **4** and **5**. Most remarkably, the *syn*-(*S*,*S*) **5** presents again a high similarity (0.3 Å for $\mathbf{5_{1A}}$) with the most abundant cluster from the **3** GaMD simulation. This comparative analysis further confirms the ability of GaMD to yield extensive sampling of the macrocycle conformations, but it also suggests that fluorination of unguisin may filter out specific areas of the phase space available to the unsubstituted peptide.

The computational prediction of the possible conformational filtering/stabilization (or complete backbone rearrangement) induced by fluorination, basing only on the knowledge of the

phase space of the unsubstituted cyclic peptide, could help in the rational design of fluorinated peptides. In this respect, we applied the MM-PBSA protocol described in Methods in order to score the stability of the structures that result after having mutated the corresponding H atoms into F atoms at the Gaba residue on the frames from the **3** GaMD trajectory. In this way, a series of *unrelaxed* fluorinated structures (**3_{GaMD}→4, 3_{GaMD}→5**, …) were obtained. Prior to the MM-PBSA calculations, the coordinates of the solute molecule and those of a 15 Å solvent cap were optimized via 1000 conjugate-gradient steps in order to relax the fluorinated Gaba residue and remove any structural influence of the GaMD bias potential. For each fluorine substitution pattern, the calculated G_{MMPBSA} energies were averaged assuming the cluster distribution derived from the **3** GaMD trajectory (only the 50 most populated clusters and 500 snapshots within each cluster were considered).

< Figure 4 here >

Using the MM-PBSA scoring function, it becomes feasible to assess the relative stability of the *fluorinated* **3** GaMD conformations grouped into the various cluster ensembles. These scorings were plotted against the minimum backbone RMSD distances between the corresponding cluster representatives in the **3** GaMD simulation and those of the fluorinated systems in the cMD simulations (see Figure 4). Interestingly, we found that, in the case of the *syn*-(*R,R*), *syn*-(*S,S*) and *anti*-(*R,S*) derivatives (**4-6**), there exists a significant correlation (r=0.4-0.6) between the cluster-averaged G_{MMPBSA} energies and the RMSD distances, in such a way that the lowest G_{MMPBSA} values correspond to those fluorinated **3** GaMD frames that effectively match the **4-7** conformers (*i.e.* **4_{1A}**, **4_{2A}**, **5_{1A}**, etc.). In the case of the *anti*-(*S,R*) (**7**) stereoisomer, however, the G_{MMPBSA} /RMSD data are essentially uncorrelated. Nonetheless, the MM-PBSA scoring of the *fluorinated* GaMD ensemble of the natural unguisin A suggests that the conformational effect of

fluorine largely consists in selecting particular conformations accessible to the unsubstituted macrocylic peptide.

DISCUSSION

Our understanding of peptide macrocycles may benefit from an extensive conformational sampling of their phase space. Many different computational tools have been proposed and/or specifically designed for this task. In this respect, our simulations further support the general-purpose accelerated MD technique as a robust and straightforward approach to calculate free energy maps of the conformational landscape of the macrocycles. Although costly simulations on the μs time scale are required for convergence, the GaMD calculations can be efficiently run on GPU devices. However, an adequate force field parameterization is a prerequisite for the application of GaMD or other methods, what is particularly relevant for fluorinated residues and/or non-standard amino acids that are generally missing in conventional force fields. Indeed it must be emphasized that the modelling of the unguisin A peptides depends critically on the development of specific torsional parameters for the Gaba residue, as they largely determine the preferred conformations for such residue. On the other hand, we note that conventional MD simulations play an equally important role in the characterization of the peptide macrocycles because they confirm and refine the conformational search performed by the GaMD calculations. It is worth mentioning that the various GaMD minima were probed by the cMD simulations in order to achieve a more exhaustive conformational search. Thus, in our modelling of unguisin A (**3**) and its difluorinated derivatives (**4-7**), the majority of the macrocyclic GaMD conformations turned out to be essentially stable along the μs cMD trajectories (*e.g.*, the conformers of **4-6**). Moreover, the extensive cMD equilibrium sampling can help unveil other alternative

configurations (*e.g.*, the $\mathbf{3_{1B}}$ conformer) as well as to assess the flexibility of the macrocycles. The cMD simulations also allowed us to estimate the relative stability of the conformational ensembles by applying approximate end-point free energy methods, which, in turn, can largely benefit from the inclusion of conformational entropies of the peptidic solutes. Therefore, we conclude that the combined GaMD + cMD protocol can be a powerful computational tool, particularly suitable to predict the most likely conformations of cyclic peptides in aqueous and organic solvents.

The experimental pieces of information about the secondary structure of the unguisin A peptide and its four difluoro-derivatives have been obtained indirectly from NMR data.[19, 20] Although static molecular models have been reported in ref. 20 (built from short 1 *ns* constrained MD simulations followed by MM relaxation), they yield little or no insight into the conformational properties of the systems. We believe that these static models are now superseded by our parameterization work and the combined GaMD + cMD simulations, which convincingly show that two or more conformations become structurally and energetically accessible to each of the **3**-**7** macrocycles.

The structural features of the cMD conformers are in consonance with the experimental NMR-based evidences. For instance, the variations in the ^{1}H-NMR chemical shifts with *T* ($\Delta\delta/\Delta T$) point out that each cyclic peptide has certain NH protons involved in intramolecular H-bonds, and several empirical and computational H-bond assignments coincide. A more direct comparison between the experimental data and our simulations is outlined in terms of the NMR *J*-based analysis of the conformations of the –CβHX-CαHX– moiety of the Gaba residue, which maintains a *gauche* orientation between the two F atoms in all the simulations. For compounds **3**,

4 and **5**, there is a good quantitative agreement in the $^{3}J_{H,H}$ and $^{3}J_{H,F}$ (F-Cβ-Cα-H and H-Cβ-Cα-F) constants . The analysis of the torsion angles during the cMD simulations further reveals the degree of conformational disorder at the Gaba residue (*e.g.*, the fast *g+/g−* rearrangement in **4$_1$**). In the case of **6** and **7**, the comparative analysis indicates that the current force-field representation underestimates the *g+/g-* disorder, what results in the overestimation of the difference between the two $^{3}J_{H,F}$ constants. However, the simulations reproduce satisfactorily the lack of an *anti*-parallel alignment of the F-Cα-C=O moiety as experimentally confirmed by the absence of the $^{4}J_{H,F}$ couplings.

The experimental proposal that the fluorinated peptides adopt dramatically different secondary structures seems, to a large extent, corroborated by the GaMD+cMD protocol given that the simulations of the **4-7** cyclic peptides lead to distinct conformational preferences. However, when we compare in detail the **4-7** cMD structures with the major cluster representatives of **3**, an important caveat arises, that is, the (*R,R*), (*R,S*) and (*S,R*) substitutions alter substantially the secondary structure of natural unguisin, but the (*S,S*) stereochemistry is more prone to select the unguisin A conformations (**3$_{1A}$** and **3$_{1B}$**). In addition, the majority of the preferred backbone structures of the fluorinated peptides are significantly sampled during the enhanced GaMD simulation of natural unguisin A. Furthermore, it may be possible to rank the GaMD conformations of **3** in terms of their propensity to become stabilized by selective fluorination as suggested by the MM-PBSA scoring calculations. These interesting results suggest that, depending on the stereochemical pattern, fluorine substitution may destabilize or stabilize particular conformations of the cyclic peptides. It turns out also that some *a priori* knowledge of the fluorine effects could be gained by *fluorinating* the enhanced conformational space of the unsubstituted peptide molecules.

In principle, several factors (*gauche effect*, intramolecular contacts, solute-solvent interactions, solute entropy, etc.) may affect the conformational changes in the unguisin A macrocycle upon fluorination. The preference for *gauche* conformations in the Gaba residue of **4**-**7** points towards the *gauche* effect as the main factor in dictating the overall conformations, what would be in line with former proposals suggesting that a low energy fluorine *gauche* interaction can trigger a drastic reorganization of the macrocycle backbone.[10] However, other factors, like steric effects, have been also shown to operate in multi-vicinal fluoroalkanes that result in an unexpected strong propensity for *anti* C-F/C-F alignments.[16] In this respect, we have found that understanding the fluorine effects in terms of conformational energies requires both a physically-based scoring function (*e.g.*, MM-PBSA with *explicit* non-polar solute-solvent interactions) and significant amounts of sampling (*e.g.*, μs-length GaMD /cMD), suggesting thus that a complex interplay among various enthalpic and entropic factors controls the macrocycle conformations.

CONCLUSIONS

In this work, we provide further evidence that the prediction of macrocyclic structures like unguisin A in its fluorinated derivatives is computationally feasible thanks to the application of enhanced-sampling MD methods using an explicit solvent representation. Our results also suggest the convenience of carrying out extensive cMD trajectories to refine the GaMD conformers and characterize their equilibrium properties. The relative stability of the cMD ensembles can be estimated using average MM-PBSA energies augmented with conformational entropy estimations. The MM-PBSA energies as computed in this work could also be useful to guess the conformational preferences of fluorine substitution utilizing the GaMD/cMD structures of the unsubstituted systems.

According to the GaMD+cMD protocol, the unguisin A macrocycle and the series of difluorinated derivatives are relatively flexible molecules in DMSO solution that can adopt two or more conformations depending on the fluorination pattern. All of the conformers present both fluorine atoms in *gauche* and, remarkably, the *anti*-planar C-F/C=O alignment is not observed. The computational predictions, which are validated by a direct comparison with the experimental NMR *J*-based analysis of the Gaba conformations, confirm that, in general, the fluorinated peptides possess secondary structures that deviate significantly from that of the natural unguisin A, excepting for the *syn*-(S,S) derivative that remains structurally close to unguisin A. Finally, we conclude that the conformational effect of the F atoms cannot be predicted only in terms of the *gauche* effect at the Gaba residue and that, most likely, the macrocycle conformations are determined by a subtle enthalpy-entropy balance Therefore, the GaMD + cMD or similar computational protocols seem required to yield reliable predictions about the fluorine substitution effects.

ACKNOWLEDGMENT

We acknowledge the financial support by FICyT (IDI2018-000177) co-financed by FEDER funds. This research was also supported by the CTQ2015-65790-P and PGC2018-095953-B-100 grants (MICINN, Spain). We also thank the reviewers for their helpful comments and suggestions.

SUPPORTING INFORMATION

Table S1 with the MM parameters for the Gaba residues. Table S2 with results of the clustering analysis of the GaMD and cMD trajectories. Figure S1 with ball-and-stick views of the selected GaMD structures. Figure S3 showing the time evolution of the MM-PBSA energies. Figure S4

with convergence plots of conformational entropy. Figure S5 with PDF and time evolution of selected torsions. Table S3 with information for the most abundant H-bonds. Tables S4 and S5 with RMSD values computed between cluster representatives. Amber *prep* and *frcmod* files for the different Gaba residues.

REFERENCES

1. Fosgerau, K.; Hoffmann, T., Peptide Therapeutics: Current Status and Future Directions. *Drug Discov. Today* **2015**, *20*, 122-128.

2. Lee, A. C.-L.; Harris, J. L.; Khanna, K. K.; Hong, J.-H., A Comprehensive Review on Current Advances in Peptide Drug Development and Design. *Int. J. Mol. Sci.* **2019**, *20*, 2383.

3. Jing, X.; Jin, K., A Gold Mine for Drug Discovery: Strategies to Develop Cyclic Peptides into Therapies. *Med. Res. Rev.* **2020**, *40*, 753-810.

4. Matsson, P.; Doak, B. C.; Over, B.; Kihlberg, J., Cell Permeability Beyond the Rule of 5. *Adv. Drug Deliver. Rev.* **2016**, *101*, 42-61.

5. Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S., Testing the Conformational Hypothesis of Passive Membrane Permeability Using Synthetic Cyclic Peptide Diastereomers. *J. Am. Chem. Soc.* **2006**, *128*, 2510-2511.

6. Vinogradov, A. A.; Yin, Y.; Suga, H., Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *J. Am. Chem. Soc.* **2019**, *141*, 4167-4181.

7. Sebastiano, M. R.; Doak, B. C.; Backlund, M.; Poongavanam, V.; Over, B.; Ermondi, G.; Caron, G.; Matsson, P.; Kihlber, J., Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5. *J. Med. Chem.* **2018**, *61*, 4189-4202.

8. Poongavanam, V.; Danelius, E.; Peintner, S.; Alcaraz, L.; Caron, G.; Cummings, M. D.; Wlodek, S.; Erdelyi, M.; Hawkins, P. C. D.; Ermondi, G.; Kihlberg, J., Conformational Sampling of Macrocyclic Drugs in Different Environments: Can We Find the Relevant Conformations? *ACS Omega* **2018**, *3*, 11742-11757.

9. Dougherty, P. G.; Sahni, A.; Pei, D., Understanding Cell Penetration of Cyclic Peptides. *Chem. Rev.* **2019**, *119*, 10241-10287.

10. Appavoo, S. D.; Huh, S.; Diaz, D. B.; Yudin, A. K., Conformational Control of Macrocycles by Remote Structural Modification. *Chem. Rev.* **2019**, *119*, 9724-9752.

11. Zhou, Y.; Wang, J.; Gu, Z.; Wang, S.; Zhu, W.; Aceña, J. L.; Soloshonok, V. A.; Izawa, K.; Liu, H., Next Generation of Fluorine-Containing Pharmaceuticals, Compounds Currently in Phase II−III Clinical Trials of Major Pharmaceutical Companies: New Structural Trends and Therapeutic Areas. *Chem. Rev.* **2016**, *116*, 422-518.

12. Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A., Applications of Fluorine in Medicinal Chemistry. *J. Med. Chem.* **2015**, *58*, 8315-8359.

13. McNaught, A. D.; Wilkinson, A., *IUPAC. Compendium of Chemical Terminology*. Blackwell Scientific Publications: Oxford, 1997.

14. Hunter, L., The C–F Bond as a Conformational Tool in Organic and Biological Chemistry. *Beilstein J. Org. Chem.* **2010**, *6* (38).

15. Briggs, C. R. S.; O'Hagan, D.; Howard, J. A. K.; Yufit, D. S., The C-F Bond as a Tool in the Conformational Control of Amides. *J. Fluor. Chem.* **2003**, *119*, 9-13.

16. Scheidt, F.; Selter, P.; Santschi, N.; Holland, M. C.; Dudenko, D. V.; Daniliuc, C.; Mück-Lichtenfeld, C.; Hansen, M. R.; Gilmour, R., Emulating Natural Product Conformation by Cooperative, Non-Covalent Fluorine Interactions. *Chem. Eur. J.* **2017**, *23*, 6142-6149.

17. Yamamoto, I.; Jordan, M. J. T.; Gavande, N.; Doddareddy, M. R.; Chebib, M.; Hunter, L., The Enantiomers of syn-2,3-Difluoro-4-Aminobutyric Acid Elicit Opposite Responses at the $GABA_C$ Receptor. *Chem. Commun.* **2012**, *48*, 829-831.

18. Ariawan, D.; Webb, J. E. A.; Howe, E. N. W.; Gale, P. A.; Thordarson, P.; Hunter, L., Cyclic peptide unguisin A is an anion receptor with high affinity for phosphate and pyrophosphate. *Org. Biomol. Chem.* **2017**, *15*, 2962-2967

19. Hunter, L.; Chung, J. H., Total Synthesis of Unguisin A. *J. Org. Chem.* **2011**, *76*, 5502-5505.

20. Hu, X.-G.; Thomas, D. S.; Griffith, R.; Hunter, L., Stereoselective Fluorination Alters the Geometry of a Cyclic Peptide: Exploration of Backbone-Fluorinated Analogues of Unguisin A. *Angew. Chem. Int. Ed.* **2014**, *53*, 6176-6179.

21. Díaz, N.; Jiménez-Grávalos, F.; Suárez, D.; Francisco, E.; Martín-Pendás, A., Fluorine Conformational Effects Characterized by Energy Decomposition Analysis. *PCCP* **2019**, *21*, 25258-25275.

22. Kamenik, A. S.; Lessel, U.; Fuchs, J. E.; Fox, T.; Liedl, K. R., Peptidic Macrocycles. Conformational Sampling and Thermodynamic Characterization. *J. Chem. Inf. Model.* **2018**, *58*, 982-992.

23. Jusot, M.; Stratmann, D.; Vaisset, M.; Chomilier, J.; Cortés, J., Exhaustive Exploration of the Conformational Landscape of Small Cyclic Peptides Using a Robotics Approach. *J. Chem. Inf. Model.* **2018**, *58*, 2355-2368.

24. D'Annessa, I.; Di Leva, F. S.; La Teana, A.; Novellino, E.; Limongelli, V.; Di Marino, D., Bioinformatics and Biosimulations as Toolbox for Peptides and Peptidomimetics Design: Where Are We? *Front. Mol. Biosci.* **2020**, *7* (66).

25. CSID:8828057. http://www.chemspider.com/Chemical-Structure.8828057.html (July 13, 2017).

26. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4* (17).

27. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

28. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Determining Atom-Centered Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

29. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357-1377.

30. Sauton, N.; Lagorce, D.; Villoutreix, B. O.; Miteva, M. A., Ms-Dock: Accurate Multiple Conformation Generator and Rigid Docking Protocol for Multi-Step Virtual Ligand Screening. *BMC Bioinf.* **2008**, *9* (184).

31. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. 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.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.;

Cammi; 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 03*, Gaussian, Inc.: Wallingford, CT, USA, 2004.

32. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247-260.

33. Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *Amber 2016*, University of San Francisco: San Francisco, CA, 2016.

34. Neese, F., The ORCA Program System. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73-78.

35. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An Overview of the Amber Biomolecular Simulation Package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198-210.

36. Fox, T.; Kollman, P. A., Application of the RESP Methodology in the Parametrization of Organic Solvents. *J. Phys. Chem. B.* **1998**, *102*, 8070-8079.

37. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

38. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

39. Miao, Y.; Feher, V.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584-3595.

40. Le Grand, S.; Götz, A. W.; Walker, R. C., SPFP: Speed without Compromise-A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374-380.

41. Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A., Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10*, 2677-2689.

42. Perez, J. J.; Tomas, M. S.; Rubio-Martinez, J., Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory. *J. Chem. Inf. Model.* **2016**, *56*, 1950-1962.

43. Roe, D. R.; Cheatham, T. E., Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.

44. Toniolo, C.; Crisma, M.; Formaggio, F.; Alemán, C.; Ramakrishnan, C.; Kalmankar, N.; Balaram, P., Intramolecular Backbone···Backbone Hydrogen Bonds in Polypeptide Conformations. The Other Way Around: ε-turn. *Biopolymers* **2017**, *108*, e22911.

45. Pérez, C.; Frank Löhr, F.; Rüterjans, H.; Schmidt, J. M., Self-Consistent Karplus Parametrization of ^{3}J Couplings Depending on the Polypeptide Side-Chain Torsion χ1. *J. Am. Chem. Soc.* **2001**, *123*, 7081-7093.

46. Thibaudeau, C.; Plavec, J.; Chattopadhyaya, J., A New Generalized Karplus-Type Equation Relating Vicinal Proton-Fluorine Coupling Constants to H-C-C-F Torsion Angles. *J. Org. Chem.* **1998**, *63*, 4967-4984.

47. Suárez, E.; Díaz, N.; Suárez, D., Entropy Calculations of Single Molecules by Combining the Rigid-Rotor and Harmonic-Oscillator Approximations with Conformational Entropy Estimations from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2638-2653.

48. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem.* **2004**, *25*, 1605-1612.

49. Peng, C.; Atilaw, Y.; Wang, J.; Xu, Z.; Poongavanam, V.; Shi, J.; Kihlberg, J.; Zhu, W.; Erdélyi, M., Conformation of the Macrocyclic Drug Lorlatinib in Polar and Nonpolar Environments: A MD Simulation and NMR Study. *ACS Omega* **2019**, *4*, 22245-22250.

50. Grochowski, P.; Trylska, J., Continuum Molecular Electrostatics, Salt Effects, and Counterion Binding—A Review of the Poisson–Boltzmann Theory and its Modifications. *Biopolymers* **2008**, *89*, 93-113.

51. Cai, Q.; Hsieh, M.-J.; Wang, J.; Luo, R., Performance of Nonlinear Finite-Difference Poisson−Boltzmann Solvers. *J. Chem. Theory Comput.* **2010**, *6*, 203-211.

52. Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *Journal of computational chemistry* **2004**, *25*, 238-50.

53. Suárez, D.; Díaz, N., Direct methods for computing single-molecule entropies from molecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015**, *5*, 1-26.

54. Macke, T. J.; Case, D. A. Modeling Unusual Nucleic Acid Structures. In *Molecular modeling of nucleic acids*; American Chemical Society: 1997; Vol. *682*, Chapter 24, pp 379-393.

55. Grossfield, A.; Zuckerman, D. M. Chapter 2 Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. In *Annual Reports in Computational Chemistry*, Ralph, A. W., Ed.; Elsevier: 2009; Vol. *5*, pp 23-48.

56. Suárez, E.; Díaz, N.; Méndez, J.; Suárez, D., CENCALC: A Computational Tool for Conformational Entropy Calculations from Molecular Simulations. *J. Comput. Chem.* **2013**, *34*, 2041-2054.

57. Suárez, E.; Suárez, D., Multibody Local Approximation: Application to Conformational Entropy Calculations on Biomolecules. *J. Chem. Phys.* **2012**, *137*, 084115.

58. King, B. M.; Tidor, B., MIST: Maximum Information Spanning Trees for Dimension Reduction of Biological Data Sets. *Bioinformatics (Oxford, England)* **2009**, *25*, 1165-72.

59. Andersen, N. H.; Neidigh, J. W.; Harris, S. M.; Lee, G. M.; Liu, Z.; Tong, H., Extracting Information from the Temperature Gradients of Polypeptide NH Chemical Shifts. 1. The Importance of Conformational Averaging. *J. Am. Chem. Soc.* **1997**, *119*, 8547-8561.

60. Daley, M. E.; Graether, S. P.; Sykes, B. D., Hydrogen Bonding on the Ice-Binding Face of a β-Helical Antifreeze Protein Indicated by Amide Proton NMR Chemical Shifts. *Biochemistry* **2004**, *43*, 13012-13017.

Table 1. Average values for the MM-PBSA energies (in *kcal/mol*) and absolute entropies (S^{RRHO} in *cal/mol K*) of the polypeptide molecules in the various cMD trajectories. The limiting values of the conformational entropies ($S_{conform}$) are also included. Block average errors are indicated in parentheses.

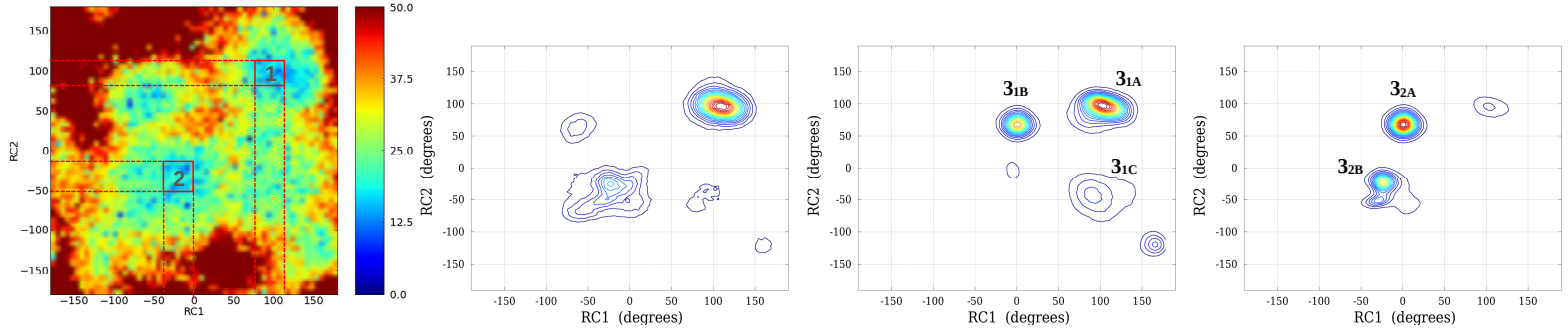
	$\bar{G}_{MMPBSA}$	$\bar{S}_{RRHO}$	$S_{conform}$	$G^{(*)}$
3₁ (8.0 μs)	-31.9 (0.9)	317.0 (0.8)	38.7	-138.6 (0.7)
3₂ (8.0 μs / last 4.0 μs)	-31.4 (1.8) / -35.0 (0.2)	312.9 (0.8) / 314.2 (0.2)	38.1 / 23.8	-136.7 (1.6) / -136.4 (0.2)
4₁ (4.0 μs)	-26.1 (0.2)	320.0 (0.1)	28.5	-130.7 (0.2)
4₂ (4.0 μs)	-27.9 (0.3)	316.3 (0.1)	25.3	-130.4 (0.2)
5₁ (4.0 μs)	-27.5 (0.2)	321.0 (0.1)	20.4	-129.9 (0.3)
5₂ (4.0 μs)	-27.2 (0.3)	321.5 (0.6)	23.2	-130.6 (0.2)
6₁ (4.0 μs)	-26.6 (0.4)	322.0 (0.1)	20.1	-129.2 (0.3)
6₂ (8.0 μs / last 3.0 μs)	-25.1 (0.8) / -26.3 (0.3)	319.9 (1.1) / 321.3 (0.3)	34.7 / 23.2	-131.5 (0.8) / -129.9 (0.6)
7₁ (8.0 μs / last 6.5 μs)	-27.4 (0.6) / -28.4 (0.2)	320.6 (0.3) / 321.0 (0.2)	33.2 / 27.7	-133.5 (0.5) / -133.0 (0.2)
7₂ (8.0 μs)	-24.2 (0.4)	322.2 (0.8)	41.4	-133.3 (1.0)
7₃ (8.0 μs / last 4.0 μs)	-25.4 (1.2) / -27.7 (0.3)	322.0 (0.9) / 321.2 (0.3)	35.9 / 29.7	-132.8 (0.2) / -133.0 (0.2)

(*) The global *G* scoring is obtained by combing the MM-PBSA and entropic terms: $G = \bar{G}_{MMPBSA} - TS_{RRHO} - TS_{conform}$

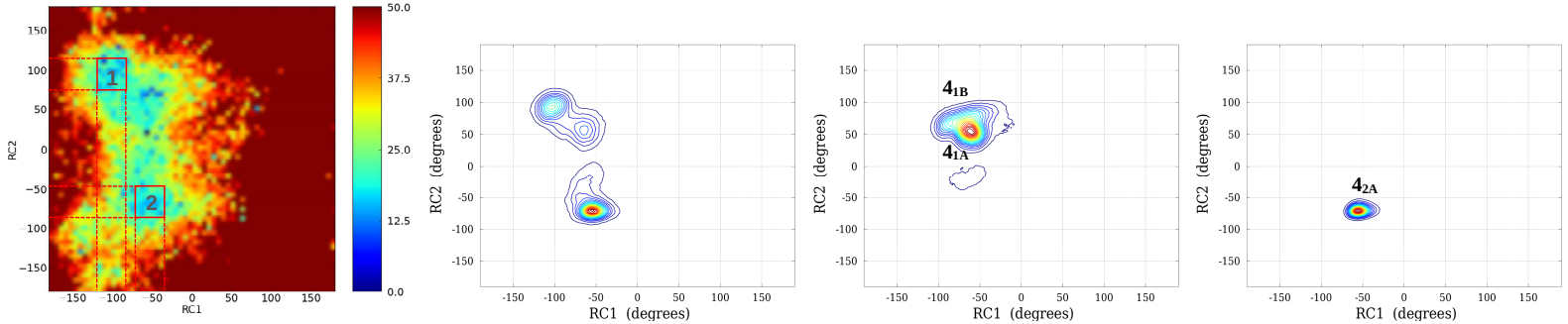
Table 2. Mean values and standard deviations of the $^3J_{H,H}$ and $^3J_{H,F}$ spin-spin coupling constants (Hz) estimated from the cMD simulations of **3-7** using Karplus relationships for selected bond and torsion angles in Gaba. The standard error of the mean values is below 0.01 Hz in all cases. The mean values correspond to the weighted average using the *G* scorings of the cMD simulations. Previously reported values for **3-7** obtained from the simulation of the ^{1}H NMR spectra (ref. 19 and 20) are also given in Italics.

	$^3J_{H,H}$ (H4-Cβ-Cα-H6)	$^3J_{H,H}$ (H4-Cβ-Cα-H7)	$^3J_{H,H}$ (H5-Cβ-Cα-H6)	$^3J_{H,H}$ (H5-Cβ-Cα-H7)
3_1	4.2 ± 2.5	8.7 ± 3.1	8.7 ± 3.1	5.0 ± 2.5
NMR	*5.2*	*8.0*	*7.7*	*5.8*
	$^3J_{H,F}$ (F-Cβ-Cα-H)	$^3J_{H,F}$ (H-Cβ-Cα-F)	$^3J_{H,H}$ (H-Cβ-Cα-H)	
4_1	16.9 ± 14.2	16.6 ± 13.6	6.1 ± 3.6	
4_2	27.7 ± 10.4	25.9 ± 10.2	3.3 ± 1.1	
Average	21.0	20.1	5.0	
NMR	*20.8*	*20.0*	*4.0*	
5_1	23.1 ± 10.1	22.7 ± 10.0	2.9 ± 1.3	
5_2	23.8 ± 10.4	23.4 ± 10.1	2.9 ± 1.2	
Average	23.6	23.2	2.9	
NMR	*27.6*	*25.0*	*1.0*	
6_2	8.0 ± 6.5	25.8 ± 10.6	3.8 ± 1.3	
NMR	*18.4*	*21.1*	--	
7_1	6.2 ± 4.4	25.2 ± 9.8	4.2 ± 1.0	
7_2	8.8 ± 6.0	25.3 ± 9.9	3.5 ± 1.0	
7_3	7.0 ± 5.2	25.3 ± 9.9	4.0 ± 1.1	
Average	7.2	25.2	3.9	
NMR	*18.3*	*21.5*	*2.6*	

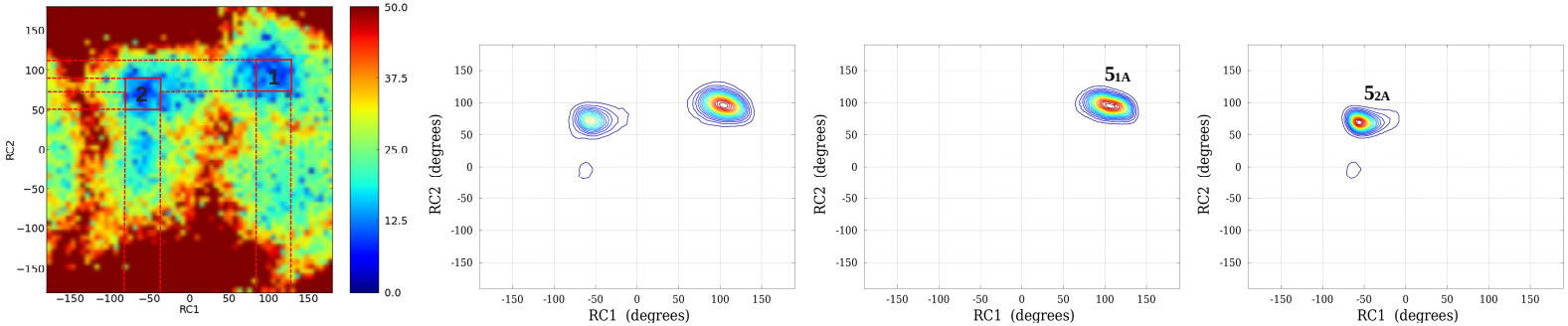
GaMD **3** cMD **3_1** cMD **3_2**



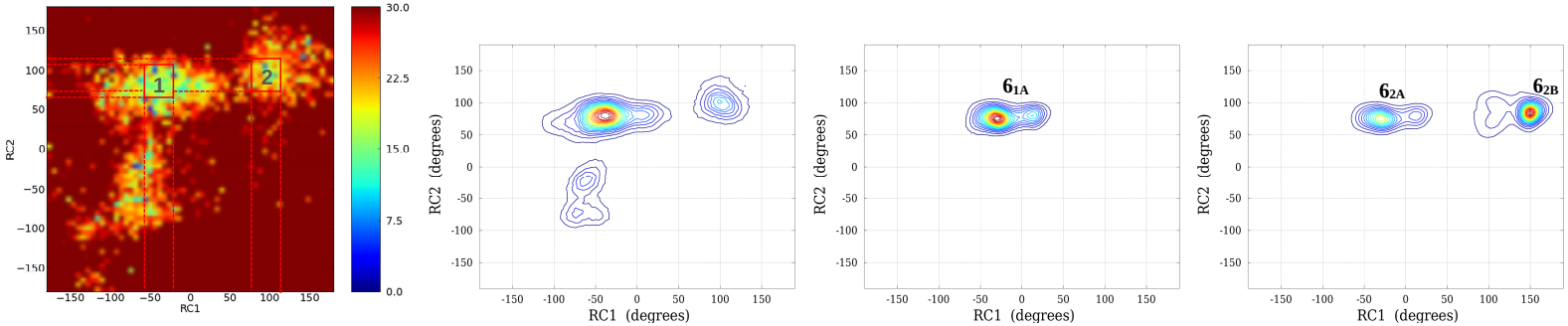
GaMD **4** cMD **4_1** cMD **4_2**



GaMD **5** cMD **5_1** cMD **5_2**



GaMD **6** cMD **6_1** cMD **6_2**



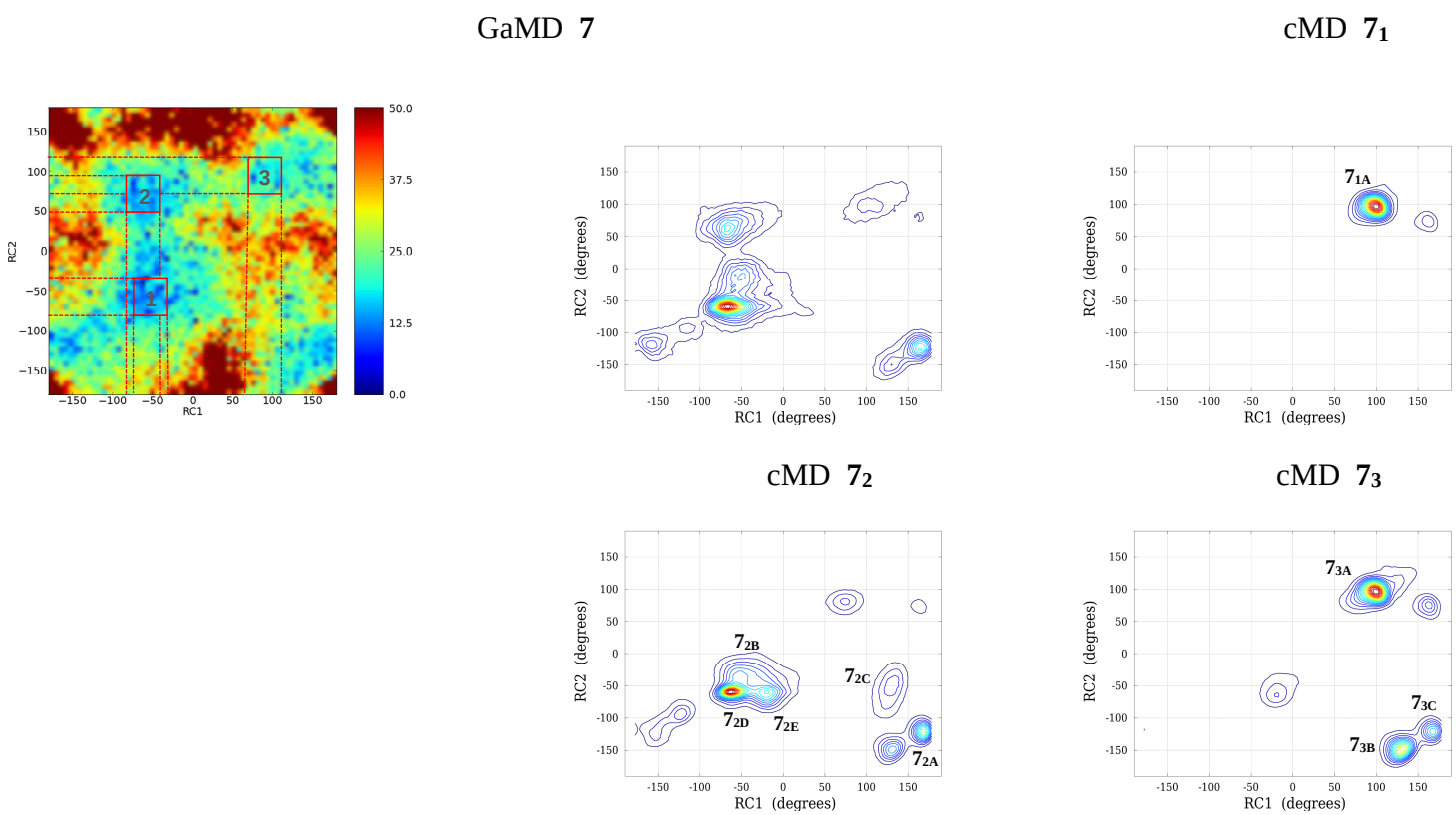
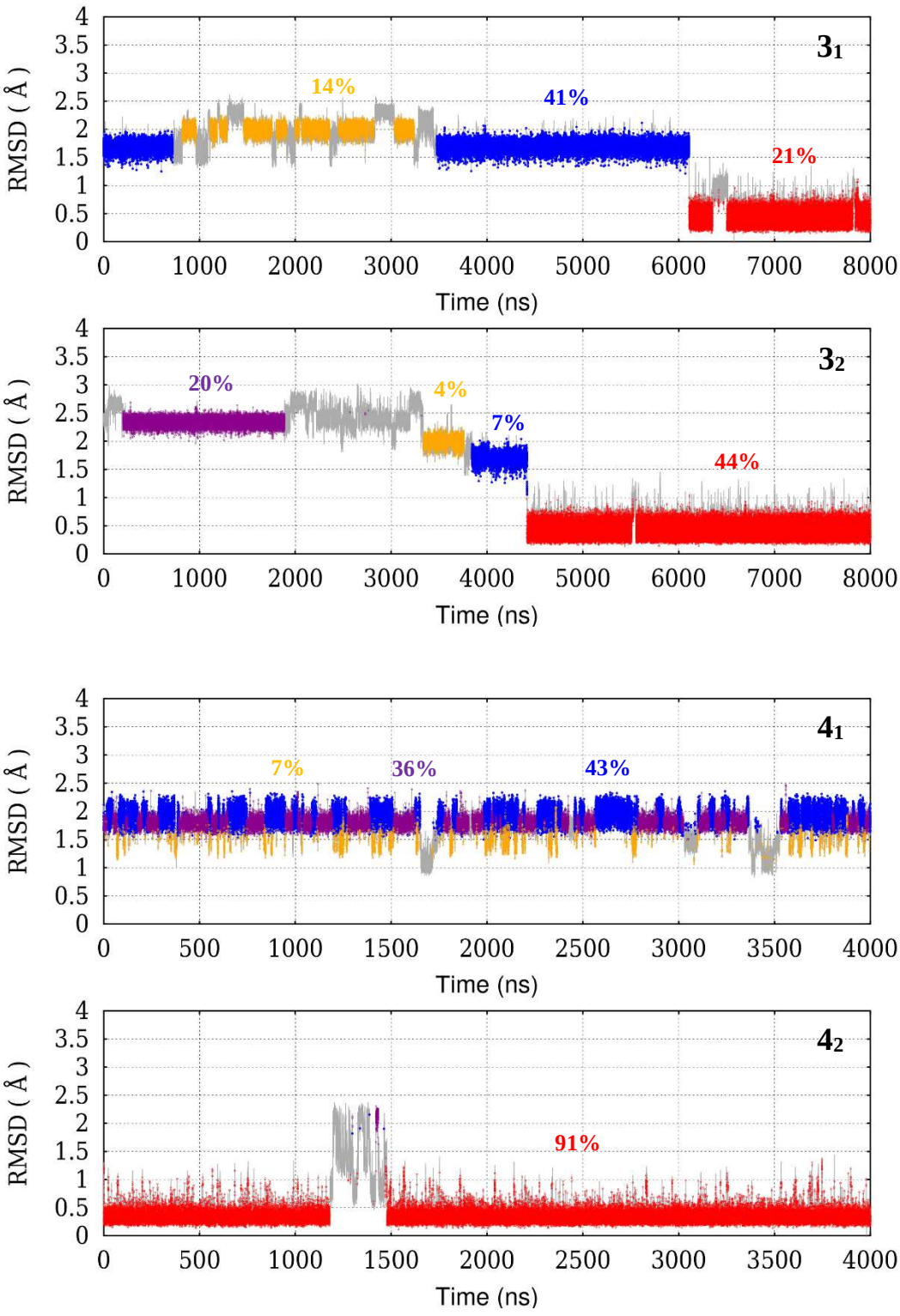
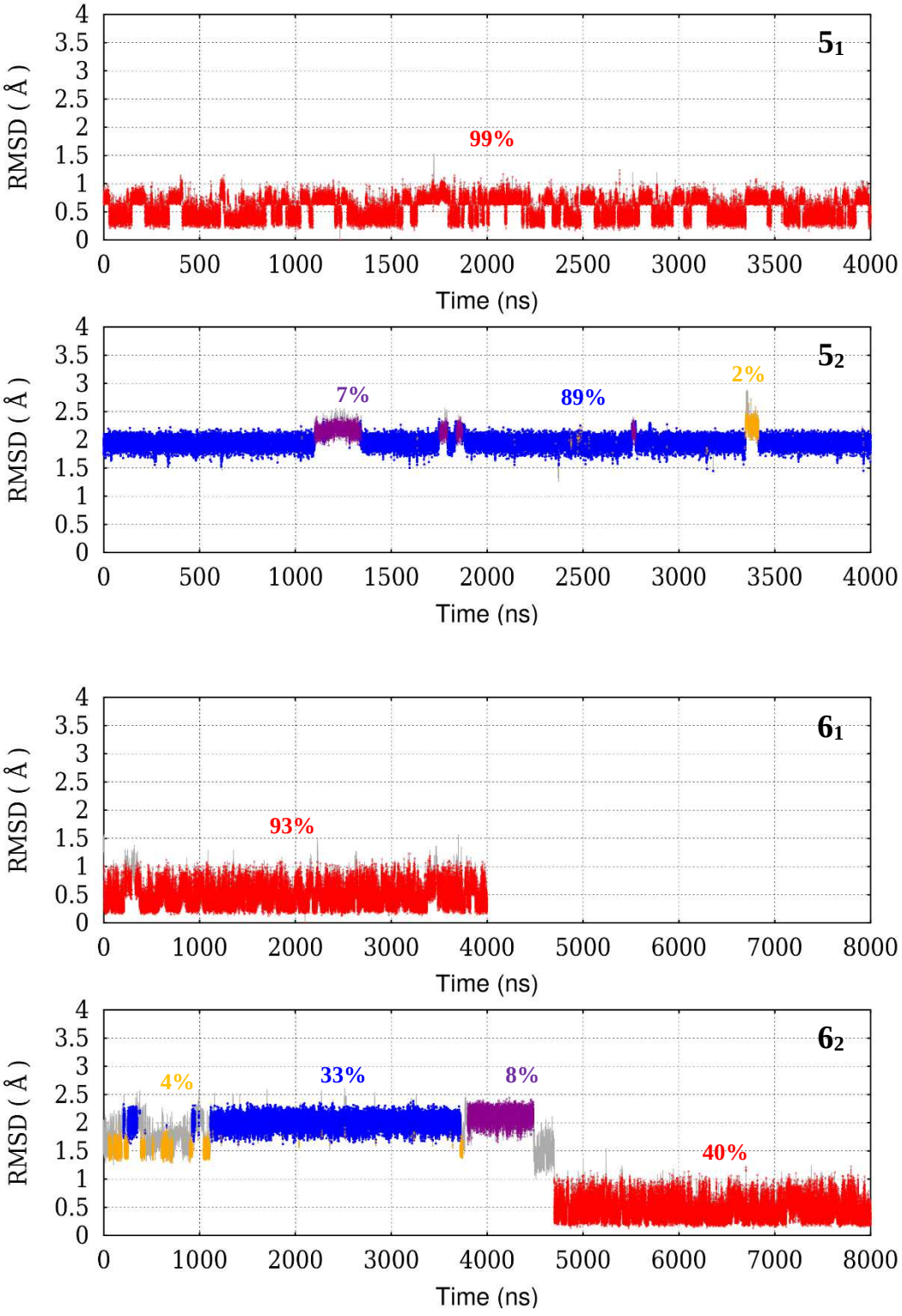


Figure 1. Left) Free energy maps obtained for compounds **3-7** after reweighting the GaMD simulations using the RC1 (Val_1@Cα–Ala_2@Cα–Trp_3@Cα–$Gaba_4$@Cγ) and the RC2 ($Gaba_4$@Cα–Ala_5@Cα–Val_6@Cα–Phe_7@Cα) coordinates. The red squares within the free energy maps delimitate the regions for which structures were randomly selected to start the cMD simulations. The square numbering correspond to the cMD numbering. Right) 2D histogram representations of the population of the RC1 and RC2 values sampled along the GaMD and the cMD simulations. For each cMD histogram, the conformational regions are labelled accordingly.


RMSD (Å)
Time (ns)
3₁
14%
41%
21%
3₂
20%
4%
7%
44%
4₁
7%
36%
43%
4₂
91%


RMSD (Å)
Time (ns)
5₁
99%
5₂
7%
89%
2%
6₁
93%
6₂
4%
33%
8%
40%

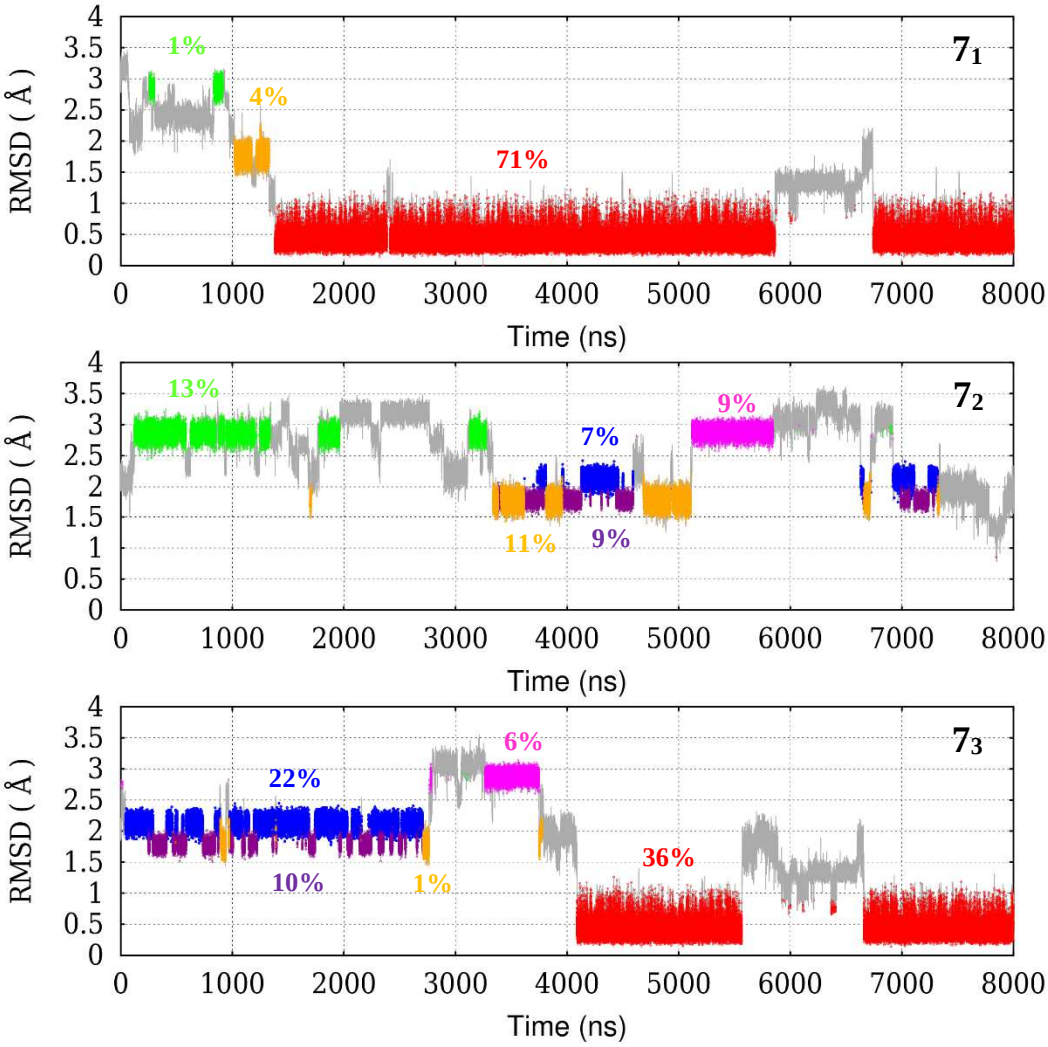
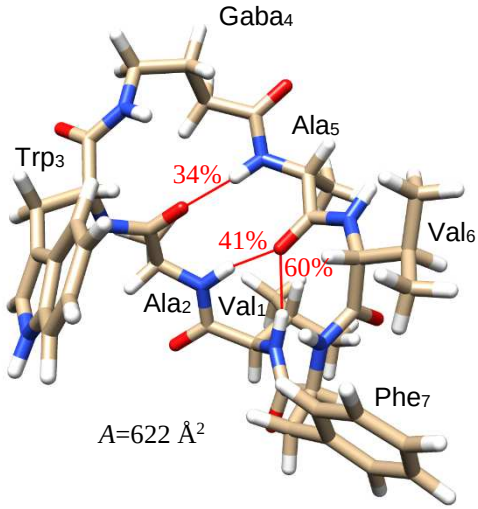
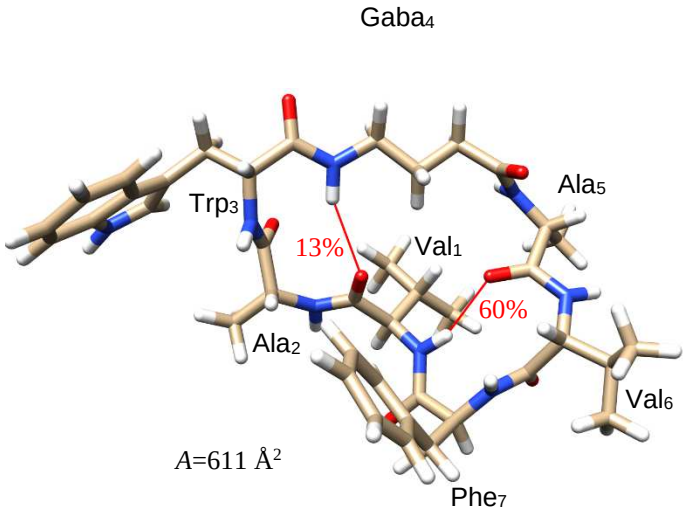


Figure 2. Time evolution of the root mean squared deviation (RMSD in Å) computed for the backbone heavy atoms along the different cMD simulations. In the RMSD calculations, the reference structure is the representative of the most populated cluster. The color of the data points denotes cluster membership ordered in terms of abundance (red>blue>dark magenta>orange>green>magenta; grey denotes other lowly populated clusters). The percentages of abundance within each individual trajectory are also indicated.

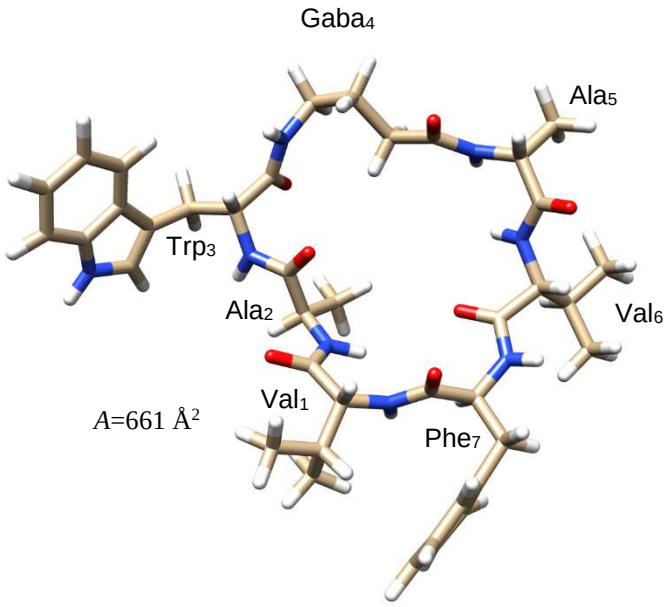
3_{1A} (41%)



3_{1B} (21%)

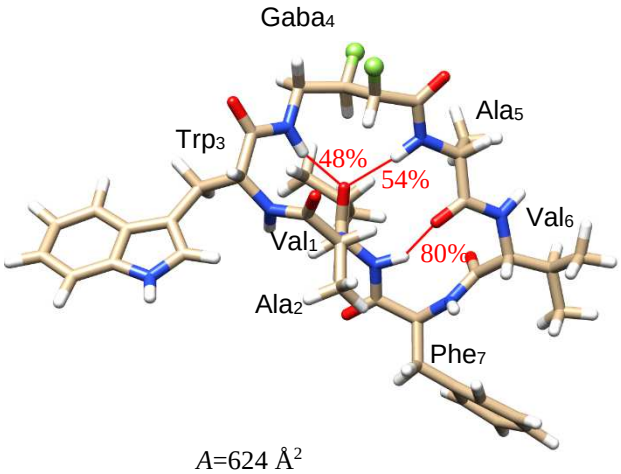


3_{1C} (14%)



4_{1A} (43%)

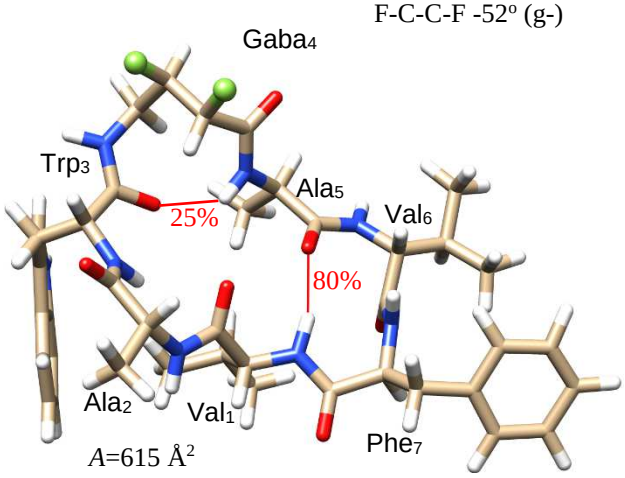
F-C-C-F 64º (g+)



A=624 Å^2

4_{1B} (36%)

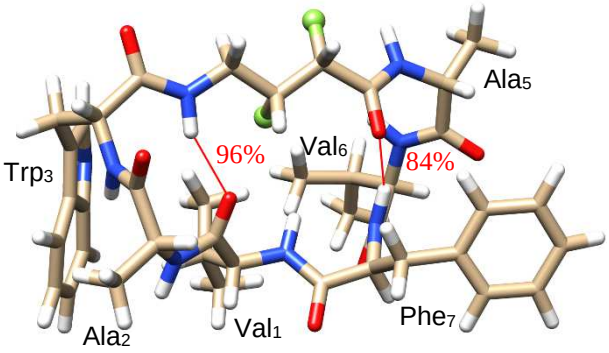
F-C-C-F -52º (g-)



A=615 Å^2

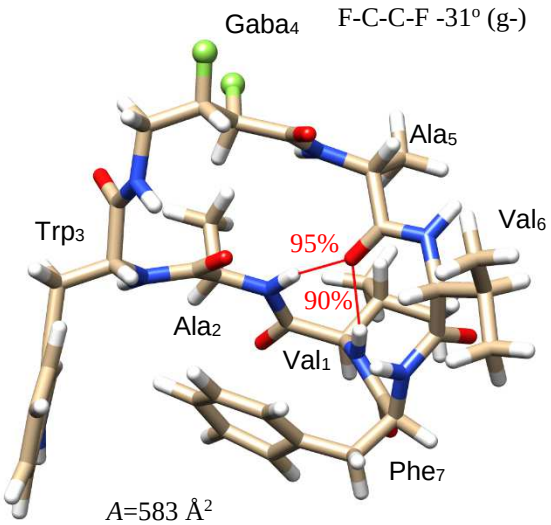
4_{2A} (91%)

F-C-C-F 68º (g+)

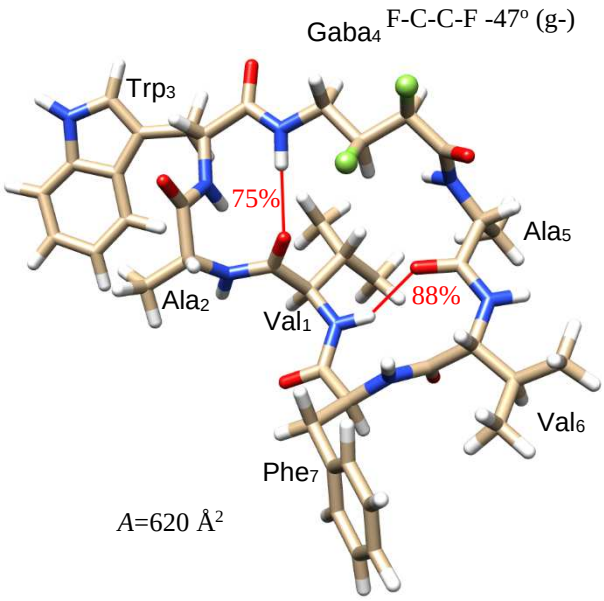


A=572 Å^2

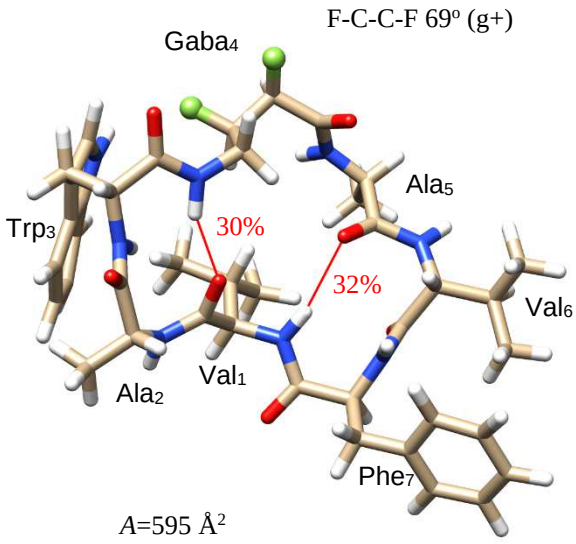
5_{1A} (99%)



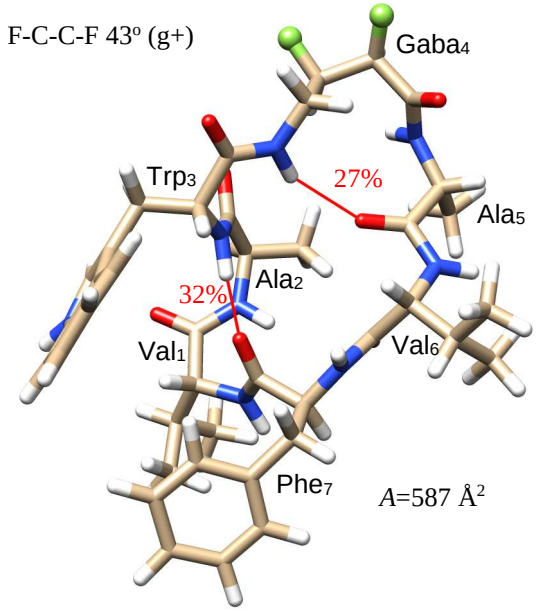
5_{2A} (89%)



6_{2A} (40 %)



6_{2B} (33 %)



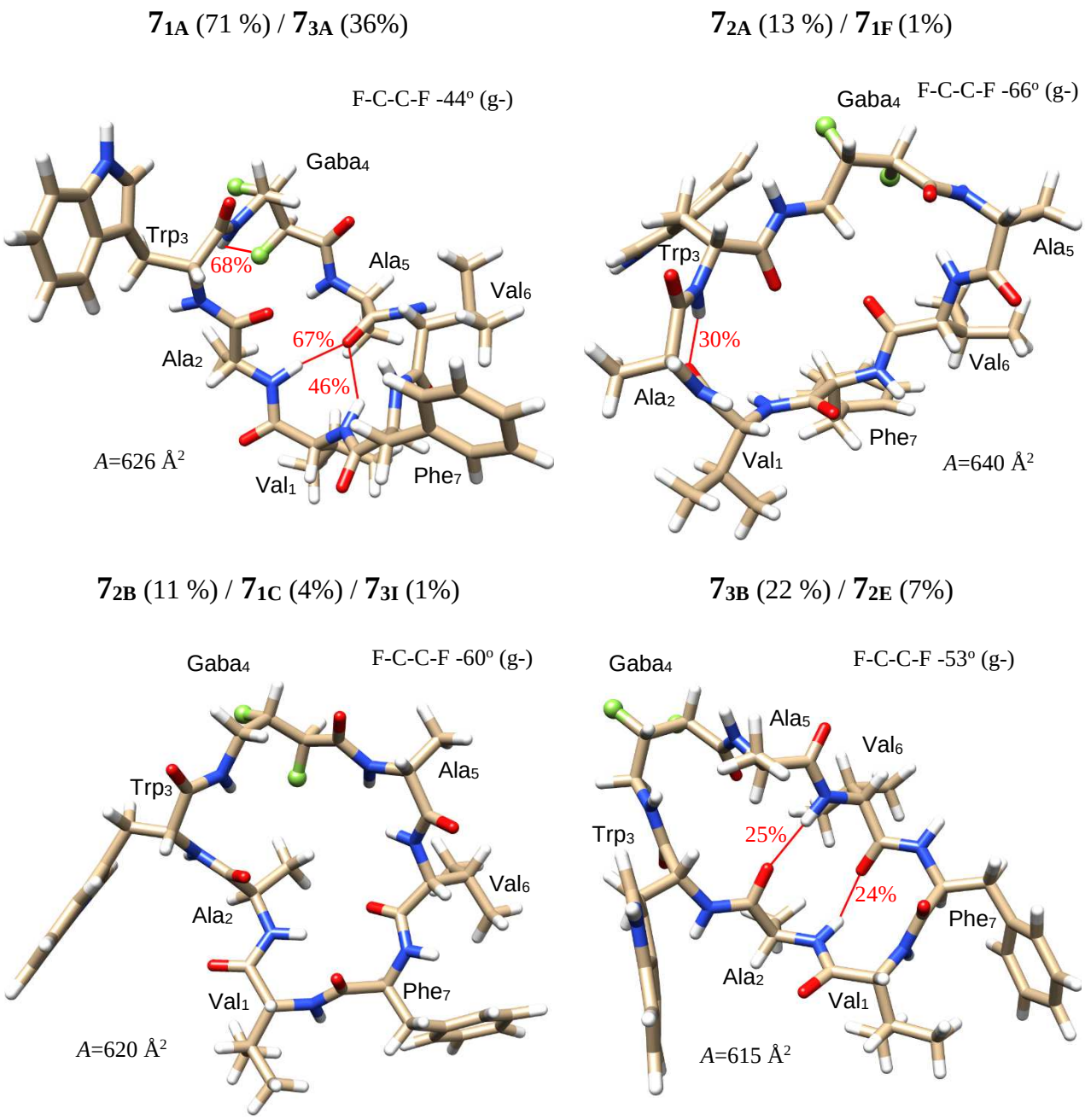


Figure 3. Stick models of the cluster representatives obtained for Unguisin A (**3**) and its four difluorinated derivatives (**4-7**). Fluorine atoms are represented as green balls. The percentage population (%) of each cluster within the corresponding cMD simulation are indicated. The more abundant intramolecular hydrogen bonds are marked with red lines and their abundances (%) in the corresponding trajectory are also displayed. The molecular surface area (A) is also *given.*

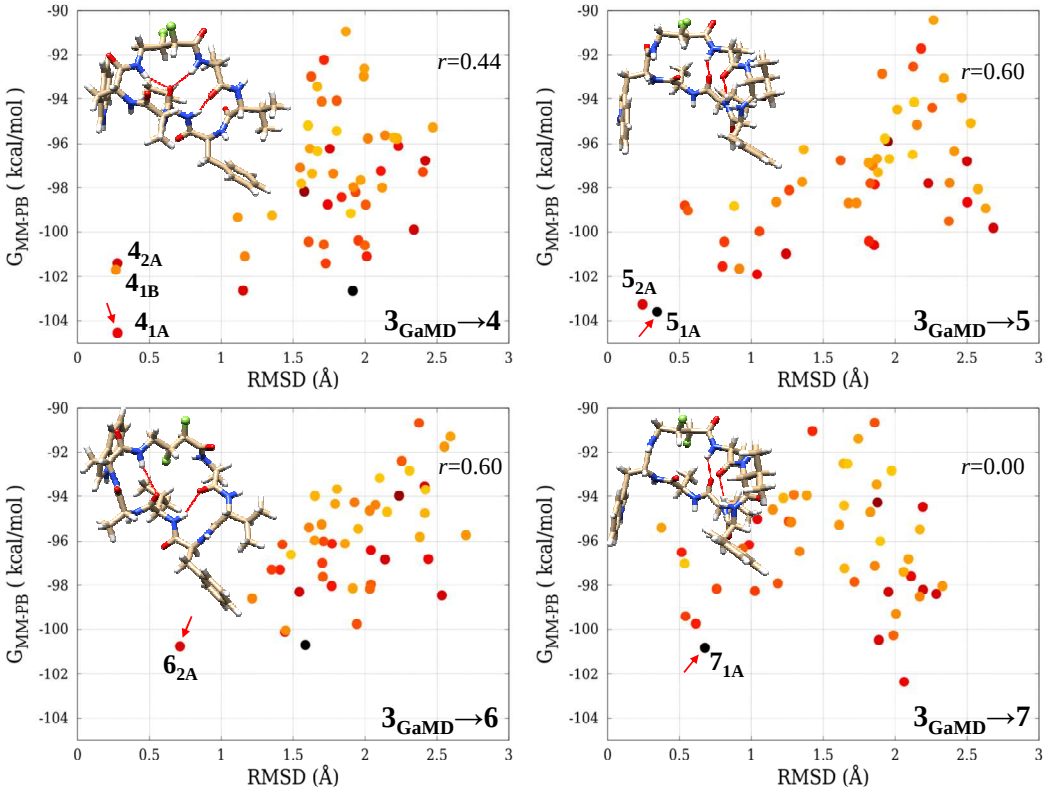


Figure 4. Correlation plots between the average $G_{MM\text{-}PBSA}$ values of the *fluorinated* $\mathbf{3_{GaMD}}$ clusters (*i.e.*, $\mathbf{3_{GaMD}}$→**4-7**, see text for details) and the minimum RMSD distances between the major **4-7** cMD clusters and the $\mathbf{3_{GaMD}}$ cluster representatives. Coloring of the data points denotes logarithmic abundance of the $\mathbf{3_{GaMD}}$ clusters (black=most populated, yellow=least). Pearson correlation coefficients (*r*) are also given. The correspondence of some data points with the structurally-closest **4-7** clusters is indicated for specific data points. The ball-and-stick models display members of the cluster signaled with a red arrow.

For Table of Contents Use Only

Understanding the Conformational Properties of Fluorinated Polypeptide: Molecular Modelling of Unguisin A Natalia Díaz and Dimas Suárez