In silico engineering of tailor made peptides: pushing the accuracy of theoretical molecular modeling beyond off-the-shelf computational procedures

by

Vincenzo Barone[1,2*], Balasubramanian Chandramouli[1,3], Sara Del Galdo[1], Giordano Mancini[1,2], Nicola Tasinato[1]

[1] Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy

[2] Istituto Nazionale di Fisica Nucleare (INFN) sezione di Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy

[3] Istituto Italiano di Tecnologia, via Morego 30, Genova, Italy

Email: vincenzo.barone@sns.it
Phone: +39050509215

ABSTRACT

Designing peptides with targeted functionalities strategically benefits by introducing amino acid scaffolds with special spectroscopic or structural properties. This demands a careful examination of the inherent behavior of such amino acids alone and/or when incorporated into the peptides. In this context, computational modeling techniques are widely utilized for unveiling the atomic-molecular properties that underpin the characteristic features; be it intrinsic or responsive to environments. To this end, statistical ensembles generated using molecular dynamics or Monte Carlo methods are fruitful to delineate, at atomic level, the conformational landscape of molecules. The accuracy of the sampling can be improved by rigorous refinement of the computational model via deriving the force field parameters with high level *ab initio* calculations. In this review, we report several show case examples that demonstrate the utility and effectiveness of robust force field parameterization to study both coded and non-coded amino acids, especially of spectroscopic interest.

1. INTRODUCTION

The identification of the conformers of a molecule (and the evaluation of its properties) under a specific set of conditions is a topic of tantamount importance in computational chemistry and has been the subject of a high number of studies.[1–3] As a possible criterion for the classification of conformational search methods one may consider the adopted physical model used to evaluate the energy and its derivatives (quantum mechanical, semi-classical or classical) and on the algorithm that generates the conformations which may be deterministic (e. g. a systematic scan over a set of internal coordinates) or stochastic, such as a Monte Carlo (MC) simulation.[4] Systematic search approaches may be effective under specific conditions but they become overwhelmingly costly as the number of internal degrees of freedom grows.[5,6] Molecular size obviously also affects the type of physical model that can be applied with available computer power. Theoretical modeling based on quantum chemical or classical approaches is nowadays widely employed to support experimental investigations, thus allowing the interpretation (or forecast) of the observed macroscopic behaviour in terms of basic molecular properties.[7] Approaches relying on quantum mechanics (QM) can actually deliver very detailed information on on-going chemical reactions and spectroscopic response to various stimuli;[8] however, the size of the systems that can be investigated is limited due to the high computational cost of QM methods. Indicatively, molecules containing a tenth of heavy (i.e. non-hydrogen) atoms can treated by coupled cluster (CC) theory; in particular, CC with singles, doubles and augmented by a perturbative estimate of the connected triples[9,10], CCSD(T), coupled to medium- or large- basis sets has become the "gold standard" for obtaining highly accurate structures, thermochemical and spectroscopic data[11–20]. Less costly but still accurate, second-order Møller-Plesset perturbation theory[21] (MP2) can be applied to molecules with about a few hundreds atoms, while Density Functional Theory (DFT) can be pushed to treat thousands of atoms.[22–32] Beyond this limits, or if an extended sampling in phase space is needed starts the realm of semi-classical and classical approaches. In the present study the focus is not on benchmarking different approaches but rather to assess the degree of accuracy and robustness that can be achieved under a specific theoretical framework, given a level of physical accuracy with which the energy and other properties of conformers are evaluated.

In the last years, QM computations have imposed themselves as powerful and widespread tools for the assignment and prediction of the experimental spectra either in the gas or condensed phases, as well as to get deep insights into the different effects which determine the observed spectroscopic properties.[33–35] In this context, the impressive growth of computational spectroscopy has prompted the development of the Virtual Multifrequency Spectrometer (VMS) that gives access to the latest developments of the field also to the non-specialist user[8,36,37]. The computational module (VMS-Comp), developed for computing the relevant spectroscopic data, has been complemented by powerful graphical user interface (GUI), VMS-Draw, which supports users in preparing and pre-organizing input and in processing the outcomes of a QM calculation by visualizing all the relevant information in an intuitive way[37]. VMS has been designed for computing a (vibrationally-resolved) and (infrared, Raman,) vibrational spectra as well as their chiral counterparts, while the support for EPR spectroscopy has been provided through an interface to the E-SPIRES code[38]. Recently, the frequency coverage of the VMS has been extended to cover (1- and 2-dimensional) NMR spectroscopy[39] and rotational spectroscopy[40].

Among spectroscopic techniques, integrated theoretical-experimental approaches to vibrational and rotational spectroscopy, are widely employed to study the structure, energetic and dynamics of the biochemical building blocks and drugs in the gas-phase, thus avoiding the complications arising from the interaction with the surrounding environment[41–44]. In fact, gas-phase studies offer the possibility of avoiding the competition between intra- and intermolecular interactions and hence accessing the inherent conformational behavior of bare biomolecules.

Furthermore, these spectroscopies are fruitfully exploited in the study of molecular complexes showing the prototypical interactions of biochemical systems or microsolvation effects, with QM calculations proving unique insights into the strength and the role played by non-covalent interactions in holding together the molecular fragments[45,46].

Quantum chemical approaches to spectroscopy and reactivity have also found a natural application in astrochemistry[47–49]: in fact, it has been recently pointed out "The ability to analyze routinely and completely the structural, spectroscopic, and electronic properties of any given molecule, regardless of its laboratory stability, make this tool a necessary component for astrochemical analysis."[50]. In this field, the interplay between laboratory measurements and theoretical investigation has led to the astronomical detection of many molecules in the interstellar medium and circumstellar shells. Special attention has been devoted to the study of prebiotic molecules, such amides, imines and carbonylic compounds, as they are essential ingredients for the origin of life, with quantum chemistry providing not only the relevant spectroscopic data for molecular detection[48–52], but also a tool to explore their possible reaction pathways, and the corresponding kinetic rates,[53–56] in the hostile conditions characterizing extraterrestrial objects.

In addition of being instrumental in guiding and supporting experimental investigations, QM calculations provide high-quality reference data for developing and testing new theoretical methods and computational strategies based on either classical or quantum mechanics. While database for thermochemical applications usually rely on correlated wavefunction methods coupled to extrapolation schemes to produce reference values[57–59], in the case of molecular structure, the semi-experimental (SE) approach can be fruitfully exploited to obtain accurate gas-phase equilibrium geometries[60,61]. According to this method, semi-experimental equilibrium structures are derived by a non-linear least-squares fit of the rotational constants measured experimentally for different isotopologues and corrected by vibrational contributions computed theoretically[62]. The interplay between experimental measurements and theoretical calculations has paved the route toward the extension of accurate structural studies to larger systems. In this respect, a structural database has been compiled in which the SE equilibrium geometries have been derived by using vibrational corrections to rotational constants evaluated at DFT level, thus allowing the treatment of biochemical building blocks such as glycine, uracil and thiouracil[63,64] Recently, this database has been extended by including sulfur-bearing molecules[65] and a new software for applying the SE approach has been implemented with a number of advanced features on error analysis, selection coordinate system and application of predicate observations[66]. Furthermore, in an ongoing effort to extend the size of the molecular system, two methods, namely the Linear Regression Approach (LRA) and the Template Molecule Approach (TMA) have been proposed[63,64] in order to improve the intrinsic accuracy of structural parameters obtained from B3LYP[67,68] and B2PLYP[69] calculations. Even though QM calculations can provide an accurate characterization structural, thermodynamic and spectroscopic properties of molecular system, also for flexible molecules[70] it should be recalled that the adopted model has to reflect the most relevant physical properties of the target molecular system. In this respect, paradigmatic is the case of the chemotherapeutic agent cisplatin, for which a model capable of reproducing its structural and vibrational features has been achieved only recently.[71]

Classical Molecular Mechanics (MM) simulations such as Molecular Dynamics (MD) and MC, have been successfully employed for studying complex systems (e.g. condensed phase systems) at an atomistic level for many years now[72–75] Among the reasons for this success their reduced computational cost and reasonable accuracy. In principle, classical methods should scale as $O(N^2)$ in a worst case scenario which must be compared to the $O(N^4)$ *with respect to the number of basis functions* scaling of simple Hartree Fock. This has to be balanced with the limited applicability of classical methods: chemical reactivity simulation is limited to a restricted range of substance and its application presents many difficulties.[76]Apparently, if the task at hand is the sampling of conformational space and no specific features of MD are of interest (e. g. the calculation of transport properties or of time correlation functions) MC should be the

approach of choice as it involves only the calculation of the energy, saving computation of all interatomic forces and the resolution of the equation of motion. However, devising an efficient scheme for generating MC moves is a daunting theoretical effort and without such an efficient scheme the ability of a MC scheme to achieve a sufficient sampling may be severely hampered;[74] move selection schemes optimized for biomolecular systems that tried to generate moves based on their specific topology (e.g. backbone moves or Cα-Cβ rotations) to increase sampling have been proposed[77]. Mezei has discussed the difficulty of generating such schemes;[78] MC has indeed been applied to the problem of conformational search of peptides with some success.[79] However, MD is currently the method of choice in biomolecular simulations. Independently of the sampling algorithm is selected, application of a classical method requires the definition of a force-field (FF) to calculate the potential energy of a simulated chemical system and the forces acting between atom pairs. A FF is a set of parameters and functional forms trained in order to reproduce experimental and/or quantum mechanical (QM) data. FFs allow to approximate the forces between interacting atoms and molecules with relatively simple analytical expressions, thus allowing the computation of chemical properties at a reasonable cost. The proper definition of the analytical functional form of the FF as well as its parameterization is crucial to reproduce properly the properties of the simulated system. The higher the accuracy of the underlying microscopic model, the higher the reliability of the estimated macroscopic properties will be; however, this accuracy may come at high cost because a very fine fitting of the properties of a specific molecular system comes at the expense of FF transferability to related molecular species. Hence, FF parameterization must always be carried out taking into proper account the phenomena that the classical approximation is meant to reproduce. In designing a FF, each atom is assigned a specific atom type depending on its chemical environment (for instance, an aromatic carbon belongs to a different atom type with respect to an aliphatic one). The FFs are usually constituted by a relative amount of simple interactions between the atom types. The set of interactions and the corresponding parameters can be derived according to different degrees of accuracy or transferability. The balance between the accuracy and transferability is fine tuned to treat a targeted class of molecules. For instance, popular FFs such as AMBER,[80] CHARMM,[81] GROMOS[82] or OPLS[83] were developed to be used for biomolecular simulations. Universal FFs (e.g. UFF[84] and Dreiding[85]), were designed to be highly transferable and include atoms over the whole periodic table. To obtain single molecule accuracy, the FF parameters should be refined to fit quantum data obtained from high level *ab initio* calculations. Some popular software tools devised for such purpose include ffTK,[86] Paramfit,[87] Joyce[88,89] and Picky[90]). Such an *ad hoc* parameterization is highly required especially when the intramolecular description achieved by standard FFs is not adequate. This is often necessary when dealing with the reproduction of spectroscopic properties or extending FFs to excited electronic states. One simple alternative route, to fine-tune accuracy versus transferability, could be to re-parameterize only few terms of an existing literature FF to fit data obtained from high level of QM theory. This can be useful when handling large molecules (for instance, biopolymers) where most of the existing parameters are fully validated, but some novel functional groups are introduced, e.g. to model the effect of an artificial modification.[91–93] In this context, often the focus is on the energetic contributions from non-bonded interactions and flexible structural degrees of freedom, as these are highly sensitive to the molecular specificity and they strongly influence the conformational equilibria of the phase space sampled via computer simulations.[94–97] In fact, they are also the most typically revised and improved terms between successive releases of generalized FFs, together with electrostatic terms. It should be also noted that often simulations of organic and biological molecules are performed by constraining the stiff internal coordinates (e.g. by means of constraining algorithms such as SHAKE,[98] LINCS[99] or using a quaternion based approach[100]), hence refinement of stretching and bending parameters is done less frequently. Given the delicate balance between transferability and accuracy, there is no one-size-fits-all solution to adapt conveniently the existing FF parameters to all occurrences. Hence, tailor made parameterization and/or refinement with high level *ab initio* calculations are highly important to obtain reliable

computational models.

In conclusion, accurate interpretation or forecast of the chemical properties of even small peptides, particularly if they include a nonstandard functionalization or are studied in specific environments, still constitutes a significant challenge. To be addressed properly a composite computational protocol, using the state-of-the-art of different physical models from high accurate QM methods to classical ones is often needed. For this reason, we tried to include here a perspective of the different components of such a composite strategy, focusing on specific case studies to achieve the necessary in-depth presentation of the methods. The paper is organized as follows: having listed the computational methods used in case studies (section 2), we first show (section 3) the level of accuracy achievable by state-of-the-art QM approaches in modeling the structural, energetic and spectroscopic properties of small biomolecules, by focusing our attentions on the results recently obtained for glycine and its dipeptide; as mentioned, we present a work based on a systematic search approach for a relatively small molecule (glycine) as the efficiency conformational search algorithms is beyond the scope of the present study. Then, (section 4) we present one of the possible approaches with which (starting from scratch or from one the widely used FF in the biomolecular simulation community) it is possible to derive tailor made FFs. The purpose of deriving a new FF may be for the sake of increasing the accuracy or because a new molecular species is investigated (it is worth to observe that this includes not only "proper new species" but also, as an example, FFs for excited states); we discuss this topic presenting a case study specifically carried out for this study, i.e. the zwitterionic form of Tyrosine which was selected to show how the parameterization of even this relatively simple molecule can still present challenges if a high accuracy is sought after. Finally, (section 5) we discuss how the presence of tetra substituted $C^{\alpha.\alpha}$ residues and nitroxide groups affect the relative stability of secondary structures in short peptides, using this systems to demonstrate the necessity for building FFs of ever increased accuracy and the performance of the presented approaches. Conclusions and future perspective are illustrated in the last section.

2. COMPUTATIONAL DETAILS

In this study we discuss a set of well-tested and yet recent computational protocols for the computational modeling of tailor made peptides and polypeptides using a set of case studies. Some of these case studies originated from recent papers while other, namely the semiclassical conformational search for glycine and the generation of a FF for the Tyrosine zwitterion, were purposely prepared for the present study. In the present section we provide computational and software details for the latter. To model the non-bonded part of the FF, atomic charges were estimated on the global minimum using population analysis based on Charge Model 5 (CM5)[101] and the LJ parameters were taken from OPLS. Tyrosine was immersed in a cubic box containing 3913 H_2O molecules (the SPC[102] model was used to mimic water/water and Tyrosine/water interactions) under periodic boundary conditions. The systems were initially brought to 0 K with a steepest descent minimization and then heated 300 K in a NVT ensemble using the velocity rescaling method [103] with an integration time step of 0.2 fs and a coupling constant of 0.1 ps for 10^6 simulation steps (200 ps). The time step and temperature coupling constant were then increased to 2.0 fs and 0.2 ps, respectively, and systems were let to converge to uniform density in a NPT ensemble (using the Parrinello–Rahman barostat and a coupling constant of 1.0 ps) for 10 ns. A final production run of 100 ns in a NVT ensemble was the carried out. All bonds were constrained using the LINCS[99] algorithm. The particle mesh Ewald[104] method was used to compute long range interactions with grid search and cut-off radii of 1.1 nm, for short range interactions a cut-off radius of 1.1 nm was employed. For each of the Tyrosine configurations sampled, the transition energies from the ground up to the fifth excited states were computed with Gaussian 09 package, by means of the TD-DFT method, using the B3LYP functional coupled with the 6-311+G(*d*) basis set, treating the water effect through C-PCM.

3. INTEGRATED STATISTICAL–QUANTUM CHEMICAL APPROACHES TO CONFORMATIONAL SEARCH

3.1 Methodological background

In this paragraph, we focus our attention to the sampling of the conformational landscape of flexible molecules, with particular attention to aminoacids, by means of statistical approaches. Such a preliminary screening is of strategic relevance for subsequent experimental or (high-level) theoretical studies. In fact, recent investigations have shown that accurate structures and spectroscopic properties of biochemical building blocks in the gas phase can be accurately obtained by joint spectroscopic, in particular rotational spectroscopy, and quantum chemical investigations. From a theoretical point of view, a thorough characterization and fruitful interpretation of experimental information requires the determination of accurate molecular structures and of the corresponding energetics, which are a fundamental prerequisite for accurate prediction of the spectroscopic properties measured experimentally. Despite its apparent simplicity, glycine may not be overlooked, in fact a systematic modeling of its conformational potential energy surface and spectroscopic behavior has been achieved only recently.

Because of its small size, the geometries, energies and rotational-vibrational spectroscopic properties of glycine have been determined to a very high accuracy by quantum chemical calculations. In a series of works, Barone and co-workers characterized the structures, relative energies and infrared spectra of the six low-energy conformers of glycine, by state-of-the-art quantum chemical calculations, which delivers an accuracy rivaling that of the most refined experimental techniques.[21,22,105] The structures of the minima and of the corresponding connecting transition states were determined from a preliminary search of the relevant stationary points on PES performed by using the B3LYP functional in conjunction with the polarized double-ζ SNSD[106] basis set.[43] For the minima, the calculations were then refined by means of composite schemes that rely on extrapolation techniques and hybrid coupled cluster/DFT computational protocols. Equilibrium structures were calculated according to two different methods accounting simultaneously for basis set and electronic-correlation effects. The first approach considers the Hartree-Fock self-consistent-field (HF-SCF) extrapolated to the complete basis set (CBS) limit ($\frac{dE^{\infty}_{HF-SCF}}{dx}$), the CCSD(T) valence correlation energy extrapolated to the CBS ($\frac{d\Delta E^{\infty}_{CCSD(T)}}{dx}$) and the core-correlation energy ($\frac{dE_{CV}}{dx}$) , and it is based on the additivity of the various contributions at the energy-gradient level (see Equation 1):[107]

$$\frac{dE_{CBS+CV}}{dx} = \frac{dE^{inf}_{HF-SCF}}{dx} + \frac{dE^{inf}_{CCSD(T)}}{dx} + \frac{dE_{CV}}{dx} \quad Eq.\ 1$$

The second computational protocol employed was the so-called “cheap” scheme, which exploits the additivity relationship directly at the level of geometrical parameter, r [46,48]:

$$r(best) = r_{CCSD(T)/VTZ} + \Delta r^{\infty} + \Delta r_{CV} + \Delta r_{aug} \quad Eq.\ 2$$

This protocol starts from the CCSD(T)/cc-pVTZ result ($r_{CCSD(T)/VTZ}$) and the different missing contributions are evaluated by using the second-order Møller-Plesset perturbation theory in conjunction with the cc-pV*n*Z (*n* = T, Q), cc-p[108–110] as detailed in Table 1. Involving geometry optimizations mainly at the MP2 level, the second scheme is characterized by a lower computational burden then the first, more rigorous, one as highlighted by the “cheap” label.

Table 1 Contributions to the CCSD(T) ansatz within "cheap" composite scheme.

Contribution	Expression	Level of theory
Complete basis set	$\Delta r^{\infty} = \frac{Y^3 r(Y) - X^3 r(X)}{Y^3 - X^3} - r(X)$	Geometry optimization at the MP2 level coupled to cc-pVQZ (Y = 4) and CC-pVTZ (X = 3) basis sets
Core-correlation	$\Delta r_{\mathrm{CV}} = r(\mathrm{CVTZ, all}) - r(\mathrm{CVTZ, fc})$	Geometry optimization at the MP2 level using the cc-pCVTZ basis set by correlating all electrons and in the frozen-core (fc), approximation.
Diffuse functions	$\Delta r_{\mathrm{aug}} = r(\mathrm{aVTZ}) - r(\mathrm{VTZ})$	Geometry optimization at MP2 level using aug-cc-pVTZ and cc-pVTZ basis sets.

The same scheme of Eq. 2 was also applied to obtain best estimates of harmonic vibrational frequencies, (ω), while a slight modification was introduced for infrared harmonic intensities (I):

$$\omega(best) = \omega_{CCSD(T)/VTZ} + \Delta\omega^{\infty} + \Delta\omega_{CV} + \Delta\omega_{aug} \quad Eq.\ 3$$

$$I(best) = I_{CCSD(T)/VTZ} + \Delta I_{CV} + \Delta I_{aug} + \Delta I_{QZ-TZ} \quad Eq.\ 4$$

where the meaning of the various terms is analogous to that of Table 1, and

$$\Delta I_{QZ/TZ} = I(MP\,2/VQZ) - I(MP\,2/VTZ) \quad Eq.\ 5$$

Electronic energies were computed at the best estimated equilibrium structures according at the CCSD(T)/CBS + CV level:

$$E^{elec}_{CBS+CV} = E^{\infty}_{HF-SCF} + \Delta E^{\infty}_{cor} + \Delta E_{cv} \quad Eq.\ 6$$

where E^{∞}_{HF-SCF} is the HF-SCF CBS limit obtained by using cc-pVnZ basis sets (n = T, Q, 5), ΔE^{∞}_{cor} represents the CCSD(T) correlation contribution extrapolated to the CBS by using cc-pVnZ (n = T, Q), and ΔE_{CV} is the correction for core correlation. In order to account for anharmonic effects in the infrared spectra and vibrational contributions to thermodynamic properties, an hybrid CCSD(T)/DFT anharmonic force field was formulated by correcting best-estimated harmonic frequencies through cubic- and quartic semi-diagonal force constants evaluated at the B3LYP/SNSD level of theory. Anharmonic vibrational frequencies and infrared intensities were obtained within the framework of generalized vibrational perturbation theory to second order (GVPT2).[111,112] For thermodynamic properties, simple perturbation theory (SPT) was adopted for calculating the partition function at the anharmonic level, with vibrational energies obtained using the hybrid degeneracy corrected second-order perturbation theory (HDCPT2).[113]

3.2 Results on glycine and glycine dipeptide analog

The adopted approach provided bond distances, conformational enthalpies and vibrational frequencies with an accuracy of (at least) 0.001 Å, 1 kJ mol^{-1} and 10 cm^{-1}, respectively, and it allowed a direct comparison between simulated and experimental IR spectra, including also weak features arising from overtone and combination bands. The experimental vibrational frequencies of the most stable conformers (see Figure 1), namely glycine Ip and IIn, were reproduced with a mean absolute error of about 7 cm^{-1}, while the accuracy reached for structural parameters turned into an agreement within 0.2% for the ground state rotational constants of glycine Ip. In a similar way, the computed vibrational

frequencies of the two elusive conformers IIIp and IVn resulted in close agreement around (mean absolute deviation of 4 cm^{-1}) with the available experimental data obtained in noble gas matrix. Furthermore, it was shown that the low temperature IR spectra of the d0 and d3-glycine isotopologues are dominated by the absorption due to the Ip, IIn and IIIp conformers, with relative populations of 0.77, 0.18 and 0.06, respectively. Yet, it was also pointed out that minor contributions, around 1%, may arise from the IVn, Vn and VIp conformers.

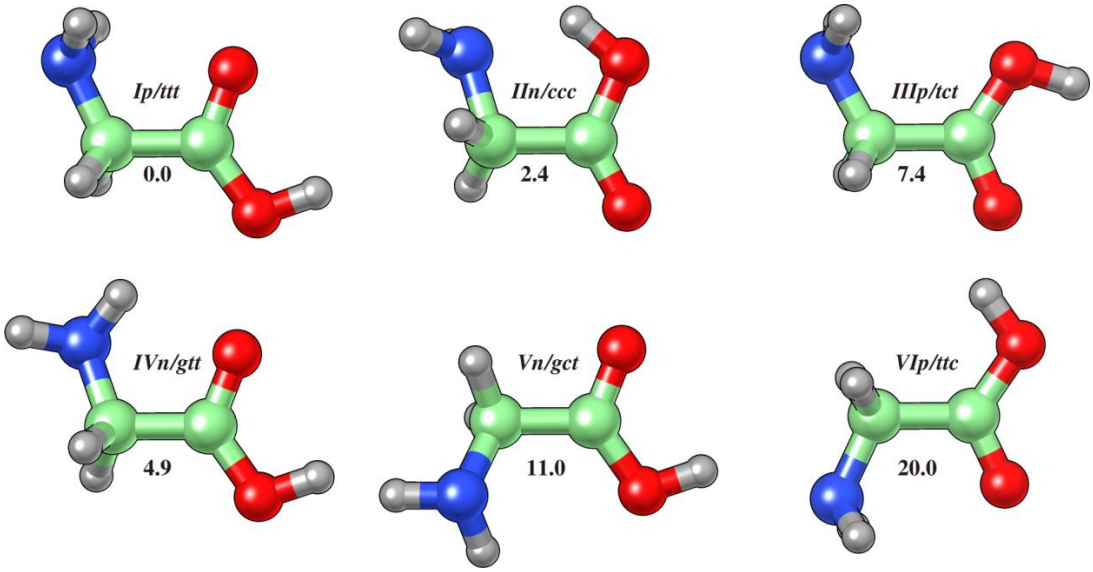

Figure 1 Glycine stable conformers

The "cheap" geometry scheme and the hybrid CCSD(T)/DFT approach were also applied to study the glycine dipeptide analogue (see), N-acetyl-glycinamide (Ac-Gly-NH2).[114] Theoretical results were completed by the experimental measurements of the rotational spectrum by chirped-pulse Fourier-transform microwave (CP-FTMW) spectroscopy with laser ablation, with the aim of investigating the structure and conformational behavior of an α-aminoacid incorporated into a peptide chain. As in the case of the glycine monomer, a conformational search was carried out at B3LYP/SNSD level, on the basis of previous studies, resulting in the C7 and C5 rotamers being the most stable ones (see Figure 2). The molecular spectroscopic parameters (rotational constants, nitrogen quadrupolar coupling constants and dipole moment components), obtained by correcting equilibrium values as obtained from the "cheap" composite scheme through B3LYP/SNSD vibrational contributions, were then employed to guide the experimental investigation of this dipeptide. In particular, the predicted ground state rotational constants resulted to agree within 0.3% and 0.6% for the C7 and C5 rotamers, respectively, while the maximum difference between theory and experiment in the nitrogen quadrupolar coupling parameters was 5%. By referring to pure electronic energy, the C7 conformer, which involves the formational of a seven-membered ring, was estimated to be more stable than the C5 one (featuring a 5 membered ring) by 4.7 kJ mol^{-1}, which reduces to 1.9 kJ mol^{-1} when considering the Gibbs free energy. The latter was in turn employed to computed the C5/C7 population ratio which resulted 0.43, in good agreement with the value of 0.32(10) estimated from the relative intensities of the rotational transitions measured in the CP-FTMW spectrum. It is interesting to note that, the good agreement reached for the population ratio appears a natural consequence of the inclusion of anharmonicity into thermodynamic properties and the proper treatment of low-frequency torsional modes, with vibrational correction computed by adopting a hindered-rotor anharmonic oscillator model within VPT2 coupled to SPT.

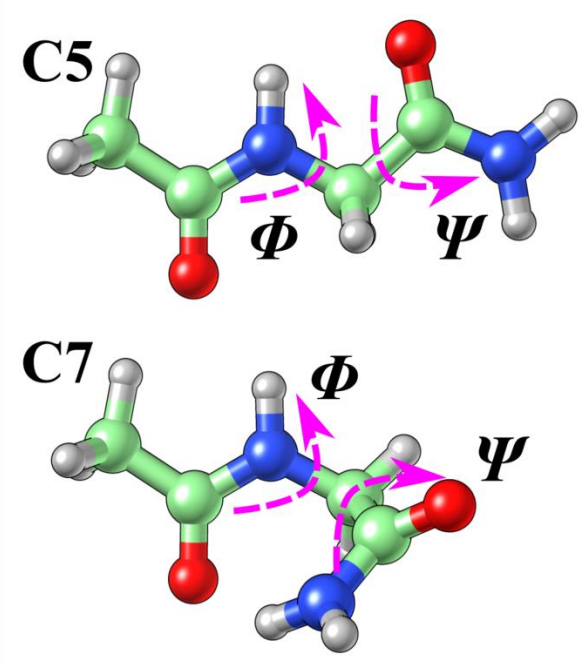

Figure 2 Glycine dipeptide

The results obtained for the glycine dipeptide analogue suggest that quantum chemical approaches are able to provide a reliable description of medium-sized molecules of biochemical interest even when flexible internal degrees of freedom are present. Yet, the majority of biochemical molecules are characterized by a variety of conformers, closely spaced in energy, due to the presence of many floppy torsional angles and interaction sites. In those cases where a rich conformational landscape is sought, a systematic scan of dihedral angles for searching the minima structures, may be inefficient from the computational point of view, even resorting to DFT based methods, or may lead to an incomplete sampling of the PES. In such instances, an integrated statistical – quantum mechanical approach can be adopted.

4. ROBUST TAILOR MADE FORCEFIELDS

4.1. Parameterization of intramolecular force fields from quantum mechanical calculations

The most common FF mathematical expression is a linear combination of functions of a set of internal coordinates (namely, the bond stretches, angle bendings, and dihedral torsions), completed by the insertion of the non-bonded interactions in the form of Coulomb and Lennard Jones (LJ) terms. Hard (or ”stiff”) internal coordinates, that is bond stretchings, angle bendings, and stiff dihedral angles (like those that drive the planarity of aromatic rings) are expressed with harmonic potentials, while sums of cosine functions are employed to model the more flexible dihedrals (or ”soft” internal coordinates). FF can be discriminated according to the degrees of complexity adopted for treating such intramolecular part. There are FFs diagonal in a given set of internal coordinates, containing only harmonic terms for bond stretching and angle bending and there are FFs including anharmonic terms and off diagonal couplings. Typically, the former are designed for simulations of large systems in condensed phase, while the latter are mostly used for studying gas phase systems. Several coupling terms (namely, stretching-stretching, stretching-bending, stretching-torsion, bending-torsion and torsion-torsion terms) are implemented in the Joyce software, which allows deriving the equilibrium internal coordinates and force constants for the target molecule by fitting QM optimized energies, gradients and Hessian matrix. Further, it supports flexible handling of non-bonded intramolecular interactions, thus tuning the FF specificity according to the user needs. The FF parameters are obtained by minimizing the following cost function, sampling N_g conformations of the molecule to be studied:

$$I = \sum I_g \quad Eq.\ 7$$

where for each g conformation

$$I_g = W_g\left[\left(E_g - E_0\right) - V_g\right]^2 + \sum_{K=1}^{3N-6} \frac{W'_{Kg}}{3N-6}\left[E'_{Kg} - V'_{Kg}\right]^2 + \sum_{K\leqslant L}^{3N-6} \frac{2W''_{KLg}}{(3N-6)(3N-5)}\left[E''_{KLg} - V''_{KLg}\right]^2 \quad Eq.\ 8$$

where the indices *K* and *L* run over the internal coordinates. E_g is the total energy obtained by a QM calculation, and E_0 is the energy at the reference geometry (g = 0). E'_{Kg}, E''_{Kg} are the energy gradient and Hessian at a given geometry with respect to the normal coordinates evaluated at the same geometry. *V*, *V*', and *V*'' are the corresponding quantities calculated by the FF. The constants *W*, *W*', and *W*'' weight the different terms at each geometry. The Hessian matrix and harmonic vibrational frequencies to be inserted in the FF equation are evaluated on the target molecule global minimum. Conversely, the "soft" dihedral parameters are derived by fitting each torsional energy profile obtained scanning QM energies along the corresponding (flexible) internal coordinate of interest, constraining all the remaining geometrical parameters. All torsional energy fittings are performed within the Frozen Internal Rotation Approximation. It consists of ignoring the changes of most of the internal coordinate not directly involved in the internal rotation, eventually retaining the changes of few pertinent degrees of freedom whose coupling term with the dihedral has to be included in the FF. In such a way, the torsional energy is mainly attributed to the torsional term, whereas in the QM calculation it is distributed over several internal coordinates, since, in principle, these are all coupled to each other. Joyce is meant for refinenment of internal degrees of freedom; non-bonded intramolecular interactions are treated as the sum of the Coulomb interaction and the LJ term. Point charges can be either assigned by the user (e.g. from the starting FF) or taken from the QM computed data (e. g. using the RESP[115] or CM5[101] methods). As far as the LJ intramolecular interactions (for pairs located three or more bonds apart) are concerned, the parameters can be assigned to literature values or separately fitted to improve the parameterization (including the possibility of switching off selected pairs of sites). Actually, parametrization of non bonded (both electrostatic and Lennard Jones in this context) terms is harder optimization problem as compared to internal degrees of freedom, at least if assuming a diagonal FF, because of the inherent higher dimensionality of the problem and the intrinsic need for a certain degree of transferability. As pointed out, atomic point charges are usually derived from QM calculations using a fit of the electrostatic potential (as in the RESP[115] and related methods) or using population analysis methods (as in the CM5 and related methods); given the relative cheap costs of such calculations, point charges are often re-assigned when dealing with new molecules. Lennard Jones parameters undergo a different treatment both because of differences in the physical model and because their adjustment is used to reproduce bulk observables. Standard force fields such as OPLS[83] often apply a cycle of parameter optimization with respect to a given target (e. g. hydration free energies) for small fragment followed by validation using extensive simulations. Recently, new approaches that try to automatically avoid overfitting while yielding a more effective sampling of the parameter space, such as Force Balance[116] (actually, one such method, called LRR-DE is actually under development in our laboratory), have been proposed but still need to reach widespread application. For this reason LJ set of parameters can be considered the hardest and hence, less often updated part of a FF, at least in the domain of biomolecular applications. Hence, in this study we will adopt the common approach that starts the refinement of FF or its extension to a new molecular species with the improvement of soft dihedrals

FFs used in biomolecular simulations have most often a diagonal form in which the total energy is partitioned into bonded (for atom pairs up to three covalent bonds apart) and non bonded interactions (for atom pairs separated by more than three covalent bonds); atoms separated by exactly three bonds (1-4 interactions) are often treated using a scaled down version of the non bonded term (with the same scaling factors of e. g. 0.5 as in OPLS or different scaling factors as in AMBER). According to the aforementioned internal coordinate contributions the total energy is portioned as in Equation 9:

$$E_{intra} = E_{\substack{bond \\ stretchings}} + E_{\substack{angle \\ bendings}} + E_{\substack{stiff \\ torsions}} + E_{\substack{soft \\ torsions}} + E_{\substack{non-bonded \\ interaction}} \quad Eq.\ 9$$

$$E_{\substack{bond \\ stretchings}} = \frac{1}{2}\sum_{\mu}^{N} k_{\mu}^{s}\left(b_{\mu} - b_{\mu}^{0}\right) \quad Eq.\ 10$$

$$E_{\substack{angle \\ bendings}} = \frac{1}{2}\sum_{\mu}^{N} k_{\mu}^{b}\left(\theta_{\mu} - \theta_{\mu}^{0}\right) \quad Eq.\ 11$$

$$E_{\substack{stiff \\ torsions}} = \frac{1}{2}\sum_{\mu}^{N} k_{\mu}^{t}\left(\phi_{\mu} - \phi_{\mu}^{0}\right) \quad Eq.\ 12$$

$$E_{\substack{soft \\ torsions}} = \sum_{\mu}^{N}\sum_{j}^{N}\left[1 + cos\left(n_{j}^{\mu}\delta_{\mu} - \gamma_{j}^{\mu}\right)\right] \quad Eq.\ 13$$

$$E_{\substack{non \\ bonded}} = \sum_{i}^{N}\sum_{j>i}\frac{q_i q_j}{r_{ij}} + 4\sum_{i}^{N}\sum_{j>i}\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad Eq.\ 14$$

In these expressions, $k_{\mu}^{s}, k_{\mu}^{b}, k_{\mu}^{t}, b_{\mu}^{0}, \theta_{\mu}^{0}, \phi_{\mu}^{0}$ are, respectively, the force constants and the corresponding equilibrium values for stretching, bending and stiff torsions. N_{μ} is the number of cosine functions used to describe the dihedral angle μ, while $n_{j}^{\mu}, \gamma_{j}^{\mu}$ are the multiplicity and the phase factor of the *j*-th cosine term. Equation 14 describes the non-bonded interactions, accounting for both Coulomb and Lennard-Jones contributions. The summations run over all the non-bonded couples, q_i is the charge on atom *i* and r_{ij} is the distance between the atoms *i* and *j*. According to the typical Lennard Jones[117] potential functional form, ε_{ij} is the well depth and σ_{ij} is the sphere radius for each *ij* couple. The general flowchart of the parameterizing/ re-parameterizing using Joyce protocol is depicted in Figure 3.

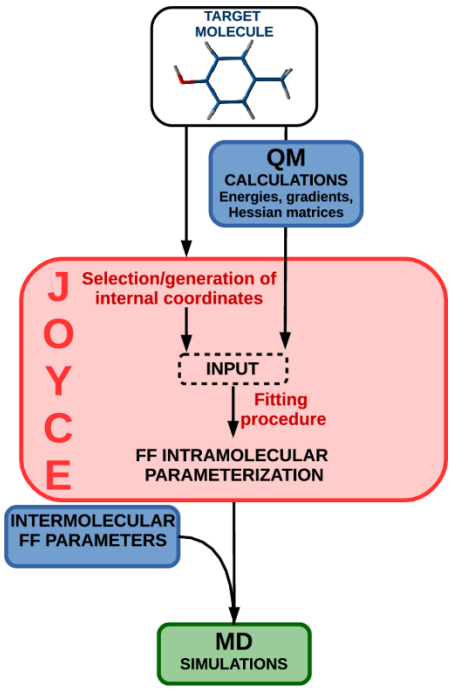

Figure 3 Flowchart of Joyce computational protocol

4.2. Generation of Tyrosine zwitterion FF

In the decade Joyce has been developed, it proved its accuracy and robustness with many different molecular systems. This section illustrates some noticeable examples, especially focusing on the studies carried out during the last five years, after the actual version of the code has been proposed. Remarkably, the Joyce protocol was used to parameterize the FF of electronic excited molecular states. The procedure was firstly applied to several coumarin dyes, obtaining accurate *ab initio* derived force fields for both ground and excited electronic states thus allowing a reliable computation of the fluorescence spectra[89,118]. Joyce was also applied for merging literature parameters and "ad hoc" derived parameters. In fact, a tailored FF was used to model the 4-naphthyloxy-1-

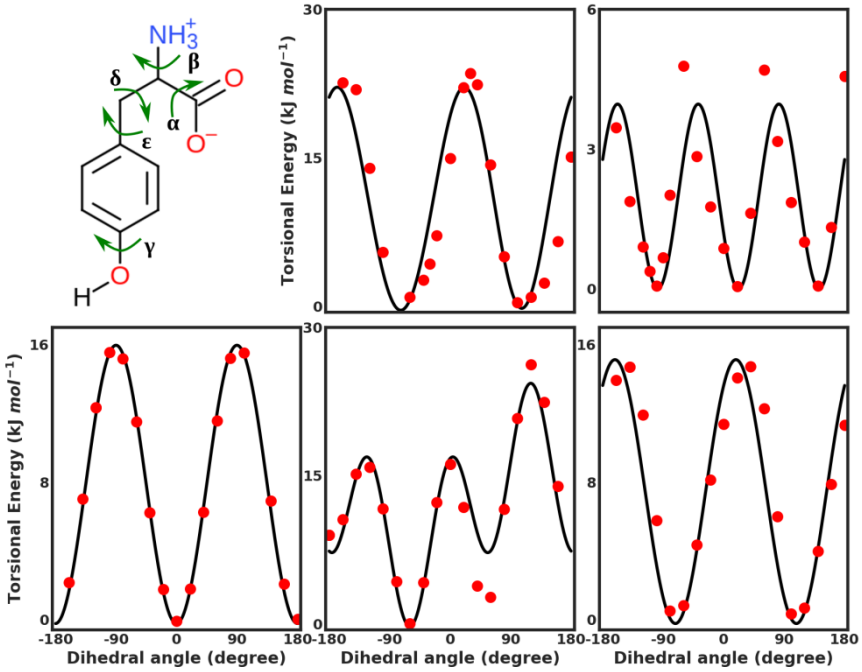

Figure 4 Structural formula of zwitterionic Tyrosine showing the definition of dihedral angles (first panel). Comparison of the torsional profiles obtained by ab initio calculations (red dots) and Joyce (lines). Panels refer to dihedrals α, β, γ, δ, ε.

methoxy-2,2,6,6-tetramethylpiperidine chromophore[119] in its ground and excited states, whereas a refined set of literature parameters was employed to mimic a surrounding polymeric environment[120–122]. Recent applications have been addressed to the FF parameterization of newly synthesized molecules, namely 5-((4-dicyanovinyl)ethynyl)-1-methyl-2-(4-nitrophenyl)-1H-imidazole, 5-((4-methoxy)ethynyl)-1-methyl-2-(4-nitrophenyl)-1H-imidazole.[123] In such cases, Joyce has proven to be very effective in reproducing reliable MD trajectories of the dyes both in solution with different polarity solvents, polymeric matrices and lipid membrane. Here, we report as an illustrative case study, the complete FF reparameterization of the Tyrosine amino acid in zwitterionic form. Tyrosine is very sensitive to the surrounding molecular environment and hence it is routinely employed as a molecular spectroscopic probe[124,125] for studying protein response to environmental changes; furthermore the zwitterionic form is (to the best of our knowledge) missing in standard biomolecular forcefields. In this regard, the correct sampling of its accessible conformational space is essential to reliably reproduce its spectroscopic properties. For each soft dihedral internal coordinate, the torsional profile to be fitted was obtained by scanning the optimized QM energy along the dihedral coordinate while constraining all the other conformational parameters to the corresponding global minimum structure values. Our first aim was to achieve the most accurate reproduction of the *ab initio* energy torsional profile of each of the five dihedral angles characterizing the amino acid. In Figure 4 we show the *ab initio* torsional profiles and the corresponding Joyce FF functions derived for each dihedral internal coordinate. As it can be seen, the FF functions derived accurately

reproduce the QM profiles. Regarding the torsions of the charged groups (i.e. dihedrals α and β) it can be observed that the agreement may still be considered satisfactory but the diagonal approximation of the FF begins to show its limits. To test the FF thus obtained we run a MD simulation in aqueous solution to (i) assess the accuracy of the FF in sampling the correct conformational basins of Tyrosine and (ii) show how sampling of configurations from a classical trajectory is an effective way of generating weighted sets of structures without a too expensive coverage of the PES at QM level. Figure 5 shows the profiles of the free energy variation as a function of each of the dihedral coordinates corresponding to the torsions of the two charged groups (i.e. α and β in Figure 1) and compares them with the corresponding QM energy.

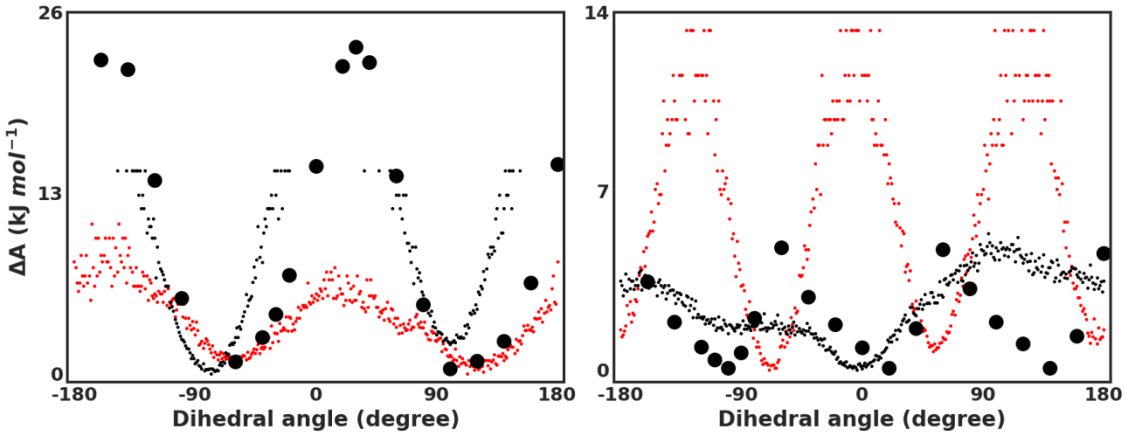

Figure 5 QM torsional energy profile (black circles) compared with the Helmholtz free energy variation as a function of the dihedral coordinate, as obtained from the MD simulations performed using Joyce (black dots) and OPLS (red dots) FFs. The two panels refer to α and β dihedrals (refer Fig1). Note that missing ΔA points imply that no sampling could be found for the corresponding geometry. The profile obtained using as statistical ensemble the MD simulation performed via the Joyce FF (black dots), is also compared with the one obtained with the use of OPLS FF (red dots).

The Helmholtz free energy variation (ΔA) was computed from the MD ensemble probability distribution of a generic dihedral coordinate, according to the following definition

$$\Delta A(\theta) = -kTln\left(\frac{\rho(\theta)}{\rho_0}\right) \quad Eq.\ 15$$

where k is the Boltzmann constant, T is the temperature, $\rho(q)$ is the probability of the q angle having ρ_0 as a maximum value. As far as the α dihedral angle is concerned, both Joyce and OPLS FFs correctly locate the energy minima (with OPLS accuracy being slightly better) but the transition barrier is clearly underestimated by the latter. However, a completely reversed trend of maxima and minima positions for the β dihedral angle is obtained from the OPLS simulation. Clearly, the most sampled conformations (along the two considered internal coordinates) yielded using the Joyce FF are closer to the QM minima with respect to the starting FF. Note that intrinsic discrepancies between QM energy and free energy profiles can be ascribed to the entropic contribution accounted for by the latter observable. The SPC-Tyrosine MD trajectory produced by using Joyce FF was used to compute the theoretical UV-VIS absorption spectrum of zwitterionic Tyrosine in solution. Two hundred statistically uncorrelated snapshots were extracted from the MD simulation. For each of the sampled conformations, the vertical transition energies were convoluted with Gaussian functions setting 0.1 eV as HWHM.[119,126] Finally, by averaging over the conformational ensemble sampled the spectrum reported in Figure 6 was obtained. As it can be seen, the spectrum is characterized by two bands in the spectral range of 180-300 nm, and experimental peaks are measured at 275 and 220 nm. The corresponding computed values resulted to be red- and blue- shifted by 17 and 10 nm, respectively.

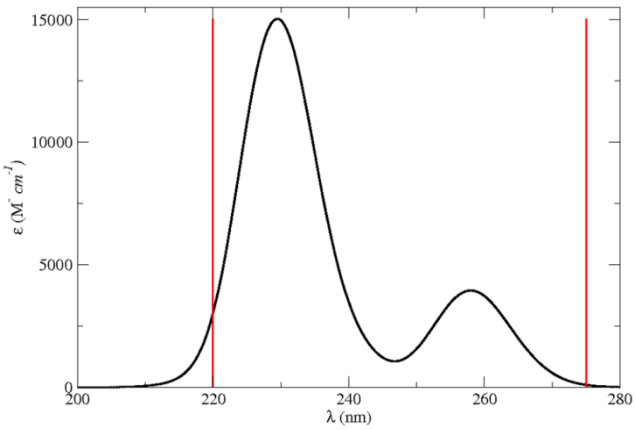

Figure 6 Theoretical-computational UV-Vis absorption spectra of zwitterionic Tyrosine in SPC. The solid red lines indicate the wavelengths of experimental maxima.

5 COMPUTATIONAL MODELING OF $C^{\alpha,\alpha}$ -TETRASUBSTITUTED AMINO ACIDS

5.1 $C^{\alpha,\alpha}$ -tetrasubstituted amino acids in peptide design

$C^{\alpha,\alpha}$-tetrasubstituted amino acids represent a versatile set of non-proteinogenic amino acids, widely used in peptide engineering. Notable and well studied members of this family include the cyclic 2,2,6,6-tetramethyl-N-oxyl-4-amino-4-carboxylic acid (TOAC), 1-aminocyclohexane-1-carboxylic acid ($A_{c6}C$), and the linear α-aminoisobutyric acid (Aib), whose structures are illustrate in Figure 4. These amino acids have an additional substitution at C^{α} position that render additional conformational rigidity to the backbone due to steric hindrance, thereby conferring specific secondary preferences to the hosting peptides (particularly 3_{10} or α-helix backbone conformations).

TOAC (Figure 7A) is an unnatural paramagnetic amino acid that has a nitroxide group rigidly attached to C^{α} carbon, and the first non-coded amino acid incorporated into model peptides to study the conformational properties via EPR spectroscopy[127]. TOAC has several advantages as a spin label for studying peptides. First, it has a similar backbone construct as of natural amino acids and hence site-specific introduction can be achieved, via peptide bond or by simple peptide synthesis. Second, rigidity of its cyclic structure eliminates ambiguities due to side chain conformations when studying peptide backbone dynamics. Use of double labeled TOAC to study the peptide secondary structure is widespread in the literature, largely through EPR spectroscopy[128]. The spin-spin interactions in the EPR spectra between the two TOAC residues permits direct estimation of the intermediate distance, thus enabling a precise understanding of the fold and secondary structure of the host peptide[129,130]. Further, TOAC is a fluorescence quencher, and has been utilized in photochemical studies on peptides containing both TOAC and additional fluorophore (be it incorporated or natural amino acid like tryptophan, tyrosine). Earlier studies mainly focused on TOAC incorporation on short peptides. Recently, Karim et al., devised a new method for the incorporation into proteins which has opened promising avenues for TOAC applications in protein research.

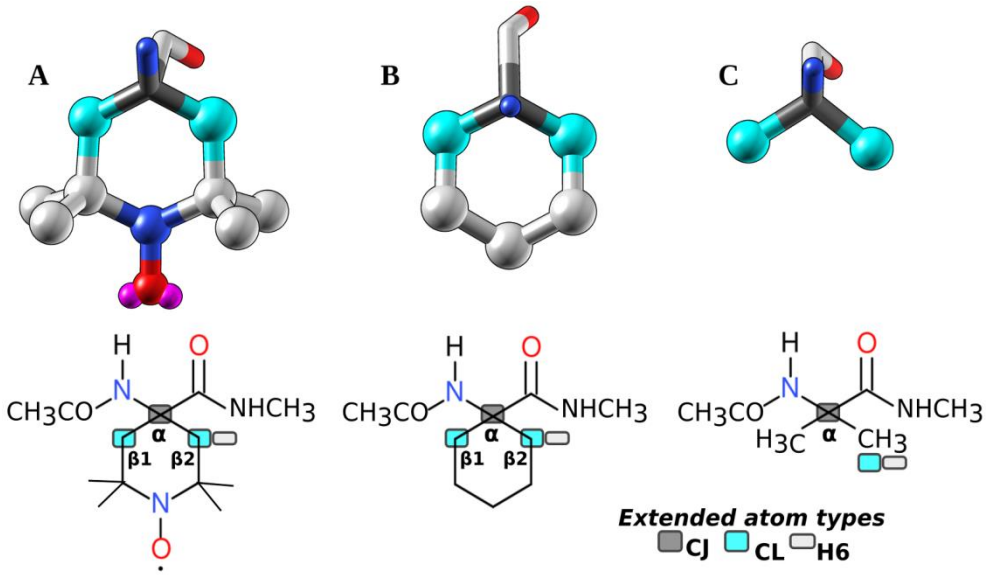

Figure 7 Structure of $C^{\alpha,\alpha}$-disubstituted amino acids; TOAC (A), $A_{c6}c$ (B) and Aib (C), where the peptide backbone is shown as sticks and side chain in ball-stick mode. Backbone C^{α} and side chain C^{β} that bear new atom types are shown in dark gray and cyan, respectively. For clarity, hydrogen were omitted. Schematic illustrations of corresponding dipeptide analogues adopted for parameterization are shown in the bottom panels. Additional atom types extended to AMBER ff99SB set are indicated.

$A_{c6}C$ (Figure 7B) is unnatural amino acid with a cyclic side chain similar to TOAC. Introduction of $A_{c6}c$ into peptides is known to increase the propensity of forming helical conformations, and has potential utilities in the design of bioactive

peptides. Aib (also known as α-methyl alanine, (Figure 7C) is simple and widely studied α-α dialkylated amino acid. Aib is present in peptides of microbial origins, and short Aib containing linear peptides such as alamethicin produce voltage gate membrane channels[131,132]. Aib constituent peptides have antibiotic effects, and hence well suited for developing novel peptide based antibiotics[133,134]

These $C^{\alpha,\alpha}$-disubstituted amino acids are used to study the conformational features of synthetic and stereo-chemically restricted analogues of bioactive peptides, and hence they have implications in peptide design. Despite their importance, computational studies detailing the effects of these amino acids incorporation into model peptides, their conformational preferences and influence of the solvent conditions are very limited. In the following text, we highlight the recent in-house developed computational protocols, that dovetails QM and MM approaches, for modeling and investigating $C^{\alpha,\alpha}$-tetrasubstituted amino acid substituted polypeptides.

5.2 Improved parameterization and assessment of AMBER force field for Aib amino acid

A plethora of studies have shown that Aib containing peptides adopt helical conformations (3_{10}/α-helix)[131,135–139], and specific factors driving the conformational propensity of poly-Aib towards 3_{10} or α-helix have been addressed[140–142] While one can transfer the parameters of natural amino acids (eg. alanine) existing in biomolecular FF like AMBER, often such FF have limitations due to their intrinsic preference towards specific secondary structures like α-helix. Further, for Aib a careful treatment of the non-bonding interactions between the methyl groups is needed to study the conformational landscape of Aib polypeptides, which clearly demonstrates the necessity for the derivation of an accurate Aib FF. To this end, we employed, on various Ac-Aib-NMe dipeptide conformers (Figure 8A), high level QM calculations with SNST basis set and B3LYP functional augmented for dispersion energy through the Grimme's DFT-D3 scheme to obtain good estimates of the conformational energies[143] Bond and angle parameters for Aib were transferred from AMBER ff99SB and those of cyclic α-α dialkylated residues (see section 5.3), whilst van der Waals and torsional parameters for the methyl groups were rederived. The atom centered RESP charges were determined at the PBE0/N07D level. The improved FF was tested by comparison to the QM data and a series of MD simulations on Aib polypeptides (Ac-$(Aib)_n$-Nme, n=1-33) in *vacuo* and in aqueous solutions. Energies of different Ac-Aib-NMe conformers obtained at MM level with the newly derived force field well agreed with QM data with an overall RMSD of 0.5 kcal mol^{-1} compared to 1.6 kcal mol^{-1} as yielded with standard AMBER parameters (Table 2). Notably, the energy difference of 4.5 kcal mol^{-1} between γ and δ was also close to the QM level estimation (4.8 kcal mol^{-1}). The standard AMBER parameters produced a difference of 0.2 kcal mol^{-1}. Similar fine agreement was observed between the MM and QM calculations performed on Aib_n polypeptides of varying length (n = 1-33) that supported the transferability of the new FF to study longer peptides.

Table 2: Energies (in kcal mol^{-1}) of Ac-Aib-NMe from QM (B3LYP/N07T) and MM calculations with newly derived Aib FF and [b]ff99SB force field. Energy values obtained from Grubisic et al.[144]

Structures [a]	QM	Derived-FF	Std-FF[b]
I γ	0.00	0.00	0.00
II β	0.89	2.50	-0.03
III 3_{10}	2.53	2.29	-2.24
IV α	2.86	2.81	-0.15
V ε	3.84	4.66	4.24
VI δ	4.81	4.54	0.16
RMSD		0.48	1.63

The parameters were tested to study solvent effects on homopolypeptides (Ac-$(Aib)_4$-Nme and Ac-$(Aib)_7$-Nme) through MD simulations in vacuo and aqueous solutions. The simulations showed the predominance of 3_{10} helix (φ = ±60°, ψ = ±30°) by non-terminal Aib residues in *vacuo* and α-helix (φ = ±60°, ψ = ±40°) in aqueous solutions[138,145,146] Further, the solvation free energy for the conversion from 3_{10} to α-helix was obtained using MM/PCM calculations by rotating the ψ angle (Figure 8A), keeping the fixed φ value, as the latter has similar values for both helices (i.e., ±60). The minima in the profile was constituted by both 3_{10} (-51°,-20°) and α-helix (-54°,-47°). The energy barrier computed for the interconversion was ~1.6 kcal mol^{-1}, a value which fairly agreed with QM/MM estimation of 0.9 kcal mol^{-1} between 3_{10} structure and the transition state. Additionally, performance of the new Aib FF was also demonstrated to accurately reproducing the secondary structures of the double spin-labeled Fmoc-$(Aib\text{-}Aib\text{-}TOAC)_2$-Aib-OMe heptapeptide in different solvents[147] in line with the EPR data[148]

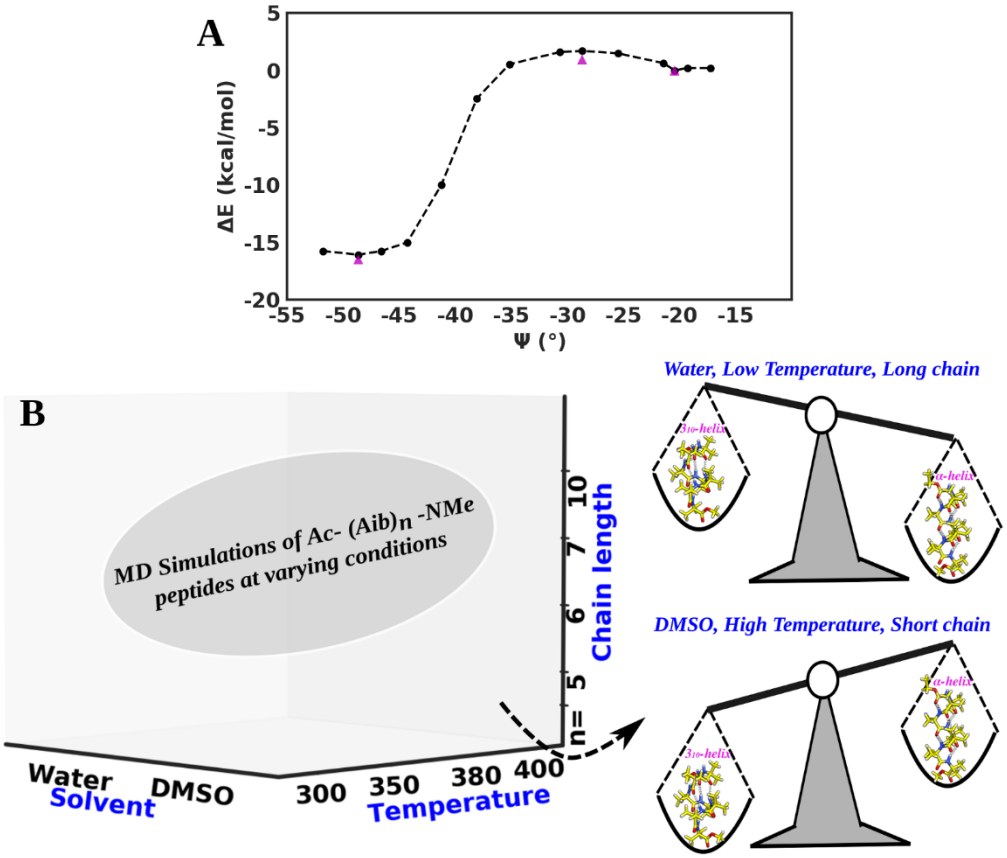

Figure 8 A) Free energy profile for interconversion from 3_{10} to α-helix of Aib heptapeptide in aqueous solution along the reaction coordinate ψ angle° as obtained from MM/PCM estimations (black curve). Magenta triangles indicate QM/PCM estimations. B) Schematic depiction of the helicogenic equilibrium shift of Aib homooligopeptides at varying physicochemical conditions, obtained from MD simulations with new Aib FF.

Incorporation of Aib residues is known to stabilize the hosting peptides in helical conformations 3_{10}/α-helix or a mixture of both the helices, whose relative stability is influenced by external factors that drive the equilibrium towards 3_{10} or α-helix. Effects of solvent polarity, temperature and peptide chain length have been reported.[130–132] For instance, Aib residues in shorter peptides tend to stabilize 3_{10} helix and distort α-helix in crystals and less polar environment. In shorter peptides, the steric contacts between the methyl groups at i and i+3 positions destabilizes the α-helix. Indeed a seminal X-ray diffraction study by Pavone et al., reported the first proof that a conversion from 3_{10} to α-helix can be achieved by mere lengthening of peptide backbone alone.[133] Following the parameterization, we addressed the influence of physicochemical conditions on the helicogenic equilibrium of Aib containing peptides with the newly derived Aib FF.[134] In this regard, extensive series of MD simulations on Aib homopolypeptides ($(Aib)_n$, n=5,6,7,10) in water and

DMSO at different temperatures (300,350,380 and 400K) were performed. Simulations in Water and DMSO at low temperature revealed rapid structural transitions between 3_{10} and α helices, while at high temperature the 3_{10} helix was predominant. For shorter peptides, predominant population of 3_{10} helix was observed at low temperature due to sterical effects mentioned above. With increase in peptide length, α helix became dominant especially in water. Overall, the results highlighted how the subtle equilibrium between the two conformations, i.e., relative helical shifts is modulated by the physicochemical conditions as summarized in Figure 5, a phenomenon that has gained enduring interest among the Aib research community.

5.3 Extended AMBER force field for cyclic $C^{\alpha,\alpha}$ -tetrasubstituted amino acids with *ab initio* parameterization

Recently we extended the popular biomolecular AMBER FF parameters to treat a wide class of cyclic α,α dialkylated amino acids.[144,147] For this purpose, an exhaustive set of Ac-A_{c6}C-NMe and Ac-TOAC-NMe dipeptide conformers were used. This included the three expected minima corresponding to the backbone dihedrals (ie., *gauche*$^+$ 60°, *trans* 180° and *gauche*$^-$ -60°) and favorable cyclic side chain arrangements (chair and twist), giving rise to 27 initial structures. The geometry optimization and subsequent structural parameters were computed at DFT level of theory with PBE0 functional and N07D basis set[149–151]. The choice of PBE0 functional was driven by the accurate estimations of several physico-chemical observables of polypeptides as shown in previous reports[96,152] Three new atom types (CJ,CL and H6, refer Figure 4) were introduced that bear most of the parameters from AMBER ff99SB[153], except the van der Waals parameters for CJ, H6 and the torsional parameters for CJ which were re-derived with high level QM calculations (refer to Grubisic et al, [147] for parameters). Atom centered charges for the peptides were obtained, following the RESP procedure as implemented in AMBER[154] The RESP procedure involves the least-squares fitting of electrostatic potential with additional hyperbolic restraints on heavy atoms, and suffice, for instance, to obtain reduced charges on buried atoms or charge equivalencing for similar atoms. The intramolecular force field parameters were initially adapted from AMBER (ff99SB) and subsequently refined till a reasonable agreement between QM and MM data was achieved. The refinement was performed with Joyce procedure and the best parameters were obtained by reducing the cost function as described in section 4.1 (*vide supra*). For TOAC, the nitroxide parameters of six-membered TEMPO skeleton ((2,2, 6,6-tetramethyl piperidin-1-yl)oxidanyl) parameterized in AMBER were adopted[150,155]. The efficiency of the newly derived FF parameters was assessed by the standard state-of-the-art comparison of the energies of different minima conformers optimized at QM and MM levels, which showed an energy RMSD of 1.32 and 1.16 kcal mol^{-1} for TOAC and A_{c6}C dipeptides, respectively (Table 3). For A_{c6}C peptides, calculations with standard AMBER parameters produced a RMSD of 4.1 kcal mol^{-1}. Superposition of the corresponding conformers (from QM and MM) yielded an overall structural RMSD of 0.04 and 0.03Å for Ac-TOAC-NMe and Ac-A_{c6}C-NMe, affirming the good agreement between the QM and MM data.

Table 3: Energies (in kcal mol^{-1}) yielded from QM (PBE0/N07D) and MM (AMBER level) optimizations of [a]Ac-TOAC-NMe and [b]Ac-A_{c6}C-Nme dipeptide conformers. [c]values from ff99SB FF. Energy values obtained from Grubisic et al.[147]

Structures[a]	QM	Derived-FF	Structures[b]	QM	Derived-FF	Std-FF[c]
I γ	0.00	0.00	I γ	0.00	0.00	0.00
II γ	0.83	0.20	II γ	0.25	0.02	0.50
III γ	0.83	0.27	III 3_{10}	2.51	3.95	0.12
IV γ	1.19	1.01	IV ε	2.92	4.23	1.25
V γ	1.81	1.87	V β	3.45	1.55	-4.48
VI β	2.39	3.58	VI α	4.00	5.15	1.16
VII ε	2.60	2.85	VII γ	5.27	5.46	7.22
VIII 3_{10}	2.60	4.86	VIII γ	5.71	4.85	7.21
IX 3_{10}	3.71	5.42	IX γ	5.76	6.40	7.94
X α	3.88	5.55	X δ	5.85	8.59	3.26
XI α	4.07	5.91	XI β	6.55	6.05	-1.20
XII ε	5.18	3.57	XII 3_{10}	7.83	8.95	7.27
XIII α	5.72	7.88	XIII β	8.32	7.41	-1.00
XIV δ	6.76	8.33	XIV 3_{10}	8.41	8.91	6.46
XV β	7.19	7.06	XV ε	9.15	9.05	7.90
RMSD		1.32			1.16	4.01

The suitability of the derived-FF to study model peptides containing TOAC residues was examined via MD simulations of TOAC dipeptide and double spin-labeled Fmoc-(Aib-Aib-TOAC)$_2$-Aib-OMe heptapeptide both gas and condensed phases. Simulations of Ac-TOAC-NMe dipeptide showed a high preference for both helix and extended conformations in water (γ, α and ε), gamma turn in *vacuo* and chloroform (γ), in line with the previous experimental and theoretical observations[156,157]. For the heptapeptide, MD simulations were performed in both *vacuo* and in aqueous solutions, where the Aib parameters were adapted from AMBER ff99SB. Solvent effects were estimated with CPCM. MM calculations showed a high preference for 3_{10} in *vacuo* over α-helix by about 3.4 kcal mol^{-1}, whilst in aqueous solution α-helix was found to be stable by about 1.32 kcal mol^{-1} over the 3_{10} helix. These results correlated with data derived from EPR and QM calculations[148.]

The new TOAC parameters were tested to study the dynamics of Phospholamban (PLB), an L-shaped amphipathic membrane peptide (52 residues) that regulates the Ca-ATPase in cardiac sarcoplasmic reticulum[158–160] PLB constitutes a membrane spanning transmembrane and surface exposed cytoplasmic domains (Figure 9). PLB derivatives in lipid bilayers, containing a TOAC residue at transmembrane (position 46) or cytoplasmic (position 11 or 24) domain, have been studied through EPR spectra. The spectra of PLB with TOAC substitutions at cytoplasmic domain revealed a direct influence of the transitions between its two distinct states on the peptide function. We performed classical simulations on PLB peptides, both native and single TOAC substituted at cytoplasmic domain (position 11), embedded in DOPC lipid bilayer in explicit solvent condition. Efficiency of the TOAC FF was assessed, computing the descriptors that highlight the secondary structural features. Analysis of the backbone φ and ψ angles of TOAC residues shows a uniform distribution around the α-helix in water (Figure 9B). Time evolution of the secondary structure of the peptide stretch with the TOAC reveals not only a stable α-helical structure but also the absence of any FF induced artefacts (Figure 9B). Further, inspection of the average α-helical content of cytoplasmic residues shows no significant change due to TOAC incorporation. Indeed, the CD spectra of native and TOAC substituted PLB revealed no apparent alterations in the secondary structure and the helical content due to TOAC incorporations.[159] These encouraging agreements support the utility of the new parameters as fruitful computational tools to design and investigate TOAC labeled model peptides, and additionally to predict theoretical spectra of such peptides.

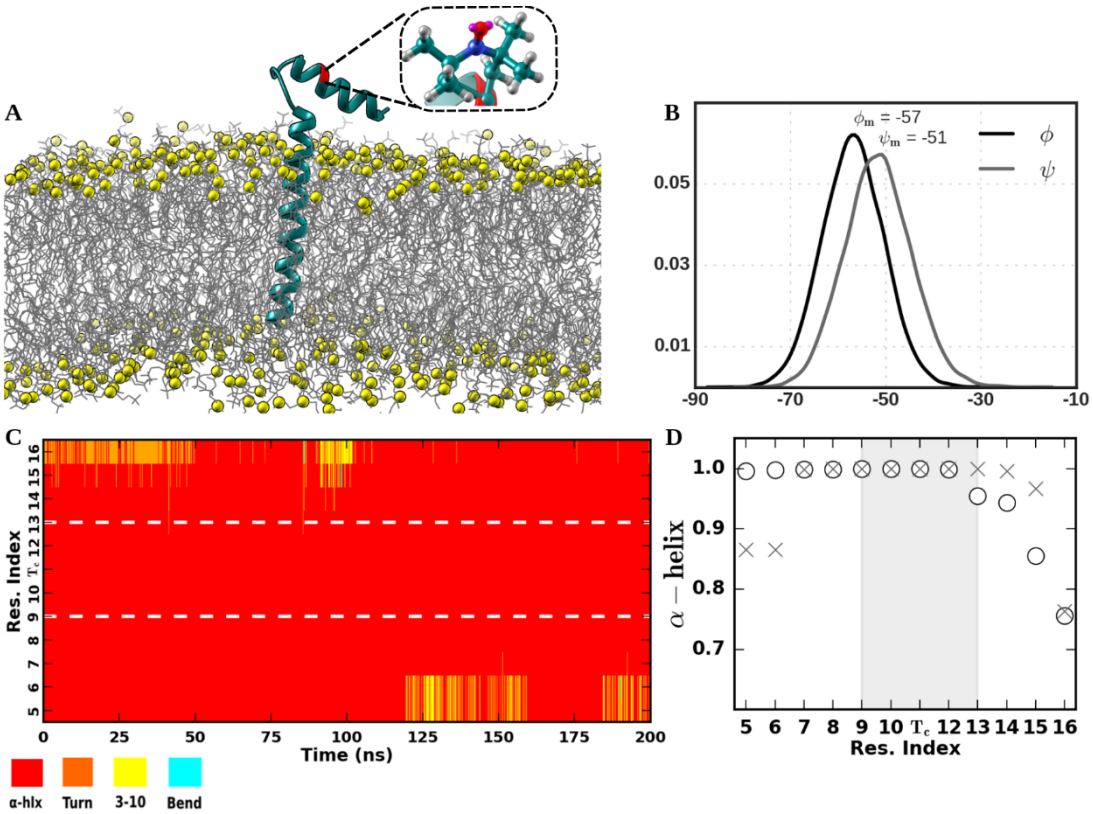

Figure 9 PLB peptide-DOPC lipid bilayer assembly. The PLB is shown as ribbon and the lipid bilayer as sticks with phosphate atoms as yellow spheres. In PLB, the TOAC introduced site (position 11) in the cytoplasmic domain is highlighted in red. B) Distribution of backbone φ and ψ angles of TOAC; along with corresponding mean values. C) Time evolved secondary structures of the cytoplasmic stretch (residue 5-16). TOAC, along with neighboring residues are enclosed in dashed lines for clarity. D) Average α-helix content as obtained from the simulations of native (o) and TOAC incorporated (x) PLB.

6. CONCLUSIONS AND PERSPECTIVES

Herein, we have presented a series of tailored-made computational protocols for a precise characterization of small biological building blocks and biopolymers through an effective combination of QM and MM. As discussed, the current state-of-art of QM methods allows to precisely describe the structure, thermodynamics and spectroscopic features of small molecules embedded in the most different types of environment. The correct application of such high level theoretical methods presents many difficulties for non specialist users but great efforts are begin take to make these techniques present in the toolbox of most computational and experimental chemists. However, the application of top-notch QM techniques greatly restricts the application field in terms of system size (number of atom and electrons) and lifetime of studied phenomena. However, a correct application of QM techniques to train more simplified parametric models, i. e. classical force fields, allows in many cases to (partly) overcome these limits, as shown in sections 4 and 5. Given their diverse features in terms of structural and spectroscopic properties, it less likely to expect a *one-size-fits-all* solution to characterize their intricate properties. As the complexity and size of the phenomena studies increases it can be foreseen that the necessity of ever more integrated quantum mechanical, semi-classical and classical protocols will be present in the future; this is necessary, among other things, to allow interdisciplinary cross fertilization of ideas and exchange of techniques (e. g. between astronomy and quantum chemistry). Hence, the methodological integration should be coupled with a development of unified frameworks and data analysis environments as discussed in reference 8.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No.ERC-2012-AdG-320951-DREAMS. The authors thank the SMART technical staff for managing the computing facilities at SNS.

REFERENCES

1. Leach, A. R. Reviews in Computational Chemistry; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1991; Vol. 2; pp 1–55.
2. Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. J Am Chem Soc 1990, 112, 1419–1427.
3. Ngo, J. T.; Karplus, M. J Am Chem Soc 1997, 119, 5657–5667.
4. Innovations in biomolecular modeling and simulations. Vol. 2: ...; RSC biomolecular sciences; Royal Soc. of Chemistry: Cambridge, 2012.
5. Bruccoleri, R. E.; Karplus, M. Biopolymers 1987, 26, 137–168.
6. Lipton, M.; Still, W. C. J Comput Chem 1988, 9, 343–355.
7. Computational strategies for spectroscopy: from small molecules to nano systems; Wiley: Hoboken, N.J, 2012.
8. Barone, V. Wiley Interdiscip Rev Comput Mol Sci 2016, 6, 86–110.
9. Purvis, G. D.; Bartlett, R. J. J Chem Phys 1982, 76, 1910–1918.
10. Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. Chem Phys Lett 1989, 157, 479–483.
11. Jurečka, P.; Hobza, P. J Am Chem Soc 2003, 125, 15608–15613.
12. Demaison, J.; Herman, M.; Liévin, J. J Chem Phys 2007, 126, 164305.
13. Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. J Chem Phys 2008, 128, 114111.

14. Feller, D.; Peterson, K. A. J Chem Phys 2009, 131, 154306.
15. Cazzoli, G.; Cludi, L.; Puzzarini, C.; Stoppa, P.; Charmet, A. P.; Tasinato, N.; Baldacci, A.; Baldan, A.; Giorgianni, S.; Larsen, R. W.; Stopkowicz, S.; Gauss, J. J Phys Chem A 2011, 115, 453–459.
16. Tasinato, N.; Regini, G.; Stoppa, P.; Charmet, A. P.; Gambi, A. J Chem Phys 2012, 136, 214302.
17. Tasinato, N.; Pietropolli Charmet, A.; Stoppa, P.; Giorgianni, S.; Gambi, A. Chem Phys 2012, 397, 55–64.
18. Barone, V.; Biczysko, M.; Bloino, J.; Egidi, F.; Puzzarini, C. J Chem Phys 2013, 138, 234303.
19. Puzzarini, C.; Biczysko, M.; Peterson, K. A.; Francisco, J. S.; Linguerri, R. J Chem Phys 2017, 147, 24302.
20. Pietropolli Charmet, A.; Stoppa, P.; Giorgianni, S.; Bloino, J.; Tasinato, N.; Carnimeo, I.; Biczysko, M.; Puzzarini, C. J Phys Chem A 2017, 121, 3305–3317.
21. Møller, C.; Plesset, M. S. Phys Rev 1934, 46, 618–622.
22. Schwabe, T.; Grimme, S. Acc Chem Res 2008, 41, 569–579.
23. Neese, F.; Hansen, A.; Wennmohs, F.; Grimme, S. Acc Chem Res 2009, 42, 641–648.
24. Neese, F. Coord Chem Rev 2009, 253, 526–563.
25. Burns, L. A.; Mayagoitia, Á. V.-; Sumpter, B. G.; Sherrill, C. D. J Chem Phys 2011, 134, 84107.
26. Antony, J.; Grimme, S.; Liakos, D. G.; Neese, F. J Phys Chem A 2011, 115, 11210–11220.
27. Burke, K. J Chem Phys 2012, 136, 150901.
28. Sure, R.; Grimme, S. J Chem Theory Comput 2015, 11, 3785–3801.
29. Tasinato, N.; Moro, D.; Stoppa, P.; Pietropolli Charmet, A.; Toninello, P.; Giorgianni, S. Appl Surf Sci 2015, 353, 986–994.
30. Fernando, A.; Weerawardene, K. L. D. M.; Karimova, N. V.; Aikens, C. M. Chem Rev 2015, 115, 6112–6216.
31. Navarro-Ruiz, J.; Ugliengo, P.; Sodupe, M.; Rimola, A. Chem Commun 2016, 52, 6873–6876.
32. Corno, M.; Delle Piane, M.; Choquet, P.; Ugliengo, P. Phys Chem Chem Phys 2017, 19, 7793–7806.
33. Barone, V.; Polimeno, A. Chem Soc Rev 2007, 36.
34. Barone, V.; Improta, R.; Rega, N. Acc Chem Res 2008, 41, 605–616.
35. Pedone, A.; Biczysko, M.; Barone, V. ChemPhysChem 2010, NA-NA.
36. Barone, V.; Baiardi, A.; Biczysko, M.; Bloino, J.; Cappelli, C.; Lipparini, F. Phys Chem Chem Phys 2012, 14.
37. Licari, D.; Baiardi, A.; Biczysko, M.; Egidi, F.; Latouche, C.; Barone, V. J Comput Chem 2015, 36, 321–334.
38. Zerbetto, M.; Licari, D.; Barone, V.; Polimeno, A. Mol Phys 2013, 111, 2746–2756.
39. Presti, D.; Pedone, A.; Licari, D.; Barone, V. J Chem Theory Comput 2017, 13, 2215–2229.
40. Licari, D.; Tasinato, N.; Spada, L.; Puzzarini, C.; Barone, V. J Chem Theory Comput 2017, 13, 4382–4396.
41. Alonso, J. L.; Pérez, C.; Eugenia Sanz, M.; López, J. C.; Blanco, S. Phys Chem Chem Phys 2009, 11, 617–627.
42. Écija, P.; Uriarte, I.; Spada, L.; Davis, B. G.; Caminati, W.; Basterretxea, F. J.; Lesarri, A.; Cocinero, E. J. Chem Commun 2016, 52, 6241–6244.
43. Puzzarini, C.; Biczysko, M.; Barone, V.; Peña, I.; Cabezas, C.; Alonso, J. L. Phys Chem Chem Phys 2013, 15, 16965.
44. Écija, P.; Cocinero, E. J.; Lesarri, A.; Fernández, J. A.; Caminati, W.; Castaño, F. J Chem Phys 2013, 138, 114304.
45. Puzzarini, C.; Biczysko, M. J Phys Chem A 2015, 119, 5386–5395.
46. Spada, L.; Tasinato, N.; Vazart, F.; Barone, V.; Caminati, W.; Puzzarini, C. Chem - Eur J

2017, 23, 4876–4883.
47. Barone, V.; Biczysko, M.; Puzzarini, C. Acc Chem Res 2015, 48, 1413–1422.
48. Puzzarini, C. Int J Quantum Chem 2017, 117, 129–138.
49. Puzzarini, C.; Baiardi, A.; Bloino, J.; Barone, V.; Murphy, T. E.; Drew, H. D.; Ali, A. Astron J 2017, 154, 82.
50. Fortenberry, R. C. Int J Quantum Chem 2017, 117, 81–91.
51. Puzzarini, C.; Biczysko, M.; Bloino, J.; Barone, V. Astrophys J 2014, 785, 107.
52. Puzzarini, C.; Ali, A.; Biczysko, M.; Barone, V. Astrophys J 2014, 792, 118.
53. Mon Not R Astron Soc Lett 2015.
54. Vazart, F.; Latouche, C.; Skouteris, D.; Balucani, N.; Barone, V. Astrophys J 2015, 810, 111.
55. Gianturco, F. A.; Satta, M.; Mendolicchio, M.; Palazzetti, F.; Piserchia, A.; Barone, V.; Wester, R. Astrophys J 2016, 830, 2.
56. Vazart, F.; Calderini, D.; Puzzarini, C.; Skouteris, D.; Barone, V. J Chem Theory Comput 2016, 12, 5385–5397.
57. Zhao, Y.; Truhlar, D. G. J Chem Theory Comput 2005, 1, 415–432.
58. Goerigk, L.; Grimme, S. J Chem Theory Comput 2011, 7, 291–309.
59. Řezáč, J.; Riley, K. E.; Hobza, P. J Chem Theory Comput 2011, 7, 2427–2438.
60. Pulay, P.; Meyer, W.; Boggs, J. E. J Chem Phys 1978, 68, 5077–5085.
61. Pawłowski, F.; Jørgensen, P.; Olsen, J.; Hegelund, F.; Helgaker, T.; Gauss, J.; Bak, K. L.; Stanton, J. F. J Chem Phys 2002, 116, 6482–6496.
62. Equilibrium molecular structures: from spectroscopy to quantum chemistry; CRC Press: Boca Raton, 2011.
63. Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. J Phys Chem A 2015, 119, 2058–2082.
64. Penocchio, E.; Piccardo, M.; Barone, V. J Chem Theory Comput 2015, 11, 4689–4707.
65. Penocchio, E.; Mendolicchio, M.; Tasinato, N.; Barone, V. Can J Chem 2016, 94, 1065–1076.
66. Mendolicchio, M.; Penocchio, E.; Licari, D.; Tasinato, N.; Barone, V. J Chem Theory Comput 2017, 13, 3060–3075.
67. Lee, C.; Yang, W.; Parr, R. G. Phys Rev B 1988, 37, 785–789.
68. Becke, A. D. J Chem Phys 1993, 98, 5648–5652.
69. Grimme, S. J Chem Phys 2006, 124, 34108.
70. Barone, V.; Biczysko, M.; Bloino, J.; Cimino, P.; Penocchio, E.; Puzzarini, C. J Chem Theory Comput 2015, 11, 4342–4363.
71. Tasinato, N.; Puzzarini, C.; Barone, V. Angew Chem Int Ed 2017.
72. Allen, T. W.; Andersen, O. S.; Roux, B. Biophys Chem 2006, 124, 251–267.
73. Rapaport, D. C. The art of molecular dynamics simulation, 2nd ed.; Cambridge University Press: Cambridge, UK ; New York, NY, 2004.
74. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Angew Chem Int Ed 2006, 45, 4064–4092.
75. Wong-ekkabut, J.; Karttunen, M. Biochim Biophys Acta BBA-Biomembr 2016.
76. van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. J Phys Chem A 2001, 105, 9396–9409.
77. Dodd, L. R.; Boone, T. D.; Theodorou, D. N. Mol Phys 1993, 78, 961–996.
78. Jedlovszky, P.; Mezei, M. J Chem Phys 1999, 111, 10770–10773.
79. Ozkan, S. B.; Meirovitch, H. J Comput Chem 2004, 25, 565–572.
80. Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. J Comput Chem 2005, 26, 1668–1688.
81. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. J Phys Chem B 2010, 114, 7830–

7843.
82. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. J Phys Chem A 1999, 103, 3596–3607.
83. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. J Chem Theory Comput 2015.
84. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. J Am Chem Soc 1992, 114, 10024–10035.
85. Mayo, S. L.; Olafson, B. D.; Goddard, W. A. J Phys Chem 1990, 94, 8897–8909.
86. Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. J Comput Chem 2013, 34, 2757–2770.
87. Betz, R. M.; Walker, R. C. J Comput Chem 2015, 36, 79–87.
88. Cacelli, I.; Prampolini, G. J Chem Theory Comput 2007, 3, 1803–1817.
89. Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. Phys Chem Chem Phys 2013, 15, 3736.
90. Cacelli, I.; Cimoli, A.; Livotto, P. R.; Prampolini, G. J Comput Chem 2012, 33, 1055–1067.
91. Chandramouli, B.; Di Maio, D.; Mancini, G.; Barone, V.; Brancato, G. PLOS ONE 2015, 10, e0120196.
92. Chandramouli, B.; Di Maio, D.; Mancini, G.; Brancato, G. Biochim Biophys Acta BBA - Biomembr 2015.
93. Di Maio, D.; Chandramouli, B.; Brancato, G. PLOS ONE 2015, 10, e0140258.
94. Chillemi, G.; Coletta, A.; Mancini, G.; Sanna, N.; Desideri, A. Theor Chem Acc 2009, 127, 293–302.
95. Dahlgren, M. K.; Schyman, P.; Tirado-Rives, J.; Jorgensen, W. L. J Chem Inf Model 2013, 53, 1191–1199.
96. Maciejewski, A.; Pasenkiewicz-Gierula, M.; Cramariuc, O.; Vattulainen, I.; Rog, T. J Phys Chem B 2014, 118, 4571–4581.
97. Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. J Comput Chem 2012, 33, 2451–2468.
98. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . J Comput Phys 1977, 23, 327–341.
99. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. J Comput Chem 1997, 18, 1463–1472.
100. Tao, P.; Wu, X.; Brooks, B. R. J Chem Phys 2012, 137, 134110.
101. Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. J Chem Theory Comput 2012, 8, 527–541.
102. Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Hermans, J. Intermolecular Forces 1981, 11, 331–342.
103. Bussi, G.; Donadio, D.; Parrinello, M. J Chem Phys 2007, 126, 14101.
104. Darden, T.; Perera, L.; Li, L.; Pedersen, L. Structure 1999, 7, R55–R60.
105. Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. Phys Chem Chem Phys 2013, 15, 10094.
106. Carnimeo, I.; Puzzarini, C.; Tasinato, N.; Stoppa, P.; Charmet, A. P.; Biczysko, M.; Cappelli, C.; Barone, V. J Chem Phys 2013, 139, 74310.
107. Heckert, M.; Kállay, M.; Tew, D. P.; Klopper, W.; Gauss, J. J Chem Phys 2006, 125, 44108.
108. Dunning, T. H. J Chem Phys 1989, 90, 1007–1023.
109. Kendall, R. A.; Dunning, T. H.; Harrison, R. J. J Chem Phys 1992, 96, 6796–6806.
110. Peterson, K. A.; Dunning, T. H. J Chem Phys 2002, 117, 10548–10560.
111. Törring, T. Berichte Bunsenges Für Phys Chem 1983, 87, 461–461.
112. Barone, V. J Chem Phys 2005, 122.
113. Bloino, J.; Biczysko, M.; Barone, V. J Chem Theory Comput 2012, 8, 1015–1036.
114. Puzzarini, C.; Biczysko, M.; Barone, V.; Largo, L.; Peña, I.; Cabezas, C.; Alonso, J. L. J Phys Chem Lett 2014, 5, 534–540.
115. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. J Comput Chem 1995, 16, 1357–1377.

116. Wang, L.-P.; Chen, J.; Van Voorhis, T. J Chem Theory Comput 2013, 9, 452–460.
117. Leach, A. R. Molecular modelling: principles and applications, 2nd ed.; Prentice Hall: Harlow, England ; New York, 2001.
118. Barone, V.; Biczysko, M.; Bloino, J.; Carta, L.; Pedone, A. Comput Theor Chem 2014, 1037, 35–48.
119. Del Frate, G.; Bellina, F.; Mancini, G.; Marianetti, G.; Minei, P.; Pucci, A.; Barone, V. Phys Chem Chem Phys 2016, 18, 9724–9733.
120. De Mitri, N.; Monti, S.; Prampolini, G.; Barone, V. J Chem Theory Comput 2013, 9, 4507–4516.
121. De Mitri, N.; Prampolini, G.; Monti, S.; Barone, V. Phys Chem Chem Phys 2014, 16, 16573–16587.
122. Prampolini, G.; Monti, S.; De Mitri, N.; Barone, V. Chem Phys Lett 2014, 601, 134–138.
123. Koenig, M.; Bottari, G.; Brancato, G.; Barone, V.; Guldi, D. M.; Torres, T. Chem Sci 2013, 4.
124. Antosiewicz, J. M.; Shugar, D. Biophys Rev 2016, 8, 151–161.
125. Antosiewicz, J. M.; Shugar, D. Biophys Rev 2016, 8, 163–177.
126. Pagliai, M.; Mancini, G.; Carnimeo, I.; De Mitri, N.; Barone, V. J Comput Chem 2016.
127. Nakaie, C. R.; Goissis, G.; Schreier, S.; Paiva, A. C. Braz J Med Biol Res Rev Bras Pesqui Medicas E Biol 1981, 14, 173–180.
128. McNulty, J. C.; Silapie, J. L.; Carnevali, M.; Farrar, C. T.; Griffin, R. G.; Formaggio, F.; Crisma, M.; Toniolo, C.; Millhauser, G. L. Biopolymers 2000, 55, 479–485.
129. Smythe, M. L.; Nakaie, C. R.; Marshall, G. R. J Am Chem Soc 1995, 117, 10555–10562.
130. Hanson, P.; Millhauser, G.; Formaggio, F.; Crisma, M.; Toniolo, C. J Am Chem Soc 1996, 118, 7618–7625.
131. Karle, I. L.; Balaram, P. Biochemistry (Mosc) 1990, 29, 6747–6756.
132. Raap, J.; Erkelens, K.; Ogrel, A.; Skladnev, D. A.; Brückner, H. J Pept Sci 2005, 11, 331–338.
133. Stoppacher, N.; Neumann, N. K. N.; Burgstaller, L.; Zeilinger, S.; Degenkolb, T.; Brückner, H.; Schuhmacher, R. Chem Biodivers 2013, 10, 734–743.
134. Michael Conlon, J.; Al-Kharrge, R.; Ahmed, E.; Raza, H.; Galadari, S.; Condamine, E. Peptides 2007, 28, 2075–2080.
135. Venkataram Prasad, B. V.; Balaram, P.; Benedetti, E. Crit Rev Biochem 1984, 16, 307–348.
136. Di Blasio, B.; Pavone, V.; Saviano, M.; Lombardi, A.; Nastri, F.; Pedone, C.; Benedetti, E.; Crisma, M.; Anzolin, M.; Toniolo, C. J Am Chem Soc 1992, 114, 6273–6278.
137. Improta, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. J Am Chem Soc 2001, 123, 3311–3322.
138. Yokum, T. S.; Gauthier, T. J.; Hammer, R. P.; McLaughlin, M. L. J Am Chem Soc 1997, 119, 1167–1168.
139. Improta, R.; Rega, N.; Aleman, C.; Barone, V. Macromolecules 2001, 34, 7550–7557.
140. Alemán, C.; Roca, R.; Luque, F. J.; Orozco, M. Proteins 1997, 28, 83–93.
141. Wang, Y.; Kuczera, K. J Phys Chem B 1997, 101, 5205–5213.
142. Yu, H.; Ramseier, M.; Bürgi, R.; van Gunsteren, W. F. ChemPhysChem 2004, 5, 633–641.
143. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. J Chem Phys 2010, 132.
144. Grubišić, S.; Brancato, G.; Barone, V. Phys Chem Chem Phys 2013, 15, 17395–17407.
145. Kennedy, D. F.; Crisma, M.; Toniolo, C.; Chapman, D. Biochemistry (Mosc) 1991, 30, 6541–6548.
146. Formaggio, F.; Crisma, M.; Rossi, P.; Scrimin, P.; Kaptein, B.; Broxterman, Q. B.; Kamphuis, J.; Toniolo, C. Chemistry 2000, 6, 4498–4504.
147. Grubišić, S.; Brancato, G.; Pedone, A.; Barone, V. Phys Chem Chem Phys 2012, 14, 15308–15320.
148. Carlotto, S.; Cimino, P.; Zerbetto, M.; Franco, L.; Corvaja, C.; Crisma, M.; Formaggio, F.; Toniolo, C.; Polimeno, A.; Barone, V. J Am Chem Soc 2007, 129, 11248–11258.

149. Adamo, C.; Barone, V. J Chem Phys 1999, 110, 6158–6170.
150. Cimino, P.; Pedone, A.; Stendardo, E.; Barone, V. Phys Chem Chem Phys 2010, 12, 3741–3746.
151. Barone, V.; Cimino, P.; Stendardo, E. J Chem Theory Comput 2008, 4, 751–764.
152. Langella, E.; Improta, R.; Barone, V. J Am Chem Soc 2002, 124, 11531–11540.
153. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Proteins Struct Funct Bioinforma 2006, 65, 712–725.
154. Wang, J.; Cieplak, P.; Kollman, P. A. J Comput Chem 2000, 21, 1049–1074.
155. Stendardo, E.; Pedone, A.; Cimino, P.; Cristina Menziani, M.; Crescenzi, O.; Barone, V. Phys Chem Chem Phys 2010, 12, 11697.
156. Madison, V.; Kopple, K. D. J Am Chem Soc 1980, 102, 4855–4863.
157. Weise, C. F.; Weisshaar, J. C. J Phys Chem B 2003, 107, 3265–3277.
158. Maclennan, D. H.; Rice, W. J.; Odermatt, A. Ann N Y Acad Sci 1997, 834, 175–185.
159. Karim, C. B.; Kirby, T. L.; Zhang, Z.; Nesmelov, Y.; Thomas, D. D. Proc Natl Acad Sci U S A 2004, 101, 14437–14442.
160. Zhang, Z.; Remmer, H. A.; Thomas, D. D.; Karim, C. B. Biopolymers 2007, 88, 29–35.