

# Well-Tempered MetaDynamics based method to evaluate universal peptidomimetics

Enrico Gandini<sup>a</sup>, Federico Dapiaggi<sup>a</sup>, Francesco Oliva<sup>a</sup>, Stefano Pieraccini<sup>\*ab</sup>, Maurizio Sironi<sup>\*ab</sup>

<sup>a</sup> Dipartimento di Chimica, Università degli Studi di Milano, Via Golgi 19, 20133, Milano, Italy. E-mail: stefano.pieraccini@unimi.it, maurizio.sironi@unimi.it

<sup>b</sup> Istituto di Scienze e Tecnologie Molecolari (INSTM), CNR, and INSTM, Udr Milano, Via Golgi 19, 20133, Milano, Italy

## **Abstract**

Peptidomimetics are a broad family of compounds designed to mimic the main structural features of peptides, avoiding their metabolic drawbacks. In particular, universal peptidomimetics use only peptide side-chain analogs to mimic different secondary structures. In this work, a novel method is proposed to identify universal peptidomimetics. It is based on a single Meta-Dynamics simulation which allows the reconstruction of the Free Energy Surface of the compound of interest. Subsequently, cluster analysis is carried out to obtain representative structures. Such conformers are then compared to ideal secondary structures to assess their mimicking ability. This protocol was validated against known universal peptidomimetics.

## **Keywords**

minimalist mimics, universal mimics, peptidomimetics, metadynamics, well-tempered metadynamics

## Introduction

Peptides are an important family of molecules, having an essential role in numerous biological processes, ranging from cell signaling to immune response. Accordingly, they are thoroughly investigated as potential bioactive agents [1-4]. Natural peptides are generally not considered as optimal drug candidates by the medicinal chemistry community [5], because of their limited stability toward proteolysis and the generally poor transport properties across body barriers. Moreover, their inherent flexibility enables interaction with multiple receptors besides the desired target, increasing the risk of undesired side effects [6]. Thus, it is expedient to design small peptide-like chains encompassing their pharmacophore while improving their pharmacokinetic and pharmacodynamic properties. Such molecules are called “peptidomimetics” [7]. Peptidomimetics have been broadly studied at theoretical and experimental level [8-10].

Many peptidomimetics closely resemble peptides, mimicking both backbone and side chains. Usually, they are actually peptides containing modified, non-natural residues to constrain their structure in the relevant active conformations, e.g. by applying peripheral groups [11,12] or by making a cyclic peptide [13,14].

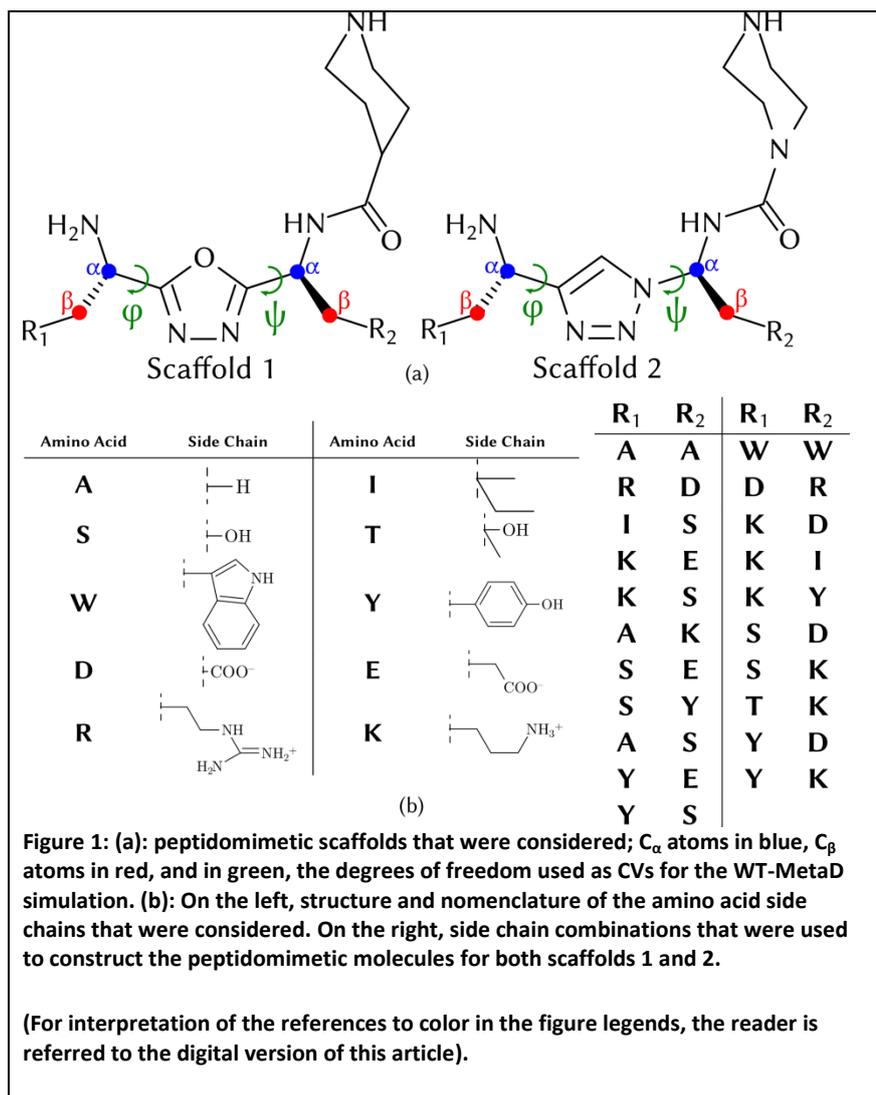
Another interesting type of peptidomimetics are compounds that achieve mimicry using only side chains analogs. The first example of this type of compounds are the  $\beta$  turn mimics by Hirschmann and Smith [15-17]. This type of peptidomimetics were thoroughly studied by several authors [18-23], and has been defined as minimalist peptidomimetics [18]. Even though minimalist peptidomimetics only mimic peptide side chains, they have the potential to mimic hot-spots residues at protein-protein interfaces [24-26], since side chain substituents account for about 80% of the interactions [27-29].

Minimalist peptidomimetics should be synthetically accessible with different side chains, should display thermodynamically and kinetically accessible conformations for induced fit and finally side chains orientation should depend on a limited number of significant degrees of freedom [18].

Recently, a subset of the minimalist peptidomimetics family, comprising molecules that are flexible enough to mimic different kinds of secondary structures, has been identified [18]. These compounds have been labeled universal peptidomimetics. Burgess and coworkers also developed a protocol to evaluate the performance of a compound as a potential universal peptidomimetic [21]. This protocol is based on Density Functional Theory (DFT) calculations to evaluate the energy barriers between relevant conformations, and on Quenched Molecular Dynamics (QMD) and cluster analysis to explore the conformational ensemble of the molecules.

In this work we propose an alternative method to identify universal peptidomimetics based on MetaDynamics (MetaD) [30-32] and cluster analysis. MetaD is an enhanced sampling technique based on phase space coarse graining through the identification of collective variables describing the system that allows the reconstruction of its free energy surface [33-37]. In particular, Well-Tempered MetaD (WTMetaD) [38] has been employed to explore the conformational space of the investigated compounds and to estimate the energy barriers between the most populated conformers, whereas subsequent cluster analysis has been used to identify representative conformations from the Free Energy Surfaces (FESs) minima. To evaluate the ability of the studied compounds to mimic typical secondary structure elements, the  $C_{\alpha}$ - $C_{\beta}$  vectors (as defined in Figure 1) of the relevant conformations were overlaid on the corresponding atoms of ideal secondary structures. Our protocol requires just one production phase simulation for every tested compound and provides thermodynamically relevant conformations that effectively mimic secondary structures elements. To validate our protocol, we tested it on two universal peptidomimetic scaffolds proposed and previously studied by Ko *et al.* We investigated their conformational and energetic properties, considering the effect of the scaffold structure and sidechains variability, defining which secondary structure elements are better reproduced. Our results, coherent with previous observations, assess the viability of the proposed protocol as a validation tool for universal peptidomimetics.

## Materials and Methods



The peptidomimetic scaffolds considered in this work are presented in Figure 1 Panel (a). Scaffold 1 is based on a 1,3,4 - oxadiazole, whereas Scaffold 2 is based on a 1,2,3 - triazole. These two scaffolds were proposed by Ko *et al.* [18] as universal peptidomimetics. In particular Scaffold 1 was also synthesized with a variety of side chains, reported in Figure 1 Panel (b). We simulated compounds based on Scaffold 1 and 2 with the combinations of side chains R<sub>1</sub> and R<sub>2</sub> reported in Figure 1 Panel (b).

In order to investigate the steric hindrance and charge effects of substituents R<sub>1</sub> and R<sub>2</sub> we tested, as limiting cases, scaffolds 1 and 2 with side chains of Alanine and Tryptophan or Arginine and Aspartic Acid respectively.

Therefore, we considered twenty-one side chain combinations, and two scaffolds, amounting to forty-two molecules Figure 1 Panel (b). We will henceforth reference to these compounds with the following nomenclature: NR<sub>1</sub>R<sub>2</sub>, where N is the number of the Scaffold, and R<sub>1</sub>, R<sub>2</sub> are letters, identical to the one-letter codes of the amino acids. (*i.e.* 1YE is the compound based on Scaffold 1, with R<sub>1</sub> and R<sub>2</sub> the side chains of Tyrosine and Glutamic Acid respectively).

The molecules were initially optimized with Gaussian 09 (Revision D.01) [39] using Restricted Hartree-Fock (RHF) with basis set 6-31G(d). Each optimized structure has been solvated in a cubic box of 4 nm with TIP3P water. Every system has been submitted to geometry optimization with the steepest descent algorithm with a convergence of 100 kJ mol<sup>-1</sup> nm<sup>-1</sup>. Then we performed a 1 ns NVT equilibration at 300 K, using velocity rescaling as temperature coupling algorithm and τ<sub>T</sub> = 0.01 ps. A 1 ns NPT equilibration followed, using the Berendsen pressure coupling algorithm with isotropic coupling, a reference pressure of 1 bar, τ<sub>p</sub> = 2 ps, and the reference isothermal compressibility of water: 4.5 × 10<sup>-5</sup> bar<sup>-1</sup>.

After the equilibration phases, we ran a 50 ns WMetaD simulation with a 20.0 kJ mol<sup>-1</sup> bias-factor, at a reference temperature of 300 K. The  $\phi$  and  $\psi$  torsion angles (defined in Figure 1 Panel (a)) have been used as Collective Variables (CVs), because they describe the relative orientation of the side chains. A Gaussian hill with  $\sigma = 0.1$  rad for both  $\phi$  and  $\psi$  and an initial height of 0.5 kJ mol<sup>-1</sup> was added once every 100 steps. The simulation was sped up by saving the Gaussian hills on a grid that ranged from  $-\pi$  to  $+\pi$  rad and spaced 0.025 rad in the directions of both CVs.

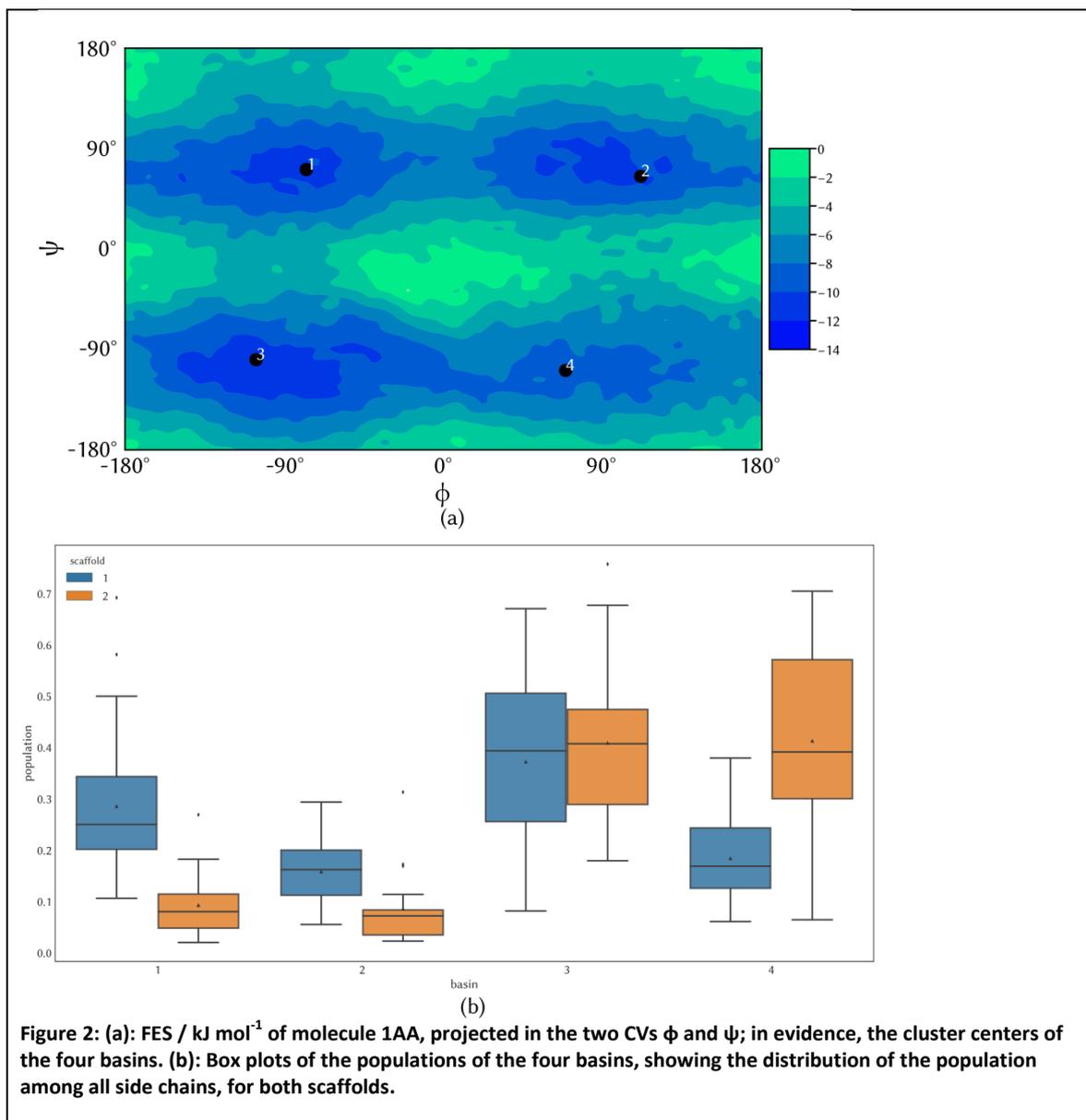
The molecules have been described using the Generalized Amber Force Field (GAFF) [40,41]. MD simulations were performed using the leap-frog algorithm with 2 fs time-step, with holonomic constraints on every bond enforced using the LINCS algorithm. Simulations and subsequent analysis were performed with the GROMACS 5.0.4 [42,43] program suite along with PLUMED version 2.2.2 [44].

Cluster analysis were performed with the GROMOS [45] algorithm.

The reference ideal secondary structures were obtained from ProBuilder On-line, a service based on VEGA [46].

## Results and Discussion

We calculated the FES for each molecule in Figure 1. In general, the FES exhibits four distinct minima corresponding to different conformations, separated by low free energy barriers. As an example, FES of molecule 1AA is shown in Figure 2 Panel (a) (FES of all other molecules are reported in Supporting Informations). This is in good agreement with the definition of minimalist mimics: the preferred conformation should be rather independent on the different type of side chains of the compound.



While the general shape of the FES is the same for every molecule, the depth and width of each basin differ among the different molecules.

First of all, we analyzed the influence of the scaffold on the population of each of the four basins (Figure 2 Panel (b)). The population of each basin is defined accordingly to Equation 1.

$$p_i = \frac{\sum_{\mu \in i} e^{\frac{-F_{\mu}}{k_B T}}}{\sum_i^N \sum_{\mu \in i} e^{\frac{-F_{\mu}}{k_B T}}}$$

Equation 1

Where  $i$  is a basin,  $N$  is the total number of basins (four in this case), and  $F_{\mu}$  is the Free energy of point  $\mu$  in basin  $i$ .

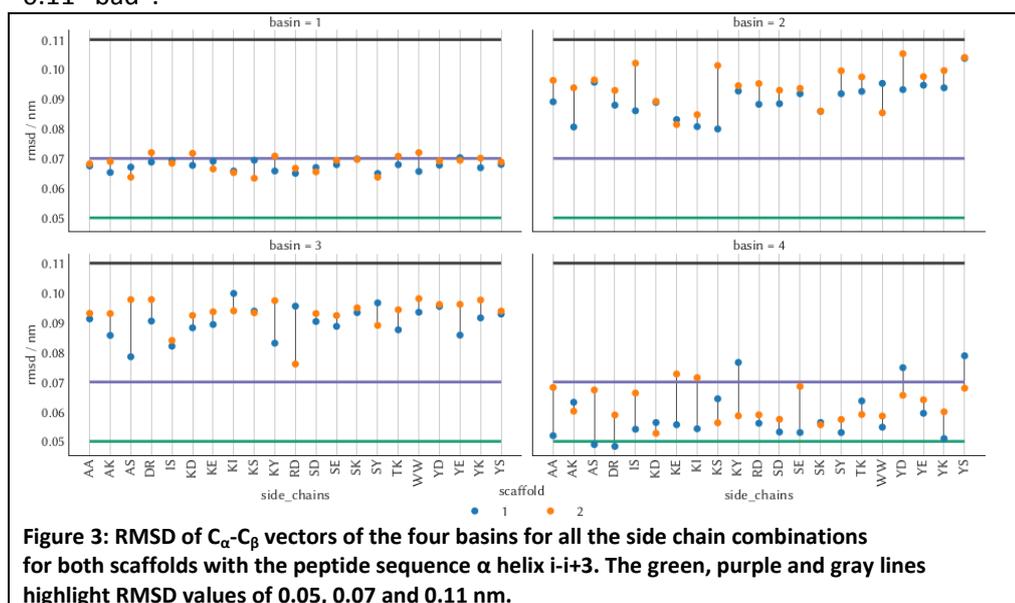
From Figure 2 Panel (b) we notice that there are significant differences in the population of the four basins depending on the scaffold type. In particular, scaffold 1 preferentially populates basin one, whereas basin 4 is preferentially populated by scaffold 2. Additionally, the third basin is similarly occupied by the two scaffolds. Moreover, a slight preference for the third basin over the second is observed. These observations suggest that one scaffold will better mimic a secondary structure whose  $C_{\alpha}$ - $C_{\beta}$  vectors have the

characteristic orientation of a particular basin. Within each basin, the side chains influence which conformation is preferred.

The representative structures of the four basins were obtained performing cluster analysis with GROMOS algorithm [45], using 0.1 nm RMSD cutoff. Structures were clustered based on  $C_{\alpha}$  and  $C_{\beta}$  atoms, highlighted in Figure 1 Panel (a).

We obtained a single cluster for each basin, and the central molecule was chosen as representative for the whole basin. We tested the ability of the aforementioned central structures to mimic ideal secondary structures, as our aim was to highlight the difference between the two scaffolds and among the side chains. To this purpose we calculated the RMSD between central structure and ideal secondary structures superimposing the  $C_{\alpha}$  and  $C_{\beta}$  atoms

To evaluate the quality of an overlay, we adopted the same criterion employed by Burgess and coworkers: from 0.03 to 0.05 nm “very good”, from 0.05 to 0.07 nm “good”, from 0.07 to 0.11 “poor” and greater than 0.11 “bad”.



At first, we assessed the effect of side chains on the superimposition quality between peptidomimetics and an archetypical  $\alpha$  helical secondary structure. In particular we evaluated the overlay between the side chains of our peptidomimetics and the side chains of the i-i+3 residues of the ideal  $\alpha$  helix (Figure 3).

From now on, we will refer to specific pairs of amino acid belonging to an ideal secondary structure using the nomenclature “i-i+n” where n is an integer  $\geq 1$  indicating the reciprocal positions along the sequence.

In Figure 3 the green, purple and dark gray lines mark the values of RMSD that act as boundaries between two overlay quality ranges. As an example, an RMSD value between the purple and the dark gray line manifests “poor” overlay, and an RMSD value under the green line stands for “very good” overlay.

RMSD values have a greater central tendency for structures of basin 1 than for structures of the other basins. The secondary structure  $\alpha$  helix i-i+3 is better mimicked by structures in basins 1 and 4, for both scaffolds, independently from the side chains. Analogously, no significant side chain dependence of the overlay quality was observed for any other secondary structure (data not shown).

These results support the hypotheses that scaffolds 1 and 2 are universal peptidomimetics, as previously inferred by Ko et. al.

Since the different side chains do not significantly affect the ability of our scaffolds to mimic a secondary structure, from now on we will focus our analysis on the simplest molecules (1AA and 2AA), to evaluate their peptidomimetic quality, by overlaying them on various ideal secondary structures (Figure 4).

On the whole, we observe that both scaffolds are able to mimic many different secondary structures, as expected for universal peptidomimetics. We recognize that our scaffolds are suited to mimic various kinds of turns. In particular, basins 1 and 4 are very good at mimicking  $\beta$  turn i-i+2, while basins 2 and 3 can mimic very well the inverse  $\beta$  turn i+1-i+2.

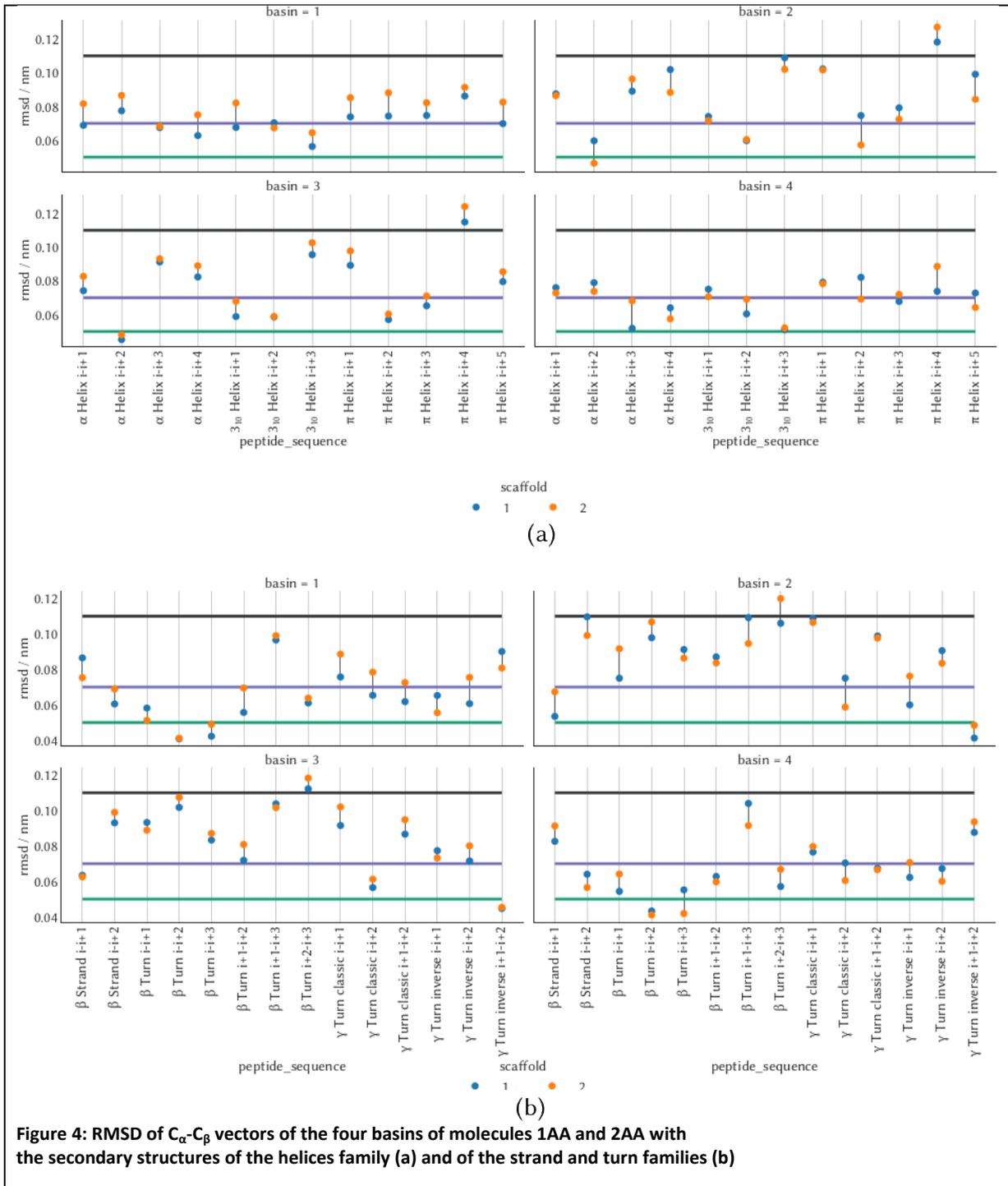
Concerning  $\alpha$  helices, secondary structures from i-i+2 to i-i+4 are well mimicked by both scaffolds.

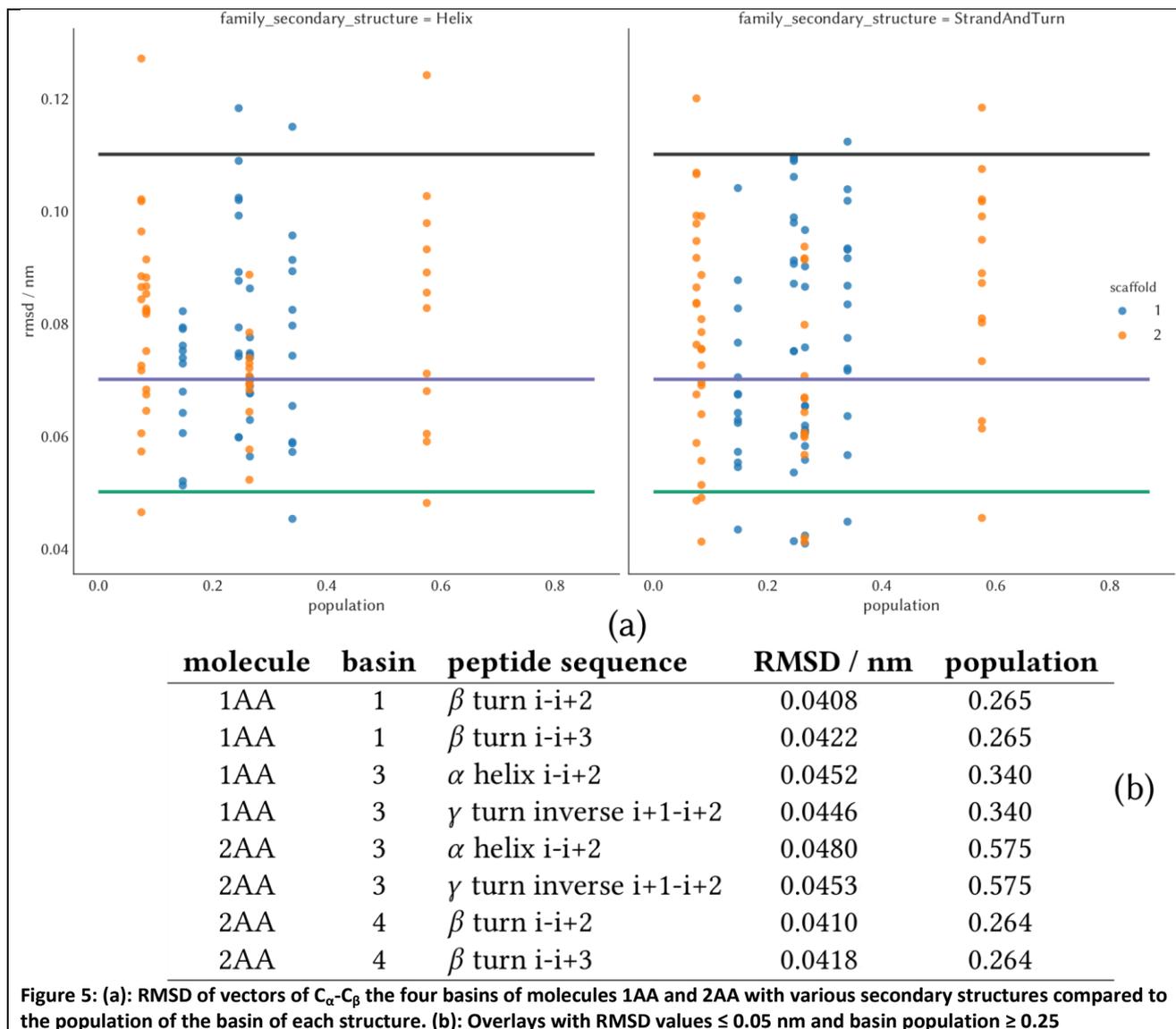
Finally, our scaffolds are able to mimic  $\beta$  strands  $i-i+1$  and  $i-i+2$  with a good RMSD.

Our protocol allowed us not only to evaluate the geometric similarities between peptidomimetics and peptide sequences, but also to correlate this information with the population of the different conformers.

In Figure 5 Panel (a) we plotted the RMSD values for the overlays of molecules 1AA and 2AA with all the secondary structures against the population of the four basins of the molecules (we made separate plots for the secondary structures of the helix family and of the strand - turn families). This analysis points out that many peptidomimetics which overlay with at least “good” RMSD values also display a significant population. Actually, we noticed that a few structures share two of the important characteristics of a promising peptidomimetic: they have both a “very good” RMSD (with respect to a certain secondary structure) and they are also highly populated ( $\geq 0.25$ ). These structures are reported in Figure 5 Panel (b). Structures highlighted in this table are the most promising to mimic the corresponding secondary structures.

As an example, Figure 6 Panel (a) shows the overlay between molecule 2AA basin 3 and  $\alpha$  helix  $i-i+2$ , whereas Figure 6 Panel (b) shows the overlay between basin 1 of molecule 1AA and secondary structure  $i-i+2$  of an ideal  $\beta$  turn.





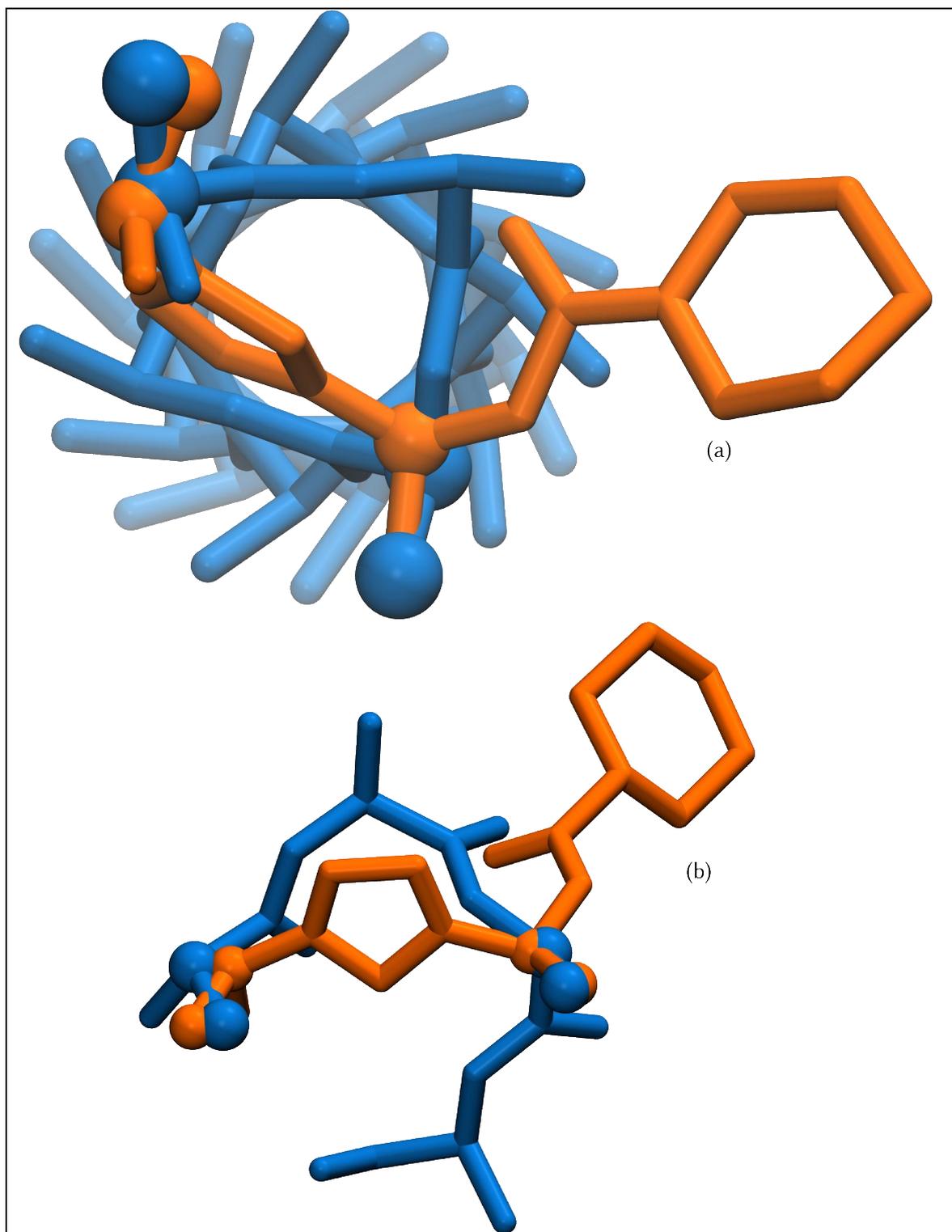


Figure 6: (a): Central structure of basin 3 of molecule 2AA (orange) overlaid on the secondary structure  $\alpha$  helix  $i-i+2$  (blue). (b): Central structure of basin 1 of molecule 1AA (orange) overlaid on the secondary structure  $\beta$  turn  $i-i+2$ . Backbones are depicted with sticks,  $C_\alpha$  and  $C_\beta$  atoms are highlighted with spheres.

## Conclusions

In this work we developed a new protocol to evaluate the ability of different classes of molecules to act as universal peptidomimetics. A WTMetaD simulation allowed us to investigate the conformational space relative to the chosen collective variables for every molecule. An approach based on cluster analysis was developed to extract structures from the dynamics that are representative of the FES minima and suitable for further analysis. Those conformers were overlaid with ideal secondary structures, confirming that scaffolds 1 and 2 are suitable to mimic them, with a clear propensity for short sequences in helices and turns. The ability of molecules to mimic secondary structures is independent on the side chains. Finally, we correlated the mimicking ability of our compounds with the populations of basins, pointing out high populated structures exhibiting a very good overlay with ideal secondary structures. This protocol allows to investigate the peptidomimetic character of a molecule using a single WTMetaD simulation.

## Aknowledgments

We acknowledge Università degli Studi di Milano for financial support (Development Plan of Athenaeum grant - line 2 action A 2017).

## Bibliography

- [1] D. J. Craik, D. P. Fairlie, S. Liras, D. Price, *Chemical Biology & Drug Design* **2013**, *81*, 136.
- [2] D. Gaspar, A. S. Veiga, M. A. R. B. Castanho, *Front. Microbiol.* **2013**, *4*.
- [3] F. Dapiaggi, S. Pieraccini, D. Potenza, F. Vasile, H. Macut, S. Pellegrino, et al., *New Journal of Chemistry* **2017**, *41*, 4308.
- [4] S. Pieraccini, G. Saladino, G. Cappelletti, D. Cartelli, P. Francescato, G. Speranza, et al., *Nature chemistry* **2009**, *1*, 642.
- [5] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *Journal of medicinal chemistry* **2002**, *45*, 2615.
- [6] G. Shankar, S. Arkin, L. Cocea, V. Devanarayan, S. Kirshner, A. Kromminga, et al., *AAPS J* **2014**, *16*, 658.
- [7] P. Farmer, E. Ariens, *Trends in Pharmacological Sciences* **1982**, *3*, 362.
- [8] Y. Niu, R. E. Wang, H. Wu, J. Cai, *Future Medicinal Chemistry* **2012**, *4*, 1853.
- [9] M. K. Pathak, V. Jha, N. K. Jain, N. Shrivastava, N. Yadav, *Indo American Journal of Pharmaceutical Research* **5**, *9*.
- [10] M. C. Bellucci, M. Frigerio, C. Castellano, F. Meneghetti, A. Sacchetti, A. Volonterio, *Organic & Biomolecular Chemistry* **2018**, *16*, 521.
- [11] D. Pei, P. G. Schultz, *Journal of the American Chemical Society* **1991**, *113*, 9398.
- [12] E. Cabezas, A. C. Satterthwait, *Journal of the American Chemical Society* **1999**, *121*, 3862.
- [13] A. Patgiri, A. L. Jochim, P. Arora, *Accounts of Chemical Research* **2008**, *41*, 1289.
- [14] D. F. Veber, R. M. Freidinger, D. S. Perlow, W. J. Paleveda, F. W. Holly, R. G. Strachan, et al., *Nature* **1981**, *292*, 55.
- [15] R. Hirschmann, K. Nicolaou, S. Pietranico, J. Salvino, E. M. Leahy, P. A. Sprengeler, et al., *Journal of the American Chemical Society* **1992**, *114*, 9217.
- [16] R. Hirschmann, J. Hynes, M. A. Cichy-Knight, R. D. van Rijn, P. A. Sprengeler, P. G. Spoor, et al., *Journal of medicinal chemistry* **1998**, *41*, 1382.
- [17] R. Hirschmann, P. A. Sprengeler, T. Kawasaki, J. W. Leahy, W. C. Shakespeare, A. B. Smith III, *Journal of the American Chemical Society* **1992**, *114*, 9699.
- [18] E. Ko, J. Liu, L. M. Perez, G. Lu, A. Schaefer, K. Burgess, *Journal of the American Chemical Society* **2011**, *133*, 462.
- [19] E. Ko, J. Liu, K. Burgess, *Chemical Society Reviews* **2011**, *40*, 4411.
- [20] D. Chen, F. Brahim, Y. Angell, Y.-C. Li, J. Moskowicz, H. U. Saragovi, et al., *ACS chemical biology* **2009**, *4*, 769.
- [21] D. Xin, E. Ko, L. M. Perez, T. R. Ioerger, K. Burgess, *Organic & Biomolecular Chemistry* **2013**, *11*, 7789.
- [22] Y. Angell, D. Chen, F. Brahim, H. U. Saragovi, K. Burgess, *Journal of the American Chemical Society* **2008**, *130*, 556.
- [23] M. S. Christodoulou, F. Zunino, V. Zuco, S. Borrelli, D. Comi, G. Fontana, et al., *ChemMedChem* **2012**, *7*, 2134.
- [24] I. Baran, R. S. Varekova, L. Parthasarathi, S. Suchomel, F. Casey, D. C. Shields, *Journal of chemical information and modeling* **2007**, *47*, 464.
- [25] M. Stucchi, G. Grazioso, C. Lammi, S. Manara, C. Zanoni, A. Arnoldi, et al., *Organic & Biomolecular Chemistry* **2016**, *14*, 9736.
- [26] L. R. Whitby, D. L. Boger, *Accounts of Chemical Research* **2012**, *45*, 1698.
- [27] L. L. Conte, C. Chothia, J. Janin, *Journal of molecular biology* **1999**, *285*, 2177.
- [28] S. Jones, J. M. Thornton, *Proceedings of the National Academy of Sciences* **1996**, *93*, 13.
- [29] I. S. Moreira, P. A. Fernandes, M. J. Ramos, *Proteins: Structure, Function, and Bioinformatics* **2007**, *68*, 803.
- [30] A. Barducci, M. Bonomi, M. Parrinello, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 826.
- [31] A. Laio, F. L. Gervasio, *Reports on Progress in Physics* **2008**, *71*, 126601.
- [32] L. Sutto, S. Marsili, F. L. Gervasio, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 771.
- [33] A. Spitaleri, M. Ghitti, S. Mari, L. Alberici, C. Traversari, G.-P. Rizzardi, et al., *Angewandte Chemie International Edition* **2011**, *50*, 1832.
- [34] C. Paissoni, M. Ghitti, L. Belvisi, A. Spitaleri, G. Musco, *Chemistry - A European Journal* **2015**, *21*, 14165.
- [35] G. Saladino, S. Pieraccini, S. Rendine, T. Recca, P. Francescato, G. Speranza, et al., *Journal of the American Chemical Society* **2011**, *133*, 2897.
- [36] G. Saladino, M. Marenchino, S. Pieraccini, R. Campos-Olivas, M. Sironi, F. L. Gervasio, *J. Chem. Theory Comput.* **2011**, *7*, 3846.
- [37] F. Doro, G. Saladino, L. Belvisi, M. Civera, F. L. Gervasio, *J. Chem. Theory Comput.* **2015**, *11*, 1354.
- [38] A. Barducci, G. Bussi, M. Parrinello, *Physical review letters* **2008**, *100*, 020603.

- [39] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, et al., *Gaussian 09, revision D. 01*, Gaussian, Inc., Wallingford CT, 2009.
- [40] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *Journal of computational chemistry* **2004**, *25*, 1157.
- [41] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *Journal of molecular graphics and modelling* **2006**, *25*, 247.
- [42] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *Journal of computational chemistry* **2005**, *26*, 1701.
- [43] M. Abraham, D. Van Der Spoel, E. Lindahl, B. Hess, others, GROMACS user manual version 5.0. 4 **2014**.
- [44] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Computer Physics Communications* **2014**, *185*, 604.
- [45] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angewandte Chemie International Edition* **1999**, *38*, 236.
- [46] A. Pedretti, L. Villa, G. Vistoli, *J Comput Aided Mol Des* **2004**, *18*, 167.