Critical assessment of force fields accuracy against NMR data for cyclic peptides containing β-amino acids

C. Paissoni,[a]‡ F. Nardelli,[a] S. Zanella,[b] F. Curnis,[c] L. Belvisi,[b] G. Musco[a]* and M. Ghitti[a]*

[a] Biomolecular NMR Unit, IRCCS Ospedale San Raffaele, Via Olgettina 60, 20132 Milan, Italy. E-mail: ghitti.michela@hsr.it, musco.giovanna@hsr.it

[b] Dipartimento di Chimica, Universita` degli Studi di Milano, Via Golgi 19, 20133 Milan, Italy

[c] Tumor Biology and Vascular Targeting Unit, IRCCS Ospedale San Raffaele, Via Olgettina 60, 20132 Milan, Italy

[‡] Current address: Dipartimento di Bioscienze, Universita` degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy.

Hybrid cyclic α/β-peptides, in which one or more β-amino acids are incorporated into the backbone, are gaining increasing interest as potential therapeutics, thanks to their ability to achieve enhanced binding affinities for a biological target through pre-organization in solution. The *in silico* prediction of their three dimensional structure through strategies as MD simulations would substantially advance the rational design process. However, it remains to be verified whether the molecular mechanics force fields are accurate in sampling highly constrained cyclopeptides containing β-amino acids. Here, we present a systematic assessment of the ability of 8 widely used force fields to reproduce 79 NMR observables (including chemical shifts and ^{3}J scalar couplings) on five cyclic α/β-peptides that contain the integrin recognition motif isoDGR. Most of the investigated force fields, which include force fields from AMBER, OPLS, CHARMM and GROMOS families, display very good agreement with experimental $^{3}J(HN,H\alpha)$, suggesting that MD simulations could be an appropriate tool in the rational design of therapeutic cyclic α-peptides. However, for NMR observables directly related to β-amino acids, we observed a poor agreement with experiments and a remarkable dependence of our evaluation on the choice of Karplus parameters. The force fields weaknesses herein unveiled might constitute a source of inspiration for further force fields optimization.

Introduction

Peptides and peptidomimetics represent an important class of compounds capable of modulating protein-protein interactions (PPIs) thanks to their ability to mimic the structure of natural protein ligands.[1] In this context, enhanced binding affinity for a biological target can be achieved through the pre-organization of the free ligand in its bound state, the so-called bioactive conformation.[2,3] Different strategies, such as backbone cyclization, introduction of β-amino acids and chemical modifications, can be applied to restrict the conformational variability of linear peptides and to fine-tune their conformational preferences.[4–6] In particular, the construction of hybrid cyclic α/β-peptides (α/β-CPs) may represent an effective method for obtaining compounds with suitable characteristics for pharmacological applications.[7,8] In fact, their favourable pharmacokinetic properties, as well as their enhanced cell permeability, proteolytic stability, oral availability and target selectivity, make them more promising PPIs inhibitors than linear α-peptides.[9] As an example, the application of α/β-CPs as therapeutics has proven to be extremely successful in the field of integrin inhibitors, where the high receptor affinity and selectivity have been mainly ascribed to their pre-organization in solution.[3,4,10] In particular, α/β-CPs containing the integrin binding motif isoAspartate-Glycine-Arginine (isoDGR), with the isoD β-amino acid, are gaining increasing interest as potential therapeutics or as ligands for drug delivery.[11–17]

Despite the promising applications offered by the rational design of these α/β-CPs, their potentialities are still largely under-explored. A major impediment to the full exploitation of CPs as PPI modulators is related to the inherent difficulty to accurately predict, both experimentally and computationally, their three-dimensional structure and the conformational effects induced by chemical modifications and by the presence of β-amino acids.[3,18–21] Because of these challenges, their optimization for specific biological targets has been mainly based on empirical approaches, requiring extensive and expensive synthesis campaigns. Herein computational techniques could represent a valuable tool to assist the rational design process. However, the ring strain of CPs creates large free energy barriers between different conformations, making conformational sampling through traditional all-atoms molecular dynamics (MD) simulation highly challenging. During the last decades, several advanced sampling techniques aiming at producing converged ensembles of biomolecules and providing reliable estimate of free energy differences have been developed.[22] These advanced conformational sampling strategies comprise thermodynamic integration,[23] free energy perturbation,[24] umbrella sampling,[25] steered molecular dynamics,[26] parallel tempering,[27] and replica exchange molecular dynamics.[28] In this context, we and others previously demonstrated that Metadynamics in its Bias Exchange variant (BE-META)[29,30] provides a perfect framework to exhaustively sample the conformations of CPs, guaranteeing a full coverage of the relevant conformational space.[18,31] Nevertheless, the accuracy of MD simulations could be strongly affected by the reliability of molecular mechanics force fields. These force fields are well established and tested in reproducing the structure and dynamics of proteins and linear peptides,[32–35] but their performance is poorly assessed in the case of highly constrained CPs containing β-amino acids as isoAspartate.

In this work we present a systematic investigation, where we tested the ability of eight popular force fields to reproduce

the conformational properties of five integrin binding isoDGR-based CPs. As shown in **Figure 1**, the five investigated molecules include two pentapeptides cycled through a disulphide bond, CisoDGRC and its acetylated variant acCisoDGRC,[13,36] and three head-to-tail cyclic hexapeptides, c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG).[12] The accuracy of each force field was assessed by comparison with NMR derived observables, that are a commonly used benchmark able to capture the dynamics of molecules in solution.[37] NMR data include chemical shifts and ^{3}J scalar couplings, the latter being particularly relevant in the study of CPs because of their strong dependence on dihedral angles. The collected experimental data were quantitatively compared with the ensemble average values of the same observables retrieved from BE-META simulations using a χ^2 metric. Our results reveal that most of the recently developed force fields reliably reproduce the dynamic of CPs backbone obtaining a good agreement with experimental data. Nevertheless we found that simulations could be not enough accurate in describing conformations and dynamics of β-amino acids.

Materials and Methods

Experimental Data. We used as benchmark a set of five isoDGR-containing CPs (**Figure 1**), including two pentapeptides cycled through disulphide bridge (CisoDGRC and its acetylated variant acCisoDGRC) and three head-to-tail cyclic hexapeptides (c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG)). For the force field evaluation, 79 NMR observables were considered, including: i. 38 chemical shifts of the C, Cα, Cβ, HN and Hα atoms and ii. 41 ^{3}J scalar couplings, subdivided in 31 ^{3}J(HN,Hα), associated to the backbone dihedral angle ϕ and 10 $^{3}J^{isoD}$(Hα,Hβ), associated to the dihedral angle ζ(N-Cα-Cβ-C) characteristic of isoAspartate backbone (**Supplementary Table S1-S6**). Peptides were chemically synthesized in-house as described in [11,12] or purchased (Biomatik, Delaware, USA). For each molecule, 1H-1D, 1H-1D TOCSY (TOtal Correlation SpectroscopY), 1H-2D TOCSY and 2D ROESY (Rotational nuclear Overhauser Effect Spectroscopy) NMR spectra (**Supplementary Figure S1-S5**) were recorded at a temperature of 280-285 K, on a Bruker Avance-600 spectrometer equipped with TCI cryoprobe (Bruker). Further details on the experimental procedure are described in the Supplementary Information.

Force Fields. Eight force fields belonging to AMBER, OPLS, CHARMM and GROMOS families have been used to simulate five different cyclic peptides containing the β-amino acid isoAspartate (**Figure 1**). Due to comparable local chemical environments between the isoAspartic residue and the C-terminal Asparagine (**Supplementary Figure S6**), no new atom types were needed. Accordingly, the atom-types of Asparagine side-chain were assigned to the backbone atoms of isoAspartate and the atom-types of C-terminal carboxylate to the side-chain atoms of isoAspartate. Parameters for Phenylglycine, which is present in c(CphgisoDGRG), were derived from Phenylalanine, applying appropriate dihedral angles. The atomic partial charges of both Phenylglycine and isoAspartate atoms to be used with the AMBER force fields were derived using the R.E.D. III (RESP ESP charge Derive) package.[38] To this aim, two different initial conformations for both Phenylglycine and isoAspartate dipeptides (i.e. the amino acid capped with the acetyl and N-methyl groups) have been

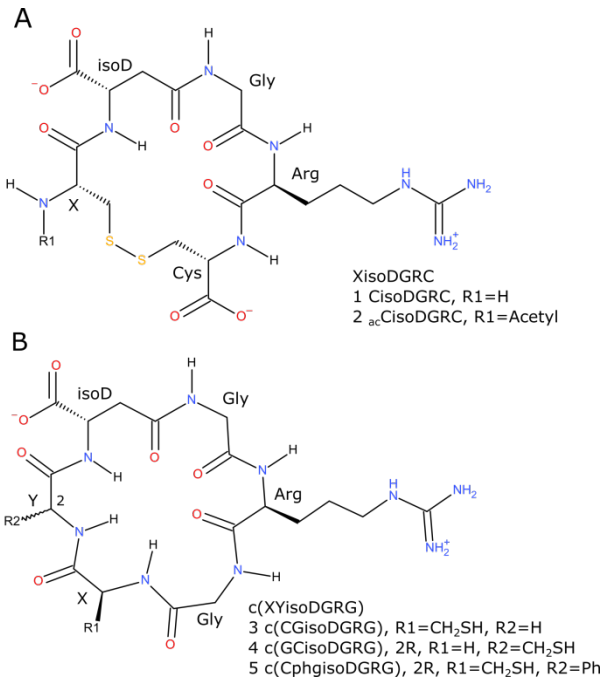

Figure 1. 2D representation of the investigated peptides. A) Pentapeptides CisoDGRC and acCisoDGRC; B) hexapeptides c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG).

generated with Maestro.[39] GAMESS package[40] was used for geometry optimization in the gas phase and for the computation of the molecular electrostatic potentials (MEP), using the HF/6-31G* level of theory and the Connolly surface algorithm. Four different molecular orientations for each optimized geometry have been considered using the rigid-body reorientation algorithm implemented in the R.E.D. tool. Finally, the two-stages RESP method was used for the fitting of the atomic charges following the procedure originally published by Kollman et al..[41] Intra-molecular constraints were imposed to set the charges of the acetyl and N-methyl capping groups to zero; additionally charges equivalences were imposed to hydrogens of methyl and methylene groups, to oxygens of carboxylate ions and to symmetric atoms in the phenyl group. The derived RESP partial charges for Phenylglycine and isoAspartate atoms are reported in **Supplementary Table S7**.

More specifically, the force fields employed include: AMBER ff99sb,[42] AMBER ff99sb-ildn,[43] AMBER ff99sb*,[44] AMBER ff14sb,[45] OPLS-AA/L,[46] its variant that we denoted as OPLS-AA/L_{STD}, CHARMM-27[47] and GROMOS-54a7.[48] OPLS-AA/L_{STD} differs from OPLS-AA/L simply for the torsional parameters used to describe the dihedral angle of isoAspartic

backbone, where parameters associated to the Asparagine atom-types described in [49] and in [46] were adopted respectively. It is worthwhile noting that the two AMBER force fields ff99sb and ff99sb-ildn only differ in the torsional parameters describing the Isoleucine, Leucine, Aspartic Acid and Asparagine residues, which are not included in the tested cyclic peptides. Therefore, in this work, we simply tested to what extent the ff99sb-ildn torsional parameters optimized for Asparagine side-chain could be transferred to isoAspartate dihedral angles. For the water molecules topology, the TIP3P model was used in all the simulations with the exception of the ones performed using the GROMOS-54a7 force field, where the SPC water model was preferred.[50,51]

Simulations Details. The initial structures of CisoDGRC and $_{ac}$CisoDGRC were generated with CNS (Crystallography & NMR System) program,[52] treating the amino and carboxyl terminals consistently with a pH value of 6.5. For the cyclic head-to-tail hexapeptides, an initial structure of c(GGisoDGRG) was generated using the Maestro 2D sketcher tool and energy minimized;[39] then CNS program was employed to replace Glycine residues with the appropriate amino acids to generate c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG). Each system was solvated in a cubic box and neutralized. The dimensions of the box were chosen so that the distance between any peptide atom and the box edges was at least 1.2 nm. The initial energy minimization has been followed by a three-stages equilibration procedure consisting in: i. 2 ns equilibration in the NVT ensemble, where the v-rescale thermostat[53] was used to maintain temperature at 280 K; ii. 2 ns in the NPT ensemble, where Berendsen barostat[54] and v-rescale thermostat were employed to control pressure and temperature (1 bar and 280 K); iii. 4 ns in the NPT ensemble where the Berendsen barostat was replaced by the Parrinello-Rahman one[55] (with the exception of simulations performed with GROMOS-54a7 in which Berendsen barostat was maintained). Positional restraints on the peptides' heavy atoms were employed in the first two steps of equilibration and then released in the last stage. The relaxation time for the barostat was of 1 ps while the thermostat coupling time constant was of 100 fs (CPs and solvent molecules were coupled to independent thermostats). During the equilibration the LINCS algorithm was applied to constraint all the bond lengths.[56] In all the cases the md integrator was used with a time step of 2 fs. A cut-off of 1.0 nm was used to truncate both the van der Waals and the electrostatic interactions; long range electrostatic interactions beyond the cut-off were treated with Particle-Mesh Ewald method (Fourier grid spacing of 0.12 nm and interpolation order of 4).[57] BE-META simulations were performed in the NVT ensemble, setting the temperature at 280 K by means of the v-rescale thermostat (coupling time constant of 100 fs) and using the same settings described for equilibration. During this production run all the bond lengths involving hydrogen atoms have been constrained. Gaussian hills were added every 1 ps (height: 0.24 kcal mol^{-1}, width: 0.2 rad) and exchanges between pairs of replica were attempted every 15 ps, with an acceptance rate between 15% and 25%. In the five molecules all the dihedral angles describing the cycle, with the exception of the ones associated to the planarity of the peptide bonds, have been used as collective variables (CVs), resulting in the bias of 13, 14 and 15 CVs for the three cyclic head-to-tail hexapeptides, CisoDGRC and $_{ac}$CisoDGRC, respectively. A complete list of CVs is reported in **Supplementary Table S8**. During the BE-META simulations, each of the listed CVs was biased in one replica, with the length of each replica ranging from 30 to 60 ns. The convergence was checked using the Metagui tool,[58] comparing the mono-dimensional free-energy profiles derived by averaging on the two halves of the simulation, after an equilibration time. For all the simulations, the equilibration time was comprised between 1 and 5 ns. In all the cases the two profiles were found to be consistent within $2k_bT$. All the MD simulations have been performed using the Gromacs-5.0.4 software[59] and the Plumed 2.1.3 plugin.[60]

Back-calculation of ^{3}J couplings and chemical shifts. To recover the equilibrium properties of the simulated systems from BE-META trajectories, we followed the method proposed by Laio and coworkers.[61] Herein the structures visited during the simulations were clustered in microstates according to the values adopted by a reduced number of CVs (**Supplementary Table S8**), then the free-energy of each microstate was estimated applying a WHAM procedure with the help of the Metagui tool.[58,62] For each microstate μ, the arithmetic average (X_μ) of the observable X (which could be either the back-calculated ^{3}J scalar couplings or chemical shifts values) over all the clustered structures have been computed. Finally, the average equilibrium value of each observable X has been calculated according to:

$$\langle \boldsymbol{X} \rangle = \frac{\sum_{\mu} \boldsymbol{X_\mu} e^{-F_\mu / k_b T}}{\sum_{\mu} e^{-F_\mu / k_b T}},$$

where F_μ is the free-energy of microstate μ, k_b is the Boltzmann constant, T is the temperature and X_μ is the arithmetic average of the observable X in the microstate μ. For each of the structures clustered in a microstate, the Karplus relation[63] $J(\theta) = A\cos^2(\theta) + B\cos(\theta) + C$ and the Sparta+ program[64] were used to back-calculate the scalar couplings (^{3}J(HN,Hα) and ^{3}J^{isoD}(Hα,Hβ)) and the chemical shifts values, respectively. Since the choice of the empirical Karplus parameters could influence our evaluation, different sets of Karplus parameters have been employed. In particular, ^{3}J(HN,Hα) scalar couplings were calculated using two sets of parameters, denoted as ORIG[65] and DFT,[66] following the example of [33,49]. ^{3}J^{isoD}(Hα,Hβ) were back-calculated with three sets of Karplus parameters: the ones of Cung et al.,[67,68] the widely used parametrization of De Marco et al.[69,70] and the more recent Asparagine-specific parametrization of Perez et al.[71] (**Supplementary Table S9**). Errors of 0.70 Hz and 1.00 Hz

were associated to the back-calculation via Karplus parameters of ^{3}J(HN,Hα) and $^3J^{isoD}$(Hα,Hβ), respectively. The chemical shifts of C, Cα, Cβ, HN and Hα atoms were back-calculated from each of the structures clustered in microstates using the Sparta+ program, whose applicability is limited to natural residues which are not neighbours of non-standard amino acids (i.e. Phenylglycine and isoAspartate). The systematic errors associated to these chemical shift predictions are the ones determined in the original paper[64] and reported in **Supplementary Table S10**.

Evaluation of force-fields accuracy. The accuracy of each force field was quantified estimating the agreement between back-calculated and experimental data according to:

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N} \frac{\left(\langle X_i \rangle_{comp} - X_{i,exp}\right)^2}{\sigma_i^2},$$

where X can be either scalar couplings or chemical shifts, N is the total number of experimental data used and σ_i is the systematic uncertainty, arising from the back-calculation of NMR observables from peptides conformations, associated to the value X_i. The χ^2 statistic is a widely used metric that allows to quantitatively compare expected data with computed ones.[32,33] A very small χ^2 test statistic (close to 1) means that computational data fits experimental data extremely well. $\chi^2 \leq 2$ indicates that the deviation from experiment is acceptable, as reported in other studies.[33,72] Finally, a very large χ^2 test statistic indicates that the agreement between experimental and computational data in not satisfactory. The standard deviation of the χ^2 values was estimated splitting in two halves each simulation and computing the deviation between the χ^2 values associated to the first or the second halves.

To further improve the precision of the standard deviations all the possible combinations of first/second halves of the simulations were considered and average standard deviations were computed. Additionally, in order to estimate the robustness of our results with respect to the set of CPs chosen as benchmark, we adopted a Jack-Knife procedure, systematically excluding in the computation of χ^2 the experimental data associated to one molecule. The Jack-Knife standard error was computed as:

$$SE_{JK} = \sqrt{\frac{M-1}{M}\sum_{i=1}^{M}\left(\chi^2_{(i)} - \chi^2_{(.)}\right)^2},$$

where M=5 is the number of molecules considered in the test,

$\chi^2_{(i)} = \frac{\sum_{j\neq i}^{M}\chi_j^2}{M-1}$, χ_i^2 is the χ^2 computed on the experimental data of molecule *i* and $\chi^2_{(.)} = \frac{\sum_{i=1}^{M}\chi^2_{(i)}}{M}$.

Results

In this work we have evaluated the ability of 8 widely used force fields to reproduce experimental equilibrium properties (chemical shifts and ^{3}J scalar couplings) of 5 isoDGR-based CPs, analysing 40 BE-META simulations for a total of ~23.8 μs. The tested force fields include 4 AMBER variants (ff99sb, ff99sb-ildn, ff99sb*, ff14sb), 2 OPLS variants (OPLS-AA/L, OPLS-AA/L_{STD}), CHARMM-27 and GROMOS-54a7. The performance of each force field was evaluated using the χ^2 metric, where the deviations between experimental and computed values were weighted with the uncertainty associated to these observables.

Chemical shifts reproducibility. We firstly evaluated the ability of the considered force fields to reproduce chemical shift values, using a set of 38 experimental data (**Supplementary Table S1, S11**). For all the force fields, we found small deviations between predicted and experimental chemical shifts, obtaining χ^2 values ranging from 0.50 to 0.75 (**Figure S7**). Herein we observed that most of the considered chemical shifts values are compatible with the ones characteristic of a random coil conformation, consistent with the fact that these CPs mainly adopt turn-like structures devoid of secondary structure elements. Conceivably, these random coil-like chemical shifts can be well-reproduced in MD simulations, irrespective of the employed force field. As the agreement between back-calculated and measured chemical shifts is similar for all the force fields and since similar χ2 values are not indicative of similar structural ensembles, this comparison has not been considered for further analysis.

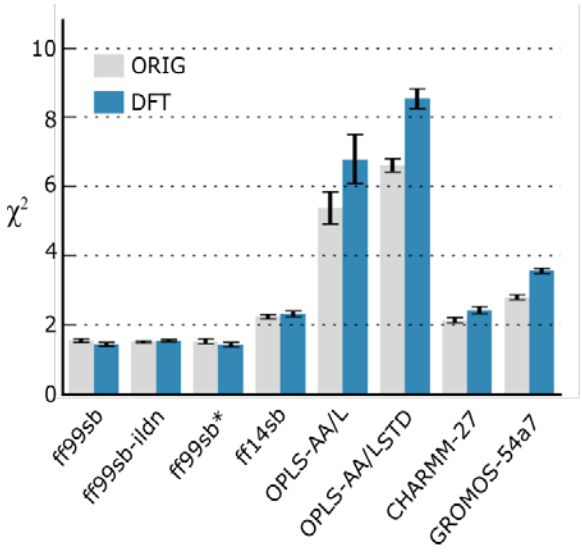

Figure 2. χ^2 values, estimating the ability of each force field to reproduce experimental ^{3}J(HN,Hα) scalar couplings, computed with both the ORIG (light gray bars) and the DFT (blue bars) sets of Karplus parameters. The error bars represent the standard deviation of the χ^2 values obtained analysing separately the first and second halves of each simulation, as described in Material and Methods.

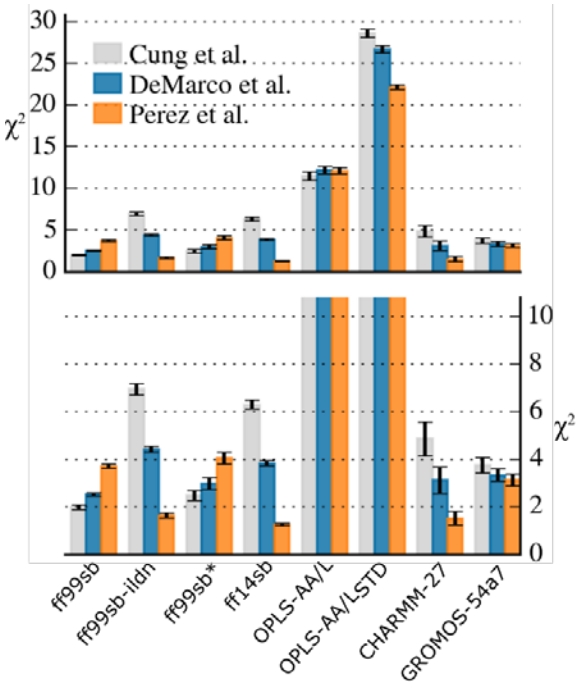

Figure 3. χ^2 values, estimating the ability of each force field to reproduce experimental ^{3}J(Hα,Hβ) scalar couplings, computed with the Karplus parameters developed by Cung et al. (light gray bars), De Marco et al. (blue bars) and Perez et al. (orange bars) are shown. The error bars represent the standard deviation of the χ^2 values obtained analysing separately the first and second halves of each simulation, as described in Material and Methods. To avoid flattening of the data a zoom on the χ^2 range [0:10] is displayed in the lower panel.

^{3}J(HN,Hα) couplings reproducibility. We next focused our attention on the comparison between experimental and back-calculated ^{3}J scalar couplings. To this aim, we analysed separately the ^{3}J(HN,Hα) and the $^3J^{isoD}$(Hα,Hβ) associated to the ζ(N-Cα-Cβ-C) dihedral angle of isoAspartate backbone (**Figure S6**). While the reproducibility of the first observables provides general information on force field accuracy in CPs simulations, the latter are more focused on the appropriateness of current force fields to treat β-amino acids, such as the isoAspartic residue. Concerning the reproducibility of ^{3}J(HN,Hα) couplings, our results reveal that the three AMBER force fields ff99sb, ff99sb-ildn, ff99sb* can accurately reproduce ^{3}J(HN,Hα) of CPs, displaying χ^2 values lower than 1.6 (**Figure 2, Supplementary Table S12**); satisfactory performances are achieved by the recently optimized AMBER ff14sb, CHARMM-27 and GROMOS-54a7 (χ^2<3). Conversely, we observed that both the variants of the OPLS-AA/L force field display a lower agreement with experiments, in line with previous observations reporting the poor ability of this force field to reproduce observables relying on torsional energetics.[32,34,73] Of note, analogous conclusions can be drawn when the ^{3}J(HN,Hα) couplings are back-calculated with either the ORIG or the DFT sets of Karplus parameters, suggesting that our force fields evaluation is robust with respect to Karplus parameters variations (**Figure 2**). We then investigated whether our assessment could depend on the choice of CPs used as benchmark. Herein, considering the Jack-Knife standard error, the differences between the performances of the tested force fields remain significant, suggesting that our evaluation is only negligibly affected by the selective exclusion of one CP from the data-set (**Figure S8**). Finally, in order to verify if the subset of ^{3}J(HN,Hα) scalar couplings of isoAspartate could affect the overall force fields performance, we excluded these data from the computation of the χ^2 values (**Figure S9**). Since no significant differences can be observed after this exclusion, we conclude that the ^{3}J(HN,Hα) of β-amino acids can be back-calculated with an accuracy comparable to the one of the scalar couplings of α-amino acids. For sake of completeness, we also reported the comparison between back-calculated and experimental ^{3}J(HN,Hα) using linear least-squares regression (**Figure S10**). We excluded from the comparison the two OPLS-AA/L variants, as they always show the worst performance. AMBER ff99sb, ff99sb-ildn, ff99sb* display a satisfactory r^2 value (>0.6), in line with the results reported in [35]. AMBER ff14sb and CHARMM-27 show a r^2 value between 0.5 and 0.6; while GROMOS-54a7 displays a lower agreement with a r^2<0.5. These results are well consistent with the ones obtained through the χ^2 metric. We then dissected the contributions of the single CP and of the single amino acid to the low agreement observed for OPLS force fields. We observed the highest deviations for glycine and arginine and found that the poor correlation between computed and experimental ^{3}J(HN,Hα) does not depend on the specific CP (**Figure S11**).

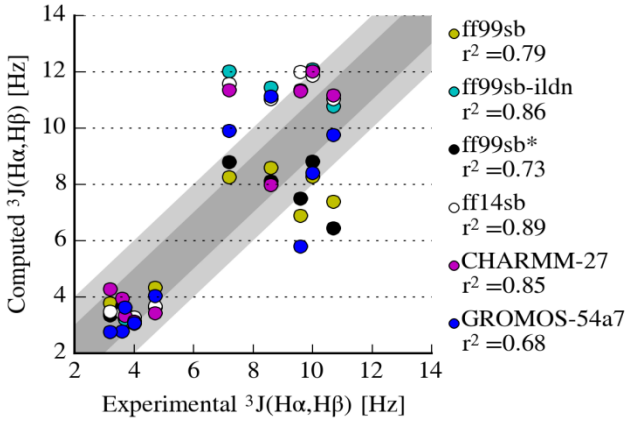

Figure 4. Comparison of the ^{3}J(Hα,Hβ) scalar couplings measured experimentally and back-calculated using the Karplus parameters developed by De Marco et al.. The dark and light grey shadows indicate a deviation of ±1 Hz and ±2 Hz from experimental data, respectively. The r^2 value for linear regression of each force field is also reported.

^{3}J(Hα,Hβ) couplings reproducibility. We then assessed the reproducibility of $^{3}J^{isoD}$(Hα,Hβ) couplings, associated to the ζ(N-Cα-Cβ-C) dihedral angle of isoAspartate backbone. In this case we observed a poor agreement with experiments and a remarkable dependence of our results on the choice of Karplus parameters (**Figure 3, Supplementary Table S13**). Overall, independently of the Karplus parametrization used, we found that OPLS-AA/L is not appropriate in sampling isoAspartate dynamics. As far as the other force fields is concerned, the results strongly depend on the Karplus parametrization: ff99sb and ff99sb* outperform all the other force fields when the Cung parameters are adopted; ff99sb, ff99sb*, CHARMM-27 and GROMOS-54a7 display comparable performance using the De Marco parametrization. Finally, ff99sb-ildn, ff14sb and CHARMM-27 force fields display the best performance with Perez parameters. The strong dependence of the back-calculated ^{3}J(Hα,Hβ) couplings on the type of Karplus parameters relies on the conspicuous variations encoded in the empirical Karplus relations, as previously discussed in [67,74] This discrepancy is particularly evident at the curve global maximum that can vary from 11.0 Hz (Perez) to 13.9 Hz (Cung) (**Figure S12**). In this context it is hard to evaluate which of the considered Karplus parametrizations is more appropriate. The Perez parameters have been developed considering only NMR data for proteins and are amino acid specific, representing a considerable improvement with respect to earlier methods. However, because of conformational averaging this curve is considerably flatter than the others (**Figure S12**), resulting in several experimental ^{3}J-coupling values lying outside the range of scalar couplings allowed.[49,67] Therefore, based on these considerations, here we present the results deriving from the widely used De Marco parametrization, which is placed in the middle of the three Karplus relations. For sake of completeness we have reported in the Supplementary Materials the results obtained with the other parameters.

The force fields evaluation based on the $^{3}J^{isoD}$(Hα,Hβ) is also influenced by the set of molecules used as benchmark, as suggested by the larger Jack-Knife standard errors (**Figure S13**) compared to the ones obtained for the ^{3}J(HN,Hα) (**Figure S8**). To get more insights into the reproducibility of $^{3}J^{isoD}$(Hα,Hβ) scalar couplings, we compared the back-calculated and the experimental ones through linear least-squares regression (**Figure 4 and S14**), excluding the two OPLS-AA/L variants that always show the worse performance. All the considered force fields display a satisfactory r^2 value (>0.6) and distinguish between small (<5 Hz) and big (>7 Hz) scalar ^{3}J couplings. However, when considering big scalar ^{3}J couplings alone, the correlation is poor and the deviations are high, suggesting that none of the considered force fields is able to accurately reproduce the conformational equilibrium of the ζ(N-Cα-Cβ-C) isoAspartate dihedral angle.

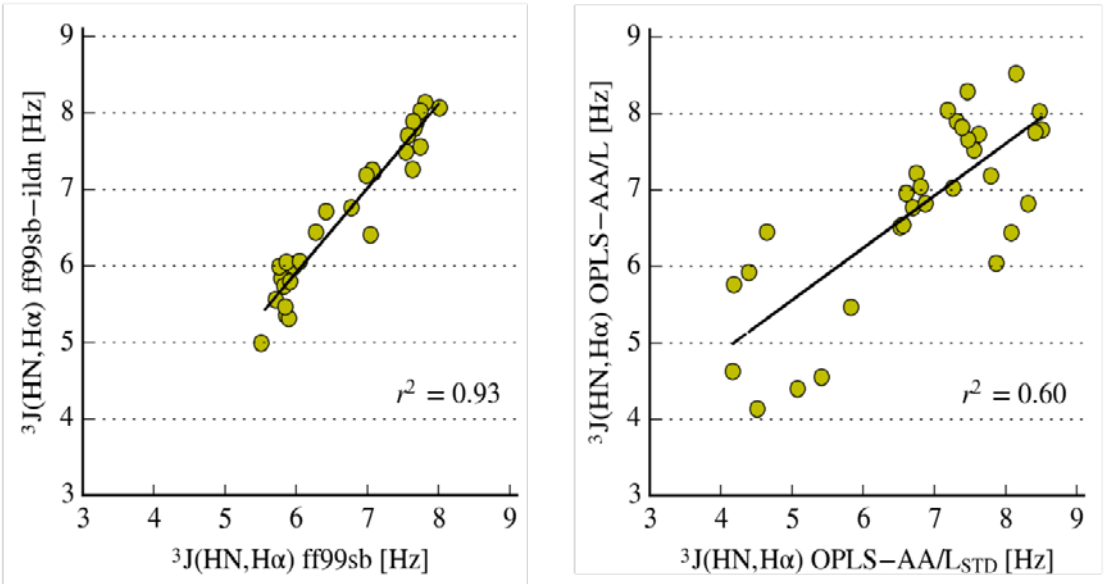

Figure 5. Correlation between ^{3}J(HN,Hα) scalar couplings computed with: i) AMBER ff99sb or ff99sb-ildn force fields, left panel; and ii) OPLS-AA/L_{STD} or OPLS-AA/L force fields, right panel. All the ^{3}J(HN,Hα) scalar couplings have been back-calculated using the ORIG set of Karplus parameters.

Influence of isoAspartate parametrization on force field accuracy. Finally, we asked whether the choice of torsional parameters associated to isoAspartate backbone dihedral angle ζ(N-Cα-Cβ-C) is limited to the local dynamic of the β-amino acid, or whether it influences the overall conformations herewith affecting the back-calculated scalar couplings of the adjacent amino acids. To this aim we compared the AMBER ff99sb and OPLS-AA/L_{STD} force fields, with the corresponding variants AMBER ff99sb-ildn and OPLS-AA/L, which differ from the previous ones only for the torsion potentials describing the backbone isoAspartate dihedral angles N-Cα-Cβ-C and Cα-Cβ-C-N_{i+1}. Correlations between the ^{3}J(HN,Hα) couplings computed with the two couples of force fields highlight different effects (**Figure 5 and S15**). The ^{3}J(HN,Hα) scalar couplings predicted with the two AMBER force fields are highly correlated (r^2=0.93), suggesting that the different torsional potentials employed for isoAspartate dihedral angles have only a local effect without any influence on the surrounding residues. Conversely, the isoAspartate parametrization in the OPLS-AA/L force fields has long-range effects, resulting in a poor correlation (r^2=0.60) between the ^{3}J(HN,Hα) scalar couplings computed with the two force field variants. To identify the residues mostly affected by the different OPLS-AA/L parameters, we repeated the same analysis systematically excluding subsets of ^{3}J(HN,Hα) couplings **(Figure S16).** We observed that the two residues following isoAspartate are the ones mostly affected by the OPLS-AA/L parameters variation, being most likely responsible of the overall low correlation (**Figure S16, Panel F**). However when we re-computed the OPLS-AA/L χ^2 values excluding the ^{3}J(HN,Hα) scalar couplings of the one/two residues following isoAspartate, we did not detect any relevant variation (**Figure 6 and S17**). These results suggest that the poor agreement between computational and experimental ^{3}J couplings displayed by OPLS-AA/L force fields is independent from the presence of isoAspartate. Conceivably, the bad performance might be ascribed to the cyclic constraint and/or to intrinsic defects of the force field, mainly related to torsional energetics dependent observables.[32,34,73]

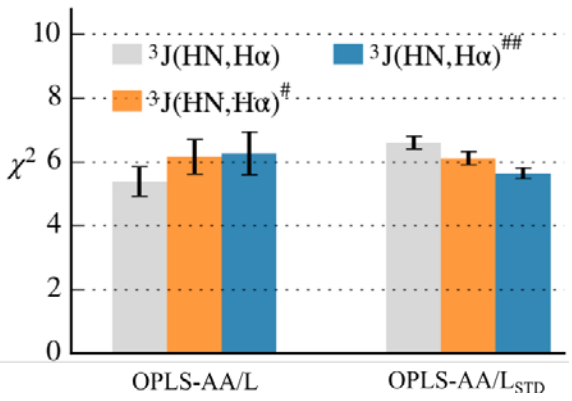

Figure 6. For the OPLS-AA/L and OPLS-AA/L_{STD} force fields, are reported the χ^2 values, estimating the ability to reproduce: i. all the experimental ^{3}J(HN,Hα) couplings (grey bars); ii. all the ^{3}J(HN,Hα) couplings except the ones of the residue following isoAspartate (^{3}J(HN,Hα)$^{\#}$, orange bars); iii. all the ^{3}J(HN,Hα) couplings except the ones of the two residues following isoAspartate (^{3}J(HN,Hα)$^{\#\#}$, blue bars). The error bars represent the standard deviation of the χ^2 values obtained analysing separately the first and second halves of each simulation, as described in Material and Methods. All the ^{3}J(HN,Hα) scalar couplings have been back-calculated using the ORIG set of Karplus parameters.

Discussion

Peptides constitute an important class of compounds able to target protein-protein interactions. Their conformational equilibrium, receptor specificity and binding affinity along with resistance to proteases degradation can be modulated through different strategies, ranging from cyclization to introduction of chemical modifications and inclusion of β-amino acids.[4,7,8] Herein, the characterization of hybrid α/β-CPs is experimentally a difficult task as they adopt multiple conformations in solution. *In silico* methods, such as BE-META, allowing an exhaustive description of CPs conformational landscape represent a valid support to their development as therapeutics.[18,31] However, it remains to be verified whether the current force fields are appropriate to predict the three-dimensional structure of highly constrained CPs.

In this paper we have addressed this issue, systematically testing the ability of eight force fields (AMBER ff99sb, ff99sb-ildn, ff99sb*, ff14sb, OPLS-AA/L, OPLS-AA/L_{STD}, CHARMM-27 and GROMOS-54a7) to reproduce experimental chemical shifts and ^{3}J scalar couplings of five CPs containing the isoAspartate β-amino acid. Our evaluation of chemical shifts reproducibility appears less informative than expected, as all the force fields perform equally well. This is probably due to the flexibility of the investigated peptides and to the absence of stably populated secondary structures, with experimental chemical shifts very similar to random coil values. The comparison of experimental and back-calculated ^{3}J scalar couplings, comprising ^{3}J(HN,Hα) and $^3J^{isoD}$(Hα,Hβ), is by far more informative. Herein we found that most of the considered force fields can well reproduce the experimental ^{3}J(HN,Hα) scalar couplings, including the ^{3}J(HN,Hα) of the β-amino acid. Of note, this result is robust with respect to both the choice of Karplus parameters and the benchmarks used. The AMBER force fields ff99sb, ff99sb-ildn and ff99sb* show the best performances in terms of ^{3}J(HN,Hα) χ^2 values. Good agreement is also achieved by the recent AMBER ff14sb, CHARMM-27 and GROMOS-54a7. Conversely, the OPLS-AA/L force fields showed the worst performance in reproducing experimental ^{3}J(HN,Hα), probably because of limitations in reproducing quantities related to torsional energetics and/or in simulating peptides with cyclic constrained conformations. A significant improvement in this sense is expected to be observed using the new OPLS-AA/M force field, in which the problem of the scarce reproducibility of torsional-related quantities was specifically addressed.[66] Unfortunately, this improved force field was not included in our test, since it is not yet available for GROMACS software. In this context, it is worthwhile

mentioning the recently developed RSFF2 force field,[35] which has a better agreement with experimental X-ray structures of CPs as compared to OPLS-AA/L and AMBER ff99sb-ildn. The optimization of some parameters for specific amino acids offered by RSFF2 represents an important improvement in the field of CPs simulations. Unfortunately, these improvements could not be exploited in our systems, as they contain non-natural amino acids d-phenylglycine and β amino acids (i.e. isoAspartic acid). Further evaluations of OPLS-AA/M and RSFF2 performance with CPs containing β-amino acids could be very informative. The analysis of $^{3}J^{isoD}$(Hα,Hβ) couplings reproducibility is less straightforward, as our evaluation of force fields performances is highly compromised by the roughness of the Karplus parameters and by a strong dependence of the results upon the molecules included in the benchmark. Despite all these aspects, our data clearly indicate that none of the considered force fields is able to accurately reproduce $^{3}J^{isoD}$(Hα,Hβ) couplings within α/β-CPs. These scalar couplings are relevant indicators of the conformation adopted by the isoAspartate ζ dihedral angle, which can mainly populates three minima around -60°, +60° and 180° (**Figure S12,** lower panel). In our NMR experiments, all the five CPs present $^{3}J^{isoD}$(Hα,Hβ1) and $^{3}J^{isoD}$(Hα,Hβ2) values higher than 7 Hz and lower than 5 Hz, respectively, indicating that the ζ dihedral angle should be in equilibrium between the ζ=-60° and ζ=+60° conformations (**Figure S12**, lower panel). Even if the majority of the investigated force fields correctly discriminates between big and small $^{3}J^{isoD}$(Hα,Hβ) scalar couplings, the correlation between computational and experimental ^{3}J scalar couplings >7 Hz is poor. In particular, despite the wide range of the experimental $^{3}J^{isoD}$(Hα,Hβ1) couplings measured for different CPs (between 7.2 and 10.7 Hz), each force field mainly reproduces its own specific $^{3}J^{isoD}$(Hα,Hβ1) value, independently from the type of molecule. Collectively, the experimental observations suggest that the five CPs should adopt different conformational equilibrium between the ζ=-60° and ζ=+60°, that the force fields are not able to reproduce. Thus we conclude that, most likely, the equilibrium between the multiple rotameric states of the β-amino acid ζ dihedral angle observed in our simulations is mainly driven by the bonded interactions (especially by the torsional energetics) rather than by the non-bonded interactions. We therefore infer that the force field torsional parameters might not be sufficient to describe the observed experimental variability and that new force fields optimization studies should refine both torsional-related parameters and non-bonded interactions.

Conclusions

In this work we have performed a critical assessment of the reliability of commonly used force fields in the characterization of α/β-CPs conformational equilibrium. Our data show that ^{3}J(HN,Hα) scalar couplings are accurately reproduced, suggesting that α-CPs can be reliably sampled with explicit solvent BE-META simulations, provided that a sufficiently accurate force field is selected. Importantly, we observed that all the force fields are less accurate in the description of the local properties of β-amino acids, such as isoAspartate. In this context, QM calculations, though computationally expensive, could be a valuable strategy to get deeper insights into the structural preferences of these CPs. It is also worth noting that force fields evaluation in terms of ^{3}J(Hα,Hβ) scalar couplings reproducibility has been hampered by the large uncertainties enclosed in the Karplus relation, whose parameters may also contain systematic errors. Supplementary investigations, aiming at the development of more appropriate ^{3}J(Hα,Hβ) Karplus parametrizations are therefore highly warranted. They are expected to contribute to both force fields optimization and to their evaluation with experimental data. In conclusion, our study has highlighted the major weaknesses of molecular mechanic force fields in accurately predicting the conformational equilibrium of CPs containing β-amino acids. We believe that these results might inspire new force fields optimization studies aiming at the refinement of both torsional-related parameters and non-bonded interactions. In particular, these improvements should be extremely relevant for isoDGR containing CPs, as the ζ(N–Cα–Cβ–C) dihedral angle value describes the orientation of the carboxylic group of the side chain of the isoAspartic acid with respect to the plane of the cyclopeptide's backbone and only a specific orientation of the carboxylate allows appropriate interactions within the target binding pocket.[13] More generally, in the field of the rational design of CPs for therapeutic purposes, these improvements are urgently needed to foster the use of MD simulation as an invaluable tool for the successful conformational prediction of α/β-CPs.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The Italian Ministry for Health (RF-2011-02350836 to GM) and Asosciazione Italiana Ricerca sul Cancro (IG-17468 to GM) are gratefully acknowledged for financial support. Professor Angelo Corti is sincerely acknowledged for the useful discussions.

References

1 H. G. Albericio, Fernando; Kruger, *Future Med. Chem.*, 2012, **4**, 1527–1531.

2 M. Pelay-Gimeno, A. Glas, O. Koch and T. N. Grossmann, *Angew. Chemie - Int. Ed.*, 2015, 54, 8896–8927.

3 A. E. Wakefield, W. M. Wuest and V. A. Voelz, *J. Chem. Inf. Model.*, 2015, **55**, 806–813.

4 T. Weide, A. Modlinger and H. Kessler, *Top. Curr. Chem.*, 2007, **272**, 1–50.

5 J. Chatterjee, F. Rechenmacher and H. Kessler, *Angew. Chemie - Int. Ed.*, 2013, 52, 254–269.

6 E. W. Guthöhrlein, M. Malešević, Z. Majer and N. Sewald, *Biopolym. - Pept. Sci. Sect.*, 2007, **88**, 829–839.

7 C. Cabrele, T. A. Martinek, O. Reiser and L. Berlicki, *J. Med. Chem.*, 2014, **57**, 9718–39.

8 R. De Marco, A. Tolomelli, E. Juaristi and L. Gentilucci, *Med. Res. Rev.*, 2016, **36**, 389–424.

9 K. Fosgerau and T. Hoffmann, *Drug Discov. Today*, 2015, **20**, 122–128.

10 L. Belvisi, T. Riccioni, M. Marcellini, L. Vesci, I. Chiarucci, D. Efrati, D. Potenza, C. Scolastico, L. Manzoni, K. Lombardo, M. A. Stasi, A. Orlandi, A. Ciucci, B. Nico, D. Ribatti, G. Giannini, M. Presta, P. Carminati and C. Pisano, *Mol. Cancer Ther.*, 2005, **4**, 1670–1680.

11 F. Curnis, R. Longhi, L. Crippa, A. Cattaneo, E. Dondossola, A. Bachi and A. Corti, *J. Biol. Chem.*, 2006, **281**, 36466–36476.

12 F. Curnis, A. Sacchi, R. Longhi, B. Colombo, A. Gasparri and A. Corti, *Small*, 2012, **9**, 1–6.

13 A. Spitaleri, M. Ghitti, S. Mari, L. Alberici, C. Traversari, G.-P. Rizzardi and G. Musco, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 1832–1836.

14 M. Mingozzi, A. Dal Corso, M. Marchini, I. Guzzetti, M. Civera, U. Piarulli, D. Arosio, L. Belvisi, D. Potenza, L. Pignataro and C. Gennari, *Chem. - A Eur. J.*, 2013, **19**, 3563–3567.

15 M. Ghitti, A. Spitaleri, B. Valentinis, S. Mari, C. Asperti, C. Traversari, G. P. Rizzardi and G. Musco, *Angew. Chemie - Int. Ed.*, 2012, **51**, 7702–7705.

16 A. Bochen, U. K. Marelli, E. Otto, D. Pallarola, C. Mas-Moruno, F. S. Di Leva, H. Boehm, J. P. Spatz, E. Novellino, H. Kessler and L. Marinelli, *J. Med. Chem.*, 2013, **56**, 1509–1519.

17 S. Zanella, S. Angerani, A. Pina, P. López Rivas, C. Giannini, S. Panzeri, D. Arosio, M. Caruso, F. Gasparri, I. Fraietta, C. Albanese, A. Marsiglio, L. Pignataro, L. Belvisi, U. Piarulli and C. Gennari, *Chem. - A Eur. J.*, 2017, **23**, 7910–7914.

18 H. Yu and Y.-S. Lin, *Phys. Chem. Chem. Phys.*, 2015, **17**, 4210–4219.

19 Z. Gattin, J. Zaugg and W. F. van Gunsteren, *ChemPhysChem*, 2010, **11**, 830–835.

20 H. Geng, F. Jiang and Y. D. Wu, *J Phys Chem Lett*, 2016, **7**, 1805–1810.

21 S. E. Allen, N. V. Dokholyan and A. A. Bowers, *ACS Chem. Biol.*, 2016, **11**, 10–24.

22 V. Spiwok, Z. Sućur and P. Hošek, *Biotechnol. Adv.*, 2014, **1977**, 1–11.

23 E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, *Chem. Phys. Lett.*, 1989, **156**, 472–477.

24 P. A. Bash, U. C. Singh, R. Langridge and P. A. Kollman, *Science*, 1987, **236**, 564–8.

25 G. N. Patey and J. P. Valleau, *J. Chem. Phys.*, 1975, **63**, 2334–2339.

26 J. Gullingsrud, R. Braun and K. Schulten, *J. Comput. Phys.*, 1999, **151**, 190–211.

27 H. Merlitz and W. Wenzel, *Chem. Phys. Lett.*, 2002, **362**, 271–277.

28 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.

29 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci.*, 2002, **99**, 12562–12566.

30 S. Piana and A. Laio, *J. Phys. Chem. B*, 2007, **111**, 4553–9.

31 C. Paissoni, M. Ghitti, L. Belvisi, A. Spitaleri and G. Musco, *Chem. - A Eur. J.*, 2015, **21**, 14165–14170.

32 K. A. Beauchamp, Y. S. Lin, R. Das and V. S. Pande, *J. Chem. Theory Comput.*, 2012, **8**, 1409–1414.

33 R. B. Best, N. V. Buchete and G. Hummer, *Biophys. J.*, , DOI:10.1529/biophysj.108.132696.

34 K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, *PLoS One*, 2012, **7**, e32131.

35 S. Li and A. H. Elcock, *J. Phys. Chem. Lett.*, 2015, **6**, 2127–2133.

36 F. Curnis, A. Sacchi, A. Gasparri, R. Longhi, A. Bachi, C. Doglioni, C. Bordignon, C. Traversari, G. Rizzardi and A. Corti, 2008, 7073–7082.

37 W. F. Van Gunsteren, J. R. Allison, X. Daura, N. Hansen, A. E. Mark, C. Oostenbrink, V. H. Rusu and L. J. Smith, 2–23.

38 F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7821.

39 G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J. Comput. Aided. Mol. Des.*, 2013, **27**, 221–234.

40 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347–1363.

41 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

42 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins Struct. Funct. Genet.*, 2006, **65**, 712–725.

43 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins Struct. Funct. Bioinforma.*, 2010, **78**, 1950–1958.

44 R. B. Best and G. Hummer, *J. Phys. Chem. B*, 2009, **113**, 9004–9015.

45 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

46 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.

47 A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

48 N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. Van Gunsteren, *Eur. Biophys. J.*, 2011, **40**, 843–856.

49 W. L. Jorgensen, D. S. Maxwell and J. Tirado-rives, 1996, **7863**, 11225–11236.

50 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

51 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren

and J. Hermans, in *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13--16, 1981*, 1981, pp. 331–342.
52 A. T. Brunger, *Nat. Protoc.*, 2007, **2**, 2728–2733.
53 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 14101.
54 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
55 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
56 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
57 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
58 X. Biarnés, F. Pietrucci, F. Marinelli and A. Laio, *Comput. Phys. Commun.*, 2012, **183**, 203–211.
59 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
60 G. a. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
61 F. Marinelli, F. Pietrucci, A. Laio and S. Piana, *PLoS Comput. Biol.*, 2009, **5**, e1000452.
62 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
63 M. Karplus, *J. Am. Chem. Soc.*, 1963, 85, 2870–2871.
64 Y. Shen and A. Bax, *J. Biomol. NMR*, 2010, **48**, 13–22.
65 P. H. Nguyen, G. Stock, H. Schwalbe, J. Wolfgang, G. V Frankfurt, M.- V Strasse, J. Graf, P. H. Nguyen, G. Stock and H. Schwalbe, *J. Am. Chem. Soc.*, 2007, **129**, 1179–1189.
66 D. A. Case, C. Scheurer and R. Bruschweiler, *J. Am. Chem. Soc.*, 2000, **122**, 10390–10397.
67 J. R. Allison and W. F. Van Gunsteren, *ChemPhysChem*, 2009, **10**, 3213–3228.
68 J. M. Schmidt and J. M. Reson, 1997, **322**, 310–322.
69 A. Demarco, M. Llinás and K. Wüthrich, *Biopolymers*, 1978, **17**, 637–650.
70 A. Demarco, M. LlináS and K. Wuthrich, *Biopolymers*, 1978, **17**, 2727–2742.
71 C. Pérez, F. Löhr, H. Rüterjans and J. M. Schmidt, *J. Am. Chem. Soc.*, 2001, **123**, 7081–7093.
72 A. Barducci, M. Bonomi and M. Parrinello, *Biophys. J.*, , DOI:10.1016/j.bpj.2010.01.033.
73 M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, *J. Chem. Theory Comput.*, 2015, **11**, 3499–3509.
74 D. Steiner, J. R. Allison, A. P. Eichenberger and W. F. Van Gunsteren, *J. Biomol. NMR*, 2012, **53**, 223–246.