Rationalization of the Membrane Permeability Differences in a Series of Analogue Cyclic Decapeptides

Jagna Witek[1], Shuzhe Wang[1], Benjamin Schroeder[1], Robin Lingwood[1], Andreas Dounas[1], Hans-Jörg Roth[2], Marianne Fouché[2], Markus Blatter[2], Oliver Lemke[3], Bettina Keller[3], and Sereina Riniker*[1]

[1]*Laboratory of Physical Chemistry, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland*
[2]*Novartis Institutes for BioMedical Research, Novartis Pharma AG, Novartis Campus, 4056 Basel, Switzerland*
[3]*Department of Biology, Chemistry, Pharmacy, Freie Universität Berlin, Takustrasse 3, 14195 Berlin, Germany*

Abstract

Cyclization and selected backbone N-methylations are found to be often necessary but not sufficient conditions for peptidic drugs to have a good bioavailability. Thus, the design of cyclic peptides with good passive membrane permeability and good solubility remains a challenge. The backbone scaffold of a recently published series of cyclic decapeptides with six selected backbone N-methylations was designed to favor the adoption of a *closed* conformation with β-turns and four transannular hydrogen bonds. Although this conformation was indeed adopted by the peptides as determined by NMR measurements, substantial differences in the membrane permeability were observed. In this work, we aim to rationalize the impact of discrete side chain modifications on membrane permeability for six of these cyclic decapeptides. The thermodynamic and kinetic properties were investigated using molecular dynamics simulations and Markov state modeling in water and chloroform. The study highlights the influence that side-chain modifications can have on the backbone conformation. Peptides with a D-proline in the β-turns were more likely to adopt even in water the *closed* conformation with transannular hydrogen bonds, which facilitates transition through the membrane. The population of the *closed* conformation in water was found to correlate positively with PAMPA log P_e.

1 Introduction

Peptides have gained interest as potential inhibitors for difficult protein-protein interactions with flat interfaces,[1–8] but they often suffer from poor bioavailability due to their size and polarity. The positive effects of cyclization and selected backbone N-methylations on passive membrane permeability were noted early on.[9–12] Cyclic peptides with N-methylated backbone amides are thus attractive alternatives to linear peptides as therapeutics.[13–15] Due to the higher rigidity and decreased polarity, their conformations are more defined and they are metabolically more stable leading to lower *in vivo* clearance.[16] Nevertheless, the membrane permeability of cyclic peptides often remains too low for the use as oral drugs. A well-known exception is the highly lipophilic natural product cyclosporine A

*Corresponding author: email: sriniker@ethz.ch, ORCID: 0000-0003-1893-4031

(CsA),[17] with a medium to good membrane permeability and oral bioavailability despite its size. CsA is a cyclic undecamer with seven backbone *N*-methylations. According to the crystal structure and the NMR structure in apolar solvents, its backbone forms two opposite β-strands stabilized by two type II' β-turns and three transannular hydrogen bonds (H-bonds) (Figure 1A). CsA binds to its target cyclophilin in the cytosol, however, in an *open* conformation, i.e. without the aforementioned conformational features of transannular H-bonds and β-turns.[18] The mechanism for membrane crossing of CsA is thus thought to rely on its chameleonic conformational behaviour, changing from an *open* conformation in water to the *closed* conformation inside the membrane interior, where the polar groups are shielded from the apolar environment.[19–21]

More recently, examples of membrane-permeable cyclic hexapeptides[19,20,22–24] and cyclic decapeptides[25,26] have been reported. The latter ones have in common a backbone scaffold with six *N*-methylations, which was designed to favor the formation of all four possible transannular H-bonds (Figure 1B). In addition, a D-Pro, D-NMe-Ala or glycine was introduced at position 4 and 9, followed by L-amino acids at position 5 and 10 in order to favor type II' β-turns, as proposed by Hutchinson *et al.*[27] This provided a model system for the rationalization of permeability differences in relation to modest variation of the side chains.

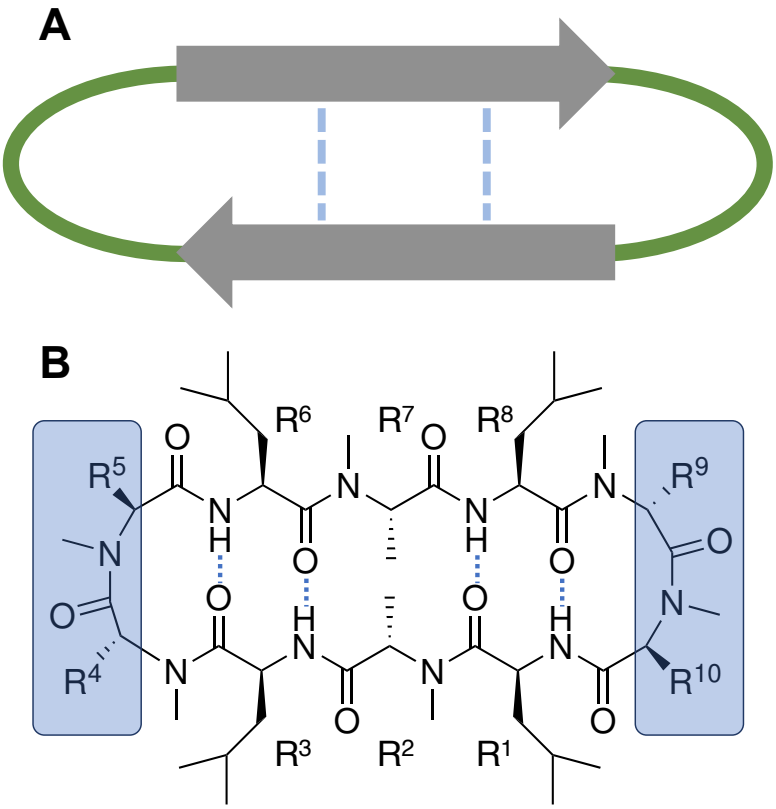

Figure 1: (A) Schematic representation of the *closed* conformation of cyclic peptides with antiparallel β-strands and type II" β-turns. β-strands are colored grey, type II" β-turns are green. (B) Scaffold of the studied series of cyclic decapeptides.[25,26] Type II' β-turns are highlighted in blue, intramolecular H-bonds are marked as blue dashed lines.

An early example to rationalize the membrane permeability of cyclic peptides is the study by Borchardt and co-workers on cyclic hexapeptides.[28–30] Based on NMR measurements and molecular dynamics (MD) simulations in water, they explained the good permeability of the cyclic peptides

versus their linear analogues through the preferred formation of a β-turn conformation (due to a more rigid backbone), which facilitates intramolecular H-bonds even in a polar environment. Recently, we have reported kinetic models of CsA based on multi-microsecond MD simulations in water and chloroform to mimic the conformational behavior outside and inside the membrane.[31] The *closed* conformation as well as a half-open conformation were found to be populated in both the apolar and polar environment (i.e. *congruent* conformations), supporting the hypothesis that CsA adopts a "permeable" conformation in the aqueous phase or head-group region before crossing the apolar core of a membrane. Such *congruent* conformations facilitate the insertion of the peptide into the membrane as the cost of desolvation is reduced. While for smaller cyclic peptides such as hexamers a rapid (i.e. not rate-limiting) interconversion between *open* and *closed* conformations in water can be assumed,[19] the same interconversion rates of larger cyclic peptides such as CsA may become slow enough to be rate-limiting due to the increased flexibility, and thus affect the permeability. The interconversion rates for the synthetic derivative cyclosporine E (CsE), which differs structurally from CsA only in one missing backbone *N*-methylation, were found to be nearly one order of magnitude slower than for CsA.[32] The slower interconversion between *open* and *closed* conformations may serve as a rationale for the lower membrane permeability of CsE compared to CsA observed experimentally. The additional N-H in CsE, which in principle increases the polarity compared to CsA, is forming transannular H-bonds in the *closed conformation* and is thus shielded from the apolar environment. Craik and co-workers showed recently with MD simulations of CsA and an explicit membrane at elevated temperature (449 K) that the chameleonic nature of CsA is important for permeation.[33] When CsA was constrained in the open conformation, the peptide resided preferentially among the lipid head groups, whereas it was located in the membrane interior when constrained in the closed conformation. Only with full flexibility, insertion, crossing and exit from the membrane was observed.[33] The chameleonic nature of CsA as the basis for the unusually good membrane permeability is further supported by the analysis of the 3D polar surface area (3D-PSA) of *open* and *closed* conformations of macrocycles.[34]

In this work, we investigated six members of the recently reported series of cyclic decapeptides[25,26] that fall into four classes (I-IV) based on membrane permeability and aqueous solubility, according to the biopharmaceutic drug classification (BCS)[35] system. Class I corresponds to peptides with high permeability and high solubility, class II to peptides with high permeability but low solubility, class III to peptides with low permeability but high solubility, and class IV to peptides with both low permeability and low solubility. The six peptides differ only in the size and polarity of the side chains at the β-turns (positions 4, 5, 9 and 10, see Figure 1B). The composition and experimental results are summarized in Table 1. We decided to focus here on experimental permeability coefficients measured in a parallel artificial membrane permeability assay (PAMPA),[36] as the conditions in PAMPA are most similar to the simulation setup used. Permeability results from a cell-based assay are reported in Ref. 25. NMR measurements in chloroform showed that the closed conformation was adopted by all six peptides, however, large differences in membrane permeability were nevertheless observed. This indicates that the ability to form intramolecular H-bonds in an apolar environment in order to shield unmethylated amide groups as such is not sufficient to explain good permeabilities. This point is emphasized by the fact that the permeability of these peptides does not correlate with their calculated 3D-PSA[34] values (Table 1). We hypothesized therefore that only considering the conformations in the apolar environment would not allow to explain the permeation behavior of such peptides.

Table 1: Side chain substitutions at positions 4, 5, 9 and 10 (Figure 1B) for the six cyclic decapeptides studied with experimental data taken from Refs. 25,26 (bioavailability (BAV) and solubility (S)) or determined for this study (PAMPA log P_e and logD at pH 7.4). In addition, the calculated octanol/water partition coefficient (clogP[37]) and 3D polar surface area (3D-PSA[34]) are given.

Class	**ID**	**R**^{4}	**R**^{5}	**R**^{9}	**R**^{10}	**PAMPA log P_e**	**BAV [%]**	**S [mM]**	**logD pH 7.4**	**clogP**	**3D-PSA [Å^2]**
I	**1**	D-Pro	MePhe	D-Pro	MeThr	-4.3	79	0.065	>6.1	10.8	223
I	**2**	D-Pro	MeAla	D-Pro	MeAla	-4.6	130*	0.794	4.6	10.1	210
I	**3**	D-Pro	MeThr	D-Pro	MeThr	-5.1	-	0.884	3.9	8.8	249
II	**4**	D-Pro	MePhe	D-Pro	MePhe	-5.3	46	<0.004	>6.4	12.9	179
III	**5**	Sar	Pro	Sar	Pro	-6.7	-	>1.000	3.3	9.0	202
IV	**6**	D-MeAla	MePhe	D-MeAla	MePhe	-5.9	42**	0.004	>6.3	13.2	201

*The value above 100% is caused by extrapolation uncertainty.[26]
**Data not reported previously.

To rationalize the membrane permeability of the six cyclic decapeptides, we follow a similar workflow as for CsA[31] and CsE[32] and compared models of the conformational kinetics of the cyclic decapeptides in water and chloroform. In the previous publications, we had used Markov state models (MSM)[38–41] to obtain the kinetic models. In this work, we use a more recent variant of this technique: core-set Markov models (CSMM).[42–45] In both techniques, the conformational kinetics is modeled as transitions between disjoint conformational states. In MSMs, these states are called microstates and cover the entire conformational space, thus yielding sharp state-to-state boundaries. Spurious recrossing events across these boundaries can distort the estimate of the microstate-to-microstate transition count, which decreases the approximation quality of the overall kinetic model. CSMMs circumvent this problem by defining states, so-called core sets, which are centered at the minima of the potential energy landscape and whose boundaries are not in direct contact with each other. The conformational space between the core sets is called intermediate region. Any point in this region has a fuzzy assignment to the core sets, which is determined by the committor functions of the core sets. Effective transition counts between these core sets are obtained from transition-path theory.[43] Suitable core sets can be identified by the common-nearest-neighbor (CNN) cluster algorithm.[45] In both approaches, MSM and CSMM, the kinetic model is simplified by grouping microstates into metastable sets. The identification of the metastable sets is based on an analysis of the transition matrix eigenvectors, and the associated eigenvalues can be converted into relaxation timescales between different metastable sets. MSMs and CSMMs have been applied successfully to study a variety of biological systems.[46–50]

2 Methods

Experimental details of the PAMPA and logD measurements, as well as the NMR NOESY measurements and the procedure to determine the NMR solution structure are given in the Supporting Information.

2.1 Clustering Methods

In our previous studies of CsA[31] and CsE,[32] the Butina-type clustering algorithm[51] implemented in the `cluster` program of the GROMOS package of analysis programs[52] was used. The algorithm takes the RMSD matrix calculated between all frames in the trajectory and a user defined distance cutoff value as the input. The RMSD was calculated based on all backbone torsional angles. Structures within the cutoff value are considered to be geometric neighbors and members of the same cluster, and the cluster center is defined by the structure having the highest number of neighbors. The procedure

is then repeated for the remaining frames in the trajectory to identify other clusters. All frames in the trajectory were then assigned to the closest microstate based on the RMSD value, resulting in the Voronoi partitioning of the conformational space. Due to the large amount of the data from MD simulations of CsA and CsE, the clustering was performed on a subset of frames from the trajectories to limit computational cost. Thus, the cluster centers may not provide an accurate representation for the whole ensemble, and assignment of the frames to the closest cluster center can lead to discretization errors. Additionally, as the whole conformational space was partitioned, also short-lived ("transition") conformations were assigned to microstates and consequently to the metastable sets. This in turn may lead to an underestimation of the ITS of interconversions between metastable sets.

In the present study, the CNN algorithm described in Ref. 53 was used to identify the core sets for a CSMM.[45] Input coordinates for the generation of the CSMMs were chosen to be the backbone ϕ and ψ torsional angles of all ten amino acids. The CNN algorithm defines clusters based on a measure of local data-point (neighbor) density. The input parameters for the CNN algorithm are the cutoffs for the nearest-neighbor distance and the number of nearest neighbors, which specify the density around a given structure. A frame is assigned to a cluster if it shares enough neighbors with one of the cluster members. Frames which do not share sufficient neighbors with any of the other frames in the data set are classified as noise points. This means that the actual CNN clusters do not fully partition the conformational space and their relative populations do not add to up to 1. In our CSMMs, the CNN clusters are used as the core sets and the noise points define the intermediate region.

2.2 Simulation Details

All MD simulations were performed using the GROMOS software package[54] and the GROMOS 54A7 force field[55]. Simulations were carried out under NpT conditions with periodic boundaries. Newton's equations of motion were integrated using the leap-frog scheme[56] with a time step of 2 fs. The simple-point-charge (SPC) water model,[57] the GROMOS chloroform[58] and DMSO[59] solvent models were used. Weak coupling to a temperature bath with a relaxation time of 0.1 ps was applied[60] to maintain constant temperature. The peptide and the solvent were coupled to separate temperature baths. The pressure was maintained close to 1.013 bar (1 atm) by weak coupling to a pressure bath with a relaxation time of 0.5 ps and using the experimental isothermal compressibility of the corresponding solvents. A twin-range cutoff scheme with 0.8 nm and 1.4 nm was used for the nonbonded interactions. A reaction-field force[61] with the relative dielectric permittivity of the corresponding solvent (61.0 for water,[62] 4.8 for chloroform,[58] and 46.7 for DMSO[63]) was applied to treat the electrostatic interactions beyond the long-range cutoff. Bond lengths were constrained to ideal values using the SHAKE algorithm[64] with a tolerance of 10^{-4} nm. The coordinates and energies were written out every 5 ps for analysis.

The *closed* conformations of the peptides determined by NMR spectroscopy in chloroform (Figure 2) were used as starting structures for the initial MD simulations of 100 ns length in water, chloroform and DMSO. In addition, a 100-ns Hamiltonian replica exchange (HRE)[65] simulation in water was performed to facilitate an opening of the *closed* conformation. For this, six replicas were used, where the force constant of backbone torsions was reduced as a function of the λ-value ranging from 0 (normal) to 1 (zero) (scaled by 1.0, 0.8, 0.6, 0.4, 0.2, and 0.0 for subsequent λ-points). Each simulation was preceded by 100 ps of thermalization and equilibration. The 100 most diverse conformations in terms of backbone atom-positional RMSD were selected for each of the 100-ns simulations using the `LazyPicker` function in the RDKit[66] cheminformatics toolkit. From the combined pool of 400 conformations, the final 100 most diverse were selected using the same procedure. They were subsequently used as the seed (starting) conformations for 100-ns simulations in water and chloroform, which resulted in total 10 μs of sampling in each solvent. If the implied timescales (ITS) in a CSMM constructed from the 10 μs data did not converge, additional resampling was performed starting from conformations selected

from the poorly sampled regions of phase space. The total amount of simulation data per peptide and solvent is reported in Table 2.

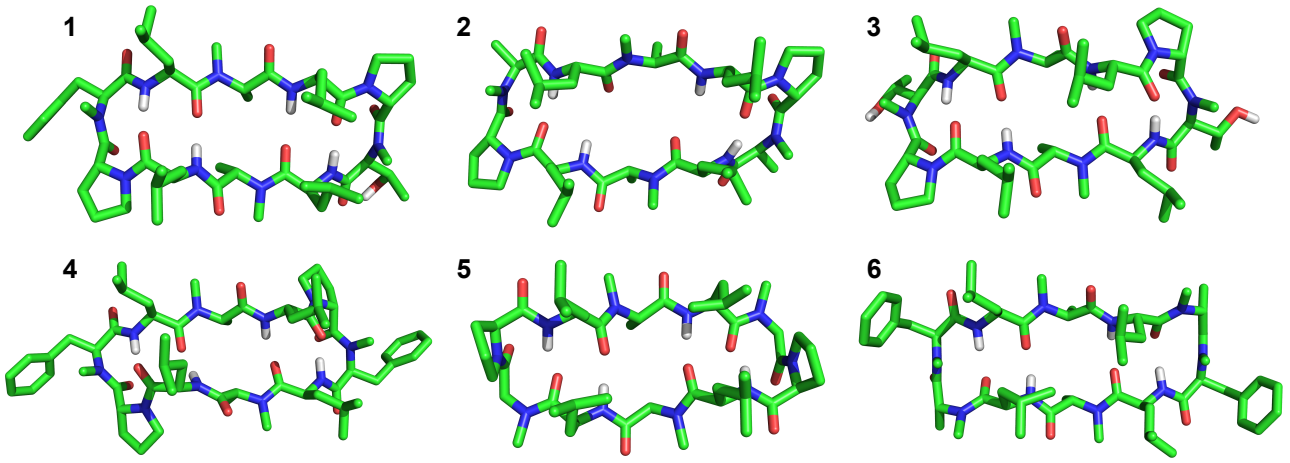

Figure 2: NMR structures of decapeptides **1** (A), **2** (B), **3** (C), **4** (D), **5** (E), **6** (F) in chloroform used as starting conformations for the MD simulations (generated using PyMOL[67]). The complete NMR bundles are shown in Figure S1 in the Supporting Information.

Table 2: Total amount of simulation data (in μs) and percentage of frames used to construct the CSMMs for each peptide in water and chloroform.

	Solvent			
Peptide ID	H_2O		$CHCl_3$	
	Time [μs]	Frame %	Time [μs]	Frame %
1	10.0	77	23.8	38
2	11.6	55	44.6	95
3	44.5	39	13.4	78
4	10.0	45	33.0	96
5	16.5	84	29.8	58
6	27.3	83	10.0	84

2.3 Markov Model Construction

Torsional angles were extracted from the trajectories using the GROMOS analysis program[52] `tser`. They were further transformed into collective coordinates by time-lagged independent component analysis (TICA).[68–70] A CNN algorithm[53] was used to identify the cores for a CSMM in the space of these collective variables. Effective transition probabilities between the core sets were obtained using the CSMM technique.[42–45] The PCCA+ analysis[71,72] implemented in the EMMA package[73] was further used to identify the metastable sets of the CSMMs. The ITS of the interconversion processes were calculated from the eigenvalues of the transition matrix $T(\tau)$ using the following relation,

$$t_i(\tau) = -\frac{\tau}{\ln \lambda_i(\tau)}\,. \tag{1}$$

The populations of metastable sets were derived from the steady-state distribution of the core sets. They correspond to the magnitude of the elements of the eigenvector, which corresponds to the largest

eigenvalue of the transition matrix $T(\tau)$. The population of a metastable state comprises the populations of the core sets, which are assigned to this metastable state, and additionally the fuzzy assignment of the intermediate space to these core sets. The parameters used for clustering, the number of TICs, and the lagtimes (τ) selected for CSMM construction are summarized in Table S1 in the Supporting Information.

Bootstrapping was performed after clustering of frames into microstates, following the procedure described in Ref. 74. Per each bootstrap iteration, the transition matrix for the CSMM was estimated by randomly picking n trajectories with replacement from the set of all trajectories, where n equals the total number of trajectories. 50 iterations were performed per peptide and solvent. From the 50 CSMMs, the average and standard deviation was estimated for the ITS and the steady-state populations. For the steady-state populations, only CSMMs were considered that had the same metastable sets as the CSMM of the whole data set. Figures S2 - S7 in the Supporting Information show the ITS of the whole datasets together with the average and standard deviation from bootstrapping.

2.4 Analysis

The initial 100-ns trajectories were analyzed by calculating the atom-positional backbone RMSD with respect to the NMR solution structure using the GROMOS analysis program `rmsd`. Metastable sets were analyzed with respect to transannular H-bonds using the program `hbond`. The criterion for a H-bond existence was a donor-acceptor distance lower than 0.25 nm and an angle larger than 135°.[75]. Experimentally determined NOE upper distance bounds were compared with corresponding distances in the simulated conformations using the programs `prep_noe`, `noe` and `post_noe`. Pseudo-atom distance corrections as given by Wüthrich *et al.*[76] were added.

2.5 Calculation of 3D Polar Surface Area

The 3D polar surface area (PSA) was calculated following the protocol of Rossi *et al.*[34] using the PyMOL.[67] The selection of polar atoms was done as described in Ref. 34: N, O, polar H, and other atoms with a partial charge above a cutoff of 0.5 e. Partial charges for the decapeptides were derived by fitting of the electrostatic potential (ESP) obtained from a QM calculation with the semiempirical method PM3 using the software tool Spartan'16 V2.0.7.[77] The radius of a water molecule (0.14 nm) was used as probe to obtain the van der Waals surface. The 3D-PSA was calculated for each solution structures in the NMR bundle (up to ten) and the minimum value is reported in Table 1.

3 Results and Discussion

3.1 Initial MD Simulations

Simulations of the six cyclic decapeptides in chloroform and water were performed for 100 ns starting from the *closed* structure derived from NMR measurements in chloroform (Figure 3). The *closed* conformation was stable for all peptides in chloroform, as indicated by a low RMSD (< 0.1 nm) with respect to the NMR structure. Only for peptide **5**, a higher RMSD of about 0.15 nm was observed due to a small backbone torsional change of the L-Pro residues, however, the four transannular H-bonds were maintained. Thus, all six decapeptides adopted a stable *closed* conformation with four transannular H-bonds in an apolar solvent, underlining that this is a mandatory but not sufficient condition for membrane permeability.

In water, differences in the conformational behavior between the six cyclic decapeptides could already be observed in the initial simulations. For the peptides with good membrane permeability, the *closed* conformation either remained stable over the course of 100 ns with low RMSD (< 0.10 nm) with

respect to the NMR structure (peptides **1** and **2**), or both opening and closing occurred with RMSD values up to 0.25 nm (peptides **3** and **4**). In contrast, peptide **6** opened (RMSD $>$ 0.20 nm) during the first 20 ns of simulation and remained open for the rest of the simulation. For peptide **5**, partial opening occurred already during equilibration (RMSD $>$ 0.15 nm) and the peptide interconverted fully to open conformations after a few nanoseconds of simulation (RMSD $>$ 0.25 nm). The initial simulations indicate that peptides characterized with low membrane permeability are more likely to adopt an open conformation. For peptides with high membrane permeability on the other hand, the *closed* conformation was relatively stable in a polar environment, or the interconversion between open and *closed* conformations occurred on a short timescale. These results confirm the anticipated role of D-Pro in the β-turns. Peptides **1** and **2** as well as **3** and **4** with a rigidifying D-Pro in position 4 and 9 showed generally a smaller conformational variation than peptide **5** with the more flexible glycine (Sar) in these positions.

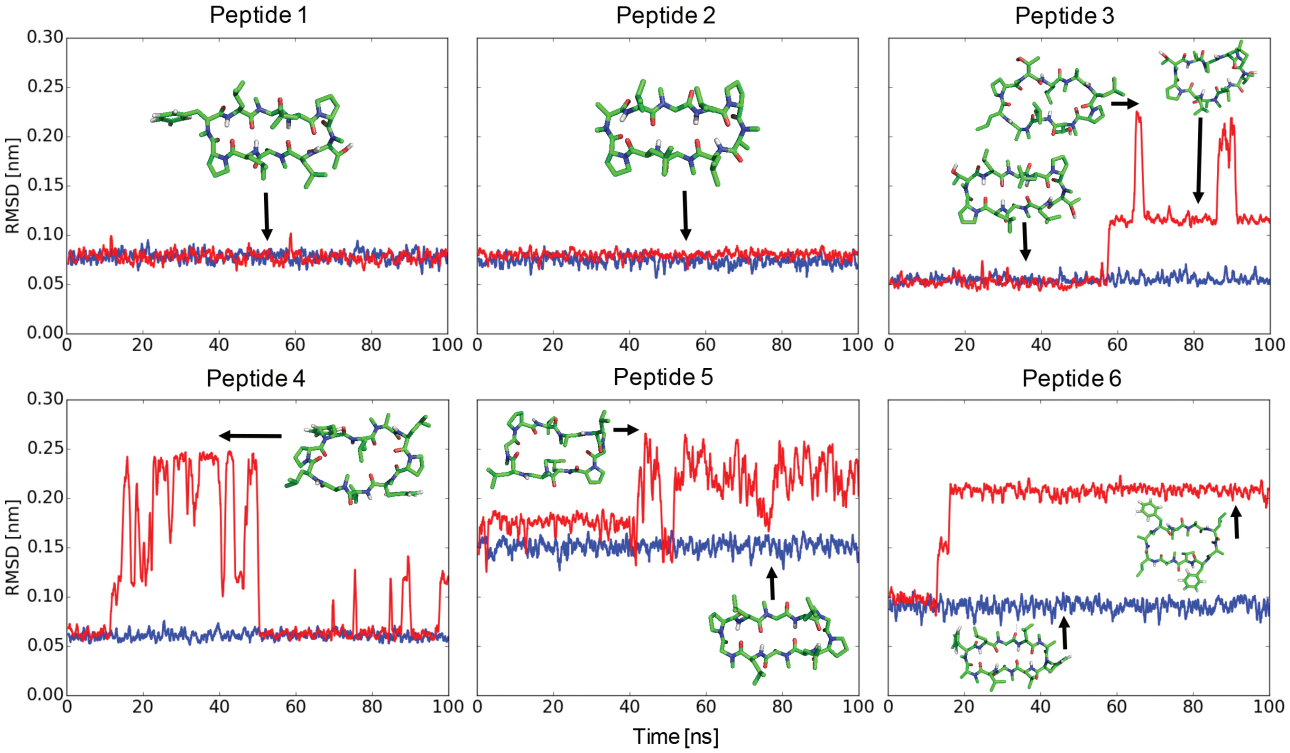

Figure 3: Backbone atom-positional RMSD with respect to the *closed* NMR structure for the 100 ns initial MD simulations (running average over 100 frames). Red lines correspond to the simulations in water, blue lines correspond to the simulations in chloroform. Representative snapshots are shown for selected RMSD values.

3.2 Kinetic Models

In the following, the CSMMs in chloroform and water are discussed for the six cylic decapeptides, divided into permeable (class I and II) and poorly permeable (class III and IV) peptides.

3.2.1 Class I and II

Although the four peptides are all considered "permeable" (PAMPA log $P_e \geq$ -5.3), the CSMMs show considerable differences in both environments. The CSMM of peptide **1** in water is composed of five metastable sets W1-W5 (Figure 4A,B). The slowest interconversion process is >150 ns and separates the

metastable sets W2 with transannular H-bonds and W5 from the open metastable sets W1, W3, W4. The steady-state population of metastable set W5, which contains the *closed* conformation, is 75±3%, making it the most populated set. The conformations in metastable set W2 form also transannular H-bonds, although with a "register shift" in the backbone arrangement (Figure 4B). Interconversion between metastable sets W2 and W5 occurs on a timescale of >120 ns. The CSMM of peptide **1** in chloroform is composed of only two metastable sets, separated on a timescale of 350 ns (Figure 4C,D). Metastable set C1 contains the *closed* conformation and is highly populated (86±2%). The remaining metastable set C2 contains mainly open conformations.

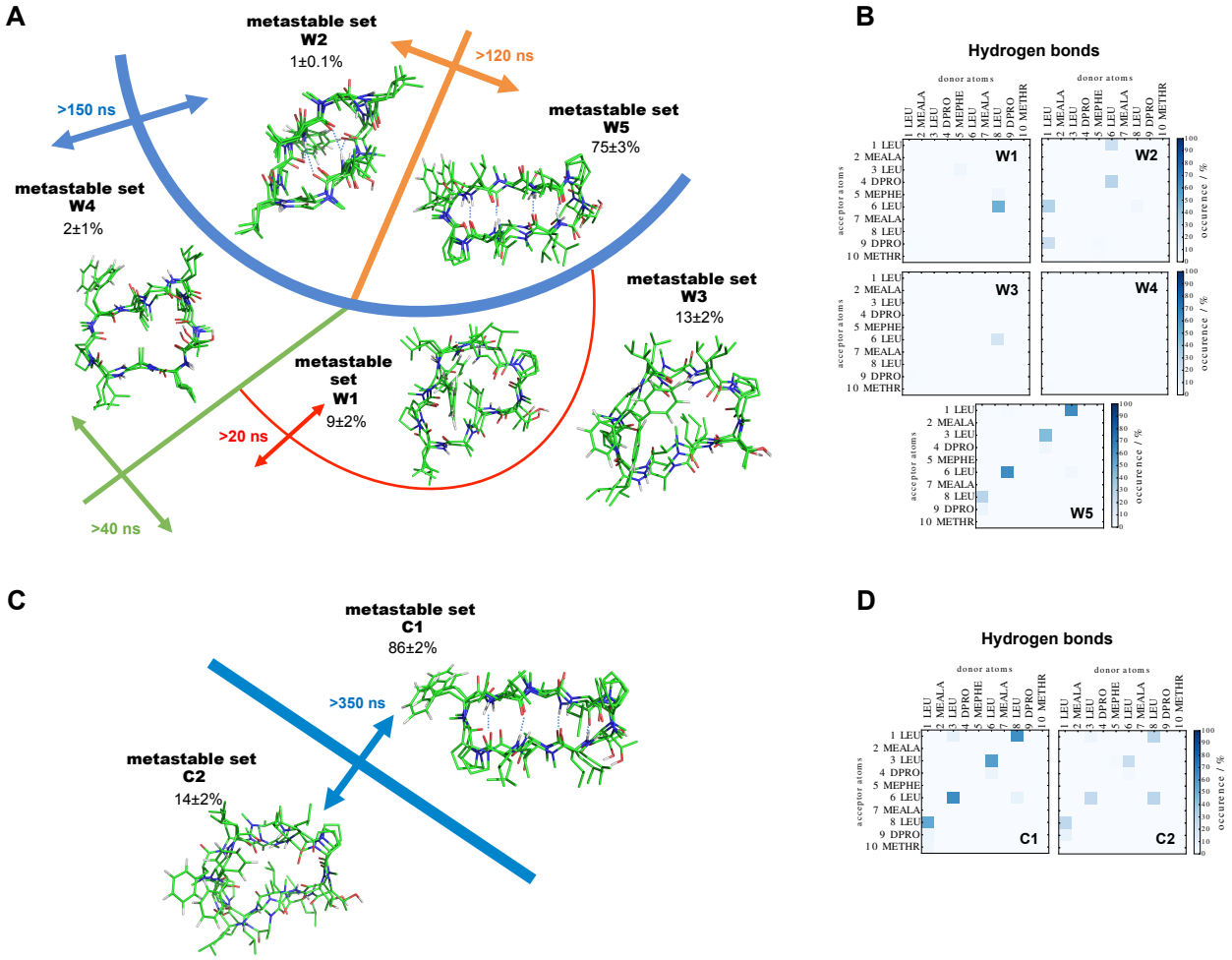

Figure 4: Metastable sets of peptide **1** in water (A,B) and in chloroform (C,D). (A,C) Schematic representation of the CSMM. The lines separating metastable sets correspond to interconversion processes occurring in the system. Three representative structures were randomly picked from each metastable set. Note that the population of a metastable state comprises that of the assigned core sets as well as the fuzzy assignment of the intermediate space to these core sets. The standard deviations of the populations were estimated with bootstrapping. Patterns of backbone-backbone H-bonds observed in water (B) and chloroform (D).

The CSMM of peptide **2** in water is largely dominated by metastable set W3 (steady-state population 74±3%), which contains the *closed* conformation (Figure 5A,B). The slowest process occurs on a >100 ns timescale, and is associated with the interconversion from *closed* (W3) to open (W1 and W2) conformational states. Interestingly, unlike the other decapeptides, peptide **2** is described by a higher number of metastable sets in chloroform than in water. This observation is discussed in more detail

below when comparing to experimental T2 relaxation times. The CSMM in chloroform is composed of five metastable sets C1-C5, where the most dominant set C5 (steady-state population 91±1%) contains the *closed* conformation (Figure 5C,D). The slowest process of >550 ns separates metastable set C2, containing conformations with transannular H-bonds and a "register shift", from the remaining metastable sets. Metastable set C1 contains half-open conformations, and metastable sets C3 and C4 contain different open conformations. Note that the sampling required to get a convergence of the ITS of the CSMM in chloroform amounted to 44.6 μs. This indicates that the transitions occurring on the longest timescale are difficult to sample, and the timescale connected with this interconversion process may be shorter with even more resampling.

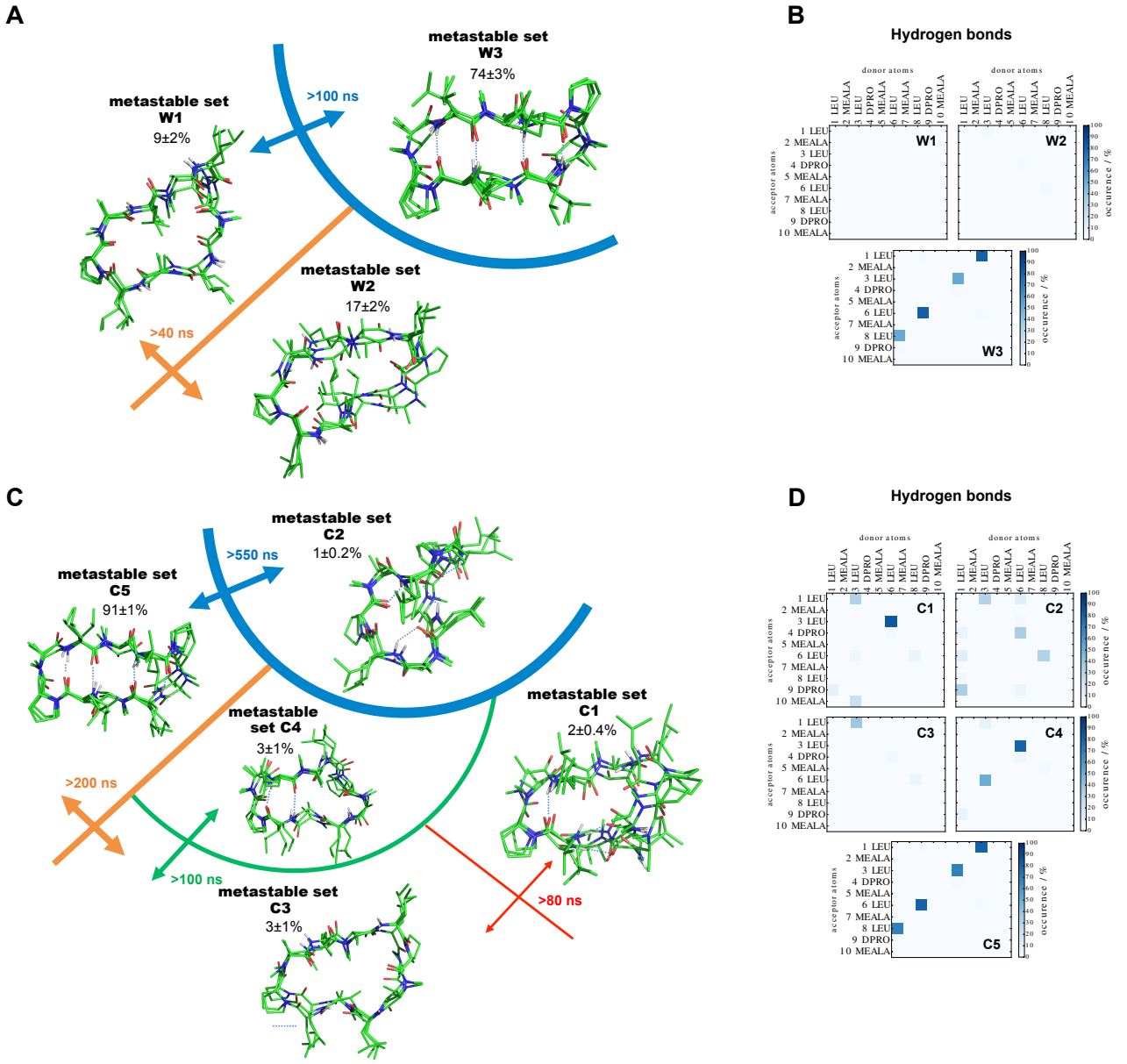

Figure 5: Metastable sets of peptide **2** in water (A,B) and in chloroform (C,D). (A,C) Schematic representation of the CSMM. The lines separating metastable sets correspond to interconversion processes occurring in the system. Three representative structures were randomly picked from each metastable set. Note that the population of a metastable state comprises that of the assigned core sets as well as the fuzzy assignment of the intermediate space to these core sets. The standard deviations of the populations were estimated with bootstrapping. Patterns of backbone-backbone H-bonds observed in water (B) and chloroform (D).

The CSMM of peptide **3** in water consists of four metastable sets W1-W4 (Figure 6A,B). Metastable set W1 contains conformations with transannular H-bonds and a "register shift", and is separated from the remaining metastable sets by the slowest interconversion process (>600 ns). Metastable sets W2 and W3 contain open conformations, whereas the most abundant set W4 (61±1%) contains the *closed* conformation. Interconversion between open and *closed* conformations occurs on a timescale >200 ns. Note that a long total simulation time (44.5 μs) was required to converge the ITS in water. The CSMM of peptide **3** in chloroform is composed of three metastable sets C1-C3 (Figure 6C,D). The

most populated metastable set C3 (steady-state population 90±1%) contains the *closed* conformation. The remaining metastable sets are populated with conformations forming different patterns of H-bonds compared to C3.

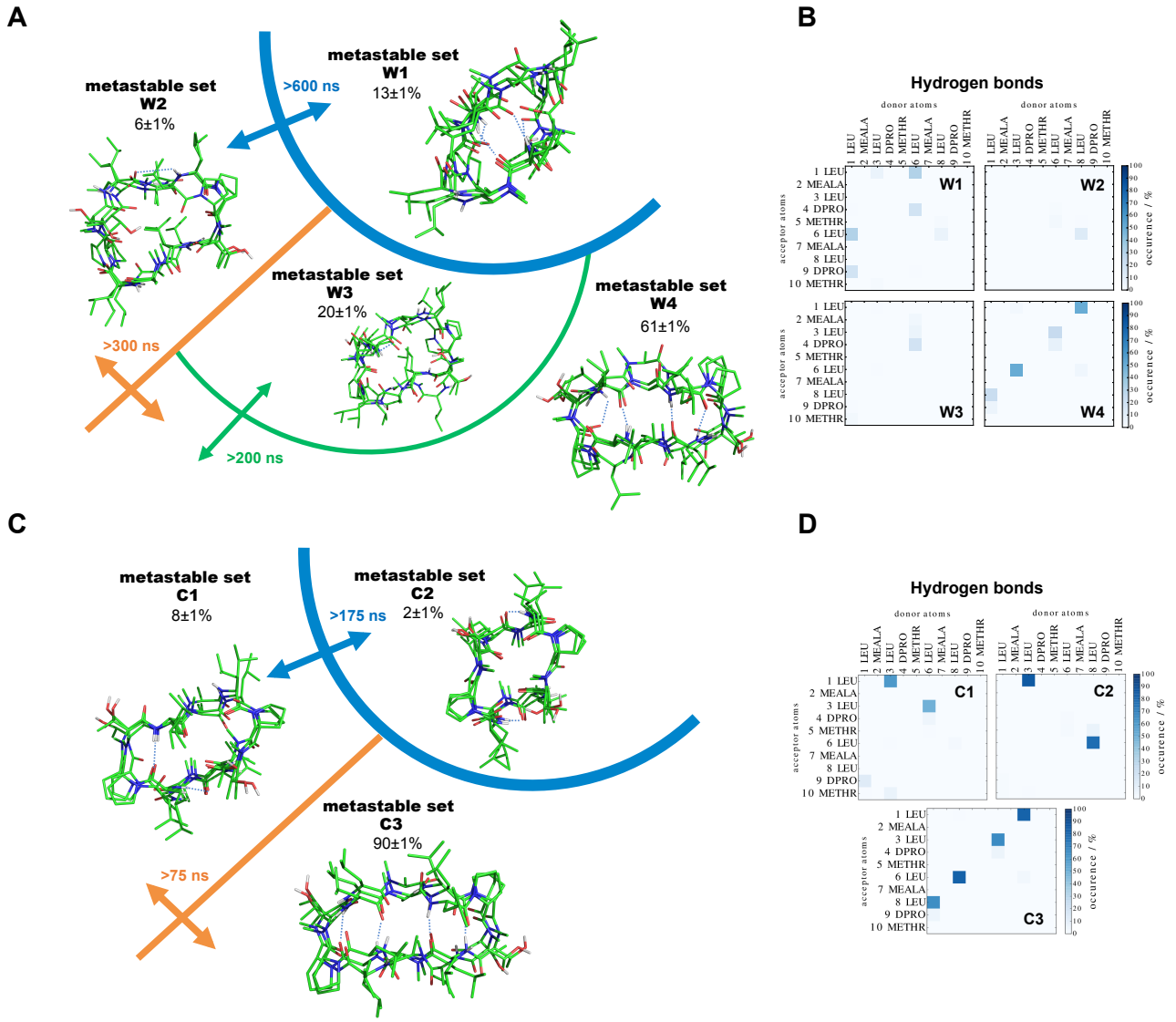

Figure 6: Metastable sets of peptide **3** in water (A,B) and in chloroform (C,D). (A,C) Schematic representation of the CSMM. The lines separating metastable sets correspond to interconversion processes occurring in the system. Three representative structures were randomly picked from each metastable set. Note that the population of a metastable state comprises that of the assigned core sets as well as the fuzzy assignment of the intermediate space to these core sets. The standard deviations of the populations were estimated with bootstrapping. Patterns of backbone-backbone H-bonds observed in water (B) and chloroform (D).

The CSMM of peptide **4** in water consists of six metastable sets W1-W6 (Figure 7A,B), with the slowest process (>175 ns) separating an ensemble of open conformations (W1, W2, W4, W5) from the metastable set W6 containing the *closed* conformation and another metastable set W3 containing open conformations. The transition from the *closed* to open conformations occurs on a timescale >100 ns. Metastable set W6 with the *closed* conformation constitutes the most populated metastable set (steady-state population 45±3%). The CSMM of peptide **4** in chloroform consists of three metastable

sets C1-C3 (Figure 7C,D). The slowest process occurs on a >900 ns timescale and separates metastable set C2 from the remaining sets. Metastable sets C1 and C3 are populated similarly (48±2% and 41±2%, respectively). Metastable set C1 contains the *closed* conformation. Note that achieving good convergence of the ITS of this model required significant amount of resampling (33.0 μs).

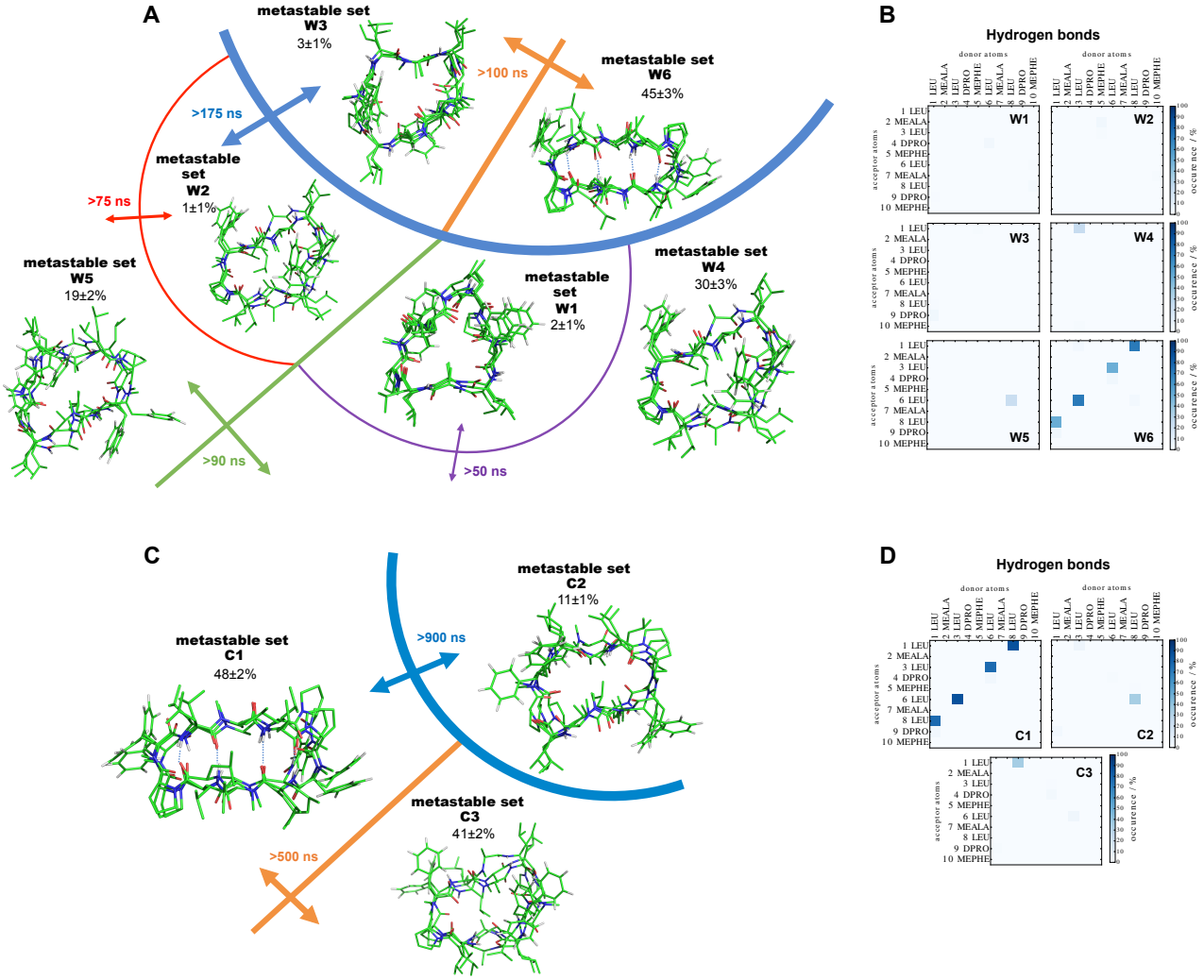

Figure 7: Metastable sets of peptide **4** in water (A,B) and in chloroform (C,D). (A,C) Schematic representation of the CSMM. The lines separating metastable sets correspond to interconversion processes occurring in the system. Three representative structures were randomly picked from each metastable set. Note that the population of a metastable state comprises that of the assigned core sets as well as the fuzzy assignment of the intermediate space to these core sets. The standard deviations of the populations were estimated with bootstrapping. Patterns of backbone-backbone H-bonds observed in water (B) and chloroform (D).

3.2.2 Class III and IV

The CSMM of peptide **5** in water consists of three metastable sets W1-W3 (Figure 8A,B). The slowest interconversion process separates W2 from the remaining metastable sets on a timescale of >120 ns. Metastable set W1 contains the *closed* conformation and is only populated to a low extent (steady-state population 16±2%). Metastable set W2 contains half-open conformations, with different H-bonding patterns compared to W1 (Figure 8B). The most populated state is W3 (steady-state population

81±2%), which contains open conformations. The transition between *closed* and open conformations occurs on a timescale of >40 ns. It has been noticed before,[25] that L-Pro in position 5 and 10 is not able to stabilize the β-turn conformation. This observation is supported by the CSMM in water, where the *closed* conformation is populated to a much smaller extent compared to the peptides with D-Pro in position 4 and 9 (peptides **1**, **2**, **3**, and **4**). In chloroform, the CSMM of peptide **5** divides the conformational space into two metastable sets C1 and C2 (Figure 8C,D), separated by a relatively short timescale of >80 ns. The most dominant set is C1 (steady-state population 90%) with the *closed* conformation (Figure 8D), while set C2 contains open conformations.

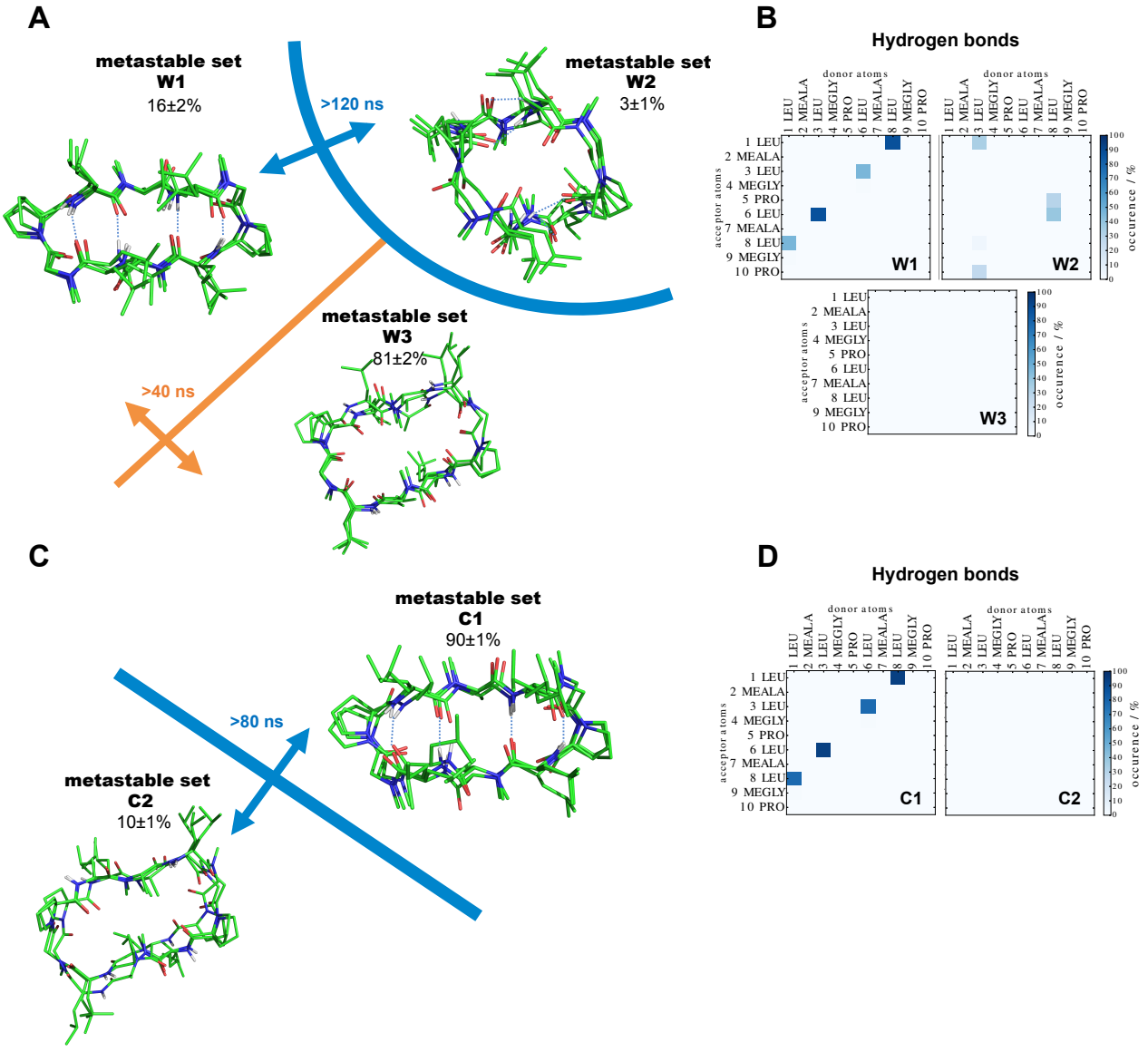

Figure 8: Metastable sets of peptide **5** in water (A,B) and in chloroform (C,D). (A,C) Schematic representation of the CSMM. The lines separating metastable sets correspond to interconversion processes occurring in the system. Three representative structures were randomly picked from each metastable set. Note that the population of a metastable state comprises that of the assigned core sets as well as the fuzzy assignment of the intermediate space to these core sets. The standard deviations of the populations were estimated with bootstrapping. Patterns of backbone-backbone H-bonds observed in water (B) and chloroform (D).

The CSMM of peptide **6** in water is composed of five metastable sets W1-W5 (Figure 9A,B). The slowest interconversion process is observed at a timescale of >400 ns. All metastable sets contain open conformations, no set with the *closed* conformation was observed. In chloroform, four metastable sets C1-C4 were found (Figure 9C,D). The slowest process occurs at a timescale >300 ns, separating metastable set C1 from remaining sets. Metastable sets C1-C3 contain open conformations, while metastable set C4 contains the *closed* conformation (steady-state population 46±4%).

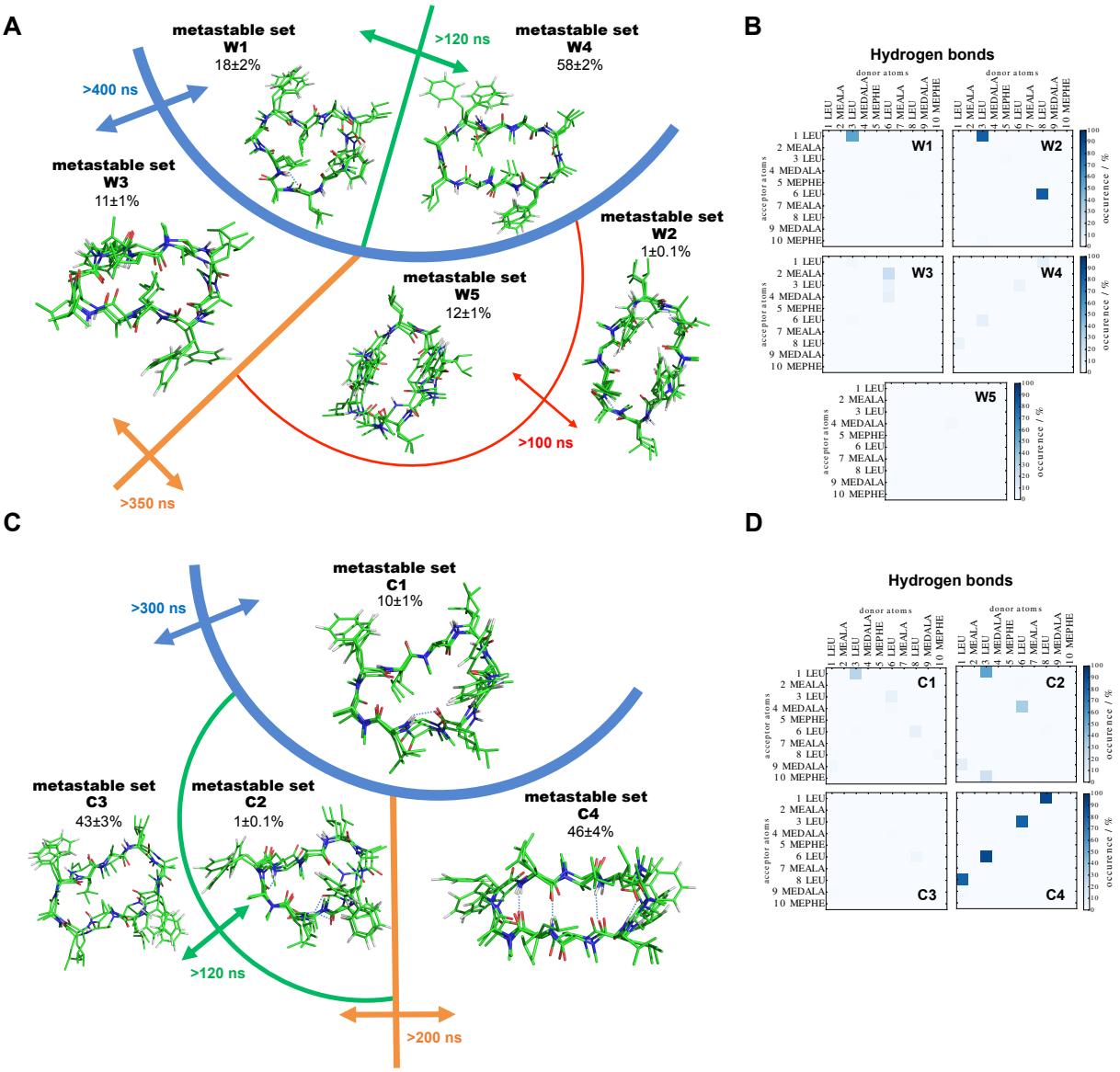

Figure 9: Metastable sets of peptide **6** in water (A,B) and in chloroform (C,D). (A,C) Schematic representation of the CSMM. The lines separating metastable sets correspond to interconversion processes occurring in the system. Three representative structures were randomly picked from each metastable set. Note that the population of a metastable state comprises that of the assigned core sets as well as the fuzzy assignment of the intermediate space to these core sets. The standard deviations of the populations were estimated with bootstrapping. Patterns of backbone-backbone H-bonds observed in water (B) and chloroform (D).

3.3 Interpretation and Comparison with Experiment

3.3.1 NOE Upper Distance Bounds

The complete simulated ensembles in chloroform were compared to NOE upper distance bounds (Figures S8–S13 in the Supporting Information). Overall, a good agreement with experimentally determined NOE bounds was observed, especially for NOE distances between protons of backbone atoms. Violations of NOE bounds larger than 0.1 nm were observed only between side chain–backbone protons or side chain–side chain protons. Interestingly, the number of NOE distances that could be determined varies substantially between the peptides. The highest number of NOE distances was obtained for peptide **5** (more than 190) and the smallest number for peptide **2** (fewer than 30). This can be interpreted as an indirect indication for the higher flexibility of the corresponding peptide in the apolar environment, resulting in a smaller number of cross peaks in the NOESY spectrum.

3.3.2 T2 Relaxation Times

Experimentally determined T2 relaxation time describes the decay of the magnetization perpendicular to the applied magnetic field caused by nonbonded interactions and conformational changes.[78] The faster these conformational changes are, the less they reduce the relaxation time. Thus, T2 relaxation times of C_α protons can provide information about the internal flexibility of the peptides. The T2 relaxation times for the six decapeptides in chloroform and DMSO are listed in Table 3. In general, the T2 relaxation times in a polar solvent such as DMSO are expected to be faster (i.e. higher) than in an apolar solvent such as chloroform. An extensive discussion of the T2 relaxation times of the decapeptides and their interpretation can be found in Ref. 25. Here, we focus on the comparison between experimental and simulation data.

For peptide **2**, however, higher T2 relaxation times in chloroform than in DMSO were observed, which is in line with the small number of NOE distances that could be determined in chloroform. This was an unusual observation that is difficult to explain with chemical intuition. It led to a hypothesis about the entropy contribution to permeability.[25,26] Higher mobility in chloroform compared to DMSO would suggest an entropic gain upon entering the membrane, resulting in a favorable contribution to the free energy difference. Of course, leaving the membrane would imply entropy loss for the same peptide, disfavoring this process. In the simulations, an indicator for flexibility can be the number of metastable sets in the CSMMs. The CSMM of peptide **2** in water consists of three metastable sets, while the conformational space in chloroform is partitioned into five metastable sets, suggesting also higher conformational diversity in the apolar environment. As the CSMM in DMSO may differ from the one in water (see e.g. for CsE in Ref. 32), we generated the CSMM in DMSO for peptide **2** (Figure S8 in the Supporting Information). Similarly to water, the CSMM in DMSO consists also of three metastable sets. The simulations thus agree qualitatively with the experimental observations. A quantitative comparison of the experimental T2 relaxation times with the MD simulations is unfortunately not possible because the individual trajectories were only 100 ns of length, i.e. much smaller than the time scales observed in the experiment. When the flexibility in the simulations was assessed by calculating the root-mean-square fluctuations (RMSF) of the ten C_α atoms, higher RMSF values in DMSO than in chloroform were observed for all peptides except a few C_α atoms of peptide **6** (Table S2 in the Supporting Information).

For peptide **1**, the experimental T2 relaxation times of some residues were also higher in chloroform than in DMSO (Table 3), but to a much smaller extent as for peptide **2** and not supported by a low number of cross peaks in the NOESY spectrum. A clear interpretation of the experimental data is therefore not possible.

Table 3: T_2 relaxation times for C_α protons in chloroform and DMSO taken from Refs. 25,26. Note that all peptides except **1** are symmetrical.

Class	Peptide ID	T2 relaxation time in chloroform/DMSO [ms] Position 1,6	2,7	3,8	4,9	5,10
I	**1**	106/48 (Leu) 84/58 (Leu)	38/33 (Ala) 37/33 (Ala)	23/23 (Leu) 20/27 (Leu)	65/16 (Pro) 65/16 (Pro)	10/13 (Phe) 48/50 (Thr)
I	**2**	31/21 (Leu)	35/17 (Ala)	25/17 (Leu)	204/18 (Pro)	42/17 (Ala)
I	**3**	25/29 (Leu)	33/24 (Ala)	20/37 (Leu)	27/45 (Pro)	33/50 (Thr)
II	**4**	11/13 (Leu)	37/26 (Ala)	16/15 (Leu)	18/20 (Pro)	25/25 (Phe)
III	**5**	-	-	-	-	-
IV	**6**	19/38 (Leu)	21/22 (Ala)	23/29 (Leu)	21/40 (Ala)	13/23 (Phe)

3.3.3 Permeability

The decapeptides in class I and II are characterized by a good membrane permeability as measured by the PAMPA assay (Table 1). As mentioned before, a *closed* conformation with β-strands and transannular H-bonds was observed in many membrane-permeable cyclic peptides. For the studied set of six cyclic decapeptides the population of the metastable set containing the *closed* conformation in water was found to show a good correlation with the experimental PAMPA log P_e values (Figure 10). These results suggest that the ability to adopt the *closed* conformation (as the *congruent* conformation) in water is essential for good membrane permeability. Interestingly, all peptides with good membrane permeability have D-Pro in position 4 and 9, which likely increases the stability of the *closed* conformation, in contrast to the poorly permeable peptides featuring D-alanine or glycine at those positions. The common backbone pattern alone, which facilitates formation of four transannular H-bonds in chloroform, does not ensure stabilization of the *closed* conformation in water.

For small molecules, a compound's lipophilicity has been found to generally correlate with its permeability,[79] although it has been noted in recent studies that the relationship can be more complicated.[80,81] The octanol/water partition coefficient logP (or distribution coefficient logD) is typically used as a measure for lipophilicity. Despite their size and flexibility, the calculated logP values (clogP[37]) and the experimental logD values show a high correlation for the six cyclic decapeptides studied. Nevertheless, no correlation between the clogP (or logD) values and the experimental PAMPA log P_e values was observed. The two peptides containing two phenylalanine residues (peptides **4** and **6**) have the highest values of clogP and logD (Table 1), but peptide **4** is permeable (class II) and peptide **6** not (class IV). Peptides with good membrane permeability (PAMPA log $P_e \geq$ -5.3) have clogP values ranging from 8.8 to 12.9 (Table 1). This indicates that simple lipophilicity measures such as clogP/logD are not sufficient to predict the permeability of larger cyclic peptides.

Recently, it was reported that the minimum 3D-PSA of compounds beyond the "rule of five"[79] correlates with passive permeability.[34] The minimum 3D-PSA value was typically observed for the "closed" conformation of the compounds. Using the same procedure, the minimum 3D-PSA in the NMR bundle in chloroform was calculated for the six decapeptides in this study. As can be seen from Table 1, no correlation between the 3D-PSA and the permeability exists for this set of peptides. Peptides **1** and **3** in class I give even the highest 3D-PSA values due to the threonine residues. As for lipophilicity, these findings indicate that the complete conformational behaviour has to be considered to understand permeability differences among structurally similar compounds.

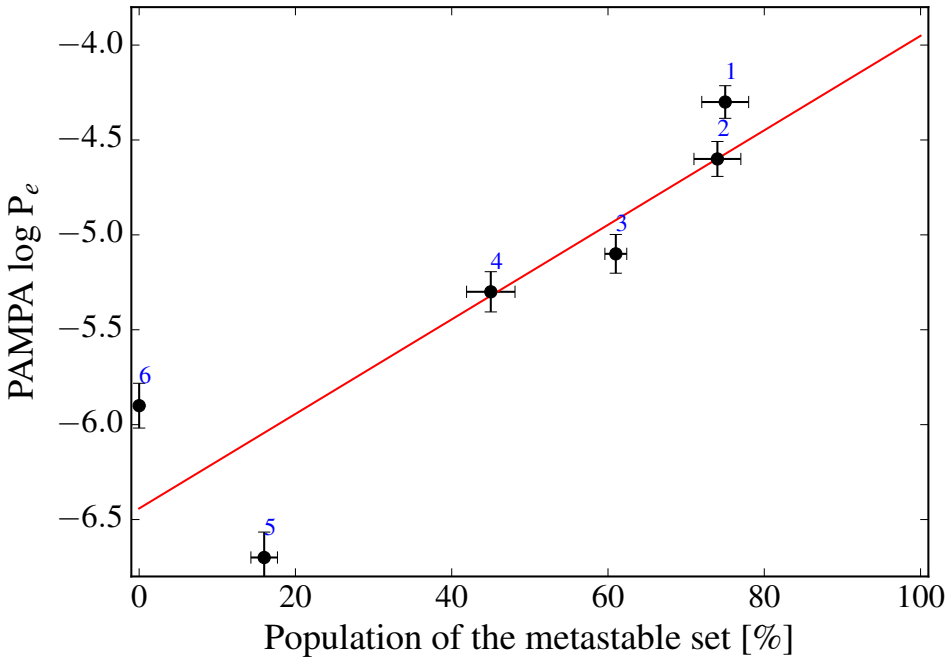

Figure 10: PAMPA log P_e as a function of the steady-state population of the metastable set containing the *closed* conformation in the CSMM in water. Linear regression (red line) gave a correlation coefficient $r^2 = 0.784$. The experimental uncertainty of log P_e is 0.2 %. The standard deviations of the populations were estimated with bootstrapping.

3.3.4 Solubility

The cyclic decapeptides in class I and III have a good aqueous solubility, whereas the peptides in class II and IV are poorly soluble in water (Table 1). From the simulation data, no correlation between the differences in the aqueous solubilities and the conformational behavior in water as described by the number and diversity of the metastable sets was found. Both peptides **4** (class II) and **6** (class IV) show a relatively high number of metastable sets in water with diverse conformations. Thus, the differences in solubility are likely dominated by the differences in the lattice energy in the solid state.

4 Conclusions

In the presented CSMMs of a series of six cyclic decapeptides, the metastable sets provide information about long-lived conformations, while ITS characterize the interconversion rates between them. It should be emphasized again at this point that the timescales observed in the simulations depend (among other aspects such as the force field and discretization strategy) on sufficient sampling of the relevant transitions. Despite the relatively long total simulation time (10 - 45 μs), the slowest ITS may not be fully converged for all peptides due to the high rigidity of the backbone for many of the studied peptides. With these limitations in mind, the interconversion rates between open and *closed* conformations of the six decapeptides in water can be considered similar, allowing a direct relative comparison of the steady-state populations of the metastable sets. Furthermore, through the relative comparison, effects from the chosen force field are expected to cancel.

Identification of *congruent* conformation(s), accessible on short enough timescales in both water

and chloroform, was the key to rationalize the difference in the membrane permeability of CsA and CsE described previously. In case of the four decapeptides with high membrane permeability (peptides **1**, **2**, **3**, and **4**), a metastable set containing the *closed* conformation with four transannular H-bonds was observed in both water and chloroform, thus presenting a *congruent* conformation. The *closed* conformation was even the most populated metastable set in water for these four peptides (steady-state population 43% - 75%). D-Pro in position 4 and 9 stabilizes the β-turns, thus favoring the *closed* conformation in water. This scaffold rigidification can reduce in the entropic penalty upon transition from a polar to an apolar environment. Inversely, a higher conformational flexibility in the apolar environment as observed for peptide **2** in the simulations, T2 relaxation and NOESY experiments can lead to an entropic gain upon entering the membrane. In case of the two peptides with low membrane permeability (peptides **5** and **6**), a different conformational behavior was observed. For peptide **6**, no *closed* or other *congruent* conformation was significantly populated in water. For peptide **5**, a *closed* conformation was populated in water, however, to a much smaller extent (steady-state population 16%) compared to the permeable peptides. Obviously, L-Pro in position 5 and 10 cannot sufficiently stabilize the β-turns in water and is overruled by the presence of flexible glycine (Sar) residues at position 4 and 9.

Overall, the population of the *closed* conformation in water was found to correlate with PAMPA permeability coefficients for this series of cyclic decapeptides. This indicates that for such large cyclic peptides, the investigation of the complete conformational behaviour is important for the rationalization of the membrane permeability. Lipophilicity (as represented by clogP or logD), the ability to adopt a *closed* conformation in an apolar environment, or the number of transannular H-bonds in the *closed* conformation failed to explain the permeability differences. Future work will further investigate the impact of side-chain variations at different positions for a possible development of cyclic peptides with desired permeability and solubility. Additionally, the development of efficient sampling approaches will be beneficial for making the workflow applicable for bigger-scale applications.

Acknowledgements

S. R. gratefully acknowledges financial support by the Swiss National Science Foundation (Grant Number 200021-178762) and by ETH Zürich (ETH-34 17-2). B. G. K gratefully acknowledges financial support by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 Scaling Cascades in Complex Systems (project B05).

References

[1] E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett. The Exploration of Macrocycles for Drug Discovery – an Underexploited Structural Class. *Nat. Rev. Drug Discov.*, 7, 608–624, **2008**.

[2] W. S. Horne. Peptide and Peptoid Foldamers in Medicinal Chemistry. *Expert. Opin. Drug Dis.*, 6, 1247–1262, **2011**.

[3] E. Marsault, M. L. Peterson. Macrocycles are Great Cycles: Applications, Opportunities and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.*, 54, 1961–2004, **2011**.

[4] J. Mallinson, I. Collins. Macrocycles in New Drug Discovery. *Future Med. Chem.*, 6, 1409–1438, **2012**.

[5] E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. P. Jr., D. Kozakov, S. Vajda, A. Whitty. How Proteins Bind Macrocycles. *Nat. Chem. Biol.*, 10, 723–731, **2014**.

[6] A. K. Yudin. Macrocycles: Lessons from the Distant Past, Recent Developments, and Future Directions. *Chem. Sci.*, 6, 30–49, **2015**.

[7] B. C. Doak, J. Zheng, D. Dobritzsch, J. Kihlberg. How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *J. Med. Chem.*, 59, 2312–2327, **2015**.

[8] Z. Qian, P. G. Dougherty, D. Pei. Targeting Intracellular Protein–Protein Interactions with Cell-Permeable Cyclic Peptides. *Curr. Opin. Chem. Biol.*, 38, 80–86, **2017**.

[9] D. K. Chalmers, G. R. Marshall. Pro-D-NMe-Amino Acid and D-Pro-NMe-Amino Acid: Simple, Efficient Reverse-Turn Constraints. *J. Am. Chem. Soc.*, 117, 5927–5937, **1995**.

[10] Y. Takeuchi, G. R. Marshall. Conformational Analysis of Reverse-Turn Constraints by N-Methylation and N-Hydroxylation of Amide Bonds in Peptides and Non-Peptide Mimetics. *J. Am. Chem. Soc.*, 120, 5363–5372, **1998**.

[11] K. Möhle, H.-J. Hofmann. Secondary Structure Formation in N-substituted Peptides. *J. Peptide Res.*, 51, 19–28, **1998**.

[12] A. C. Gibbs, T. C. Bjorndahl, R. S. Hodges, D. S. Wishart. Probing the Structural Determinants of Type II' β-Turn Formation in Peptides and Proteins. *J. Am. Chem. Soc.*, 124, 1203–1213, **2002**.

[13] N. Tsomaia. Peptide Therapeutics: Targeting the Undruggable Space. *Eur. J. Med. Chem.*, 94, 459–470, **2015**.

[14] M. R. Naylor, A. T. Bockus, M. J. Blanco, R. S. Lokey. Cyclic Peptide Natural Products Chart the Frontier of Oral Bioavailability in the Pursuit of Undruggable Targets. *Curr. Opin. Chem. Biol.*, 38, 141–147, **2017**.

[15] A. Zorzi, K. Deyle, C. Heinis. Cyclic Peptide Therapeutics: Past, Present and Future. *Curr. Opin. Chem. Biol.*, 38, 24–29, **2017**.

[16] C. Adessi, C. Soto. Converting a Peptide into a Drug: Strategies to Improve Stability and Bioavailability. *Curr. Med. Chem.*, 9, 963–978, **2002**.

[17] S. L. Schreiber, G. R. Crabtree. The Mechanism of Action of Cyclosporin A and FK506. *Immunol. Today*, 13(4), 136–142, **1992**.

[18] K. Kajitani, M. Fujihashi, Y. Kobayashi, S. Shimizu, Y. Tsujimoto, K. Miki. Crystal Structure of Human Cyclophilin D in Complex with Cyclosporin A. *Proteins*, 7, 1635–1639, **2008**.

[19] T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey, M. P. Jacobson. Conformational Flexibility, Internal Hydrogen Bonding, and Passive Membrane Permeability: Successful in Silico Prediction of the Relative Permeabilities of Cyclic Peptides. *J. Am. Chem. Soc.*, 128, 14073–14080, **2006**.

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

[21] P. Matsson, J. Kihlberg. How Big Is Too Big for Cell Permeability? *J. Med. Chem.*, 60, 1662–1664, **2017**.

[22] E. Biron, J. Chatterjee, O. Ovadia, D. Langenegger, J. Brueggen, D. Hoyer, H. A. Schmid, R. Jelinek, C. Gilon, A. Hoffman, *et al.*. Improving Oral Bioavailability of Peptides by Multiple N-Methylation: Somatostatin Analogues. *Angew. Chem. Int. Ed.*, 47, 2595–2599, **2008**.

[23] T. R. White, C. M. Renzelman, A. C. Rand, T. Rezai, C. M. McEwen, V. M. Gelev, R. A. Turner, R. G. Linington, S. S. F. Leung, A. S. Kalgutkar, *et al.*. On-Resin N-Methylation of Cyclic Peptides for Discovery of Orally Bioavailable Scaffolds. *Nat. Chem. Biol.*, 7, 810–817, **2011**.

[24] W. M. Hewitt, S. S. F. Leung, C. R. Pye, A. R. Ponkey, M. Bednarek, M. P. Jacobson, R. S. Lokey. Cell-Permeable Cyclic Peptides from Synthetic Libraries Inspired by Natural Products. *J. Am. Chem. Soc.*, 137, 715–721, **2015**.

[25] M. Fouché, M. Schäfer, J. Berghausen, S. Desrayaud, M. Blatter, P. Piéchon, I. Dix, A. M. Garcia, H.-J. Roth. Design and Development of a Cyclic Decapeptide Scaffold with Suitable Properties for Bioavailability and Oral Exposure. *ChemMedChem.*, 11, 1048–1059, **2016**.

[26] M. Fouché, M. Schäfer, M. Blatter, J. Berghausen, S. Desrayaud, H.-J. Roth. Pharmacokinetic Studies around the Mono- and Difunctionalization of a Bioavailable Cyclic Decapeptide Scaffold. *ChemMedChem.*, 11, 1060–1068, **2016**.

[27] E. G. Hutchinson, J. M. Thornton. A Revised set of Potentials for β-turn Formation in Proteins. *Protein Sci.*, 3, 2207–2216, **1994**.

[28] S. Gangwar, S. D. Jois, T. J. Siahaan, D. G. V. Velde, V. J. Stella, R. T. Borchardt. The Effect of Conformation on Membrane Permeability of an Acyloxyalkoxy-linked Cyclic Prodrug of a Model Hexapeptide. *Pharm. Res.*, 13, 1657–1662, **1996**.

[29] F. W. Okumu, G. M. Pauletti, D. G. V. Velde, T. J. Siahaan, R. T. Borchardt. Effect of Restricted Conformational Flexibility on the Permeation of Model Hexapeptides Across Caco-2 Cell Monolayers. *Pharm. Res.*, 14, 170–175, **1997**.

[30] O. S. Gudmundsson, S. D. S. Jois, D. G. V. Velde, T. J. Siahaan, B. Wang, R. T. Borchardt. The Effect of Conformation on the Membrane Permeation of Coumarinic Acid- and Phenylpropionic Acid-Based Cyclic Prodrugs of Opioid Peptides. *J. Pept. Res.*, 53, 383–392, **1999**.

[31] J. Witek, B. G. Keller, M. Blatter, A. Meissner, T. Wagner, S. Riniker. Kinetic Models of Cyclosporin A in Polar and Apolar Environments Reveal Multiple Congruent Conformational States. *J. Chem. Inf. Model.*, 56, 1547–1562, **2016**.

[32] J. Witek, M. Mühlbauer, B. G. Keller, M. Blatter, A. Meissner, T. Wagner, S. Riniker. Interconversion Rates Between Conformational States as Rationale for the Membrane Permeability of Cyclosporines. *ChemPhysChem.*, 18, 3309–3314, **2017**.

[33] C. K. Wang, J. E. Swedberg, P. J. H. Harvey, Q. Kaas, D. J. Craik. Conformational Flexibility Is a Determinant of Permeability for Cyclosporin. *J. Phys. Chem. B*, 122, 2261–2276, **2018**.

[34] M. R. Sebastiano, B. C. Doak, M. Backlund, V. Poongavanam, B. Over, G. Ermondi, G. Caron, P. Matsson, J. Kihlberg. Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5. *J. Med. Chem.*, 2018, 4189–4202, **2018**.

[35] G. L. Amidon, H. Lennernäs, V. P. Shah, J. R. Crison. A Theoretical Basis for a Biopharmaceutic Drug Classification: The Correlation of *in Vitro* Drug Product Dissolution and *in Vivo* Bioavailability. *Pharma. Res.*, 12, 321–341, **1995**.

[36] M. Kansy, F. Senner, K. Gubernator. Physicochemical High Throughput Screening: Parallel Artificial Membrane Permeation Assay in the Description of Passive Absorption Processes. *J. Med. Chem.*, 41, 1007–1010, **1998**.

[37] A. Leo. Calculating log P_{oct} from Structures. *Chem. Rev.*, 93, 1281–1306, **1993**.

[38] C. Schütte, A. Fischer, W. Huisinga, P. Deuflhard. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.*, 151, 146–168, **1999**.

[39] W. C. Swope, J. W. Pitera, F. Suits. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *J. Phys. Chem. B*, 108, 6571–6581, **2004**.

[40] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, F. Noé. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.*, 134, 174105, **2011**.

[41] J. D. Chodera, F. Noé. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struc. Biol.*, 25, 135–144, **2014**.

[42] N. V. Buchete, G. Hummer. Coarse Master Equations for Peptide Folding Dynamics? *J. Phys. Chem. B*, 112, 6057–6069, **2008**.

[43] E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, R. Elber. On the Assumptions Underlying Milestoning. *J. Chem. Phys.*, 129, 174102, **2008**.

[44] C. Schütte, F. Noé, J. Lu, M. Sarich, E. Vanden-Eijnden. Markov State Models Based on Milestoning. *J. Chem. Phys.*, 134, 204105, **2011**.

[45] O. Lemke, B. G. Keller. Density-based Cluster Algorithms for the Identification of Core Sets. *J. Chem. Phys.*, 145, 164104, **2016**.

[46] R. D. Malmstrom, A. P. Kornev, S. S. Taylor, R. E. Amaro. Allostery Through the Computational Microscope: cAMP Activation of a Canonical Signalling Domain. *Nature Commun.*, 6, 7588, **2015**.

[47] K. M. Thayer, B. Lakhani, D. L. Beveridge. Molecular Dynamics Markov State Model of Protein Ligand Binding and Allostery in CRIB-PDZ: Conformational Selection and Induced Fit. *J. Phys. Chem. B*, 121, 5509–5514, **2017**.

[48] N. Plattner, S. Doerr, G. D. Fabritiis, F. Noé. Complete Protein–Protein Association Kinetics in Atomic Detail Revealed by Molecular Dynamics Simulations and Markov Modelling. *Nat. Chem.*, 9, 1005–1011, **2017**.

[49] R. P. Wiewiora, S. Chen, M. Luo, J. D. Chodera. Conformational Dynamics of Histone Methyltransferase SET8 Probed by Millisecond-Timescale Molecular Dynamics, Markov State Modeling and Biochemical Experiments. *Biophys. J.*, 114, 398a, **2018**.

[50] Y. Meng, C. Gao, D. K. Clawson, S. Atwell, M. Russell, M. Vieth, B. Roux. Predicting the Conformational Variability of Abl Tyrosine Kinase using Molecular Dynamics Simulations and Markov State Models. *J. Chem. Theory Comput.*, 14, 2721–2732, **2018**.

[51] X. Daura, W. F. van Gunsteren, A. E. Mark. Folding–Unfolding Thermodynamics of a β-Heptapeptide From Equilibrium Simulations. *Proteins*, 34, 269–280, **1999**.

[52] A. P. Eichenberger, J. R. Allison, J. Dolenc, D. P. Geerke, B. A. C. Horta, K. Meier, C. Oostenbrink, N. Schmid, D. Steiner, D. Wang, *et al.*. GROMOS++ Software for the Analysis of Biomolecular Simulation Trajectories. *J. Chem. Theory Comput.*, 7, 3379–3390, **2011**.

[53] B. Keller, X. Daura, W. F. van Gunsteren. Comparing Geometric and Kinetic Cluster Algorithms for Molecular Simulation Data. *J. Chem. Phys.*, 132, 074110, **2010**.

[54] N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger, W. F. van Gunsteren. Architecture, Implementation and Parallelisation of the GROMOS Software for Biomolecular Simulation. *Comput. Phys. Commun.*, 183, 890–903, **2012**.

[55] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, W. F. van Gunsteren. Definition and Testing of the GROMOS Force-Field Version 54A7 and 54B7. *Eur. Biophys. J.*, 40, 843–856, **2011**.

[56] R. W. Hockney. Potential Calculation and Some Applications. *Methods Comput. Phys.*, 9, 135–211, **1970**.

[57] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans. Interaction Models for Water in Relation to Protein Hydration. *Struct. Bond*, 14, 331–342, **1981**.

[58] I. G. Tironi, W. F. van Gunsteren. A Molecular Dynamics Simulation Study of Chloroform. *Mol. Phys.*, 83, 381–403, **1994**.

[59] H. Liu, F. Müller-Plathe, W. F. van Gunsteren. A Force Field for Liquid Dimethyl Sulfoxide and Physical Properties of Liquid Dimethyl Sulfoxide Calculated Using Molecular Dynamics Simulation. *J. Am. Chem. Soc.*, 117, 4363–4366, **1995**.

[60] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, 81, 3684–3690, **1984**.

[61] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.*, 102, 5451–5459, **1995**.

[62] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger. Comparison of Four Methods to Compute the Dielectric Permittivity of Liquids from Molecular Dynamics Simulations. *J. Chem. Phys.*, 115, 1125–1136, **2001**.

[63] J. A. Riddick, W. B. Bunger, T. K. Sakano. *Organic Solvents: Physical Properties and Methods of Purification*. John Wiley and Sons, New York, United States of America, **1986**.

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

[65] C. J. Woods, J. W. Essex, M. A. King. The Development of Replica-Exchange-Based Free-Energy Methods. *J. Phys. Chem. B*, 107, 13703–13710, **2003**.

[66] RDKit: Open-Source Cheminformatics Toolkit. URL `http://www.rdkit.org`.

[67] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8, **2015**. PyMOL The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC.

[68] L. Molgedey, H. G. Schuster. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Lett.*, 72, 3634–3637, **1994**.

[69] G. Pérez-Hernández, F. Paul, T. Giorgino, G. D. Fabritiis, F. Noé. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.*, 139, 015102, **2013**.

[70] C. R. Schwantes, V. S. Pande. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.*, 9, 2000–2009, **2013**.

[71] P. Deuflhard, M. Weber. Robust Perron Cluster Analysis in Conformation Dynamics. *Linear Algebra Appl.*, 398, 161–184, **2005**.

[72] M. Weber. *Improved Perron Cluster Analysis*. ZIB, Berlin, Germany, **2003**.

[73] M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, C. Schütte, F. Noé. EMMA: A Software Package for Markov Model Building and Analysis. *J. Chem. Theory Comput.*, 8, 2223–2238, **2012**.

[74] F. Vitalini, A. S. J. S. Mey, F. Noé, B. G. Keller. Dynamic Properties of Force Fields. *J. Chem. Phys.*, 142, 084101, **2015**.

[75] J. Lautz, H. Kessler, R. Kaptein, W. F. van Gunsteren. Molecular Dynamics Simulations of Cyclosporin A: The Crystal Structure and Dynamic Modelling of a Structure in Apolar Solution Based on NMR Data. *J. Comput. Aid. Mol. Des.*, 1, 219–241, **1987**.

[76] G. Wagner, A. Pardi, K. Wüthrich. Hydrogen Bond Length and Proton NMR Chemical Shifts in Proteins. *J. Am. Chem. Soc.*, 105, 5948–5949, **1983**.

[77] Spartan'16 V2.0.7, WaveFunction, Inc. Irvine, CA, **2018**.

[78] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton. In J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton (Eds.), *Protein NMR Spectroscopy*, 333 – 404. Academic Press, Burlington, 2nd edition ed., **2007**.

[79] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.*, 46, 3–25, **1997**.

[80] T. Potter, G. Ermondi, G. Newbury, G. Caron. Relating Caco-2 Permeability to Molecular Properties Using Block Relevance Analysis. *Med. Chem. Commun.*, 6, 626–629, **2015**.

[81] M. R. Naylor, A. M. Ly, M. J. Handford, D. P. Ramos, C. R. Pye, A. Furukawa, V. Klein, R. P. Noland, Q. Edmondson, A. C. Turmon, *et al.*. Lipophilic Permeability Efficiency (LPE) Reconciles the Opposing Roles of Lipophilicity in Membrane Permeability and Aqueous Solubility. *J. Med. Chem.*, DOI: 10.1021/acs.jmedchem.8b01259, **2018**.